



NACIONES UNIDAS

CEPAL

# Temas estadísticos de la CEPAL

No. 4 | Diciembre de 2021

## Estimaciones subnacionales de la pobreza para América Latina

### Contenidos

1. Introducción .....	2
2. Antecedentes de las mediciones de pobreza en CEPAL .....	2
3. Fuentes de información ...	3
4. Modelo de estimación en áreas pequeñas .....	3
5. Procedimiento de generación de mapas de pobreza .....	4
6. Resultados .....	6

Los mapas de pobreza a nivel subnacional son herramientas apropiadas para que los gobiernos puedan diseñar, aplicar y monitorear de manera más efectiva las políticas de desarrollo, dirigiéndolas a los lugares o grupos de población en donde se necesitan con mayor urgencia.

En la presente nota se presenta la metodología de mapeo de la pobreza basada en métodos de estimación en áreas pequeñas utilizada por la División de Estadísticas de la CEPAL. Esta metodología permite enlazar la información de un censo y una encuesta de hogares para realizar estimaciones a niveles provinciales, comunales o municipales, para los cuales las estimaciones directas que provienen de las encuestas de hogares no cuentan generalmente con suficiente precisión.

A manera de ilustración, se presentan los resultados obtenidos por esta metodología para los países de Chile, Colombia y Perú.

---

## 1. Introducción

---

Las encuestas de hogares son diseñadas y ejecutadas por las Oficinas Nacionales de Estadística para generar estadísticas representativas a un nivel de agregación determinado; generalmente supeditado a grandes subdivisiones geográficas, sexo, o dominios socioeconómicos de la población. Sin embargo, cuando se requiere obtener la estimación directa de distintos indicadores en subdivisiones más pequeñas de las que originalmente estaba planificado producir (por ejemplo, estimaciones a nivel provincial o municipal), la inferencia resultante de las encuestas resulta poco precisa y exacta. En general, a medida que aumenta la desagregación de las estimaciones se pierde eficacia en las mismas y su confiabilidad disminuye ostensiblemente, generando inclusive, para algunos indicadores complejos, problemas de sesgo en la estimación directa y en la estimación de su error estándar.

La estimación en áreas pequeñas (SAE, por sus siglas en inglés) es un conjunto de técnicas estadísticas que permiten obtener la estimación desagregada de parámetros poblacionales y que buscan mejorar la calidad de la inferencia cuando la desagregación de las encuestas de hogares no alcanza los criterios de calidad adecuados para su publicación. Este documento resume una serie de pasos para la aplicación del modelo de unidad con errores anidados para la estimación de indicadores de interés relacionados con los ingresos de los hogares, como son la incidencia, la brecha y la intensidad de la pobreza y la pobreza extrema, a nivel provincial en el Perú, comunal en Chile y municipal en Colombia. Los resultados permiten evidenciar una ganancia en precisión para los indicadores en áreas geográficas más pequeñas donde las encuestas no alcanzan la representatividad adecuada. Lo anterior es posible mediante el cálculo de los errores cuadrados medios para cada uno de los modelos establecidos.

---

## 2. Antecedentes de las mediciones de pobreza en CEPAL

---

La CEPAL produce periódicamente estimaciones de pobreza extrema y pobreza para 18 países de América Latina, utilizando una metodología que tiene como objetivo lograr la comparabilidad regional. La metodología general para medir la pobreza absoluta clasifica a una persona como pobre cuando el ingreso per cápita de su hogar es menor que la línea de pobreza, es decir, el costo de cubrir sus necesidades alimentarias y otras necesidades básicas no alimentarias<sup>1</sup>.

El costo de las necesidades alimentarias se estima mediante la construcción de canastas básicas de alimentos, que brindan las cantidades recomendadas de energía y nutrientes al tiempo que reflejan los hábitos de consumo de la población. Los requisitos provienen de las recomendaciones internacionales vigentes para mantener una vida saludable. Los hábitos de consumo se capturan a través de encuestas de ingresos y gastos de los hogares y corresponden a los de un subconjunto particular de la población, que se adopta como población de referencia con base en los criterios establecidos por la metodología. El costo mensual de la canasta básica de alimentos se conoce como la 'línea de pobreza extrema'. La línea de pobreza se obtiene como el producto de la línea de pobreza extrema por el cociente entre el gasto total y el gasto en alimentos de la misma población de referencia utilizada para definir la canasta básica de alimentos.

Los indicadores comúnmente utilizados para medir la pobreza corresponden a la familia de índices paramétricos propuestos por Foster, Greer y Thorbecke<sup>2</sup>. Estos índices (denominados FGT) corresponden a la función:

---

<sup>1</sup> Comisión Económica para América Latina y el Caribe (CEPAL), "Medición de la pobreza por ingresos: actualización metodológica y resultados", Metodologías de la CEPAL, No. 2 (LC/PUB.2018/22-P), Santiago, 2019.

<sup>2</sup> J. Foster, J. Greer and E. Thorbecke, "A class of decomposable poverty measures", *Econometrica*, vol. 52, No. 3, 1984.

$$F_{\alpha d} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left( \frac{z - y_{di}}{z} \right)^{\alpha} I(y_{di} < z)$$

En esta expresión,  $N_d$  es la población total del grupo de interés  $d$ ,  $y_{di}$  es el ingreso por persona del hogar,  $z$  es la línea de pobreza, y  $\alpha$  es un parámetro mayor o igual a 0 que determina las propiedades que cumple el índice. Cuando  $\alpha = 0$  el índice FGT corresponde al tradicional “índice de recuento” (denotado por H); es decir, la proporción de personas que se encuentran bajo la línea de pobreza. Si  $\alpha = 1$ , el índice FGT corresponde a la “brecha de pobreza”, que indica la distancia promedio entre los ingresos de los pobres y la línea de pobreza, ponderada por la incidencia de la pobreza. Si  $\alpha = 2$ , el coeficiente asigna un mayor peso relativo a las observaciones cuyo ingreso se encuentra a una mayor distancia de la línea de pobreza y, por lo tanto, hace que el indicador sea sensible a la distribución del ingreso entre las personas por debajo de la línea de pobreza.

---

### 3. Fuentes de información

---

La implementación del modelo a nivel de unidad SAE presentado en este documento requiere dos fuentes de información. Las primeras son las encuestas nacionales de hogares. En este caso, se obtienen del Banco de Datos de Encuestas de Hogares (BADEHOG) de la CEPAL, un repositorio de encuestas de hogares de 18 países de América Latina que mantiene la División de Estadística. La segunda fuente de información son los censos nacionales de población, a los que se ha accedido a través de los sitios web correspondientes de las Oficinas Nacionales de Estadística.

Para Chile se utilizó la Encuesta Nacional de Caracterización Socioeconómica (CASEN) 2017, que corresponde a una muestra representativa a nivel nacional, regional, nacional urbano y nacional rural, junto con el Censo de Población y Vivienda de 2017. Para Colombia se utilizó la Gran Encuesta Integrada de Hogares (GEIH) 2018, que es representativa a nivel nacional, nacional urbano, nacional rural, regional, departamental y para las capitales de los Departamentos del país, junto con el Censo Nacional de Población y Vivienda 2018. Para el Perú se utilizó la Encuesta Nacional de Hogares (ENAHOG) 2017, la cual es representativa a nivel nacional, urbano, rural y departamental; además del XII Censo de Población, VII de Vivienda y III de Comunidades Indígenas 2017.

---

### 4. Modelo de estimación en áreas pequeñas

---

Para la estimación de los ingresos promedio, se utiliza un modelo de unidad con ajuste al diseño de muestreo complejo. Este modelo induce una aproximación del mejor predictor empírico (Pseudo-EBP) basado en el modelo con errores anidados<sup>3</sup> propuesto por Guadarrama, Molina y Rao (2018)<sup>4</sup>. Este método asume que las variables transformadas del ingreso  $y_{di}^* = \log(y_{di} + c)$  siguen el modelo descrito a continuación (por simplicidad llamaremos  $y_{di}$  a la variable transformada):

$$y_{di}^* = x_{di}^T \beta + u_d + e_{di}, \quad i = 1, \dots, N_d, \quad d = 1, \dots, D,$$

<sup>3</sup> I. Molina, I. and J.N.K. Rao, “Small Area Estimation of Poverty Indicators”. *Canadian Journal of Statistics*, vol. 38, No. 3, 2010.

<sup>4</sup> M. Guadarrama, I. Molina and J. N. K. Rao, “Small area estimation of general parameters under complex sampling designs”. *Computational Statistics & Data Analysis*, No. 121, 2017.

donde  $\beta$  es el vector de coeficientes de las covariables,  $u_d$  es el efecto aleatorio del área tal que  $u_d \sim N(0, \sigma_u^2)$  y  $e_{di} \sim N(0, \sigma_e^2)$  son los errores a nivel individuo independientes de los efectos aleatorios. Según Molina (2019)<sup>5</sup>, para aquellos indicadores FGT que puedan ser definidos como una función de  $y_d$  el mejor predictor lineal es aquel que minimiza el Error Cuadrático Medio (ECM) y está dado por la esperanza de los elementos no seleccionados en la muestra dentro de la subdivisión de interés  $d$  condicionado a los valores observados para los elementos seleccionados en la muestra.

Dado que la información disponible no permite identificar y vincular las unidades de las muestras con las unidades de los censos, entonces el acercamiento que se utiliza es del tipo "Census-EB", que asume que todos los elementos del censo están asociados a las observaciones fuera de la muestra, tal como se presenta a continuación:

$$\tilde{F}_{ad}^B(\theta) = \frac{1}{N_d} \left( \sum_{i \in r_d} \tilde{F}_{\alpha, di}^B(\theta) \right)$$

Dado que, en las encuestas de hogares de la región, y en particular en las usadas en este documento, la fracción entre las unidades seleccionadas en las muestras y el número de personas del país es muy cercano a cero, el predictor Census-EB tiene un desempeño bastante similar a Pseudo-EBP.

---

## 5. Procedimiento de generación de mapas de pobreza

---

Este trabajo tiene como objetivo producir indicadores de pobreza desglosados geográficamente y producir mapas para visualizar las estimaciones resultantes. Esto ayuda a los formuladores de políticas a tener una perspectiva clara sobre la incidencia del indicador estimado en diferentes dominios geográficos, utilizando diferentes matices o colores para representar la magnitud de los indicadores de ingresos y pobreza. Las etapas del proceso de mapeo de la pobreza son las siguientes:

Etapa 1:

- » Estandarización y homologación de las bases de datos.
- » Elaboración de indicadores relacionados con desigualdad, ingresos y pobreza.

Etapa 2:

- » Estimación del modelo SAE para indicadores relacionados con ingresos, desigualdad, o pobreza.
- » Definición de las interacciones y selección de variables auxiliares.

Etapa 3:

- » Simulación por Bootstrap paramétrico para la estimación del ECM.

Etapa 4:

- » Validación de los supuestos del modelo.
- » Benchmarking con las estimaciones de la encuesta.

---

<sup>5</sup> I. Molina, "Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas", *Statistical Studies series*, No. 97 (LC/TS.2018/82/Rev.1), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC), 2019.

Etapa 5:

- » Generación de mapas con la estimación del indicador FGT y su respectivo ECM.

En la primera etapa, los modelos a nivel de unidad ajustados a los datos de la encuesta se replican utilizando los microdatos censales respectivos. Por ello, es necesario estandarizar las variables relevantes aplicando definiciones y categorías homogéneas en ambas fuentes de datos. De esta forma, se descartan posibles sesgos inducidos por las diferentes medidas en las covariables o errores en la predicción debido a diferentes variables con nombres similares. Para ello, se generan estructuras estandarizadas, junto con un diccionario de variables que describe las categorías y otras especificaciones necesarias para cada una de ellas.

A manera de ejemplo, para construir la variable "años de estudio", la ENAHO de Perú permite identificar el último año aprobado para todos los niveles de estudios aprobados; en cambio, el censo de Perú solo permite identificar esta desagregación tan específica hasta la secundaria, pues a partir de ahí las opciones de respuesta se abordan de manera mucho más general pues solo se consulta por educación superior completa o incompleta, pero no por el número de años cursados. Por otro lado, en el caso de Chile y Colombia la variable de años de estudio se encontraba implícita en los microdatos, por lo cual solo se realizó el proceso de categorización respectiva.

En la segunda parte de esta etapa se procede a transformar la variable de interés con el fin de garantizar la estructura del modelo con errores anidados definido en la sección 4. Por ejemplo, el modelo considera una transformación de la variable ingreso per cápita en el hogar que garantice una distribución aproximadamente normal; para tal fin se revisaron las familias de transformaciones Box-Cox y Log-Shift. Esta última fue escogida para realizar la transformación del ingreso en los modelos de los tres países, aunque los parámetros asociados a cada transformación resultaron diferentes.

Se crea una nueva variable  $f(y_{di}) = \log(y_{di} + c)$  para una grilla de valores predefinidos de  $c$ . El valor seleccionado de  $c$  es aquel que garantice la normalidad de la variable transformada, junto con un coeficiente de asimetría de Fisher más cercano a cero. Al ajustar el proceso a los países bajo análisis se obtuvieron los siguientes valores para la constante  $c$  expresados en moneda local: 8.600 CLP (Chile), 81.958 COP (Colombia) y 10,96 PEN (Perú).

En la segunda etapa de la metodología, se considera un procedimiento de simulación de Montecarlo para estimar los indicadores de pobreza considerados, pues a menudo la esperanza que define el mejor predictor no se puede calcular de forma analítica. Además, una parte importante de esta etapa consiste en identificar la capacidad predictiva de las variables auxiliares empleadas.

Una primera opción consiste en generar modelos lineales distintos a partir de diferentes combinaciones de las covariables (con y sin intercepto) y comparar sus respectivas medidas de bondad de ajuste. Adicional a lo anterior, el número de variables significativas, y las medidas de bondad de ajuste, como los criterios de información AIC o BIC son algunos de los elementos utilizados para el análisis y comparabilidad de los modelos.

Además, como primer paso para establecer la viabilidad de un conjunto de covariables, se utilizaron las regresiones Ridge y Lasso adaptado para analizar el ajuste de las covariables.

En la tercera etapa de la metodología propuesta se considera el método Bootstrap paramétrico para estimar el ECM del predictor Census-EB, el cual consiste en generar poblacionales completas empleando los hogares del censo ubicados en las provincias, comunas y municipios seleccionados en las encuestas de hogares de cada país y la ecuación del modelo, seleccionar una muestra con las mismas características que la origina y estimar el indicador FGT de interés a partir de cada población completa para toda subdivisión de interés. Luego de repetir este procedimiento muchas veces, se obtendrá el error cuadrático medio estimado. En esta etapa se recomienda que las provincias, comunas y municipios con un coeficiente de variación mayor a 30% sean excluidas del mapa al considerarse que no tienen la precisión requerida.

En la cuarta etapa del procedimiento se realiza el Benchmarking con las estimaciones de CEPAL obtenidas de la encuesta del indicador FGT bajo análisis. Este proceso se realiza al mismo nivel al que las estimaciones de las encuestas son representativas (insesgadas y precisas); es decir a nivel nacional, urbano, rural y por departamento. Lo que se busca con este proceso es llevar las agregaciones de provincias, comunas y municipios a la cifra reportada para los diferentes niveles de desagregación con el fin de: i) eliminar el sesgo producido por una mala especificación del modelo; ii) mejorar las estimaciones en las provincias, comunas y municipios con base en las estimaciones oficiales insesgadas y consistentes y iii) permitir que el mapa de pobreza construido tenga comparabilidad con las cifras publicadas.

Posteriormente, resulta importante realizar un proceso de validación de los supuestos del modelo de errores anidados descrito en la sección 4. En particular se realizan pruebas de normalidad (Kolmogorov–Smirnov, Jarque–Bera); también se emplean diagnósticos gráficos en forma de histograma con kernel de densidades y los gráficos cuantil–cuantil. Luego, se realizan pruebas de heteroscedasticidad (White y Breusch–Pagan) y se identifican valores atípicos e influyentes (distancias de Cook y los  $dfbetas$ ).

En la última etapa del procedimiento se hace uso de los sistemas de información geográfica (GIS, por sus siglas en inglés) y las capas de interés en cada país a nivel de provincia, comuna y municipios con el fin de generar los cartogramas que se presentan en el apéndice de este documento.

---

## 6. Resultados

---

Los mapas de pobreza son instrumentos que permiten condensar en una sola imagen un vasto número de cifras sobre ciudades o municipios. La visualización de la pobreza no solo es útil como una herramienta de comunicación, sino que facilita el análisis de las relaciones espaciales entre distintos indicadores, para permitir una mejor comprensión de la pobreza en los países de la región, y contribuye a identificar zonas de atención prioritaria y la focalización geográfica del gasto público, mejorar las coberturas de programas sociales, entre otros usos.

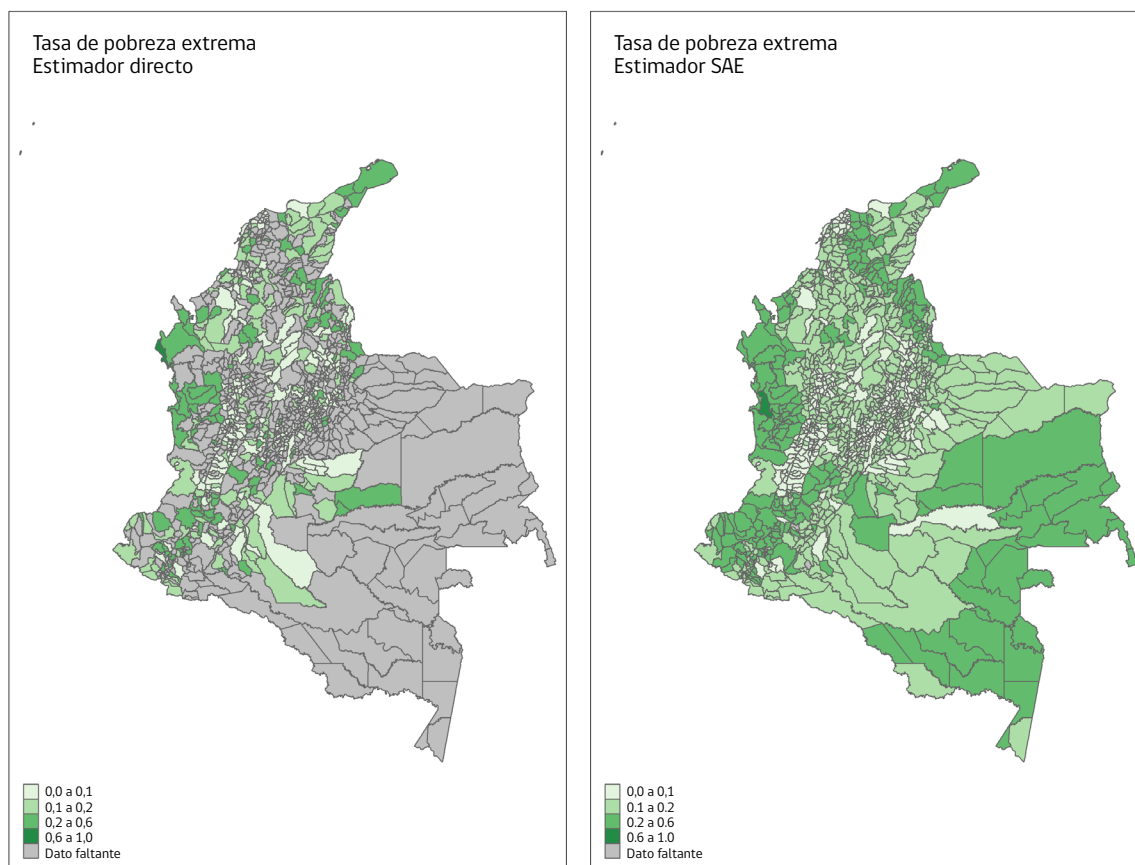
A continuación, se presentan algunos de los mapas de pobreza generados para Chile, Colombia y Perú. Todos fueron elaborados empleando 4 puntos de corte y una escala que va del color verde claro (menor pobreza) a rojo (mayor pobreza), con el fin de ilustrar la distribución de la incidencia de pobreza por provincias, comunas y municipios según el país bajo análisis<sup>6</sup>. Para ilustrar el efecto de la inclusión de la información auxiliar en la estimación de la pobreza y de la pobreza extrema en Colombia, el mapa 1 muestra las estimaciones directas, usando únicamente la información de las encuestas de hogares, y las estimaciones basadas en el modelo SAE, integrando la información de censos y encuestas. Nótese que los municipios para los cuales no se tienen estimaciones precisas, denotados con color gris, predominan en el mapa de las estimaciones directas.

---

<sup>6</sup> Dada la menor incidencia de la pobreza en Chile respecto a Colombia y Perú, los puntos de corte para el mapa de este país fueron modificados para reflejar de una mejor manera la dispersión del fenómeno en las comunas del país.



» **Mapa 1. Estimación Directa (izquierda) y estimación SAE (derecha) de la tasa de incidencia de la pobreza extrema en Colombia para el año 2019**



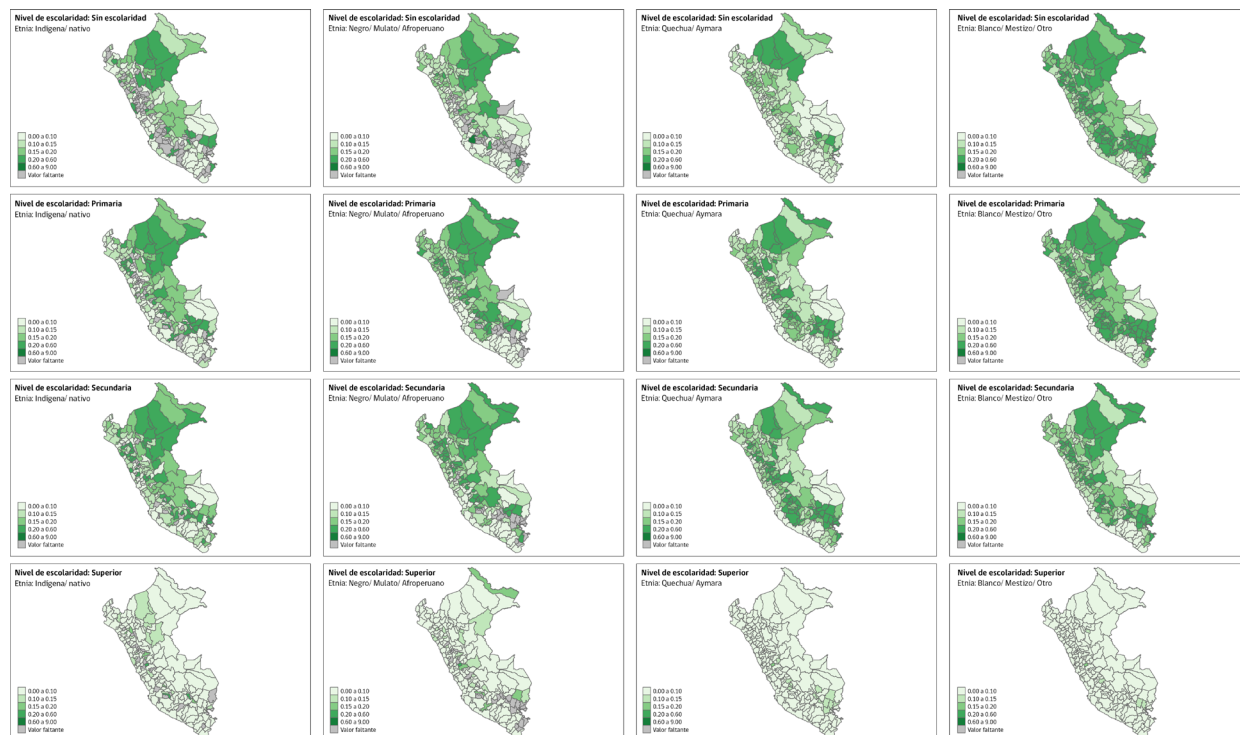
Fuente: Elaboración propia.

Nota: Los límites y los nombres que figuran en este mapa no implican su apoyo o aceptación oficial por las Naciones Unidas.

En Colombia, los municipios más pobres se encuentran ubicados en la periferia del país; puntualmente, en la región Pacífico (Chocó, Cauca y Nariño), región Caribe (La Guajira, Magdalena, Bolívar, Sucre, Córdoba y Cesar), departamentos de frontera (Norte de Santander y Arauca), y la región Orinoquía y Amazonía (Vichada, Guainía, Vaupés, Amazonas y Caquetá). En Perú, las mayores tasas de pobreza se observan en las zonas limítrofes del norte y sur del país. Específicamente, al norte la mayor incidencia de pobreza se concentra en las regiones de Loreto, Amazonas, la parte sur de Cajamarca y occidental de La libertad; de forma similar en la zona limítrofe con Bolivia se observa un segundo foco de pobreza en las regiones de Puno y Cuzco.

Las desagregaciones relevantes no se refieren solamente a áreas geográficas, sino también a los grupos poblacionales de interés para las políticas públicas y el seguimiento y monitoreo de la agenda 2030, según su sexo, edad, situación laboral, grupo étnico, situación de discapacidad, etc. A manera de ilustración, el mapa 2 muestra la distribución de la pobreza en las provincias de Perú, según el grupo étnico y nivel de educación. En particular se observa que la tasa de pobreza entre los grupos indígenas es mayor que entre los afrodescendientes y, esta a su vez es mayor que la de los individuos que no se auto-identifican con ninguna etnia. De la misma manera, a medida que se avanza en el logro educativo, las personas tienen una menor posibilidad de estar en situación de pobreza.

## » Mapa 2. Estimación SAE de la tasa de incidencia de pobreza en Perú para el año 2019 desglosada por etnia y nivel de escolaridad



Fuente: Elaboración propia.

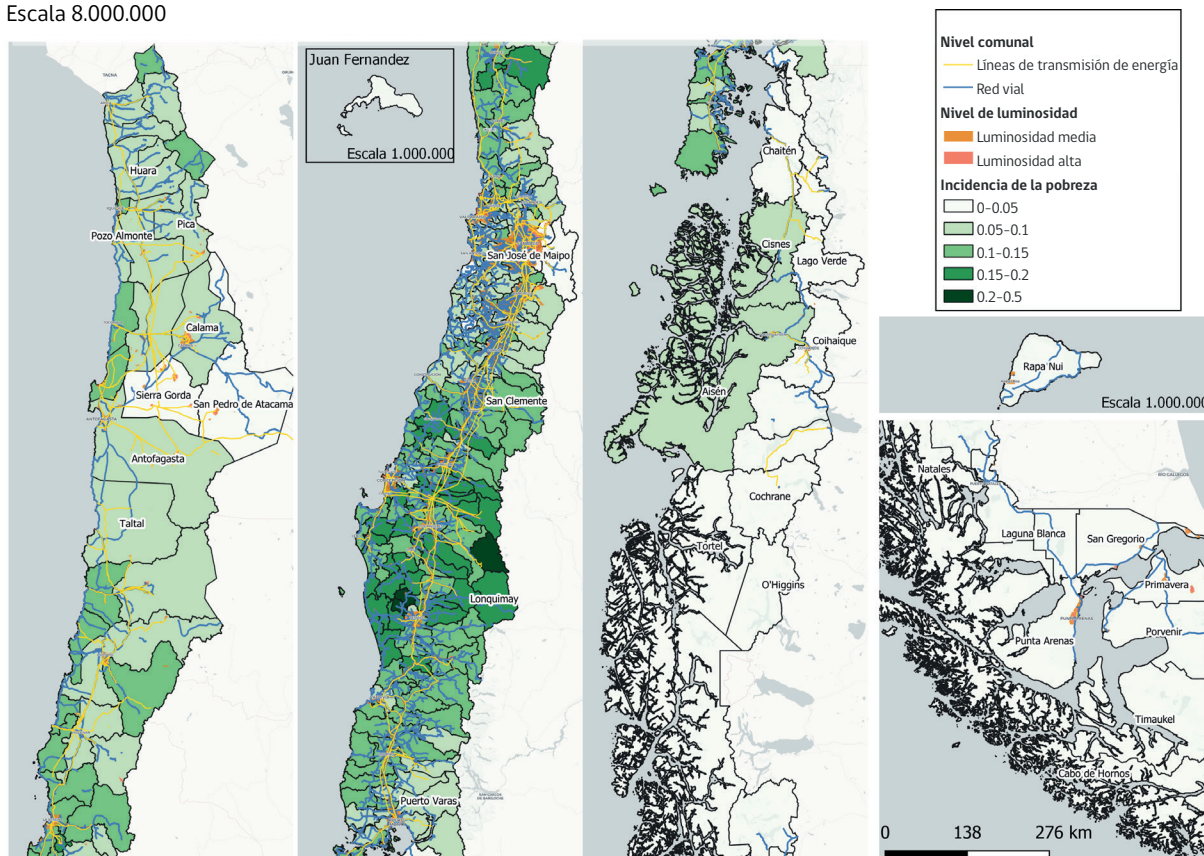
Nota: Los límites y los nombres que figuran en este mapa no implican su apoyo o aceptación oficial por las Naciones Unidas.

Los resultados de los mapas de pobreza pueden ser contrastados con las redes de infraestructura de los países, por ejemplo, la cobertura vial y el acceso a energía eléctrica constituyen un elemento central en la integración del sistema económico, facilitando las transacciones entre empresas y ciudadanos. El mapa 3 muestra que, en general, las zonas con menor incidencia de pobreza se encuentran en centros urbanos, los cuales concentran mayor infraestructura vial y una prestación servicios más eficiente.



### » Mapa 3: Contraste de la infraestructura local con el resultado de la estimación SAE de la tasa de incidencia de pobreza en Chile

Incidencia de la pobreza a nivel comunal y acceso a infraestructura, Chile.  
Escala 8.000.000



Fuente: Elaboración propia

Nota: Los límites y los nombres que figuran en este mapa no implican su apoyo o aceptación oficial por las Naciones Unidas.

Esta publicación, de carácter bimestral, es elaborada por la División de Estadísticas de la Comisión Económica para América Latina y el Caribe (CEPAL). Este número fue preparado por Andrés Gutiérrez, Asesor Regional en Estadísticas Sociales; Xavier Mancero, Jefe de la Unidad de Estadísticas Sociales; Felipe Molina y Diego Lemus, consultores de la División de Estadísticas.

Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad de los autores y pueden no coincidir con las de la Organización o las de los países que representa.

[www.cepal.org/es/estadisticas](http://www.cepal.org/es/estadisticas)



Comisión Económica para América Latina y el Caribe (CEPAL)  
Economic Commission for Latin America and the Caribbean (ECLAC)  
[www.cepal.org](http://www.cepal.org)