



ESTIMACIÓN

ÁREAS PEQUEÑAS

INFORME METODOLÓGICO

ESTIMACIÓN DE ÁREAS PEQUEÑAS

INFORME METODOLÓGICO

**Instituto Nacional de Estadísticas
Enero / 2026**

INSTITUTO NACIONAL DE ESTADÍSTICAS

Dirección Nacional

Departamento de Metodologías e Innovación Estadística
Subdepartamento de Investigación Estadística

Subdirección Técnica

Departamento de Estadísticas Demográficas y Sociales
Subdepartamento de Estadísticas de Condiciones de Vida

ENCUESTA NACIONAL URBANA DE SEGURIDAD CIUDADANA 2024

Estimación de Áreas Pequeñas
Informe Metodológico
Enero 2026

En este documento se presentan resultados comunales obtenidos mediante SAE para la Victimización de hogares por delitos violentos (VHDV) de la ENUSC 2024. Estos datos se consideran estadísticas experimentales, vale decir, corresponden a resultados y/o productos estadísticos que se generan a partir de metodologías y/o fuentes de datos que se encuentran en una etapa inicial de madurez dentro de la producción de estadísticas oficiales. Su elaboración y difusión busca ampliar, mejorar y fortalecer la oferta estadística vigente, resguardando la transparencia respecto de sus alcances y limitaciones metodológicas.

Dirección: Morandé 801, Santiago de Chile

Teléfono: (56) 228924000

Sitio web: www.ine.gob.cl

Correo electrónico: ine@ine.gob.cl

X: @INE_Chile

Instagram: Chile.INE

Linkedin: Instituto Nacional de Estadísticas de Chile

Facebook: @ChileINE

ÍNDICE

ÍNDICE	4
Índice de tablas	5
Índice de figuras	5
1. INTRODUCCIÓN	6
2. EXPERIENCIA Y CONCEPTOS PRELIMINARES	9
2.1 Sobre las experiencias en estimaciones de áreas pequeñas	9
2.2 Estimaciones directas	11
2.3 Estimación indirecta y estimación en áreas pequeñas	12
3. METODOLOGÍA	14
4. ESPECIFICACIÓN	17
4.1 Necesidad de los usuarios	17
4.2 Fuentes de información auxiliar	18
4.2.1 Registro Social de Hogares	24
4.2.2 Registros Administrativos del Ministerio de Desarrollo Social y Familia	25
4.2.3 Sistema Nacional de Información Municipal	26
4.2.4 Imágenes satelitales – Radiancia nocturna	26
4.2.5 Registros Administrativos de Carabineros de Chile	27
4.2.6 Información estadística del INE	29
5. ANÁLISIS Y ADAPTACIÓN	32
5.1. Estimador directo	32
5.1.1. Suavizamiento de la varianza directa	34
5.1.2. Criterios de calidad para la inclusión o exclusión de comunas en el modelo	37
5.2. EBLUP basado en el modelo Fay–Herriot	39
5.3. Transformación arcoseno	42
5.4. Selección de covariables	43
6. EVALUACIÓN	47
6.1. Tratamiento de valores atípicos e influyentes	49
6.2. Bondad de ajuste	50
6.3. Benchmarking	51
6.4. Error Cuadrático Medio y Coeficiente de Variación Logarítmico	54
7. ESTIMACIONES	57
8. DISCUSIÓN Y TRABAJO FUTURO	60
9. BIBLIOGRAFÍA	63
10. ANEXOS	69
Anexo 1: Fuentes de información, dimensión de análisis y covariables consideradas en la implementación de modelos Fay-Herriot para VHDV	69

ÍNDICE DE TABLAS

Tabla 1. Componentes del indicador	18
Tabla 2. Dimensiones, fuentes y covariables relacionadas con la victimización	23
Tabla 3. Resumen de criterios de inclusión y exclusión	38
Tabla 4. Covariables en modelos SAE para VHDV	45
Tabla 5. Resumen de test de normalidad, homocedasticidad y no autocorrelación	48
Tabla 6. Coeficientes λ_{ir} Benchmarking	53
Tabla 7. Medidas de resumen para comparar desempeño del estimador directo y F-H	56

ÍNDICE DE FIGURAS

Figura 1. Marco metodológico para la producción de estadísticas en áreas pequeñas	16
Figura 2. Análisis de calidad de ajuste de FVG para VHDV	36
Figura 3. Flujo de calidad para inclusión de estimaciones directas en VHDV.....	38
Figura 4. Residuos estandarizados vs. valores predichos para VHDV	47
Figura 5. Q-Q plot. Residuos Estandarizados Izq. - Efectos aleatorios Der.....	48
Figura 6. Distancias de Cook según comunas para VHDV.....	50
Figura 7. Verificación pre benchmarking VHDV	53
Figura 8. RMSE comunales SAE vs. Directas para VHDV	56
Figura 9. Estimaciones comunales SAE vs. Directas para VHDV.....	57
Figura 10. Estimación directa vs. Estimación EBLUP para VHDV.....	58
Figura 11. Mapa comunal según zona geográfica para estimación SAE de VHDV	59

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)**. El objetivo general de esta operación estadística 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 las zonas urbanas del país ([Instituto Nacional de Estadísticas - Chile, 2025b](#)). Esto 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. Sin embargo, en los últimos años existe una demanda creciente de información para grupos específicos de la población o para áreas geográficas menores. De hecho, **el marco global de indicadores para el seguimiento de los Objetivos del Desarrollo Sostenible plantea que la información debe ser desagregada no solo de manera geográfica, sino también por grupos de ingreso, sexo, edad, raza, origen étnico, estatus migratorio y condición de discapacidad**. En ese sentido, es relevante avanzar en la implementación de metodologías estadísticas que permitan, por ejemplo, la obtención de estimaciones sobre victimización, en subgrupos de poblaciones específicas o dominios no considerados en el diseño de la ENUSC.

Las estimaciones obtenidas a través de la encuesta, en adelante estimaciones directas, en general, mantienen su precisión siempre y cuando se realicen para dominios de estudio que han sido planificados previamente en el diseño muestral. En consecuencia, esta precisión va decreciendo cuando se realizan inferencias para niveles de desagregación para los cuales la encuesta no fue diseñada, denominadas áreas pequeñas. La ENUSC 2023 tuvo un diseño muestral con representatividad a nivel nacional urbano, regional urbano y comunal urbano (136 comunas), por lo que se logró producir estimaciones con una precisión aceptable para la mayor parte de estas desagregaciones. Sin embargo, un diseño muestral como el de la ENUSC 2023, implica tamaños de muestra elevados, mayores costos y exigencias operativas importantes. La ENUSC 2024 tuvo un diseño con representatividad nacional y regional en zonas urbanas, por tanto, el nivel comunal

urbano deja de ser un dominio planificado y puede ser considerado como un área pequeña. En este escenario, cobra relevancia el uso de estimadores indirectos basados en modelos, que permiten extender el sistema inferencial a desagregaciones de interés no planificadas en el diseño. Dichos estimadores se construyen combinando la información de la encuesta con fuentes auxiliares externas —por ejemplo, censos, registros administrativos e imágenes satelitales— lo que contribuye a obtener estimaciones más precisas y exactas en estas áreas pequeñas.

Entre los estimadores indirectos, se encuentran aquellos basados en modelos de regresión explícitos, llamados estimadores basados en modelos. [Molina y Marhuenda \(2015\)](#) expresan que estos estimadores se basan en asumir una relación entre la variable objetivo y algunas variables explicativas que son constantes en todas las áreas. Los parámetros comunes del modelo se estiman utilizando el conjunto completo de datos de la muestra, lo que a menudo conduce a estimadores de áreas pequeñas con una eficiencia apreciablemente mejor que los estimadores directos, siempre que se mantengan los supuestos del modelo. **Este conjunto de técnicas que proporcionan cifras estadísticas a un nivel desagregado tiene lugar en una metodología estadística conocida como estimación en áreas pequeñas (en inglés, *small area estimation* o SAE).** Los modelos SAE generalmente se clasifican en dos grandes clases basadas en la disponibilidad de las fuentes de información auxiliar: (i) modelos a nivel de área donde la información sobre la variable respuesta está disponible agregada al nivel de área pequeña o dominio de interés; y (ii) modelos a nivel de unidad donde la información sobre la variable respuesta está disponible a nivel de unidad y puede asociarse a ese nivel con el conjunto de variables auxiliares.

En ese contexto, el INE junto con la SPD, acordaron la implementación de la metodología SAE para la obtención de estimaciones a nivel comunal en el indicador de **Victimización a Hogares por Delitos Violentos (VHDV) de la ENUSC 2024**. Por tanto, este documento presenta el marco conceptual y metodológico sobre la implementación del proceso de estimación de áreas pequeñas. Luego de presentar la introducción, antecedentes y algunas definiciones preliminares, la [sección 3](#) describe el marco metodológico sobre el cual se basa el proceso de estimación de áreas pequeñas; en la [sección 4](#) se describe las necesidades de los usuarios en términos de objetivos de estimación y área geográfica objetivo, así como la descripción general de las fuentes de información auxiliar usadas, su calidad y especificaciones sobre el proceso de adaptación de estas. En la [sección 5](#) se presenta una revisión del Mejor Estimador Lineal Empírico Insesgado (EBLUP, por sus siglas en

inglés), basado en el modelo Fay–Herriot y la adaptación de los modelos considerados para la estimación de la tasa de victimización por delitos violentos¹. En la [sección 6](#) se presenta la evaluación el diagnóstico del modelo implementado, y en la [sección 7](#) se presentan los principales resultados. Finalmente, en la [sección 8](#) se discuten elementos de la metodología implementada, los alcances y aplicaciones en futuros trabajos.

1 En el presente documento, el término “tasa de victimización” se utiliza como equivalente operativo de “proporción de hogares victimizados en el periodo de referencia (últimos 12 meses)”, en concordancia con la definición adoptada por la ENUSC.

2. EXPERIENCIA Y CONCEPTOS PRELIMINARES

En las últimas décadas, la estimación en áreas pequeñas se ha incorporado de manera sistemática en los sistemas estadísticos oficiales de distintos países, principalmente para responder a la demanda de estimaciones precisas y confiables de indicadores desagregados, sin necesidad de recurrir a cambios en el diseño muestral (aumento tamaño de la muestra, redistribución eficiente del tamaño muestral en ciertas áreas, entre otros). En la [sección 2.1](#) de forma general, se mencionan algunas de estas experiencias. Por otro lado, antes de introducir la metodología SAE y su implementación en la ENUSC 2024, se hace necesario tener claro el concepto de estimación directa, estimación indirecta y estimación en área pequeña, en la [sección 2.2](#) y [2.3](#), se presentan estos conceptos.

2.1 SOBRE LAS EXPERIENCIAS EN ESTIMACIONES DE ÁREAS PEQUEÑAS

En materia de indicadores de pobreza, el Banco Mundial acumula más de tres décadas de trabajo en estimación en áreas pequeñas orientada a la elaboración de mapas de pobreza. Desde finales de los años noventa han desarrollado aplicaciones basadas en modelos a nivel de unidad y de área, donde el enfoque de [Elbers, Lanjouw y Lanjouw \(2003\)](#) se convirtió durante largo tiempo en el método de referencia, complementado luego con software específico (*PovMap*, comandos en *Stata*) y con directrices que ordenan las alternativas metodológicas disponibles ([Corral et al., 2022](#)). En la misma línea, el U.S. Census Bureau implementa el programa Small Area Income and Poverty Estimates (SAIPE), que produce anualmente estimaciones de pobreza e ingreso para estados, condados y distritos escolares mediante modelos a nivel de área de: número total de personas en situación de pobreza, pobreza infantil, mediana del ingreso de los hogares y recuentos de población y de niños en edad escolar, entre otros; que se utilizan como insumo en la administración y en las fórmulas de asignación de recursos de diversos programas federales y estatales². En el

² <https://www.census.gov/programs-surveys/saipe/about.html>

ámbito del mercado laboral, Statistics Canada ha desarrollado una metodología de SAE asociada a la encuesta de fuerza laboral (Labour Force Survey) para obtener estimaciones mensuales de empleo y desempleo en desagregaciones subprovinciales, como las áreas metropolitanas censales, las aglomeraciones censales y las áreas laborales autocontenidoas (self-contained labour areas). El enfoque se basa en modelos a nivel de área, cuyos resultados se generan mediante el módulo de SAE del software generalizado *G-EST*³. En Europa, por su parte, se han elaborado directrices específicas para el uso de SAE en estadísticas de ciudades y áreas urbanas funcionales ([Münich et al., 2019](#)), con el objetivo de mejorar la precisión de indicadores sociales y económicos en dominios no planificados del diseño muestral y avanzar hacia marcos armonizados que aseguren coherencia y comparabilidad entre países.

En cuanto a la criminalidad y la percepción de seguridad, la experiencia reciente de México constituye un referente para el uso de SAE con encuestas de victimización. A partir de la Encuesta Nacional de Victimización y Percepción de Seguridad Pública 2021 (ENVIPE 2021), se estimó, para los 2.469 municipios del país, el porcentaje de personas de 18 años o más que perciben inseguridad en su lugar de residencia, utilizando predictores a nivel de área y, además, incorporando una componente espacial, ambos basados en el modelo de Fay–Herriot ([Santillana et al., 2023](#)). En la literatura internacional, destaca también el trabajo de David Buil-Gil, doctor en criminología y especialista en métodos cuantitativos, quien ha desarrollado múltiples aplicaciones de estimación en áreas pequeñas para estudiar tanto la cifra oculta de la delincuencia como la percepción por el delito y la confianza en la policía a partir de encuestas de victimización en Europa y el Reino Unido⁴.

Finalmente, en cuanto a la experiencia a nivel nacional, en Chile, la estimación en áreas pequeñas se ha utilizado principalmente para la medición de la pobreza y otros indicadores sociales a nivel comunal. Desde el 2009, el Ministerio de Desarrollo Social y Familia aplica modelos de tipo Fay–Herriot a partir de la encuesta CASEN para obtener tasas de pobreza a nivel de comuna y, más recientemente, indicadores comunales de pobreza multidimensional e inseguridad alimentaria,

3 <https://www150.statcan.gc.ca/n1/pub/75-005-m/75-005-m2024003-eng.htm>

4 <https://research.manchester.ac.uk/en/persons/david.builgil>

cuyas estimaciones se emplean como insumo en la asignación de recursos y en el seguimiento de brechas territoriales ([MDSF-CEPAL, 2020](#)). En este contexto, las experiencias internacionales y nacionales muestran que la metodología SAE se ha consolidado como una herramienta útil para mejorar la precisión y confiabilidad de las estimaciones en desagregaciones que no han sido previamente planificadas en el diseño muestral. Sobre esta base, el presente trabajo busca aplicar estos enfoques al ámbito de la seguridad ciudadana, utilizando modelos de tipo Fay–Herriot para mejorar la precisión del indicador de victimización a hogares por delitos violentos a nivel comunal y aportar información confiable para el seguimiento de brechas territoriales y la planificación de políticas públicas a ese nivel de desagregación.

2.2 ESTIMACIONES DIRECTAS

Las estimaciones directas son aquellas que se realizan siguiendo el diseño de muestreo propuesto para la encuesta sin usar otro tipo de información auxiliar y para las cuales es necesario validar y presentar las medidas de incertidumbre como varianza, intervalos de confianza, coeficientes de variación, efecto de diseño (*DEFF*), entre otros, con el fin de determinar la calidad de la estimación y su validez.

Para la desagregación de variables a través de encuestas por muestreo existen dos tipos de grupos: i) los estratos son subgrupos de la población, mutuamente excluyentes entre sí, estos son definidos desde el diseño muestral, y ii) los dominios que son similares a los estratos y corresponden a subgrupos poblacionales, para el cual se contempla contar con estadísticas representativas producidas por el estudio. Por ejemplo, un dominio podría consistir en un área geográfica como una región o un centro de población importante. También podría comprender una categoría de población específica, como un grupo nacional o étnico importante. La cantidad y tipo de dominios tiene una relación importante con el tamaño y la distribución de la muestra, ya que influye en la posibilidad de obtener estimaciones representativas para cada uno de ellos.

Sin embargo, pueden existir subdivisiones (desagregaciones) para las cuales se desea obtener estimaciones y que no fueron consideradas en el diseño de la encuesta, por tanto, el tamaño muestral en algunos casos puede ser demasiado insuficiente, lo cual se traduce en errores de muestreo excesivamente grandes para los estimadores directos. Las subdivisiones o áreas en las que esto

ocurre, independientemente de su tamaño poblacional, se conocen en la literatura como “áreas pequeñas”.

2.3 ESTIMACIÓN INDIRECTA Y ESTIMACIÓN EN ÁREAS PEQUEÑAS

De forma general, la metodología de estimación en áreas pequeñas es usada cuando a través de las estimaciones directas no es posible obtener resultados para desagregar a los dominios de interés, usualmente no planificados y que tienen un tamaño de muestra pequeño o nulo. Generalmente, los dominios están definidos como áreas geográficas (provincias, comunas, barrios) o grupos poblacionales (sexo, edad, grupos étnicos, LGBTIQ+, víctimas, etc.). Para la aplicación de SAE se hace uso de **estimadores indirectos** los cuales consideran la información de la encuesta, variables auxiliares de otras fuentes de información y la relación existente con otras áreas o dominios. Según [Rao & Molina \(2015\)](#) la ganancia en eficiencia de los estimadores indirectos respecto de los directos es mayor cuanto menor es el tamaño de muestra del dominio. Al usar estimadores indirectos se está “sacrificando” el insesgamiento de los estimadores basados en el diseño⁵ y es necesario escoger un modelo adecuado que se ajuste correctamente a los datos, buscando mejorar la precisión de las estimaciones.

Los métodos de estimación basados en modelos asumen un modelo estadístico tanto para la población como para los datos muestrales, y a partir de este se construyen predictores óptimos de los parámetros de interés bajo los supuestos del modelo. En este contexto, convencionalmente se utiliza el término “predictor” en lugar de “estimador” debido a que el modelo utilizado supone que los parámetros objetivos 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 combinación de los componentes: 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

⁵ Insesgado bajo el diseño significa que, promediando sobre todas las muestras posibles generadas por el diseño, el estimador tiene esperanza igual al parámetro poblacional (Thompson, 2012).

datos de encuestas, a nivel del área geográfica de interés, pero también incorpora datos de otras áreas (como entrada a los parámetros del modelo) y datos auxiliares, provenientes de fuentes externas.

Es importante considerar que la información auxiliar debe provenir de una fuente independiente de la encuesta y 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 solo en los resultados de la estimación, sino también en la carga de trabajo, ya que toda información susceptible de uso debe ser previamente revisada y adaptada según las necesidades del modelo.

A partir de esta clasificación general, cabe destacar que en la literatura se reconocen dos enfoques principales de modelamiento SAE: los modelos a nivel de área (o agregados) y los modelos a nivel de unidad, los cuales difieren según el nivel al que se dispone la información de la variable respuesta y las covariables [\(Chatrchi, 2019\)](#). En este estudio, dado que las fuentes auxiliares más relevantes —como los datos policiales, por una parte y por otra, los indicadores sociodemográficos— se encuentran agregadas a nivel comunal, se opta por el enfoque de área, cuya formulación teórica se presenta más adelante.

3. METODOLOGÍA

La estimación de áreas pequeñas es un área de investigación en las estadísticas oficiales y de encuestas de gran relevancia práctica para Oficinas Nacionales de Estadística (ONE) y organizaciones afines, que ha tenido rápidos avances en metodología y *software*. Para implementarla es necesario contar con directrices prácticas para el proceso de estimación de áreas pequeñas, por tanto, se ha seguido el marco general para la producción de estadísticas de áreas pequeñas (ver Figura 1), propuesto por [Tzavidis et al. \(2018\)](#), regido por el principio de parsimonia y basado en tres etapas generales: especificación, análisis y adaptación, y evaluación.

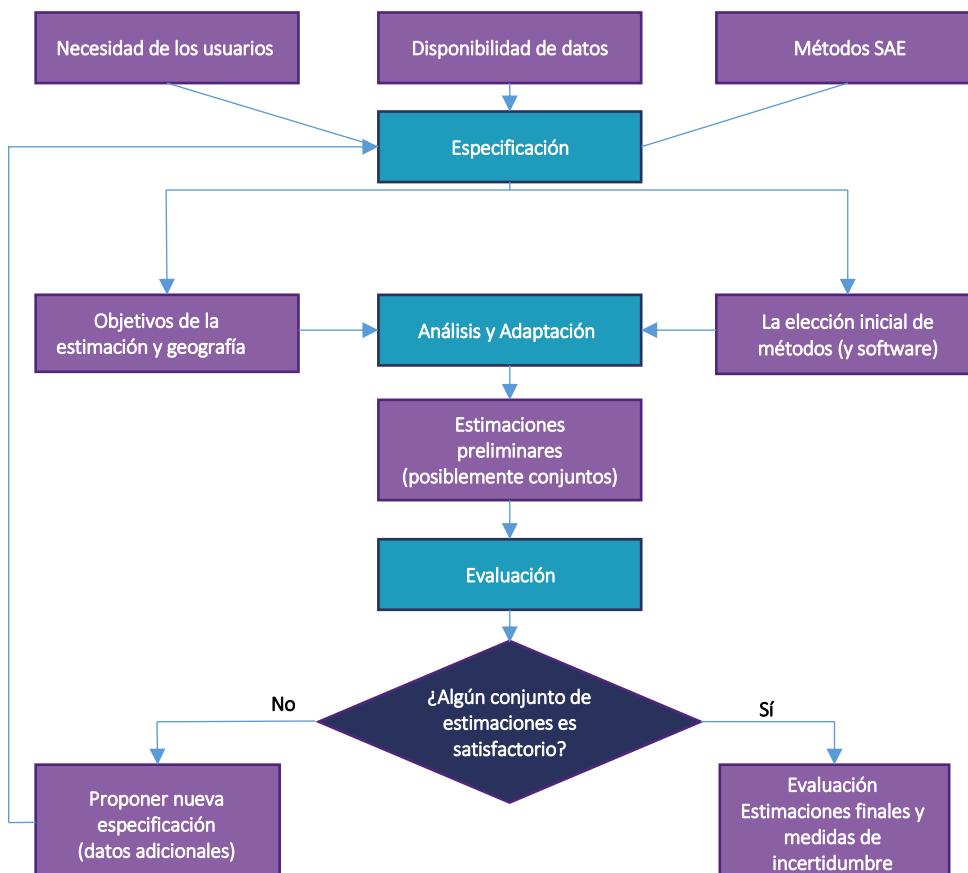
Una descripción de las necesidades del usuario, revisión de los datos disponibles y de los métodos SAE existentes, son los insumos más importantes para la primera etapa de especificación. Con el apoyo de los analistas técnicos, el usuario define un conjunto de áreas o dominios de interés e indicadores objetivos y luego identifica los posibles métodos de áreas pequeñas existentes que son aplicables, en función de la disponibilidad, calidad y tipo de datos. Estos son los insumos necesarios para la segunda etapa. En particular, el usuario principal de las estadísticas oficiales de victimización comunal es la Subsecretaría de Prevención del Delito, dependiente del Ministerio de Seguridad Pública.

La segunda etapa, análisis y adaptación del modelo, es donde se desarrollan los estimadores. Desde un punto de vista analítico, es útil si este proceso se rige por el principio de parsimonia, es decir, optar por **el método más simple que proporcione una precisión aceptable**. La parsimonia puede definirse en términos de una jerarquía de métodos de estimación en orden creciente de complejidad. Esto puede incluir estimadores directos, sintéticos y compuestos. Comúnmente, estos estimadores se pueden mejorar mediante el uso de modelos de nivel de unidad o modelos de nivel de área ([Molina, 2019](#)). Sin embargo, esta etapa implica un mayor grado de complejidad técnica, pues implica la construcción de modelos estadísticos adecuados y la realización de diagnósticos. Finalmente, las especificaciones del modelo pueden incluir: consideraciones sobre los dominios a ser excluidos de los modelos dada la falta de precisión, transformaciones, efectos aleatorios correlacionados en el tiempo y espacio, efectos aleatorios no normales y estimadores robustos, implementaciones de modelos semiparamétricos o no paramétricos, entre otros. El principio de parsimonia dicta que tal esfuerzo solo debe introducirse para superar las deficiencias específicas

que se han identificado en los métodos más básicos, y la mejora potencial debe sopesarse frente a la complejidad adicional y los posibles inconvenientes. Si bien esta definición de parsimonia no es exacta, proporciona un marco útil para guiar el proceso de producción de estimaciones de áreas pequeñas.

El objetivo de la tercera etapa del marco general es la evaluación de los múltiples conjuntos de estimaciones que se producen en la etapa anterior, **esto implica tanto la evaluación de la incertidumbre como la evaluación del método**. Concretamente, **el proceso SAE finaliza cuando al menos un conjunto de estimaciones se considere de precisión aceptable, es decir, alcance el umbral de precisión previamente establecido**. Sin embargo, es fundamental definir qué es lo que se entiende por precisión aceptable, y comparando de manera relativa una gama de métodos en términos de ganancias en precisión, sensibilidad a los supuestos subyacentes del modelo, los recursos adicionales para su implementación, así como los costos y riesgos operativos asociados. Si después de seguir estos pasos no se ha encontrado un conjunto de estimaciones aceptables de área pequeña, el proceso puede necesitar volver a la etapa de especificación para definir geografías o subpoblaciones alternativas, indicadores objetivos y/o nuevas fuentes de información auxiliar.

Figura 1. Marco metodológico para la producción de estadísticas en áreas pequeñas



Fuente: Instituto Nacional de Estadísticas (INE) basado en [Tzavidis et al. \(2018\)](#).

4. ESPECIFICACIÓN

4.1 NECESIDAD DE LOS USUARIOS

Una tarea importante en esta etapa es la especificación del nivel objetivo de la geografía o grupo poblacional y los objetivos de la estimación, lo que impactará en todo el proceso SAE posterior. Además del nivel objetivo de la geografía y los objetivos de la estimación, las propiedades más importantes del método de estimación también deben definirse, por ejemplo, si el usuario está más interesado en estimaciones transversales o estimaciones longitudinales, lo cual determinará tanto los datos que se requieren como los modelos que se utilizan.

La ENUSC es una encuesta nacional que produce estimaciones anuales sobre percepción de inseguridad, reacción frente a delitos y la victimización según lo informado por las personas de 15 años o más residentes en viviendas particulares ocupadas de las zonas urbanas del país, utilizando como periodo de referencia los últimos 12 meses desde la aplicación de la entrevista. En el caso de la ENUSC 2024, se utilizó un diseño muestral trietápico⁶ cuya población objetivo está conformada por las personas de 15 años o más y los hogares de las viviendas particulares ocupadas del área urbana de 136 comunas en las que tiene cobertura geográfica la ENUSC. Los datos de la ENUSC 2024 han sido proporcionados por los informantes seleccionados y encuestados, su tamaño de muestra efectivo fue de 24.472 viviendas, en 136 de 346 comunas, abarcando las 16 regiones del país. Esta encuesta fue diseñada para proporcionar estimaciones con precisión aceptable a nivel nacional-urbano y regional-urbano, por tanto, en este caso, el usuario (SPD) está interesado en obtener las estimaciones de corte transversal de tasas de victimización para el nivel communal-urbano. Para eso, en una primera etapa se considera el indicador principal de la encuesta, victimización a hogares por delitos violentos.

⁶ Para más detalles, véase Informe metodológico ENUSC 2024:
<https://www.ine.gob.cl/estadisticas/sociales/seguridad-publica-y-justicia/seguridad-ciudadana>

Tabla 1. Componentes del indicador

Indicador	Delitos que lo compone
Victimización a hogares por delitos violentos (VHDV)	Robo de vehículo con violencia; robo con fuerza en la vivienda con violencia, robo con violencia e intimidación, robo por sorpresa con violencia; agresiones y lesiones; amenaza con violencia; extorsión con violencia.

Fuente: Manual de usuario de la base de datos - ENUSC 2023.

Dado que se tiene acceso tanto a la base de datos de la ENUSC 2024 como a diversas fuentes de información auxiliar agregadas a nivel de comuna, es factible implementar algún modelo SAE a nivel de área, en este caso, a nivel comunal. Es importante tener presente que el uso de métodos SAE se justifica si: i) pueden producir estimaciones comunales más eficientes que las estimaciones directas y ii) pueden producir estimaciones aceptables para comunas no incluidas en la muestra⁷, siendo el caso particular de este trabajo el punto i). Finalmente, es importante comunicar cuidadosamente el efecto potencial de los supuestos del modelo y evaluar adecuadamente los métodos y las estimaciones.

4.2 FUENTES DE INFORMACIÓN AUXILIAR

Identificar qué datos se necesitan tiene impacto no solo en los resultados de la estimación, sino también en términos de costo y eficiencia del personal que lo trabaja. La metodología SAE básicamente consiste en solucionar un problema de predicción y generalmente se basa en uso de datos de encuestas y datos del censo o registros administrativos (RR.AA.). El acceso a las fuentes de datos censales y administrativos suele ser un desafío debido a las restricciones de confidencialidad.

⁷ Para la ENUSC 2024, más del 50% de las comunas con área urbana no son consideradas en la muestra, lo que hace imposible su estimación directa. Según el Censo de Población y Vivienda 2017, hay 27 comunas que son 100% urbanas, 27 comunas 100% rurales y 292 comunas mixtas.

Comúnmente, el acceso a los datos del nivel agregado (área o dominio) es posible, pero el acceso a los microdatos puede no ser posible. La pregunta es cómo el tipo de datos disponibles afecta a SAE, en ese sentido, si se está interesado en estimar estadísticas lineales, por ejemplo, promedios de área pequeña, los datos agregados serán suficientes para SAE ([Tzavidis et al., 2018](#)). En relación con lo mencionado en el párrafo anterior, tres aspectos importantes deben ser considerados con respecto a las fuentes de información auxiliar, a saber:

- (a) Identificación de indicadores o covariables relacionados con los fenómenos de victimización que se desean estimar.
- (b) Disponibilidad de fuentes de información que provean los indicadores identificados o provean información para su construcción.
- (c) Validación de la calidad de las fuentes de información en dimensiones como completitud, consistencia, oportunidad, etc.

Con respecto al [punto \(a\)](#), de acuerdo con las investigaciones revisadas, es posible distinguir las siguientes dimensiones de análisis para la indagación de fuentes de información:

- **Características sociodemográficas personales:** Estas características determinan un estilo de vida particular relacionado con la exposición a lugares y situaciones con diferentes grados de peligrosidad. Estos atributos llevan consigo expectativas de los sujetos sobre el comportamiento apropiado en estos espacios ([Armas Montalvo & Herrera, 2018](#)). En este sentido es relevante la variable sexo, donde hombres y mujeres tienen distintas probabilidades de ser víctima dependiendo del tipo de delito⁸. Por otro lado, el efecto de la edad en la victimización tiene relación con las diferencias en el grado de exposición a los espacios públicos, de acuerdo con las conductas en estos. Según la investigación del [INEGI \(2017\)](#), la edad tiende a aumentar el riesgo de victimización, sin embargo, existe un punto donde esta asociación comienza a disminuir; o bien, los

⁸ Si bien, algunas investigaciones respaldan que los hombres son más victimizados que las mujeres a nivel macro (INEGI, 2017), esto debe ser revisado con mayor profundidad respecto a delitos específicos, como son los robos en el transporte público, la violencia en el contexto intrafamiliar o la violencia sexual, o desde una perspectiva interseccional que integre distintas variables sociodemográficas al análisis por sexo.

jóvenes tienden a ser más victimizados por delitos violentos, fundamentalmente por la mayor exposición a lugares y actividades de mayor riesgo ([Armas Montalvo & Herrera, 2018](#)). En general las variables relacionadas a esta dimensión son: edad, sexo/género, etnia/nacionalidad, estado civil y características del hogar.

- **Características socioeconómicas personales:** En el marco de la teoría de los estilos de vida, las características socioeconómicas desempeñan un papel crucial para entender los factores que determinan las probabilidades de victimización de las personas. Los bajos ingresos, por ejemplo, limitan significativamente las opciones de vivienda, transporte y actividades de ocio ([Meier & Terance, 1993](#)). Aunque las personas con mejores ingresos suelen ser más atractivas como víctimas, también tienen la posibilidad de optar por entornos más seguros, con mayores medidas de protección, lo que les permite reducir situaciones de riesgo y realizar actividades de ocio en áreas más seguras. Las variables asociadas a esta dimensión son: educación, ingresos/nivel socioeconómico y situación en el mercado laboral.
- **Características de la infraestructura urbana y del entorno:** Aquí se agrupan variables vinculadas a la teoría de la desorganización social y describen el entorno físico y social en el que se desenvuelve la comunidad. Incluye aspectos como la infraestructura urbana, por ejemplo, iluminación pública o densidad de edificios y vegetación; y aspectos institucionales, como la vigilancia policial. En este sentido, las áreas con mayor densidad de población no solo pueden experimentar mayores tasas de crímenes, sino también tienen menor probabilidad de que estos sean denunciados, debido a una mayor probabilidad de anonimato y una menor vigilancia comunitaria ([Buil-Gil, Moretti, Shlomo, & Medina, 2020](#)). Situación similar ocurre con la presencia de la policía y de las instituciones encargadas de la seguridad pública, que no solo incide en la capacidad de prevenir situaciones o actuar frente al delito, sino que también su presencia afecta la percepción de las personas sobre su efectividad. Las variables relacionadas a esta dimensión son: calidad de las viviendas, dotación policial, tránsito, características de la ciudad (como urbanidad y vegetación) y acceso a servicios.
- **Factores relacionados con el crimen y la delincuencia:** Similar al punto anterior, esta dimensión está vinculada a la desorganización social y se enfoca en variables directamente relacionadas con la ocurrencia de delitos y la actividad criminal en una

comunidad, en las cuales se incluyen tasas de homicidios, robos, violaciones, agresiones sexuales, entre otros. Muchas de estas variables provienen de registros administrativos de denuncias realizadas a las policías, lo que las convierte en un insumo clave para los actores a nivel local, ya que actualmente son la fuente más utilizada para monitorear y analizar el comportamiento del crimen en los territorios ([Fay, Planty, & Diallo, 2013](#)). Sin embargo, las estadísticas policiales presentan limitaciones importantes para analizar tendencias delictivas y diferencias en el crimen, tanto en áreas grandes como pequeñas. Por ejemplo, los registros de denuncias suelen ser más bajos en vecindarios con desventajas económicas significativas, altas tasas de crimen, concentración de inmigrantes y bajos niveles de cohesión social. Estos factores pueden afectar de manera desigual la cifra oculta⁹ del crimen en determinadas áreas ([Buil Gil D., 2019](#)). A pesar de estas limitaciones, estas variables son esenciales para comprender el nivel de riesgo de victimización en áreas específicas y contribuyen de manera significativa a entender el fenómeno delictivo tanto en contextos locales como en regiones más amplias. Las principales variables asociadas a esta dimensión son: registros de denuncias/casos policiales, percepción de inseguridad, consumo de alcohol o drogas y tenencia de armas.

- **Características culturales y cohesión social:** Finalmente, desde la perspectiva de la teoría de la desorganización social, se aborda la incapacidad de una comunidad para mantener un control social efectivo, atribuida a la ausencia de instituciones sociales sólidas como escuelas, iglesias, organizaciones comunitarias y sistemas de apoyo familiar ([Olavarriá Gambi M., 2006](#)). Por otro lado, la cohesión social se entiende como un concepto más amplio que la seguridad, y hace referencia al grado en que los miembros de una sociedad o comunidad están unidos, compartiendo un sentido de solidaridad y pertenencia ([Díaz & Meller, 2012](#)). El crimen, en este marco, representa una amenaza directa para la cohesión social, ya que las comunidades que desarrollan desconfianza

⁹ Actos delictivos que no son denunciados a la policía o autoridad competente, o que no son objeto de una investigación y, por tanto, no figuran en ninguna estadística.

hacia sus miembros están más expuestas a la fragmentación y la estigmatización ([Dammert, 2012](#)). Un argumento clave en esta relación es que una mayor cohesión social puede reducir el crimen al fortalecer los vínculos comunitarios, facilitando la organización para abordar problemas de comportamiento oportunista y mitigando los conflictos sociales. En contraste, en grandes ciudades, donde predominan la movilidad y el anonimato, el comportamiento oportunista tiende a ser más prevalente. La falta de vínculos sociales sólidos y el anonimato que caracteriza los entornos urbanos pueden favorecer a los delincuentes, al protegerlos del estigma social ([Fajnzylber, Lederman, & Loayza, 2001](#)). Además, esta dimensión considera características culturales e ideológicas de la población, las cuales son fundamentales en la configuración de una identidad compartida. Estas características adquieren mayor relevancia al promover valores comunes que fortalecen la solidaridad y la confianza entre los miembros de la comunidad ([Alvira & Rubio, 1982](#)). Las variables asociadas a esta dimensión son: ideología política, práctica religiosa, confianza interpersonal y participación en organizaciones.

En relación con el [punto \(b\)](#), la Tabla 2 presenta las principales fuentes de información consultadas como potenciales insumos para la implementación SAE, y algunos ejemplos de covariables que contienen. Estas fuentes incluyen el Censo de Población y Vivienda 2024, proyecciones de población, registros administrativos provenientes de instituciones públicas e imágenes satelitales. Cada una de estas fuentes está asociada a alguna de las dimensiones expuesta anteriormente y en su mayoría corresponden al periodo 2024, lo que implica un resguardo del criterio de inclusión de información relevante y actualizada, buscando así enriquecer el modelo de estimación tanto en lo temático como en la eficiencia estadística. Respecto al [punto \(c\)](#), se establece un marco de evaluación de la calidad de estas fuentes, considerando criterios como la **cobertura geográfica y temporal, la precisión en los registros y la consistencia en las definiciones operativas de las variables, además del nivel de actualización de los datos**. En esta línea, cada variable derivada de estas bases fue analizada, determinando si debía o no imputarse o si debía descartarse el uso de esta o incluso de toda la base de datos.

Tabla 2. Dimensiones, fuentes y covariables relacionadas con la victimización

Dimensión	Fuente	Ejemplos de registro(s) que contiene
Sociodemográfica	Registro Social de Hogares	Personas carentes en tramo 0% - 40%
	RR.AA. del Ministerio de Desarrollo Social y Familia	Personas adscritas a Fonasa en tramo 0% - 40%
	Proyecciones de población	Mujeres extranjeras en tramo 0% - 40%
	CENSO 2024	Tasa de matrícula de personas entre 0 y 24 años
		Sexo
		Tramos etarios
		Personas sin escolaridad
		Viviendas inadecuadas según su materialidad
		Personas ocupadas
Socioeconómica	Registro Social de Hogares /CENSO 2024	Total de hogares/viviendas hacinadas(os)
	RR.AA. del Ministerio de Desarrollo Social y Familia	Promedio ingreso imponible de asalariados formales dependientes
		Mediana del ingreso imponible de los asalariados dependientes
		Tasa de ocupados formales dependientes
Infraestructura y del entorno	Carabineros de Chile	Infraestructura (comisarías, retenes, puestos de control, garitas comunitarias, subcomisarías y tenencias)
	Imágenes satelitales	Barrios peligrosos
		Radiancia nocturna
		Índice de Urbanización
		Índice de Vegetación Urbana
	Precenso 2023-TECHO-Ministerio de Vivienda y Urbanismo de Chile	Total de campamentos
		Superficie campamentos
		Total drones
	Sistema nacional de información municipal	Total casetas de seguridad
		Total cámaras de vigilancia
Crimen y delincuencia	Carabineros de Chile	Total Víctimas por tipo de delito
		Total Casos por tipo de delito
		Central de Comunicaciones de Carabineros de Chile (CENCO)
Culturales y cohesión social	Servicio Electoral	Total de votos plebiscito

Es importante mencionar que no todas las fuentes de información listadas en la Tabla 2 son finalmente usadas. Si bien fueron consideradas en un principio, se descarta su uso, ya que, por ejemplo, no cuentan con metadatos asociados, cobertura para las 136 comunas de la muestra

ENUSC 2024, las covariables que componen dichas fuentes en la fase exploratoria no evidencian relación con el indicador a modelar o no constituyen, junto con otras covariables, modelos multivariados aptos para generar estimaciones SAE. En el [Anexo 1](#), se listan todas las variables auxiliares consideradas para la implementación de modelos SAE. A continuación, se describen las fuentes de información que fueron consideradas durante la etapa de especificación de los modelos, aunque como se mencionó antes, algunas de ellas no son usadas en el modelo finalmente seleccionado.

4.2.1 Registro Social de Hogares

El Registro Social de Hogares (RSH) es un sistema que recopila información proporcionada por los propios hogares y las instituciones del Estado, con el objetivo de identificar a las personas beneficiarias de distintos programas sociales y ayudas económicas. A partir de esta información, se calcula una Calificación Socioeconómica (CSE) que clasifica a cada hogar en un tramo específico. Para determinar los tramos, se consideran diversos factores como ingresos, composición familiar, condiciones de vivienda, nivel educativo y estado de salud, entre otros. La información disponible para uso público¹⁰ está alojada en diferentes fuentes: indicadores CSE personas, indicadores CSE hogares, indicadores CSE viviendas, indicadores CSE educación, indicadores CSE salud, indicadores CSE trabajo. Estos datos poseen un dominio temporal que incluye periodicidades mensuales, semestrales y anuales, abarcando principalmente el periodo comprendido entre 2023 y 2024 (fueron descargados durante abril del 2025). Por tanto, dadas las diferentes estructuras entre las bases de datos (temporalidad y unidad de información) este RR.AA. requirió un esfuerzo importante de adaptación, y así mismo, fue necesario realizar la imputación de algunos registros¹¹,

¹⁰ <https://bidat.gob.cl/directorio/Registro%20Social%20de%20Hogares>

¹¹ El trabajo de imputación consistió en: registros codificados como “Valores entre 1 y 9” se imputa a 4; para las agregaciones se decidió omitir valores NA.

dada la identificación de datos faltantes o información inconsistente (en ningún caso superior al 10% de la totalidad del registro) antes de ser habilitado para su uso en los modelos SAE y, en otros casos, dado el nivel de inconsistencia se decidió prescindir del mismo.

4.2.2 Registros Administrativos del Ministerio de Desarrollo Social y Familia

La Subsecretaría de Evaluación Social, a través de los registros administrativos del Ministerio de Desarrollo Social y Familia (MDSF), implementa un sistema de indicadores con registros administrativos¹² para caracterizar el contexto socioeconómico del país. Este sistema permite monitorear con mayor frecuencia y detalle áreas clave como empleo, ingresos, salud y educación, ofreciendo un diagnóstico más preciso y eficiente, especialmente en poblaciones y territorios que no son fácilmente captados por encuestas tradicionales. Los datos más actualizados de este registro corresponden principalmente a los períodos 2023-2024, los cuales fueron descargados durante abril del 2025. Aunque no se detectan inconsistencias, se aplica un filtro comunal y se identifican datos faltantes en al menos una comuna de la muestra ENUSC 2024. La base de datos de ingresos incluye una serie de indicadores (2023) clave que permiten analizar el comportamiento del mercado laboral formal en términos de ingresos imponibles para los trabajadores dependientes. Estas variables permiten evaluar el ingreso promedio, la distribución del ingreso y su relación con el ingreso mínimo, así como la proporción de trabajadores con ingresos por debajo del 50% de la mediana. Además, se incluyen medidas agregadas que permiten entender el ingreso medio de la población formalmente empleada en relación con la población en edad de trabajar.

Las variables relacionadas con educación incluyen la tasa de matrícula de la población de 0 a 24 años, definida como el porcentaje de personas de ese grupo etario matriculadas en establecimientos educacionales. Esta tasa considera la educación regular, especial y de adultos, y corresponde al año 2023. Cabe señalar que se prescindió de algunos registros dado que por falta de información no fue posible imputar los valores faltantes con los datos de años anteriores.

¹² <https://bidat.gob.cl/directorio/Indicadores%20de%20registros%20administrativos/indicadores-con-registros-administrativos-2024>

4.2.3 Sistema Nacional de Información Municipal

El Sistema Nacional de Información Municipal (SINIM), administrado por la Subsecretaría de Desarrollo Regional y Administrativo (SUBDERE), presenta un compendio de información referida al quehacer de las municipalidades, correspondiente a diferentes períodos, siendo de particular interés el año 2024 a nivel comunal, regional y nacional. El contenido presentado, responde a información remitida por las propias municipalidades, en los sucesivos procesos de captura de datos anuales realizados a través del portal www.sinim.gov.cl, de otros organismos públicos y también información de su elaboración propia. Se consulta la información disponible en el SINIM, específicamente en el área Social y Comunitaria, sub-área Prevención del Delito y Seguridad¹³, obteniendo, por ejemplo, información sobre número de cámaras de vigilancia en la comuna o número de camionetas de seguridad municipal o patrullaje. Dado que algunas variables de interés presentan datos faltantes, en caso de contar con información para el año 2023, se imputa ese valor; en otro caso, se utiliza el dato del año más cercano con información disponible.

4.2.4 Imágenes satelitales – Radiancia nocturna

La Radiancia nocturna proporciona una medición indirecta de la actividad humana mediante la detección de la emisión de luz artificial, lo que permite identificar patrones de urbanización, densidad poblacional y posibles correlaciones con la incidencia delictiva. La base de radiancia nocturna¹⁴ utilizada en el análisis proviene de datos públicos disponibles en Google Earth Engine (GEE), proporcionando mediciones satelitales que capturan la emisión de luz artificial durante la noche, permitiendo una aproximación a la actividad humana en diferentes escalas geográficas. Cada fila de la base representa la media de los datos descargados en una celda de 52x52 metros, basada en la grilla denominada rhealpix, la cual permite una cobertura uniforme del territorio, para este estudio se descargó información del mes de octubre del 2024. Para cada comuna, pueden

13 https://datos.sinim.gov.cl/informacion_municipal.php

14 https://developers.google.com/earth-engine/datasets/catalog/NOAA_DMSP-OLS_CALIBRATED_LIGHTS_V4?hl=es-419

existir múltiples valores, ya que el satélite proporciona varias observaciones dentro de una misma unidad geográfica. Dado que la precisión de los datos puede verse afectada por la nubosidad en algunas observaciones, se construye una variable de radiancia ajustada por calidad. Para ello, se aplica un promedio ponderado, donde la radiancia promedio de cada celda se ajusta por su indicador de calidad, el cual refleja la cantidad de observaciones libres de nubosidad. De este modo, las celdas con mejores condiciones de observación tienen mayor influencia en el valor final, mientras que aquellas con alta nubosidad aportan menos al promedio. El resultado es un único valor de radiancia nocturna ajustada por calidad para cada comuna.

4.2.5 Registros Administrativos de Carabineros de Chile

4.2.5.1 Base de Casos Policiales

Carabineros de Chile (CCh) y el INE cuentan con un convenio de transferencia de información vigente desde 2006, el cual incluye registros administrativos de Casos Policiales, definidos como hechos delictuales conocidos por la policía, conformado tanto por las denuncias, como por las detenciones en flagrancia. Las denuncias son los reportes de delitos que realizan las personas en las unidades policiales, y las detenciones son hechos conocidos por las policías mediante un hecho flagrante, es decir, durante la comisión del delito y hasta 12 horas después. Ambos hechos policiales tratan de un trámite, por lo tanto, la unidad de observación de la base de datos son hechos policiales. **En cuanto a la temporalidad de los datos, la base de datos que se conforma contiene todos los casos policiales registrados entre el 01 de enero al 31 de diciembre del 2024. La cobertura geográfica es nacional, regional y comunal.** En el marco del aseguramiento de calidad dentro de las Estadísticas de Seguridad y Justicia, la base de casos policiales cuenta con un proceso de revisión integral que incluye la verificación de completitud y la validación de cifras mediante procedimientos manuales y automatizados a través del software R. Esta base no presenta datos faltantes y, si un delito vigente no está registrado, se debe a la ausencia de denuncias formales o detenciones en el periodo analizado, lo que implica que no se ha generado una frecuencia para dicho delito.

Es importante mencionar que, si bien esta base de datos fue analizada y explorada, se decide usar para la etapa de especificación de los modelos SAE, la base de víctimas que se describe más

abajo, dado que ambas bases presentan bondades similares, pero la última tiene mayor consistencia con la medición de la ENUSC.

4.2.5.2 Base de Víctimas

Como se mencionó previamente, el INE y CCh mantienen un convenio de cooperación vigente, y a partir de la solicitud anual de 2022, **se incorporó la categoría de víctimas**, lo que permitió disponer de información geográfica, temporal, de codificación penal y socioeconómica sobre las víctimas de delitos, estableciendo esta como la unidad de observación. **Estos datos abarcan el período comprendido entre el 1 de enero y el 31 de diciembre de cada año, registrando un total de 1.330.348 víctimas para 2024.** La cobertura geográfica incluye los niveles nacional, regional y comunal, lo que facilita su vinculación con la información de la ENUSC en el contexto del modelamiento SAE. Al igual que la base de casos policiales, esta base no presenta datos faltantes y, si un delito vigente no está registrado, ello se debe a la ausencia de denuncias formales en el periodo analizado, lo que implica que no se ha generado una frecuencia para dicho delito.

Desde el preprocesamiento y en el ámbito temático, se establece la misma correspondencia entre los delitos registrados en la ENUSC y la codificación nacional penal, esto permite identificar a las víctimas de delitos cuya tipificación es coincidente con la empleada en ENUSC 2024. Dado que el delito constituye el eje central del registro, esta correspondencia se aplica tanto a la base de casos policiales como a la de víctimas. El Anexo [1](#) indica todas las variables consideradas provenientes de CCh, incluyendo las que pertenecen a las bases de datos que se describen más abajo.

4.2.5.3 Base de Infraestructura física de Carabineros

En el marco del convenio¹⁵ entre el INE y CCh, se dispone de información geográfica que incluye la ubicación de comisarías y sus áreas de acción (representadas en capas de puntos y polígonos). Esto permite contar con datos detallados sobre la ubicación de comisarías, retenes, puestos de control, garitas comunitarias, subcomisarías y tenencias, junto con sus respectivos cuadrantes, para

¹⁵ Convenio de colaboración entre el INE y CCh, del 24 de agosto del 2023.

todas las comunas del país. Esta base fue actualizada al periodo entre mayo y julio del 2024. Para poder trabajar con esta información se decide agregar a nivel de comuna la información y contabilizar el número de unidades de carabineros. No se detectan inconsistencias ni tampoco información faltante para las 136 comunas pertenecientes a la muestra ENUSC 2024.

4.2.5.4 Base de Sectores Peligrosos: método multicriterio

Con el propósito de optimizar el despliegue de los recursos policiales y garantizar su disponibilidad de manera oportuna y efectiva, así como de contribuir al cumplimiento del Plan Nacional de Operaciones (PNO), la Dirección Nacional de Orden y Seguridad solicita al Departamento de Análisis Criminal (DAC) la implementación de una metodología enfocada en la correcta focalización de los sectores que requieren prioridad. Esta metodología concentra los esfuerzos policiales en áreas específicas para abordar problemáticas de alta complejidad, como el tráfico de drogas y los delitos con armas de fuego. A través de este enfoque, se identifican y priorizan sectores (definidos como unidades base de análisis) donde se registra una alta concentración de violencia asociada al uso de armas de fuego y al tráfico de droga¹⁶. La base de datos contiene información para la caracterización de 536 sectores a nivel nacional correspondiente al periodo entre mayo y julio del 2024. Para trabajar con esta información, se agrega a nivel comunal el número de sectores clasificados como “Crítico” y “Muy Crítico”. No se detectan inconsistencias ni tampoco información faltante para las 136 comunas pertenecientes al marco de selección ENUSC 2024.

4.2.6 Información estadística del INE

4.2.6.1 Censo de población y vivienda

Los censos de población y vivienda constituyen uno de los pilares para la recopilación de información estadística sobre la cantidad, características y condiciones de vida de la población de

¹⁶ Para mayor detalle puede consultar <https://pazciudadana.cl/biblioteca/analisis-criminal/metodo-de-analisis-multicriterio-para-determinar-sectores-de-violencia-asociadas-al-uso-de-armas-de-fuego-y-drogas/>

un país, y permiten responder las preguntas “¿cuántos somos?” y “¿dónde y cómo vivimos?”. Este proceso es de alta complejidad debido, entre otras cosas, a la planeación y diseño, al financiamiento, al despliegue de los diferentes recursos disponibles, a los operativos de campo y demás; se considera que un censo es la operación estadística más grande e importante que puede desarrollar un país, ya que se debe alcanzar una cobertura del 100%, es decir, debe llegar a cada hogar del territorio nacional.

Si bien el operativo censal se enfoca en comprender los cambios poblacionales, territoriales y el crecimiento demográfico del país, su realización es crucial para elaborar el marco muestral de futuras encuestas que profundicen en diversas materias socioeconómicas. Es decir, es la base de la construcción estadística, ya que constituye el eje estructural para mediciones de carácter coyuntural. La información de un censo es un insumo fundamental para actualizar las necesidades de la población, planificar acciones y tomar decisiones en beneficio de la ciudadanía. Las definiciones sobre cómo usar los recursos, tanto públicos como privados, para mejorar la vida de las personas en el país se toman teniendo en cuenta la información de quienes habitan en el territorio. Además, una de las características fundamentales de un censo radica en que es capaz de producir datos detallados de localidades grandes y pequeñas, así como de distintos grupos de la población, siendo un punto de partida para la toma de decisiones de los próximos años ([INE, 2025d](#)).

El Censo de Población y Vivienda 2024 (CPV 2024) se desarrolló entre marzo y julio del 2024 y presentó dos innovaciones centrales. Una de ellas fue la implementación de la metodología de derecho, la cual se centró en recolectar información basada en el lugar de residencia habitual de las personas, permitiendo obtener una imagen más precisa de la distribución poblacional. La segunda innovación fue la introducción de tecnologías a la operación de manera global, es decir, tanto a la recolección de los datos como al seguimiento en línea, lo que resultó fundamental para lograr la oportuna y correcta gestión del operativo durante el periodo de recolección. Inició su preparación en 2019, y culmina con la diseminación y entrega de resultados en 2025. Durante este período, el proyecto realizó la detección de necesidades de información, el diseño y construcción de instrumentos, la implementación de las herramientas y plataformas tecnológicas, el desarrollo y ejecución del plan de pruebas (tres de integración de sistemas y cinco pilotos), la Actualización Precensal, la realización del operativo a nivel nacional en 2024 y, finalmente, el procesamiento y

difusión de los datos ([INE, 2025d](#)). En este punto es importante mencionar que, al momento de realizar este trabajo, se dispone de información sociodemográfica a nivel comunal, sin contar con la apertura urbano-rural; y que los totales poblacionales a nivel comunal, regional y nacional corresponden a la población efectivamente censada y no a la estimación actualizada de la población que reside en Chile.

4.2.6.2 Proyecciones de población e información sobre campamentos

Las proyecciones de población a nivel comunal proporcionan una referencia del futuro tamaño y estructura por edades de una población, basados en un conjunto de supuestos sobre el comportamiento de los componentes demográficos: fecundidad, mortalidad y migración. Así, para diferentes desagregaciones de interés (hombre, mujer, tramo etario), se explora esta fuente de información para el año 2024¹⁷. En este punto, es importante señalar que también son usadas para relativizar por el tamaño comunal de la mayoría de las variables, por ejemplo, las variables provenientes de la base de víctimas de CCh, que se refieren al total de víctimas en algún delito específico, fueron transformadas a tasas, dividiendo en cada caso por el total proyectado a nivel comunal para el año 2024.

Finalmente, otro insumo utilizado proviene de la actualización cartográfica requerida para la planificación e incorporación de los campamentos en el Censo 2024, con lo que se obtiene información sobre la superficie y la cantidad, para todas las comunas del país donde fueron identificados¹⁸.

¹⁷ No se usa la información proveniente del Censo de Población y Vivienda 2024, dado que, al momento de elaboración de este trabajo, la población oficial en Chile es la remitida por las proyecciones de población.

¹⁸ Este insumo también se nutre de información proveniente del catastro de campamentos del Ministerio de Vivienda y Urbanismo, y del catastro de TECHO-Chile.

5. ANÁLISIS Y ADAPTACIÓN

La segunda etapa en SAE involucra el análisis de los datos y la adaptación de los modelos. Como se explicó anteriormente, el proceso debe regirse por el principio de parsimonia. En la [sección 5.1](#) se establecen definiciones para el estimador directo y su error estándar, así como puntos clave sobre el tratamiento de la varianza y los criterios para inclusión o exclusión de las comunas pertenecientes a la muestra. En las secciones EBLUP basado en el modelo Fay–Herriot ([sección 5.2](#)), transformación arcoseno y selección de covariables ([sección 5.3 y 5.4](#)), se ilustran las actividades centrales de análisis y adaptación, incluidos los temas relevantes de la construcción del modelo y sus adaptaciones finales mediante ajustes.

5.1. ESTIMADOR DIRECTO

El estimador directo, que se denota por $\hat{\theta}_d^{DIR}$, usa solamente los datos del área d , por lo que solo está disponible para un área en la muestra. Es la primera estimación cuando no hay datos auxiliares disponibles y la estimación se basa directamente en los factores de expansión del diseño muestral.

Considere 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 lo fue. Asumamos que la población está dividida en pequeñas áreas denotadas por $d = 1, \dots, D$. En nuestro caso, estas áreas corresponden a las zonas urbanas de 136 comunas de Chile en la ENUSC 2024. Sea s una muestra extraída al azar de la población de acuerdo con un diseño muestral dado. La población y la muestra son denotadas por $U = \bigcup_{d=1}^D U_d$ y $s = \bigcup_{d=1}^D s_d$, respectivamente. Sean N_d y n_d los tamaños de U_d y s_d , respectivamente, con $d = 1, \dots, D$. El objetivo es estimar la tasa de victimización para cada indicador, $\theta_d = N^{-1} \sum_{j \in U_d} y_j$, con su respectivo error estándar. El estimador directo de θ_d propuesto por [Hájek \(1971\)](#), y el correspondiente error estándar estimado son,

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

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

donde $\widehat{N}_d = \sum_{j \in s_d} w_j$ 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 2024 se obtuvieron a partir de un diseño de muestreo complejo en tres etapas: (1) unidad primaria de muestreo, (2) viviendas ocupadas en áreas urbanas, y (3) persona (informante Kish)¹⁹.

Los estimadores directos son aproximadamente insesgados para $\theta_d, d = 1, \dots, D$, pero no son precisos para la estimación de parámetros en áreas pequeñas. Debido al pequeño tamaño de la muestra, dan lugar a errores cuadráticos medios (MSE, por sus siglas en inglés) altos y coeficientes de variación considerablemente altos (por ejemplo, mayores al 30%)²⁰. En consecuencia, dado los tamaños muestrales pequeños que tenemos para las comunas y los coeficientes de variación altos, la estimación de las tasas de victimización comunal es un problema de estimación de área pequeña.

Como se señaló anteriormente, los estimadores de áreas pequeñas pueden basarse en modelos a nivel de unidad o de área. En la práctica, no existe evidencia de que un enfoque provea los mejores resultados, y su aplicación dependerá del nivel de desagregación de la información externa disponible. Por lo tanto, ante la posibilidad de implementar ambos, la recomendación es hacerlo de las dos maneras y comparar resultados. En este caso, dado el nivel de detalle de la información externa disponible, la aplicación de un modelo a nivel de áreas es la única factible.

Bajo el enfoque de modelo de áreas, se ajusta un modelo a los datos agregados $(\widehat{\theta}_d^{DIR}, x_d)$, $d = 1, \dots, D$, donde x_d es un vector que contiene los valores de p-variables auxiliares. El modelo básico de áreas es el propuesto por [Fay y Herriot \(1979\)](#), que se describe a continuación.

¹⁹ Para más detalle revisar https://www.ine.gob.cl/docs/default-source/seguridad-ciudadana/metodologia/2024/informe-metodologico---enusc-2024.pdf?sfvrsn=651010da_2

²⁰ Ver Särndal et al. (1992) para obtener más detalles sobre la inferencia en la población finita y las propiedades de los estimadores directos.

5.1.1. Suavizamiento de la varianza directa

El modelo Fay–Herriot asume que las varianzas muestrales son conocidas en el modelo. En la práctica, sin embargo, las variaciones muestrales son desconocidas y se estiman directamente a partir de los datos de la encuesta. Siendo este el caso, al igual que las estimaciones directas del parámetro de interés, las estimaciones directas de las varianzas muestrales también están sujetas a errores. De hecho, las estimaciones directas $\hat{\theta}_d$ son muy imprecisas debido a los tamaños muestrales deficientes para algunas o todas las áreas, y por tanto, también se espera que las estimaciones de la varianza muestral, $\tilde{\psi}_d$, sean muy imprecisas debido a las mismas limitaciones del tamaño de la muestra (Bell, 2008). Una de las soluciones para este problema es modelar $\tilde{\psi}_d$ para obtener estimaciones mejoradas de ψ_d y cuantificar su error de estimación. Modelar las varianzas muestrales estimadas puede resultar en medidas mejoradas del MSE. Además, Wolter (1985, p.203) introdujo el método de función de varianza generalizada (FVG), que se usa comúnmente para obtener estimaciones suavizadas de las varianzas muestrales. Es un modelo matemático que describe la relación entre la varianza o la varianza relativa de un estimador de encuestas y su esperanza. Adicionalmente, Valliant (1987) destacó que la principal ventaja de este método consiste en obtener estimaciones rápidas de las varianzas muestrales. Así, estas estimaciones son más estables y precisas que las directas, eliminando el problema de volatilidad que puede existir a nivel comunal.

Una forma de obtener un estimador eficiente es modelando $\tilde{\psi}_d$ a través de un modelo log-lineal como sigue

$$\log(\tilde{\psi}_d) = \mathbf{x}_d^T \boldsymbol{\alpha} + \varepsilon_d, d = 1, \dots, D, \quad (1)$$

donde $\varepsilon_d \sim N(0, \sigma_\varepsilon^2)$ independientes, $d = 1, \dots, D$; \mathbf{x}_d^T es un vector de variables explicativas, $\boldsymbol{\alpha}$ es un vector de parámetros del modelo que deben ser estimados.

El procedimiento usado para la estimación de la varianza a partir del modelo en (1) es el siguiente:

1. Se descartan las comunas con estimaciones directas y varianzas igual a 0 y las comunas que presentan un *DEFF* menor a 1. Lo que implica hacer el modelamiento de la varianza con 68 comunas.

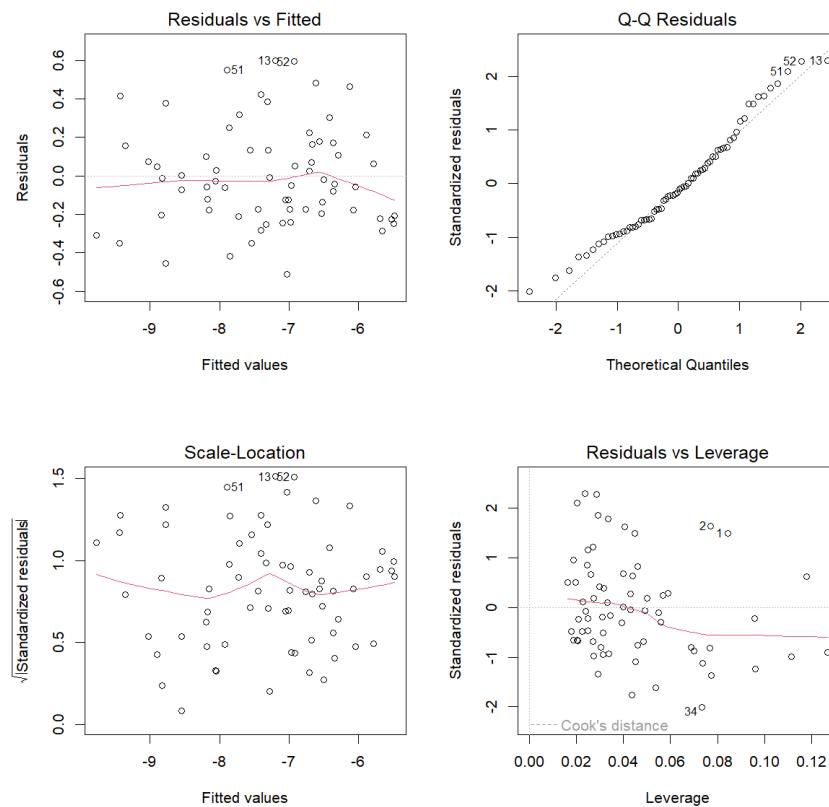
2. Se realiza un análisis exploratorio para determinar cuáles covariables guardan mejor relación con el logaritmo de la varianza estimada. Adicionalmente, se exploran transformaciones para conseguir mejores ajustes, en particular, la transformación logaritmo natural arroja buenos resultados. Finalmente, se resguarda que el modelo cumpla con los supuestos necesarios para realizar inferencia, a saber, homocedasticidad, normalidad y no presencia de valores influyentes. En consecuencia, el modelo de FVG para VHDV está compuesto por las siguientes variables:
 - Logaritmo natural de VHDV.²¹
 - Logaritmo natural del tamaño muestral.²²

La Figura 2, muestra que no hay evidencia de no normalidad, heterocedasticidad o comunas influyentes. Además el modelo elegido presenta un buen ajuste ($R^2_{ajustado} = 94.2\%$).

²¹ Utiliza función log1p del Software R.

²² Utiliza función log1p del Software R.

Figura 2. Análisis de calidad de ajuste de FVG para VHDV



A fin de obtener estimaciones para ψ_d , los valores predichos, $\hat{\psi}_d$, del modelo en (1) se obtienen empleando la siguiente expresión

$$\hat{\psi}_d = \exp(x_d^T \hat{\alpha}) \times \hat{\Delta}(\hat{\alpha}), d = 1, \dots, D, \quad (2)$$

donde $\hat{\Delta}(\hat{\alpha}) = \frac{\sum_{d=1}^D \tilde{\psi}_d}{\sum_{d=1}^D \exp(x_d^T \hat{\alpha})}$ es un estimador insesgado de un término de corrección de sesgo propuesto por [Hidiroglou et al. \(2019\)](#) [$\Delta = E(\exp(\varepsilon_d))$] y $\hat{\alpha}$ es el estimador de α obtenido mediante el método de mínimos cuadrados, con $\hat{\alpha} = (\sum_{d=1}^D x_d x_d^T)^{-1} \sum_{d=1}^D x_d \log(\hat{\psi}_d)$.

[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:

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

Independientemente del procedimiento de estimación que se utilice, existe la posibilidad de obtener estimaciones de varianza negativa. Así que siguiendo la sugerencia de [Wolter \(2007\)](#) se han introducido restricciones sobre el intercepto para evitar esta situación. Finalmente, las estimaciones $\hat{\hat{\psi}}_d$ son conectadas en lugar de ψ_d en el modelo Fay Herriot ([Sección 5.2](#)) para generar las estimaciones EBLUP y las respectivas estimaciones del error cuadrático medio.

El modelo de varianza que se genera es también utilizado como insumo para poder predecir el *DEFF* en las comunas donde es menor a 1. Esto permite pasar a la etapa siguiente, de establecer un flujo de criterios de inclusión y exclusión de comunas, previo a iniciar la implementación de los modelos Fay-Herriot.

5.1.2. Criterios de calidad para la inclusión o exclusión de comunas en el modelo

Se establece un flujo de calidad para las estimaciones directas del indicador de interés con el fin de incluir comunas con un muestreo de calidad, que no vayan a inducir a sesgos en las estimaciones SAE comunales. Esto, siguiendo el trabajo realizado por el MDSF y la Comisión Económica para América Latina y el Caribe (CEPAL), en el marco de la implementación de la metodología SAE para las estimaciones comunales de pobreza en Chile, señalando que “es importante destacar que para el ejercicio SAE, el objetivo no es incluir comunas precisas, ya que el mismo modelo Fay-Herriot incluye una ponderación por precisión, sino más bien, el objetivo es incluir comunas con un muestreo de calidad que indique que los insumos que entran al modelo (precisos o no precisos) son confiables” [\(División Observatorio Social MDSF – CEPAL, 2020\)](#).

El flujo se presenta en la Figura 3. Primero se verifican las comunas con 14 o más grados de libertad (*gl*), en cuyo caso se incluyen directamente; si tienen menos de 14 *gl*, pasan a la siguiente fase donde se verifica: tamaño de muestra logrado (*n*) < 50, *DEFF* < 1, casos no ponderados (Y_{np}) ≤ 2 , $gl \leq 2$. Si la comuna cumple con alguno de estos criterios es excluida. La Tabla 3 muestra la cantidad de comunas incluidas o excluidas al aplicar de forma secuencial los criterios.

Así, bajo este conjunto de criterios de calidad, son excluidas 21 comunas: 1 en Tarapacá; 1 en Atacama; 1 en Coquimbo; 1 en Valparaíso; 1 en Metropolitana; 2 en O'Higgins; 1 en Maule; 5 en

Ñuble; 1 en Biobío; 4 en La Araucanía; y 3 en Los Lagos. Esto implica que, **para las comunas excluidas, la estimación SAE solo tendrá el componente sintético asociado al modelo derivado de las fuentes de información utilizadas en cada caso**. En los cuadros estadísticos donde se presentan los resultados de las estimaciones, se indica si la estimación SAE está compuesta por estimación directa y sintética, es decir la comuna aprobó los criterios de calidad fijados; o si solamente considera la parte del modelo, por tanto, la comuna no aprobó los criterios.

Figura 3. Flujo de calidad para inclusión de estimaciones directas en VHDV

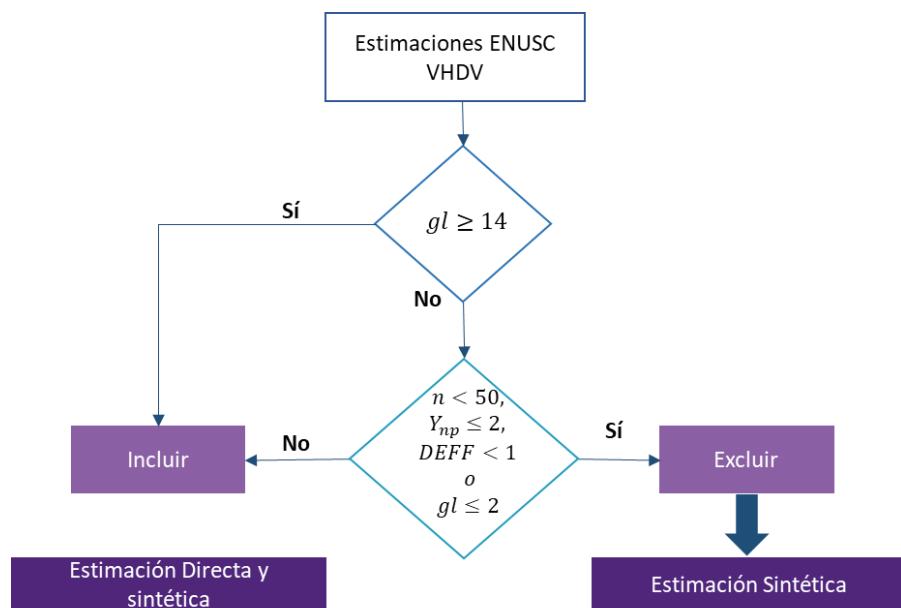


Tabla 3. Resumen de criterios de inclusión y exclusión

	Inclusión	Exclusión					Total incluir	Total excluir
		grados de libertad	tamaño muestral	efecto de diseño	conteo casos no ponderados	grados de libertad		
	$gl \geq 14$		$n < 50$	$deff < 1$	$Y_{np} \leq 2$	$gl < 2$		
Nº Comunas	46		6	0	18	0	115	21

5.2. EBLUP BASADO EN EL MODELO FAY–HERRIOT

La metodología SAE es uno de los tópicos en el muestreo de encuestas donde el uso de enfoques basados en modelos es ampliamente aceptado como necesario. Estos métodos asumen un modelo para la población y los datos de la muestra, y construyen predictores óptimos de los parámetros objetivo. Aquí describimos el modelo de Fay–Herriot que se usará para estimar la tasa de victimización a hogares por delitos violentos. Un modelo de nivel de área básico, conocido como Fay-Herriot ([Fay y Herriot, 1979](#)) consta de dos componentes. Sea $\hat{\theta}_d^{DIR}$ un estimador directo de θ_d . En el primer nivel, se asume que, dado θ_d , $\hat{\theta}_d^{DIR}$ es un estimador insesgado de θ_d , más concretamente,

$$\hat{\theta}_d^{DIR} = \theta_d + e_d, e_d \stackrel{ind}{\sim} N(0, \psi_d), d = 1, \dots, D, \quad (3)$$

donde ψ_d es la varianza muestral del estimador directo $\hat{\theta}_d^{DIR}$ dado θ_d que se asume conocida para todo $d = 1, \dots, D$.

En un segundo nivel, se asume que los parámetros del área θ_d están linealmente relacionados con un vector p -dimensional x_d de variables auxiliares a nivel de áreas como sigue,

$$\theta_d = x_d^T \beta + u_d, u_d \stackrel{iid}{\sim} N(0, A), d = 1, \dots, D. \quad (4)$$

El [modelo \(4\)](#) se llama modelo de vinculación porque relaciona todas las áreas mediante los coeficientes de regresión comunes β , lo que permite aprovechar la fuerza de todas las áreas. El [modelo \(3\)](#) se llama modelo de muestreo porque representa la incertidumbre debido al hecho que θ_d es no observable y, en lugar de θ_d , se observa su estimador directo basado en la muestra, $\hat{\theta}_d^{DIR}$.

Combinando los dos componentes, el modelo de Fay-Herriot puede verse como un modelo de regresión mixto a nivel de área como sigue:

$$\hat{\theta}_d^{DIR} = x_d^T \beta + u_d + e_d, e_d \stackrel{ind}{\sim} N(0, \psi_d), u_d \stackrel{iid}{\sim} N(0, A), d = 1, \dots, D, \quad (5)$$

Donde u_d es independiente de e_d para todo d^{23} . La normalidad no es necesaria para la estimación puntual, pero es necesaria para la estimación del MSE.

[Henderson \(1975\)](#) obtuvo el mejor predictor lineal insesgado (BLUP, por sus siglas en inglés) de un efecto mixto, es decir, una combinación lineal de los efectos fijos y aleatorios β y $u = (u_1, \dots, u_D)^T$. El BLUP de θ_d bajo el modelo Fay-Herriot en [\(3\)](#) es dado por:

$$\tilde{\theta}_d^{BLUP} = x_d^T \tilde{\beta}(A) + \tilde{u}_d(A), \quad (6)$$

donde $\tilde{u}_d(A) = \gamma_d(A) \left(\hat{\theta}_d^{DIR} - x_d^T \tilde{\beta}(A) \right)$ es el efecto aleatorio predicho, $\gamma_d(A) = A/(A + \psi_d) \in (0, 1)$ y $\tilde{\beta}(A) = [\sum_{d=1}^D (A + \psi_d)^{-1} x_d x_d^T]^{-1} \sum_{d=1}^D (A + \psi_d)^{-1} x_d \hat{\theta}_d^{DIR}$ es el estimador de β por mínimos cuadrados ponderados.

El BLUP asume que A es conocida. El EBLUP $\hat{\theta}_d^{EBLUP}$ se obtiene reemplazando A en el BLUP [\(4\)](#) por un estimador consistente \hat{A} . El EBLUP puede ser expresado como una combinación convexa entre el estimador directo y el estimador sintético de regresión como sigue,

$$\hat{\theta}_d^{EBLUP} = \begin{cases} \hat{\gamma}_d \hat{\theta}_d^{DIR} + (1 - \hat{\gamma}_d) x_d^T \hat{\beta}, & \text{si } d \in s \\ x_d^T \hat{\beta}, & \text{si } d \notin s \end{cases} \quad (7)$$

²³ En nuestro caso d hace referencia a la comuna d y D al total de comunas en la muestra ENUSC 2024, es decir, 136.

Donde $\hat{\gamma}_d = \gamma_d(\hat{A}) = \hat{A}/(\hat{A} + \psi_d) \in (0,1)$ y $\hat{\beta} = \tilde{\beta}(\hat{A})$. En (7) podemos ver que, para $d \in s$, cuando el estimador directo es confiable, es decir, ψ_d es pequeño comparado con A , entonces el EBLUP se acerca al estimador directo. **En contraste, cuando el estimador directo no es confiable, es decir, ψ_d es grande comparado con A , entonces el EBLUP se acerca al estimador sintético de regresión.** Por lo tanto, el EBLUP utiliza el supuesto de regresión solo para las áreas donde se necesita prestar fuerza. Además, podemos ver que para $d \notin s$, es decir una comuna que no pertenece a la muestra o que ha sido excluida, según los criterios de calidad que mencionamos en la sección 5.1.2, la estimación solo depende del estimador sintético de regresión.

Los métodos comunes de ajuste de modelos que ofrecen estimadores consistentes para A son el método de Fay-Herriot (F-H) (Fay y Herriot, 1979), máxima verosimilitud (ML) y ML restringido (REML), donde este último representa los grados de libertad debidos a la estimación de β y, por lo tanto, tiene un sesgo muestral finito reducido. Molina y Marhuenda (2015) expresan que si el estimador \hat{A} es una función par e invariante de traslación del vector de estimaciones directas, que se cumple para los métodos de ajuste F-H, ML y REML, entonces bajo distribuciones simétricas de efectos y errores aleatorios, el $\hat{\theta}_d^{EBLUP} = \tilde{\theta}_d^{BLUP}(\hat{A})$ permanece insesgado.

Por lo general, los modelos se comparan en función de medidas de bondad de ajuste, como log-verosimilitud, el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC). Bajo el modelo F-H en (5), la log-verosimilitud viene dada por:

$$\ell(A, \beta) = -\frac{1}{2} \left[D \log(2\pi) + \sum_{d=1}^D \log(A + \psi_d) + \sum_{d=1}^D \log(A + \psi_d)^{-1} (\hat{\theta}_d^{DIR} - x_d^T \beta)^2 \right].$$

AIC y BIC se obtienen respectivamente como:

$$AIC = -2\ell(A, \beta) + 2(p + 1), \quad BIC = -2\ell(A, \beta) + 2(p + 1)\log(D).$$

En la sección 6.4, se observa que como medida de evaluación de la incertidumbre de los modelos se usa el MSE del EBLUP. Para este trabajo, las estimaciones SAE y sus respectivos MSE fueron estimadas a través de la librería **EMDI** (Estimating and Mapping Disaggregated Indicators), desarrollada por Kreutzmann et al. 2019, del software R, utilizando el método de Bootstrap paramétrico.

5.3. TRANSFORMACIÓN ARCOSENO

En aplicaciones de datos reales pueden ocurrir problemas que teóricamente no se esperaban o pueden violarse los supuestos del modelo F-H 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 de la raíz cuadrada, que se usa frecuentemente cuando el estimador directo del modelo F-H 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:

- a. En su concepción más básica, el modelo F-H 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 $\theta_d \in (0,1)$.
- b. El estimador de la varianza de la estimación directa para una proporción en un diseño complejo depende de la proporción estimada, por ende, habría problemas de endogeneidad²⁴ 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 $* \text{arcsin}$ 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 efecto de diseño ([Jiang et al. 2001](#)). El modelo F-H es estimado usando la ecuación (7) y debe verificarse que las

²⁴ La endogeneidad puede entenderse como un problema estadístico fundamental que ocurre cuando una variable explicativa (independiente) en un modelo de regresión está correlacionada con el término de error.

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 propuesta por Sugawara y Kubokawa (2017). En este análisis se utiliza la **transformación inversa *naive*** que consiste en lo siguiente:

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

5.4. SELECCIÓN DE COVARIABLES

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 de los modelos, su diagnóstico, el análisis de sensibilidad y la validación, ocupan un lugar central dentro de la metodología SAE. Si bien no existe un enfoque único para la construcción de modelos, se listan algunas pautas de buenas prácticas que hemos usado para la estimación de las tasas de victimización a hogares por delitos violentos de la ENUSC 2024. La construcción del modelo se basa en un proceso de selección de covariables que brinden el mayor poder predictivo, minimizando los efectos aleatorios con un espíritu de parsimonia. Un punto de partida, que se ha considerado adecuado, es realizar un análisis exploratorio de los datos que permita observar relaciones existentes en ellos, en particular, relaciones entre las covariables disponibles y el indicador de interés.

Una vez definido el set de covariables, se procede con una selección automática de variables mediante el algoritmo *stepwise*, utilizando la función *step* implementada en el paquete *EMDI*. Esta función permite explorar distintas combinaciones de covariables dentro del modelo Fay-Herriot ajustado, evaluando el ajuste de cada modelo a partir de criterios de información. Se parte del modelo ajustado con las especificaciones definidas previamente, aplicando el procedimiento de selección en ambas direcciones (hacia adelante y hacia atrás), lo que permite identificar un subconjunto de variables explicativas que optimiza el desempeño del modelo sin incorporar complejidad innecesaria.

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 para evaluar la bondad de ajuste del modelo (ver Tabla 4). 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. Por otra parte, **a pesar de que el procedimiento descrito provee de un proceso robusto para la selección de covariables, la construcción del modelo 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, siendo deseable que cubran la mayoría de dimensiones identificadas en el marco teórico y que sus coeficientes tengan el signo esperado.**

En el caso de la implementación de SAE para estimar la tasa de victimización a hogares por delitos violentos, estudios consultados como, por ejemplo, [Taton et al. \(2001\)](#), [Fay y Diallo \(2012\)](#) y [Li et al. \(2012\)](#), muestran el uso de covariables relacionadas con tasas de denuncias (en particular por delitos cometidos contra la propiedad), así como las relacionadas con grupos poblacionales específicos: población perteneciente a pueblos originarios, población en cierto tramo etario, población en determinada clasificación socioeconómica, etc. Así, para todos los modelos candidatos, se consideró necesaria la inclusión de covariables relacionadas con denuncias y víctimas de delitos, que guardan correspondencia con los indicadores que componen a VHDV. 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 dicotómicas (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 al ir en pos del “préstamo de fuerza” entre áreas que son similares, principio base en la estimación de áreas pequeñas. Estas variables dicotómicas permiten introducir un efecto fijo diferenciador a nivel regional para cada una de las comunas objetivos, bajo el supuesto de que en Chile la dinámica de las regiones del norte difiere de las regiones centrales, en particular de la Región 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. **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.**

A partir de los criterios estadísticos evaluados en el proceso y las consideraciones temáticas anteriormente descritas, el modelo finalmente incluye las variables:

- Proporción de personas que denuncian ser victimizadas por robo con intimidación.
- Mediana del ingreso de los asalariados dependientes.
- Proporción de viviendas en condiciones inadecuadas según la materialidad.
- Proporción de campamentos.

Tabla 4. Covariables en modelos SAE para VHDV

Variable	Coeficientes	Error estándar	Valor t	Valor p
(Intercepto)	0.4982	0.0779	6.3929	0.0000
Prop.RoboConIntimidación	0.1498	0.0482	3.1049	0.0019
MedingAsaDep	-0.1134	0.0455	-2.4897	0.0128
PropVivInadecuada	-0.1916	0.0636	-3.0119	0.0026
PropCampamentos	0.0114	0.0054	2.1032	0.0354
dummy Tarapacá	-0.0204	0.0294	-0.6928	0.4884
dummy Antofagasta	-0.1238	0.0339	-3.6556	0.0003
dummy Atacama	-0.1066	0.0344	-3.1008	0.0019
dummy Coquimbo	-0.1178	0.0424	-2.7751	0.0055
dummy Valparaíso	-0.1371	0.0457	-2.9991	0.0027
dummy Metropolitana	-0.1180	0.0457	-2.5844	0.0098
dummy O'Higgins	-0.1367	0.0454	-3.0092	0.0026
dummy Maule	-0.1646	0.0482	-3.4123	0.0006
dummy Ñuble	-0.1655	0.0502	-3.2973	0.0010
dummy Biobío	-0.1580	0.0474	-3.3318	0.0009
dummy La Araucanía	-0.1180	0.0485	-2.4314	0.0150
dummy Los Ríos	-0.1716	0.0475	-3.6143	0.0003
dummy Los Lago	-0.1565	0.0433	-3.6176	0.0003
dummy Aysén	-0.1634	0.0479	-3.4116	0.0006
dummy Magallanes	-0.1816	0.0485	-3.7415	0.0002
AIC	-359.96			
BIC	-302.31			
<i>R</i> ² <i>Hidiroglo</i> <i>lou</i>	84%			

Si bien tratar de cubrir todas las dimensiones temáticas y asegurar máxima calidad en las fuentes de información utilizadas es un aspecto relevante, el principal objetivo de la metodología SAE es conseguir la máxima precisión posible en las estimaciones de área pequeña, asegurando que el modelo implementado cumple los supuestos para la inferencia, y por tanto, es imperativo considerar los criterios de i) bondad de ajuste y significancia de las variables, ii) nivel de significancia para test de evaluación y diagnóstico; y iii) medidas de precisión tales como el CVL y RMSE, asociadas a la amplitud del intervalo de confianza. A continuación, se presentan diferentes medidas para la evaluación y diagnóstico del modelo implementado.

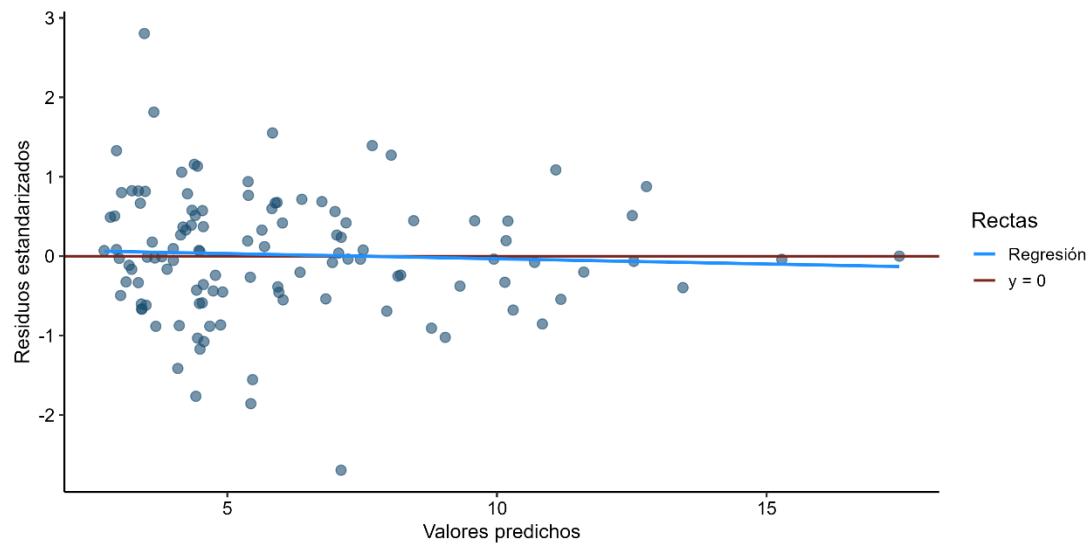
6. EVALUACIÓN

La evaluación de las propiedades del conjunto de estimaciones de áreas pequeñas, como el rango de las estimaciones, es un tema relevante cuya discusión detallada está más allá del alcance de este documento. Los niveles de precisión de las estimaciones SAE están supeditados a la confiabilidad del modelo de Fay-Herriot, por lo tanto, es esencial hacer una evaluación cuidadosa de la validez del modelo antes de disponibilizar las estimaciones. Por ejemplo, es importante verificar que se mantenga una relación lineal entre $\hat{\theta}_d$ y \mathbf{x}_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 $\mathbf{x}_d^T \hat{\beta}$.

$$\hat{a}_d = \frac{\hat{\theta}_d - \mathbf{x}_d^T \hat{\beta}}{\sqrt{\hat{A} + \hat{\psi}_d}}.$$

La Figura 4 muestra que la dispersión de los residuos no sigue un patrón específico, siendo esto un indicativo de una buena especificación de los modelos ajustados para el indicador.

Figura 4. Residuos estandarizados vs. valores predichos para VHDV



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 notar alejamiento de la normalidad leve, [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. Siguiendo la guía proporcionada por [Tzavidis et al. \(2018\)](#), a continuación, se describen algunos aspectos de la evaluación (método e incertidumbre), que deben tenerse en cuenta en cualquier aplicación SAE. Se presentan las gráficas q-q en la Figura 5 para corroborar la normalidad de los residuos estandarizados (en azul) y efectos aleatorios (en gris). Se observa que los residuos siguen un patrón normal con variaciones en las colas. En la Tabla 5 se presenta un resumen de los diferentes test estadísticos para corroborar normalidad (Shapiro-Wilk), homocedasticidad (Breusch-Pagan) y no autocorrelación (Durbin-Watson), se verifica que el modelo ajustado cumple con todos los supuestos requeridos.

Figura 5. Q-Q plot. Residuos Estandarizados Izq. - Efectos aleatorios Der.

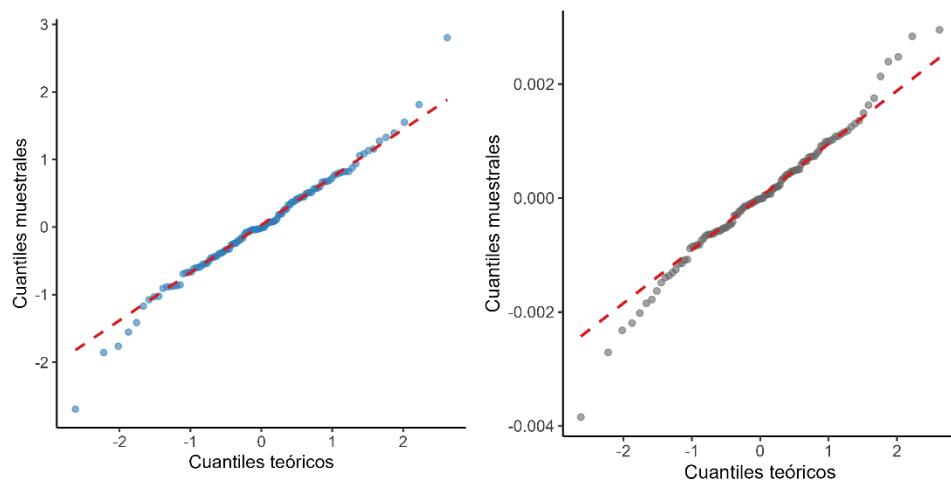


Tabla 5. Resumen de test de normalidad, homocedasticidad y no autocorrelación

	Test Shapiro-Wilk		Test Breusch-Pagan		Test Durbin-Watson	
	Efectos Aleatorios	Estandarizados	Estandarizados	Estandarizados	Estandarizados	Estandarizados
	VHDV					
Estadístico	0.982	0.982	0.449	1.826		
Valor <i>p</i>	0.134	0.133	0.503	0.173		

6.1. 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 $\hat{a}_d^2 \sim \chi^2_1$. 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

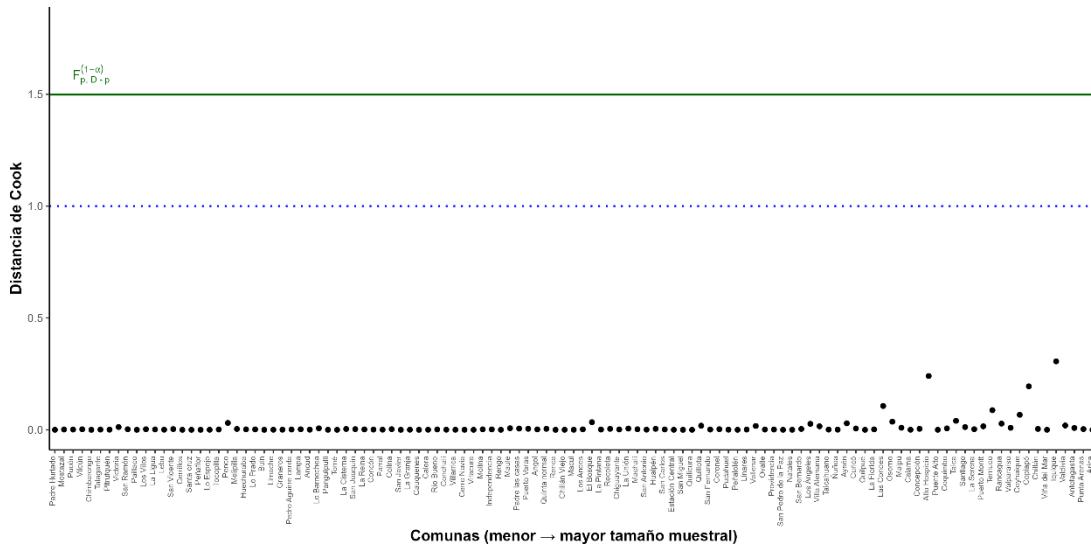
$(\hat{a}_{dm}^2 \leq c) = 1 - \alpha$, para un α dado. Si el residuo cuadrado más grande es mayor que c (es decir, $\hat{a}_{dm}^2 > c$), entonces el dominio correspondiente se considera un valor atípico. Se reserva y se retendrá la estimación directa para este dominio. A partir de los dominios restantes, se vuelve a calcular el modelo y 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

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

donde q es la dimensión de \mathbf{x}_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](#))²⁵. La Figura 6 muestra que no hay valores influyentes.

²⁵ Otra aproximación consiste en comparar $Dist_d$ con $F(p, D - p, \alpha)$, donde p es el número de variables y n el número de observaciones. Si $Dist_d > F$ es un posible outlier.

Figura 6. Distancias de Cook según comunas para VHDV



6.2. BONDAD DE AJUSTE

También es informativo evaluar el desempeño de un estimador a partir de alguna medida que dé cuenta de la fuerza del conjunto de covariables incluidas en los modelos para la predicción de θ_d . Para este fin, hemos considerado y calculado como métrica de evaluación un coeficiente de determinación o R^2 propuesto por [Hidiroglou et al. \(2019\)](#). Este coeficiente de determinación está asociado al modelo de enlace $\theta_d = x_d^T \beta + u_d$. Tenga en cuenta que el coeficiente de determinación asociado con el modelo combinado, $\hat{\theta}_d^{DIR} = x_d^T \beta + u_d + e_d$, no es de interés ya que el objetivo no es la predicción de $\hat{\theta}_d^{DIR}$, sino la predicción de θ_d . El coeficiente de determinación viene dado por:

$$R_{Hidiroglou}^2 = 1 - \frac{\hat{A}}{\frac{(D - q)}{(D - 1)} \hat{A} + S^2(\hat{\beta})}, \quad (8)$$

donde q es la dimensión de x_d y $S^2(\hat{\beta})$ es la varianza muestral de $x_d^T \hat{\beta}$. Los detalles de la derivación del coeficiente de determinación en (8) son provistos por [Hidiroglou et al. \(2019\)](#).

[pp.122-123](#)). El modelo seleccionado tiene una bondad de ajuste del 84% lo que indica una adecuada capacidad explicativa del conjunto de covariables utilizadas.

6.3. BENCHMARKING

Finalmente, se espera que el conjunto de estimaciones de área pequeña sea numéricamente consistente y más eficiente que las estimaciones directas insesgadas. Para verificar esta consistencia, se comparan 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, es posible utilizar técnicas de *benchmarking* para lograr dicha consistencia. Las estimaciones por *benchmarking* de áreas pequeñas ofrecen una propiedad atractiva para las ONE, en el sentido que garantizan consistencia entre las estimaciones obtenidas mediante un modelo estadístico y las estadísticas oficiales obtenidas bajo el diseño muestral de la encuesta, y que se consideran confiables. En este caso, los tamaños muestrales en las regiones de ENUSC 2024 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 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 2024; 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.

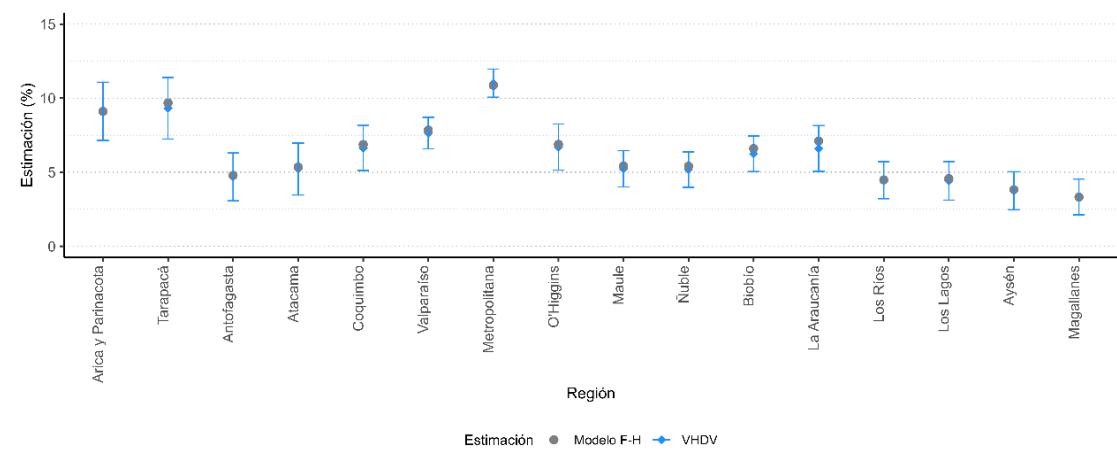
Generalmente, las estimaciones basadas en modelos no cumplen con esta propiedad de *benchmarking*. Para conseguirlo, se sigue el ajuste propuesto por [Fay y Herriot \(1979\)](#), multiplicando las estimaciones de las tasas comunales por un factor de consistencia de la forma:

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

Así, la estimación de VHDV, 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 (Ver Tabla 6). Adicionalmente, y como medida de evaluación de la consistencia interna de las estimaciones comunales SAE de cada indicador, **se verifica que antes de la implementación del *bechkmarking*, al agregar las estimaciones a nivel regional, estas se encuentran contenidas dentro del intervalo de confianza de las estimaciones regionales directas, para las cuales la ENUSC 2024 si tiene representatividad**. La Figura 7 muestra que, en todas las regiones, la estimación SAE previa al benchmarking, está contenida dentro del intervalo de confianza de la estimación regional oficial 2024.

Tabla 6. Coeficientes λ_r , Benchmarking

Región	VHDV
Arica y Parinacota	1.000
Tarapacá	0.963
Antofagasta	0.979
Atacama	0.975
Coquimbo	0.966
Valparaíso	0.973
Metropolitana	1.013
O'Higgins	0.973
Maule	0.965
Ñuble	0.959
Biobío	0.946
La Araucanía	0.929
Los Ríos	0.996
Los Lagos	0.967
Aysén	0.980
Magallanes	1.003

Figura 7. Verificación pre benchmarking VHDV


6.4. ERROR CUADRÁTICO MEDIO Y COEFICIENTE DE VARIACIÓN LOGARÍTMICO

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 $\hat{\theta}_d^{DIR}$ es el MSE, que se define como:

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

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:

$$MSE(\hat{\theta}_d^{EBLUP}) = g_{1i}(A) + g_{2i}(A) + 2g_{3i}(A), \quad (9)$$

donde $g_{1i}(A) = \hat{\gamma}_d \widehat{\psi}_d$, $g_{2i}(A) = (1 - \hat{\gamma}_d)^2 x_d^T \left[\sum_{d=1}^D \frac{x_d x_d^T}{A + \psi_d} \right]^{-1} x_d$, y

$$g_{3i}(A) = (1 - \hat{\gamma}_d)^2 (A + \widehat{\psi}_d)^{-1} \overline{V}(A)$$

donde $\overline{V}(A)$ es la varianza asintótica del estimador de A.

De igual forma, 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 directa. 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.

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

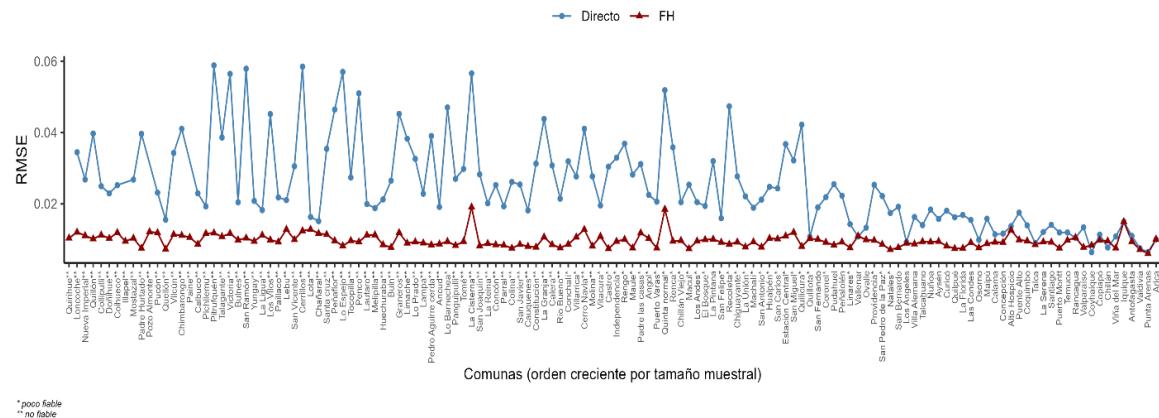
Sin embargo, el CV es una medida para la que se requiere hacer algunas consideraciones importantes. Por una parte, si se fija un umbral para el CV, esto tiene un impacto directo sobre la amplitud relativa del intervalo de confianza. Por ejemplo, se fija como umbral un CV del 30% ([Gutiérrez, et. al., 2020](#)), la amplitud relativa del intervalo de confianza estará en torno al 118%²⁶. 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:

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

La Figura 8 muestra el RMSE²⁷ (error cuadrático medio de la raíz) estimados de las estimaciones directas y EBLUP. Se observa que, en general, las estimaciones de áreas pequeñas obtenidas a través de EBLUP (rojo) son significativamente más eficientes que las estimaciones directas (azul), especialmente en las áreas de menor tamaño muestral.

²⁶ Esto se verifica escribiendo el intervalo de confianza en términos del coeficiente de variación y posteriormente, haciendo la relación con la amplitud relativa.

²⁷ RMSE (Root Mean Squared Error).

Figura 8. RMSE comunales SAE vs. Directas para VHDV


La Tabla 7 muestra medidas resumen de los dos tipos de estimadores, directo y F-H; en cuanto a las estimaciones se observa un efecto de suavizamiento, y adicionalmente, para el CVL y el RMSE estimado, evidencian la ganancia en precisión al implementar SAE, pasando por ejemplo de un CVL máximo comunal del 28% a uno del 8.7%.

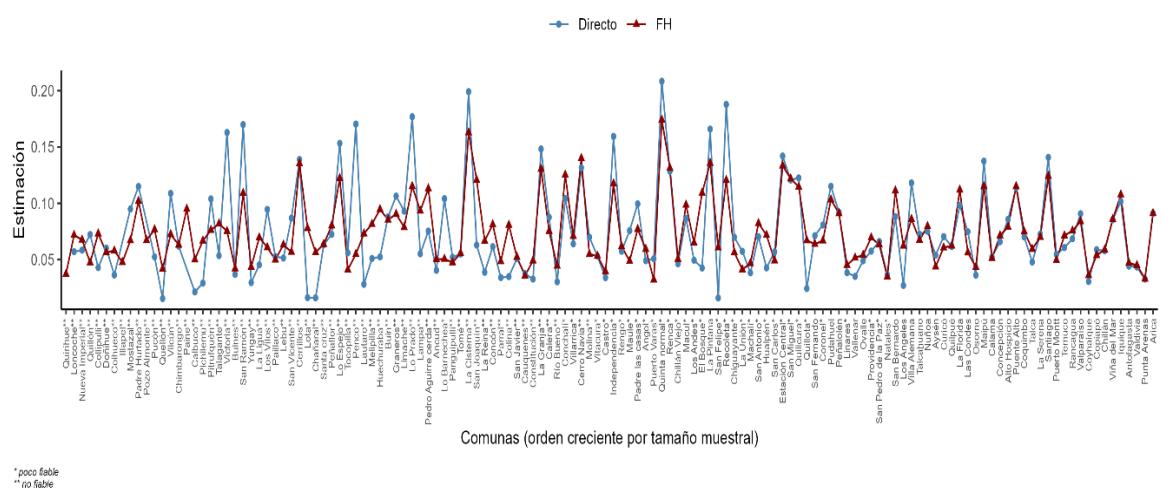
Tabla 7. Medidas de resumen para comparar desempeño del estimador directo y F-H

	Tipo	Mín.	Cuartil 1	Media	Mediana	Cuartil 3	Máx.
Estimación	Directo	0.0000	0.0430	0.0720	0.0610	0.0920	0.2080
	F-H	0.0310	0.0510	0.0733	0.0652	0.0868	0.1764
CVL	Directo	4.5931	9.6476	14.1473	14.3474	17.5850	28.0718
	F-H	3.0575	4.5377	5.4389	5.6223	6.3185	8.7489
RMSE	Directo	0.0000	0.0157	0.0244	0.0218	0.0312	0.0589
	F-H	0.0061	0.0085	0.0097	0.0095	0.0106	0.0191

7. ESTIMACIONES

Como se señaló anteriormente, para la construcción del EBLUP del indicador VHDV, se utilizó la transformación arcoseno, por lo tanto, se corroboró que $x_d \widehat{\beta} \in [0, \pi/2]$, es decir, que las estimaciones sintéticas se encuentren en el rango válido. Además, las estimaciones EBLUP fueron ajustadas por el factor λ_r , descrito en la [subsección 6.3](#). La Figura 9, muestra las estimaciones para VHDV, 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 comunas con los tamaños de muestra más grandes, por ejemplo, Punta Arenas y Arica, ambas estimaciones son muy similares.

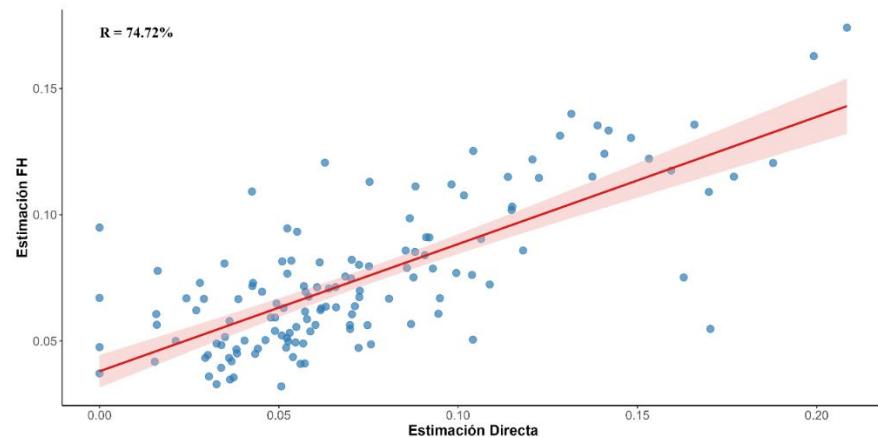
Figura 9. Estimaciones comunales SAE vs. Directas para VHDV



La Figura 10 muestra el gráfico de dispersión, que 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 dos casos.

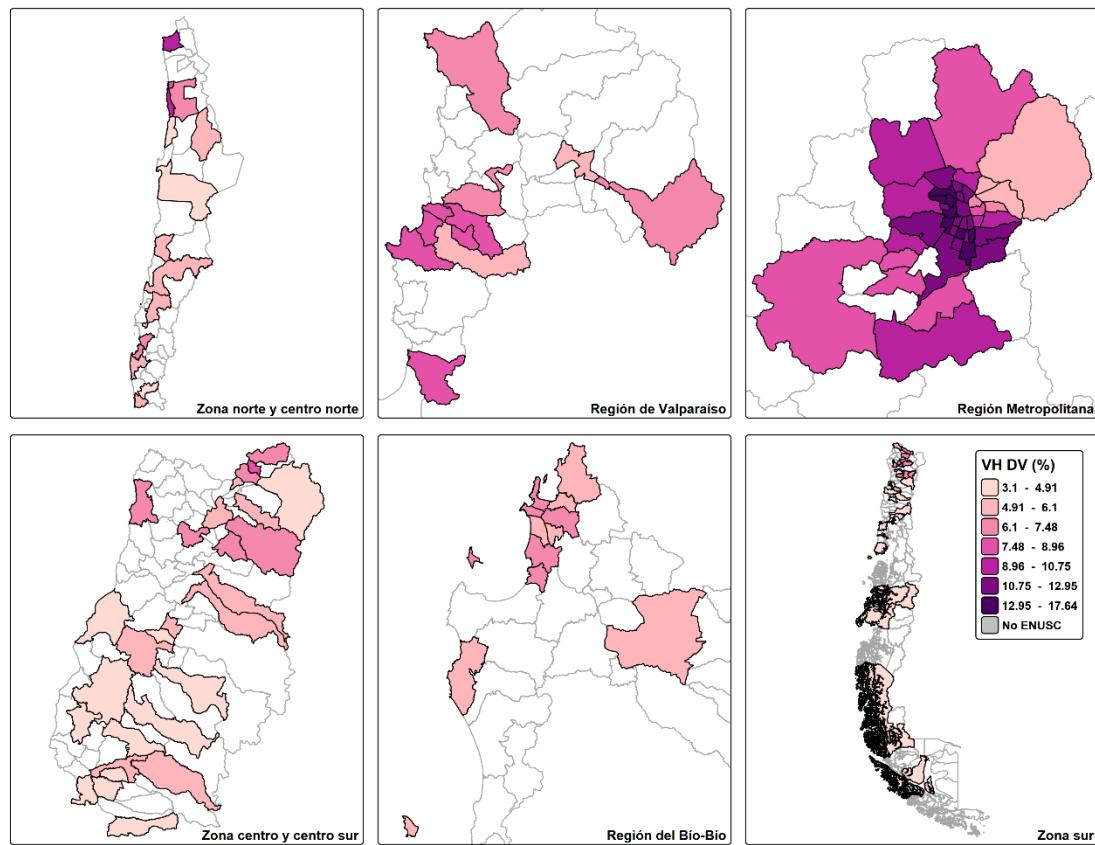
Para evaluar más a fondo la presencia de sesgo en las estimaciones basadas en modelos, se ajustó una línea de regresión para las estimaciones directas y basadas en modelos. La estimación del coeficiente de pendiente fue significativa con *valor p* < 0.005, lo que da cuenta de la consistencia entre las estimaciones obtenidas mediante el muestreo y el modelo F-H.

Figura 10. Estimación directa vs. Estimación EBLUP para VHDV



Finalmente, es posible visualizar la distribución del fenómeno mediante mapas que evidencian patrones de diferenciación geográfica entre comunas en términos de victimización por delitos violentos. Esta representación permite distinguir agrupamientos de comunas con baja o alta victimización, así como comunas que se ubican en un extremo de la distribución respecto de su entorno inmediato. Para visualizar el mapa interactivo del indicador, visitar el siguiente enlace: [\[LINK\]](#). La versión fija se presenta en la Figura 11. Cabe señalar que la clasificación en categorías se construye a partir de estimaciones puntuales y debe interpretarse con cautela, dado que cada estimación posee incertidumbre (intervalo de confianza); por ello, comunas cercanas a los puntos de corte pueden ser compatibles con categorías adyacentes.

Figura 11. Mapa comunal según zona geográfica para estimación SAE de VHDV



8. DISCUSIÓN Y TRABAJO FUTURO

Este documento ha presentado la implementación de una metodología para obtener estimaciones EBLUP confiables, para el indicador de Victimización a Hogares por Delitos Violentos (VHDV), en las áreas urbanas de las 136 comunas pertenecientes a la muestra de la ENUSC 2024. El análisis de los resultados permite concluir que los métodos de estimación de áreas pequeñas basados en modelos —y en particular, el EBLUP bajo la metodología de Fay-Herriot— han mostrado ser herramientas eficaces para estimar, a nivel comunal, el comportamiento del indicador. Esto resulta especialmente relevante cuando los tamaños de muestra no son lo suficientemente grandes como para obtener estimaciones directas confiables.

Para garantizar la consistencia de los resultados, es imperativo contar con datos auxiliares fiables: precisos, confiables, accesibles, claros y coherentes, entre otros aspectos que se encuentran definidos por la Guía para la implementación del marco de aseguramiento de la calidad para procesos y productos estadísticos (2022). En ese sentido, previo a la implementación de modelos F-H, se establece un marco de evaluación de la calidad para las fuentes de información, considerando criterios como la cobertura geográfica y temporal, la precisión en los registros y la consistencia en las definiciones operativas de las variables, además del nivel de actualización de los datos. Igual de importante ha sido la elaboración del marco teórico para identificar dimensiones temáticas relacionadas con la victimización y asociar a cada una de estas dimensiones, fuentes de información y variables disponibles para la especificación del estimador de interés.

En relación con la etapa de evaluación y validación, que 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, constituye una parte crucial para asegurar la precisión y utilidad de las estimaciones comunales SAE del indicador de victimización de la ENUSC 2024 que se ha modelado. En ese sentido, en primera instancia, se han establecido un conjunto de criterios de calidad para las estimaciones directas, previo a la implementación del SAE, garantizando que las estimaciones comunales de la encuesta que no cumplen con la calidad fijada, no contribuyan en la generación del modelo para el resto de comunas (se excluyen 21 comunas). En segundo lugar, al realizar la implementación del modelo F-H, se hace una revisión exhaustiva de los modelos para garantizar: i) consistencia teórica, es decir, que las covariables usadas tienen relación con VHDV, ii)

buenas bondades de ajuste mediante R^2 de Hidiroglou, significancia estadística para los parámetros estimados; y además, para juzgar la complejidad del modelo, se emplean los criterios de información de Akaike (*AIC*) e información Bayesiana (*BIC*), iii) análisis que verifican supuestos para los efectos aleatorios y los residuales estandarizados, como normalidad, homocedasticidad y no autocorrelación y iv) se verifica la no presencia de puntos atípicos o influyentes. Adicionalmente, para garantizar consistencia interna, previo a la etapa de *benchmarking*, se verifica que las estimaciones comunales SAE agregadas a nivel regional, se encuentran contenidas dentro del intervalo de confianza de las estimaciones regionales directas, para las cuales, la ENUSC 2024 si tiene representatividad. En cuanto a las medidas de precisión, se ha verificado que el CVL promedio obtenido mediante SAE en ENUSC 2024 (5,4%), es menor que el CVL promedio de estimación directa 2024 (14,1%).

Respecto a los alcances y limitaciones, se puede señalar que los resultados obtenidos están supeditados tanto a la calidad de las fuentes de información auxiliar, como a aspectos relacionados con las estimaciones directas, en particular sus medidas de error. Respecto a las limitaciones de los registros administrativos utilizados, se identificaron limitaciones de cobertura, ya que, por ejemplo, en el caso del Registro Social de Hogares, su cobertura depende de la cantidad de hogares y personas que se auto reportan en el mismo. Por otro lado, algunas variables obtenidas del Servicio Nacional de Información Municipal, dada la ausencia de datos para todas las comunas, o algunas inconsistencias identificadas, fue necesario realizar imputación utilizando información de períodos anteriores. **La inclusión del Censo de Población y Vivienda 2024 mitigó estas limitaciones, fortaleciendo la especificación y la consistencia del modelo final.**

En relación con los problemas del levantamiento muestral, se observa que, por ejemplo, para las estimaciones comunales directas del indicador VHDV fueron identificadas 64 comunas con un $DEFF < 1$, lo que podría sugerir una subestimación de la varianza. Por tanto, para subsanar este problema, fue necesario incorporar una Función de Varianza Generalizada. Adicionalmente, dentro del flujo de calidad fijado para este tipo de estimaciones, se identificaron, por ejemplo, comunas que debían ser excluidas debido a un tamaño de muestra logrado menor al umbral establecido (50), lo que plantea desafíos y consideraciones en el diseño muestral para futuras versiones de la encuesta.

Finalmente, un aspecto relevante en este punto es que, **aunque la metodología SAE permite obtener estimaciones confiables a nivel comunal para el periodo 2024, no es posible recurrir a las pruebas estadísticas habituales para comparar las estimaciones con el periodo anterior, 2023, ya que, en esa versión, las estimaciones comunales fueron producidas exclusivamente bajo métodos directos o muestreo.** Esta misma limitación se presentaría en futuras estimaciones si se especifican diferentes modelos SAE en cada periodo (ya que cada modelo se ajusta específicamente a los datos del momento y además se requiere asegurar la disponibilidad de las mismas fuentes y covariables a lo largo de los periodos que se desean comparar). **Así, se evidencia la necesidad de investigar y profundizar en metodologías estadísticas que permitan comparar los resultados para áreas pequeñas, en periodos donde la estimación es obtenida mediante estimadores directos y periodos donde la estimación es obtenida mediante SAE, o bien, bajo diferentes especificaciones de modelos.** Esto requiere establecer un trabajo metodológico coordinado y sostenido en el tiempo, que permita generar una serie interanual comparable. Además, es fundamental asegurar la disponibilidad continua de las fuentes de información necesarias para mantener esa coherencia a lo largo del tiempo.

Como trabajo futuro en la implementación SAE para la ENUSC, además del desafío sobre la comparabilidad que se menciona anteriormente, será necesario definir, junto a la Subsecretaría de Prevención del Delito, un nuevo conjunto de indicadores de interés. Entre ellos, podrían considerarse la percepción de la inseguridad, la cifra oculta o delitos específicos, como robo con violencia o intimidación.

9. BIBLIOGRAFÍA

- Alvira, F., & Rubio, M. (1982). *Victimización e inseguridad: La perspectiva de las encuestas de victimización en España*.
- 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.
- Bell, W. R. (2008). Examining sensitivity of small area inferences to uncertainty about sampling error variances. . Joint Statistical Meetings-Section on Survey Research Methods.
- 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.
- Buil Gil, D., Moretti, A., Shlomo, N., & Medina, J. (2020). Applying the spatial EBLUP to place-based policing. Simulation study and application to confidence in police work. *Applied spatial analysis and policy*, 13(4), 901-924.
- 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.
- Comisión Económica para América Latina y el Caribe [CEPAL]. (2022). *Guía para la implementación del marco de aseguramiento de la calidad para procesos y productos estadísticos (LC/CEA.11/19)*. Santiago, Chile.
<https://repositorio.cepal.org/server/api/core/bitstreams/a15254d0-5402-4391-82e9-2243ac036746/content>
- Comisión Económica para América Latina y el Caribe [CEPAL]. (2022). *Marco de referencia para los sistemas estadísticos de seguridad y justicia penal en América Latina y el Caribe*.
<https://www.cepal.org/es/publicaciones/47928-marco-referencia-sistemas-estadisticos-seguridad-justicia-penal-america-latina>
- Corral, P., Molina, I., Cojocaru, A., & Segovia, S. (2022). Guidelines to small area estimation for poverty mapping. Washington: World Bank.

- 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> .
- Cook, R. D., & Weisberg, S. (1994). An introduction to regression graphics. Wiley.
- Dammert, L. (2012). Fear and crime in Latin America: redefining state-society relations. Routledge.
- División Observatorio Social MDSF – CEPAL. (2020). Informe Metodológico. Estimaciones Comunales de Pobreza por ingresos en Chile Mediante Métodos de Estimación en Áreas Pequeñas. Obtenido de
https://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/pobreza-comunal/2020/Informe_SAE_2020.pdf
- Díaz, F. J., & Meller Bock, P. (2012). Violencia y cohesión social en América Latina. CIEPLAN.
- Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, 71(1), 355-364.
- Fajnzylber, P., Lederman, D., & Loayza, N. (2001). Crimen y violencia en América Latina. Washington, D.C: Banco Mundial. Obtenido de
<https://documentos.bancomundial.org/es/publication/documents-reports/documentdetail/522791468749792478/crimen-y-violencia-en-america-latina>
- Fay, R. E., & Diallo, M. S. (2012). Small area estimation alternatives for the national crime victimization survey. *Proceedings of the Section on Survey Research Methods*. American Statistical Association, 3742–3756.
- 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.

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>

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.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 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.

Instituto Nacional de Estadísticas - Chile. (2025a). *Informe de Calidad ENUSC 2024*. Manual. Obtenido de https://www.ine.gob.cl/docs/default-source/seguridad-ciudadana/metodologia/2024/informe-de-calidad---enusc-2024.pdf?sfvrsn=80446245_2

Instituto Nacional de Estadísticas - Chile. (2025b). *Informe Metodológico ENUSC 2024*. Manual. Obtenido de https://www.ine.gob.cl/docs/default-source/seguridad-ciudadana/metodologia/2024/informe-metodologico---enusc-2024.pdf?sfvrsn=651010da_2

Instituto Nacional de Estadísticas - Chile. (2025c). *Manual de Usuario Base de datos 2024*. Manual. Obtenido de https://www.ine.gob.cl/docs/default-source/seguridad-ciudadana/manuales/2024/manual-de-usuario-de-base-de-datos---enusc-2024.pdf?sfvrsn=bd9be134_2

Instituto Nacional de Estadísticas - Chile. (2025d). *Memoria CENSO 2024*. Manual. Obtenido de <https://censo2024.ine.gob.cl/wp-content/uploads/2025/05/MEMORIA-CENSO-2024.pdf>

Instituto Nacional de Estadísticas - Chile. (Febrero de 2020). Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares. Obtenido de <https://www.ine.gob.cl/docs/default-source/institucionalidad/buenas-pr%C3%A1cticas/clasificaciones-y-estándares/est%C3%A1ndar-evaluaci%C3%B3n-de-calidad-de-estimaciones-publicaci%C3%B3n-27022020.pdf>

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-interanual---enusc-2008-2018.pdf?sfvrsn=41a813c5_4

International Organization for Standardization. (2008). *ISO 25000: Calidad de software y datos*. Obtenido de <https://iso25000.com/index.php/normas-iso-25000/iso-25012>

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, J., Diallo, M.S., and Fay, R.E., (2012) "Rethinking the NCVS: Small Area Approaches to Estimating Crime," presented at the 2012 Federal Committee on Statistical Methodology Conference, Washington, DC, Jan. 10-12, 2012, available at http://www.fcsm.gov/12papers/Li_2012FCSM_I-B.pdf.

Meier, R. F., & Miethe, T. D. (1993). Understanding theories of criminal victimization. *Crime and justice*, 17, 459-499.

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>

Münnich, R., Burgard, J. B., Ertz, F., Lenau, S., Manecke, J., & Merkle, H. (2019). *Small area estimation for city statistics and other functional areas* (Eurostat Statistical Working Papers, KS-TC-19-006-EN). Luxembourg: Publications Office of the European Union.

Olavarría Gambi, M. (2006). El crimen en Chile: Una mirada desde las víctimas. Santiago: CESC.

R Core Team. (2019). A language and environment for statistical computing. *R Foundation for Statistical Computing*.

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.

Santillana, M. A., Gallegos, J. A., García, A. I., Díaz, E., Gutiérrez, D., & González, N. L. (2023). Perception of insecurity in municipalities in Mexico: A small area estimation approach. *Statistical Journal of the IAOS*, 39(2), 491-500.

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

Sugawara, 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

Tanton, R., Jones, R., & Lubulwa, G. (2001). Analyses of the 1998 Australian national crime and safety survey. Character, Impact; Prevention of Crime in Regional Australia Conference.

Thompson, S. K. (2012). *Sampling* (Vol. 755). John Wiley & Sons.

- Tseloni, A., Wittebrood, K., Farrell, G., & Pease, K. (2010). The effectiveness of burglary security devices. *European Journal of Criminology*. Disponible en https://www.researchgate.net/publication/271909133_The_effectiveness_of_burglary_security_devices.
- 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.
- Valliant, R. (1987). Generalized variance functions in stratified two-stage sampling. *Journal of the American Statistical Association*(82), 499-508.
- Van Dijk, J., van Kesteren, J., & Smit, P. (2007). Criminal Victimization in International Perspective: Key findings from the 2004–2005 ICVS and EU ICS. UNODC/UNICRI. Disponible en https://unicri.org/sites/default/files/2021-06/ICVS2004_05report.pdf
- Wolter, K. M. (1985). Introduction to variance estimation. Springer-Verlag.
- Wolter, K. M. (2007). *Statistics for social and behavioral sciences. Introduction to variance estimation* (2nd ed.). Springer Science + Business Media.

10. ANEXOS

ANEXO 1: FUENTES DE INFORMACIÓN, DIMENSIÓN DE ANÁLISIS Y COVARIABLES CONSIDERADAS EN LA IMPLEMENTACIÓN DE MODELOS FAY-HERRIOT PARA VHDV

Fuente	Dimensión	Covariable
RSH	Características sociodemográficas/socioeconómicas	Proporción de hombres del RSH en CSE 1 Proporción de mujeres del RSH en CSE 1 Proporción de personas del RSH en CSE 1 Proporción de personas del área urbana del RSH en CSE 1 Proporción de personas adscritas a Fonasa del RSH en CSE 1 Proporción de personas adscritas a Isapre del RSH en CSE 1 Proporción de personas adscritas a Fonasa del RSH Proporción de personas adscritas a Isapre del RSH Proporción de hombres carentes del RSH en CSE 1 Proporción de mujeres carentes del RSH en CSE 1 Proporción de hombres del RSH en CSE 1 Proporción de hombres extranjeros del RSH en CSE 1 Proporción de mujeres extranjeras del RSH en CSE 1 Proporción de extranjeros del RSH en CSE 1 Proporción de hogares del RSH en CSE 1 Proporción de hogares del área urbana del RSH en CSE 1 Proporción de hogares hacinados del RSH en CSE 1 Proporción de hogares hacinados del RSH
RRAA MDSF	Características sociodemográficas/socioeconómicas	Tasa de matrícula de personas entre 0 y 24 años Mediana del ingreso imponible de los asalariados dependientes
	Características socioeconómicas	Promedio ingreso imponible de la población en edad de trabajar
	Características socioeconómicas	Tasa de ocupados formales dependientes
SINIM	Características de la infraestructura urbana y del entorno	Proporción de casetas de seguridad por comuna Proporción de cámaras de vigilancia por comuna Proporción de guardias, inspectores o vigilantes por comuna personas pagadas para seguridad (recursos)
Imagen Satelital	Características de la infraestructura urbana y del entorno	Promedio ponderado de radiancia nocturna
CCh	Factores relacionados con el crimen y la delincuencia	Proporción de agresiones y lesiones casos de víctimas, es la suma de los delitos: Lesiones contra profesionales y funcionarios de salud y manipuladores de alimentos, Lesiones graves, Lesiones con arma blanca, lesiones con fuegos artificiales, lesiones con armas de fuego, Lesiones menos graves, Lesiones graves gravísimas, Lesiones graves y menos graves a miembros de bomberos en el ejercicio de sus funciones, Lesiones leves,

Fuente	Dimensión	Covariable
		Lesiones a las personas o daño a la propiedad con motivo u ocasión de un espectáculo de fútbol profesional.
		Proporción de amenaza casos de víctimas, es la suma de los delitos: Amenazas simples contra personas y propiedades, Amenazas condicionales contra personas y propiedades, Amenazas contra profesionales y funcionarios de salud y manipuladores de alimentos, Amenazas a carabineros, Amenaza con arma (falta), Amenazas con arma blanca.
		Proporción de amenaza casos de víctimas, diferenciado para la VHDV, es la suma de los delitos: Amenaza con arma (falta) y Amenazas con arma blanca.
		Proporción de hackeo de correo electrónico y redes sociales, infección de software malicioso o ciberbullying casos de víctimas, es la suma de los delitos: Infracción Ley general. de telecomunicaciones, Ataque a la integridad de los datos informáticos, Sabotaje informático., Ataque a la integridad de un sistema informático., Acceso ilícito, Interceptación ilícita, Falsificación informática, Receptación de datos informáticos, Otros fraudes informáticos, Abuso de los dispositivos.
		Proporción de Robo con violencia o intimidación casos de víctimas, es la suma de los delitos: Robo con intimidación y Robo con violencia.
		Proporción de Robo de vehículos casos de víctimas, es la suma de los delitos: Robo de vehículo motorizado por sorpresa, violencia o intimidación y Robo con intimidación, violencia o sorpresa vehículo Motorizado (encerrona).
		Proporción Vandalismo de vehículo casos de víctimas, es la suma de los delitos: Daños simples y Atentado a vehículo motorizado en circulación con objeto contundente u otro semejante.
		Proporción Vandalismo a la vivienda casos de víctimas, es la suma de los delitos: Usurpación u ocupación no violenta, pero con daño en las cosas., Destrucción o alteración de deslindes, Incendio con peligro para las personas, Incendio solo con daños o sin peligro de propagación.
		Proporción de Amenazas a Carabineros
		Proporción de Lesiones leves
		Proporción de Amenazas con arma blanca
		Proporción de Hurto Falta
		Proporción de Amenaza falta con arma
		Proporción de Otros fraudes informáticos
		Proporción de Amenazas simples contra personas y propiedades
		Proporción de Amenazas condicionales contra personas y propiedades
		Proporción de Amenazas contra profesionales y funcionarios de la salud y manipuladores de alimentos
		Proporción de Lesiones contra profesionales y funcionarios de la salud y manipuladores de alimentos
		Proporción de Lesiones graves
		Proporción de Lesiones menos graves
		Proporción de Lesiones con arma blanca
		Proporción de Lesiones con arma de fuego
		Proporción de Lesiones graves gravísimas

Fuente	Dimensión	Covariable
	Características sociodemográficas	Proporción por persona de infraestructura de Carabineros

Fuente	Dimensión	Covariable
		Proporción de personas de sectores críticos y muy críticos identificados por carabineros.
CPV 2024	Características sociodemográficas	Proporción de hombres
		Proporción de mujeres
		Proporción de personas censadas por tramo etario (0-14, 15-24, 25-34, 35-44, 45-54, 55-64 y 65 o más años)
		Proporción de personas extranjeras
		Proporción de personas ocupadas
		Proporción de personas desocupadas
		Proporción de personas inactivas
		Proporción de personas sin nivel de escolaridad
		Proporción de personas con nivel de escolaridad primaria
		Proporción de personas con nivel de escolaridad secundaria
		Proporción de personas con nivel de escolaridad terciaria
		Proporción de viviendas particulares categorías 6 a 8 pregunta 2
		Proporción de viviendas en condiciones inadecuadas según su materialidad
		Proporción de viviendas en condiciones de hacinamiento
Precenso 2023-TECHO-MINVU	Características de la infraestructura urbana y del entorno	Proporción por persona de número de campamentos



ENCUESTA
NACIONAL
URBANA DE
SEGURIDAD
CIUDADANA

ENUSC
2024