



FIED 2023

Primer Foro Internacional de
Evaluación para el Desarrollo

Innovación en la generación de estadísticas: Métodos de estimación en áreas pequeñas para la desagregación de datos



Xavier Mancero

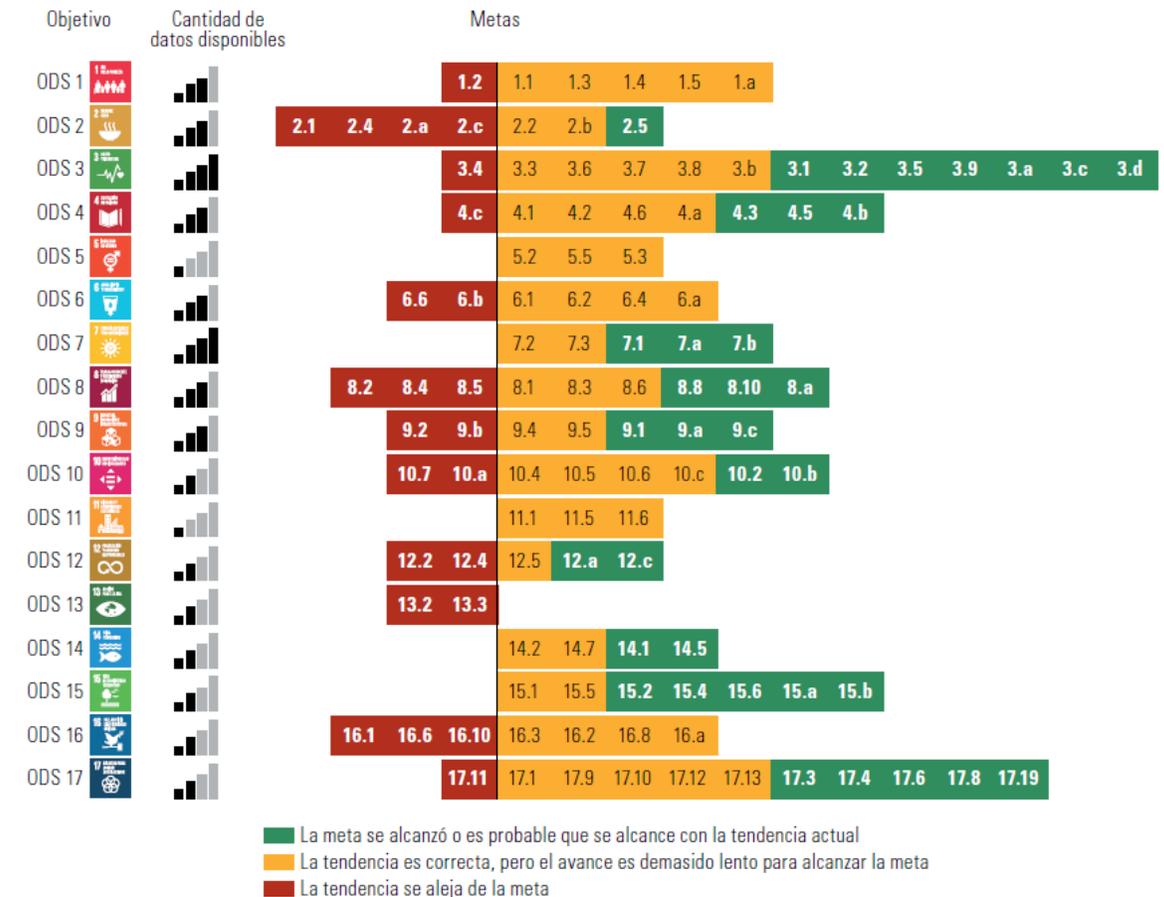
División de Estadísticas, CEPAL

19/oct/2023



Datos para políticas públicas

- Las metas de desarrollo establecidas por los Estados requieren de información adecuada para su seguimiento.
- Objetivos de Desarrollo Sostenible: “no dejar a nadie atrás”.
 - “Los indicadores de los ODS deben desagregarse, cuando corresponda, por ingreso, sexo, edad, raza, etnia, estatus migratorio, discapacidad y ubicación geográfica, u otras características” (E/CN.3/2020/2).
- Desafío:**
 - Estadísticas oportunas y confiables ...**
 - ... que además sean representativas de grupos y áreas de interés.**



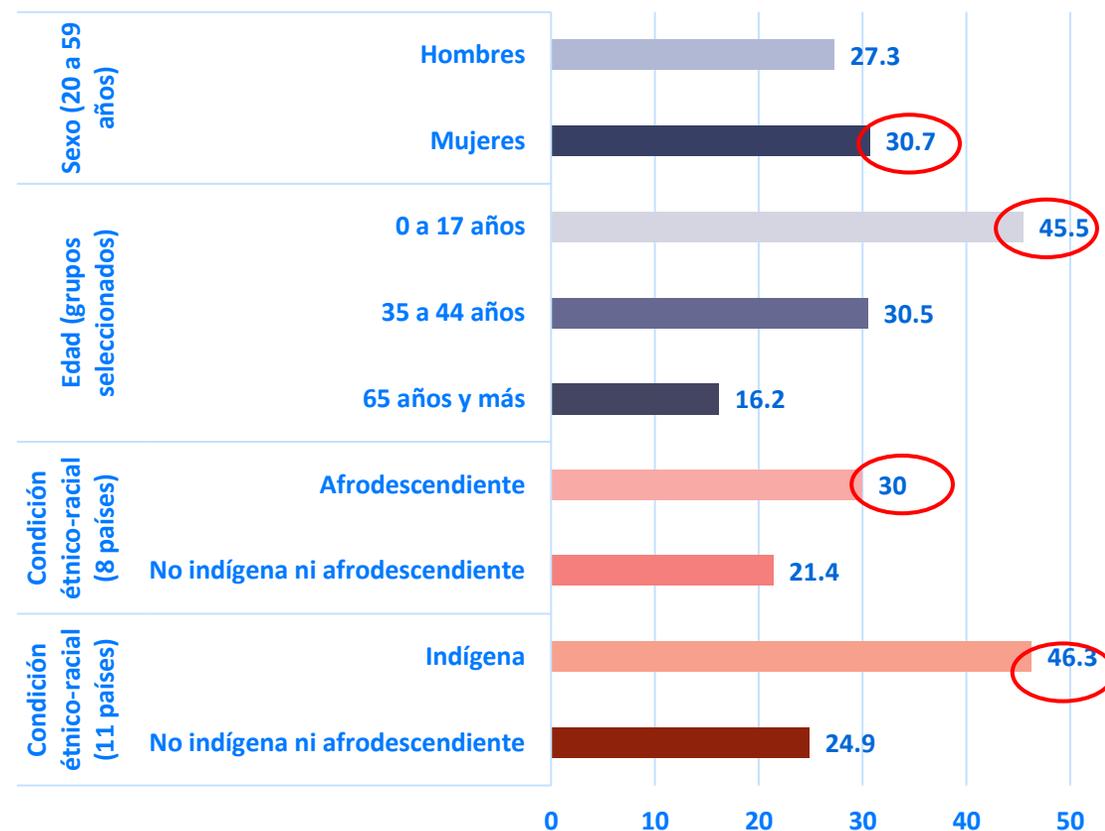
Fuente: Comisión Económica para América Latina y el Caribe (CEPAL), Una década de acción para un cambio de época (LC/FDS.5/3), Santiago, 2022.



Ejemplo: La probabilidad de estar en situación de pobreza no es igual para todos ...

- La tasa de pobreza de las **mujeres** de 20 a 59 años de edad es más alta que la de los hombres en todos los países de la región.
- Las **niñas, niños y adolescentes** enfrentan una mayor incidencia de la pobreza que el resto de grupos etarios.
- La pobreza es considerablemente más alta para la **población indígena o afrodescendiente**.

AMÉRICA LATINA: INCIDENCIA DE LA POBREZA SEGÚN DISTINTAS CARACTERÍSTICAS SOCIODEMOGRÁFICAS, 2021
(En porcentajes)



Fuentes de datos para las estadísticas sociales

- **Censos**

- Encuestas exhaustivas de población
- Baja cobertura temática
- Alta capacidad de desagregación
- Aprox. cada 10 años, alto costo

- **Encuestas por muestreo**

- Muestras representativas de población
- Alta cobertura temática
- Capacidad de desagregación limitada
- Frecuencias variadas, dependiendo del tema

- **Registros administrativos**

- Creciente prioridad para la generación de estadísticas
- Alta capacidad de desagregación y costo reducido
- Limitaciones de acceso
- Cobertura de población puede ser incompleta

- **Fuentes no tradicionales**

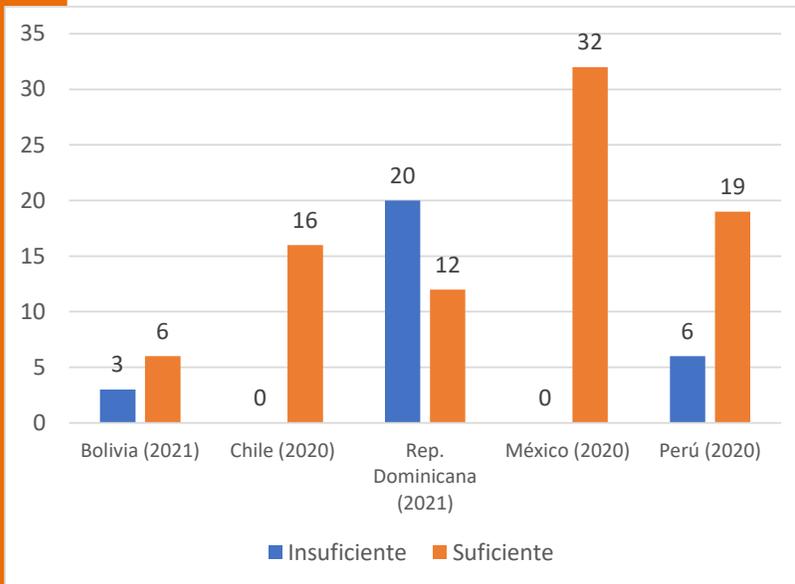
- Alto potencial, pero uso actual muy limitado
- Muy alta frecuencia
- Costo bajo
- Cobertura de población puede ser incompleta



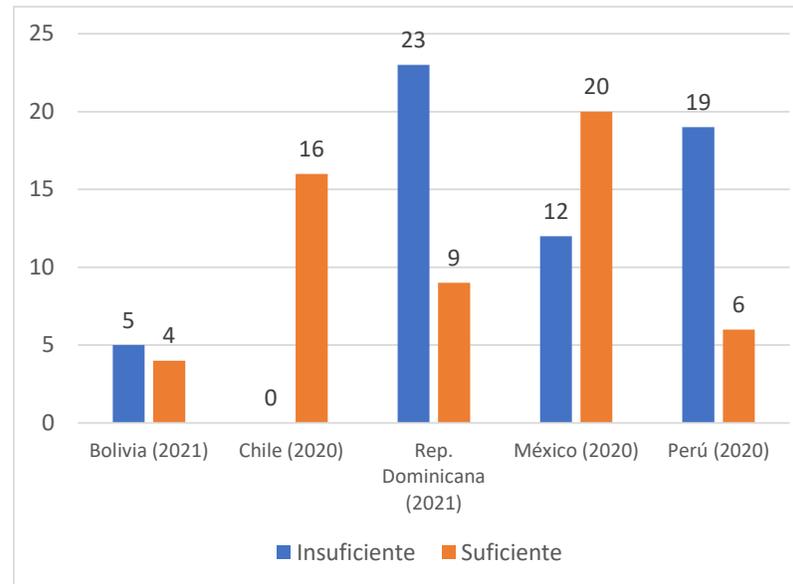
Las encuestas de hogares son la fuente de datos más utilizada para producir estadísticas sociales ...

Pero su capacidad de desagregación es limitada

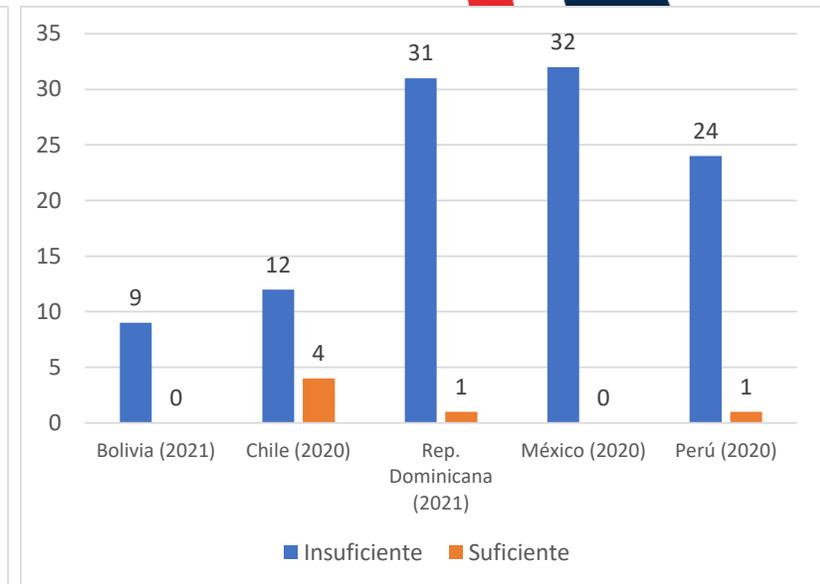
América Latina (5 países): Número de DAM¹ según calidad de las estimaciones de la tasa de desocupación



Tasa de desocupación, 15 y más años



Tasa de desocupación femenina, 15 y más años



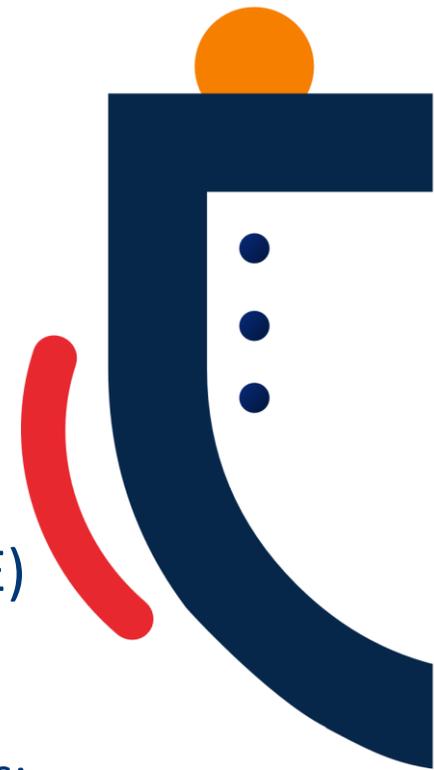
Tasa de desocupación femenina, 25 a 29 años

(*) DAM = División Administrativa Mayor / MAD = Major Administrative Division



¿Cómo podemos generar estadísticas con la desagregación requerida para las políticas públicas?

- Ampliar tamaño de la muestra de la encuesta
 - Puede tener un costo muy elevado.
 - Incluso encuestas muy grandes tienen muestra insuficiente para algunos grupos.
- Métodos de “estimación en áreas pequeñas” (Small Area Estimation o SAE)
 - Modelos estadísticos que “toman prestada” la capacidad de desagregación de fuentes auxiliares.
 - Ofrecen una posibilidad para ir más allá de lo que las encuestas permiten por sí solas:
 - Generar cifras para áreas donde la encuesta no alcanza la calidad requerida.
 - Reducir el error de estimación de las cifras.
 - Generar cifras para áreas no cubiertas por la encuesta.
 - “Área pequeña”:
 - Dominio para el cual no es posible obtener estimaciones directas confiables.
 - La subpoblación de interés puede ser un área geográfica o un subgrupo socioeconómico.



¿Cómo funciona SAE?

- Fuentes auxiliares:
 - Censos
 - Registros administrativos
 - Datos satelitales
 - Otras encuestas
 - Misma encuesta en otro período
 - Etc.
- Proceso
 - Identificar las variables auxiliares x disponibles en la encuesta y en las fuentes auxiliares.
 - Estimar los parámetros β de un modelo para predecir la variable de interés y con la encuesta ($y = f(x)$).
 - Aplicar los parámetros β estimados a la fuente auxiliar, para estimar la variable y para el nivel de desagregación deseado.
 - Medir el error de estimación y evaluar la confiabilidad de los resultados.



Características básicas de los métodos SAE

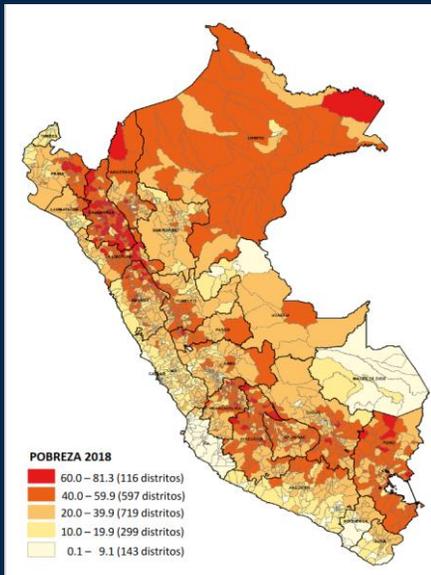
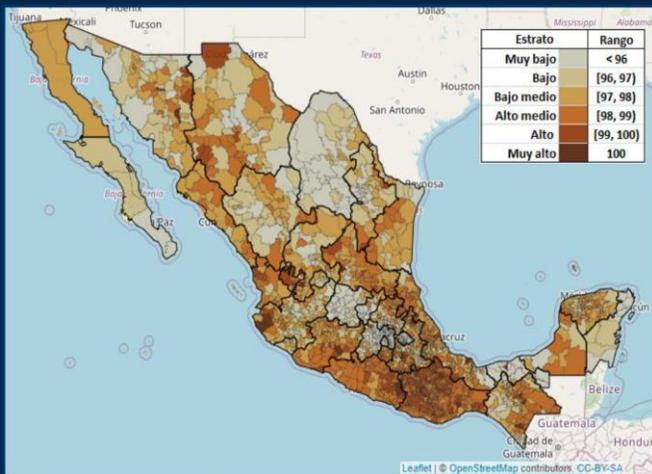
- Modelos de área
 - Predicen la variable de interés con variables auxiliares agregadas al nivel de área.
 - Fay & Herriot (1979): estimación de ingresos per cápita en áreas pequeñas de EEUU.
 - Usado en SAIPE (Small Area Income and Poverty Estimates) de la Oficina del Censo de EE.UU: pobreza para varios grupos de edad a nivel de estados, condados y distritos escolares.
- Modelos de unidad
 - Predicen la variable de interés con variables auxiliares para cada individuo (o sub-área) y luego agregan al nivel de área.
 - Battese, Harter & Fuller (1988): modelo de errores anidados con efectos aleatorios al nivel del área deseada.
 - Elbers, Lanjouw & Lanjouw (2003): método usado ampliamente por el Banco Mundial para mapas de pobreza.
 - Molina & Rao (2010): introduce innovación para reducir error cuadrático medio respecto a ELL.



Algunas experiencias regionales

- Varios países de la región han avanzado en la implementación de modelos SAE, aplicados a diversos indicadores de interés.
- Seminario regional ([Jun 2023, Sao Paulo, Brasil](#)):
 - Pobreza, acceso a internet por unidades de Federación (IBGE, Brasil)
 - Diversos indicadores sociodemográficos y percepción de inseguridad a nivel municipal (INEGI Mexico)
 - Pobreza, migración y otros indicadores a nivel municipal (DANE Colombia)
 - Pobreza a nivel distrital (INEI Perú)
 - Pobreza a nivel comunal (MDSF Chile)
 - Seguridad ciudadana a nivel comunal (INE Chile)



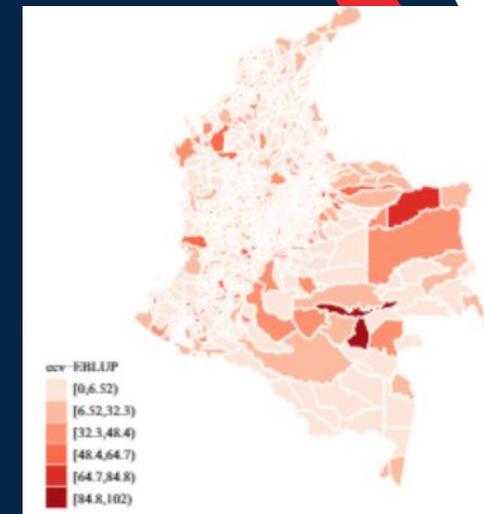
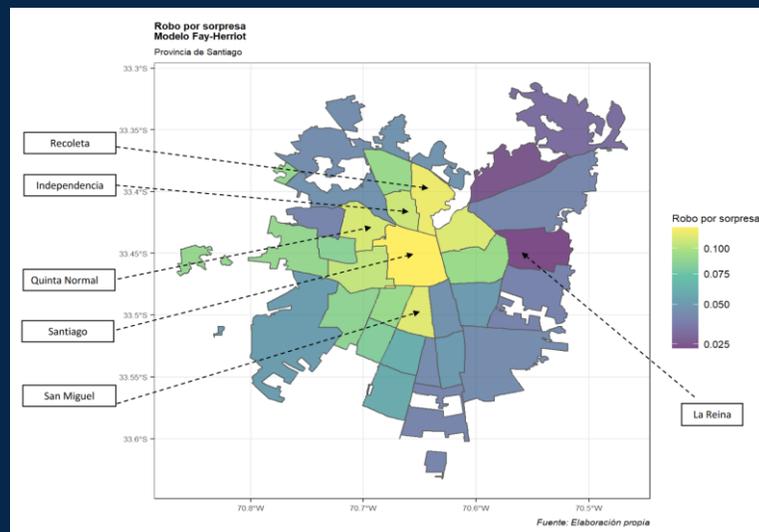
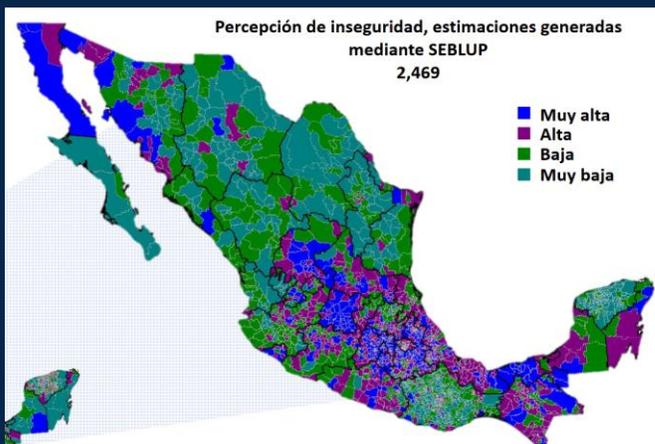
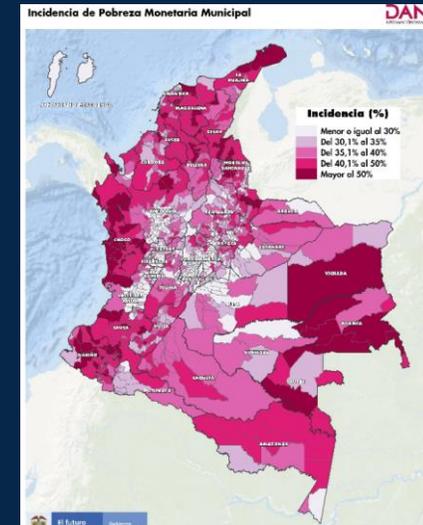


INEI

DOCUMENTO METODOLÓGICO

Aproximación al Mapa de Desnutrición Crónica Infantil a Nivel Provincial, 2019

(Niños menores de cinco años de edad)
Patrón de la Organización Mundial de la Salud - OMS



Experiencias internacionales y recursos de información: Toolkit on Using Small Area Estimation for SDGs (SAE4SDG)

UN Statistics Wiki Spaces

SAE4SDG

Created by UNSD Clarence Lio, last modified by Haoyi Chen on Jan 06, 2023

Inter-Agency and Expert Group on SDG Indicators

United Nations | DESA | Statistics Division

THE INTER-SECRETARIAT WORKING GROUP ON HOUSEHOLD SURVEYS

Welcome to the Toolkit on Using Small Area Estimation for SDGs!

In committing to the realization of the 2030 Agenda for Sustainable Development, Member States recognized that the dignity of the individuals is fundamental and that the Agenda's Goals and targets should be met for all nations and people and for all segments of society. Ensuring that these commitments are translated into effective action requires a precise understanding of the target populations and progress made in addressing their particular priorities.

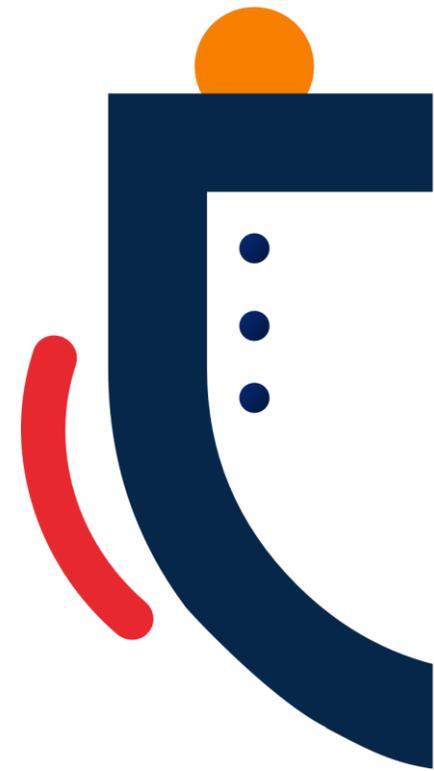
To properly measure this, statistics need to be presented for different population groups and geographical areas. The Sustainable Development Goal (SDG) indicator framework has included an overarching principle of data disaggregation: SDG indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics, in accordance with the Fundamental Principles of Official Statistics.

As sound statistical methods are vital to overcome this challenge, Small Area Estimation (SAE) constitutes an important topic in the way forward. It covers a variety of methods used to produce survey based estimates for geographical areas or domains of study in which the sample sizes are too



Experiencia con SAE en CEPAL

- Proyecto “Innovative approaches for examining inequality through integration of different data sources in Latin America and the Caribbean” (2021-2024, Development Account 13th Tranche)
- Colaboraciones con UNSD, UNFPA, FAO
- Cursos
 - Presenciales
 - eLearning: español + inglés (con UNSD y UNFPA)
- Asistencias técnicas
- Seminarios nacionales y regionales
- Publicaciones
- Estimaciones SAE de CEPAL



ISSN 1680-8789

SERIE

ESTUDIOS ESTADÍSTICOS 105

Modelos de unidad para la generación de mapas de pobreza a nivel subnacional

Andrés Gutiérrez
Xavier Mancero
Gabriel Nieto
Felipe Molina
Diego Lamas



Estimación desagregada para indicadores de planificación familiar en Colombia y Perú

Utilizando técnicas de estimación en áreas pequeñas para el caso de Perú y Colombia

Andrés Gutiérrez
Gabriel Nieto
Felipe Molina
Sabrina Jurán
Walter Mendez
Carlos Ramírez
Juliana Guerrero



Poverty mapping in Latin America: ECLAC experiences on small area estimation

Andrés Gutiérrez*, Xavier Mancero and Stalyn Guerrero
Social Statistics Unit, Statistics Division, Economic Commission for Latin America and the Caribbean, Santiago, Chile

Abstract. Poverty mapping is a valuable tool for governments and international organizations to identify the geographical areas and population groups that are most deprived or vulnerable. This approach can lead to designing and monitoring development policies more effectively. This article presents the recent experience of the Statistics Division of the UN Economic Commission for Latin America and the Caribbean (ECLAC) in using Small Area Estimation (SAE) methods to combine information from household surveys, censuses and satellite imagery to deliver poverty estimates at the provincial and municipal levels, which could not be attained using the household surveys alone.

Keywords: Poverty mapping, Pseudo-EBF, small area estimation, unit-level models

1. Introduction

Most Latin American and Caribbean countries regularly implement nationally representative household surveys to measure living conditions indicators, including poverty and income inequality. These surveys can generally be disaggregated geographically by urban and rural areas and the first-level administrative division (i.e. departments, provinces or regions). However, when it comes to obtaining a more disaggregated direct estimation of such indicators, results are subject to lower levels of accuracy and precision, which might be below the established quality criteria for their use. Small Area Estimation (SAE) techniques allow obtaining such disaggregated estimates while improving inference quality.

This article presents the recent experience of the Statistics Division of the United Nations Economic Commission for Latin America and the Caribbean (ECLAC) in applying SAE techniques to estimate geographically disaggregated poverty indicators based on household disposable income in seventeen (17) coun-

tries in Latin America. The results show a significant improvement in the precision of FGT-family indicators for geographical areas where the surveys do not attain adequate representativeness.

ECLAC regularly produces standardized national estimates of extreme poverty and poverty for Latin American countries, using a methodology that aims to achieve regional comparability. Even though countries in the region publish their official poverty statistics, the diversity of procedures and assumptions used in these estimates prevent direct comparisons, possibly leading to erroneous conclusions by not considering their methodological differences.

The ECLAC approach for measuring poverty classifies a person as poor when the per capita income of their household is lower than the poverty line, based on the cost of meeting their food needs and other basic non-food needs (1). The cost of food needs is estimated through the construction of basic food baskets, which provide the recommended amounts of energy and nutrients while reflecting the consumption habits of the population. The requirements come from current international recommendations to sustain a healthy life. Consumption habits are captured through household income and expenditure surveys and correspond to

*Corresponding author: Andrés Gutiérrez, Social Statistics Unit, Statistics Division, Economic Commission for Latin America and the Caribbean, Dag Hammarskjöld 3477, Vitacura, Región Metropolitana, Chile. E-mail: andres.gutierrez@ecp.un.org

ISSN 1798-8028

Temas estadísticos de la CEPAL

No. 4 | Diciembre de 2023

Estimaciones subnacionales de la pobreza para América Latina

Contenido

- 1. Introducción
- 2. Metodología
- 3. Resultados
- 4. Conclusiones
- 5. Anexos

Los mapas de pobreza a nivel subnacional son herramientas esenciales para que los gobiernos puedan identificar y evaluar las áreas más vulnerables y las poblaciones de los hogares más pobres, dirigirlas a los lugares y grupos de población en donde se necesitan mayor apoyo.

En el presente artículo se presenta la metodología de mapas de la pobreza basada en técnicas de estimación de áreas pequeñas que se aplicó en el estudio de la pobreza de la CEPAL. Esta metodología permite reducir la dispersión de los datos y obtener estimaciones de pobreza geográficamente desagregadas, con mayor precisión, que las estimaciones directas que provienen de los censos de hogares y encuestas representativas de hogares particulares.

El artículo de la División de Estadística de la Comisión Económica para América Latina y el Caribe (CEPAL) se presenta con resultados obtenidos por esta investigación para los países de Chile, Colombia y Perú.

ISSN 1680-8789

SERIE

ESTUDIOS ESTADÍSTICOS 97

Desagregación de datos en encuestas de hogares

Metodologías de estimación en áreas pequeñas

Isabel Molina

CEPAL
POR UN DESARROLLO SOSTENIBLE CON IGUALDAD

Desagregación de indicadores regionales en CEPAL

- Fuentes de información
 - Banco de Datos de Encuestas de Hogares (BADEHOG)
 - Variables armonizadas para comparabilidad regional de indicadores
- Banco de datos censales de CELADE
 - Procesamiento mediante librería REDATAM para R
- Imágenes satelitales
 - Acceso a Google Earth Engine desde R con librería rgee
 - Luces nocturnas, suelo cultivado, suelo urbano, índice de modificación humana, accesibilidad hospitales, etc.
- Proceso:
 - Estandarización y homologación de covariables
 - Actualización de conteos de población según marginales de la encuesta de hogares
 - Definición de los modelos para los indicadores de interés
 - Predicción en áreas pequeñas y estimación del ECM
 - Bootstrap o modelo bayesiano
 - Validación de supuestos del modelo y “benchmarking”
 - Generación de mapas



Presentación

Indicadores

Comparaciones entre grupos

Métodos y Recursos



Información estadística



Capas geográficas



Capas seleccionadas

Población en situación de pobreza según DAM

Porcentaje

Años: 2020

de 0,261 a 14,778	(57)
de 14,778 a 29,295	(91)
de 29,295 a 43,812	(83)
de 43,812 a 58,329	(28)
de 58,329 a 72,846	(13)

Cambiar dimensiones Cambiar clasificación Ver en dashboard Datos más recientes

Etiquetas Colores
 Incluir nombre
 Incluir valor
Personalizar esquema de colores

Población en situación de pobreza según DAM

Porcentaje

Años: 2014



Source: Esri, Maxar, Earthstar Geographics, and the GIS User Community



Presentación

Indicadores

Comparaciones entre grupos

Métodos y Recursos

Información estadística

Capas geográficas

Capas seleccionadas



Pobreza por Municipio y año

Porcentaje

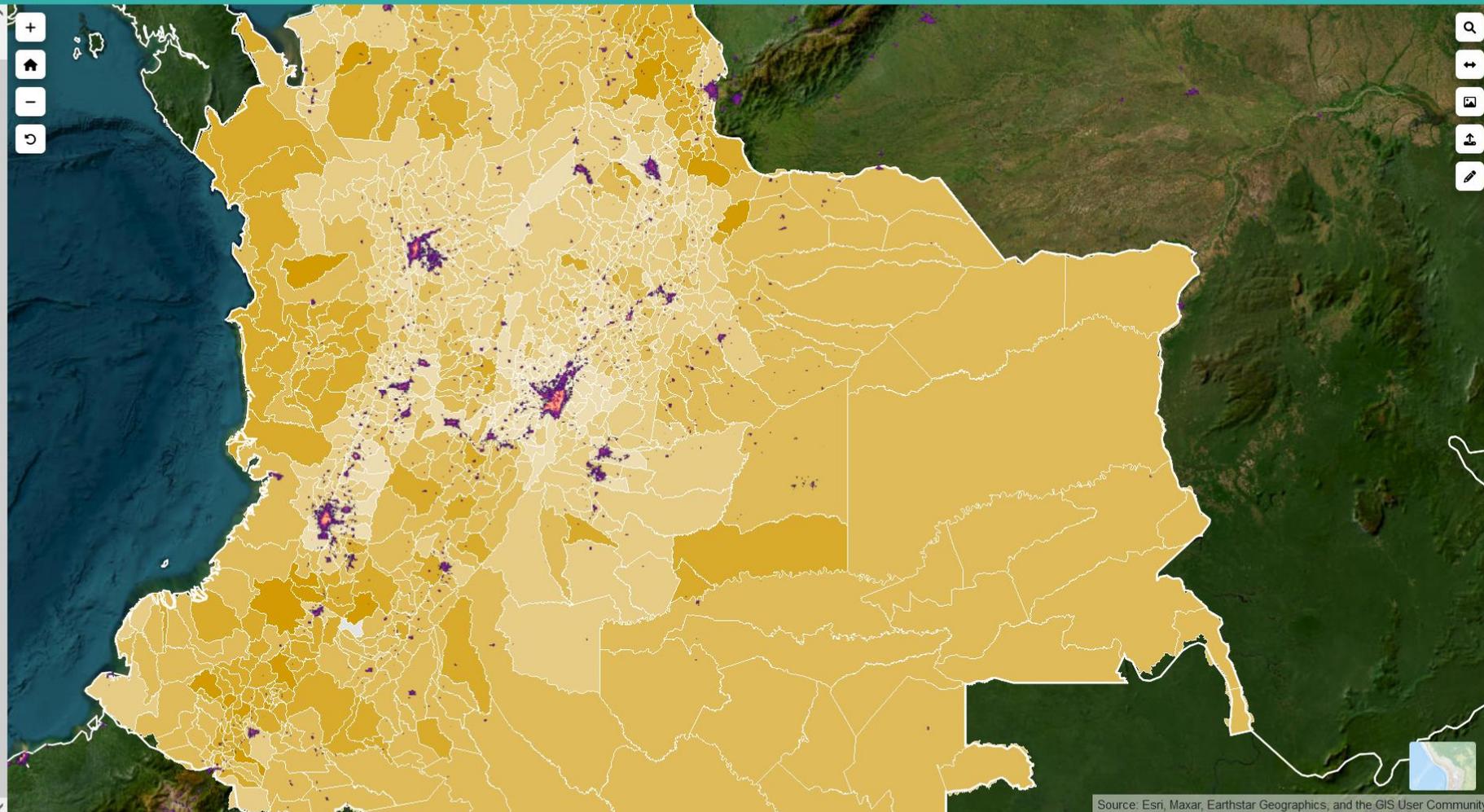
Años: 2018

de 3,592 a 21,577	(134)
de 21,577 a 39,562	(437)
de 39,562 a 57,547	(352)
de 57,547 a 75,532	(177)
de 75,532 a 93,517	(21)

Cambiar dimensiones Cambiar clasificación Ver en dashboard Datos más recientes

Etiquetas Incluir nombre Incluir valor

Colores Personalizar esquema de colores



Source: Esri, Maxar, Earthstar Geographics, and the GIS User Community

Presentación

Indicadores

Comparaciones entre grupos

Métodos y Recursos



Configuración Indicadores

Población en situación de pobreza según DAM, grupos de edad y etnia

Porcentaje

- de 3.7 a 18.7
- de 18.7 a 30.1
- de 30.1 a 43.3
- de 43.3 a 59.2
- de 59.2 a 95.5

País:

Ecuador

Filas:

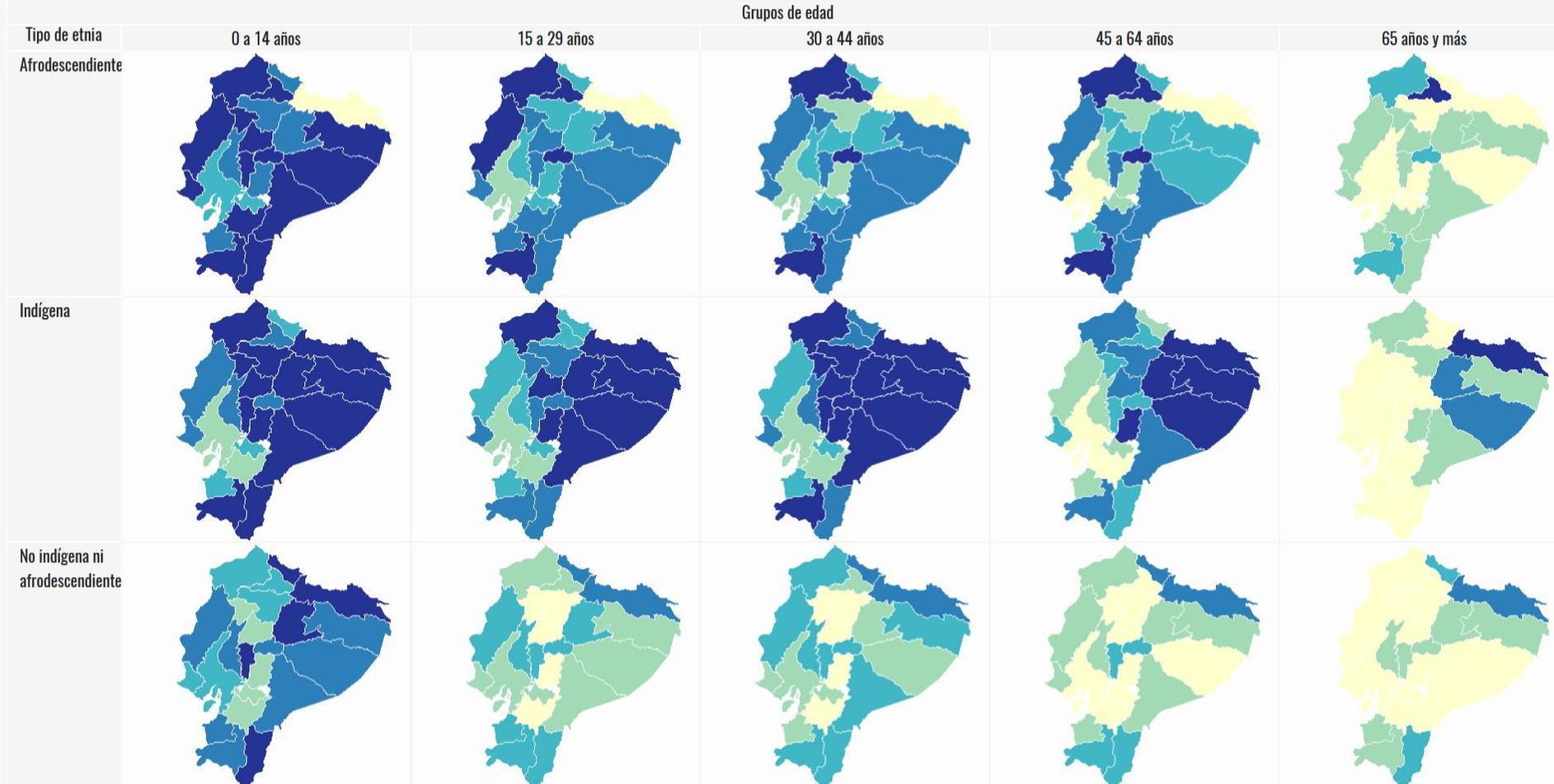
Tipo de etnia

Columnas:

Grupos de edad

Año:

2020



Hacia un mayor aprovechamiento de SAE para producir estadísticas desagregadas

- Es un buen momento para el desarrollo de SAE
 - Tema cuenta con creciente atención.
 - Mayor visibilidad en conferencias, foros, publicaciones.
- Desarrollo de capacidades técnicas
 - Amplia disponibilidad de cursos, guías, códigos (R, Stata).
 - Incentivar la experimentación y evaluación de los resultados.
- Acceso a fuentes de datos auxiliares
 - Posibilidad de aplicar SAE depende de contar con buenas covariables.
 - Ampliar acceso a registros administrativos y otras fuentes de datos auxiliares.
 - Creciente exploración con datos satelitales.
- Pasar de estadísticas experimentales a estadísticas oficiales
 - Mayor aceptación de estadísticas basadas en modelos
 - Adopción de mecanismos de evaluación de resultados
- Mejor acceso y difusión de estadísticas desagregadas





Muchas gracias por su atención!



© <https://issuu.com/> <https://mepyd.gob.do/>

#sosomeconomía #somosplanificación #somosdesarrollo

