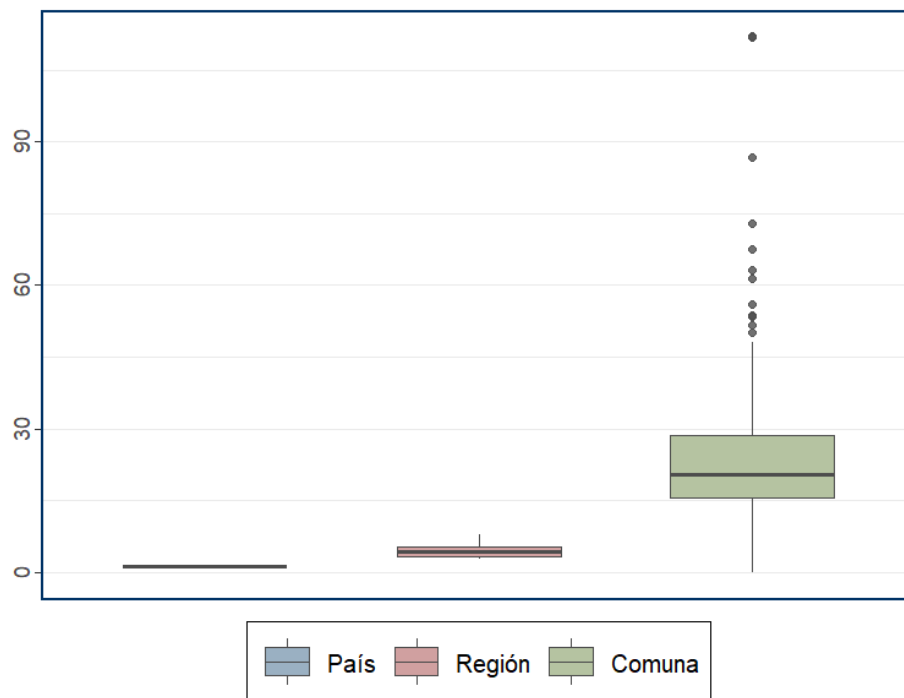


## Estimaciones Comunales de Pobreza por Ingresos y Multidimensional en base a Casen 2024

### Antecedentes

La Encuesta de Caracterización Socioeconómica Nacional 2024 (Casen 2024), 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. Esta limitación se ilustra en el siguiente gráfico, donde el coeficiente de variación (error relativo) de las estimaciones directas aumenta drásticamente al descender al nivel comunal.

### Coefficiente de Variación Pobreza por Ingresos para distintos niveles de desagregación. Casen 2024



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

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 y los pasos que se realizan para las estimaciones SAE se recomienda revisar el informe metodológico disponible en la página del BIDAT:

<https://bidat.gob.cl/details/ficha/dato/informe-metodologico-sae-2020>

Las estimaciones de pobreza por ingresos comunal y del índice de pobreza multidimensional comunal basadas en Casen 2024, 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 2024 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 aproxima 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)**

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

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

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

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

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

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  $\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\psi}_d^2}$  corresponde al ponderador.

En términos prácticos, la integración de estos dos niveles resulta en una única estimación final que pondera ambas fuentes de información. El ponderador estadístico actúa dinámicamente para cada comuna: compara el error de muestreo de la encuesta frente a la capacidad predictiva del modelo. Así, en comunas con estimaciones directas precisas, el resultado final se acercará fuertemente al dato observado en Casen; mientras que, en comunas con muestras pequeñas o sin presencia en la encuesta, el resultado dependerá en mayor medida, o en su totalidad, de la estimación sintética construida a partir de los registros administrativos y censales.

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 EMDI<sup>1</sup>
- 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

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

## Fuentes de Información

La construcción del modelo de Estimación de Áreas Pequeñas (SAE) requiere de la interacción de dos componentes de información. En primer lugar, es indispensable contar con una encuesta probabilística que mida el fenómeno a modelar para obtener las estimaciones directas, rol que cumple la Encuesta Casen 2024 al proveer las mediciones de pobreza por ingresos y pobreza multidimensional. En segundo lugar, el modelo predictivo requiere contar con un conjunto de

---

<sup>1</sup> 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.

variables auxiliares (covariables) que estén fuertemente correlacionadas con la pobreza y que se encuentren disponibles y actualizadas para todas las comunas del país.

Para el ejercicio de estimación 2024, una de las mejoras metodológicas más relevantes es la incorporación de los microdatos del Censo de Población y Vivienda 2024. Su uso permite mejorar la calidad de la estimación, ya que provee información demográfica, territorial y socioeconómica estandarizada, actualizada y universal para todas las comunas, mitigando los posibles sesgos de cobertura inherentes a la interacción de la población con los servicios públicos.

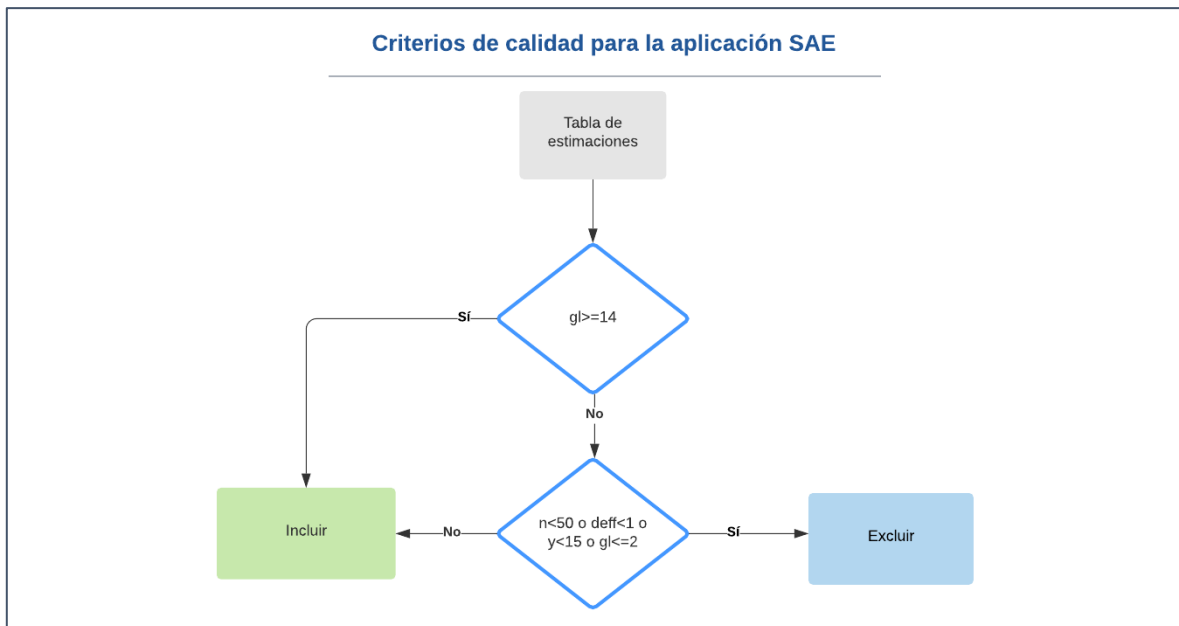
Este ecosistema de información auxiliar censal se complementa con un amplio volumen de registros administrativos consolidados a nivel comunal. Entre las principales fuentes utilizadas destacan:

- **Censo de Población y Vivienda 2024 (INE):** Provee variables demográficas (composición etaria, índices de dependencia), migración, educación, discapacidad, pertenencia a pueblos originarios, características deficitarias de la vivienda, ocupación y rama de actividad, entre otras.
- **Sistema de Indicadores con Registros Administrativos (Subsecretaría de Evaluación Social):** La Subsecretaría ha implementado este sistema basado en registros administrativos del MDSF con el objetivo de fortalecer la caracterización socioeconómica del país y apoyar el diseño y evaluación de políticas sociales. Este sistema permite monitorear con mayor frecuencia y desagregación territorial dimensiones clave utilizadas en este modelo, tales como ingresos, empleo, y educación. De aquí se extraen indicadores como la remuneración promedio y mediana imponible, proporción de trabajadores formales, tasas de ocupación, tasas de matrícula y rezago escolar.
- **Registro Social de Hogares (MDSF /ADIS):** Mediante el sistema de estadísticas del Analista Digital de Información Social (ADIS), se extraen datos agregados de cobertura poblacional del instrumento y la distribución de hogares, destacando la proporción ubicada en el tramo del 40% de mayor vulnerabilidad a diciembre de 2024.
- **Índice de Vulnerabilidad Escolar (JUNAEB):** Proporción de estudiantes clasificados según sus prioridades de vulnerabilidad en el sistema educativo.
- **Salud (FONASA, ISAPRE y DEIS):** Proporción de la población afiliada al sistema público de salud (desagregada por tramos A, B, C y D) y al sistema privado, junto a la tasa de mortalidad infantil.
- **Gestión e Ingresos Municipales (SUBDERE - SINIM):** Ingresos propios permanentes por habitante, distribución del ingreso municipal respecto al total nacional y proporción del Fondo Común Municipal (FCM) con respecto al ingreso total de la comuna.
- **Entorno y Seguridad (Subsecretaría de Prevención del Delito):** Tasas de victimización reportadas a nivel comunal.
- **Datos Satelitales:** Cubrimiento de cultivo, cubrimiento urbano, modificación humana, accesibilidad a hospitales, accesibilidad a hospitales caminando, luces nocturnas.

## 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): Se incluyen aquellas con  $gl \geq 14$ . Se excluyen aquellas con  $gl \leq 2$ .
- Muestra de personas lograda (n): Se excluyen comunas con menos de 50 personas encuestadas.
- Efecto diseño (deff): Se excluyen comunas con un DEFF menor a 1.
- Número de personas en situación de pobreza (y): Se excluyen comunas con menos de 15 casos de personas en situación de pobreza en la muestra.



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

Del total de 335 comunas incluidas en Casen 2024, un total de 299 comunas en el caso de pobreza por ingresos, y de 294 en el caso de pobreza multidimensional, cumplieron los criterios de calidad y entraron directamente como insumo para ajustar el modelo Fay-Herriot. Aquellas comunas que no cumplieron los criterios o que no tuvieron muestra en Casen, reciben directamente la estimación sintética del modelo.

## Selección del modelo

Dado el gran volumen de covariables candidatas provenientes del Censo 2024 y los diversos registros administrativos, se implementó un proceso automatizado para la selección del modelo sintético. El objetivo de este procedimiento es identificar el subconjunto de variables que mejor prediga la tasa de pobreza comunal, por ingresos y multidimensional, evitando problemas de sobreajuste, multicolinealidad y ausencia de normalidad en los residuos.

El algoritmo de selección incorporó las siguientes etapas metodológicas:

- **Exclusión de variables altamente correlacionadas:** Se analizó la matriz de correlaciones del conjunto completo de variables candidatas, eliminando de forma preliminar aquellas que presentaran una correlación cruzada extrema (mayor a 0.90) para evitar redundancias de información.
- **Depuración mediante Factor de Inflación de Varianza (VIF):** Sobre la base reducida, se aplicó un filtro iterativo basado en el Factor de Inflación de la Varianza (VIF), descartando variables que generaran inestabilidad en los coeficientes.
- **Optimización iterativa del modelo (Algoritmo *Stepwise*):** Para evitar que el modelo dependiera de un único punto de partida, se ejecutó un algoritmo *stepwise* sobre 150 iteraciones mediante un muestreo aleatorio (reteniendo el 90% de los predictores en cada ciclo). La selección interna se optimizó utilizando el Criterio de Información Bayesiano (BIC), el cual penaliza la inclusión excesiva de parámetros, favoreciendo modelos parsimoniosos.
- **Validación de supuestos y elección del modelo final:** De los modelos generados, la selección final correspondió al modelo con menor BIC dentro aquellos que cumplieran con tres condiciones estadísticas:
  - Normalidad de los residuos (Test de Shapiro-Wilk no significativo).
  - Ausencia de multicolinealidad (VIF máximo menor a 10).
  - Baja sensibilidad a valores atípicos (menos del 10% de comunas superando el umbral de influencia de la Distancia de Cook).

Como resultado de este proceso, se definieron los modelos predictivos finales tanto para la pobreza por ingresos como para la pobreza multidimensional.

*Tabla 1. Variables predictoras seleccionadas para el modelo de Pobreza por Ingresos 2024*

<b>Variable</b>	<b>Fuente</b>	<b>Relación (Signo)</b>
Proporción de hogares en el tramo de 51% a 60% de vulnerabilidad	Registro Social de Hogares	(-) Negativo
Proporción de personas ocupadas en categoría independiente	Censo 2024	(+) Positivo
Proporción de personas ocupadas que utiliza automóvil como transporte	Censo 2024	(-) Negativo
Proporción de la población que habla lengua originaria	Censo 2024	(+) Positivo
Proporción de ocupados en la rama de Comercio (CAENES G)	Censo 2024	(+) Positivo
Proporción de ocupados en la rama de Otras actividades de servicios (CAENES S)	Censo 2024	(-) Negativo
Proporción de ocupados en org. y órganos extraterritoriales (CAENES U)	Censo 2024	(-) Negativo
Tasa de matrícula bruta en educación media	MINEDUC 2024	(+) Positivo
Proporción de trabajadores formales de 25 a 34 años con ingresos menores a la mediana	AFC 2024	(+) Positivo
Proporción de trabajadores formales de 65 o más años con ingresos menores a la mediana	AFC 2024	(-) Negativo
Proporción de trabajadores formales dependientes de 60 años o más	AFC 2024	(-) Negativo

*Nota: El modelo incorpora además variables dicotómicas regionales como controles de efectos fijos territoriales.*

*Tabla 2. Variables predictoras seleccionadas para el modelo de Pobreza Multidimensional 2024*

<b>Variable</b>	<b>Fuente</b>	<b>Relación (Signo)</b>
Proporción de mujeres	Censo 2024	(-) Negativo
Tasa de ocupación formal dependiente en la población de 60 años o más	AFC 2024	(-) Negativo
Accesibilidad a Hospitales (Tiempo de viaje a hospital o clínica más cercana)	Satelital	(-) Negativo
Porcentaje de migración interna (5 años)	Censo 2024	(-) Negativo
Proporción de mujeres en Fonasa B	Fonasa	(-) Negativo
Proporción de la población que habla lengua originaria	Censo 2024	(+) Positivo
Proporción de hogares en el tramo de 51% a 60% de vulnerabilidad	Registro Social de Hogares	(-) Negativo
Tasa de asistencia neta educación media	Censo 2024	(-) Negativo
Índice de Vulnerabilidad Escolar	Junaeb 2024	(+) Positivo
Proporción de ocupados en la rama de actividades de servicios administrativos y de apoyo (CAENES N)	Censo 2024	(+) Positivo
Proporción de ocupados en la rama de actividades artísticas, de entretenimiento y recreativas (CAENES R)	Censo 2024	(-) Negativo

*Nota: El modelo incorpora además variables dicotómicas regionales como controles de efectos fijos territoriales.*

## Validación del Modelo

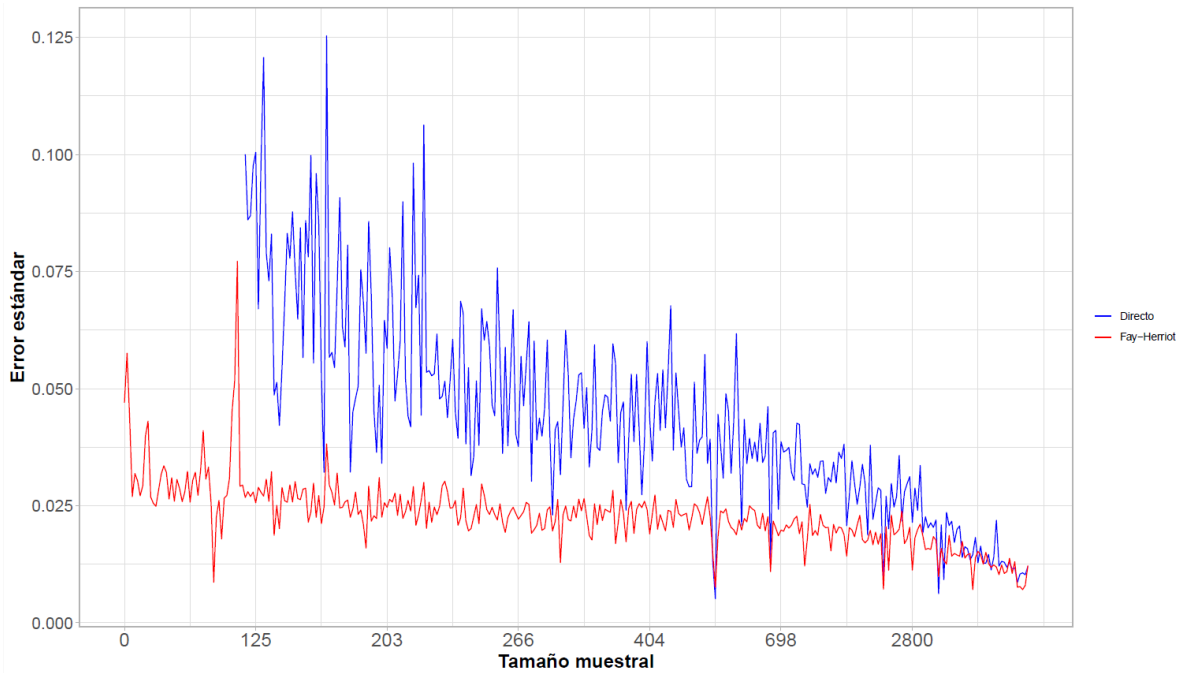
Las estimaciones de área pequeña son desarrolladas para lograr estimaciones a nivel comunal, que tengan una mayor precisión que las estimaciones directas que se obtienen de Casen. Para validar los resultados del modelo se realizan varios análisis entre los que destacan:

- Bondad de ajuste del modelo y test de significancia de los parámetros
- Evaluación de la homocedasticidad y normalidad de los residuos
- Análisis de influencia y valores atípicos (Distancia de Cook).
- Error Cuadrático Medio Relativo de la Raíz (RRMSE) y Error estándar

Los resultados de estas pruebas confirman que los modelos sintéticos, tanto para la pobreza por ingresos como para el índice de pobreza multidimensional, cumplen adecuadamente con los supuestos teóricos requeridos por la metodología de Fay-Herriot. Asimismo, se verificó la consistencia de las estimaciones: los resultados del modelo mantienen una alta correlación y se agrupan simétricamente en torno a las estimaciones directas originales, lo que demuestra que el modelo corrige la varianza extrema sin introducir sesgos poblacionales.

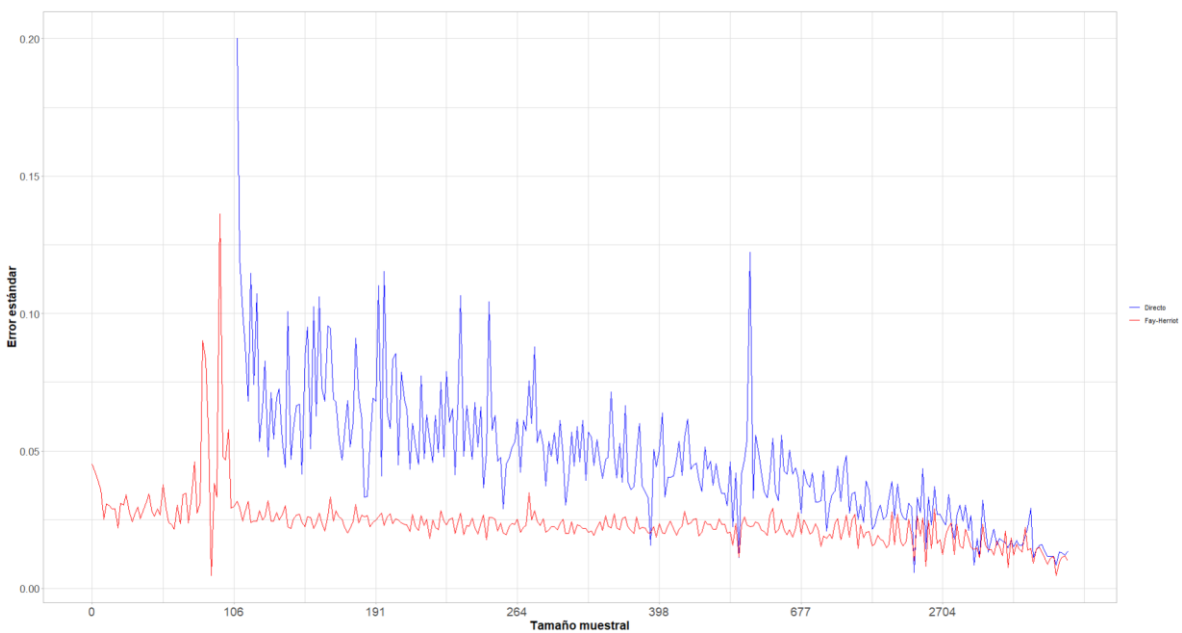
El principal logro metodológico se visualiza en la ganancia de eficiencia. En los gráficos a continuación se contrasta el Error Estándar de la estimación directa (línea azul) frente al error del estimador Fay-Herriot (línea roja), ordenados de menor a mayor tamaño muestral. Se observa una reducción sistemática en la incertidumbre estadística de las estimaciones del modelo, especialmente en aquellas comunas con tamaños de muestra más reducidos.

### Error Estándar para estimaciones directa y estimador Fay-Herriot Casen 2024 Pobreza por Ingresos



Fuente: Ministerio Desarrollo Social y Familia en base a Casen 2024.

### Error Estándar para estimaciones directa y estimador Fay-Herriot Casen 2024 Índice Pobreza Multidimensional



Fuente: Ministerio Desarrollo Social y Familia en base a Casen 2024.

## 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://bidat.gob.cl/directorio/Pobreza%20comunal/>

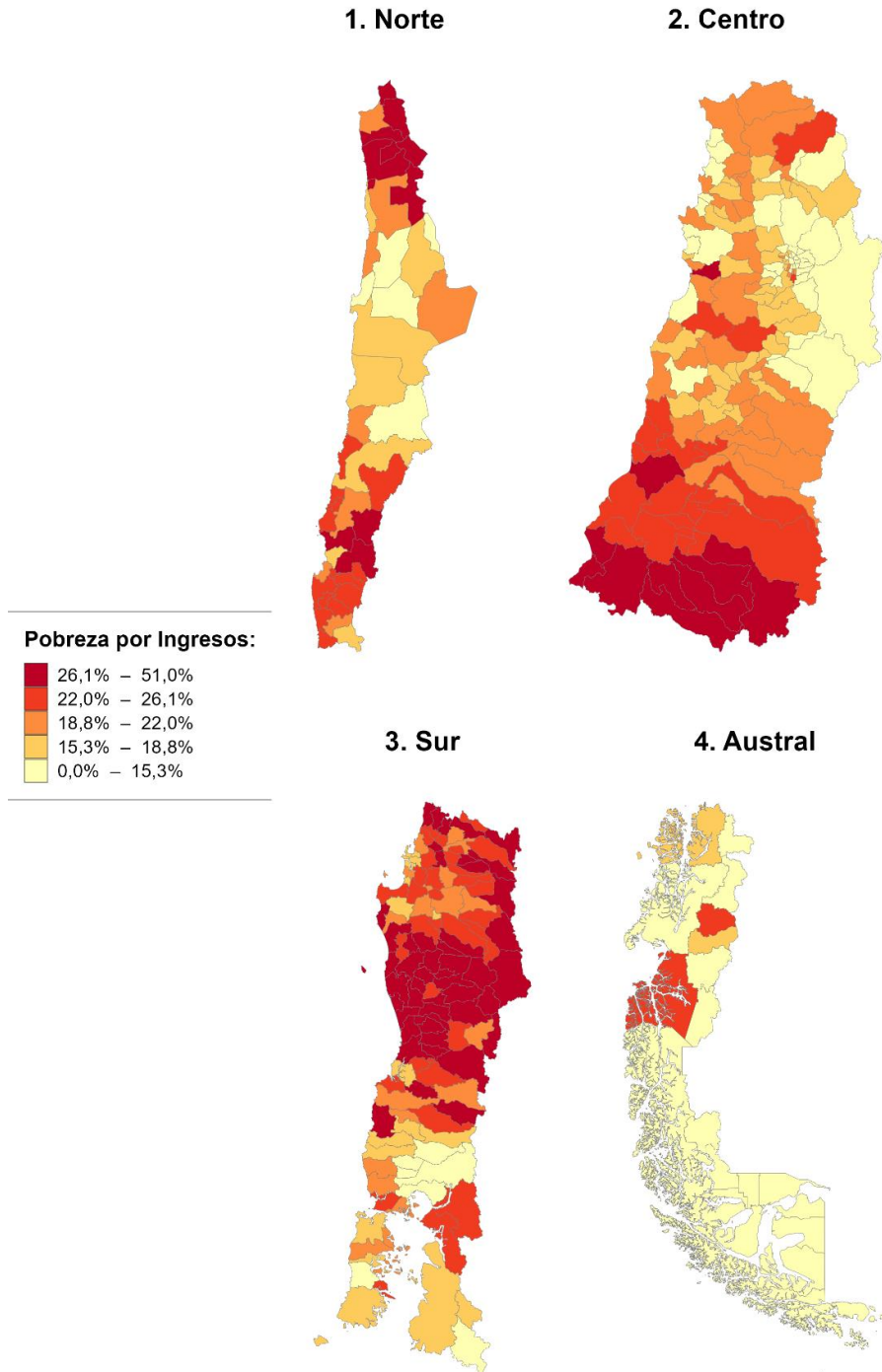
Es importante destacar que, para obtener estimaciones nacionales, regionales y por área urbana/rural de la pobreza por ingresos y el índice de pobreza multidimensional se debe utilizar la Encuesta de Caracterización Socioeconómica Nacional (Casen). La metodología SAE fue diseñada para responder a la necesidad creciente de tener datos comunales, pero no reemplaza las estimaciones directas de la Encuesta Casen en las desagregaciones para las cuales la encuesta tiene representatividad.

Asimismo, se destaca que para cada versión de las estimaciones SAE se hacen mejoras metodológicas y se ajusta el mejor modelo posible para la estimación sintética, dado esto, las estimaciones SAE no son comparables entre distintos años.

Finalmente, en los anexos se presenta una aplicación de las estimaciones SAE mediante la generación de mapas de pobreza. Estos representan las estimaciones comunales de la tasa de pobreza por ingresos y el índice de pobreza multidimensional utilizando la metodología SAE en base a Casen 2024.

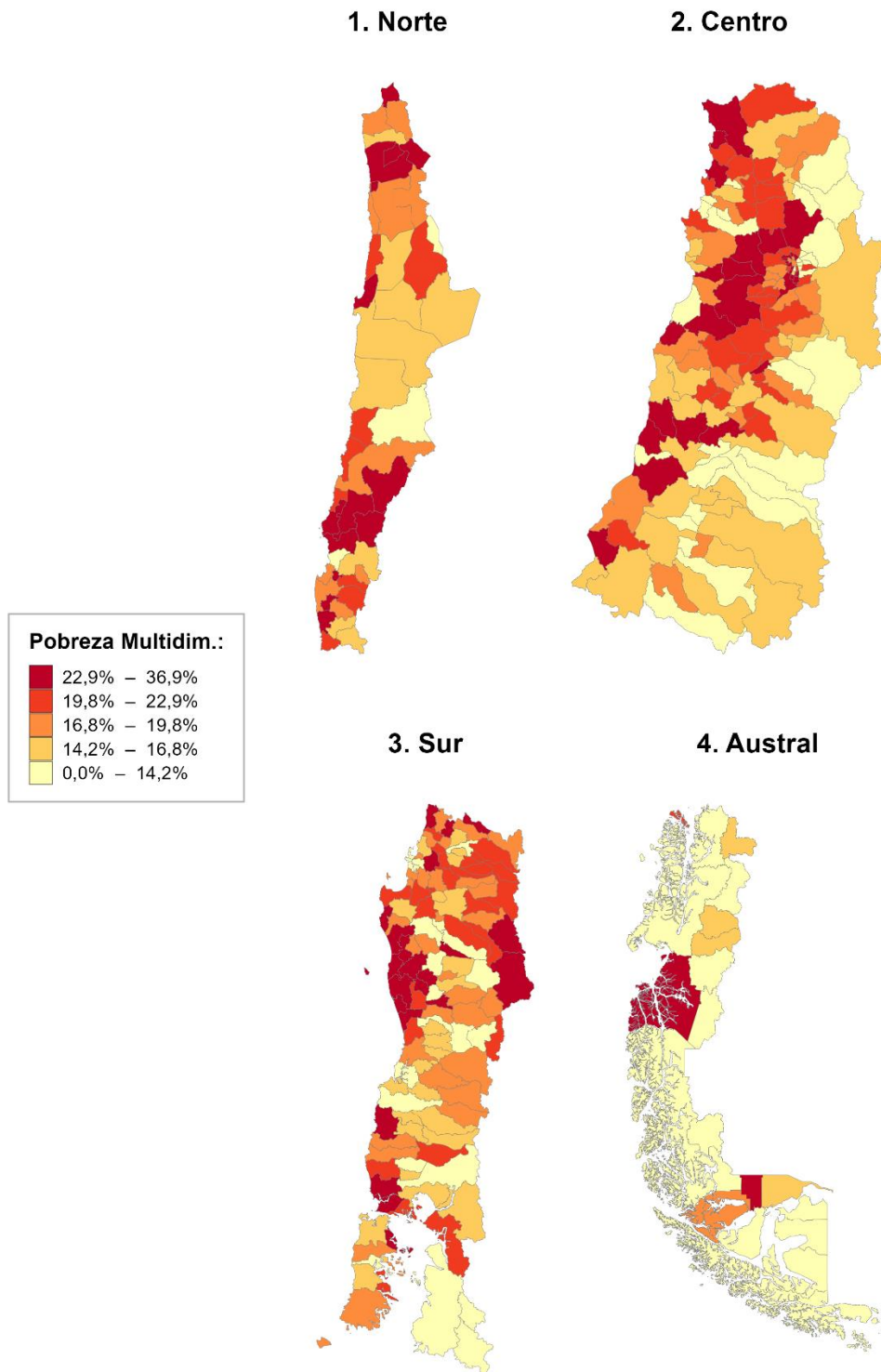
ANEXOS

Anexo 1: Estimación comunal pobreza por ingresos Casen 2024



Fuente: Ministerio Desarrollo Social y Familia en base a Casen 2024.

## Anexo 2: Estimación comunal índice pobreza multidimensional Casen 2024



Fuente: Ministerio Desarrollo Social y Familia en base a Casen 2024.