

Documento de trabajo

Estimación en áreas pequeñas: una aplicación como caso de estudio en la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC 2018)

Autores:

Andrea Marín García
Millaray Tralma Herrera
Julio Guerrero Rojas

Octubre de 2024



INSTITUTO NACIONAL DE ESTADÍSTICAS

Morandé 801, Santiago de Chile

Teléfono: 56 232461010

Correo: ine@ine.cl

Facebook: [@ChileINE](#)

X: [@INE_Chile](#)

Departamento de Metodologías e Innovación Estadística

Andrea Marín García

Millaray Tralma Herrera

Julio Guerrero Rojas

Agradecimientos

Los autores agradecen a Andrés Gutiérrez (Experto Regional en Estadísticas Sociales-CEPAL) por la asesoría y acompañamiento brindado en todas las etapas de este proyecto, ya que las capacitaciones recibidas y posteriormente los comentarios y reorientaciones permitieron dar forma a este documento de resultados.

Se agradece al equipo del Departamento de Estudios de la División de Programa y Estudios de la Subsecretaría de Prevención del Delito (SPD) por facilitar parte de la información auxiliar, la cual es punto central para el desarrollo de este tipo de trabajo de investigación; así como de participar en diferentes sesiones de revisión de resultados preliminares que permitieron realizar ajustes pertinentes a la metodología aplicada. También se agradecen las oportunidades de futuras mejoras, planteadas a partir de la instancia de intercambio de conocimiento con el equipo técnico de la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública del Instituto Nacional de Estadística y Geografía (INEGI, México).

Finalmente, a nivel del Instituto Nacional de Estadísticas (INE) se extiende el agradecimiento al equipo de Sistemas de Información Geográfica (SIG) del Subdepartamento de Geografía, quienes diseñaron la visualización interactiva de los resultados, la cual consideramos clave e indispensable para la diseminación efectiva de este tipo de resultados, y al equipo de la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) del Subdepartamento de Condiciones de Vida por la colaboración en el análisis de los resultados, que permitió vislumbrar diferentes instancias de mejora para trabajos futuros; y finalmente, al equipo del Subdepartamento de Investigación Estadística (SIE) del Departamento de Metodologías e Investigación Estadística (DMIE), por la revisión y sugerencias realizados a este documento.

Los Documentos de Trabajo del INE están dirigidos a investigadores, académicos, estudiantes y público en general, y tienen como objetivo proporcionar un análisis exhaustivo sobre aspectos conceptuales, analíticos y metodológicos claves de los productos estadísticos que elabora la institución, y de esta forma, contribuir al intercambio de ideas entre los distintos componentes del Sistema Estadístico Nacional.

Las interpretaciones y opiniones que se expresan en los Documentos de Trabajo pertenecen en forma exclusiva a los autores y colaboradores, y no reflejan necesariamente el punto de vista oficial del INE ni de la institución a la que pertenecen los colaboradores de los documentos.

Estimación en áreas pequeñas: una aplicación como caso de estudio en la Encuesta Nacional Urbana de Seguridad Ciudadana 2018

Resumen

La Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) ha proporcionado estimaciones anuales de victimización durante casi dos décadas, centrándose en los resultados de las áreas urbanas a nivel nacional y regional. Sin embargo, se requiere realizar esfuerzos para producir estimaciones para las comunas urbanas en Chile. Las estimaciones directas sobre victimización a nivel comunal se pueden obtener de la ENUSC, pero no serían confiables debido al pequeño tamaño de las muestras. Para abordar este problema, se utiliza una metodología de Small Area Estimation (SAE) basada en un modelo Fay-Herriot para mejorar la precisión de las estimaciones comunales, usando los datos de registros policiales y del Censo de Población y Vivienda 2017.

En este documento, se informan los resultados de este primer esfuerzo, específicamente para tres indicadores de victimización en la versión 2018 de la encuesta, a saber: robo con violencia o intimidación, robo por sorpresa y victimización agregada por delitos de mayor connotación social. La metodología propuesta produce estimaciones de prevalencia que son consistentes con los tamaños de muestra y la dinámica demográfica en Chile. Los coeficientes de variación logarítmicos logrados son en general inferiores al 20%, por lo tanto, las estimaciones pueden considerarse confiables, sin embargo, debe atenderse que es importante la validación de los resultados, no solamente desde una perspectiva de eficiencia estadística, sino también evaluando cuidadosamente su consistencia con otras fuentes de información tales como registros administrativos y encuestas, para garantizar así que reflejen de manera precisa la realidad comunal. Finalmente, a modo de lecciones para trabajos futuros, se destaca la importancia de estudiar la movilidad del delito, el esfuerzo por obtener información auxiliar de calidad para fortalecer los modelos utilizados e implementar una metodología adecuada para realizar las evaluaciones y validaciones de las estimaciones SAE.

Palabras clave: ENUSC, Modelo Fay – Herriot, Victimización.

Abstract

The National Urban Citizen Security Survey (ENUSC) has provided annual estimates of crime victimizations for nearly two decades, mainly focusing on national and regional results. However, efforts are required to produce estimates for urban communes in Chile. Direct estimates of victimization at the community level can be obtained from the ENUSC, but they would not be reliable due to the small size of the samples. To address this problem, a Small Area Estimation (SAE) methodology based on a Fay-Herriot model is used to improve the quality of communes' estimates, using data from police records and the 2017 population and housing census.

In this working paper, the results from this first effort are reported, specifically for three victimization indicators in the 2018 version of the survey, namely: assault with violence and intimidation, surprise robbery and High Social Connotation Crime. The proposed methodology produces prevalence estimates that are consistent with sample sizes and demographic dynamics in Chile. Almost all the logarithmic variation coefficients achieved are less than 20%, therefore, the estimates can be considered reliable, however, it should be noted that it is important to validate the results, not only from a statistical efficiency perspective, but also by carefully assessing their consistency with other sources of information such as administrative records and surveys, to ensure that they accurately reflect the reality of communes. Finally, as lessons for future work, the importance of studying crime mobility, the effort to obtain quality auxiliary information to strengthen the models and implement an adequate methodology to carry out evaluations and validations of SAE estimates are highlighted.

Key words: ENUSC, Fay – Herriot model, SAE, Victimization, SAE.

Contenido

1. Introducción	10
2. Acerca de la ENUSC 2018	11
3. Estimación en áreas pequeñas.....	12
3.1. Modelo a nivel de áreas	13
3.2. Suavizamiento de las varianzas directas	15
3.3. Transformación arcoseno	16
3.4. Ajustes al modelo EBLUP.....	17
4. Evaluación de las estimaciones de áreas pequeñas.....	18
4.1. Tratamiento de valores atípicos e influyentes	19
4.2. Estimación del error cuadrático medio	20
4.3. Coeficiente de variación.....	21
4.4. Bondad de ajuste	21
4.5. Benchmarking	22
5. Aplicación en la ENUSC 2018	23
5.1. Estimadores directos	23
5.2. Información auxiliar	25
5.3. Etapas generales en la adaptación del modelo.....	27
5.3.1. Suavizamiento de varianzas directas.....	28
5.4. Modelo a nivel de áreas	30
5.4.1. Selección de variables.....	31
5.4.2. Ajustes de modelos sobre indicadores de victimización para la estimación de áreas pequeñas	36
5.4.3. Estimaciones de áreas pequeñas para los indicadores de victimización a nivel comunal	40
5.5. Evaluación de la fiabilidad de las estimaciones	44
5.6. Diagnósticos de los modelos.....	47
5.7. Análisis de los resultados comunales	51
5.7.1. Comparación con resultados ENUSC.....	51
5.7.2. Contraste con registros administrativos.....	59
5.7.3. Evaluación de características comunales	63
5.8. Reflexiones sobre modelamiento de la tasa de victimización de hogares por robo desde vehículo	67
6. Conclusiones y trabajo futuro	71
7. Referencias	73

8. ANEXOS	79
8.1. Anexo A: Preparación de datos auxiliares	79
8.2. Anexo B: Áreas o dominios de interés	81
8.3. Anexo C: Opciones probadas con los modelos SAE	83
8.4. Anexo D: Indicadores comunales	85
8.5. Anexo E: Denuncias comunales por región.....	88

Índice de tablas

Tabla 1. Covariables relacionadas con la victimización.....	25
Tabla 2. Covariables en modelos de función de varianza generalizada.	29
Tabla 3. Covariables en modelos SAE para RVI.....	37
Tabla 4. Covariables en modelos SAE para RPS.	38
Tabla 5. Covariables en modelos SAE para VA.	39
Tabla 6. Resumen de estimaciones de áreas pequeñas de indicadores de victimización y CVL promedio.	45
Tabla 7. Resumen de test de normalidad y de homocedasticidad.	47
Tabla 8. Covariables en modelos SAE para robo desde vehículo.....	68
Tabla A. 9. Covariables para construcción de modelos EBLUP en ENUSC 2018.....	79
Tabla A. 10. Áreas o dominios de interés.	81
Tabla A. 11. Indicadores de comunas que integran el marco de ENUSC.	85

Índice de figuras

Figura 1. Covariables en plano principal (primera y segunda componentes principales) para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior).....	33
Figura 2. Covariables con mayor poder predictivo para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior), según el algoritmo de Random Forests.....	34
Figura 3. Estimaciones comunales para RVI.	41

Figura 4. Estimaciones comunales para RPS.....	41
Figura 5. Estimaciones comunales para VA.....	42
Figura 6. Estimación directa vs. Estimación EBLUP para la tasa de victimización por RVI.	42
Figura 7. Estimación directa vs. Estimación EBLUP para tasa de victimización por RPS. .	43
Figura 8. Estimación directa vs. Estimación EBLUP para tasa de VA.....	43
Figura 9. Estimaciones RMSE comunales para tasa de victimización por RVI en (comunas ordenadas por tamaño muestral) en hogares.....	45
Figura 10. Estimaciones RMSE comunales para tasa de victimización por RPS (comunas ordenadas por tamaño muestral) en hogares.....	46
Figura 11. Estimaciones RMSE comunales para tasa de VA (comunas ordenadas por tamaño muestral) en hogares.	46
Figura 12. Gráficos q-q plots para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior).	48
Figura 13. Gráficos de residuos estandarizados versus valores predichos para para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior).	49
Figura 14. Distancias de Cook para para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior).	50
Figura 15. Estimaciones SAE y ENUSC del indicador VA para las regiones de la zona norte.	52
Figura 16. Estimaciones SAE y ENUSC del indicador VA para la región de Valparaíso.....	53
Figura 17. Estimaciones SAE y ENUSC del indicador VA para las regiones de la zona centro y centro sur.	54
Figura 18. Estimaciones SAE y ENUSC del indicador VA para la región del Biobío.	56
Figura 19. Estimaciones SAE y ENUSC del indicador VA para las regiones de la zona sur.	56
Figura 20. Estimaciones SAE y ENUSC del indicador VA para la Región Metropolitana...	58
Figura 21. Correlación de denuncias por DMCS registradas por policías y estimación SAE de VA.	59

Figura 22. Estimaciones SAE para VA y Tasa de denuncia para DMCS, Región de Coquimbo.	60
Figura 23. Estimaciones SAE para VA y Tasa de denuncia para DMCS, Región de Valparaíso.	61
Figura 24. Estimaciones SAE para VA y Tasa de denuncia para DMCS, Región de Biobío.	62
Figura 25. Estimaciones SAE para VA y Tasa de denuncia para DMCS, Región Metropolitana.	63
Figura 26. Porcentaje de hogares en la muestra que fueron víctima de algún DMCS dentro de su propia comuna, Región Metropolitana.	66
Figura 27. Diferencia entre lugar de residencia y lugar de ocurrencia de RVI, en la Región Metropolitana.	67
Figura 28. Estimaciones comunales para robo desde vehículo.	70
Figura A. 29. Porcentaje de hogares en la muestra que fueron víctima de algún DMCS dentro de su propia comuna, Región de Coquimbo.....	88
Figura A. 30. Porcentaje de hogares en la muestra que fueron víctima de algún DMCS dentro de su propia comuna, Región de Valparaíso.	88
Figura A. 31. Porcentaje de hogares en la muestra que fueron víctima de algún DMCS dentro de su propia comuna, Región del Biobío.	89

1. Introducción

Desde el año 2003, el Instituto Nacional de Estadísticas (INE), a través de un convenio de colaboración con la Subsecretaría de Prevención del Delito (SPD), realiza la “Encuesta Nacional Urbana de Seguridad Ciudadana” (ENUSC). La encuesta permite contar con información útil tanto para la toma de decisiones, como para generar conocimiento sistemático sobre los niveles de victimización y percepción de inseguridad en Chile.

El objetivo de la ENUSC es obtener información sobre la percepción de inseguridad, la reacción frente al delito y la victimización de personas y hogares, a partir de una muestra de viviendas particulares ocupadas, que permite producir estadísticas oficiales con un nivel de confiabilidad aceptable sólo para los niveles nacional y regional en la zona urbana del país (Instituto Nacional de Estadísticas - Chile, 2019).

En los últimos años, debido a que estas estadísticas oficiales son utilizadas por las oficinas gubernamentales en la articulación, implementación y evaluación de programas en materia de seguridad ciudadana y distribución de los recursos de las policías en el territorio nacional, la demanda de estimaciones sobre victimización en niveles geográficos más desagregados, como el nivel comunal, ha aumentado. Las estimaciones directas para las tasas de victimización en hogares se pueden obtener de la ENUSC, no obstante, la precisión de estas disminuye con el tamaño muestral, lo que impide lograr estimaciones confiables para los niveles de desagregación requeridos; esto es, estimaciones a nivel comunal. Una forma de abordar este problema es utilizar la metodología de estimación en áreas pequeñas (en inglés, *small area estimation* o SAE) con el objetivo de mejorar la calidad de las estimaciones comunales, a través de un sistema inferencial preciso y exacto, construido a partir de la combinación entre los datos de la encuesta con información auxiliar existente a nivel poblacional como censos o registros administrativos. En este caso, se combina la información muestral de la ENUSC 2018 con indicadores sociodemográficos derivados del Censo de Población y Vivienda 2017 (CPV 2017) y registros administrativos sobre casos policiales del 2018.

El propósito de este documento es describir el primer esfuerzo del INE por producir estimaciones comunales para algunos de los indicadores de victimización a hogares medidos por la ENUSC, utilizando un modelo de estimación en áreas pequeñas (SAE). Específicamente, este trabajo se centra en producir estimaciones para tres indicadores en la

versión 2018 de la encuesta, a saber: robo con violencia o intimidación, robo por sorpresa y victimización agregada por delitos de mayor connotación social.

El documento está organizado de la siguiente manera: En la **Sección 2** se describen brevemente las características de la ENUSC 2018. La introducción a la metodología SAE con enfoque en el modelo de área se realiza en la **Sección 3**. La **Sección 4** describe cómo procede la evaluación de las estimaciones obtenidas a partir de los modelos SAE. En la **Sección 5**, se presenta la aplicación en la ENUSC 2018: descripción de la información auxiliar, pasos de modelado, evaluación de los modelos y análisis de los resultados. En la **Sección 6** se dan conclusiones, algunas ideas futuras y un resumen de las opciones probadas con el modelo SAE.

2. Acerca de la ENUSC 2018

Las estimaciones comunales se elaboran a partir de los datos de la ENUSC 2018. La ENUSC es una encuesta nacional que se ha realizado en Chile desde el 2003, y está diseñada para estudiar la proporción de personas y hogares que han sido víctimas de delitos, así como la percepción de inseguridad en la población. Se utilizó la versión 2018 en lugar de una más reciente con el fin de realizar este estudio en un contexto más típico¹ del fenómeno registrado en la encuesta, donde se realizaron análisis exhaustivos para garantizar la calidad de las estimaciones muestrales.

La ENUSC es una encuesta anual que utiliza un diseño muestral trietápico², esto ya que se seleccionan tres unidades: (1) manzanas, (2) hogares particulares ocupadas en áreas urbanas, y (3) persona de 15 años o más (informante Kish). Esta encuesta fue diseñada para proporcionar estimaciones con precisión confiable a nivel nacional y regional en el área urbana, pero todos los niveles geográficos inferiores son dominios (áreas) no planificados.

El objetivo de este trabajo es explorar, por primera vez en la institución, el uso de la teoría de estimación de áreas pequeñas para producir estimaciones comunales para las tasas de tres indicadores de victimización en hogares calculados en la encuesta, a saber: robo con

¹ En 2019, el levantamiento de la encuesta se vio afectado por el “Estallido Social”, y entre 2020 y 2022, por las medidas adoptadas debido al covid-19.

² Para más detalles, véase documento metodológico de diseño muestral XV ENUSC, disponible en https://www.ine.cl/docs/default-source/seguridad-ciudadana/metodolog%C3%ADa/2018/documento-metodol%C3%B3gico-de-dise%C3%B1o-muestral-xv-enusc-2018.pdf?sfvrsn=ea63c94e_4

violencia o intimidación, robo por sorpresa y victimización agregada por delitos de mayor connotación social. La selección de estos indicadores se basó en la factibilidad de construir los modelos SAE, con los datos auxiliares disponibles (ver **Subsección 5.2**). Adicionalmente, a modo de reflexión se incluyen algunos de los análisis realizados al tratar de modelar la tasa de victimización en hogares por robo desde vehículo.

Los datos de la ENUSC 2018 han sido proporcionados por los informantes Kish encuestados y comprenden 27.025 casos muestrales, distribuidos en el área urbana de 102 comunas, abarcando las 16 regiones del país.

3. Estimación en áreas pequeñas

Los métodos de estimación basados en modelos asumen un modelo tanto para la población como para los datos de la muestra, y construyen predictores óptimos de los parámetros objetivo bajo su alcance. En este contexto, convencionalmente se utiliza el término “predictor” en lugar de “estimador” debido a que el modelo utilizado supone que los parámetros objetivo son aleatorios.

Como se mencionó anteriormente, los métodos SAE intentan producir estimaciones confiables cuando el tamaño de la muestra en el área o dominio es pequeño. En esta aplicación de la metodología, la estimación de área pequeña es una función de dos cantidades: la estimación directa a partir de los datos de la encuesta y una predicción basada en un modelo, a veces denominada estimación indirecta o sintética. El modelo SAE involucra datos de encuestas del área geográfica de interés, pero también incorpora datos de otras áreas (como entrada a los parámetros del modelo) y datos auxiliares. Es importante considerar que la información auxiliar debe provenir de una fuente independiente de la encuesta y debe estar disponible para todas las áreas en los niveles geográficos requeridos. Identificar qué datos son necesarios para la implementación de un modelo SAE tiene implicancias no sólo en los resultados de la estimación, sino también en la carga de trabajo, puesto que toda información susceptible de uso debe ser previamente revisada y adaptada según las necesidades del modelo. En relación con las fuentes de información auxiliar, hay aspectos importantes que deben ser considerados, a saber:

- (a) Identificación de indicadores o covariables relacionados con los indicadores de victimización que se pretenden estimar.

- (b) Disponibilidad de fuentes de información que provean los indicadores identificados o proporcionar insumos para su construcción.

Hay dos tipos de modelos SAE: modelos a nivel de área (o agregados) que relacionan medidas de área pequeña con variables auxiliares específicas del área, y modelos a nivel de unidad que relacionan los valores unitarios de la variable de estudio con medidas de las variables auxiliares específicas de la unidad (Chatrchi, 2019). Dado que la información auxiliar disponible que guarda mayor relación (datos policiales e indicadores sociodemográficos) con los indicadores de victimización a modelar está agregada a nivel comunal, se utiliza el modelo a nivel de área. La teoría del modelo a nivel de área se describe brevemente en la siguiente subsección.

3.1. Modelo a nivel de áreas

Un modelo de nivel de área básico, conocido como modelo de Fay-Herriot (Fay y Herriot, 1979) consta de dos partes:

- 1. Modelo de muestreo:** $\hat{\theta}_d = \theta_d + e_d$, con $d=1, \dots, D$,

Donde D representa el número total de áreas o dominios, $\hat{\theta}_d$ es el estimador directo de la variable de interés, e_d representa los errores de muestreo, con $E_p(e_d) = 0$ y $\text{var}_p(e_d) = \psi_d$. El subíndice p indica que la esperanza y la varianza se toman con respecto al diseño muestral y se asume que el estimador directo es insesgado bajo el diseño de muestreo, mientras que ψ_d representa la varianza de $\hat{\theta}_d$ con respecto al diseño muestral y comúnmente desconocida. Se denota por $\hat{\psi}_d$ a un estimador directo de ψ_d .

- 2. Modelo de vínculo:** $\theta_d = \mathbf{z}_d^T \boldsymbol{\beta} + b_d v_d$,

donde el vector \mathbf{z}_d representa la información auxiliar específica del área d , v_d el efecto aleatorio del área d con $E_m(v_d | \mathbf{z}_d) = 0$ y $\text{var}_m(v_d | \mathbf{z}_d) = \sigma_v^2$, b_d es una constante positiva, $\boldsymbol{\beta}$ y σ_v^2 son los parámetros desconocidos del modelo. El subíndice m indica que la esperanza y la varianza se toman con respecto al modelo.

Además de los supuestos anteriores, los errores e_d y v_d , $d=1, \dots, D$, se suponen normalmente distribuidos y mutuamente independientes. Como se mencionó anteriormente, D indica el número de áreas utilizadas para el modelado, en este caso 102. Combinando el modelo de muestreo y de vinculación, obtenemos el modelo de Fay-Herriot:

$$\hat{\theta}_d = \mathbf{z}_d^T \boldsymbol{\beta} + a_d, \quad (1)$$

donde $a_d = b_d v_d + e_d$, $E_{mp}(a_d | \mathbf{z}_d) = 0$, $var_{mp}(a_d | \mathbf{z}_d) = b_d^2 \sigma_v^2 + \tilde{\psi}_d$ y $\tilde{\psi}_d = E_m(\psi_d | \mathbf{z}_d)$ es la varianza de diseño suavizada³. El subíndice *mp* indica que la esperanza y la varianza se toman con respecto tanto al modelo como al diseño muestral.

Suponiendo que $\tilde{\psi}_d$ y σ_v^2 son conocidas, la teoría de los modelos lineales mixtos generalizados proporciona un marco para obtener el predictor óptimo para θ_d llamado **mejor predictor lineal insesgado** (Best Linear Unbiased Predictor o **BLUP**, por su sigla en inglés).

El BLUP se puede expresar como la combinación ponderada del estimador directo ($\hat{\theta}_d$) y el estimador “sintético de regresión” ($\mathbf{z}_d^T \tilde{\boldsymbol{\beta}}$):

$$\tilde{\theta}_d = \gamma_d \hat{\theta}_d + (1 - \gamma_d) \mathbf{z}_d^T \tilde{\boldsymbol{\beta}},$$

donde $\tilde{\boldsymbol{\beta}} = [\sum_{d=1}^m \mathbf{z}_d \mathbf{z}_d^T / (b_d^2 \sigma_v^2 + \tilde{\psi}_d)]^{-1} [\sum_{d=1}^m \mathbf{z}_d \hat{\theta}_d / (b_d^2 \sigma_v^2 + \tilde{\psi}_d)]$ y $\gamma_d = \frac{b_d^2 \sigma_v^2}{b_d^2 \sigma_v^2 + \tilde{\psi}_d}$.

Cuando $\tilde{\psi}_d$ y σ_v^2 se desconocen, como suele ocurrir en la práctica, se reemplazan por sus estimadores y se obtiene el **mejor predictor lineal insesgado empírico (EBLUP)**:

$$\hat{\theta}_d^{EBLUP} = \hat{\gamma}_d \hat{\theta}_d + (1 - \hat{\gamma}_d) \mathbf{z}_d^T \hat{\boldsymbol{\beta}}. \quad (2)$$

Existen diferentes procedimientos para estimar σ_v^2 y $\boldsymbol{\beta}$, como la máxima verosimilitud restringida (REML), el procedimiento de Fay-Herriot (FH) como se describe en Fay y Herriot (1979), la maximización de densidad ajustada (ADM) de Li y Lahiri (2010), y el método de ajuste de constantes (método de Henderson). La principal diferencia entre estos métodos es cómo se calcula σ_v^2 , utilizando un algoritmo de *scoring* iterativo. Utilizamos el método REML, que se basa en el siguiente algoritmo de *scoring* (iterativo):

$$\hat{\sigma}_v^{2(a+1)} = \hat{\sigma}_v^{2(a)} + \left[\frac{1}{2} tr(\mathbf{PBPB}) \right]^{-1} \left[\frac{1}{2} \mathbf{y}^T \mathbf{PBP} \mathbf{y} - \frac{1}{2} tr(\mathbf{PB}) \right],$$

³ La estimación de suavizamiento de la varianza de diseño se discute en la **Subsección 3.2**.

donde $\mathbf{B} = \text{diag}(b_d^2)$ es una matriz diagonal de dimensión D , $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{Z}(\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{V}^{-1}$,

$\mathbf{V} = \text{diag}(b_d^2\hat{\sigma}_v^{2(a)} + \tilde{\psi}_d)$, \mathbf{y} es una matriz con entradas de columna igual a $\hat{\theta}_d$, y \mathbf{Z} es una matriz con entradas de fila iguales a \mathbf{z}_d^T .

3.2. Suavizamiento de las varianzas directas

Dado que se desconoce la varianza de diseño ψ_d no se puede calcular la varianza suavizada de diseño $\tilde{\psi}_d$. En su lugar, se asume que está disponible un estimador insesgado $\hat{\psi}_d$ de la varianza del diseño a partir del cual se puede calcular la varianza suavizada. Aunque $\hat{\psi}_d$ es insesgado, puede ser inestable cuando el tamaño de la muestra en el área es pequeño. Para abordar este problema, se modeló la estimación de la varianza $\hat{\psi}_d$ y se usó su valor predicho como la estimación de la varianza suavizada. Más precisamente, se tiene:

$$E_{mp}(\hat{\psi}_d) = E_m(\psi_d) = \tilde{\psi}_d.$$

Un estimador simple e insesgado de la varianza suavizada del diseño $\tilde{\psi}_d$ es $\hat{\psi}_d$. Sin embargo, este último puede ser bastante ineficiente cuando el tamaño de la muestra del dominio es pequeño. Se obtiene un estimador más estable modelando $\hat{\psi}_d$ dado \mathbf{z}_d . Dick (1995), Rivest y Belmonte (2000) y Beaumont y Bocci (2016) utilizaron el método de Funciones de Varianza Generalizada (GVF, por sus siglas en inglés) y consideraron el siguiente modelo:

$$\log(\hat{\psi}_d) = \mathbf{x}_d^T \boldsymbol{\alpha} + \varepsilon_d,$$

donde \mathbf{x}_d es un vector de variables explicativas que son funciones de \mathbf{z}_d , $\boldsymbol{\alpha}$ es un vector de parámetros desconocidos del modelo a estimar, y ε_d es un error aleatorio con $E(\varepsilon_d|\mathbf{Z}) = 0$ y varianza constante $\sigma_\varepsilon^2 = \text{var}(\varepsilon_d|\mathbf{Z})$, y los errores ε_d se distribuyen de manera idéntica condicionalmente a \mathbf{Z} .

A fin de obtener estimaciones para ψ_d , los valores predichos, $\hat{\psi}_d$, del modelo anterior se obtienen empleando la siguiente expresión, en términos de $\hat{\psi}_d$, la estimación de la varianza suavizada:

$$\hat{\psi}_d = \exp(\mathbf{x}_d^T \hat{\boldsymbol{\alpha}}) \times \hat{\Delta}(\hat{\boldsymbol{\alpha}}), d = 1, \dots, D,$$

donde $\hat{\Delta}(\hat{\boldsymbol{\alpha}}) = \frac{\sum_{d=1}^D \hat{\psi}_d}{\sum_{d=1}^D \exp(\mathbf{x}_d^T \hat{\boldsymbol{\alpha}})}$ es un estimador insesgado de un término de corrección de sesgo

propuesto por Hidioglou et al. (2019) $[\Delta = E(\exp(\varepsilon_d))]$ y $\hat{\alpha}$ es el estimador de α obtenido mediante el método de mínimos cuadrados, con $\hat{\alpha} = (\sum_{d=1}^D \mathbf{x}_d \mathbf{x}_d^T)^{-1} \sum_{d=1}^D \mathbf{x}_d \log(\hat{\psi}_d)$. Rivest y Belmonte (2000) concluyeron que este estimador no sobreestima ni subestima la varianza suavizada, debido a que el promedio del estimador de varianza suavizada, $\hat{\psi}_d$, es igual al promedio de la varianza del estimador directo, $\hat{\psi}_d$, es decir:

$$\frac{\sum_{d=1}^D \hat{\psi}_d}{D} = \frac{\sum_{d=1}^D \hat{\psi}_d}{D}.$$

3.3. Transformación arcoseno

En aplicaciones de datos reales pueden ocurrir problemas que teóricamente no se esperaban o pueden violarse los supuestos del modelo FH estándar, por ejemplo, la normalidad y la independencia de los términos de error. Una manera de remediar estos problemas es acudir al uso de transformaciones en la variable de interés.

Específicamente, se ha utilizado la transformación arcoseno, que se usa frecuentemente cuando el estimador directo del modelo FH es una razón o proporción (Casas-Cordero et al. 2016; Schmid et al. 2017). Además de los problemas relacionados con el cumplimiento de los supuestos del modelo, existen otros dos grandes argumentos que justifican el uso de esta transformación, antes de ajustar el modelo de áreas:

1. En su concepción más básica, el modelo FH es una combinación lineal de covariables que pueden tomar valores que se salen del rango aceptable en el que puede estar una proporción; es decir, $\hat{\theta}_d^{FH} \in \mathbb{R}$, mientras que $\theta_d \in (0,1)$.
2. El estimador de la varianza de la estimación directa para una proporción en un diseño complejo depende de la proporción estimada y, por ende, habría problemas de endogeneidad en los parámetros.

En este caso, se utiliza la propuesta de Jiang et al. (2001) que permite evadir las dificultades de trabajar con la estimación directa, que consiste en lo siguiente:

$$\hat{\theta}_d^{*arcsin} = \sin^{-1}\left(\sqrt{\hat{\theta}_d}\right),$$

$$VAR(\hat{\theta}_d^{*arcsin}) = \frac{1}{4\tilde{n}_d},$$

donde $*arcsin$ denota la transformación a la escala arcoseno y \tilde{n}_d el tamaño de muestra efectivo que puede describirse como el tamaño de la muestra ajustado por el diseño de muestreo (Jiang et al. 2001). El modelo FH es estimado usando la ecuación (2) y debe verificarse que las estimaciones estén en el intervalo $[0, \pi/2]$, de tal forma que los resultados al aplicar la transformación inversa estén entre 0 y 1. Existen diferentes formas de realizar la transformación inversa, como una transformación inversa *naive* (Jiang et al. 2001; Chen y Lahiri, 2002) o una transformación inversa con corrección de sesgo (bc) (Sugawasa y Kubokawa (2017); Hadam et al. (2020)). En este análisis se utilizó la transformación *naive* que consiste en lo siguiente:

$$\hat{\theta}_d^{EBLUP,naive} = \sin^2(\hat{\theta}_d^{EBLUP*arcsin}).$$

3.4. Ajustes al modelo EBLUP

Con el tiempo y los recursos adecuados, es razonable esperar que la persona que investiga pueda identificar un buen modelo de trabajo, que es el paso clave para mejorar las estadísticas de áreas pequeñas, independientemente del método basado en modelos que se utilice. No obstante, esta es una tarea bastante extensa en un entorno de encuestas complejo que requiere estadísticas de áreas pequeñas para un gran número de variables de forma continua. Un enfoque a adoptar podría ser encontrar un modelo mixto robusto relativamente simple, que pueda estar científicamente justificado y que probablemente funcione para muchas variables en diferentes momentos. Sin embargo, incluso después de tomar tal precaución, no hay garantía de que funcione razonablemente bien en todas las situaciones y todo el tiempo (Jiang y Lahiri, 2006). Por lo tanto, es esperable que quien realiza la encuesta tenga algunas precauciones adicionales, antes de producir estadísticas oficiales que son de vital importancia para comprender el estado de la economía, la salud, la educación y el bienestar general de un país. A continuación, se describe un ajuste al modelo

EBLUP que se ha considerado en la literatura, por ejemplo, en Fay y Herriot (1979), y que ha sido empleado en este trabajo.

Aunque se espera que el desempeño general del EBLUP sea mejor que el del estimador ponderado de la encuesta cuando el modelo asumido es aceptable, puede que para un área pequeña en particular no funcione del todo bien debido a una variedad de razones, por ejemplo, datos auxiliares deficientes para el área pequeña o valores atípicos. Efron y Morris (1972) sugirieron el siguiente tipo de ajuste para el EBLUP:

$$\hat{\theta}_d^{EBLUP'} = \begin{cases} \hat{\theta}_d^{EBLUP}, & \text{si } y_d - c_d\sqrt{\psi_d} \leq \hat{\theta}_d^{EBLUP} \leq y_d + c_d\sqrt{\psi_d} \\ y_d - c_d\sqrt{\psi_d}, & \text{si } \hat{\theta}_d^{EBLUP} < y_d - c_d\sqrt{\psi_d} \\ y_d + c_d\sqrt{\psi_d}, & \text{si } \hat{\theta}_d^{EBLUP} > y_d + c_d\sqrt{\psi_d} \end{cases}$$

Donde c_d es una constante adecuadamente elegida. Tal como Fay y Herriot (1979), eligiendo $c_d = 1$. La idea principal es limitar la desviación del EBLUP del estimador ponderado de la encuesta. Esto reduce la pérdida de eficiencia si el modelo falla para un área particular sin perder el beneficio de modelar las áreas donde el modelo funciona bien.

4. Evaluación de las estimaciones de áreas pequeñas

La precisión de las estimaciones de áreas pequeñas depende de la confiabilidad del modelo de Fay-Herriot. Por lo tanto, es esencial hacer una evaluación cuidadosa de la validez del modelo antes de publicar las estimaciones. Por ejemplo, es importante verificar que se mantenga una relación lineal entre $\hat{\theta}_d$ y \mathbf{z}_d , al menos aproximadamente. Una forma sencilla de verificar la suposición de linealidad es graficar los residuos estandarizados, \hat{a}_d , contra los valores predichos $\mathbf{z}_d^T \hat{\boldsymbol{\beta}}$.

$$\hat{a}_d = \frac{\hat{\theta}_d - \mathbf{z}_d^T \hat{\boldsymbol{\beta}}}{\sqrt{b_d^2 \hat{\sigma}_v^2 + \hat{\psi}_d}}$$

Los residuos estandarizados son estadísticas clave que también se pueden usar para verificar otros supuestos del modelo, como la normalidad de los errores o la homocedasticidad. No

obstante, en el caso de un modesto alejamiento de la normalidad, Rao y Molina (2015, p.138) argumentaron que las estimaciones de EBLUP y sus correspondientes estimaciones de MSE son generalmente robustas a estas desviaciones. Por el contrario, si se observa un alejamiento importante del cumplimiento de la normalidad, siguiendo lo expresado por Jiang y Lahiri (2006), se deben considerar métodos Bootstrap y Jackknife para la estimación del MSE.

4.1. Tratamiento de valores atípicos e influyentes

Los valores atípicos (es decir, las áreas que no siguen el mismo modelo que las otras áreas) del modelo a nivel de área deben identificarse y, si es necesario, el modelo debe reajustarse. Los valores atípicos se identifican iterativamente examinando los residuos estandarizados de ese modelo.

Si \hat{a}_i 's son normales entonces $\hat{a}_d^2 \sim \chi_1^2$. Sea \hat{a}_{dm}^2 el residuo estandarizado al cuadrado más grande entre los D dominios utilizados en el modelamiento. Se define el valor c tal que $(\hat{a}_{dm}^2 \leq c) = 1 - \alpha$, para un α dado. Si el residuo cuadrado más grande es mayor que c (es decir, $\hat{a}_{dm}^2 > c$), entonces el dominio correspondiente se considera un valor atípico. Se reserva y se retendrá la estimación directa para este dominio. A partir de los dominios restantes, se vuelve a calcular el modelo y se vuelve a identificar un valor atípico, si lo hay. El proceso iterativo se repite hasta que no se encuentran más valores atípicos. En esta investigación, se intentó $\alpha=0.05$. En resumen, si el residuo cuadrado más grande resultante de un modelo es demasiado grande (más grande que algún valor crítico), el dominio correspondiente se marca como un valor atípico.

Por otra parte, se ha considerado calcular las distancias de Cook (Cook, 1977) para identificar áreas que podrían tener una influencia significativa sobre la estimación de $\hat{\beta}$ y, por tanto, sobre las estimaciones SAE y sus correspondientes estimaciones de MSE. La distancia de Cook para el área d viene dada por

$$Dist_d = \frac{1}{q} (\hat{\beta} - \hat{\beta}^{(-d)})^T \sum_{d=1}^D \frac{\mathbf{z}_d \mathbf{z}_d^T}{\sigma_u^2 + \hat{\psi}_d} (\hat{\beta} - \hat{\beta}^{(-d)}),$$

donde q es la dimensión de \mathbf{z}_d y $\hat{\beta}^{(-d)}$ es la estimación de $\hat{\beta}$ obtenida después de eliminar el área d .

Se considera que una observación es influyente si $Dist_d > 1$ (Cook y Weisberg, 1994, p.118) ⁴.

4.2. Estimación del error cuadrático medio

En general, la precisión de una estadística se estudia analizando su intervalo de confianza. Por tanto, es importante contar con una medida que dé cuenta de la incertidumbre asociada a la estimación por SAE y así se verifique la factibilidad de su uso para producir estadísticas oficiales.

Un concepto que es útil para evaluar las ganancias de eficiencia resultantes del uso de la estimación de área pequeña $\hat{\theta}_d^{EBLUP}$ respecto de la estimación directa $\hat{\theta}_d$ es el error cuadrático medio (MSE, por sus siglas en inglés):

$$MSE(\hat{\theta}_d^{EBLUP}) = E_{mp}(\hat{\theta}_d^{EBLUP} - \theta_d)^2.$$

El MSE es desconocido, pero se puede estimar. El MSE de los estimadores compuesto y sintético se calcula por separado. El MSE estimado de los estimadores de nivel de área depende del procedimiento utilizado para estimar los parámetros (por ejemplo, REML o ADM). Para REML, el MSE de la estimación de área pequeña $\hat{\theta}_d^{EBLUP}$ se puede calcular usando:

$$mse(\hat{\theta}_d^{EBLUP}) = \begin{cases} g_{1d} + g_{2d} + 2g_{3d} & \text{si } d = 1, \dots, j \\ \mathbf{z}_d^T \text{var}(\hat{\boldsymbol{\beta}}) \mathbf{z}_d + b_d^2 \hat{\sigma}_v^2 & \text{si } d = j + 1, \dots, D \end{cases}$$

donde

$$g_{1d} = \hat{\gamma}_d \hat{\psi}_d,$$

$$g_{2d} = (1 - \hat{\gamma}_d)^2 \mathbf{z}_d^T \left[\sum_{d=1}^j \mathbf{z}_d \mathbf{z}_d^T / (b_d^2 \sigma_v^2 + \hat{\psi}_d) \right]^{-1} \mathbf{z}_d,$$

$$g_{3d} = \frac{(b_d^2 \hat{\psi}_d)^2}{(b_d^2 \sigma_v^2 + \hat{\psi}_d)^3} \text{var}(\hat{\sigma}_v^2),$$

⁴ Otra aproximación consiste en comparar $Dist_d$ con $F(p, n - p, \alpha)$, donde p es el número de variables y n el número de observaciones. Si $Dist_d > F$ es un posible outlier.

$$\text{var}(\hat{\beta}) = \left[\sum_{d=1}^j \frac{\mathbf{z}_d \mathbf{z}_d^T}{b_d^2 \sigma_v^2 + \hat{\psi}_d} \right]^{-1}.$$

4.3. Coeficiente de variación

Para evaluar la calidad de las estimaciones, se utiliza una medida que está basada en el MSE y es similar al coeficiente de variación (CV) de la estimación SAE. Los valores de este CV propuesto se obtienen dividiendo la raíz cuadrada de los errores cuadráticos medios estimados (MSE) de las estimaciones de SAE entre las estimaciones, y se expresan como un porcentaje.

$$CV_d = \frac{\sqrt{\text{mse}(\hat{\theta}_d^{EBLUP})}}{\hat{\theta}_d^{EBLUP}} \times 100\%.$$

Sin embargo, el CV es una medida para la que se requiere hacer algunas consideraciones importantes. Por una parte, si se fija un umbral para el CV, esto tiene un impacto directo sobre la amplitud relativa del intervalo de confianza. Por ejemplo, se fija como umbral un CV del 30% (Gutiérrez, et. al., 2020), la amplitud relativa del intervalo de confianza estará en torno al 118%⁵.

Por otra parte, en el caso de las proporciones, el CV no es una medida simétrica alrededor de $P = 0,5$, como sí lo es la varianza o el error estándar y, por su definición, cuando la proporción es pequeña (que es el caso frecuente en este trabajo), el CV tiende a ser muy alto, indicando baja precisión. Como alternativa, se consideró el uso del CV logarítmico (CVL) del área denotado por $CV_d(L)$, que para $P \leq 0,5$ (Gutiérrez, et. al., 2020), se calcula como:

$$CV_d(L) = \frac{CV_d}{-\log(\hat{\theta}_d^{EBLUP})} \times 100\%.$$

4.4. Bondad de ajuste

Además de la evaluación de la incertidumbre descrita anteriormente, también es importante evaluar el desempeño de un estimador bajo algunas medidas de rendimiento a partir de los supuestos del modelo, por ejemplo, se podría utilizar alguna medida que dé cuenta de la

⁵ Esto se verifica escribiendo el intervalo de confianza en términos del coeficiente de variación y posteriormente, haciendo la relación con la amplitud relativa.

fuerza del conjunto de covariables incluidas en los modelos para la predicción de θ_d . Para este fin, se ha considerado y calculado como métrica de evaluación un coeficiente de determinación o R^2 propuesto por Hidioglou et al. (2019). Este coeficiente de determinación está asociado al modelo de vínculo $\theta_d = \mathbf{z}_d^T \boldsymbol{\beta} + b_d v_d$. Téngase en cuenta que el coeficiente de determinación asociado con el modelo combinado, $\hat{\theta}_d = \mathbf{z}_d^T \boldsymbol{\beta} + b_d v_d + e_d$, no es de interés ya que el objetivo no es la predicción de $\hat{\theta}_d$, sino la predicción de θ_d . El coeficiente de determinación viene dado por:

$$R^2 = 1 - \frac{\sigma_v^2}{\frac{(d-q)}{(d-1)}\sigma_v^2 + S^2(\hat{\boldsymbol{\beta}})}, \quad (3)$$

Donde q es la dimensión de \mathbf{z}_d y $S^2(\hat{\boldsymbol{\beta}})$ es la varianza muestral de $\mathbf{z}_d^T \hat{\boldsymbol{\beta}}$. Los detalles de la derivación del coeficiente de determinación en (3) son provistos por Hidioglou et al. (2019, pp.122-123).

4.5. Benchmarking

Finalmente, se espera que un conjunto de estimaciones de área pequeña sea numéricamente consistentes y más eficientes que las estimaciones directas, que son insesgadas bajo el diseño muestral. Se pueden comparar las estimaciones del área agregada con las estimaciones directas correspondientes para el mismo propósito. Si las estimaciones (indirectas) basadas en modelos agregados no concuerdan con las estimaciones directas correspondientes, se puede utilizar técnicas de *benchmarking* para lograr consistencia. Las estimaciones por *benchmarking* de áreas pequeñas ofrecen una propiedad atractiva para las ONE.

En este caso, los tamaños muestrales en las regiones de ENUSC 2018 se establecieron para garantizar que las estimaciones de las tasas de victimización del área urbana regional, $\hat{\theta}_r$, cumplan los estándares de calidad del INE (INE, 2020), y así las estimaciones oficiales puedan ser consideradas confiables. Por ello, conviene asegurar que exista una consistencia entre las tasas de victimización regional estimadas de manera directa y las tasas de victimización comunal estimadas mediante el modelo EBLUP. En otras palabras, dado que las estimaciones directas a nivel regional son fiables, se espera que las estimaciones SAE del

nivel comunal sean consistentes con la estimación regional correspondiente. De manera que si la región r contiene M_r comunas y $\hat{\theta}_1^{EBLUP',TOT}, \dots, \hat{\theta}_{M_r}^{EBLUP',TOT}$, son las estimaciones de los totales de hogares victimizados en estas comunas mediante EBLUP, con $\hat{\theta}_d^{EBLUP',TOT} = \hat{\theta}_d^{EBLUP'} \times N_d$, donde N_d es el total de la población en la comuna d , $d = 1, \dots, M_r$, de acuerdo con las proyecciones de población usadas para la calibración de los factores de expansión en ENUSC 2018; es deseable que se verifique:

$$\hat{\theta}_r^{TOT} = \sum_{d=1}^{M_r} \hat{\theta}_d^{EBLUP',TOT},$$

Donde $\hat{\theta}_r^{TOT} = \hat{\theta}_r \times N_r$ es el total de hogares victimizados en la región r , $r = 1, \dots, 16$, estimado utilizando el factor de expansión regional, y N_r corresponde al total de la población en la región r de acuerdo con las proyecciones de población usadas para la calibración de los factores de expansión en ENUSC 2018.

Generalmente, las estimaciones basadas en modelos no cumplen con esta propiedad de *benchmarking*. Para conseguirlo, se sigue el ajuste propuesto por Fay y Herriot (1979), multiplicando las estimaciones de las tasas comunales por un factor de consistencia de la forma:

$$\hat{\theta}_d^{EBLUP''} = \lambda_r \times \hat{\theta}_d^{EBLUP'}, \text{ donde } \lambda_r = \frac{\hat{\theta}_r^{TOT}}{\sum_{d=1}^{M_r} \hat{\theta}_d^{EBLUP',TOT}}.$$

Así, la estimación de la tasa de victimización de cada comuna en la región r queda ajustada por su correspondiente factor de consistencia regional λ_r . Note que este factor de cierta forma da cuenta de la calidad de las estimaciones del EBLUP, puesto que, si el modelo es adecuado, el factor de ajuste regional estará en torno a uno.

5. Aplicación en la ENUSC 2018

5.1. Estimadores directos

El estimador directo, que se denota por $\hat{\theta}_d$, usa solamente los datos del área d , por lo que sólo está disponible para dicha área en la muestra. Esta estimación se basa en el diseño

muestral; es decir, no involucra la adopción de ningún modelo, por cuanto, la medida de error bajo este enfoque recoge el error de estimación a través de la aleatoriedad de la selección muestral.

Si considera una población finita de tamaño N , donde las unidades son hogares. Para cada hogar j , la variable de interés es $y_j = 1$ si al menos una persona en el hogar fue victimizada, y $y_j = 0$ si no. Asumamos que la población está dividida en pequeñas áreas denotadas por $d = 1, \dots, D$, con $D = 102$, que corresponden a las 102 comunas de Chile presentes en ENUSC 2018. Sea s una muestra extraída de la población de acuerdo con un diseño muestral dado. La población y la muestra son denotadas por $U = \cup_{d=1}^D U_d$ y $s = \cup_{d=1}^D s_d$, respectivamente. Sean N_d y n_d los tamaños de U_d y s_d , con $d = 1, \dots, 102$. **El objetivo es estimar la tasa de victimización para cada indicador objetivo, $\theta_d = N^{-1} \sum_{j \in U_d} y_j$, con su respectivo error estándar.** El estimador directo de θ_d propuesto por Hájek (1971), y el correspondiente error estándar estimado son

$$\hat{\theta}_d = \frac{1}{\hat{N}_d} \sum_{j \in s_d} w_j y_j,$$

$$\widehat{e\hat{e}}(\hat{\theta}_d) = \sqrt{\frac{1}{\hat{N}_d^2} \times \sum_{j \in s_d} w_j (w_j - 1) (y_j - \hat{\theta}_d)^2},$$

donde $\hat{N}_d = \sum_{j \in s_d} w_j$ es el estimador directo del tamaño de la población de la comuna d y w_j es el peso de la muestra o factor de expansión del hogar j . Los pesos muestrales de la ENUSC 2018 se obtuvieron a partir de un diseño de muestreo complejo en tres etapas: (1) manzanas, (2) viviendas ocupadas en áreas urbanas, y (3) persona (informante Kish). Los cuartiles de los tamaños muestrales de las 102 comunas representadas en ENUSC 2018 son $n^{(0)} = 80$, $n^{(1)} = 142$, $n^{(2)} = 191$, $n^{(3)} = 304$.

Los estimadores directos son aproximadamente insesgados para θ_d , $d = 1, \dots, 102$, pero no son precisos para la estimación de parámetros en áreas pequeñas. Debido al pequeño tamaño de la muestra, dan lugar a medidas de precisión poco confiables, con coeficientes de variación altos (por ejemplo, mayor al 30%). Ver Särndal et al. (1992) para obtener más detalles sobre la inferencia en poblaciones finitas y las propiedades de los estimadores directos. **En consecuencia, dado los pequeños tamaños muestrales con que se cuenta para las comunas, las medidas de precisión mostrarán estimaciones**

poco confiables que no soportan un proceso de inferencia estadística adecuado; por tanto, la estimación de las tasas de victimización comunal se convierte en un problema de estimación de área pequeña.

5.2. Información auxiliar

Según la revisión sustancial de la literatura realizada, los indicadores de victimización, en especial variables que han sido reflejo de relaciones significativas en diversos estudios, se pueden dividir en los cinco grandes grupos que se resumen en el **Tabla 1**.

Tabla 1. Covariables relacionadas con la victimización.

Grupo	Covariables	Fuente
Criminológicas	<ul style="list-style-type: none"> — Delitos de connotación social (violaciones sexuales y robos con intimidación) — Delitos a la propiedad, como robos, hurtos, robos de vehículos motorizados, robos con violencia y robos con violencia en el hogar. 	(Fay, Planty y Diallo, 2013)
Socioeconómicas	Educación, ingresos, índices de desarrollo, etc.	(ISUC, 2013); (Armas y Herrera, 2018); (Buil Gil, 2019); (Olavarría, 2006)
Sector más victimizado	Posición política y/o ideológica, dimensión valórica, etc.	(Alvira y Rubio, 1982) (INEGI, Características de la victimización y victimización múltiple de la población en México, 2010-2015, 2017); (Armas y Herrera, 2018); (Buil Gil, 2019); (Olavarría, 2006)
Demográficos	Rango etario, sexo, condición de actividad, densidad poblacional, etc.	(Armas y Herrera, 2018)
Medidas de seguridad	Medidas comunitarias de seguridad y de política pública desde el enfoque de seguridad y desde el enfoque social.	(Armas y Herrera, 2018)

Fuente: Elaboración propia.

Las referencias consultadas y resumidas en el **Tabla 1** sugieren la construcción de una serie de indicadores o covariables relacionados con los indicadores de victimización o variables de respuesta a modelar.

Como desafío para las futuras investigaciones, queda la revisión y validación de la información auxiliar obtenida, ya que es importante que dicha información se convierta en registros estadísticos confiables, para ser utilizadas en los modelos. Existen diferentes criterios de calidad que son ejecutables para esta validación, como las cinco dimensiones que el Centro de Políticas de Salud de Manitoba propone: precisión, validez interna, validez externa, actualización de los datos y la interpretabilidad (Smith, y otros, 2017). También existen otros criterios de calidad, como exactitud, completitud, consistencia, credibilidad, eficiencia, comprensibilidad, entre otros que se encuentra definidos por ISO/IEC 25012 (International Organization for Standardization, 2008) .

Teniendo en cuenta que las variables auxiliares no pueden presentar datos faltantes para ningún área, para la construcción de covariables se cuenta con información proveniente de los microdatos del Censo de Población y Vivienda 2017 (CPV 2017) y de registros administrativos de Carabineros de Chile provistos por la SPD agregados a nivel comunal para el año 2018. A continuación, se describe brevemente los tipos de covariables disponibles para la construcción de los modelos de áreas pequeñas:

- **Covariables derivadas del CPV 2017:** Describen características socioeconómicas y demográficas de la población, en términos de proporciones (p)⁶, tales como sexo, tramos etarios, niveles de educación, situación ocupacional, condiciones de la vivienda, entre otros.
- **Covariables derivadas de registros policiales:** La SPD tiene a disposición registros policiales de Carabineros de Chile referidos por una parte a denuncias de diferentes tipos de delitos, que corresponden al 95% del total de denuncias realizadas durante el 2018⁷. Estos registros proveen información acerca del total de casos policiales, que se utilizan para analizar la ocurrencia de hechos delictivos. La cifra considera las denuncias de delitos que realiza la comunidad en las unidades policiales. En este punto, es importante tener en cuenta que no

⁶ En el caso del indicador de victimización agregada por delitos de mayor connotación social, las proporciones fueron calculadas usando como denominador el total de personas en la comuna, para los otros dos indicadores, se usó como denominador el total de personas exclusivamente en el área urbana de la comuna. En el primer caso, se buscó guardar consistencia con los totales y las tasas de las covariables de registros policiales, las cuales son reportadas para toda la comuna.

⁷ El 5% restante corresponde a denuncias registradas por la Policía de Investigaciones de Chile (PDI).

siempre la comuna donde ocurre el delito es la misma comuna donde reside la persona que hizo la denuncia, por ende, este es un factor que debe estar presente al momento de la homologación y análisis de los datos, ya que algunos registros de casos policiales puede que no coincidan con la comuna de residencia de la persona encuestada. Por otra parte, este registro también considera las detenciones que realizan las policías ante la ocurrencia de delitos flagrantes. Internacionalmente este indicador es conocido como "delitos conocidos por las policías". Para aislar algún efecto posible del tamaño poblacional en la frecuencia de denuncias, se han construido, en base a los registros provistos, indicadores que dan cuenta de la tasa de denuncias por cada 100.000 habitantes (ts).

En el **Anexo A 8.1** se encuentra el detalle de las covariables y su preparación, previo a su inclusión en la construcción de los modelos de áreas pequeñas.

Además de los criterios de calidad de los datos mencionados anteriormente, también se debe tener presente, en particular al momento de interpretación de los resultados, la cifra oculta o dark figure en inglés⁸, la cual se define como la cantidad de delitos que no han sido denunciados a las autoridades competente, ya sea por diferentes factores, como temor, falta de tiempo, entre otras (Centro de Formación Estudio Criminal, 2018).

5.3. Etapas generales en la adaptación del modelo

Como se mencionó anteriormente, la información de los casos policiales de Carabineros de Chile y los indicadores sociodemográficos construidos a partir del CPV 2017, se utilizan como información auxiliar en el modelo SAE para cada indicador. Más precisamente, los datos auxiliares junto con las estimaciones directas de la encuesta se utilizan para obtener las estimaciones de áreas pequeñas, en este caso, las estimaciones comunales.

Los dominios de interés corresponden a las 102 comunas de la muestra de la ENUSC 2018. En el **Anexo B 8.2** se puede ver el detalle.

⁸ Existen estudios que dan cuenta de la existencia de una desigualdad socioeconómica en el registro de delitos. A grandes rasgos, reflejan que las personas de comunas con mayor nivel socioeconómico denuncian. En los trabajos futuros este es un factor que debe considerarse en los modelos para evitar posibles sesgos en las estimaciones. Véase https://www.bcn.cl/obtienearchivo?id=documentos/10221.1/29929/1/PDF_librodesigualdad_ultima_version.pdf&descargar

Las etapas de análisis en esta investigación son las mismas para los tres indicadores y fueron las siguientes:

1. Estimación de las tasas de victimización comunal para el indicador objetivo por el método de estimación directa.
2. Estimación de la varianza del estimador directo, mediante función de varianza generalizada (FVG), para cada uno de los indicadores.
3. Ajuste de los modelos FH para estimar las tasas de victimización comunal. Esta etapa es un ejercicio iterativo.
4. Evaluación de la fiabilidad de las estimaciones.
5. Diagnóstico de los modelos.

5.3.1. Suavizamiento de varianzas directas

Como se mencionó anteriormente, se suavizan las varianzas directas para reducir la variabilidad de $\hat{\psi}_d$. Las varianzas resultantes generalmente tienen menos variabilidad y menos valores atípicos. Esto generalmente conduce a un mejor ajuste de los modelos de área pequeña. El estimador de la varianza suavizado $\tilde{\psi}_d$ se obtiene siguiendo el proceso descrito en la **Subsección 3.2.**, con las siguientes consideraciones:

1. Las comunas con estimaciones directas y varianzas igual a cero han sido descartadas, ya que estas estimaciones directas no se consideraron suficientemente fiables. Estas comunas fueron tratadas como fuera de muestra y se obtuvieron sus valores predichos exclusivamente de la estimación sintética. Tanto para el robo con violencia e intimidación (RVI) como para el robo por sorpresa (RPS), el modelo FVG se construyó a partir de la información de cien comunas. En el caso de victimización agregada por delitos de mayor connotación social (VA) no se encontraron comunas con estimaciones directas o varianzas estimadas iguales cero, por tanto, se consideraron todas las comunas que forman parte de la muestra ENUSC.

2. Se utilizó la transformación de logaritmo neperiano para algunas covariables, en particular, de la esperanza del indicador de victimización y el tamaño de muestra comunal.
3. Para la selección de covariables en el modelo se usó el algoritmo de Random Forests, con el fin de identificar las covariables con mayor poder predictivo⁹. Como resultado, el modelo para cada indicador está compuesto por el logaritmo neperiano del indicador, el logaritmo neperiano del tamaño de muestra y la interacción entre ambos. En la **Tabla 2** se resumen los resultados de los modelos de función de varianza generalizada. Se observa que para robo con violencia o intimidación la bondad de ajuste del modelo es 98,65% y para el robo por sorpresa es de 98,73%. En el caso de la tasa de victimización agregada, cabe señalar que se ajusta un modelo sin intercepto, con una bondad de ajuste de 98,12%. En los tres casos, los resultados dan cuenta de una buena bondad de ajuste.

Tabla 2. Covariables en modelos de función de varianza generalizada.

Indicador	Variabes en el modelo	R^2_{adj}
Robo con violencia o intimidación	log RVI, log n, y log RVI * log n	98,65%
Robo por sorpresa	log RPS, log n, y log RPS * log n	98,73%
Victimización Agregada	log VA, log n y log VA* log n	98,12%

Fuente: Elaboración propia.

Adicionalmente, se tienen presentes las siguientes consideraciones importantes sobre la estimación de la varianza del estimador:

1. Un peligro evidente que debe evitarse, independientemente del procedimiento de estimación que se utilice, es la posibilidad de estimaciones de varianza negativa. Wolter (2007) sugiere introducir alguna restricción sobre el intercepto para evitar esta situación.
2. Hidiroglou et al. (2019) sugieren, como regla de oro, asumir como suficientemente estables las estimaciones directas de la varianza cuando el tamaño de muestra del área es grande, es decir, fijamos

⁹ Ver Breiman (2001) para más detalles acerca de la implementación de esta técnica.

$\hat{\psi}_d = \hat{\psi}_d$. Esta estrategia evita posibles sesgos del modelo en $\hat{\psi}_d$ para las áreas grandes, lo que podría resultar en estimaciones de EBLUP que se vuelven significativamente diferentes de las estimaciones directas. Esta no es una propiedad deseable para las áreas con un tamaño de muestra grande. Para efectos de este trabajo, se ha asumido como tamaño grande¹⁰ $n = 472$ para el robo con violencia o intimidación, $n = 395$ para el robo por sorpresa y $n=465$ para la tasa de victimización agregada. En el caso de robo con violencia o intimidación, significó un reemplazo en 87 áreas de la varianza de las estimaciones directas por las obtenidas a partir del modelo FGV; en el caso de robo por sorpresa, este reemplazo ocurrió en 81 áreas y finalmente, en el caso de victimización agregada en 86 comunas.

De esta forma, en el paso anterior, si se asume n grande, se utiliza $\hat{\psi}_d$ en el modelo de áreas; en caso contrario, las estimaciones $\hat{\psi}_d$ son utilizadas en lugar de $\hat{\psi}_d$ para generar las estimaciones EBLUP y sus MSE asociados.

5.4. Modelo a nivel de áreas

Esta sección adapta un modelo de Fay-Herriot a nivel comunal a las estimaciones de victimización para los tres indicadores objetivo, realiza un análisis de diagnóstico y presenta los resultados de la metodología empleada. Para el modelo de Fay-Herriot introducido en la **Subsección 3.1**, el uso de covariables que proporcionen el mayor grado posible de información a nivel comunal es muy importante. Los hallazgos se organizan de la siguiente manera:

1. Selección de covariables: analizamos el poder predictivo potencial de las covariables a través de análisis exploratorios (simples y multivariados). Por ejemplo, se verificó la distribución y

¹⁰ La determinación de tamaño grande para n se realizó usando criterios que incluyen el cálculo de coeficiente de variación logarítmico, grados de libertad y tamaño de muestra efectivo.

correlación de los datos, con respecto a los indicadores objetivo, a fin de decidir qué covariables incluir en los modelos.

2. Presentación de los resultados de cada modelo.
3. Generación de gráficos de dispersión con el objetivo de analizar la consistencia de las estimaciones obtenidas por EBLUP y visualizaciones en mapas de calor para una mejor diseminación de los resultados.
4. Se examinan los RMSE¹¹ de todas las estimaciones para comprobar su fiabilidad.
5. Se presentan los diagnósticos de cada modelo EBLUP.

Para la implementación del modelo de áreas existen varios paquetes en R (R Core Team, 2019), que incluyen sae (Molina y Marhuenda, 2018), hbsae (Boonstra, 2012), msae (Permatasari y Ubaidillah, 2021) y emdi (Kreutzmann et al., 2019), entre otros. En esta aplicación, se decidió emplear el paquete emdi debido a que facilita la estimación de indicadores utilizando métodos de estimación en área pequeña, proporciona extensiones del modelo a nivel de área (Fay y Herriot, 1979), que incluye transformaciones a nivel de área, como la transformación arcoseno, con sus retro transformaciones; herramientas para el diagnóstico de modelos, presentación y exportación de los resultados.

En cuanto al uso de la transformación arcoseno, fue usada como se describe en la **Subsección 3.3**, para la construcción de los modelos EBLUP de los indicadores de robo con violencia o intimidación y robo por sorpresa.

5.4.1. Selección de variables

El uso de modelos tiene como objetivo mejorar la precisión de las estimaciones de áreas pequeñas haciendo un uso óptimo de los datos disponibles. Por lo tanto, la construcción y diagnóstico de modelos, análisis de sensibilidad y la validación ocupan un lugar central en el SAE basado en modelos. No existe un enfoque único para la construcción de modelos. A continuación, se describen algunas pautas de buenas prácticas que se pueden seguir, y que han sido usadas para la estimación de las tasas de victimización en ENUSC 2018.

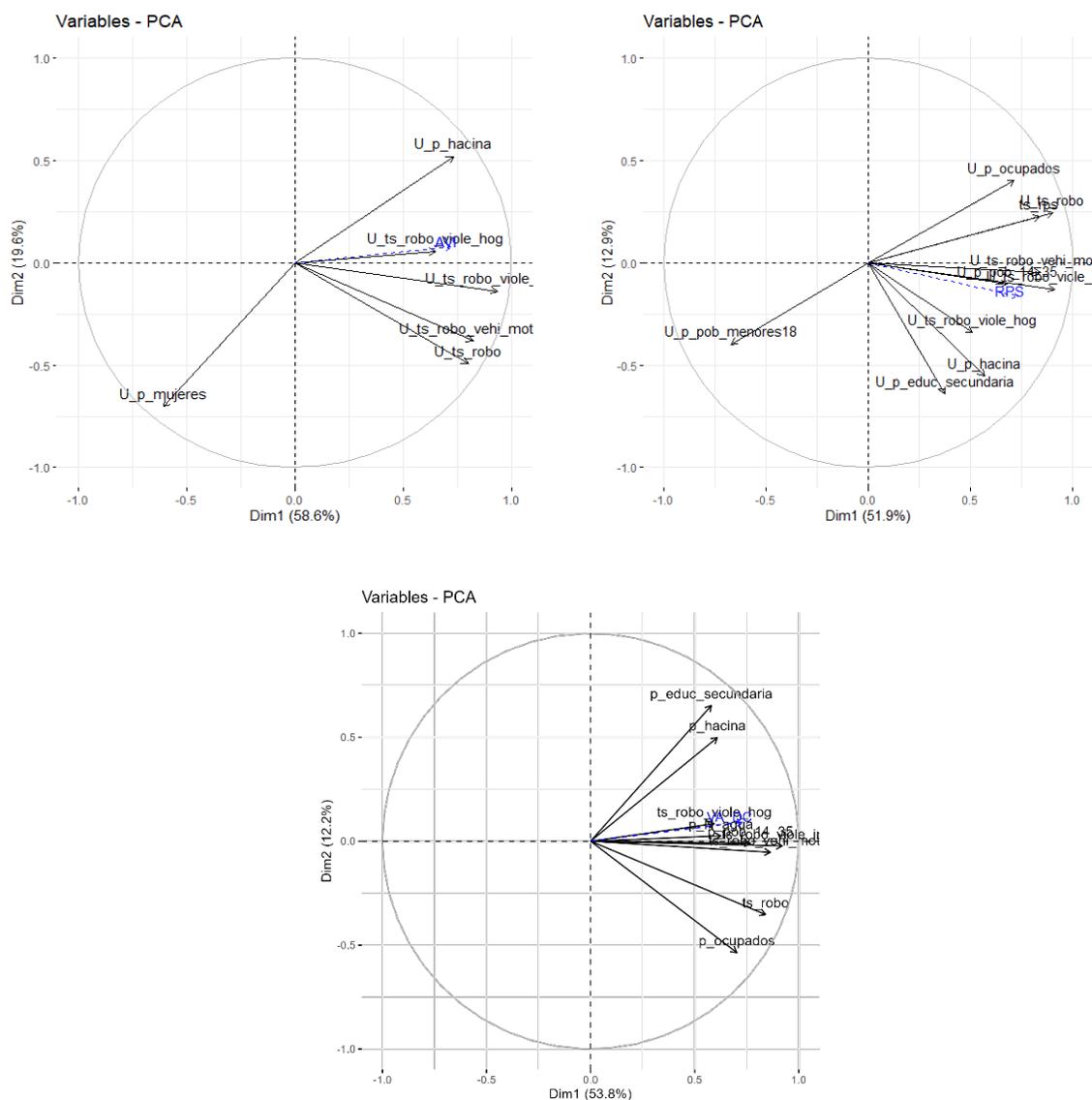
¹¹ Se utiliza RMSE (Raíz Cuadrada del Error Cuadrático Medio, con $RMSE = \sqrt{MSE}$) para hacer que la escala de los errores sea igual a la escala de las variables objetivo y facilitar la comparación entre las estimaciones EBLUP y las estimaciones directas.

La construcción del modelo se basa en un proceso de selección de covariables que brinden el mayor poder predictivo, minimizando los efectos aleatorios y conservando el espíritu de parsimonia. Un punto de partida razonable es realizar un análisis exploratorio (univariado y multivariado) de los datos que permita observar relaciones existentes en ellos, en particular, son de interés las relaciones entre las covariables disponibles y los indicadores objetivo.

En la **Figura 1**, se pueden ver representadas en el primer plano factorial del análisis en componentes principales (ACP), las covariables que mejor correlacionan ($r > |0,4|$) con los indicadores de RVI, RPS y VA. Para RVI y RPS se observa que, entre todas las covariables, la covariable que muestra la más fuerte relación es la tasa de robo con violencia e intimidación (RVI: $r = 0,78$, con $valor\ p < 2.2e - 16$ y RPS: $r = 0,71$, con $valor\ p < 2.2e - 16$). Para el indicador de VA, se observa que la tasa de robo con violencia o intimidación y de robo con violencia en el hogar prácticamente se le superponen cuando este es representado como variable suplementaria (no interviene en la construcción del plano factorial, pero puede ser representado para analizar cómo se relaciona con las otras variables), lo que es indicativo de una fuerte correlación positiva entre estas tasas y el indicador (0,72 y 0,62 respectivamente). Respecto a las variables censales, destacan la proporción de personas entre 14 y 35 años, y la proporción de hogares en condiciones de hacinamiento¹².

¹² La variable de proporción de hogares con acceso a la red pública y alcantarillado se superpone al indicador de la tasa de VA. Sin embargo, esta variable podría no ser una fuente de información relevante al ser considerada para comunas predominantemente urbanas. De hecho, el primer quintil es de 0,72 y el tercero de 0,94.

Figura 1. Covariables en plano principal (primera y segunda componentes principales) para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior).



Fuente: Elaboración propia.

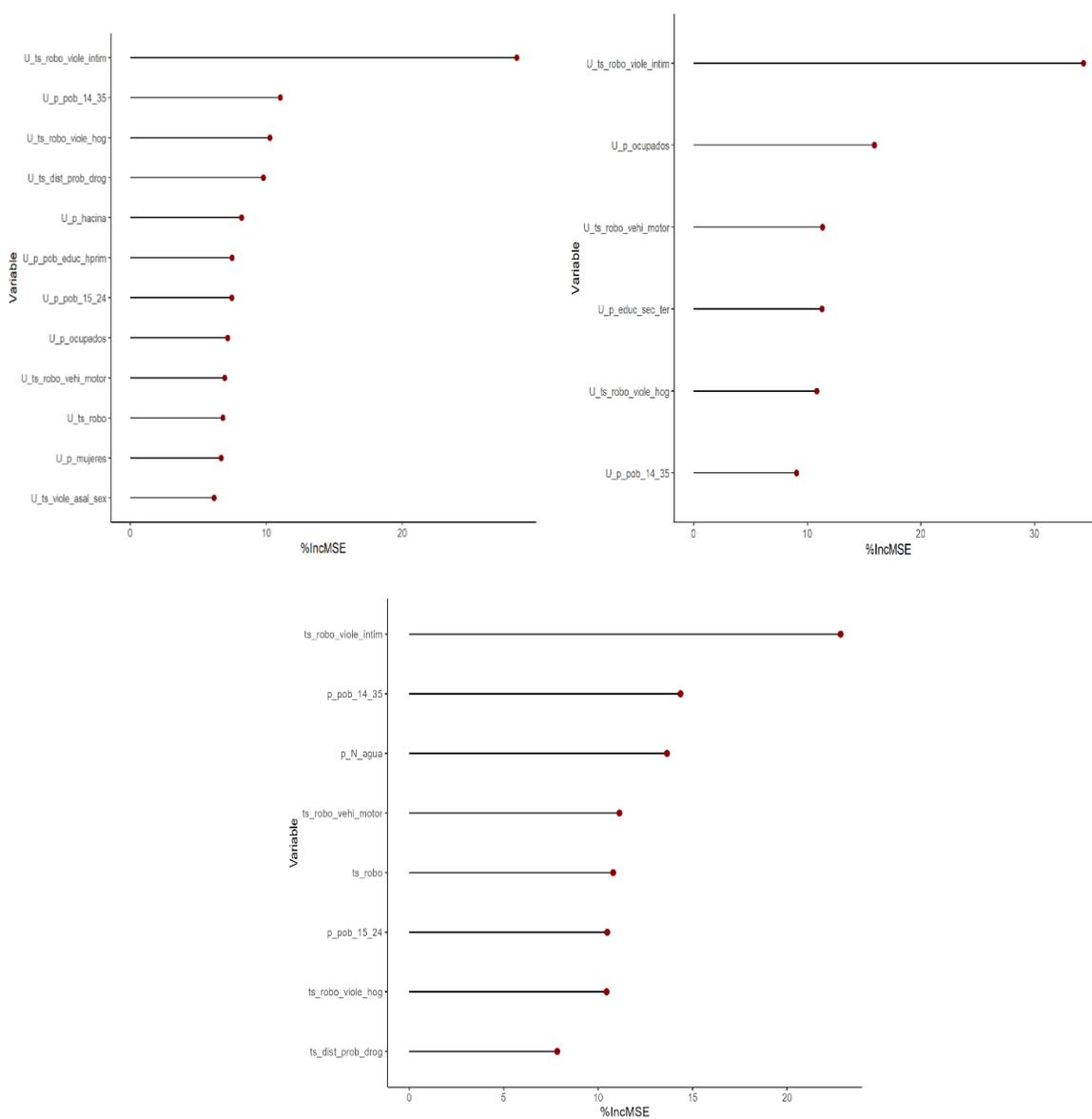
Por otra parte, el algoritmo de Random Forests (RF), que es un poderoso algoritmo de aprendizaje basado en el ensamble de modelos de árboles de decisión¹³ para problemas de clasificación y regresión¹⁴, fue empleado para obtener la importancia del poder predictivo de las covariables en términos del incremento relativo del error cuadrático medio (MSE). En la

¹³ También se usaron árboles de decisión individuales, que permitieron estudiar interacciones entre covariables.

¹⁴ También se usaron regresiones lineales por mínimos cuadrados ordinarios, con proceso de selección de variables *stepwise* bajo un esquema Bootstrap (con B=200). Tanto para robo con violencia o intimidación como para robo por sorpresa, destaca la tasa de robo con violencia o intimidación como principal predictor.

Figura 2 se observa que tal como se había detectado en el estudio del primer plano factorial del análisis de componentes principales, tanto para robo con violencia o intimidación como para robo por sorpresa, la tasa de robo con violencia o intimidación es la covariable con mayor poder predictivo, de la misma forma ocurre en el caso de la victimización agregada, siendo la tasa de robo con violencia o intimidación y la proporción de personas entre 14 y 35 años las de mayor poder predictivo. Los análisis anteriores condujeron a la selección de un conjunto menor de covariables (con mayor poder predictivo), que fueron consideradas en la implementación de los modelos de áreas para cada indicador.

Figura 2. Covariables con mayor poder predictivo para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior), según el algoritmo de Random Forests.



Fuente: Elaboración propia.

En términos estadísticos, para decidir si una covariable se incluye o no en el modelo, se usó la estadística t , con su *valor p* para evaluar la significancia estadística de los parámetros estimados. Además, para juzgar la complejidad del modelo se emplean los criterios de información de Akaike (AIC) e información Bayesiana (BIC), y la medida R^2 de Hidiroglou (ver **Subsección 4.4**) para evaluar la bondad de ajuste del modelo. Adicionalmente, el análisis de los residuos de los modelos debe corroborar el cumplimiento de los supuestos de normalidad y homocedasticidad, sin presencia de puntos atípicos o influyentes, y no presentar problemas de multicolinealidad (ver diagnósticos en la **Subsección 5.6**).

Por otra parte, a pesar de que el procedimiento descrito provee de un proceso robusto para la selección de covariables, la construcción del modelo EBLUP no debe estar supeditada exclusivamente a criterios estadísticos, pues es necesario que el modelo tenga consistencia temática, es decir, que las covariables finalmente seleccionadas tengan una suerte de relación conceptual con la variable respuesta y que sus coeficientes tengan el signo esperado.

En el caso de la victimización, estudios realizados en el país, como por ejemplo, el trabajo desarrollado por Olavarría (2006), señalan que delitos como robos con violencia son más prevalentes en la población menor de 30 años, por lo que es necesario considerar en los modelos covariables que den cuenta de estas características demográficas.

Además, la inclusión de variables de control en los modelos puede tener un efecto positivo en los mismos, pues favorece el control de observaciones atípicas o influyentes, o el cumplimiento de supuestos en los que se basa el modelo. En particular, el uso de variables indicadoras (*dummies*) que representan agregaciones mayores a las áreas pequeñas, para las cuales se obtienen estimaciones fiables bajo el diseño muestral (en este caso las regiones), puede resultar beneficioso como control al ir en pos del “préstamo de fuerza” entre áreas que son similares, principio base en la estimación de áreas pequeñas. De cualquier forma, se debe examinar que la inclusión de variables temáticas o de control no atenten contra el desempeño y la validez del modelo predictivo, en términos de la significancia de los predictores, ajustes y cumplimiento de supuestos. Estas variables *dummies* permiten introducir un efecto fijo diferenciador a nivel regional para cada una de las comunas objetivas, bajo el supuesto de que en Chile la dinámica de las regiones del Norte difiere de las regiones centrales, en particular de la Metropolitana (que conserva la mayor parte de la población) y las regiones del Sur. Al ser un efecto fijo y considerando la representatividad de la encuesta, permite acercar la media de cada región al producir una regresión para cada una

de ellas. Dicho efecto puede ser notorio al momento de realizar el *benchmarking*, donde, si las estimaciones comunales logran ser precisas considerando que el modelo posee un buen ajuste, los ponderadores estarán más cerca de uno. Esto se logra bajo la comprensión ya mencionada, donde en cada región se puede encontrar dinámicas que se ven reflejadas en cada una de sus comunas.

A partir de los criterios estadísticos evaluados en el proceso propuesto y las consideraciones temáticas descritas en el párrafo anterior, para el robo con violencia o intimidación se incluyeron diecisiete covariables, a saber:

- Tasa para robo con violencia o intimidación.
- Proporción de personas entre 14 y 35 años.
- Quince variables indicadoras para regiones.

Por su parte, para el robo por sorpresa se incluyeron diecinueve covariables, a saber:

- Tasa para robo con violencia o intimidación.
- Proporción de personas entre 14 y 35 años.
- Proporción de personas con educación secundaria o terciaria.
- Variable indicadora para tasa de robo por sorpresa.
- Quince variables indicadoras para regiones.

Finalmente, para el indicador de tasa de victimización agregada se incluyeron veinte covariables, a saber:

- Tasa para robo con violencia o intimidación.
- Tasa de robo con violencia en el hogar.
- Proporción de personas con educación secundaria o terciaria.
- Proporción de desocupados.
- Variable indicadora para comunas outlier.
- Quince variables indicadoras para regiones.

5.4.2. Ajustes de modelos sobre indicadores de victimización para la estimación de áreas pequeñas

Para producir estimaciones confiables de EBLUP, es necesario ajustar modelos a nivel de área. Aunque el objetivo principal de los modelos de estimación de áreas pequeñas es mejorar la confiabilidad de las estimaciones directas, los resultados del modelo

proporcionan un conjunto consistente de información sobre la explicación a nivel macro de la victimización y, por lo tanto, analizamos estos resultados a continuación.

En la **Tabla 3** se observa que la tasa de denuncias por robo con violencia o intimidación es la covariable más explicativa ($\beta^{EBLUP} = 0,0795$, con *valor p* < 0,001). Así, una mayor tasa de denuncias por robo con violencia explica una mayor tasa de robo con violencia o intimidación en los hogares. La segunda covariable más explicativa es la proporción de personas entre 14 y 35 años ($\beta^{EBLUP} = 0,0170$, con *valor p* < 0,07). Además, se observa que el modelo posee buen ajuste, con un R^2 de Hidiroglou de 92% aprox.

La **Tabla 4** muestra que, entre todas las covariables, la proporción de personas entre 14 y 35 años es la más explicativa ($\beta^{EBLUP} = 0,06309$, con *valor p* < 0,000001). Una mayor proporción de personas en este tramo etario explica una mayor tasa de robo por sorpresa en los hogares. La bondad de ajuste del modelo de 95% aprox., da cuenta de un buen ajuste.

Finalmente, en la **Tabla 5** se muestra que, entre todas las covariables, la tasa de robo con violencia en el hogar es la más explicativa ($\beta^{EBLUP} = 0,002$, con *valor p* < 0,001). La bondad de ajuste del modelo de 90% aprox., da cuenta de un buen ajuste.

Tabla 3. Covariables en modelos SAE para RVI.

	Coefficientes	Error estándar	Valor t	Valor p
(Intercepto)	0,1291	0,0206	6,2729	0,0000
Tasa de robo con violencia o intimidación	0,0795	0,0158	5,0224	0,0005
Proporción de personas entre 14 y 35 años	0,0170	0,0092	1,8390	0,0659
Dummy Reg. Tarapacá	0,0755	0,0326	2,3144	0,0206
Dummy Reg. Antofagasta	0,0237	0,0276	0,8590	0,3903
Dummy Reg. Atacama	0,0105	0,0332	0,3161	0,7520
Dummy Reg. Coquimbo	0,0326	0,0275	1,1852	0,2359
Dummy Reg. Valparaíso	0,0234	0,0239	0,9791	0,3275
Dummy Reg. O'Higgins	0,0060	0,0271	0,2206	0,8254
Dummy Reg. Maule	-0,0371	0,0258	-1,4385	0,1503
Dummy Reg. Biobío	0,0532	0,0247	2,1528	0,0313
Dummy Reg. Araucanía	0,0007	0,0280	0,0264	0,9790
Dummy Reg. Los Lagos	0,0077	0,0290	0,2663	0,7900
Dummy Reg. Aysén	-0,0389	0,0351	-1,1090	0,2675
Dummy Reg. Magallanes	-0,0934	0,0376	-2,4846	0,0130
Dummy Reg. Metropolitana	0,0913	0,0217	4,2111	0,0254
Dummy Reg. Los Ríos	-0,0243	0,0299	-0,8142	0,4155
Dummy Reg. Arica y Parinacota	0,0534	0,0377	1,4180	0,1562
AIC	-335,3177			

	Coeficientes	Error estándar	Valor t	Valor p
BIC	-286,0104			
$R^2_{Hidiroglou}$	0,9199			

Fuente: Elaboración propia.

Tabla 4. Covariables en modelos SAE para RPS.

	Coeficientes	Error estándar	Valor t	Valor p
Tasa de robo con violencia o intimidación	0,00005	0,00002	2,66000	0,00781
Proporción de personas con educación secundaria o terciaria	0,11292	0,03663	3,08320	0,00205
Variable indicadora para robo por sorpresa (tasas más altas)	0,08029	0,01604	5,00600	0,00056
Proporción de personas entre 14 y 35 años	0,06309	0,01076	5,86130	0,00000
Dummy Reg. Tarapacá	0,09762	0,02489	3,92230	0,08770
Dummy Reg. Antofagasta	0,05430	0,02331	2,32890	0,01986
Dummy Reg. Atacama	0,01889	0,02318	0,81510	0,41503
Dummy Reg. Coquimbo	0,03809	0,02485	1,53310	0,12525
Dummy Reg. Valparaíso	0,07168	0,02058	3,48390	0,00049
Dummy Reg. O'Higgins	0,02745	0,02077	1,32160	0,18629
Dummy Reg. Maule	0,02202	0,02028	1,08560	0,27766
Dummy Reg. Biobío	0,02408	0,01901	1,26700	0,20517
Dummy Reg. Araucanía	0,06189	0,02399	2,57970	0,00989
Dummy Reg. Los Lagos	0,02644	0,02285	1,15710	0,24722
Dummy Reg. Aysén	-0,05762	0,02472	-2,33080	0,01976
Dummy Reg. Magallanes	-0,07398	0,03014	-2,45460	0,01411
Dummy Reg. Metropolitana	0,05216	0,01923	2,71220	0,00668
Dummy Reg. Los Ríos	0,03860	0,02595	1,48790	0,13679
Dummy Reg. Arica y Parinacota	0,01903	0,03076	0,61870	0,53614
AIC	-347,5282			
BIC	-296,034			
$R^2_{Hidiroglou}$	0,94781			

Fuente: Elaboración propia.

Tabla 5. Covariables en modelos SAE para VA.

	Coefficientes	Error estándar	Valor t	Valor p
Tasa de robo con violencia e intimidación	0,000	0,000	1,981	0,048
Tasa de robo con violencia en el hogar	0,002	0,001	3,437	0,001
Proporción de personas desocupadas	2,318	1,146	2,023	0,043
Proporción de personas con educación secundaria o terciaria	0,146	0,062	2,362	0,018
Dummy Reg. Tarapacá	0,048	0,034	1,415	0,157
Dummy Reg. Antofagasta	-0,034	0,031	-1,078	0,281
Dummy Reg. Atacama	-0,013	0,033	-0,412	0,680
Dummy Reg. Coquimbo	-0,052	0,027	-1,928	0,054
Dummy Reg. Valparaíso	0,001	0,026	0,032	0,975
Dummy Reg. O'Higgins	0,005	0,028	0,178	0,859
Dummy Reg. Maule	-0,057	0,027	-2,127	0,033
Dummy Reg. Biobío	-0,023	0,025	-0,895	0,371
Dummy Reg. Araucanía	-0,023	0,028	-0,808	0,419
Dummy Reg. Los Lagos	-0,051	0,028	-1,781	0,075
Dummy Reg. Aysén	-0,037	0,035	-1,064	0,287
Dummy Reg. Magallanes	-0,108	0,035	-3,047	0,002
Dummy Reg. Metropolitana	-0,008	0,027	-0,292	0,771
Dummy Reg. Los Ríos	-0,013	0,029	-0,433	0,665
Dummy Reg. Arica y Parinacota	0,029	0,037	0,767	0,443
Variable suplementaria	0,093	0,016	5,642	0,000
AIC	-346,688			
BIC	-288,9386			
$R^2_{Hidiroglou}$	0,890704			

Fuente: Elaboración propia.

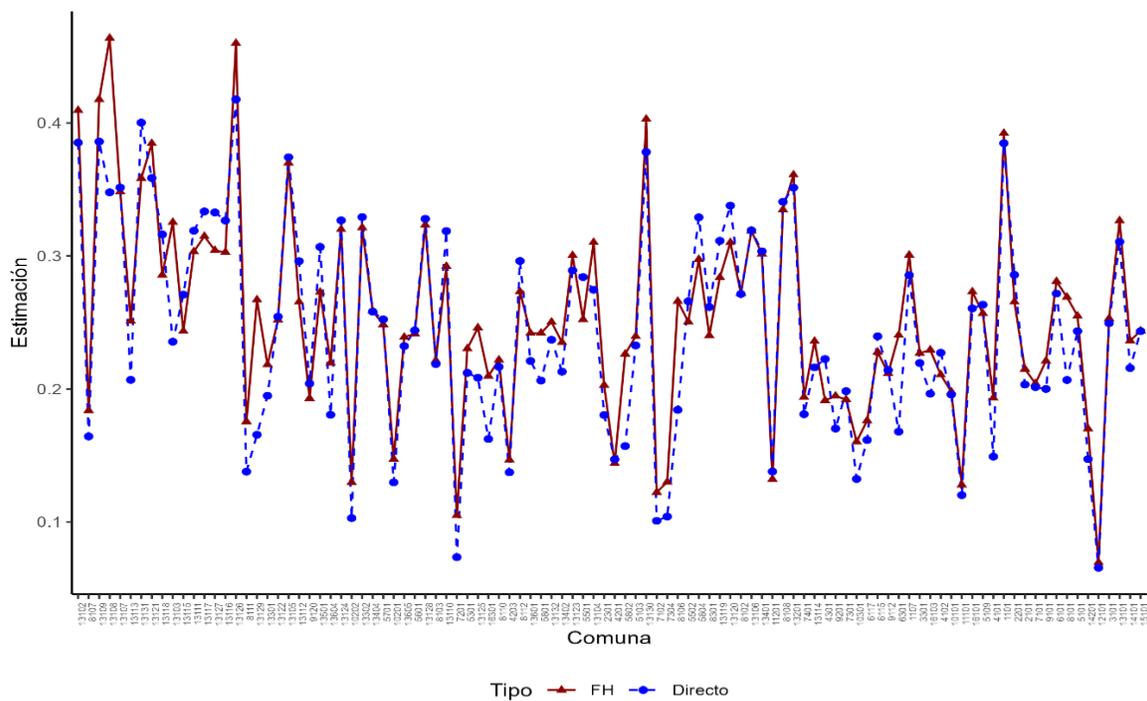
5.4.3. Estimaciones de áreas pequeñas para los indicadores de victimización a nivel comunal

Como se señaló anteriormente, para la construcción del EBLUP de los indicadores RVI y RPS, se utilizó la transformación arcoseno, por lo tanto, se corroboró que $\mathbf{z}_d^T \hat{\boldsymbol{\beta}} \in [0, \pi/2]$, es decir, que las estimaciones sintéticas se encuentren en el rango válido. Además, las estimaciones EBLUP para los tres indicadores trabajados, fueron ajustadas según el truncamiento descrito en la **Subsección 3.4** y por el factor λ_r .¹⁵ descrito en la **Subsección 4.5**, con valores entre 0,98 y 1,14. La aplicación de este factor asegura una consistencia entre las estimaciones comunales y las estimaciones regionales, oficialmente publicadas.

En las **Figura 3**, **Figura 4** y **Figura 5** se muestran las estimaciones para robo con violencia o intimidación, robo por sorpresa y victimización agregada, respectivamente, con las comunas ordenadas según tamaño muestral de menor a mayor en el eje de las abscisas. Se observa que, en general, las estimaciones EBLUP (en rojo) son más conservadoras que las obtenidas utilizando los estimadores directos (azul) en las áreas con menor tamaño de muestra. Además, como se esperaba, para las áreas con los tamaños de muestra más grandes, ambas estimaciones son muy similares.

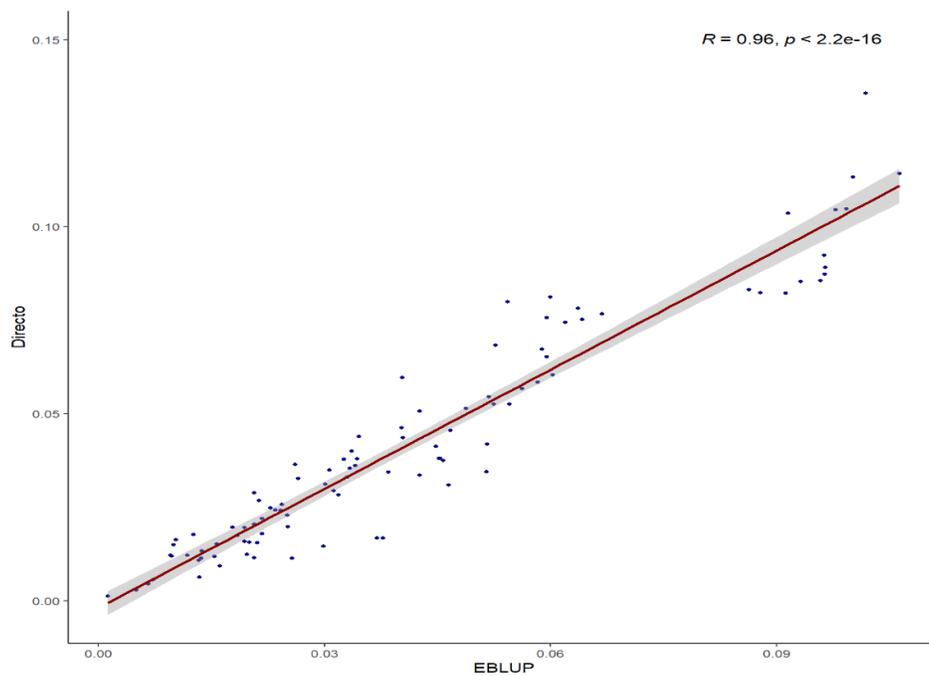
¹⁵ En estricto rigor, el ajuste del factor regional para conseguir la propiedad de *benchmarking* se aplica una vez que han sido realizados los diagnósticos del modelo (ver **Subsección 5.6**). Aquí se presentan las estimaciones con dicho ajuste a fin de que el lector aprecie las estimaciones finales.

Figura 5. Estimaciones comunales para VA.



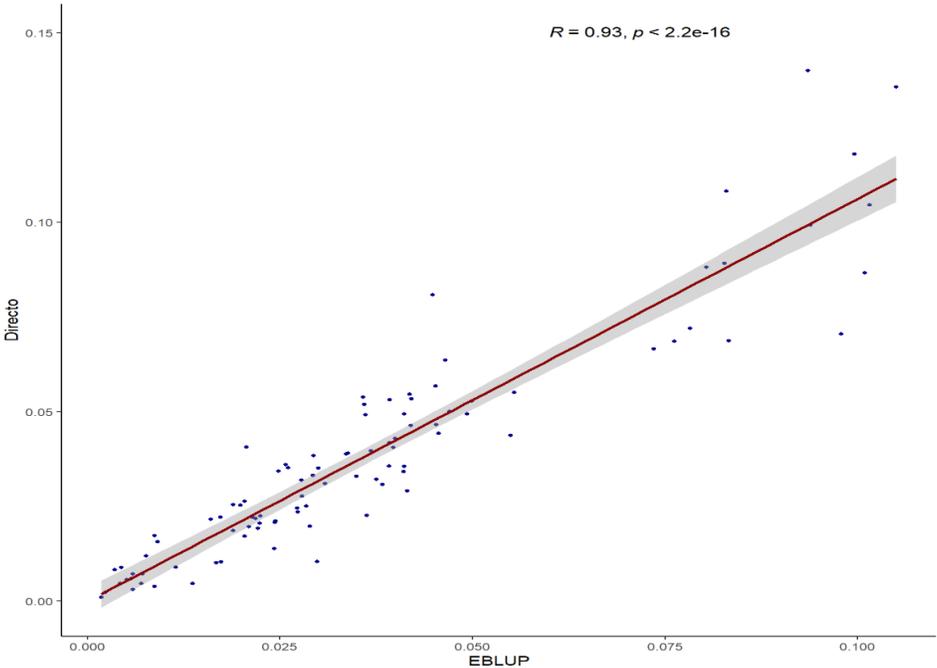
Fuente: Elaboración propia.

Figura 6. Estimación directa vs. Estimación EBLUP para la tasa de victimización por RVI.



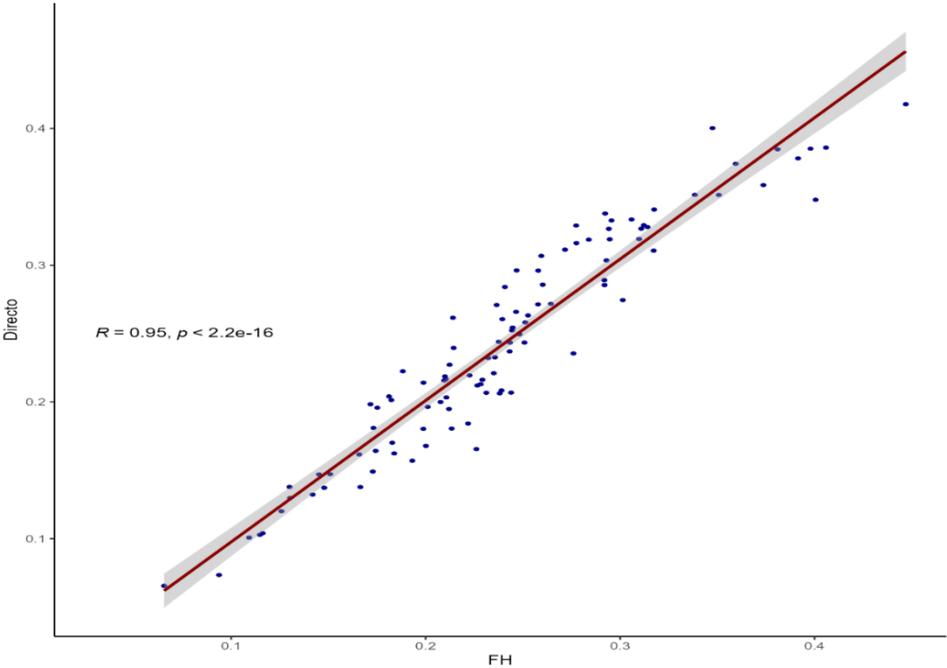
Fuente: Elaboración propia.

Figura 7. Estimación directa vs. Estimación EBLUP para tasa de victimización por RPS.



Fuente: Elaboración propia.

Figura 8. Estimación directa vs. Estimación EBLUP para tasa de VA.



Fuente: Elaboración propia.

Las **Figura 6**, **Figura 7** y **Figura 8** muestran gráficos de dispersión de las estimaciones para la tasa de victimización por robo con violencia o intimidación, robo por sorpresa y victimización agregada, respectivamente. Estos gráficos permiten investigar si las estimaciones basadas en el modelo son aproximadamente insesgadas. El objetivo del procedimiento de estimación adoptado aquí es lograr un equilibrio entre varianza y sesgo, produciendo estimaciones con buena precisión y con el menor sesgo posible. Los resultados muestran que las estimaciones directas y basadas en el modelo parecen seguirse entre sí en los tres casos.

Para evaluar más a fondo la presencia de sesgo en las estimaciones basadas en modelos, se ajustó una línea de regresión para las estimaciones directas y basadas en modelos. La estimación del coeficiente de pendiente fue significativa para ambos indicadores (para robo con violencia o intimidación $r = 0,96$, con $valor\ p < 2,2e - 16$ y para robo por sorpresa $r = 0,93$, con $valor\ p < 2,2e - 16$) y tasa de victimización $r = 0,95$, con $valor\ p < 2,2e - 16$), lo que da cuenta de la consistencia entre las estimaciones obtenidas mediante el muestreo y el modelo EBLUP.

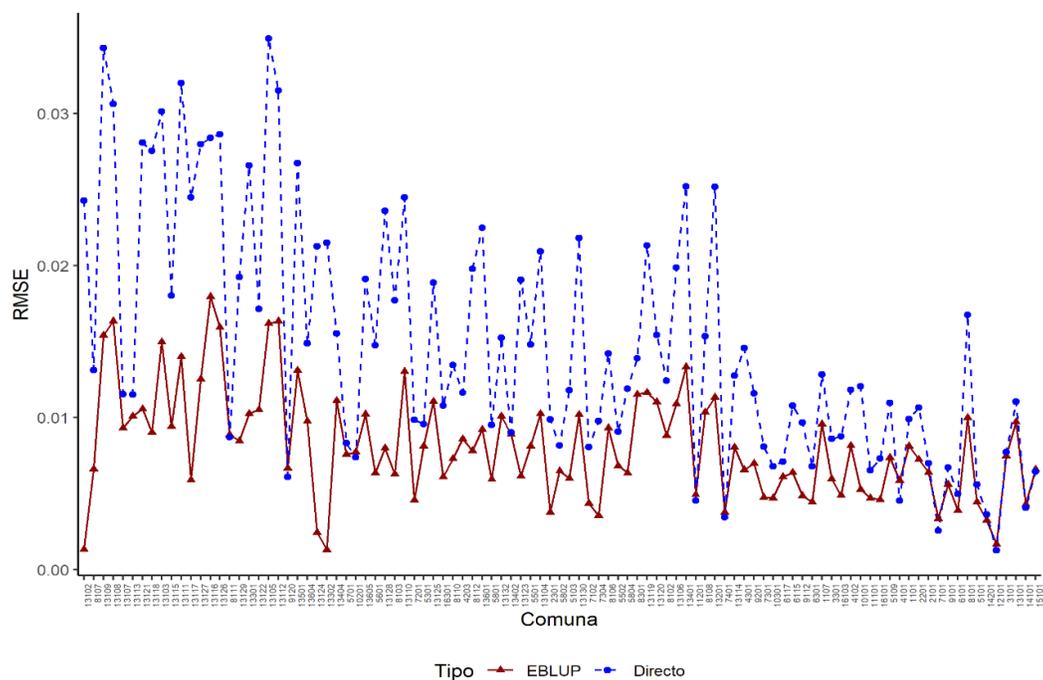
5.5. Evaluación de la fiabilidad de las estimaciones

Para comprobar la fiabilidad de las estimaciones, las **Figura 9**, **Figura 10** y **Figura 11** muestran los RMSE estimados de las estimaciones directas y EBLUP. En ambas figuras se observa que, en general, las estimaciones de áreas pequeñas obtenidas a través de EBLUP (rojo) fueron significativamente más eficientes que las estimaciones directas (azul), especialmente para las áreas con el tamaño de muestra más pequeño. En la **Tabla 6** se muestran estadísticas de resumen para las estimaciones obtenidas. Como era de esperar, el CVL promedio del EBLUP es menor al de las estimaciones directas. Para la tasa de victimización por robo con violencia o intimidación, el CVL promedio se reduce del 28,5% de las estimaciones directas al 16,7% del EBLUP. Esta reducción también se muestra en el robo por sorpresa: del 29,2% de las estimaciones directas se reduce al 16,1% de las EBLUP. Finalmente, para el indicador de victimización agregada el CVL pasa de ser 22,7% en la estimación directa a 15,6% en la estimación EBLUP. Un $CVL < 20\%$ da cuenta de fiabilidad de las estimaciones.

Tabla 6. Resumen de estimaciones de áreas pequeñas de indicadores de victimización y CVL promedio.

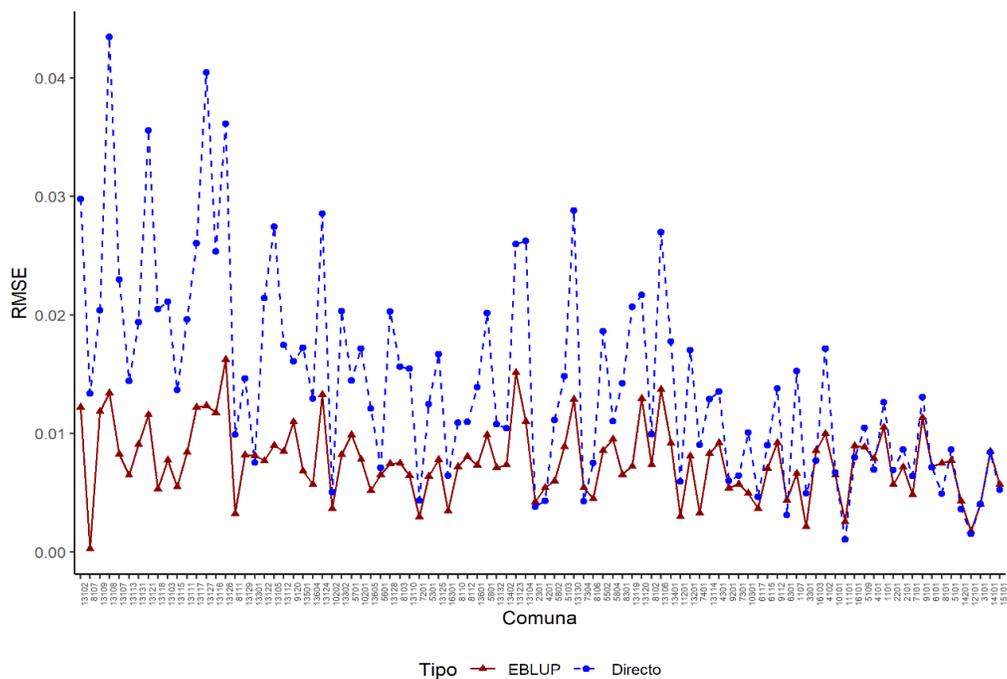
	Mín.	Cuartil 1	Media	Mediana	Cuartil 3	Máx.	Promedio CVL
Robo con violencia o intimidación							
Directo	0,0000	0,0167	0,0423	0,0344	0,0602	0,1506	0,2846
EBLUP	0,0013	0,0211	0,0535	0,0355	0,0613	0,1109	0,1671
Robo por sorpresa							
Directo	0,0000	0,0188	0,0382	0,0325	0,0514	0,1400	0,2922
EBLUP	0,0024	0,0230	0,0411	0,0351	0,0495	0,1225	0,1606
Victimización agregada							
Directo	0,0654	0,0195	0,2362	0,2431	0,3097	0,4178	0,2272
EBLUP	0,0695	0,2109	0,2472	0,2537	0,3000	0,4638	0,1558

Fuente: Elaboración propia.

Figura 9. Estimaciones RMSE comunales para tasa de victimización por RVI en (comunas ordenadas por tamaño muestral) en hogares.

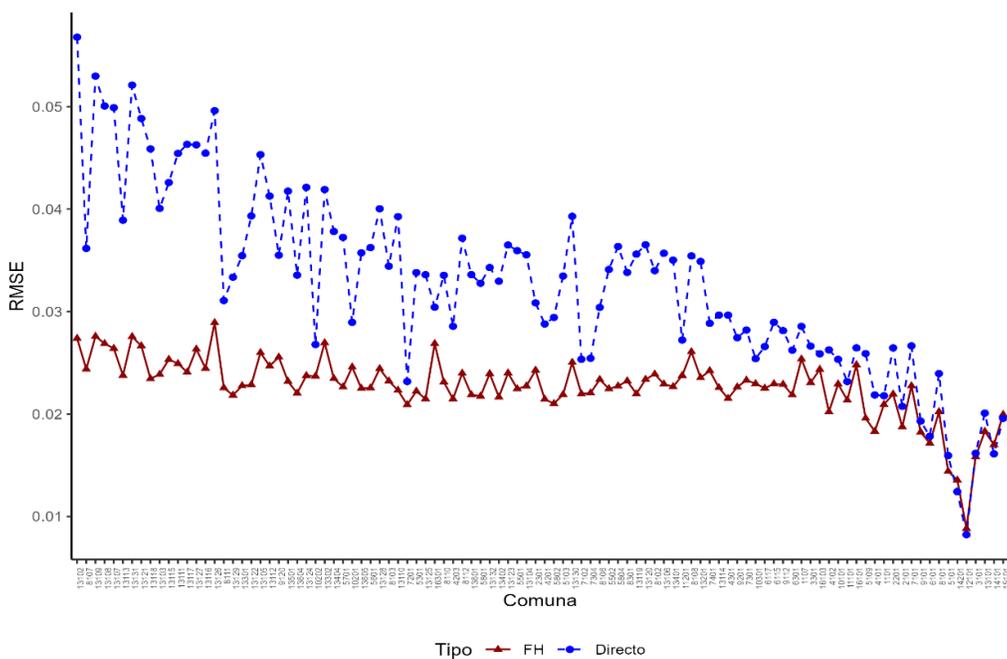
Fuente: Elaboración propia.

Figura 10. Estimaciones RMSE comunales para tasa de victimización por RPS (comunas ordenadas por tamaño muestral) en hogares.



Fuente: Elaboración propia.

Figura 11. Estimaciones RMSE comunales para tasa de VA (comunas ordenadas por tamaño muestral) en hogares.



Fuente: Elaboración propia.

5.6. Diagnósticos de los modelos

Se presentan diagnósticos de los modelos EBLUP para examinar si las estimaciones están o no sesgadas por los modelos y verificar la validez de estos. Se presentan las gráficas q-q en la **Figura 12** para corroborar la normalidad de los residuos estandarizados. Se observa que los residuos siguen un patrón normal con variaciones en las colas. En la **Tabla 7** se presenta un resumen de tres test estadísticos de normalidad aplicados, los que sugieren que no se rechaza la hipótesis nula de la distribución normal para los tres indicadores. Además, el test para homocedasticidad de Breusch-Pagan da $BP = 17,56$, con $valor p = 0,41$ para robo con violencia o intimidación y $BP = 14,62$, con $valor p = 0,69$ para robo por sorpresa y $BP = 12,93$, con $valor p = 0,88$ para victimización agregada, lo que sugiere que no se rechaza la hipótesis nula de varianza homocedástica.

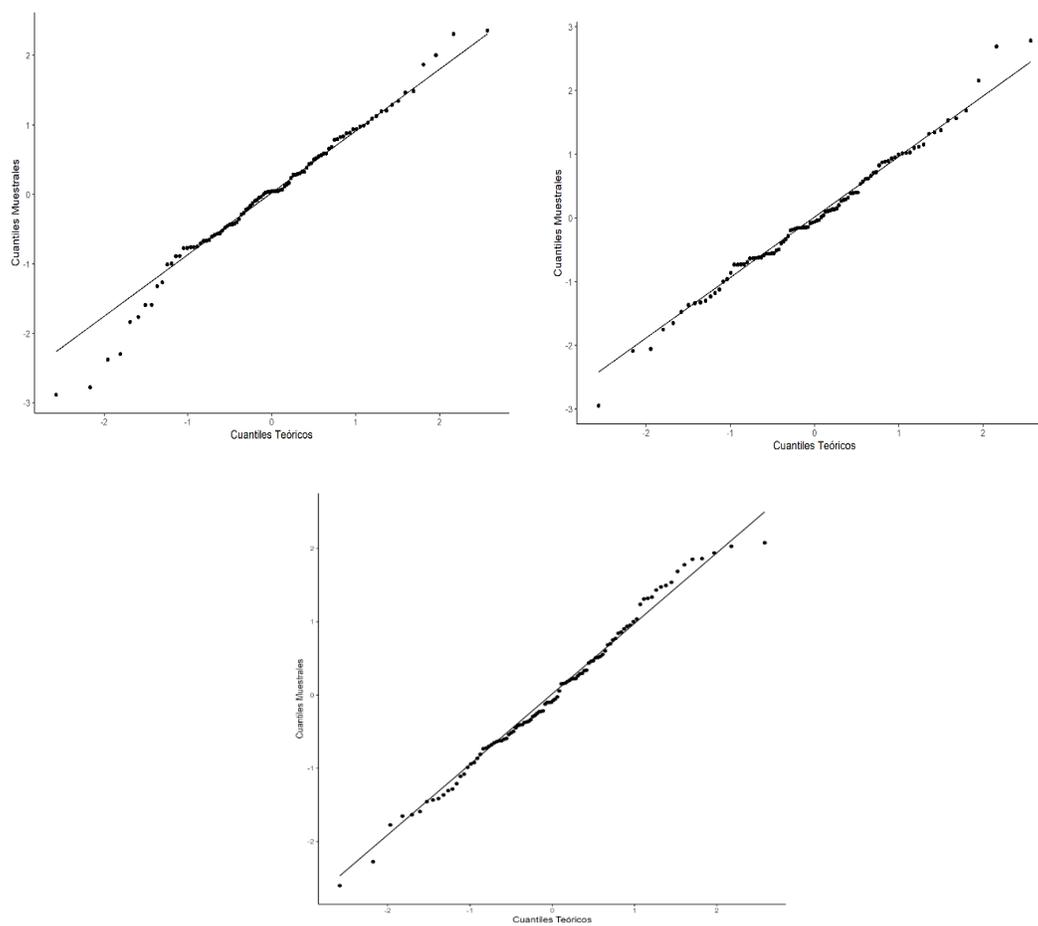
Finalmente, la **Figura 13** muestra que para ningún indicador hay presencia de valores atípicos mayores a 3, y la **Figura 14** muestra que no hay valores influyentes, es decir, no hay distancias de Cook mayores a 1.

Tabla 7. Resumen de test de normalidad y de homocedasticidad.

	Tests de normalidad			Test de homocedasticidad
	Jarque-Bera	Shapiro	D'Agostino	Breusch-Pagan
<i>Robo con violencia o intimidación</i>				
Estadístico	JB=4,1022	W=0,98005	z=-1,57638	BP=17,561
Valor p	0,0805	0,1382	0,1149	0,4171
<i>Robo por sorpresa</i>				
Estadístico	JB=1,2635	W=0,99066	z=0,341926	BP=14,619
Valor p	0,472	0,7366	0,7324	0,6879
<i>Victimización agregada</i>				
Estadístico	JB=0,5023	W=0,99012	z=0,125663	BP=12,922
Valor p	0,7779	0,6612	0,9	0,8807

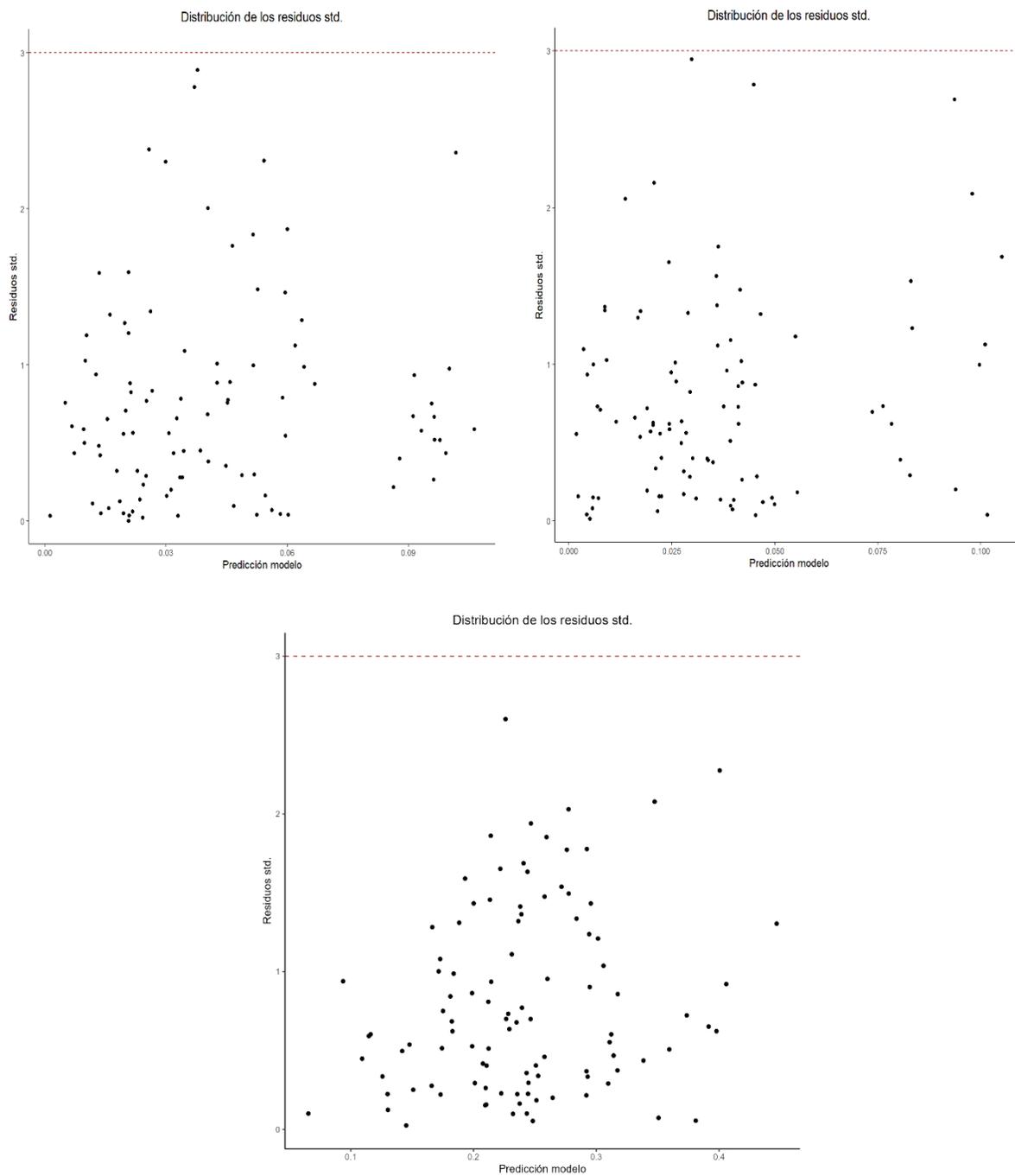
Fuente: Elaboración propia.

Figura 12. Gráficos q-q plots para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior).



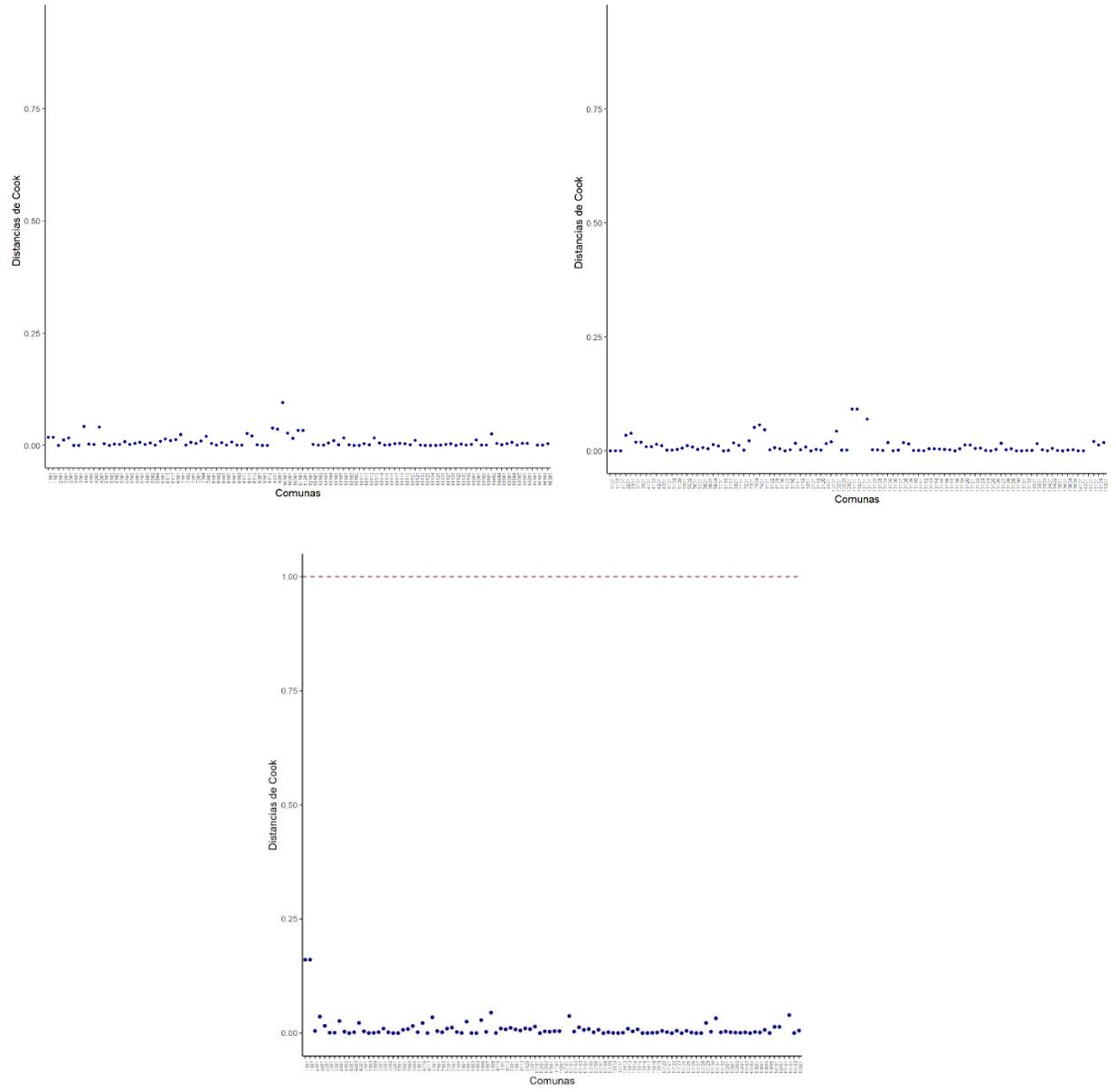
Fuente: Elaboración propia.

Figura 13. Gráficos de residuos estandarizados versus valores predichos para para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior).



Fuente: Elaboración propia.

Figura 14. Distancias de Cook para para RVI (sup. izquierda), RPS (sup. derecha) y VA (inferior).



Fuente: Elaboración propia.

5.7. Análisis de los resultados comunales

Los resultados de las estimaciones a nivel comunal para los indicadores de robo con violencia o intimidación, robo por sorpresa y victimización agregada por delitos de mayor connotación social, puede ser visualizados de manera efectiva en un mapa interactivo que permita rápidamente obtener conclusiones sobre el comportamiento de la tasa de victimización de cada uno de estos indicadores, a lo largo del país (102 comunas pertenecientes a la muestra ENUSC 2018). Puede acceder a través del siguiente [link](#).

Adicionalmente, el equipo ENUSC del Subdepartamento de Condiciones de Vida (SCV) realiza un análisis exhaustivo de los resultados derivados de la estimación SAE. Este análisis se realiza con un enfoque basado en el conocimiento experto, para evaluar la congruencia de las estimaciones con las expectativas regionales. En este sentido, se busca identificar aquellas comunas donde las estimaciones SAE exhiban un comportamiento inusual en relación con sus características sociodemográficas dentro de la región. Posteriormente, se formulan hipótesis destinadas a comprender mejor estos resultados.

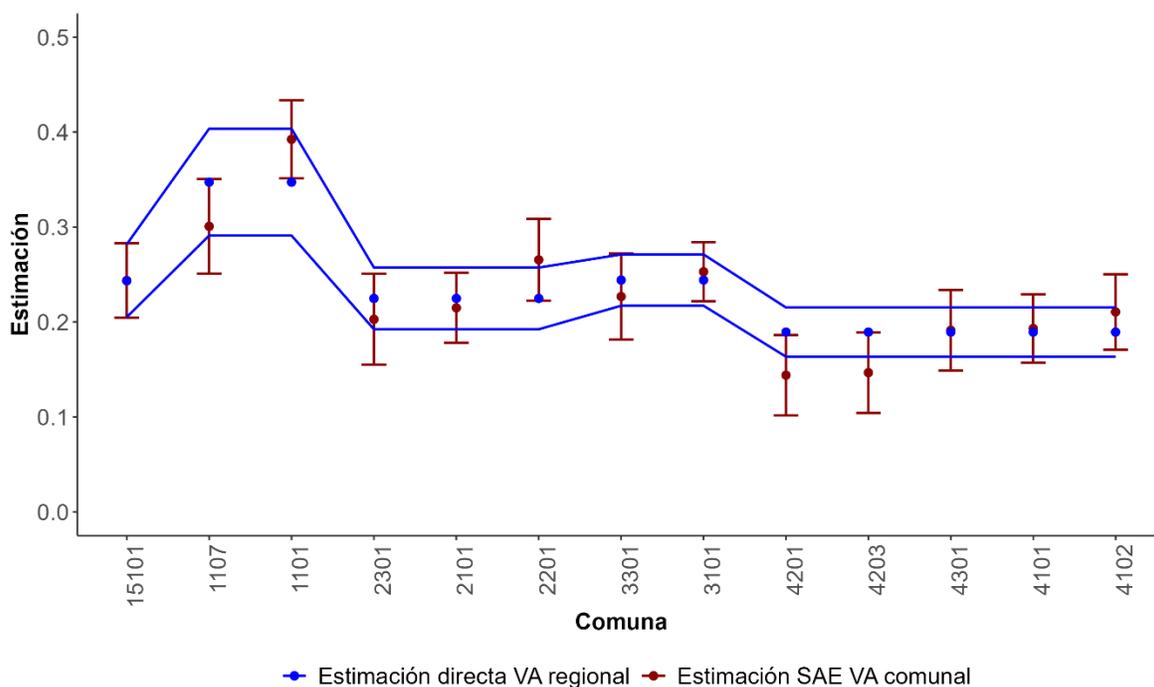
5.7.1. Comparación con resultados ENUSC

En esta sección se contrastaron las estimaciones directas a nivel regional con las SAE a nivel comunal para el indicador VA. Para realizar las comparaciones entre ambas estimaciones se consideró el intervalo de confianza de ambos datos al momento de evaluar las similitudes o diferencias, identificando alertas en los casos comunales respecto a las demás de la región. Es importante señalar que este criterio no es definitivo; más bien, representa un primer paso para evaluar la efectividad de las predicciones SAE dentro del marco propuesto y utilizando las covariables seleccionadas.

1. Regiones de la zona norte y centro norte

En las regiones del norte del país (**Figura 15**), las estimaciones de VA obtenidas por SAE y ENUSC regional, están cercanas entre sí, es decir, no es posible identificar diferencias importantes entre las comunas.

Figura 15. Estimaciones SAE y ENUSC del indicador VA para las regiones de la zona norte.



Fuente: Elaboración propia.

- **Región de Tarapacá¹⁶:** las estimaciones SAE se encuentran dentro del intervalo de confianza de la estimación regional. Sin embargo, a nivel específico, Iquique-1101¹⁷ (39,2%) sobrepasa la estimación muestral (34,7%). De acuerdo con el conocimiento del fenómeno, se espera que Iquique tenga una tasa de VA mayor que Alto Hospicio-1107 (30,1%), debido a su mayor población (Cadena & Letelier, 2018). Bajo este criterio, no se identifican alertas en la región.
- **Región de Antofagasta:** Calama-2201 (26,6%) presentó una estimación SAE mayor que Antofagasta-2101 (21,5%). La estimación SAE para Calama supera la estimación ENUSC regional (22,5%), la cual no se encuentra dentro de su intervalo de confianza. Aunque se ha identificado a Calama como una comuna con alerta bajo este criterio, según la experiencia previa, este comportamiento era previsible, por lo tanto, no se identifican alertas en esta región.

¹⁶ Se excluye el análisis de la región de Arica y Parinacota, ya que Arica es la única comuna en la muestra de ENUSC 2018.

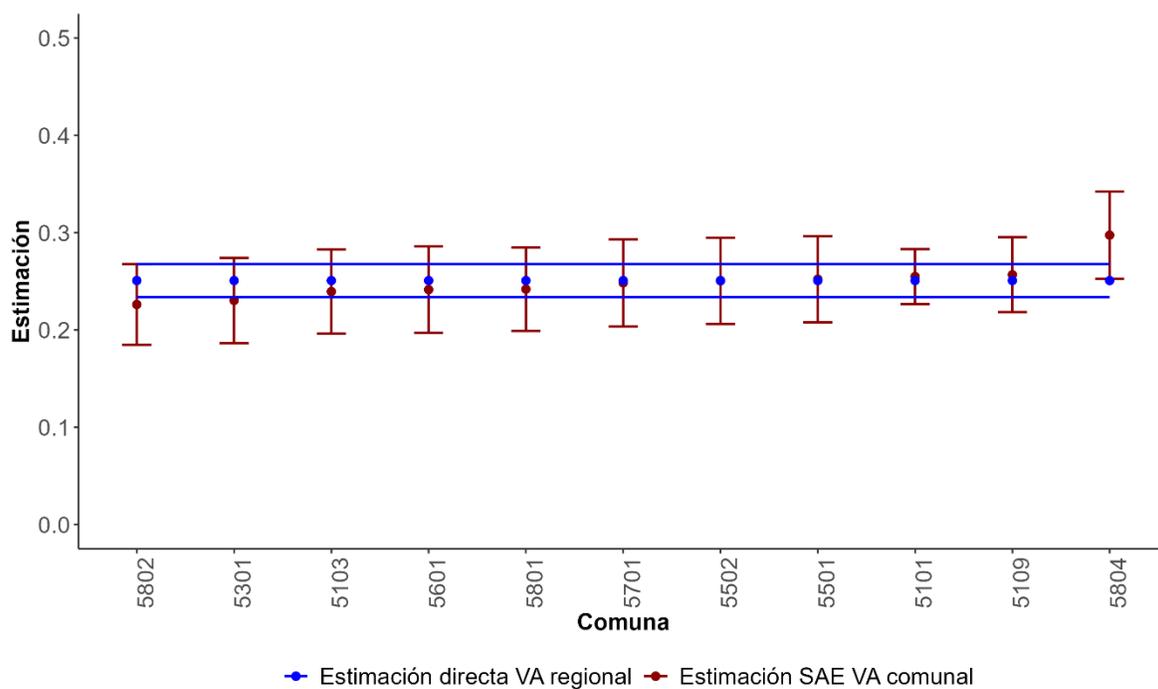
¹⁷ Para facilitar la interpretación de los resultados en esta sección, se indica la comuna y su respectivo CUT (Código Único Territorial).

- **Región de Atacama:** la estimación SAE en Copiapó-3101 (25,3%) supera tanto a Vallenar-3301 (22,7%) como a la estimación ENUSC regional (24,4%). Ambas estimaciones están dentro del intervalo de confianza regional. Bajo este criterio, no se identifican alertas en la región.
- **Región de Coquimbo:** considerando la experiencia, es lógico que las comunas con mayor población presenten estimaciones SAE más elevado en comparación con otras comunas, tal como ocurre con Illapel-4201 (14,4%) y Los Vilos-4203 (14,7%), ambas por debajo de la banda inferior del intervalo de confianza regional. En contraste, Coquimbo-4102 (21,1%) registra el indicador de VA más alto, rozando el límite superior del intervalo. Por ende, estas tres comunas se identifican como alerta en la región.

2. Región de Valparaíso

De manera exploratoria, al comparar las estimaciones SAE con las estimaciones regiones de la ENUSC 2018 para el indicador VA en la región de Valparaíso (**Figura 16**), no se observan diferencias notables.

Figura 16. Estimaciones SAE y ENUSC del indicador VA para la región de Valparaíso.



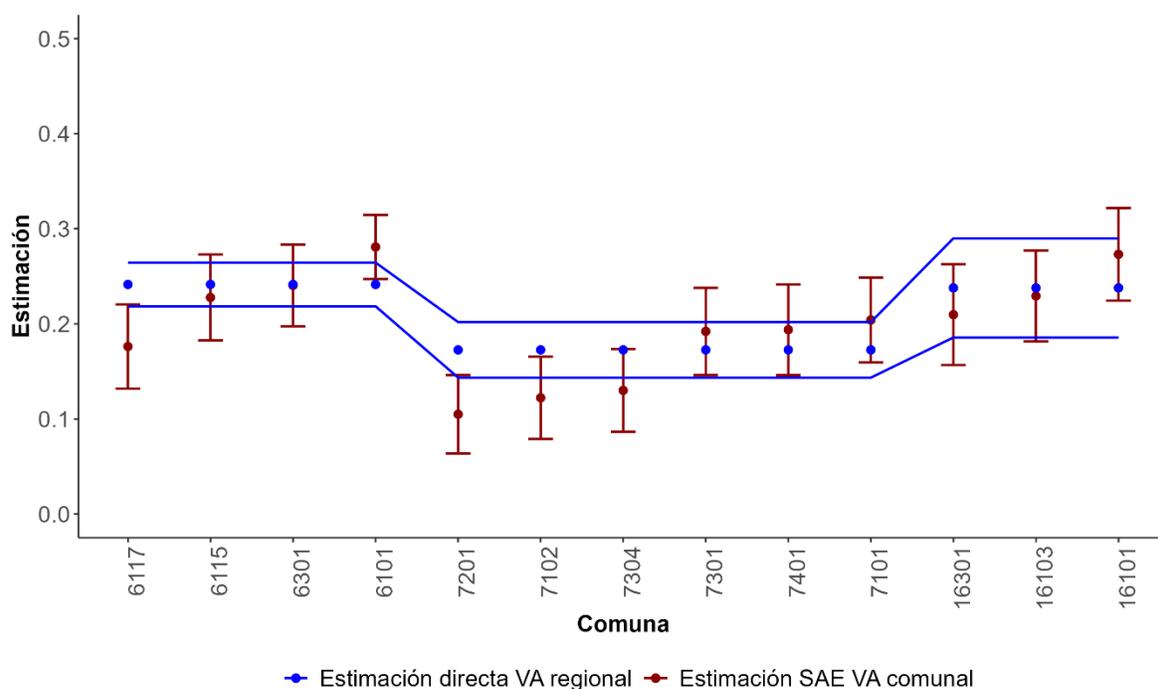
Fuente: Elaboración propia.

Las estimaciones SAE se mantienen en los intervalos de la región. Sin embargo, destaca la alta estimación SAE para Villa Alemana-5804 (29,8%), considerablemente superior a las comunas cercanas como Quilpué-5801 (24,2%) y Limache-5802 (22,6%), e incluso por encima de la estimación ENUSC para la capital regional de Valparaíso-5101 (25,5%). Las comunas identificadas como alerta según el equipo temático de la ENUSC son Limache, Los Andes y Villa Alemana. No obstante, estas comunas tienen niveles aceptables de precisión bajo la metodología SAE, con CV del 9,43%, 10,3% y 8,01% respectivamente.

3. Regiones de centro y centro sur

En la zona centro sur del país, en general, se observa que las estimaciones SAE (**Figura 17**) no difieren considerablemente de la estimación regional.

Figura 17. Estimaciones SAE y ENUSC del indicador VA para las regiones de la zona centro y centro sur.



Fuente: Elaboración propia.

- **Región de O'Higgins:** destaca la notable diferencia en la baja estimación SAE de la comuna de San Vicente-6117 (17,6%), en comparación con la estimación regional ENUSC (24,1%). En contraste, la comuna de Rancagua-6101 registró una tasa del

28,08%, superando la estimación regional, ambos comportamientos considerados dentro de lo esperado. Por otro lado, las comunas de Rengo-6115 y San Fernando-6301, se encuentran dentro del intervalo.

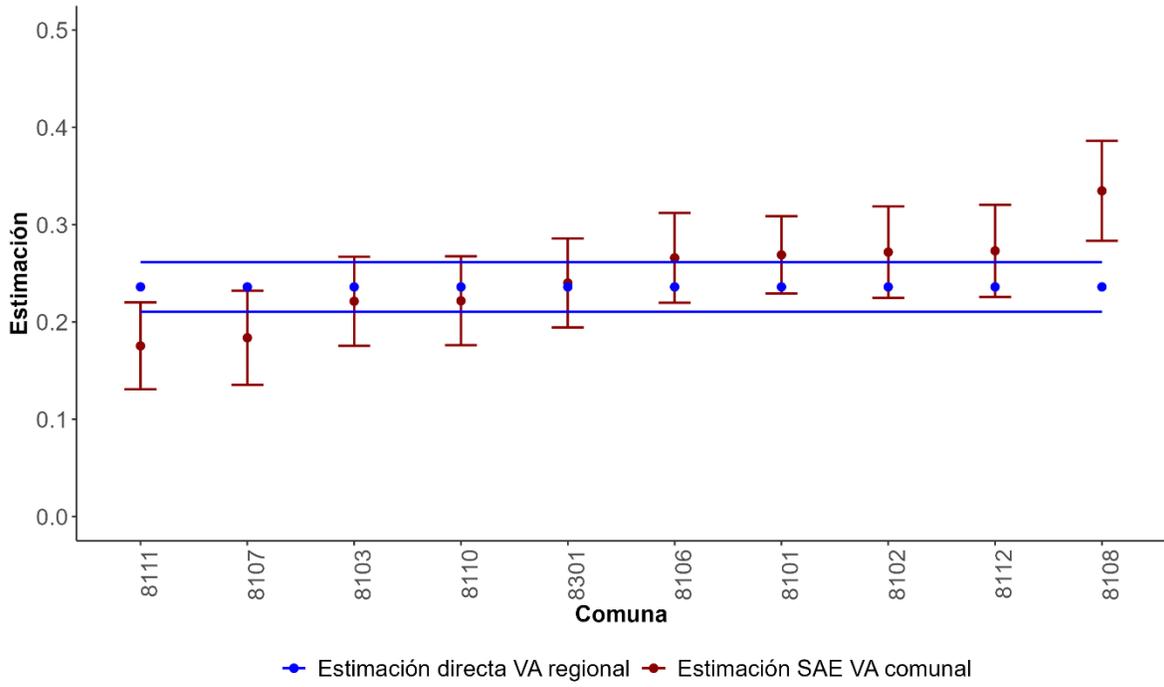
- **Región de Maule:** las comunas de Cauquenes-7201 (10,5%), Constitución-7102 (12,2%), Molina-7304 (13,0%) y Talca-7101 (20,4%), se sitúan fuera del intervalo, por lo tanto, se califican como comunas en alerta.
- **Región de Ñuble:** las estimaciones de las comunas en esta región se encuentran completamente dentro del intervalo. Bajo este criterio, no se identifican alertas en la región.

4. Región de Biobío

Para el Gran Concepción (**Figura 18**), se observa una tendencia bastante particular caracterizada por la presencia de valores bajos y altos, que se encuentran fuera de la banda de confianza. Por un lado, Tomé-8111 (17,6 %) y Penco-8107 (18,4%), muestran estimaciones SAE considerablemente inferiores a la estimación directa regional (23,6%). Por otro lado, las comunas de Lota-8106 (26,6%), Concepción-8101 (26,9%), Coronel-8102 (27,2%), Hualpén-8112 (27,3%) y San Pedro de la Paz-8108 (33,5%) exhiben estimaciones SAE superiores a la estimación regional y, en ambos casos, están fuera del intervalo de confianza. En contraste, las comunas de Chiguayante-8103 (22,1%), Talcahuano-8110 (22,2%) y Los Ángeles-8301 (24,0%) presentaron estimaciones dentro de los intervalos.

Las comunas identificadas como alerta en esta región son: Tomé, Penco, Lota, Concepción, Coronel, Hualpén y San Pedro de la Paz. Destaca que sólo esta última comuna, se desvía del comportamiento esperado de la región.

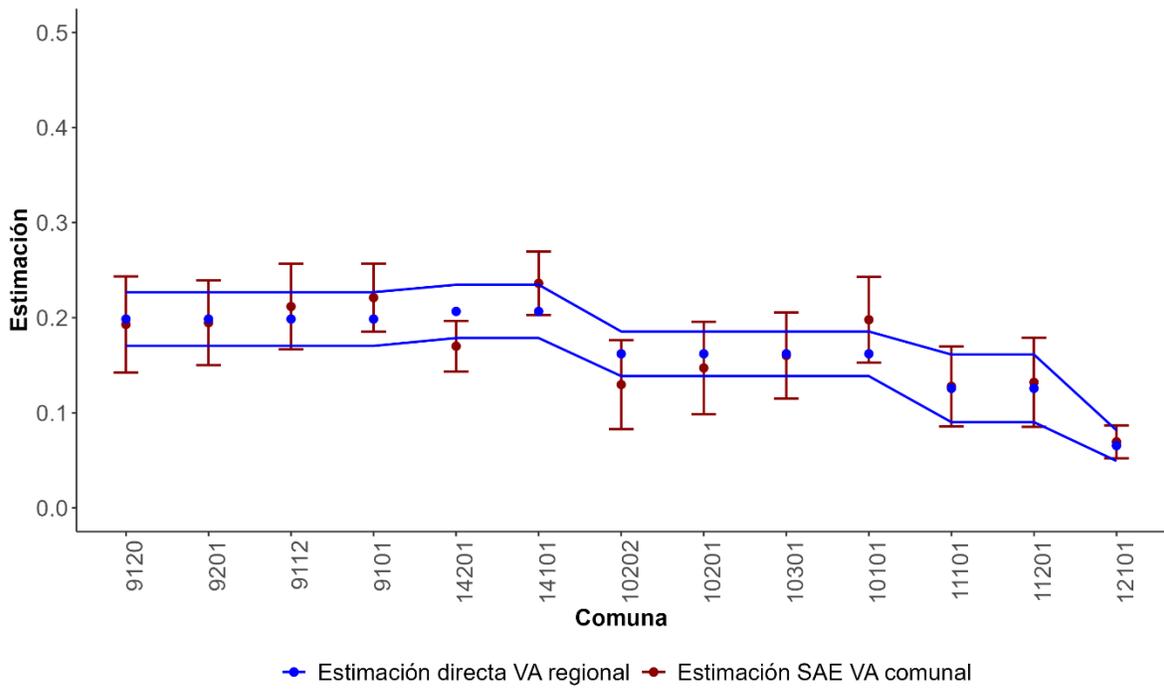
Figura 18. Estimaciones SAE y ENUSC del indicador VA para la región del Biobío.



Fuente: Elaboración propia.

5. Regiones de la zona sur

Figura 19. Estimaciones SAE y ENUSC del indicador VA para las regiones de la zona sur.



Fuente: Elaboración propia.

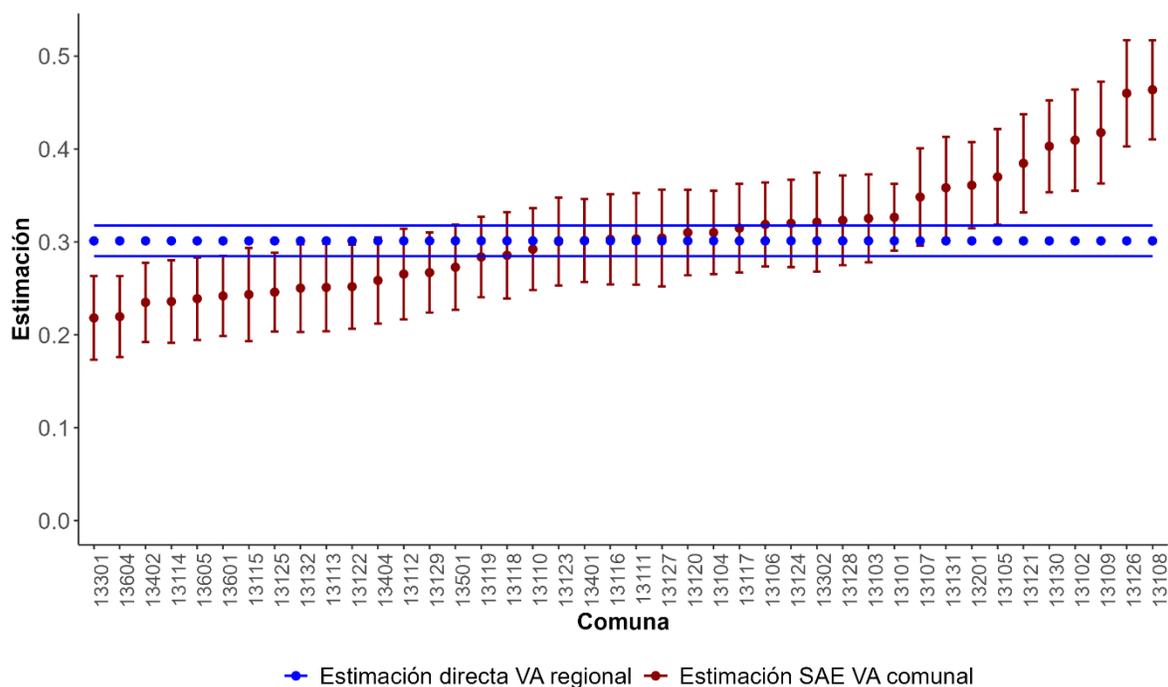
- **Región de La Araucanía:** las estimaciones SAE para las comunas de Villarrica-9120 (19,3%), Angol-9201 (19,5%), Padre las casas-9112 (21,2%) y Temuco-9101 (22,1%) se encuentran dentro del intervalo de confianza regional (**Figura 19**).
- **Región de Los Ríos:** la estimación SAE para la comuna de La Unión-14201 (17,0%) se sitúa ligeramente por debajo del intervalo (20,7%), mientras que la estimación de la comuna de Valdivia-14101 (23,6%) logra estar dentro del intervalo de confianza. Ambos comportamientos están dentro de lo esperado.
- **Región de Los Lagos:** la estimación SAE para Ancud-10202 (13,0%) se ubicó por debajo del intervalo de confianza regional. Por otro lado, la comuna de Puerto Montt-10101 (19,8%) sobrepasó el límite superior de estas bandas de confianza. Tanto Castro-10201 (14,7%) como Osorno-10301 (16,0%) se encuentran dentro del intervalo de la estimación regional. Con estos resultados, la región muestra un comportamiento esperado.
- **En la región de Aysén¹⁸:** la estimación SAE para la comuna de Punta Arenas-12101 (13,2%) se encuentra dentro de las bandas de confianza. Dadas sus características, las estimaciones obtenidas para esta comuna se consideró como un comportamiento esperable.

6. Región Metropolitana

En la Región Metropolitana, las estimaciones SAE muestra una tendencia similar a la observada en el gran Concepción, como se ilustra en la **Figura 20**. Los valores bajos de la estimación SAE, en su mayoría, se ubicaron por debajo del intervalo de confianza, mientras que los valores altos sobrepasaron este intervalo. Dentro de este contexto las comunas de Macul-13118 (28,6%), La Florida-13110 (29,2%), Providencia-13123 (30,1%), San Bernardo-13401 (30,2%), Lo Espejo-13116 (30,3%), La Granja-13111 (30,3%), Recoleta-13127 (30,4%), Ñuñoa-13120 (31,0%), Conchalí-13104 (31,0%) y Lo Prado-13117 (31,5%) se mantuvieron dentro del intervalo de confianza de la estimación regional.

¹⁸ Se excluye el análisis de la región de Magallanes, ya que Punta Arenas es la única comuna en la muestra de ENUSC 2018.

Figura 20. Estimaciones SAE y ENUSC del indicador VA para la Región Metropolitana.



Fuente: Elaboración propia.

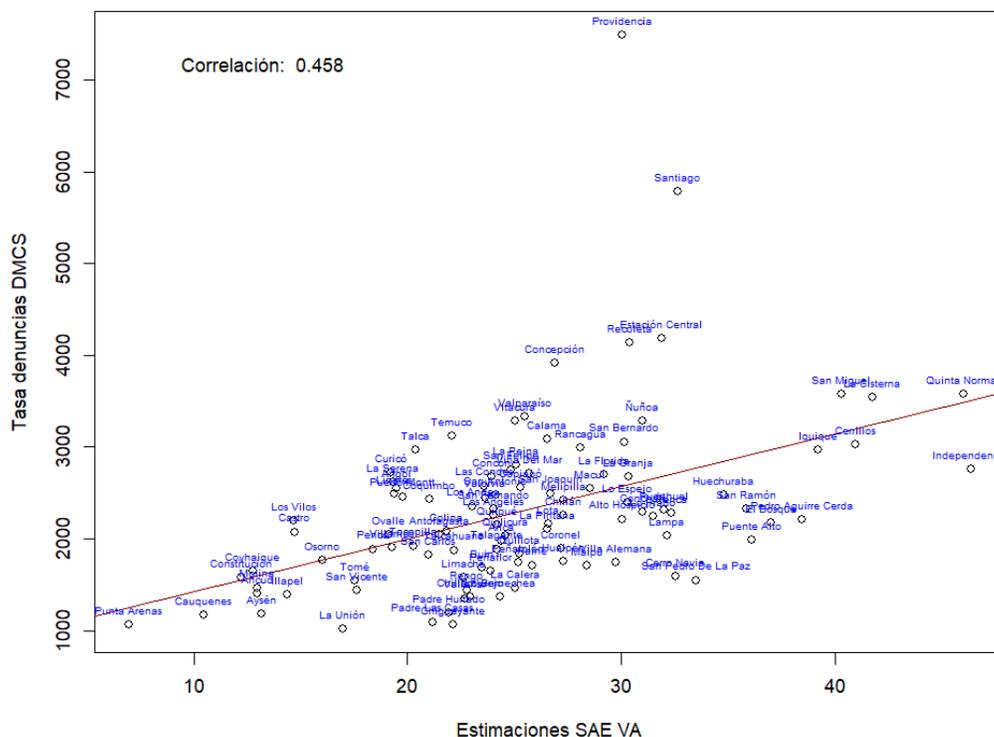
Según el criterio inicial de ubicación respecto a los intervalos de confianza proporcionados por las estimaciones regionales de la ENUSC 2018, se identificó con alerta a varias comunas. Aquellas que quedaron por debajo del límite inferior incluyen: Colina-13301 (21,8%), Padre Hurtado-13604 (22,0%), Buin-13402 (23,5%), Las Condes-13114 (23,6%), Peñaflores-13605 (23,9%), Talagante-13601 (24,2%), Lo Barnechea-13115 (24,3%), Quilicura-13125 (24,6%), Vitacura-13132 (25,0%), La Reina-13113 (25,1%), Peñalolén-13122 (25,2%), Paine-13404 (25,9%), La Pintana-13112 (26,5%), San Joaquín-13129 (26,7%), Melipilla-13501 (27,3%) y Maipú-13119 (28,4%). En contraste, las comunas que superaron el límite superior fueron: Estación Central-13106 (31,9%), Pudahuel-13124 (32,0%), Lampa-13302 (32,1%), Renca-13128 (32,3%), Cerro Navia-13103 (32,5%), Santiago-13101 (32,7%), Huechuraba-13107 (34,8%), San Ramón-13131 (35,8%), Puente Alto-13201 (36,1%), El Bosque-13105 (37,1%), Pedro Aguirre Cerda-13121 (38,5%), San Miguel-13130 (40,3%), Cerrillos-13102 (41,0%), La Cisterna-13109 (41,8%), Quinta Normal-13126 (46,0%) e Independencia-13108 (46,4%). De estas últimas, Quinta Normal e Independencia clasificaron como comunas en alerta por el equipo temático de la ENUSC.

En consecuencia, basándose en este análisis exploratorio, se han identificado las regiones que incluyen comunas con alertas: Coquimbo, Valparaíso, Maule, Biobío y Metropolitana. Estas regiones engloban aquellas estimaciones SAE que muestran un comportamiento atípico en comparación con la experiencia previa del fenómeno criminal.

5.7.2. Contraste con registros administrativos

En esta sección se revisa la correlación entre las denuncias reportadas por las fuerzas policiales (Carabineros de Chile y PDI) para los delitos de mayor connotación social (DMCS) y las estimaciones SAE del indicador VA. Como se puede apreciar en la **Figura 21**, se confirma la existencia de una asociación entre los delitos denunciados y la estimación SAE, lo que respalda el uso de los registros policiales (en este caso exclusivamente denuncias) como una referencia para evaluar la precisión de las estimaciones.

Figura 21. Correlación de denuncias por DMCS registradas por policías y estimación SAE de VA.



Fuente: Elaboración propia.

Confirmada la relación entre las denuncias y las estimaciones SAE, se examinó la tasa de denuncias policiales para cada comuna y su relación con otras comunas de características geográficas similares. Durante este análisis, se identificaron los comportamientos

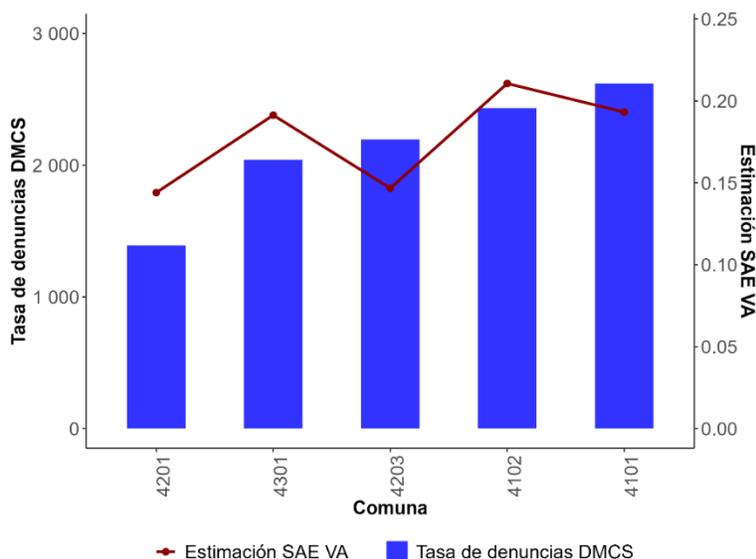
considerados atípicos o contrarios a lo esperado en términos de la cantidad de denuncias registradas.

Basándose en los hallazgos del punto anterior, este análisis se enfoca en las regiones donde identificaron comunas con estimaciones comunales SAE atípicas, o que se desvían de lo esperado, dadas sus características y su relación con otras comunas dentro de cada región. Por lo tanto, para el análisis comunal de las tasas de denuncias policiales, sólo se consideran las regiones de Coquimbo, Valparaíso, Biobío y Metropolitana¹⁹.

Al examinar la tasa de denuncias de la región de Coquimbo (**Figura 22**), se ratifica el comportamiento similar al observado en las estimaciones SAE. En este sentido, la comuna de Illapel-4201, que es considerablemente más pequeña en tamaño y lejana a la capital regional, muestra las tasas de denuncias más bajas. Mientras tanto, La Serena-4101 y Coquimbo-4102, que forman parte de la conurbación de la capital regional, tienen tasas de denuncias significativamente más altas y cercanas entre sí.

Un caso particular para destacar es el de la comuna de los Vilos-4203, que, a pesar de tener una población menor que Illapel (Ver **Anexo D 8.4.**), presenta una tasa de denuncias más elevada. Esta variación se incorpora en un grado mínimo en la estimación SAE, ya que en Los Vilos la estimación es de 14,7%, mientras que en Illapel es de 14,1%.

Figura 22. Estimaciones SAE para VA y Tasa de denuncia para DMCS, Región de Coquimbo.



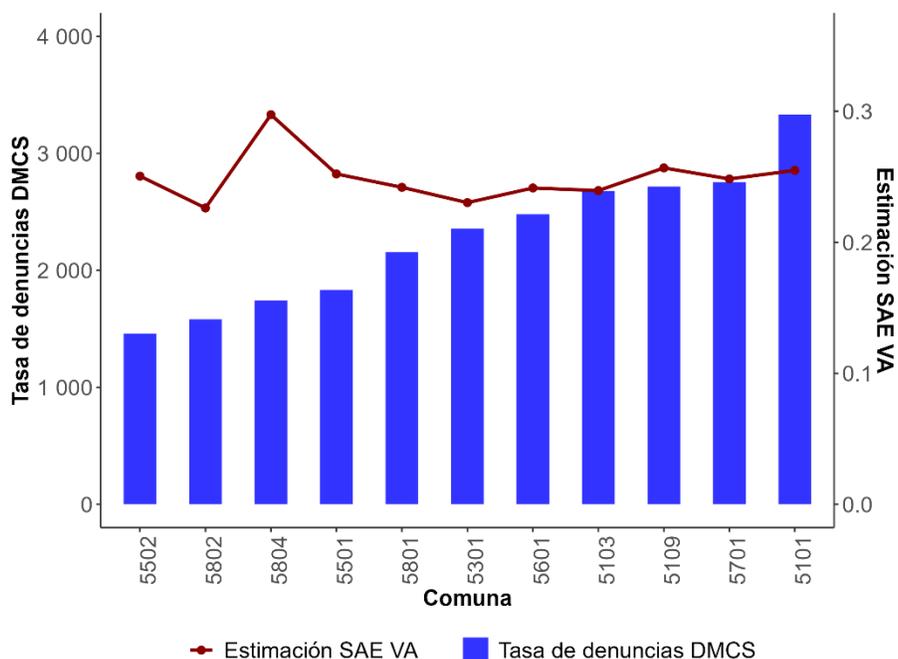
Fuente: Elaboración propia.

¹⁹ Pese a los hallazgos en Maule en el capítulo anterior, se excluyó a esta región del análisis dada la coherencia de los registros de policías respecto al indicador SAE. Ver tabla en **Anexo D 8.4.**

Al examinar los datos de la región de Valparaíso (**Figura 23**), se observa que la comuna de Villa Alemana-5804, que registró la mayor estimaciones SAE, tiene una tasa de denuncias cercana a los 2.000 delitos. Esta cifra es similar a las tasas de denuncias de sus comunas vecinas, Limache-5802 y Quilpué-5801, todas ubicadas en la Provincia de Marga Marga y con poblaciones urbanas significativamente menores que la conurbación Valparaíso - Viña del Mar. En contraste, esta última presenta tasas de denuncias cercanas a los 4.000 delitos por cada 100.000 habitantes.

Estos hallazgos refuerzan las alertas previamente mencionadas sobre el resultado aparentemente contradictorio de Villa Alemana en las estimaciones SAE, donde se muestra con la mayor tasa de victimización, según advierte el equipo temático de la ENUSC.

Figura 23. Estimaciones SAE para VA y Tasa de denuncia para DMCS, Región de Valparaíso.

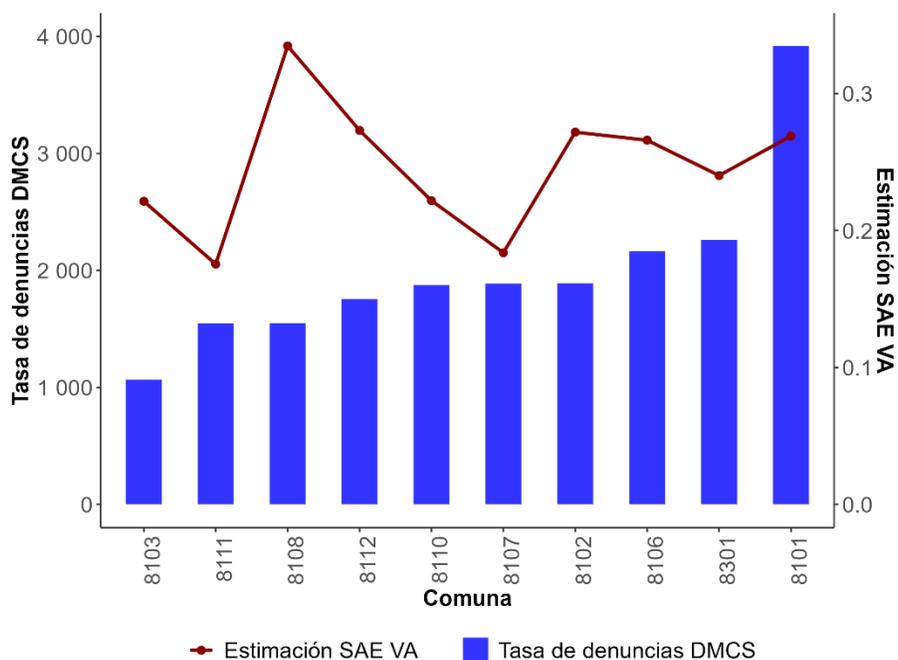


Fuente: Elaboración propia.

En cuanto a la región del Biobío (**Figura 24**), se destaca que la mayoría de las comunas que forman parte del Gran Concepción tienen tasas de denuncias por debajo de los 2.000 delitos por cada 100.000 habitantes, mientras que la comuna de Concepción registra una tasa cercana a los 4.000. Esta relación se refleja en las estimaciones SAE de la mayoría de las comunas, excepto San Pedro de la Paz-8108. A pesar de ser una comuna con una población

menor que Concepción-8101 y una tasa de denuncias de 1.549,6, San Pedro de la Paz presenta una estimación SAE (33,5%) considerablemente más alta que la de Concepción (22,9%), siendo esta última la comuna con la más alta tasa de denuncias de la región.

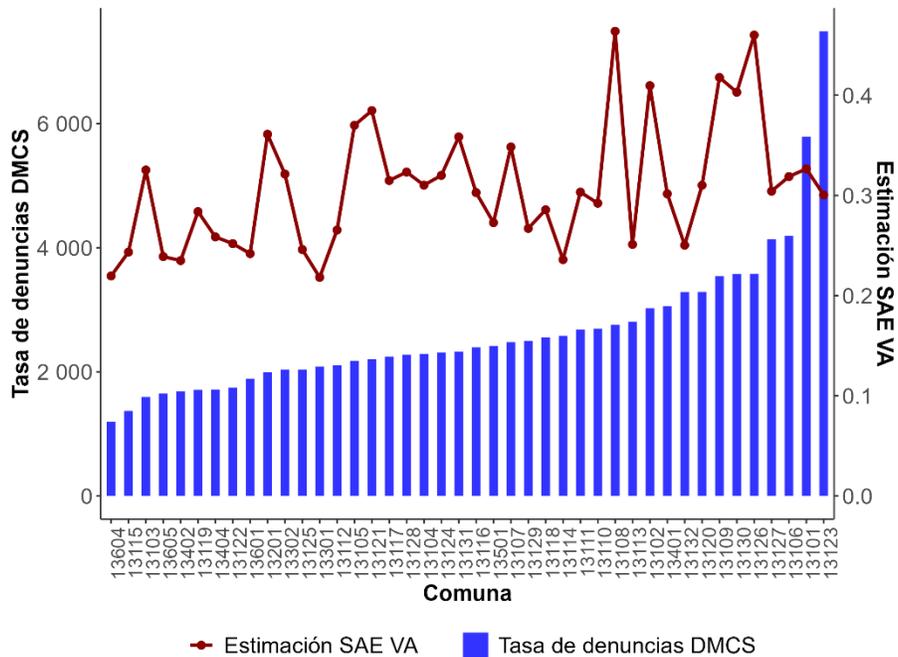
Figura 24. Estimaciones SAE para VA y Tasa de denuncia para DMCS, Región de Biobío.



Fuente: Elaboración propia.

En la Región Metropolitana, el análisis basado en los registros de denuncias revela algunas discrepancias respecto a las estimaciones SAE (**Figura 25**). En primer lugar, las comunas con las tasas de denuncias más altas, como Santiago-13101 y Providencia-13123, muestran estimaciones SAE cercanas a la estimación directa regional de ENUSC. La comuna de Independencia destaca por tener mayor estimación SAE con un 46,4%, aunque su tasa de denuncia es cercana al promedio de la región. Resulta llamativo también que la comuna de La Reina-13113, con una tasa de denuncias similar a Independencia-13108, tenga una estimación SAE considerablemente más baja. Esta discrepancia se repite en comunas como Colina-13301 y El Bosque-13105, que poseen tasas de denuncias similares, pero estimaciones SAE muy diferentes. Además, otras comunas que muestran estimaciones SAE elevadas dentro de la región y que no se reflejan en las tasas de denuncias son Cerro Navia-13103, Puente Alto-13201, Pedro Aguirre Cerda-13121, Cerrillos-13102 y Quinta Normal-13126.

Figura 25. Estimaciones SAE para VA y Tasa de denuncia para DMCS, Región Metropolitana.



Fuente: Elaboración propia.

En resumen, se han identificado relaciones atípicas entre los registros de denuncias policiales y las estimaciones comunales SAE en varias comunas. Estas incluyen San Pedro de la Paz, en la región del Biobío, Villa Alemana en Valparaíso, Los Vilos en Coquimbo, así como las comunas de Cerro Navia, Puente Alto, Pedro Aguirre Cerda, Cerrillos, Quinta Normal, Santiago, Providencia, Independencia y La Reina en la Región Metropolitana.

5.7.3. Evaluación de características comunales

Con base en los dos análisis previos, se ha identificado un conjunto de comunas que muestran resultados SAE atípicos en relación con otras comunas de la región y/o con las tasas de denuncia reportadas por las policías. Para este grupo de comunas se realizó un análisis detallado con el fin de identificar elementos que permitan explicar este comportamiento divergente al esperado.

Para este análisis, se evaluaron las tasas de denuncia y cifra oculta, así como la relación entre lugar de ocurrencia del delito y lugar de residencia de la persona informante, **en base a datos muestrales ENUSC 2018**²⁰.

- **Delitos denunciados y cifra oculta**

De los hallazgos observados en los puntos anteriores, hay dos aspectos relevantes que podrían explicar los resultados. En primer lugar, se destacan las diferencias entre las estimaciones de las encuestas de victimización y los registros de denuncias. Uno de los aspectos más significativos es que los registros policiales no reflejan la totalidad de delitos, sino únicamente la proporción que efectivamente se denuncia. La cifra oculta de delitos representa aquellos que no son denunciados y que no quedan registrados en los informes administrativos de las fuerzas policiales. Esta cifra oculta puede ser estimada mediante encuestas de victimización. Según los datos publicados de ENUSC 2018, del total de delitos de mayor connotación social reportados, el 35,1% fue denunciado ante alguna autoridad competente. De manera similar, la tasa de denuncia en la región de Coquimbo fue 40,9%, en Valparaíso fue 36,1%, en Biobío fue de 33,1% y en la Región Metropolitana fue 33,8%²¹.

Desde la perspectiva de la criminología, se señala que diversos factores influyen en la decisión de denunciar a la policía, entre los cuales se incluyen la gravedad del delito, la confianza o relación con las fuerzas policiales y las instituciones, la percepción de inseguridad, así como factores socioeconómicos, entre otros (Ariel & Bland, 2019). En este sentido, la diversidad socioeconómica y territorial, tanto entre regiones como entre comunas, también implica diferencias en la disposición de las personas para denunciar algún delito.

- **Lugar de residencia y lugar de ocurrencia del delito**

Otro punto relevante que podría explicar la diferencia entre registros de las fuerzas policiales y las estimaciones SAE de la ENUSC 2018, es el lugar en donde se registra el delito. En los registros policiales, según el estándar de producción estadística de registros de policías (CEPAL, 2022), la ubicación del registro corresponde al lugar donde ocurrió el delito, con el

²⁰ Dado que las la ENUSC 2018 no brinda estimaciones comunales para la tasa de denuncias con un nivel adecuado para hacer inferencia, los análisis son realizados con un objetivo descriptivo y exploratorio, para indagar sobre la movilidad de los delitos y la necesidad de considerar este factor en las estimaciones SAE de indicadores de victimización para un trabajo futuro.

²¹ Ver Síntesis de Resultados ENUSC 2018, disponible en: <https://www.ine.gob.cl/estadisticas/sociales/seguridad-publica-y-justicia/seguridad-ciudadana>

objetivo de generar información sobre las zonas de mayor riesgo. Por otro lado, las encuestas de victimización como ENUSC las cifras se centran en el lugar de residencia de la víctima.

Esto es especialmente relevante en ciudades conurbadas y altamente segregadas, como Santiago, y en menor medida en el Gran Concepción y el Gran Valparaíso, donde es común que las personas trabajen, estudien o se recreen en comunas diferentes a las de su residencia. Esta realidad implica una alta movilidad cotidiana dentro de la ciudad (Rodríguez, 2008). En la práctica, esto significa que, si una persona que vive en Puente Alto y es víctima de algún delito en La Florida, la denuncia policial se registrará en la Florida, pero si es encuestada, se contabilizará como una víctima de Puente Alto.

Aunque en la ENUSC 2018 los delitos se registran según la comuna de residencia de las víctimas, aún es posible identificar el lugar donde ocurrieron los hechos en los delitos específicos. Para estimar la proporción de personas que fueron víctimas de algún delito de mayor connotación social en su propia comuna, se calcula la proporción variable utilizando la pregunta: “La última vez ¿dónde le sucedió esto a usted o a algún miembro de su hogar?” para cada delito específico, en relación con el indicador de VA.

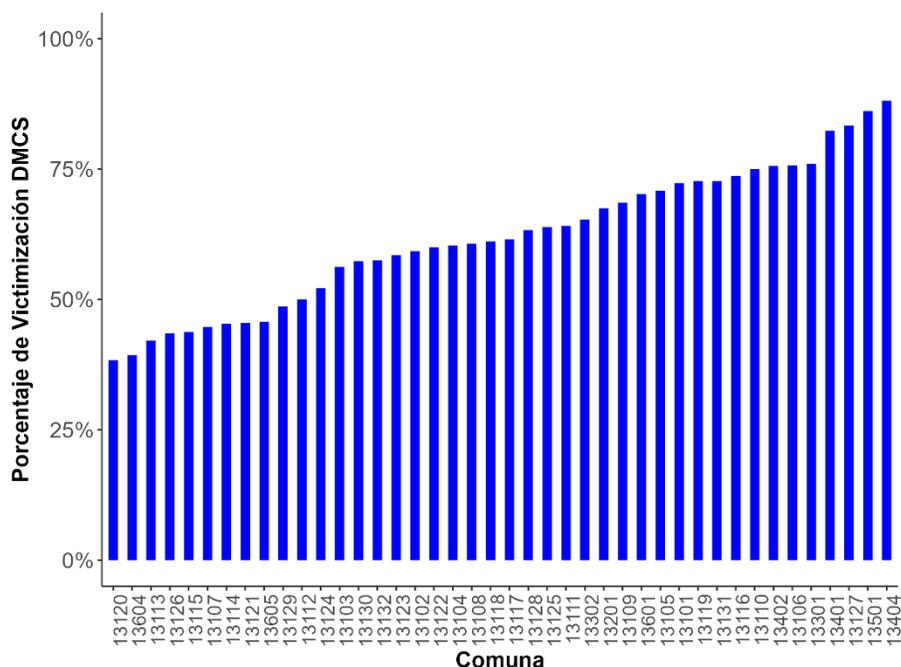
La región de Coquimbo tuvo en la ENUSC 2018 un 18,9% de victimización en VA, según los datos publicados. **Los resultados muestrales** dan cuenta que, en todas las comunas, la proporción de hogares victimizados de algún DMCS dentro de la misma comuna donde residen es cercana al 90% (**Anexo D 8.4**), lo que indica que no hay movilidad de los delitos entre las comunas de la región, por lo que el indicador no se vería afectado por el lugar donde se registra el delito, al menos a nivel muestral.

En la región de Valparaíso el indicador VA en ENUSC 2018 fue de 25,1%. En tanto, la proporción de hogares victimizados en la misma comuna en Quillota, La Calera, Los Andes y San Felipe es superior al 90%, lo que tiene relación con su ubicación más alejada de las comunas que componen el Gran Valparaíso. El resto de las comunas rondan entre el 70% y el 80%, con particular mención a Villa Alemana, en donde solo el 63,4% de los hogares fueron victimizados en su propia comuna. Esto daría cuenta de conurbaciones en donde el desplazamiento de las personas entre distintas comunas incide en esta variación.

Situación similar ocurre en la región del Biobío (23,6% de victimización según ENUSC 2018), en donde las comunas que pertenecen al Gran Concepción varían de un 69,4% en Chiguayante a un 88,1% en Concepción, mientras que Los Ángeles es de 96,4%.

Por último, en el caso de la RM, con una proporción de victimización en hogares de 30,1% en ENUSC 2018, la variabilidad de este indicador en la muestra es mucho mayor que en el resto de las regiones. Como se aprecia en el **Figura 26**, en las comunas de Ñuñoa-13120 y Padre Hurtado-13604, este indicador no supera el 40% a nivel muestral, mientras que, en comunas menos conectadas al Gran Santiago como Paine-13404 o Melipilla-13501, las estimaciones en el indicador son del 90%. Que existan comunas con una proporción tan baja de hogares victimizados dentro de la propia comuna, no solo da indicios de la gran movilidad que hay dentro de la comuna, sino que también de una mayor complejidad para relacionar registros de policías y encuestas de victimización.

Figura 26. Porcentaje de hogares en la muestra que fueron víctima de algún DMCS dentro de su propia comuna, Región Metropolitana.

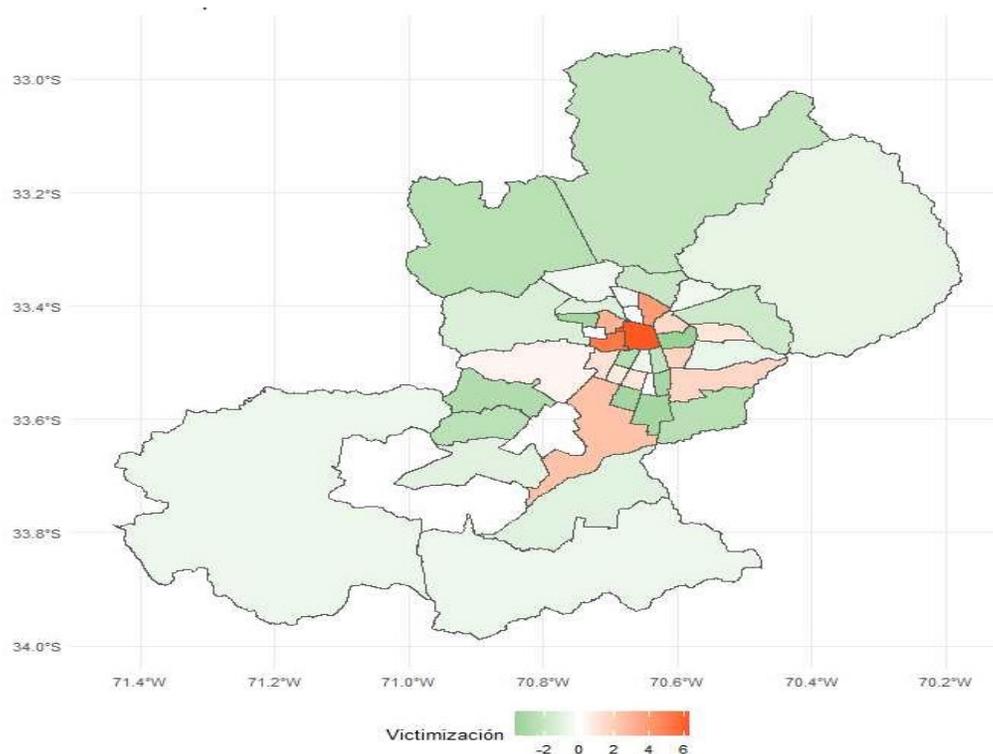


Fuente: Elaboración propia.

Una limitación adicional de este análisis es que solo da cuenta de los delitos que persisten dentro de la comuna, pero no identifica donde se mueven los delitos. Para complementarlo, es posible identificar en los delitos específicos, en qué comuna o región el hogar fue victimizado. En el caso de Robo con violencia o intimidación, el mapa de la RM (**Figura 27**) muestra en verde las comunas en donde los hogares son victimizados en otras comunas, mientras que las zonas en rojo son comunas que concentran victimización a hogares de otras comunas. Este desplazamiento de los robos con violencia o intimidación de las comunas más periféricas hacia el centro y centro sur de la RM responde a las características de alta

movilidad de las personas dentro de la región, cuestión relevante para considerar en el uso de registros de policías para explicar las tasas de victimización de ENUSC 2018.

Figura 27. Diferencia entre lugar de residencia y lugar de ocurrencia de RVI, en la Región Metropolitana.



Fuente: Elaboración propia.

5.8. Reflexiones sobre modelamiento de la tasa de victimización de hogares por robo desde vehículo

Durante el desarrollo de esta investigación, se examinó un indicador adicional, el robo desde vehículo, además de los ya mencionados. Este delito se define en la sección H de la ENUSC 2018 como la apropiación de algún objeto dejado dentro del vehículo o partes de él, sin el uso de violencia o intimidación en contra del usuario (Instituto Nacional de Estadísticas - Chile, 2018), por ejemplo, el robo de la radio del auto, el espejo, o algún bolso que se encuentre dentro del vehículo. Es importante destacar que esta sección abarca solo los

hogares que reportan tener vehículo o vehículos, lo cual reduce el tamaño de la muestra para el análisis.

Al tratarse de un indicador evidentemente más detallado, ya que específicamente se refiere a delitos ocurridos en el vehículo, y que además la violencia o intimidación no es parte de su definición, se reducen considerablemente las posibles covariables que podrían explicar este delito. Para las covariables disponibles en esta investigación, que en su mayoría se refieren a factores de intimidación y/o violencia al momento del robo, no se encontraron correlaciones altas ($r < |0,6|$) con el robo desde vehículo, misma situación con respecto a las variables sociodemográficas, con coeficientes de correlación, $r < |0,4|$. Esto posiblemente se deba a que su naturaleza guarda una mayor relación con la definición del indicador de victimización por robo por sorpresa y por robo con violencia o intimidación.

Este hecho representó el primer obstáculo para modelar el robo desde vehículo. Para abordar esta limitación, se incorporó como variable auxiliar la cantidad de permisos de circulación por comuna (Instituto Nacional de Estadísticas, 2018), que posteriormente se incluyó en la base como una tasa por cada 100.000 habitantes.

Pese a la inclusión de esta nueva covariable, no se estableció una correlación coherente con el indicador, más aún, al agregarla al modelo, su coeficiente no resultó significativo, aunque se aplicaron múltiples transformaciones. No obstante, la limitación de las covariables y con el objetivo de avanzar en el análisis, se probaron numerosos modelos con diferentes combinaciones, transformaciones e interacciones de variables, hasta que finalmente se seleccionó un modelo que cumpliera con los supuestos para la inferencia presentados en la **Subsección 5.6**, utilizando para ello la transformación arcoseno, como se muestra en la **Tabla 8**.

Tabla 8. Covariables en modelos SAE para robo desde vehículo.

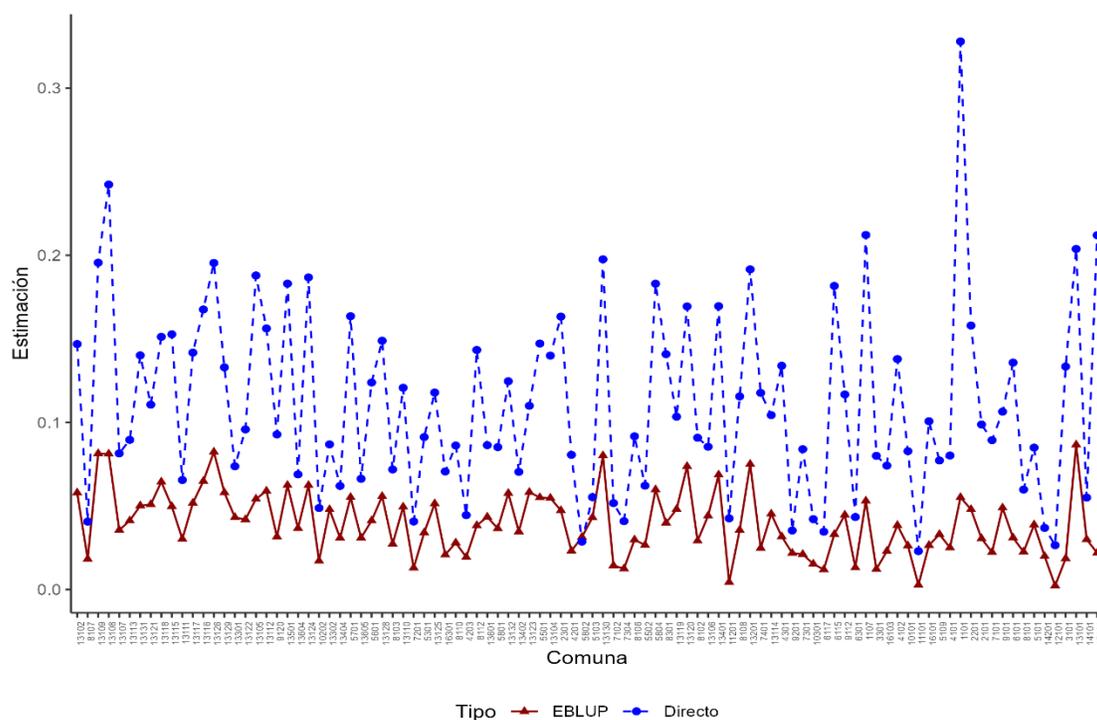
	Coefficientes	Error estándar	Valor t	Valor p
(Intercepto)	0,3087	0,0584	5,2837	0,0000
Tasa de robo	0,0000	0,0000	3,5088	0,0004
Proporción población adulta que tiene a lo sumo educación primaria incompleta	-0,5318	0,3426	-1,5521	0,1206
Tasa de hurtos	-0,0000	0,0000	-2,5142	0,0119
Dummy Reg. Tarapacá	0,2281	0,0662	3,4438	0,0005
Dummy Reg. Antofagasta	0,0669	0,0521	1,2836	0,1992
Dummy Reg. Atacama	0,0294	0,0610	0,4825	0,6294
Dummy Reg. Coquimbo	-0,0003	0,0498	-0,0063	0,9949
Dummy Reg. Valparaíso	0,0111	0,0445	0,2500	0,8025

	Coefficientes	Error estándar	Valor t	Valor p
Dummy Reg. O'Higgins	0,0356	0,0515	0,6920	0,4889
Dummy Reg. Maule	0,0005	0,0505	0,0103	0,9917
Dummy Reg. Biobío	0,0021	0,0455	0,0475	0,9621
Dummy Reg. Araucanía	-0,0086	0,0543	-0,1591	0,8736
Dummy Reg. Los Lagos	-0,0050	0,0576	-0,0883	0,9296
Dummy Reg. Aysén	-0,0501	0,0588	-0,8522	0,3940
Dummy Reg. Magallanes	-0,0913	0,0714	-1,2782	0,2011
Dummy Reg. Metropolitana	0,0188	0,0423	0,4459	0,6556
Dummy Reg. Los Ríos	-0,0511	0,0591	-0,8640	0,3875
Dummy Reg. Arica y Parinacota	0,1597	0,0703	2,2701	0,0231
AIC	-226,6103			
BIC	-174,9110			
$R^2_{Hidiroglou}$	0,6909			

Fuente: Elaboración propia.

Luego de ajustar y evaluar los supuestos del modelo, se obtuvieron las estimaciones comunales (ver **Tabla 8**), sin embargo, como se observa en la **Figura 28**, las estimaciones del modelo EBLUP no se asemejan a las estimaciones directas cuando el tamaño de la muestra es grande, lo que nos lleva a sospechar de un mal ajuste, consecuencia presumiblemente, de la falta de covariables válidas para la estimación del robo desde vehículo.

Figura 28. Estimaciones comunales para robo desde vehículo.



Fuente: Elaboración propia.

De esta forma, si el modelo está mal especificado (debido presumiblemente a la falta de datos auxiliares que se relacionen fuertemente con el indicador de interés, en este caso, tasa de victimización por robo desde vehículo), las estimaciones obtenidas vía EBLUP pueden resultar no confiables y/o no útiles para su uso en la política pública. Lo que releva la necesidad de recopilar de forma exhaustiva, información auxiliar válida y de calidad que se relacione con el fenómeno que se desea medir. Para este indicador en particular, se considera que pudo haber sido útil contar con información de los tiempos de espera en los estacionamientos, o el tiempo en que cambian los semáforos de rojo a verde por comuna.

En relación con los delitos denunciados, es importante considerar que no siempre se producen en la misma comuna de residencia de la persona afectada, por ejemplo, delitos que ocurren durante el día laboral o en salidas ocasionales, lo que implica algunas veces estar fuera de la comuna de residencia. Es crucial reconocer que la recopilación de datos auxiliares debe contemplar este factor como influyente al estimar cualquier indicador de este tipo.

6. Conclusiones y trabajo futuro

Este trabajo de investigación ha enfocado sus esfuerzos en la implementación de una metodología que permita obtener estimaciones EBLUP, para las tasas de victimización por robo con violencia o intimidación, robo por sorpresa y victimización agregada de delitos de mayor connotación social, en las 102 comunas de las 16 regiones de Chile pertenecientes a la muestra de la ENUSC 2018. De esta forma, se releva que los métodos de estimación de áreas pequeñas basados en modelos, y concretamente para este trabajo, el EBLUP basado en la metodología de Fay-Herriot, son herramientas eficaces para estimar y visualizar en mapas variables de interés criminológico a nivel de área pequeña, especialmente cuando los tamaños de muestra no son lo suficientemente grandes como para permitir estimaciones directas confiables. Para garantizar la consistencia de los resultados, es imperativo contar con datos auxiliares fiables: exactos, completos, consistentes, creíbles, eficientes, comprensibles, entre otros aspectos que se encuentra definidos por ISO/IEC 25012 (International Organization for Standardization, 2008); y verificar que las estimaciones del EBLUP para dichas variables son más fiables que las directas.

Aunque el objetivo principal de SAE es producir estimaciones confiables, los modelos analizados brindan un conjunto significativo de información para avanzar en la comprensión de la prevalencia de estos delitos. En general, para los tres indicadores analizados, se ha encontrado que las variables derivadas de los registros policiales de carabineros presentan una mayor correlación y, consecuentemente, resultan ser los mejores predictores en los modelos evaluados. Así, bajo el modelo especificado, la tasa de robo con violencia o intimidación resultó ser el mejor predictor para el indicador de robo con violencia o intimidación, covariable que está íntimamente ligada a la variable objetivo, dando cuenta de una complementación entre registro policial y encuesta. Respecto a las variables sociodemográficas, provenientes del CPV 2017, en el caso del robo por sorpresa, se encuentra una relación directa con la proporción de personas con estudios secundarios o terciarios, donde probablemente se encuentren personas con trabajo y pertenecientes a estratos socioeconómicos medios o superiores, que serían propensos a este tipo de delito (Olavarría Gambi, 2006). En ese mismo sentido, el modelo finalmente ajustado para el indicador de robo con violencia o intimidación muestra que los robos o robos violentos serían más prevalentes en la población que se encuentra entre los 14 y 35 años.

Es importante señalar que, durante el año de publicación de este trabajo, la SPD y el INE, han puesto a disposición de las personas usuarias, los resultados de la nueva versión de la encuesta, ENUSC 2023. Esta nueva versión es el resultado de un proceso de actualización que comenzó en 2019, desde la necesidad de capturar las nuevas características de la victimización y percepción de inseguridad, ampliando así el repertorio de delitos e indicadores y mejorando el diseño estadístico²². También durante este año es realizado el Censo de Población y Vivienda 2024, por tanto, se genera el desafío futuro de incorporarlo como fuente de información auxiliar, ya que constituye el principal y más completo insumo de información sociodemográfica de Chile.

En relación con los criterios de evaluación y validación de las estimaciones SAE discutidos en este trabajo, destacamos que constituyen una etapa crucial para asegurar la precisión y utilidad de los indicadores comunales derivados de la encuesta ENUSC. Esta etapa forma parte integral del proceso de producción estadística bajo el GSBPM (Generic Statistical Business Process Model) adoptado por el INE para todas sus operaciones estadísticas. Si bien la comparación con las estimaciones directas a nivel regional ofrece un punto de partida inicial, uno de los grandes desafíos a futuro es el desarrollo de una metodología adecuada para llevar a cabo estas evaluaciones. Esta metodología debe incluir la selección cuidadosa de indicadores y el uso de registros administrativos pertinentes, garantizando que los criterios de comparación sean consistentes y reflejen fielmente la realidad comunal.

Los modelos SAE descritos en las secciones anteriores se obtienen después de probar múltiples alternativas para el tratamiento de los datos (consulte el Anexo C 8.3 para ver la lista de opciones probadas con los modelos SAE y sus resultados). Sin embargo, se considera que todavía hay aspectos que se pueden mejorar en los modelos. Precisamente, este es el marco de desarrollo de este trabajo de investigación, considerado un estudio experimental, que requiere nutrirse y perfeccionarse con información auxiliar atinente y de calidad, y con el análisis de otras alternativas de modelamiento como los métodos bayesianos. Se reitera entonces, que los modelos ajustados en este trabajo están limitados por la disponibilidad de información auxiliar confiable, por ejemplo, algunas variables exploradas en estudios anteriores (desigualdad de ingresos, inversión en salud/educación, medidas de seguridad, características de barrio) no pudieron ser probadas en este trabajo. Por lo tanto, el acceso a este tipo de fuentes, así como imágenes satelitales y la extensión del potencial del EBLUP a

²² Para mayor información sobre las características de la ENUSC 2023, revisar Informe Metodológico e Informe de Calidad, disponibles en la sección de Metodologías en <https://www.ine.cl/estadisticas/sociales/seguridad-publica-y-justicia/seguridad-ciudadana>

través de la incorporación de efectos de área aleatorios correlacionados espacialmente (SEBLUP), puede contribuir en mejoras a la especificación de los modelos de áreas pequeñas y sus resultados.

Por último, este trabajo busca exponer y reflejar que el acceso, por parte del Instituto Nacional de Estadística, a registros administrativos u otras fuentes permite que se exploren nuevas metodologías para la generación de indicadores necesarios para la formulación de políticas públicas, de forma eficiente y buscando resguardar la calidad de las estadísticas oficiales del país. Es un desafío entonces acceder a más información, pero más aún es el desafío de discriminar en un mar de datos aquellos que permitan una mejor caracterización de nuestra población, sin perder de vista la consistencia, oportunidad y validez de la información estadística que se genera.

7. Referencias

- Alvira, F., & Rubio, M. (1982). *Victimización e inseguridad: La perspectiva de las encuestas de victimización en España*.
- Ariel, B., & Bland, M. (Agosto de 2019). Is Crime Rising or Falling? A Comparison of Police-Recorded Crime and Victimization Surveys. *Sociology of Crime, Law and Deviance*, 24, 7-31. Obtenido de <https://doi.org/10.1108/S1521-613620190000024004>
- Armas, C., & Herrera, J. (2018). *¿El lugar importa en la victimización? El caso de Lima Metropolitana. Documento de trabajo N°120*. Asociación peruana de economía, Lima.
- Beaumont, J.-F., & Bocci, C. (Mayo de 2016). Small Area Estimation in the Labour Force Survey. (S. Canada, Ed.)
- Bell, W. R. (2008). Examining sensitivity of small area inferences to uncertainty about sampling error variances. . Joint Statistical Meetings-Section on Survey Research Methods.
- Boonstra, H. (2012). hbsae: Hierarchical Bayesian Small Area Estimation. 1-21. Obtenido de <https://cran.r-project.org/web/packages/hbsae/hbsae.pdf>
- Breiman, L. (2001). Random Forests. 1-33.
- Buil Gil, D. (2019). *Small Area Estimation in Criminological Research: Theory, methods and applications. Thesis for the degree of PhD. in Criminology in the Faculty of Humanities*. University of Manchester, Department of Criminology, Manchester.

- Cadena, P., & Letelier, L. (2018). Factores determinantes de los Delitos de Mayor Connotación Social en la Región Metropolitana. Análisis en base a un modelo de regresión logística. *Política criminal*, 13(26). Obtenido de <https://dx.doi.org/10.4067/S0718-33992018000201170>
- Casas-Cordero Valencia, C., Encina, J., & Lahiri, P. (2016). *Poverty mapping for the Chilean Comunas. Analysis of Poverty Data by Small Area Estimation*. . Wiley Online Library.
- Centro de Formación Estudio Criminal. (27 de 11 de 2018). *CEFEC*. Obtenido de <https://www.estudiocriminal.eu/blog/que-es-la-cifra-negra/>
- CEPAL. (2022). *Marco de referencia para los sistemas estadísticos de seguridad y justicia penal en América Latina y el Caribe*. Obtenido de <https://www.cepal.org/es/publicaciones/47928-marco-referencia-sistemas-estadisticos-seguridad-justicia-penal-america-latina>
- Chatrchi, G. (2019). *Methodology Guide Small Area Estimation for Visitor Travel Survey* . Social Statistics Methods Division.
- Chen, S., & Lahiri, P. (2002). A Weighted Jackknife MSPE Estimator in Small-Area Estimation. *In Proceeding of the Section on Survey Research Methods*, 473-477.
- Cook, D. (Febrero de 1977). Detection of Influential Observation in Linear Regression. *Technometrics*, 1(19), 15-18. Obtenido de <http://www.jstor.org/stable/1268249> .
- Datta, D., Rao, J., & Smith, D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, 92(1), 183-196.
- Datta, G., & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10(2), 613–627.
- Dick, P. (1995). Modeling Net Under-coverage in the 1991 Canadian Census. *Survey Methodology*, 21, 44-55.
- Efron, & Morris. (1972). Limiting the Risk of Bayes and Empirical Bayes Estimators- Part II: The Empirical Bayes Case. *Journal of the American Statistical Association*(67), 130-139.
- Fay, R., & Diallo, M. (2015). sae2: Small Area Estimation: Time-Series Models.

- Fay, R., & Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74(366), 269-277.
- Fay, R., Planty, M., & Diallo, M. (2013). Small Area Estimates from the National Crime Victimization Survey. *Section on Survey Research Methods*, 1-14.
- Fúquene, J., Cristancho, C., Ospina, M., & Morales, D. (2018). Prevalence of international migration: an alternative for small area estimation. . 1-14. Obtenido de arXiv:1905.00353v1
- Gutiérrez, A., Mancero, X., Fuentes, A., López, F., & Molina, F. (2020). Coeficientes de variación. En *Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional* (Vol. 101, págs. 16-18). America Latina. Obtenido de <https://repositorio.cepal.org/server/api/core/bitstreams/4e2e55fc-3429-4619-8b6a-0bd5357b23b5/content>
- Hadam, S., Würz, N., & Kreutzmann, A. (2020). Estimating Regional Unemployment with Mobile Network Data for Functional Urban Areas in Germany. *Refubium - Freie Universität Berlin Repository*, 1-28. doi:doi:10.17169/refubium-26791
- Hájek, J. (1971). *Comment on \An Essay on the Logical Foundations of Survey Sampling, Part One In: The Foundations of Survey Sampling*. Godambe, V.P. y Sprott, D.A. eds.
- Haslett, S. J., Isidro, M. C., & Jones, G. (2010a). Comparison of survey regression techniques in the context of small area estimation of poverty. *Survey Methodology*. 36(2), 157-170.
- Henderson, C. (s.f.). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2), 423-447.
- Hidiroglou, M., Beaumont, J.-F., & Yung, W. (2019). Development of a small area estimation system at Statistics Canada. (C. N.-O.-X. Statistics Canada, Ed.) 45(1), 101-126.
- INEGI. (Abr-Jun de 2017). Características de la victimización y victimización múltiple de la población en México, 2010-2015. (A. F. Barrios, Ed.) *En números, Documento de análisis y estadísticas*, 1(9), 1-84.

- INEGI. (Abr-Jun de 2017). Características de la victimización y victimización múltiple de la población en México, 2010-2015. (A. F. Barrios, Ed.) *En números, Documento de análisis y estadísticas*, 1(9), 1-84.
- Instituto Nacional de Estadísticas - Chile. (2018). *ine.gob.cl*. Obtenido de https://www.ine.gob.cl/docs/default-source/seguridad-ciudadana/formularios/2018/cuestionario-xv-enusc-2018.pdf?sfvrsn=fa5d9acd_2
- Instituto Nacional de Estadísticas - Chile. (2019). *Manual de Usuario Base de datos 2008-2018*. Manual. Obtenido de https://www.ine.gob.cl/docs/default-source/seguridad-ciudadana/bbdd/2018/manual-de-usuario-de-base-de-datos-interannual---enusc-2008-2018.pdf?sfvrsn=41a813c5_4
- Instituto Nacional de Estadísticas - Chile. (Febrero de 2020). *ine.gob.cl*. Obtenido de <https://www.ine.gob.cl/docs/default-source/institucionalidad/buenas-pr%C3%A1cticas/clasificaciones-y-estandares/est%C3%A1ndar-evaluaci%C3%B3n-de-calidad-de-estimaciones-publicaci%C3%B3n-27022020.pdf>
- Instituto Nacional de Estadísticas. (2018). *ine.gob.cl*. Obtenido de <https://www.ine.gob.cl/estadisticas/economia/transporte-y-comunicaciones/permiso-de-circulacion>
- International Organization for Standardization. (2008). *ISO 25000: Calidad de software y datos*. Obtenido de <https://iso25000.com/index.php/normas-iso-25000/iso-25012>
- ISUC. (2013). *Estudio determinantes de la percepción y preocupación social ante el delito*. Informe de resultados, Pontificia Universidad Católica de Chile, Instituto de Sociología, Santiago de Chile.
- ISUC. (2013). *Estudio determinantes de la percepción y preocupación social ante el delito*. Instituto de Sociología, Pontificia Universidad Católica de Chile, Santiago de Chile.
- Jiang, J., & Lahiri, P. (Junio de 2006). Mixed model prediction and small area estimation. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 15(1), 1-96.
- Jiang, J., Lahiri, P., Wan, S.-M., & Wu., C.-H. (2001). Jackknifing in the. *In Proceedings of the Seminar on Funding Opportunity in Survey Research Council of Professional Associations on Federal Statistics*(36), 75-97.

- Kreutzmann, A., Pannier, S., Rojas-Perilla, N., Schmid, T., Templ, M., & Tzavidis, N. (2019). The {R} Package {emdi} for Estimating and Mapping Regionally Disaggregated Indicators. *Journal of Statistical Software*, 91(7), 1-33. doi:10.18637/jss.v091.i07
- Li, H., & Lahiri, P. (2010). Adjusted maximum method in the small area estimation problem. *Journal of Multivariate Analysis*, 101, 882-892.
- Molina, I. (2019). Desagregación de datos en encuestas de hogares: Metodología de estimación en áreas pequeñas. (97), 1-97. Obtenido de https://repositorio.cepal.org/bitstream/handle/11362/44214/5/S1900419_es.pdf
- Molina, I., & Marhuenda, Y. (Junio de 2015). sae: An R Package for Small Area Estimation. *The R Journal*, 7(1), 81-98.
- Molina, I., & Marhuenda, Y. (2018). sae: Small Area Estimation. 1-45. Obtenido de <https://cran.r-project.org/web/packages/sae/sae.pdf>
- Neves, A., Silva, N., & Correa, S. (2013). Small domain estimation for the Brazilian service sector survey. *Instituto Interamericano de Estadística.*, 65(185), 3-37.
- Olavarría Gambi, M. (2006). El crimen en Chile: Una mirada desde las víctimas. Santiago: CESC.
- Permatasari, N., & Ubaidillah, A. (Diciembre de 2021). msae: An R Package of Multivariate Fay-Herriot Models for Small Area Estimation. *The R Journal*, 13(2), 111-122.
- Prasad, N. G., & Rao, J. N. (s.f.). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85(409), 163-171.
- R Core Team. (2019). A language and environment for statistical computing. *R Foundation for Statistical Computing*.
- Rao, J. N. (2014). *Small-area estimation*. Wiley StatsRef: Statistics Reference Online.
- Rao, J., & Molina, I. (2015). *Small area estimation (2nd ed.)*. Wiley Series in Survey Methodology.
- Rivest, L.-P., & Belmonte, E. (2000). A Conditional Mean Squared Error of Small Area Estimators. *Survey Methodology*, 26(1), 67-78.

- Rodríguez, J. (Diciembre de 2008). Movilidad cotidiana, desigualdad social y segregación residencial en cuatro metrópolis de América Latina. *Eure*, 34(103), 49-71. Obtenido de <https://dx.doi.org/10.4067/S0250-71612008000300003>
- Särndal, C., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. Springer-Verlag.
- Schmid, T., Bruckschen, F., Salvati, N., & Zbiranski, T. (2017). Constructing Sociodemographic Indicators for National Statistical Institutes Using Mobile Phone Data: Estimating Literacy Rates in Senegal. *Journal of the Royal Statistical Society Series A*, 180(4), 1163-1190. doi:doi:10.1111/rssa.12305
- Smith, M., Lix, L. M., Azimae, M., Enns, J. E., Orr, J., Hong, S., & Roos, L. L. (Septiembre de 2017). Assessing the quality of administrative data for research: a framework from the Manitoba Centre for Health Policy. *Journal of the American Medical Informatics Association*, 25(3), 224-229. Obtenido de <https://doi.org/10.1093/jamia/ocx078>
- Sugawasa, S., & Kubokawa, T. (2017). Transforming Response Values in Small Area Prediction. *Computational Statistics and Data Analysis*, 114, 47-60. doi:doi:10.1016/j.csda.2017.03.017
- Tzavidis, N., Zhang, L., Luna, A., Schmid, T., & Rojas-Perilla, N. (2018). From start to finish: a framework for the production. *J. R. Statist. Soc. A*(181), 1-33.
- Valiant, R. (1987). Generalized variance functions in stratified two-stage sampling. *Journal of the American Statistical Association*(82), 499-508.
- Wolter, K. M. (2007). *Statistics for social and behavioral sciences. Introduction to variance estimation (2nd ed.)*. Springer Science + Business Media.

8. ANEXOS

8.1. Anexo A: Preparación de datos auxiliares

Para la generación de covariables provenientes del CPV 2017, se realizaron los siguientes tratamientos preliminares en la base de datos:

- Solo se consideran viviendas particulares con moradores presentes.
- Se excluyen aquellas personas que, al momento de ser censadas no se encontraban en la comuna donde residen.
- Dado que se trabaja con información completa, se precisa eliminar aquellos registros con información faltante para el set de variables con el que se construirán las covariables, a saber: material de paredes (P03A), materialidad del techo (P03B), materialidad del piso (P03C), número de dormitorios (P04) y acceso al agua (P05), sexo (P08), edad (P09), nivel de educación (P15), término de nivel de educación (P15A) y situación laboral (P17).
- Para la construcción del índice de hacinamiento, se decidió imputar el número de piezas a 1 cuando este valor era cero o NA.
- Para el caso de los indicadores RVI y RPS solo se consideran personas del área urbana, esto buscando guardar consistencia con los datos de la ENUSC, que solo se levanta en el área urbana. En el caso de caso del indicador de VA, se consideró toda la población de la comuna, esto buscando guardar consistencia con la otra información auxiliar, proveniente de los registros de las policías, que no hacen distinción entre el área urbana o rural de la comuna. Esta precisión sobre la cobertura debe analizarse con mayor profundidad para trabajos futuros.

Tabla A. 9. Covariables para construcción de modelos EBLUP en ENUSC 2018.

Nombre	Descripción	Fuente
U_p_pob_menores15	Proporción de la población menor de 15 años	CPV 2017
U_p_pob_15_24	Proporción de la población entre 15 y 24 años	CPV 2017
U_p_pob_25_49	Proporción de la población entre 25y 49 años	CPV 2017
U_p_pob_50_64	Proporción de la población entre 50 y 64 años	CPV 2017
U_p_pob_65ymas	Proporción de la población de 65 y más	CPV 2017
U_p_pob_14_35	Proporción de la población entre 14 y 35 años	CPV 2017
U_p_hombres	Proporción de la población de hombres	CPV 2017
U_p_mujeres	Proporción de la población de mujeres	CPV 2017
U_p_ocupados	Proporción de la población ocupada	CPV 2017
U_p_desocupados	Proporción de la población desocupada	CPV 2017
U_p_inactivos	Proporción de la población inactiva	CPV 2017

Nombre	Descripción	Fuente
U_p_pob_menores18	Proporción de la población menor de 18 años	CPV 2017
U_p_educ_o	Proporción de la población sin nivel de escolaridad	CPV 2017
U_p_educ_primaria	Proporción de la población con nivel de escolaridad primaria	CPV 2017
U_p_educ_secundaria	Proporción de la población con nivel de escolaridad secundaria	CPV 2017
U_p_educ_terciaria	Proporción de la población con nivel de escolaridad terciaria	CPV 2017
U_p_pob_25a65_educ_terciaria	Proporción de la población entre 25 y 65 años con nivel de educación terciaria	CPV 2017
U_p_educ_sec_ter	Proporción de la población adulta con al menos educación secundaria completa	CPV 2017
U_p_pob_educ_sup	Proporción de la población adulta que alcanza la educación superior (terciaria)	CPV 2017
U_p_pob_educ_hprim	Proporción de la población adulta que tiene a lo sumo educación primaria incompleta	CPV 2017
U_p_N_agua	Proporción de viviendas con acceso a la red pública de alcantarillado	CPV 2017
U_p_N_viv_inade	Proporción de viviendas en condiciones inadecuadas	CPV 2017
U_p_hacina	Proporción de viviendas en condiciones de hacinamiento. Se considera hacinamiento cuando el cociente entre el número de piezas y el número de personas es superior a tres	CPV 2017
U_ts_viole_asal_sex	Tasa de violaciones/robos sexuales por cada 100.000 habitantes	SPD
U_ts_robo	Tasa de robos por cada 100.000 habitantes	SPD
U_ts_hurtos	Tasa de hurtos por cada 100.000 habitantes	SPD
U_ts_robo_vehi_motor	Tasa de robo de vehículos por cada 100.000 habitantes	SPD
U_ts_robo_viole_intim	Tasa de robo con violencia o intimidación por cada 100.000 habitantes	SPD
U_ts_robo_viole_hog	Tasa de robos con violencia en el hogar por cada 100.000 habitantes	SPD
U_ts_asal_simple	Tasa de robos simples por cada 100.000 habitantes	SPD
U_ts_homi	Tasa de homicidios por cada 100.000 habitantes	SPD
U_ts_robo_loc_resi_priv	Tasa de robos en locales residenciales privados por cada 100.000 habitantes	SPD
U_ts_dist_prob_drog	Tasa de denuncias por distribución de drogas por cada 100.000 habitantes	SPD
U_ts_rps	Tasa de robo por sorpresa por cada 100.000 habitantes	SPD
U_ts_rdv	Tasa de robo desde vehículo por cada 100.000 habitantes	SPD

Fuente: Elaboración propia.

8.2. Anexo B: Áreas o dominios de interés

Tabla A. 10. Áreas o dominios de interés.

CUT	Comuna	Región	Tipo
1101	Iquique	Región de Tarapacá	Capital Regional
1107	Alto Hospicio	Región de Tarapacá	
2101	Antofagasta	Región de Antofagasta	Capital Regional
2201	Calama	Región de Antofagasta	Capital Provincial
2301	Tocopilla	Región de Antofagasta	Capital Provincial
3101	Copiapó	Región de Atacama	Capital Regional
3301	Vallenar	Región de Atacama	Capital Provincial
4101	La Serena	Región de Coquimbo	Capital Regional
4102	Coquimbo	Región de Coquimbo	Capital Provincial
4201	Illapel	Región de Coquimbo	Capital Provincial
4203	Los Vilos	Región de Coquimbo	
4301	Ovalle	Región de Coquimbo	Capital Provincial
5101	Valparaíso	Región de Valparaíso	Capital Regional
5103	Concón	Región de Valparaíso	
5109	Viña del Mar	Región de Valparaíso	Más de 200.000 habitantes
5301	Los Andes	Región de Valparaíso	Capital Provincial
5501	Quillota	Región de Valparaíso	Capital Provincial
5502	Calera	Región de Valparaíso	
5601	San Antonio	Región de Valparaíso	Capital Provincial
5701	San Felipe	Región de Valparaíso	Capital Provincial
5801	Quilpué	Región de Valparaíso	Capital Provincial
5802	Limache	Región de Valparaíso	
5804	Villa Alemana	Región de Valparaíso	
6101	Rancagua	Región de O'Higgins	Capital Regional
6115	Rengo	Región de O'Higgins	
6117	San Vicente	Región de O'Higgins	
6301	San Fernando	Región de O'Higgins	Capital Provincial
7101	Talca	Región del Maule	Capital Regional
7102	Constitución	Región del Maule	
7201	Cauquenes	Región del Maule	Capital Provincial
7301	Curicó	Región del Maule	Capital Provincial
7304	Molina	Región del Maule	
7401	Linares	Región del Maule	Capital Provincial
8101	Concepción	Región del Biobío	Capital Regional
8102	Coronel	Región del Biobío	
8103	Chiguayante	Región del Biobío	
8106	Lota	Región del Biobío	
8107	Penco	Región del Biobío	

CUT	Comuna	Región	Tipo
8108	San Pedro de la Paz	Región del Biobío	
8110	Talcahuano	Región del Biobío	
8111	Tomé	Región del Biobío	
8112	Hualpén	Región del Biobío	
8301	Los Ángeles	Región del Biobío	Capital Provincial
9101	Temuco	Región de la Araucanía	Capital Regional
9112	Padre Las Casas	Región de la Araucanía	
9120	Villarrica	Región de la Araucanía	
9201	Angol	Región de la Araucanía	Capital Provincial
10101	Puerto Montt	Región de Los Lagos	Capital Regional
10201	Castro	Región de Los Lagos	Capital Provincial
10202	Ancud	Región de Los Lagos	
10301	Osorno	Región de Los Lagos	Capital Provincial
11101	Coyhaique	Región de Aysén	Capital Regional
11201	Aysén	Región de Aysén	
12101	Punta Arenas	Región de Magallanes y de la Antártica Chilena	Capital Regional
13101	Santiago	Región Metropolitana	Capital Regional
13102	Cerrillos	Región Metropolitana	Alta Victimización
13103	Cerro Navia	Región Metropolitana	
13104	Conchalí	Región Metropolitana	
13105	El Bosque	Región Metropolitana	Alta Victimización
13106	Estación Central	Región Metropolitana	Alta Victimización
13107	Huechuraba	Región Metropolitana	Alta Victimización
13108	Independencia	Región Metropolitana	Alta Victimización
13109	La Cisterna	Región Metropolitana	Alta Victimización
13110	La Florida	Región Metropolitana	Más de 200.000 habitantes
13111	La Granja	Región Metropolitana	Alta Victimización
13112	La Pintana	Región Metropolitana	
13113	La Reina	Región Metropolitana	
13114	Las Condes	Región Metropolitana	Más de 200.000 habitantes
13115	Lo Barnechea	Región Metropolitana	
13116	Lo Espejo	Región Metropolitana	Alta Victimización
13117	Lo Prado	Región Metropolitana	Alta Victimización
13118	Macul	Región Metropolitana	Alta Victimización
13119	Maipú	Región Metropolitana	Más de 200.000 habitantes
13120	Ñuñoa	Región Metropolitana	Más de 200.000 habitantes
13121	Pedro Aguirre Cerda	Región Metropolitana	Alta Victimización
13122	Peñalolén	Región Metropolitana	Más de 200.000 habitantes
13123	Providencia	Región Metropolitana	
13124	Pudahuel	Región Metropolitana	Más de 200.000 habitantes
13125	Quilicura	Región Metropolitana	Más de 200.000 habitantes

CUT	Comuna	Región	Tipo
13126	Quinta Normal	Región Metropolitana	Alta Victimización
13127	Recoleta	Región Metropolitana	Alta Victimización
13128	Renca	Región Metropolitana	Alta Victimización
13129	San Joaquín	Región Metropolitana	
13130	San Miguel	Región Metropolitana	Alta Victimización
13131	San Ramón	Región Metropolitana	Alta Victimización
13132	Vitacura	Región Metropolitana	
13201	Puente Alto	Región Metropolitana	Capital Provincial
13301	Colina	Región Metropolitana	Capital Provincial
13302	Lampa	Región Metropolitana	Alta Victimización
13401	San Bernardo	Región Metropolitana	Capital Provincial
13402	Buín	Región Metropolitana	
13404	Paine	Región Metropolitana	
13501	Melipilla	Región Metropolitana	Capital Provincial
13601	Talagante	Región Metropolitana	Capital Provincial
13604	Padre Hurtado	Región Metropolitana	
13605	Peñaflor	Región Metropolitana	
14101	Valdivia	Región de Los Ríos	Capital Regional
14201	La Unión	Región de Los Ríos	Capital Provincial
15101	Arica	Región de Arica y Parinacota	Capital Regional
16101	Chillán	Región del Ñuble	Capital Regional
16103	Chillán Viejo	Región del Ñuble	Capital Provincial
16301	San Carlos	Región del Ñuble	Capital Provincial

Fuente: Elaboración propia.

8.3. Anexo C: Opciones probadas con los modelos SAE

1. Transformaciones en las covariables

- Idea: Para verificar si transformaciones en las variables mejoran los resultados o no.
- Descripción breve: Utilizar la transformación raíz cuadrada en las covariables del tipo proporción y la transformación logaritmo en las covariables del tipo tasas.
- Resultado: No hubo mejoras en los resultados.
- ¿Se implementó? NO

2. Utilizar variables indicadoras

- Idea: Para controlar valores atípicos o influyentes.

- Descripción breve: Transformar covariables en variables indicadoras para controlar valores atípicos o influyentes.
- Resultado: Brindó resultados satisfactorios.
- ¿Se implementó? SÍ

3. Cambiar la cobertura de las covariables

- Idea Para verificar si ampliar la cobertura de los registros contenidos en las variables mejoran los resultados o no.
- Descripción breve: Utilizar el registro completo de la comuna, que incluye conteos de las áreas urbana y rural para la construcción de los indicadores.
- Resultado: No hubo mejoras en los resultados, pues la mayor parte de la población se concentra en el área urbana ($\geq 85\%$).
- ¿Se implementó? NO

4. Intercepto

- Idea Para verificar si eliminar la intercepción mejora los resultados o no.
- Descripción breve: Se deshabilitó la opción de intercepción en la función fh de emdi y se compararon estimaciones con/sin intercepción.
- Resultado: Los resultados fueron peores que la configuración original para robo con violencia o intimidación (es decir, con B_0), pero funcionó para robo por sorpresa, donde incluso, el intercepción no resultó estadísticamente significativo.
- ¿Se implementó? SÍ

5. Opciones alternativas para el suavizamiento de la varianza

- Idea Para verificar el efecto de incluir comunas con estimación de varianza directa igual a 0.
- Descripción breve: A las comunas con estimación de varianza directa igual a 0 se le asigna el valor 0,0001 para que pueda ingresar al modelo de FVG.
- Resultado: Los resultados fueron peores que la configuración original (es decir, quitando las comunas con estimaciones igual a 0).
- ¿Se implementó? NO

8.4. Anexo D: Indicadores comunales

Tabla A. 11. Indicadores de comunas que integran el marco de ENUSC.

CUT	Comuna	Estimación SAE	ENUSC 2018 regional	Proyección de población 2018 (2017)	Tasa DMCS denuncias (CEAD)
1101	Iquique	39,23	34,73	209,409	2970,3
1107	Alto Hospicio	30,07	34,73	118,379	2215,8
2101	Antofagasta	21,49	22,48	395,387	2043,1
2201	Calama	26,55	22,48	177,642	3086,5
2301	Tocopilla	20,29	22,48	27,224	1924,8
3101	Copiapó	25,3	24,41	167,242	2562,8
3301	Vallenar	22,68	24,41	56,064	1352
4101	La Serena	19,32	18,94	238,659	2620,5
4102	Coquimbo	21,06	18,94	245,142	2434,1
4201	Illapel	14,4	18,94	32,411	1391,5
4203	Los Vilos	14,67	18,94	22,641	2195,1
4301	Ovalle	19,13	18,94	118,563	2042
5101	Valparaíso	25,49	25,08	310,570	3330,7
5103	Concón	23,95	25,08	44,335	2679,6
5109	Viña del Mar	25,68	25,08	353,000	2716,7
5301	Los Andes	23,03	25,08	67,071	2355,7
5501	Quillota	25,22	25,08	95,032	1833,1
5502	La Calera	25,06	25,08	52,996	1458,6
5601	San Antonio	24,15	25,08	95,130	2479,8
5701	San Felipe	24,84	25,08	81,120	2753,9
5801	Quilpué	24,2	25,08	162,464	2155,6
5802	Limache	22,62	25,08	48,633	1581,2
5804	Villa Alemana	29,75	25,08	134,099	1742,7
6101	Rancagua	28,08	24,14	258,738	2995,3
6115	Rengo	22,77	24,14	62,193	1443,9
6117	San Vicente	17,62	24,14	49,404	1441,2
6301	San Fernando	24,04	24,14	76,875	2327,2
7101	Talca	20,41	17,26	232,672	2969
7102	Constitución	12,22	17,26	49,932	1582,2
7201	Cauquenes	10,49	17,26	43,774	1174,2
7301	Curicó	19,21	17,26	158,795	2724,9
7304	Molina	13	17,26	48,412	1468,6
7401	Linares	19,38	17,26	99,056	2491,5
8101	Concepción	26,9	23,61	236,400	3919,2
8102	Coronel	27,18	23,61	123,634	1891,1
8103	Chiguayante	22,13	23,61	90,438	1067
8106	Lota	26,6	23,61	45,845	2166
8107	Penco	18,38	23,61	49,531	1887,7

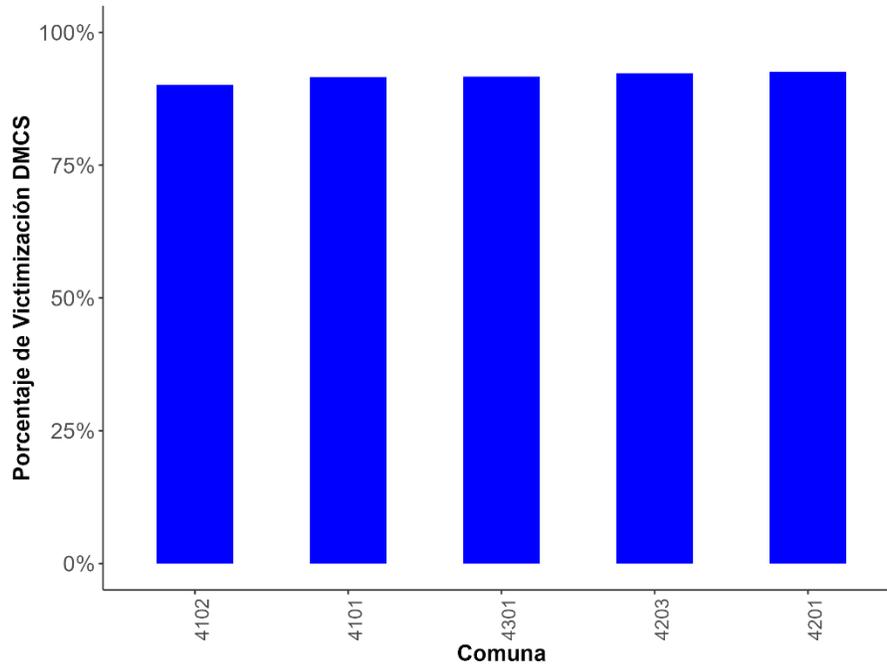
CUT	Comuna	Estimación SAE	ENUSC 2018 regional	Proyección de población 2018 (2017)	Tasa DMCS denuncias (CEAD)
8108	San Pedro De La Paz	33,48	23,61	140,877	1549,6
8110	Talcahuano	22,18	23,61	158,087	1876,2
8111	Tomé	17,55	23,61	58,294	1547,3
8112	Hualpén	27,32	23,61	96,499	1754,4
8301	Los Ángeles	24,02	23,61	214,799	2261,2
9101	Temuco	22,11	19,86	298,239	3123,3
9112	Padre Las Casas	21,18	19,86	80,067	1094,1
9120	Villarrica	19,29	19,86	58,025	1906,1
9201	Angol	19,47	19,86	55,451	2555,4
10101	Puerto Montt	19,79	16,2	262,245	2461,4
10201	Castro	14,72	16,2	46,805	2076,7
10202	Ancud	12,97	16,2	42,301	1401,9
10301	Osorno	16,03	16,2	171,233	1766,6
11101	Coyhaique	12,78	12,57	60,410	1648,7
11201	Aysén	13,2	12,57	24,749	1179,8
12101	Punta Arenas	6,95	6,55	138,248	1061,1
13101	Santiago	32,66	30,12	467,865	5786,1
13102	Cerrillos	40,95	30,12	86,451	3023,7
13103	Cerro Navia	32,53	30,12	140,355	1595,2
13104	Conchalí	31,02	30,12	135,099	2291,7
13105	El Bosque	37	30,12	171,032	2178
13106	Estación Central	31,88	30,12	166,174	4188,4
13107	Huechuraba	34,83	30,12	106,706	2478,8
13108	Independencia	46,38	30,12	117,277	2757,6
13109	La Cisterna	41,78	30,12	97,125	3541,8
13110	La Florida	29,22	30,12	390,218	2698,5
13111	La Granja	30,33	30,12	122,392	2681,5
13112	La Pintana	26,54	30,12	188,255	2105,7
13113	La Reina	25,11	30,12	97,810	2809,5
13114	Las Condes	23,59	30,12	315,183	2578,8
13115	Lo Barnechea	24,34	30,12	114,322	1370,7
13116	Lo Espejo	30,28	30,12	103,454	2394,3
13117	Lo Prado	31,48	30,12	101,803	2241,6
13118	Macul	28,56	30,12	126,804	2552
13119	Maipú	28,38	30,12	556,715	1710
13120	Ñuñoa	31,01	30,12	230,808	3287,6
13121	Pedro Aguirre Cerda	38,46	30,12	106,605	2205,3
13122	Peñalolén	25,19	30,12	257,714	1746,5
13123	Providencia	30,05	30,12	151,042	7486,7
13124	Pudahuel	32	30,12	244,526	2310,6

CUT	Comuna	Estimación SAE	ENUSC 2018 regional	Proyección de población 2018 (2017)	Tasa DMCS denuncias (CEAD)
13125	Quilicura	24,6	30,12	232,342	2034,9
13126	Quinta Normal	46	30,12	123,648	3580,3
13127	Recoleta	30,42	30,12	173,464	4134,6
13128	Renca	32,33	30,12	156,567	2278,3
13129	San Joaquín	26,71	30,12	100,566	2496,9
13130	San Miguel	40,3	30,12	120,174	3578,1
13131	San Ramón	35,84	30,12	86,575	2325,2
13132	Vitacura	25,03	30,12	91,198	3286,3
13201	Puente Alto	36,1	30,12	615,557	1991,4
13301	Colina	21,82	30,12	163,779	2085,1
13302	Lampa	32,13	30,12	115,058	2032,9
13401	San Bernardo	30,16	30,12	323,415	3055,8
13402	Buín	23,5	30,12	104,338	1684,9
13404	Paine	25,86	30,12	78,650	1713,9
13501	Melipilla	27,28	30,12	135,945	2417,2
13601	Talagante	24,19	30,12	79,158	1888,6
13604	Padre Hurtado	21,96	30,12	69,538	1196,5
13605	Peñaflor	23,9	30,12	97,255	1651,3
14101	Valdivia	23,62	20,66	173,420	2449,5
14201	La Unión	17	20,66	39,358	1016,3
15101	Arica	24,37	24,34	237,412	1976,3
16101	Chillán	27,31	23,78	195,042	2258,5
16103	Chillán Viejo	22,93	23,78	32,735	1365,5
16301	San Carlos	20,97	23,78	55,608	1827,1

Fuente: Elaboración propia.

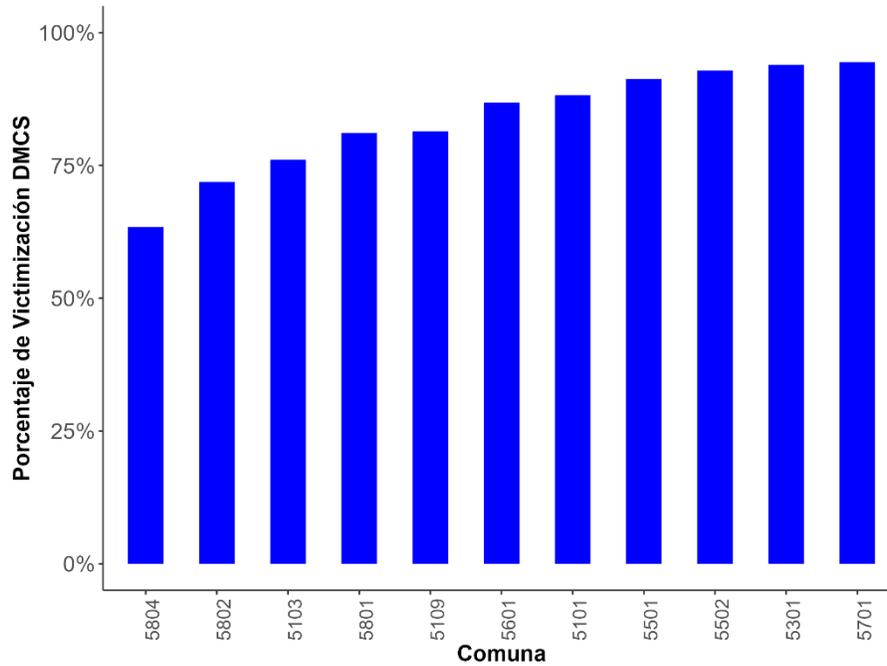
8.5. Anexo E: Denuncias comunales por región

Figura A. 29. Porcentaje de hogares en la muestra que fueron víctima de algún DMCS dentro de su propia comuna, Región de Coquimbo.



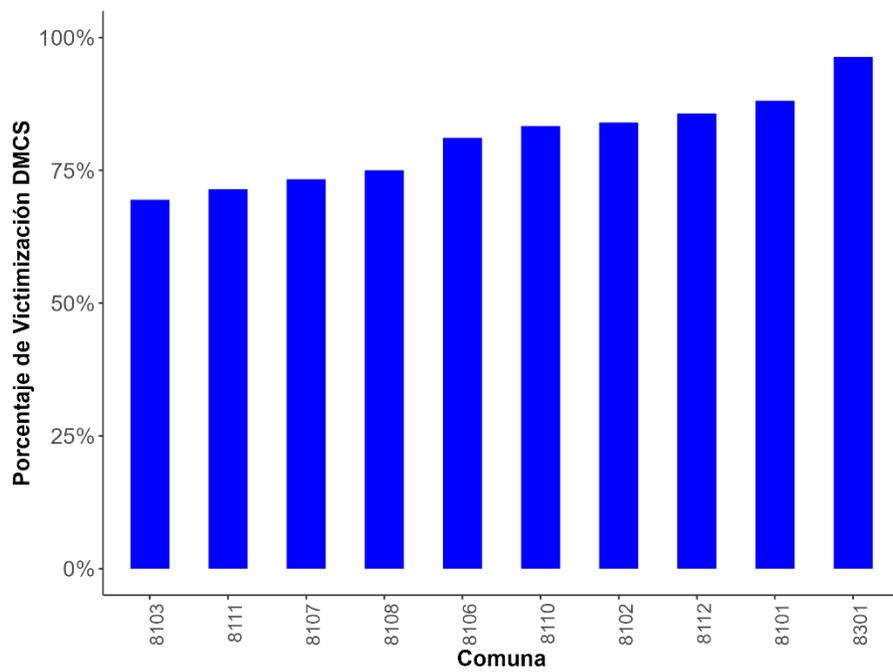
Fuente: Elaboración propia.

Figura A. 30. Porcentaje de hogares en la muestra que fueron víctima de algún DMCS dentro de su propia comuna, Región de Valparaíso.



Fuente: Elaboración propia.

Figura A. 31. Porcentaje de hogares en la muestra que fueron víctima de algún DMCS dentro de su propia comuna, Región del Biobío.



Fuente: Elaboración propia.