

Estimaciones Comunales de Pobreza por Ingresos y Multidimensional en base a Casen 2022

Antecedentes

La Encuesta de Caracterización Socioeconómica Nacional 2022 (Casen 2022), tiene representatividad a nivel nacional, regional (16 regiones del país) y en las áreas geográficas urbana y rural. Sin embargo, las comunas no son un dominio de estudio por lo que las estimaciones a ese nivel pueden tener una baja precisión y no se recomienda realizarlas.

Con el objetivo de entregar información a nivel comunal, el Ministerio de Desarrollo Social y Familia, aplica la metodología de áreas pequeñas o SAE, por sus siglas en inglés (*Small Area Estimation*), por medio de la cual se ha logrado obtener, desde 2011, estimaciones comunales para la tasa de pobreza por ingresos y, desde 2015, estimaciones comunales para el índice de pobreza multidimensional.

La metodología fue actualizada entre 2019 y 2020, con apoyo de CEPAL. Para más información sobre esta actualización, se recomienda revisar el informe metodológico disponible en:

https://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/pobreza-comunal/2020/Informe_SAE_2020.pdf

Las estimaciones de pobreza por ingresos comunal y del índice de pobreza multidimensional comunal basadas en Casen 2022, utilizan la metodología actualizada el año 2020, de manera que este documento permite conocer los principales aspectos metodológicos que están siendo utilizados en la actualidad, en cuanto a la aplicación de la metodología SAE.

Metodología

Las estimaciones comunales desarrolladas en base a Casen 2022 utilizan la metodología de estimación de áreas pequeñas, basadas en el modelo de áreas de Fay-Herriot, que plantea el uso de un estimador compuesto, definido como una combinación lineal entre un estimador sintético y el estimador directo, mediante una ponderación obtenida como función de la estimación de las varianzas de cada estimador. Mientras más pequeña sea la estimación de la varianza asociada a la estimación directa, en comparación con la heterogeneidad no explicada por esta, mayor será la ponderación que se le otorgará a dicha estimación y viceversa.

El modelo de Fay-Herriot utilizado en Chile, supone dos niveles de estimación: en el primer nivel se asume que la pobreza a nivel de la comuna se puede explicar por covariables observables x_d y un error comunal, denominado efecto aleatorio, u_d ; mientras que en el segundo nivel se asume una interacción con la estimación directa, proveniente del modelo de muestreo a través de Casen, en el cual se approxima la tasa de pobreza mediante una estimación que tiene asociado un error de muestreo e_d , que se supone conocido.

- **Nivel 1: Modelo de vínculo (estimación sintética)**

$$\theta_d = \mathbf{x}_d' \beta + u_d, \text{ con } u_d \stackrel{iid}{\sim} (0, \sigma_u^2)$$

- **Nivel 2: Modelo de muestreo (estimación directa)**

$$\hat{\theta}_d^{Dir} = Y_d + e_d, \text{ con } e_d \stackrel{ind}{\sim} (0, \psi_a^2)$$

Estos dos niveles pueden ser especificados en una sola expresión algebraica:

$$\hat{\theta}_d^{FH} = \mathbf{x}_d' \beta + u_d + e_d, \text{ con } u_d \stackrel{iid}{\sim} (0, \sigma_u^2), \text{ } e_d \stackrel{ind}{\sim} (0, \psi_a^2)$$

La cual, a su vez, puede expresarse como una ponderación de las estimaciones directas y las estimaciones sintéticas:

$$\hat{\theta}_d^{FH} = \hat{\gamma}_d \hat{\theta}_d^{Dir} + (1 - \hat{\gamma}_d) \hat{\theta}_d^{Syn}, \text{ con } d = 1, \dots, D$$

En donde $\hat{\theta}_d^{Dir}$ denota la estimación directa, $\hat{\theta}_d^{Syn} = \mathbf{x}_d' \hat{\beta}$ corresponde a la estimación sintética usando variables auxiliares, y $\hat{\gamma}_d = \widehat{\sigma_u^2} / (\widehat{\sigma_u^2} + \widehat{\psi}_d^2)$ corresponde al ponderador.

En la práctica, existen varios pasos para asegurar la calidad de los insumos que se utilizan en las estimaciones comunales, tanto de la estimación directa de Casen como de las fuentes administrativas utilizadas. Los pasos involucrados en la estimación de áreas pequeñas son:

- Generación de base de datos de fuente administrativas para la estimación sintética
- Suavización (*Trimming*) de los factores de expansión (metodología Potter-1993)
- Criterios de calidad para la estimación de las estimaciones directas
- Estimación de la varianza directa usando Función Generalizada de Varianza
- Selección del modelo sintético
- Aplicación programa EMDI¹
- Cálculo de los Intervalos de Confianza en base al Error Cuadrático Medio (ECM) usando *Bootstrap*
- Calibración (*benchmark*) de estimaciones comunales con estimaciones regionales
- Validación del modelo

¹ Desde 2020 se utiliza el paquete EMDI (*Estimating and Mapping Disaggregated Indicators*) de R, que permite centrar el trabajo en la calidad de los insumos que entran al modelo.

Los detalles de cada uno de estos pasos se pueden encontrar en el informe metodológico, disponible en:

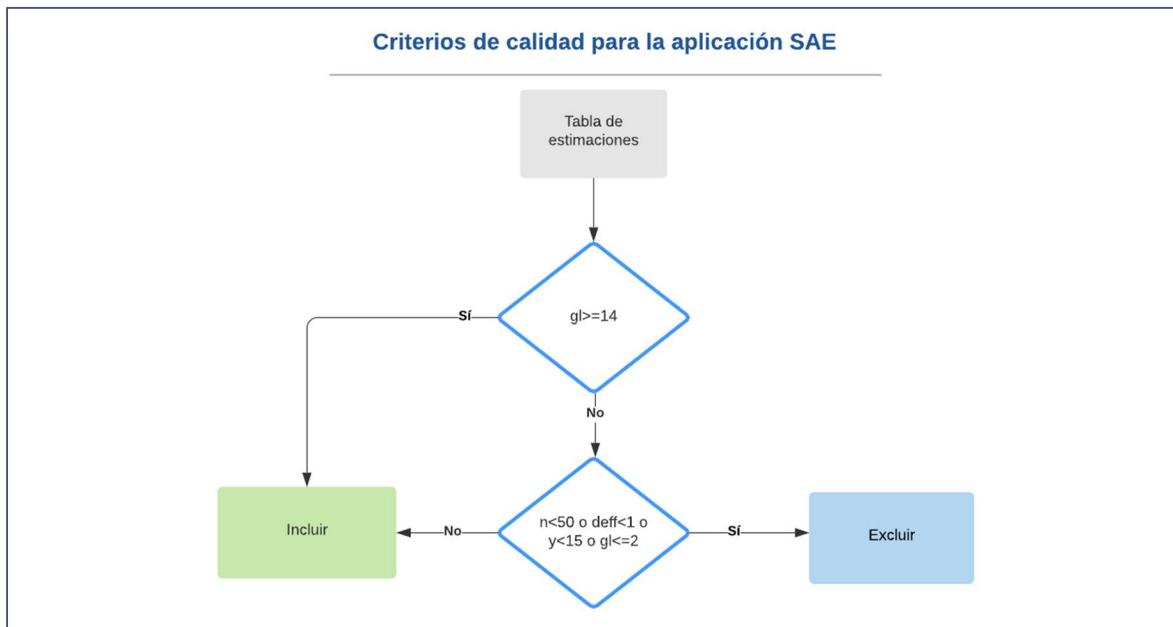
https://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/pobreza-comunal/2020/Informe_SAE_2020.pdf

En la siguiente sección se hace énfasis en los criterios de calidad aplicados y en la validación del modelo.

Criterios de Calidad

Para asegurar que las estimaciones directas de pobreza que se utilizan en la estimación SAE cumplan con las propiedades estadísticas esperadas, se utilizan criterios de calidad para evaluar las estimaciones directas. Los criterios utilizados son:

- Grados de libertad (gl)
- Muestra de personas lograda (n)
- Efecto diseño ($deff$)
- Número de personas en situación de pobreza (y)



Fuente: Informe Metodológico SAE 2020. Ministerio Desarrollo Social y Familia.

Del total de 335 comunas incluidas en Casen 2022, un total de 247 comunas en el caso de pobreza por ingresos, y de 298 en el caso de pobreza multidimensional, cumplen los criterios de calidad. Estas comunas entrarán al modelo de Fay-Herriot y su estimación será un indicador compuesto por la estimación directa de Casen y la estimación sintética del modelo de regresión. Las 10 comunas

sin muestra Casen y las comunas que no cumplen los criterios de calidad tendrán como estimación de área pequeña la estimación sintética del modelo de regresión.

En el caso de la estimación de la pobreza comunal por ingresos, de las primeras 247 comunas inicialmente seleccionadas, 246 fueron incluidas en el modelo Fay-Herriot en las estimaciones de pobreza comunal, dejando fuera a la comuna de Puchuncaví. Las razones para su exclusión son las siguientes:

- Si bien cumple el criterio de los 14 grados de libertad, lo que implica su selección automática, no cumple el criterio de número de personas en pobreza. Además, al tener una estimación puntual de cero, no permite evaluar estadísticos como el coeficiente de variación, el error estándar o el efecto de diseño.
- Incluir esta comuna empeora la calidad de la estimación de Fay-Herriot, arrojando un ajuste más bajo; esta comuna aparece como una comuna *outlier* en la regresión sintética y ocasiona que otras comunas sin muestra tengan un coeficiente de variación muy alto.
- La estimación directa de pobreza por ingreso para esta comuna no es esperable en base a sus valores históricos.

Se debe mencionar que en el caso de la pobreza multidimensional se mantienen las 298 comunas incluidas, pues todas cumplieron con los criterios de calidad establecidos.

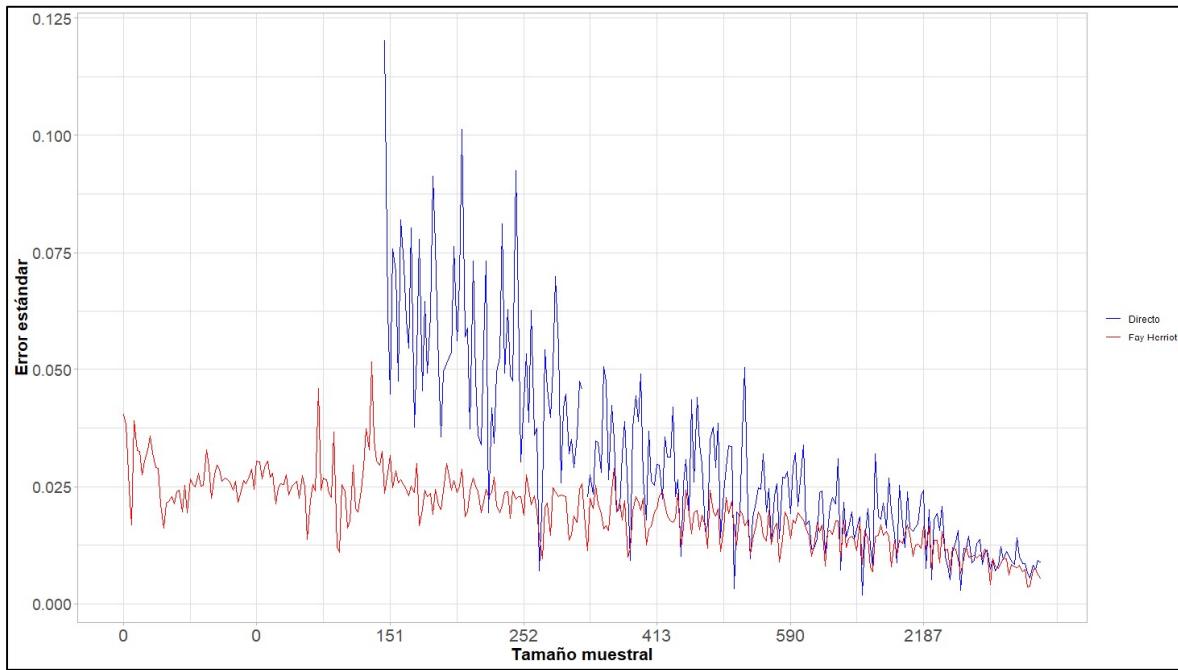
Validación del Modelo

Las estimaciones de área pequeña son desarrolladas para lograr estimaciones a nivel comunal, que tengan una mayor precisión que las estimaciones directas que se obtienen de Casen. Para validar los resultados del modelo se realizan varios análisis entre los que destacan:

- Bondad de ajuste del modelo y *tes-t* de los parámetros
- Evaluación de la homocedasticidad de los residuos
- Evaluación de normalidad de los errores
- Distancia de Cook (observaciones incluyentes)
- Error Cuadrático Medio Relativo de la Raíz (RRMSE) y Error estándar

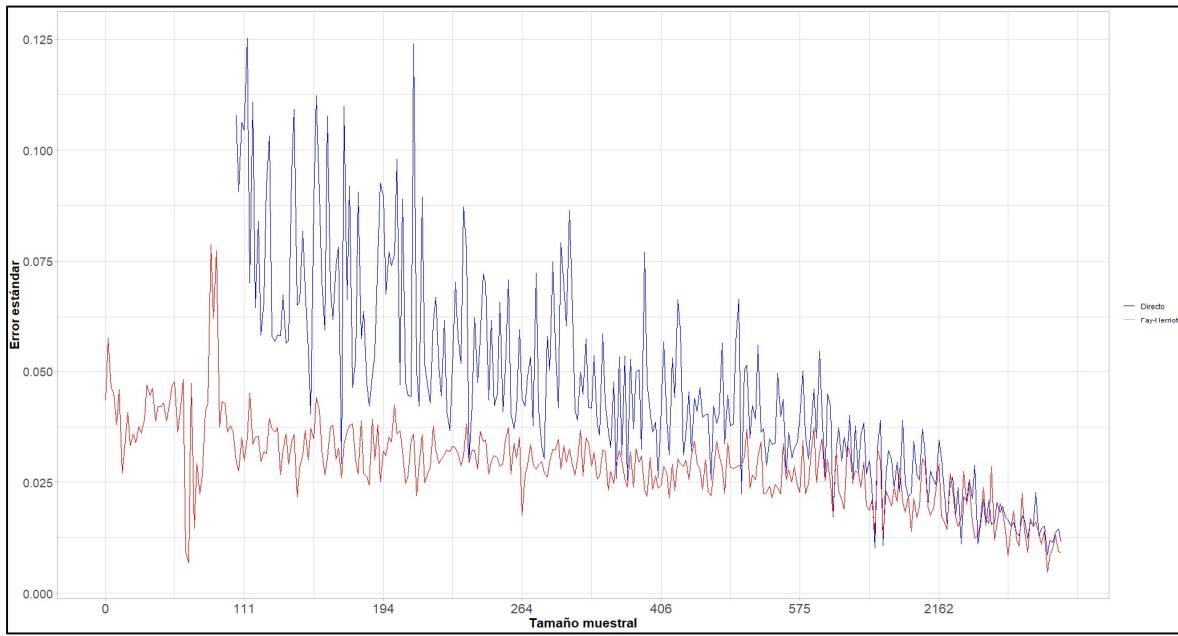
Los resultados, tanto para la estimación de pobreza por ingresos como para el índice de pobreza multidimensional, muestran que el modelo es adecuado en su funcionamiento. En el gráfico a continuación se muestra el Error Estándar para las estimaciones directa y para el estimador Fay-Herriot en base a Casen 2022, observándose una reducción importante del error de las estimaciones.

Error Estándar para estimaciones directa y estimador Fay-Herriot Casen 2022
Pobreza por Ingresos



Fuente: Ministerio Desarrollo Social y Familia en base a Casen 2022.

Error Estándar para estimaciones directa y estimador Fay-Herriot Casen 2022
Índice Pobreza Multidimensional



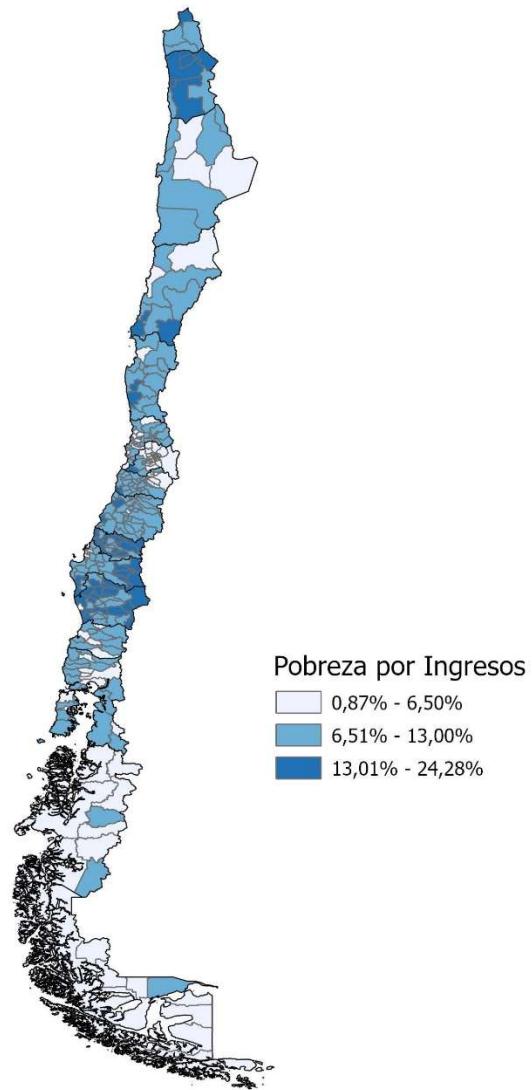
Fuente: Ministerio Desarrollo Social y Familia en base a Casen 2022.

Resultados

Los resultados de la estimación comunal de la pobreza por ingresos y del índice de pobreza multidimensional se pueden encontrar en el siguiente link:
<https://observatorio.ministeriodesarrollosocial.gob.cl/pobreza-comunal>

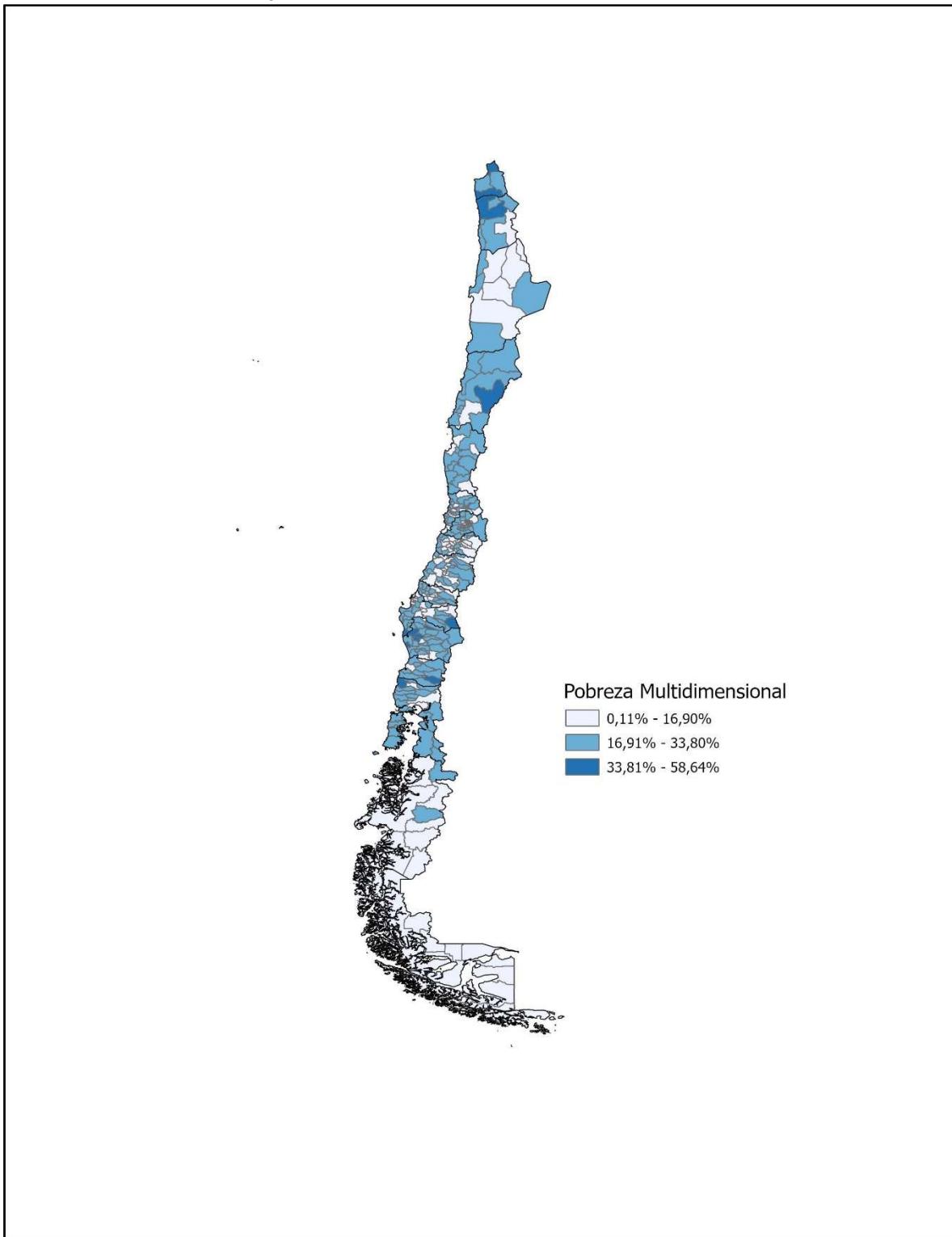
A continuación, se presentan 2 mapas del país con las estimaciones comunales para la tasa de pobreza según ingresos y el índice de pobreza multidimensional en base a la utilización de la metodología SAE en Casen 2022.

Estimación comunal pobreza por ingresos Casen 2022



Fuente: Ministerio Desarrollo Social y Familia en base a Casen 2022.

Estimación comunal índice pobreza multidimensional Casen 2022



Fuente: Ministerio Desarrollo Social y Familia en base a Casen 2022.