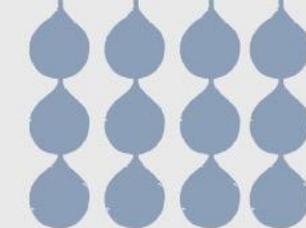


Estimación de áreas pequeñas para la pobreza comunal en Chile

SEMINARIO REGIONAL SOBRE METODOLOGÍAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS
SAO PAULO
5 y 6 de junio 2023



Subsecretaría de Evaluación Social
Ministerio de Desarrollo Social y Familia
Junio 2023

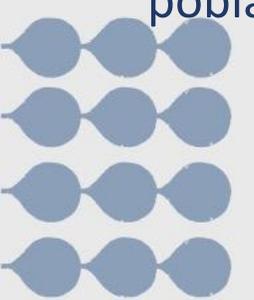


Observatorio Social: Objetivos

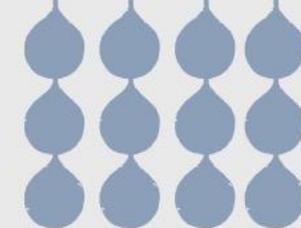
General:

Caracterizar y analizar la realidad social, a través de la recopilación, procesamiento, análisis y difusión de datos, de modo de identificar las necesidades sociales de la población para el oportuno diseño de políticas sociales.

Específicos:

1. Analizar la realidad social nacional y regional, de modo de detectar las necesidades sociales de la población para el oportuno diseño de políticas sociales.
 2. Estudiar y **proponer metodologías** para la recolección y procesamiento de información de encuestas sociales.
 3. Sistematizar y analizar registros de datos, información y estadísticas que permitan caracterizar la población nacional, regional y comunal y sus necesidades sociales.
- 

Proyectos Destacados



Encuesta de Caracterización Nacional (1987-2022)



Nueva encuesta Bifásica de CASEN, medición Bienestar Social



4 rondas encuesta. Busca medir consecuencias de la pandemia



Encuesta Nacional sobre Actividades de niños, niñas y adolescentes



Encuesta Longitudinal de Primera Infancia



Rol de Coordinación Sub Comisión Social del Grupo Técnico de Indicadores



Encuesta Nacional de Discapacidad y Dependencia



Actualización mensual de CBA y Línea de la Pobreza en base a IPC



Estimación de pobreza comunal, metodología SAE



Sistema de Indicadores en base a registros administrativos



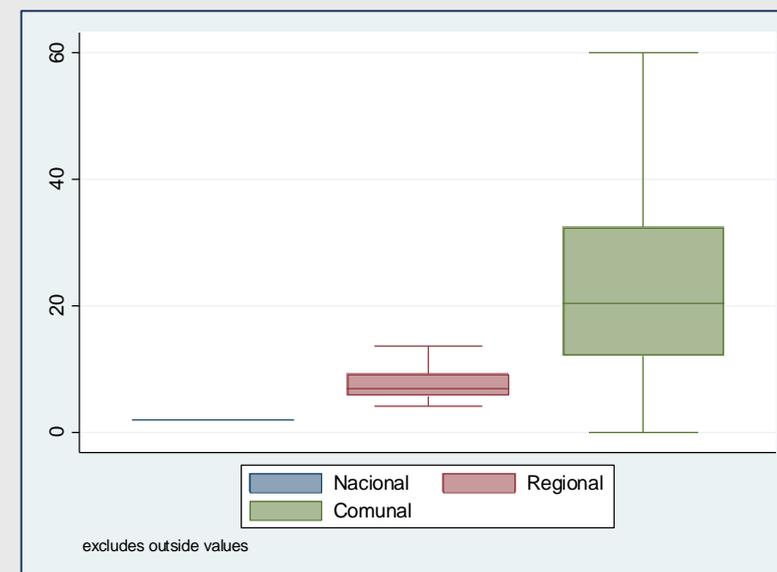
Portal de visualización de datos. Casen, RRAA, BIPS, Donaciones Sociales, RSH

ENCUESTA CASEN

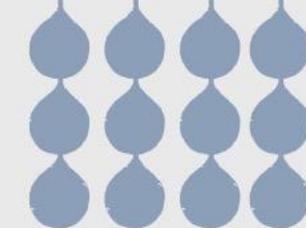
- La encuesta Casen es la fuente oficial para la estimación de las cifras de pobreza a nivel nacional y regional.
- Esta encuesta se ha realizado cada 2 o 3 años desde 1987.
- En promedio se entrevista a cerca de 70.000 hogares, con presencia en 324 de las 345 comunas del país.
- Las estimaciones de la tasa de pobreza a nivel comunal son utilizadas para la asignación de recursos públicos entre las comunas (Fondo Común Municipal), por lo que es fundamental su calidad y precisión.
- La comuna no es un dominio de estudio de la encuesta, por lo que estimaciones realizadas a dicha escala de desagregación pueden tener niveles no adecuados de precisión.



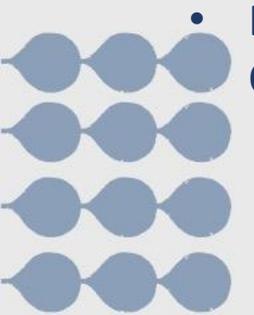
Coefficiente de Variación Pobreza por Ingresos para distintos niveles de desagregación. Casen Pandemia 2020



Fuente: Ministerio Desarrollo Social y Familia.



MOTIVACIÓN

- El año 2010 se inició una consultoría internacional con el Dr. Partha Lahiri y PNUD, para estimar la pobreza por ingresos a nivel comunal.
 - La metodología utilizada corresponde a un modelo de área en base a Fay-Herriot
 - Estas cifras se publican desde Casen 2011.
 - Esta metodología se eligió por la ventaja comparativa del Ministerio al tener acceso a registros administrativos y ser el encargado de levantar la encuesta Casen.
 - Mediante un convenio de colaboración realizado con CEPAL, durante los años 2019 y 2020 se realizó una actualización de la metodología utilizada por el Ministerio, destacando:
 - Indicadores de calidad para evaluar las estimaciones directas de Casen que son utilizadas como insumo en las estadísticas SAE
 - Estimación de la varianza de estimaciones directas utilizando la FGV
 - Estimación de Intervalos de confianza utilizando la estimación del ECM mediante Bootstrap
 - El documento con la revisión de resultados de esta revisión se puede obtener desde el sitio web del Observatorio Social ([link](#))
- 

MODELO DE ESTIMACIÓN

- El modelo de áreas de Fay-Herriot (1979) plantea una estimación en 2 niveles:

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

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

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

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

- A partir de estas 2 ecuaciones se especifica el modelo de estimación de área de F-H:

- $\hat{\mathbf{F}}_d^{FH} = \hat{\mathbf{Y}}_d \cdot \hat{Y}_d^{dir} + (1 - \hat{\mathbf{Y}}_d) \cdot \hat{Y}_d^{syn}$, con $d = 1, \dots, D$

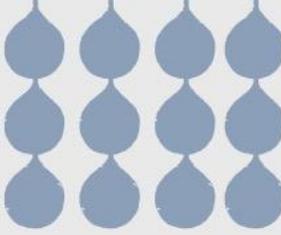
Donde:

$\hat{Y}_d^{syn} = \mathbf{x}_d' \hat{\boldsymbol{\beta}}$, corresponde a la estimación sintética usando variables auxiliares

$$\hat{\mathbf{Y}}_d = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \hat{\psi}_d)$$

- Los parámetros de interés que requieren ser estimados son: $(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\psi}_d)$

MODELO DE ESTIMACIÓN



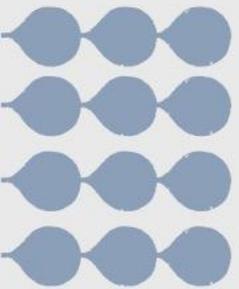
- La estimación SAE pondera dos estimaciones:

Estimación sintética: Utiliza registros administrativos y del Censo de Población para estimar un modelo de regresión que explique la pobreza comunal.

- Información disponible, actualizada y de calidad
- Para los niveles de desagregación de interés
- Relacionadas con el fenómeno que se busca estimar.

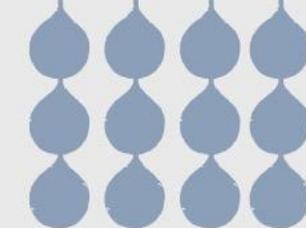
Estimación directa: Corresponde a la estimación que proviene de Casen. En los dominios de la encuesta asume insesgamiento y consistencia. A nivel comunal debe evaluarse:

- Tamaños muestrales apropiados que entreguen estimaciones con cierto nivel de precisión.



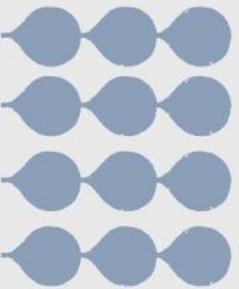
MODELO DE ESTIMACIÓN

- Las estimaciones de pobreza por ingresos que realiza el Ministerio involucran los siguientes pasos:
 - Transformación arcoseno de las tasas de pobreza directas
 - Generación de base de datos administrativos para la estimación sintética
 - Suavización de factores de expansión (modelo Potter-1993)
 - Calidad de las estimaciones directas
 - Estimación varianza directa usando Función Generalizada de Varianza ($\hat{\psi}_d$)
 - Selección del modelo sintético ($\hat{\beta}, \hat{\sigma}_u^2$)
 - Transformación inversa de estimaciones y cálculo de IC en base al ECM usando Bootstarp
 - Calibración (*benchmark*) con estimaciones regionales
 - Validación del modelo
- Hasta las estimaciones SAE 2017 el Ministerio utilizó una programación propia que realizaba todos los pasos. 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.



CALIDAD DE LAS ESTIMACIONES - RRAA

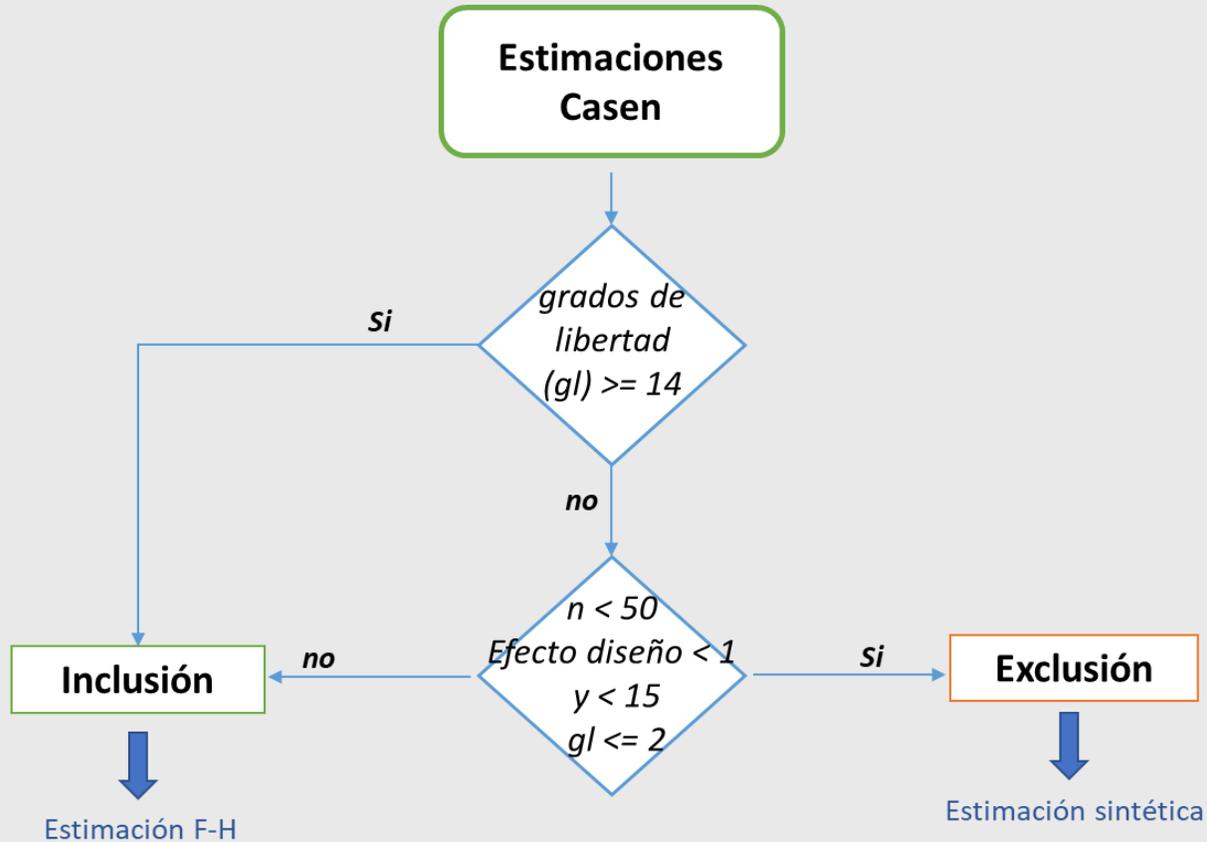
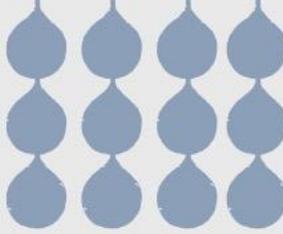
- Para el modelo sintético de las estimaciones SAE que realiza el Ministerio de Desarrollo Social y Familia, se cuenta con información de registros administrativos que levanta el sector público a nivel de comuna. Además, se dispone de la información del Censo de Población.
- Una vez definido el conjunto de variables disponibles, se analiza el poder explicativo de estas variables ajustando un modelo de regresión para obtener la estimación sintética.
- Un tema importante a destacar es que en cada ejercicio SAE, se ajusta el mejor modelo explicativo, por lo que las variables seleccionadas cambian en cada ejercicio, no permitiendo una comparación en el tiempo. Esto se ha hecho así para priorizar una buena estimación cada año, dado el uso en la asignación de recursos.



CALIDAD DE LAS ESTIMACIONES - Directa

- En el caso de las estimaciones directas de Casen, se desarrollaron criterios de calidad que permiten identificar las comunas para las cuales la estimación de Casen es un buen insumo para incorporar a las estimaciones de F-H.
- Estos indicadores son los mismos que utiliza la Oficina Nacional de Estadísticas (INE) para definir publicación de estadísticas, pero fueron adaptados a los objetivos de SAE.
- Los indicadores fueron seleccionados como una medida de:
 - Resultados del trabajo de campo en comparación con el diseño inicial
 - Calidad de la estimación comunal
- Se eligen 4 indicadores:
 - Tamaño muestral (n)
 - Grados de libertad (gl)
 - Número de personas en situación de pobreza (y)
 - Efecto diseño (deff)

CALIDAD DE LAS ESTIMACIONES - Directa

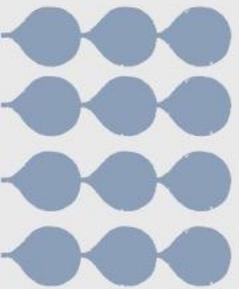


Número de comunas según los criterios de inclusión

Casen	Incluidas	Excluidas	Total comunas
2011	232	92	324
2013	279	45	324
2015	234	90	324
2017	242	82	324
2020	256	68	324

Fuente: Ministerio Desarrollo Social y Familia

Las comunas que no pasan los criterios de inclusión o que no tienen muestra Casen se les asigna la estimación sintética.



CALIDAD DE LAS ESTIMACIONES - Directa

- Ejemplo de 2 comunas que para 2017 su estimación arrojó 0% de pobreza, pero cuyos indicadores de calidad entregan conclusiones diferentes.

Casen	Comuna	p	n	gl	y	deff	tamaño efectivo	número UPM	número estratos	Inclusión
2009	Camiña	12,3%	416	5	50	10	41	6	1	1
2011	Camiña	3,4%	132	2	12	1	185	3	1	0
2013	Camiña	31,1%	103	2	43	1	148	3	1	0
2015	Camiña	40,8%	132	1	54	0	5372	2	1	0
2017	Camiña	0,0%	82	1	0			2	1	0
2020	Camiña	24,5%	86	2	27	1	74	3	1	0
2009	Vitacura	1,9%	359	15	7	4	94	16	1	1
2011	Vitacura	0,0%	375	14	0			15	1	1
2013	Vitacura	0,3%	477	32	2	3	186	33	1	1
2015	Vitacura	0,0%	550	37	0			41	4	1
2017	Vitacura	0,0%	867	87	0			92	5	1
2020	Vitacura	1,5%	659	62	9	8	80	67	5	1

Fuente: Ministerio Desarrollo Social y Familia.

Estimación de la varianza de la estimación directa utilizando FGV

- La variable objetivo de la estimación es la **proporción de personas en pobreza**.

- La estimación de proporciones tiene 2 consideraciones:

1. **Endogeneidad:** La varianza de una proporción depende del parámetro de interés (tasa de pobreza)

2. **Variabilidad** de la varianza estimada en áreas de pequeño tamaño muestral

Arcsin transformation
(Jiang, Lahiri, Wan & Wu, 2001)

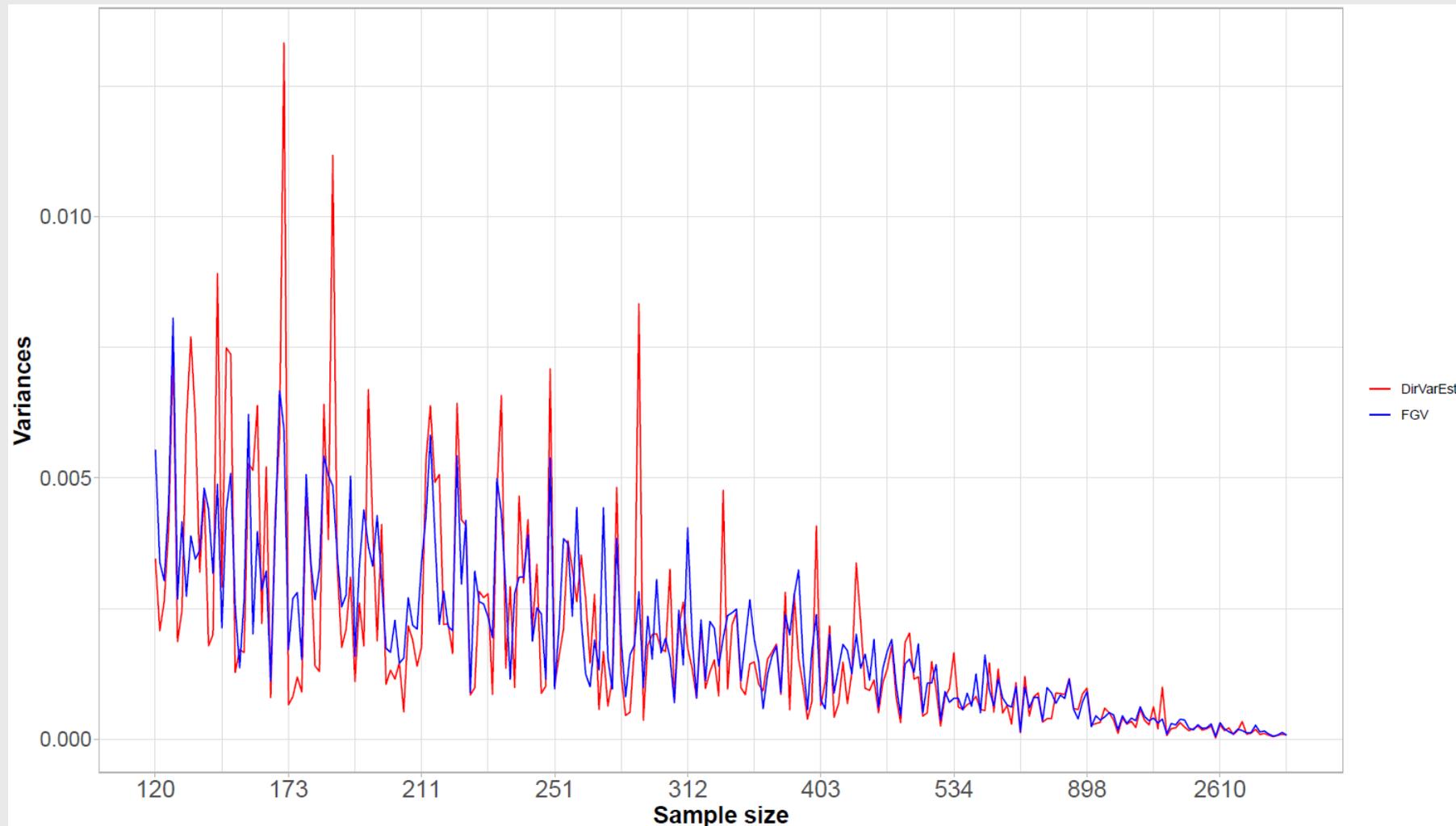
$$\hat{z}_d = \arcsin \left(\sqrt{\hat{\theta}_d^{Dir}} \right)$$

$$Var_p(\hat{z}_d) = \frac{DEFF_d}{4 \times n_d}$$

1. Generalized Variance Function
(Hidioglou, 2019; Wolter, 2007)

GVFs tend to smooth out the noisiness that may be present in the underlying data.

Estimación de la varianza de la estimación directa utilizando FGV



Fuente: Ministerio Desarrollo Social y Familia en base a Casen Pandemia 2020.

Estimación Mean Squared Error (MSE)

- Para estimación 2020 se implementó una estimación de Bootstrap que permite obtener intervalos par todas las comunas.
- Se incorporó la corrección de sesgo por el uso de arco seno y transformación a escala original (Hadam, Würz & Kreutzmann (2020)).
- Las estimaciones se calcularon usando el paquete de R **emdi** (Kreutzmann et al., 2018).

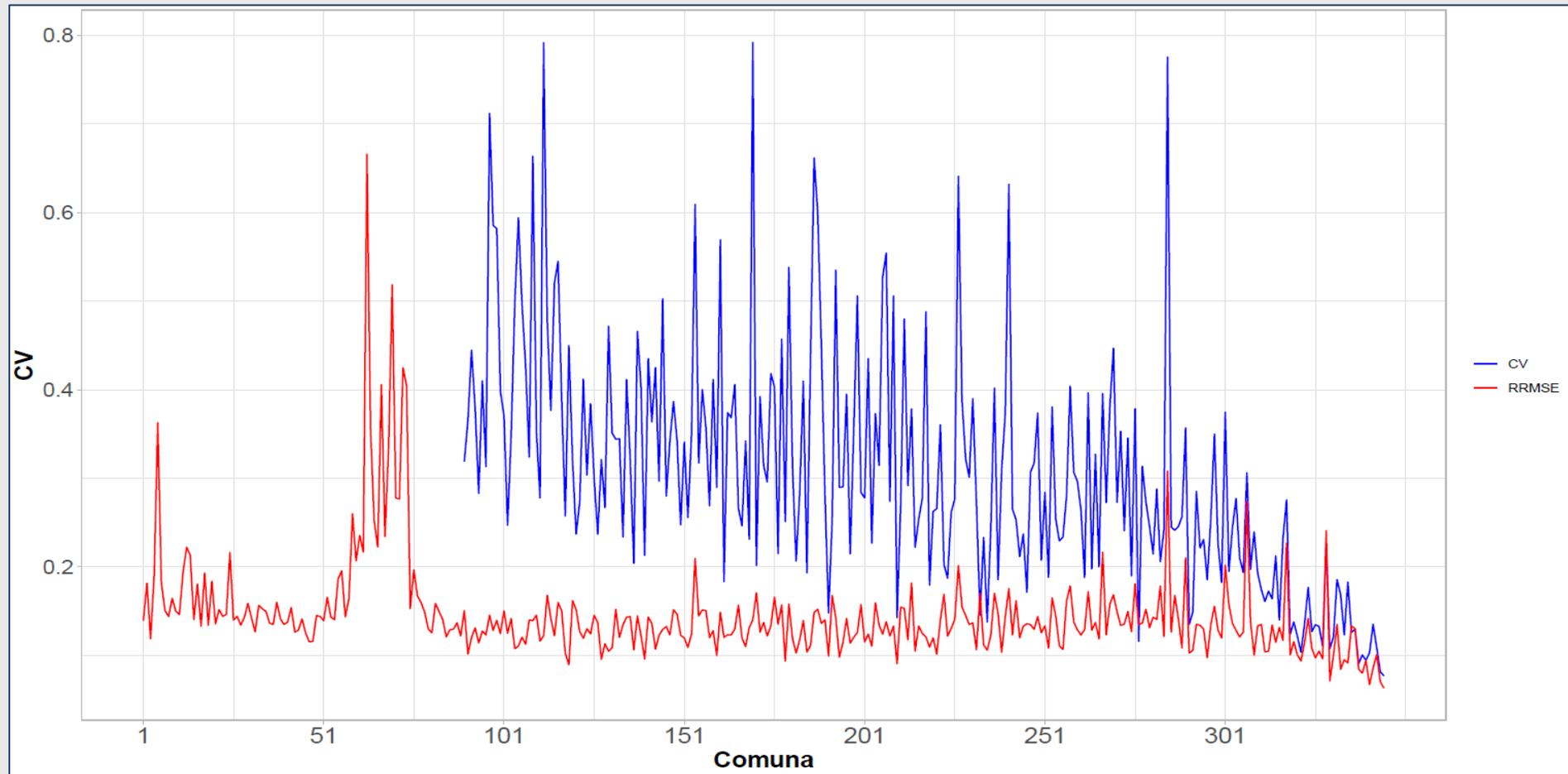
$$MSE_B(\hat{\theta}_d^{FH,trans}) = \frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}_{d,(b)}^{FH,trans} - \theta_{d,(b)}^{trans} \right)^2$$

- Con la estimación del MSE se puede obtener una medida similar al coeficiente de variación (Relative Root Mean Squared error), que permita evaluar la calidad de las estimaciones F-H.

$$RRMSE = \widehat{CV} = \frac{\sqrt{MSE_B(\hat{\theta}_d^{FH,trans})}}{\hat{\theta}_d^{FH,trans}} * 100$$

Direct CV versus Relative RMSE

Estimaciones del CV de la estimación directa y SAE

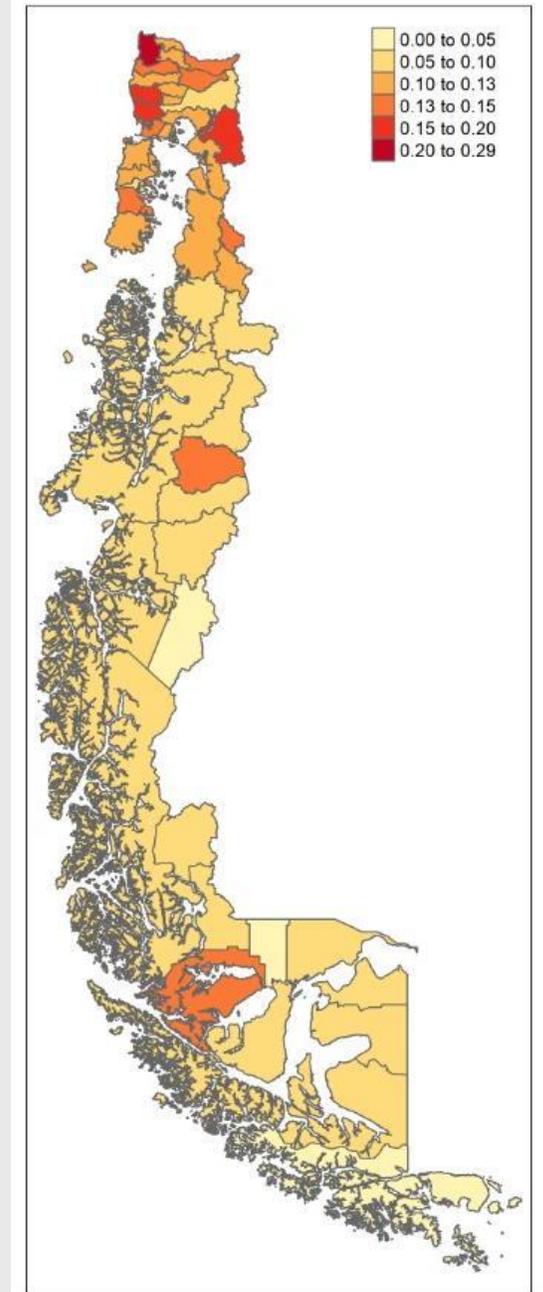
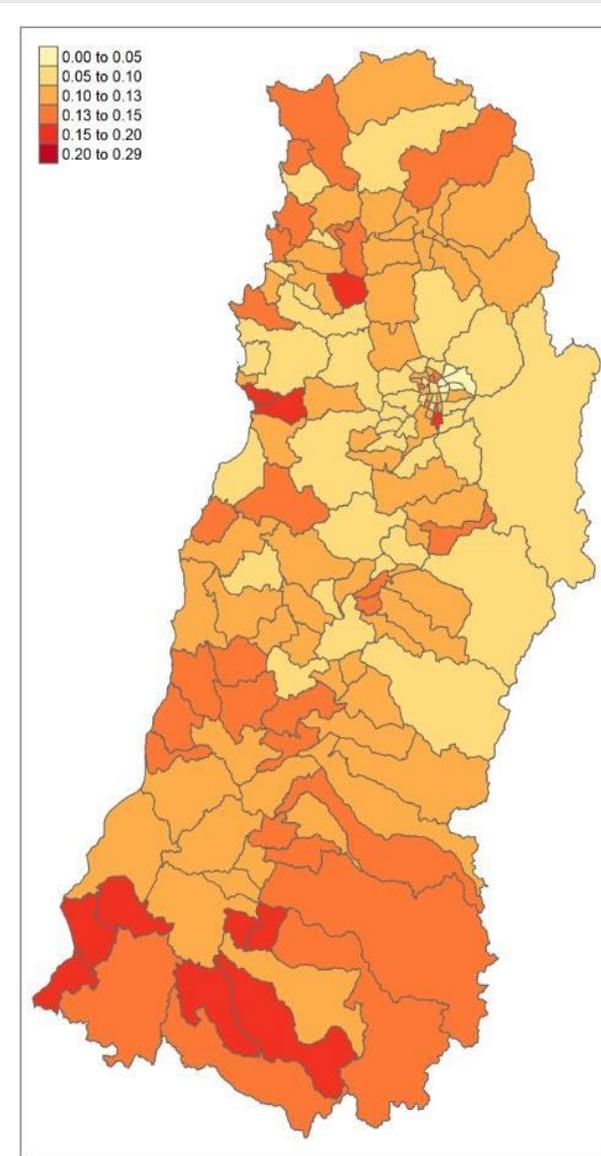
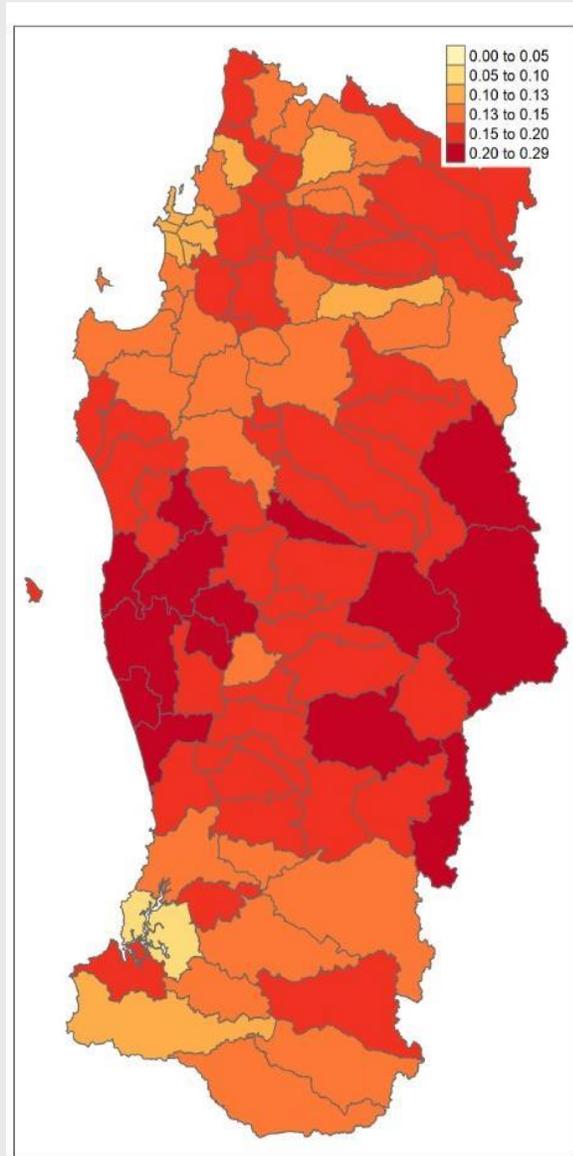
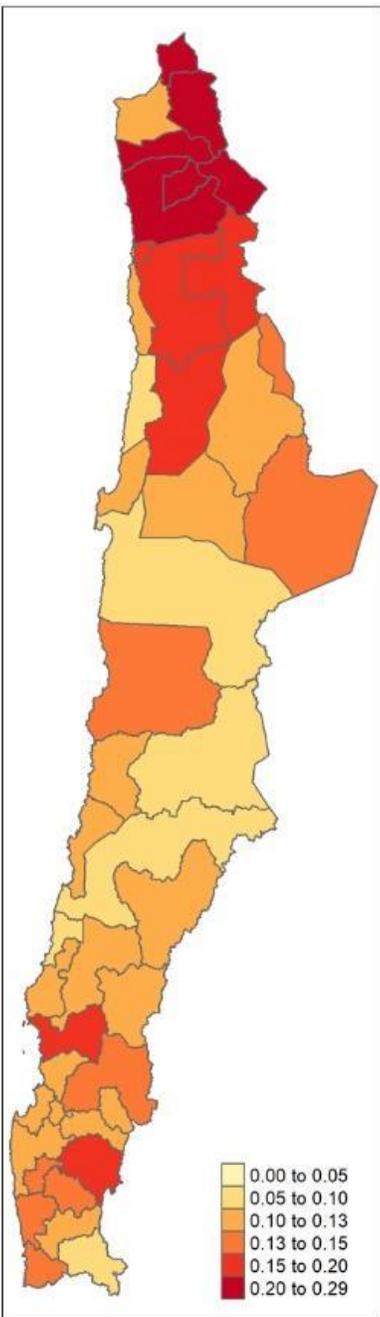


Fuente: Ministerio Desarrollo Social y Familia en base a Casen Pandemia 2020.

Próximos Avances

- Revisión de los resultados del trabajo de campo de Casen 2022:
 - Cambios en el diseño muestral
 - Cambio de marco muestral
- Publicación de estimaciones de pobreza por ingresos y pobreza multidimensional en diciembre 2023.
- Trabajo en conjunto con FAO y Cepal para obtener estimaciones a nivel de comuna de inseguridad alimentaria.
- Publicación de reportes territoriales utilizando registros administrativos y pobreza SAE.
- Evaluación de posible estimación de otros indicadores asociados a grupos prioritarios como infancia y discapacidad.

RESULTADOS



REFERENCIAS

- Hadam, Sandra, Würz, Nora, & Kreuzmann, Ann-Kristin (2020), Estimating regional unemployment with mobile network data for Functional Urban Areas in Germany. Working paper. Freie Universität.
- Hidiroglou, Michel A. 2019. “Development of a Small Area Estimation System at Statistics Canada.” *Survey Methodology* 45 (1): 101–26.
- Jiang, J., Lahiri, P. y Wan, S. (2002), A unified jackknife theory for empirical best prediction with Mestimation. *The Annals of Statistics*. 30 (6), 1782-1810, <https://doi.org/10.1214/aos/1043351257>
- Molina, I. (2019), Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas, *Series Estudios Estadísticos*, No 97, (LC/TS.2018/82/Rev.1), Santiago, Comisión Económica para América Latina y el Caribe, (CEPAL).
- Prasad, N. N. y Rao, J. N. (1990), The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163–171.
- Rao, J.N.K. y Molina (2015), *Small area estimation*, Second Ed., Hoboken, NJ: Wiley.