



ESTADÍSTICAS
MIGRATORIAS



REPORTE DE CALIDAD DE LA
ESTIMACIÓN DE PERSONAS EXTRANJERAS RESIDENTES EN CHILE
AL 31 DE DICIEMBRE DE 2023

DEPARTAMENTO DE ESTADÍSTICAS DEMOGRÁFICAS Y SOCIALES
SUBDEPARTAMENTO DE DEMOGRAFÍA
INSTITUTO NACIONAL DE ESTADÍSTICAS

DEPARTAMENTO DE ESTUDIOS
SERVICIO NACIONAL DE MIGRACIONES

Diciembre 2024

INSTITUTO NACIONAL DE ESTADÍSTICAS

Subdirección Técnica

Departamento de Estadísticas Demográficas y Sociales

Subdepartamento de Demografía

Período de información: 2023

Metodología-Estimación de personas extranjeras

Versión 1.0

Dirección: Morandé 801, piso 22, Santiago de Chile

Casilla de correo: 498 – Correo 3

Teléfono: (56) 232461010 - 232461018

Sitio web: www.ine.gob.cl

Correo electrónico: ine@ine.gob.cl

Facebook: @ChileINE

Twitter: @INE_Chile

Instagram: @chile.ine

LinkedIn: Instituto Nacional de Estadísticas de Chile

YouTube: @inechileoficial

SERVICIO NACIONAL DE MIGRACIONES

Departamento de Estudios

Dirección: San Antonio 580

Sitio web: <https://serviciomigraciones.cl/>

Facebook: @Migraciones Chile

Twitter: @MigracionesCL

Instagram: @migracionescl

LinkedIn: Servicio Nacional de Migraciones

YouTube: Migraciones Chile

INSTITUTO NACIONAL DE ESTADÍSTICAS

Dirección Nacional: Ricardo Vicuña Poblete

Subdirección Técnica: Leonardo González Allendes

Departamento de Estadísticas Demográficas y Sociales: Alejandro Henríquez

Subdepartamento de Demografía: Miguel Ojeda, Julibeth Rodríguez, Francisco González, Luis Rodríguez y Felipe Hugo.

SERVICIO NACIONAL DE MIGRACIONES

Dirección Nacional: Luis Thayer Correa

Departamento de Estudios: Felipe Mallea, Gabriel Santander, Marisol Opazo, Pablo Roessler y Consuelo Salas.

LISTADO DE ABREVIATURAS DEL DOCUMENTO

COD EXT: Código Extranjero

DNI: Documento Nacional de Identidad

INE: Instituto Nacional de Estadísticas

Mineduc: Ministerio de Educación

Minrel: Ministerio de Relaciones Exteriores

PDI: Policía de Investigaciones

RNE: Registro Nacional de Extranjeros

RRAA: Registros Administrativos

RUN: Rol Único Nacional

RUN_ALU(IPE): Identificador Provisorio Estudiante

Sermig: Servicio Nacional de Migraciones

SRCel: Servicio de Registro Civil e Identificación

HISTORIAL DE VERSIONES			
Versión	Fecha	Autores / Responsables	Cambios efectuados
1.0	Diciembre 2024	Instituto Nacional de Estadísticas (Subdirección Técnica, Departamento de Estadísticas Demográficas y Sociales, Subdepartamento de Demografía) Servicio Nacional de Migraciones Departamento de Estudios	No aplica

ÍNDICE GENERAL

ÍNDICE GENERAL.....	5
1 NOTA TECNICA.....	6
2 INTRODUCCIÓN.....	6
3 PERTINENCIA.....	9
4 PRECISIÓN Y CONFIABILIDAD.....	10
4.1 Evaluación censal.....	10
4.1.1 Controles al proceso censal.....	11
4.1.2 Tasa de no respuesta al ítem por variable.....	11
4.2 Evaluación de las fuentes administrativas.....	13
4.2.1 Tasa de completitud de las variables administrativas.....	13
4.2.2 Porcentaje de errores por variable del conjunto de datos.....	15
4.2.3 Porcentaje de observaciones duplicadas (POD).....	17
4.3 Evaluación del proceso de integración.....	18
5 ACCESIBILIDAD Y CLARIDAD.....	21
5.1 Utilización de métodos de Anonimización.....	21
6 COHERENCIA Y COMPARABILIDAD.....	21
6.1 Grado de coherencia con otras Fuentes.....	21
7 CONCEPTOS CLAVES.....	26
8 ANEXOS.....	28
9 REFERENCIAS.....	33

1 NOTA TECNICA

Los resultados entregados en esta presentación corresponden a información estimada al 31 de diciembre 2023 y utilizan, entre otras fuentes, de información el pasado Censo 2017.

La próxima publicación de resultados del Censo de Población y Vivienda 2024 dará cuenta de la información oficial sobre la población extranjera residente en Chile para dicho año. De esta forma, **la presente edición de la estimación de población extranjera a diciembre de 2023 termina un ciclo productivo, completando seis publicaciones intercensales entre 2018 y 2023.**

Por lo anterior, **durante el año 2025** el INE y el Sermig harán una revisión de la metodología y de las fuentes de información, necesarias para seguir entregando continuamente estadísticas relevantes sobre la población inmigrante internacional. Para más información ver documento Separata Técnica¹.

2 INTRODUCCIÓN

El Instituto Nacional de Estadísticas (INE) y el Servicio Nacional de Migraciones (Sermig), en colaboración con la Policía de Investigaciones de Chile (PDI), el Ministerio de Relaciones Exteriores (Minrel), el Servicio de Registro Civil e Identificación (SRCel) y el Ministerio de Educación (Mineduc), realizan una estimación anual del número de personas extranjeras que residen habitualmente en el país. Dicha estimación se realiza combinando datos censales con registros administrativos postcensales, alcanzándose una cifra estimada de 1.918.583 millones de personas extranjeras residiendo en Chile al 31 de diciembre de 2023.

Las fuentes de datos de la estimación elaborada se clasifican en dos grandes fuentes, primero el Censo 2017, y, segundo, la integración de los siguientes registros administrativos según institución²:

- Sermig: Registro Nacional de Extranjeros (RNE) incluye diversos trámites tales como permisos de residencias solicitados dentro y fuera del territorio, prórrogas de turismo vencidas y empadronamiento biométrico.
- Minrel: permisos de residencia otorgados a nivel consular.

¹ Publicado en: <https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/demografia-y-migracion>

² Para más detalle ver documento metodológico publicado en: <https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/demografia-y-migracion>

- Mineduc: registro administrativo de matrícula escolar para estudiantes de educación primaria y secundaria de nacionalidad extranjera que al momento de inscribirse al sistema no cuentan con RUN.
- PDI: (1) registro de entradas y salidas al país mediante el control migratorio de las fronteras. (2) Registro de residentes (domicilios)* utilizado hasta 2022.
- SRCel: registro de defunciones en Chile.

Para la realización de este reporte de calidad se presentan análisis y evaluaciones a la información utilizada del Censo de población 2017, así como también de aquellos registros administrativos que el Sermig recibe a través de una base de datos para su procesamiento.

Por tanto, ha quedado excluido del reporte el registro del control de fronteras de la PDI³. Si bien es cierto que, el registro de control de frontera forma parte importante de la estimación, en la práctica la PDI no envía una base de datos para su procesamiento, sino que, a partir la información de registros Sermin-Minrel_Mineduc, vinculados por Sermig, es esta última institución quien envía una base a PDI, quien a su vez la complementa con información de fecha del viaje y rumbo (entrada o salida) en caso de que se puedan vincular.

Dado que se recibió solo una base de datos por Servicio, en adelante, y a modo de facilitar la lectura, los análisis e indicadores expuestos se denotarán con el nombre de la institución proveedora⁴.

En este contexto, dado el uso intensivo que se hace de registros administrativos de distintas instituciones, surge la necesidad de evaluar su calidad: “la relevancia de evaluar la calidad de los registros administrativos obedece a su amplio e intensivo uso como fuente de información estadística en la actualidad, siendo considerada como una fuente de información para la generación de estadísticas oficiales en complemento a encuestas, censos y otros tipos de procesos estadísticos. No obstante, **la incorporación y aprovechamiento estadístico de nuevas fuentes de información, como los registros administrativos, en las**

³ La información del procesamiento realizado por PDI fue descrita por su equipo técnico y se encuentra disponible en el documento metodológico en el apartado 7.3 procesamiento (etapa 4). Allí se señalan controles en la fase de procesamiento, realizados por Sermig respecto de la precisión de la información recibida desde PDI, en particular respecto de personas sin viajes identificados, diferencias en las fechas de viaje con respecto a las mismas personas en relación a la estimación 2022, no señalan rumbo del movimiento (entrada o salida) etc. A la fecha de elaboración de este reporte tanto el INE como el Sermig no acceden directamente a la base de datos PDI para aplicar el procesamiento.

⁴ Si bien en la teoría una institución puede tener muchas bases de datos sobre distintos eventos o servicios que presta, para la estimación cada institución compila, selecciona o filtra información con base a las necesidades específicas de la estimación como por ejemplo seleccionar solo a población extranjera o sólo las variables de interés de identificación y demográficas básicas.

estadísticas oficiales, deben someterse a un proceso de análisis y tratamiento que permita resguardar criterios mínimos de calidad, aun cuando se hayan capturado con diversos controles y validaciones de carácter administrativo” (INE, 2024).

Siendo que la estimación considera dos grandes componentes (censo y registros administrativos) que se agregan a nivel macro⁵, es importante señalar aspectos de calidad relativos al Censo 2017 y aquellos referidos al uso de registros administrativos a nivel micro⁶ para la construcción de la base de registros postcensal. Dichos aspectos corresponden a acciones e indicadores aplicados para garantizar tanto la precisión como la comparabilidad y coherencia de la estimación, dando cuenta, de acuerdo con las etapas de elaboración de la estimación, de los principales análisis y controles de calidad realizados.

De acuerdo con el uso del componente censal utilizado para la construcción de la estimación, se puede señalar que el Censo de Población y Vivienda 2017 (CPV2017) como operación estadística, presentó diversos controles de calidad a lo largo de las distintas fases y que han sido heredados como componente en la estimación, los cuales pueden ser consultadas en la Memoria Censal⁷ y en el Manual de Uso de la base de datos del Censo⁸. La Memoria Censal entrega información de toda la Operación Estadística, el manual de usuario de la base se refieren al procesamiento y la base de datos final.

Para la estimación de personas extranjeras la base de datos final del Censo se filtró considerando solo a las personas que declararon que al momento de nacer su madre vivía en otro país (país de nacimiento extranjero), y que declaran ser residentes habituales del país (se excluyen transeúntes o personas que declaran que su lugar de residencia es un país extranjero). De igual forma para el análisis regional se construye a partir de la suma de las comunas, por lo que en los casos con comuna ignorada se imputó la información de la comuna de empadronamiento, y posteriormente las comunas se agregaron a nivel regional.

A esta base se le aplica un factor de ajuste en la variable edad y otro por omisión, el detalle del procesamiento puede ser consultado en el documento metodológico en la sección 7.2.

Por su parte, en cuanto al uso de los registros administrativos, existen dos grandes etapas en la evaluación de la calidad, por un lado, la evaluación de la calidad de entrada, mientras que por otro lado se encuentra la medición de la calidad del proceso y resultados. Si bien en ambas etapas de evaluación de la calidad existe un set de indicadores recomendados (INE,

⁵ La integración de datos a nivel macro combina datos que ya están agregados para producir nuevas estadísticas.

⁶ La integración de datos a un nivel micro se refiere a la integración de datos a nivel de registro (individual).

⁷ Ver (INE, 2018)

⁸ Ver INE (2018a)

2024, pág. 9), en este documento se muestran solo los indicadores seleccionados para esta versión de la estimación de personas extranjeras.

Cada registro cumple una función en la estimación⁹, por lo que se han priorizado indicadores específicos por registro, lo que explica que existan distintas cantidades de variables, así como algunos indicadores para algunos registros en detrimento de otros. Para esta versión de reporte de calidad se seleccionaron principalmente indicadores asociados con la cobertura de variables, porcentajes de error y niveles de vinculación entre bases de datos, debido a que en la estimación de personas extranjeras la vinculación cumple la función fundamental de integrar subpoblaciones que pueden estar contenidas en un solo registro y deben ser consideradas en el universo de estudio. Por el principio de mejora continua, se espera generar mayor documentación en el futuro para aquellos registros y/o indicadores que no se incluyen en la presente versión.

Es importante señalar que la cantidad de datos recibidos desde las instituciones ha variado en el tiempo, obteniendo más información y una estructura de base de datos en los años más recientes. Por lo anterior, este reporte de calidad hace referencia a los registros acumulados al año de medición (diciembre 2023), es decir, además incluyen información de trámites anteriores los cuales se utilizan para las reestimaciones históricas. La tabla 1 muestra el resumen de variables utilizadas en la estimación, las cuales se encuentran contenidas en cada base de datos recibida procesada por Sermig para la compilación del componente de registros administrativos post-censales.

3 PERTINENCIA

A raíz de los análisis realizados a los resultados del Censo 2017, surge la necesidad de establecer una Mesa de trabajo entre diferentes organismos del Estado, con el objetivo de abordar el análisis de las tendencias y magnitudes de los flujos migratorios en el país. Las sesiones de la Mesa de Estadísticas Migratorias¹⁰ en el año 2018 levantaron la necesidad de

⁹ Los registros de Sermig, Minrel y Mineduc se suman para tener una base de personas únicas sobre la cual se agregan algunas variables desde PDI y SRCel para el descuento de residentes. Por lo anterior, no se analiza todo el registro administrativo, sino que se accede a algunas variables de caracterización. Para revisar más detalles del procesamiento, se recomienda ver el documento metodológico, apartado 7.3, disponible en: <https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/demografia-y-migracion>

¹⁰ Mesa de trabajo Inter agencial liderada por el INE y el Sermig, que busca orientar el trabajo mancomunado de las distintas instituciones participantes, en vista a identificar limitaciones y potencialidades respecto de la demanda de información sobre migración.

contar con una fuente información que permitiese estimar la cantidad de inmigrantes internacionales, de allí que surge como Operación Estadística esta Estimación de Personas extranjeras.

Desde su primera publicación en 2019 a la fecha, la estimación es usada con diversos propósitos, entre los que podemos destacar, el aporte a las políticas nacionales en materia migratoria tal y como quedó en evidencia en la Política Nacional de Migraciones¹¹. También, sirve de insumo para la construcción de estimaciones y proyecciones de población (al proveer información actualizada y continua de los flujos migratorios a partir de los registros administrativos integrados y sus características demográficas básicas¹²).

Por otro lado, esta estimación aporta información relevante en materia de estadísticas migratorias a nivel regional de América Latina en grupos de trabajo como el Grupo de Expertos en Migración Internacional de Naciones Unidas (del cuál Chile es miembro¹³).

4 PRECISIÓN Y CONFIABILIDAD

La estimación de personas extranjeras como operación estadística multifuente ha permitido aproximarse al estudio de la migración internacional de forma tal que entrega información actualizada en los periodos intercensales. Considerando que los censos de población y vivienda suelen realizarse cada 10 años, estas estimaciones permiten evidenciar posibles cambios en las tendencias y perfiles de la población inmigrante internacional en el país proveyendo insumos para la toma de decisiones, la generación de políticas migratorias¹⁴ y reportes internacionales¹⁵.

4.1 Evaluación censal

Respecto de controles de calidad de la recolección de datos la Memoria Censal señala que hubo controles de calidad en las diversas fases:

¹¹ Ver (Sermig 2024: 30) Disponible en <https://serviciomigraciones.cl/wp-content/uploads/2024/01/PNME-25-01-2024.pdf>

¹² Al momento de la elaboración de las proyecciones se contaba con información Censal y de registros del Sermig (entonces Departamento de Extranjería y Migración) Ver capítulo 3.3. Estimación y proyección de la migración internacional en documento: "Estimaciones y proyecciones de la población de Chile 1992-2050 total país metodología y principales resultados" ([enlace](#)) no obstante en las estimaciones y proyecciones con base al Censo 2024, se podrá utilizar información de registros integrados que provee la estimación.

¹³ Ver <https://unstats.un.org/unsd/demographic-social/sconcerns/migration/index.cshtml>

¹⁴ Ver Política Nacional de Migraciones de Chile ([enlace](#)).

¹⁵ Por ejemplo, para el Foro Internacional sobre Estadísticas de Migración Internacional (IFMS), o en grupos de trabajos tales como los asociados al Grupo de Expertos en Migración Internacional de Naciones Unidas del cual Chile es parte ([enlace](#)).

4.1.1 Controles al proceso censal

- Fases de levantamiento precensal y de actualización cartográfica (actividad pre-censal que se realiza para diseñar las cargas de trabajo): se generaron reportes a nivel central que fueron enviados a los encargados comunales para su revisión justificación e implementación de mejoras. Se resumen en la memoria los siguientes: Registro de vivienda colectivas, reportes de sectores sobre el total de viviendas y personas, reportes de viviendas con 7 o más hogares, reporte de inconsistencias entre cargas de plataformas tecnológicas, informes de cierre operativos con indicadores de viviendas esperadas y controles de calidad geográfica.
- Pruebas Piloto del Cuestionario Censal: se definieron cuatro pruebas efectuadas secuencialmente: las tres primeras conformaron un plan para la elaboración del cuestionario censal. La última, en cambio, consistió en un ensayo operativo del proceso de recolección del 19 de abril de 2017.
- Operativos del levantamiento en terreno: el personal de los centros operativos que trabajó el día del levantamiento censal, se hizo cargo de la distribución del material y registro de los censistas en cada local. Además, participó en la supervisión, recepción y revisión del material una vez concluido el trabajo del censista.
- Procesamiento: en esta etapa se realizaron validaciones de primer y segundo nivel. En la fase de lectura óptica el INE hizo control y seguimiento a la calidad, revisando las imágenes de los cuestionarios censales y las bases de datos, usando criterios como la coincidencia de portafolios y cuestionarios con la base de datos. En la fase realizada con la base de datos, el proceso de validación tuvo por objetivo revisar la validez y coherencia de los datos mediante la revisión de las preguntas del cuestionario censal (coherencia de las respuestas frente a lo definido como respuesta válida).

4.1.2 Tasa de no respuesta al ítem por variable

Las tasas de no respuesta de las preguntas del cuestionario censal, consideran el total de casos que no contestaron respecto al universo total de cada variable. Los universos difieren debido que el cuestionario tenía filtros y saltos conforme las subpoblaciones que debían contestar cada pregunta. En el caso de las variables P08 y P09, en el módulo de persona, se presenta la tasa de no respuesta final y la original. Esto se debe a que fueron las únicas variables a las cuales se les completó la información en los casos sin respuesta, debido a su relevancia para el cumplimiento de los objetivos del censo (INE, 2008, pág. 61). A continuación, se expone la tasa de no respuesta por variable respecto a la información proveniente del componente censal utilizadas en la estimación.

Respecto de la completitud de las variables de la base original del censo (previo a la estimación de personas extranjeras) para las variables de interés se tiene que en ningún caso supera el 4%. Las variables de edad y sexo son variables que previo a la integración tenían porcentajes de no respuestas inferiores al 0,2% y que para efectos de validación de todo el cuestionario censal se imputan hasta dejarlas completas lo que hace que aparezca como 0 en casos no declarado en la tabla 2. Adicionalmente, cabe señalar que dicho documento señala como universo de las preguntas P07, P09, P10 y P12 el total de población empadronada, no obstante, la estimación usa un universo menor siendo que se centra en residentes habituales y no en la población censada¹⁶.

Esta selección de variables del CPV2017 se explican porque: a) no es un objetivo de la estimación caracterizar por otras variables, b) el censo no tiene identificadores únicos para vincular a nivel de personas con los registros c) los registros administrativos usados en la estimación, actualmente, no están diseñados para medir la totalidad de información que los censos de población y vivienda tradicionales.

Tabla 1. No respuesta al ítem por variable de interés para la estimación. Censo de Población y Vivienda 2017

Variable	Total de casos no declarados	Tasa de no respuesta	Universo de la pregunta	tasa de no respuesta original
Sexo (P08)	-	0,00%	17.574.003	0,13%
Edad (P09)	-	0,00%	17.574.003	0,88%
Residencia habitual (P10)	202.508	1,15%	17.574.003	-
Residencia habitual (P10COMUNA*)	9.461	2,03%	466.271	-
Residencia habitual (P10PAIS*)	442	1,00%	44.303	-
Lugar de nacimiento (P12)	317.575	1,81%	17.574.003	-
Lugar de nacimiento (P12COMUNA*)	276.081	3,30%	8.108.952	-
Lugar de nacimiento (P12PAIS*)	3.960	0,50%	784.685	-

Nota: * Todas estas son preguntas abiertas en las que debía consignarse un texto en el cuestionario. Los casos no declarados incluyen los que debían haber contestado y no declararon nada, así como también los casos que declararon información, pero no pudieron codificarse

Fuente: adaptado de (INE, 2008, pág. 61)

¹⁶ Dependiendo el tipo de censo, se puede clasificar en censos de hecho (censos históricos de Chile) cuando se censa según el lugar donde la población estaba presente en un momento del tiempo originando un total de población censada, mientras que los censos de derecho (tal y como es el CPV24) la recolección se diseña con base al lugar de residencia habitual de las personas, las cuales podrían no estar presentes al momento de la entrevista. En los censos de hecho se suelen incluir preguntas sobre la residencia habitual de las personas, en el CEPV2017 el total de personas residentes 17.327.192, de los cuales 44.303 declararon residir en otro país y 202.508 no respondieron la pregunta.

4.2 Evaluación de las fuentes administrativas

Por su parte la tabla 2, muestra el resumen de variables contenidas en los registros administrativos utilizados en la integración de nivel. A partir de la tabla se observa que en el caso de las bases de datos de Mineduc y Minrel las variables de interés cuentan con tasas de completitud muy altas, considerando variables tanto de identificación como de caracterización.

4.2.1 Tasa de completitud de las variables administrativas

La tasa de completitud, es la razón entre el número de valores con datos y el número de valores totales requeridos de alguna variable específica. Este indicador es fundamental en el análisis de la calidad de los registros administrativos, pues a mayor nivel de completitud de los datos que forman parte de una variable, mayor es su precisión conforme se tendrá toda la información disponible para analizarse. El numerador corresponde al conjunto de valores con información para la variable de análisis. El denominador refiere al total de valores pertinentes de la variable (solo se consideran dentro del denominador aquellos valores en los que la presencia de información sea un requisito).

$$TCV = \frac{\text{Número de valores con datos}}{\text{Número de valores de datos pertinentes}} \times 100$$

En el caso de Sermig, que es la fuente de información administrativa con mayor cantidad de datos iniciales (10.961.105- ver anexo 3) y con mayor peso relativo en los resultados de la estimación¹⁷ la mayoría de las variables tienen altos porcentajes de completitud con valores superiores al 96% (ver tabla 2).

¹⁷ Ver documento metodológico apartado 8.3 procesamiento de registros administrativos.

Tabla 2. Tasa de completitud por variable del registro según fuente. Estimación de personas extranjeras 2023

Variable	Mineduc	Minrel	Sermig	SRCel	PDI
RUN			83,9	75,5	0,5
Pasaporte		100,0	97,3		
RUN_ALU (IPE)	100,0				
DNI			9,3		
Número del Doc. Identidad				25,5	4,6
COD EXT			100,0		100,0
Nombres	100,0	100,0	100,0	100,0	4,6
Apellidos	100,0	99,8	100,0	100,0	4,6
Sexo	100,0	100,0	100,0		
Fecha de nacimiento	100,0	100,0	100,0	100,0	4,6
País de origen	100,0		99,9		
País de nacionalidad		100,0	99,9	100,0	4,6
Comuna de residencia	100,0		96,3		100,0

Nota: las celdas vacías corresponden a los casos de variables que no aplican en la fuente.

Fuente: INE - Sermig

En el caso de la variable RUN, donde el porcentaje de completitud es 83,9%, su información es consistente con el tipo de registro administrativo y la población objetivo del mismo, ya que las personas extranjeras sujetas a realizar trámites en el Servicio no tienen RUN provisto por el SRCel hasta no tener aprobado un permiso de residencia. Por ejemplo, esto también ocurre con las personas en situación migratoria irregular que forman parte de la base de datos del Sermig sin ser solicitantes de permisos de residencia, pero habiéndose registrado en el empadronamiento biométrico. Misma situación ocurre con la variable DNI, que tiene el porcentaje de completitud más bajo con 9,3%, la cual no es obligatoria, aunque las personas extranjeras suelen entregar información como respaldo de su identidad.

Por su parte, en el caso de la base de fallecidos de SRCel el porcentaje de completitud de la variable RUN es 75,5%, mientras que la tasa de completitud del documento de identidad es 25,5%. En este registro también hay personas extranjeras que al momento de fallecer en el país no contaban con un RUN, en cuyo caso se registra el número de documento de identidad con el que contaba la persona difunta.

Finalmente, debe señalarse del registro de direcciones de la PDI que originalmente se recibía solo la variable de caracterización (comuna de residencia), al igual que en el caso del registro del control de fronteras. Esta característica se asignaba a los datos Minrel mediante la variable llave “COD EXT”, por lo que ambas variables tienen 100% de completitud en PDI. No obstante, en la Estimación 2022 la PDI envió información de las variables adicionales, las cuales cuentan

con porcentajes de completitud muy bajos (inferiores al 5%). Dichas variables adicionales son RUN, Número del Doc. Identidad, Nombres, Apellidos, fecha de Nacimiento y País de Nacionalidad (ver Tabla 3), siendo coincidentemente centrales para el análisis, aunque en el caso de los registros de PDI se utilizan sólo para efectos de la integración de la base de datos.

Este conjunto de indicadores se focaliza en evaluar los procesos iniciales, como los relativos a la base de datos o las variables provenientes de la oficina administrativa, así como la evaluación de los metadatos que la acompañan en los casos en que corresponda. Para el caso de la estimación de personas extranjeras, se realizan los cálculos considerando las bases de datos recibidas en el Sermig previo a la integración y procesamiento para la consolidación de la estimación de personas extranjeras.

Las variables denominadas en las bases de datos de interés para la identificación de personas y su posterior uso estadístico son las siguientes: RUN; Pasaporte; RUN_ALU (IPE); DNI; Número del Doc. Identidad; COD EXT; Nombres; Apellidos; Sexo; Fecha de nacimiento; País de origen; País de nacionalidad; Comuna de residencia.

4.2.2 Porcentaje de errores por variable del conjunto de datos

Es la razón entre la cantidad de errores por variable y el total de valores de la variable, expresada en porcentaje. Consiste en verificar el cumplimiento de los datos para su uso estadístico, validando que la información cumple los rangos, longitud y dominio previstos para la variable en su descripción.

$$PEV = \frac{\text{Número de valores con errores de la variable}}{\text{Número total de valores de la variable}} \times 100$$

Para la estimación de personas extranjeras se definieron como posibles errores los siguientes: en el caso de las variables RUN, Pasaporte, RUN_ALU, DNI y Numero de documento de identidad los errores son principalmente por longitud, aunque también pueden responder a por ejemplo que se haya rellenado solo con letras “sin sentido”, que denoten pruebas (ejemplo “AAA”, “Test”) o que se haya completado con números rápidos (ejemplo “00”, “1111111-1”). En el caso de los nombres y apellidos, que si bien son difíciles de definir como errores puesto que un mismo nombre puede tener diversas formas de escribirlo, se refiere a letras agrupadas “sin sentido aparente” o pruebas (ejemplo “AAA”, “Test”); en el caso del Sexo se refiere a casos que no hubiesen podido ser clasificables en la estimación como hombre-mujer, o masculino-femenino (pero no se encontraron errores de este tipo en ninguna de las bases); respecto de la fecha de nacimiento se definió como error la longitud de caracteres y que tenga o no el formato DD_MM_AAAA, y que correspondan a fechas poco probables o que

vienen de forma predeterminada en los sistemas de información (01-01-1900) o fechas en el “futuro” superiores al año 2024. Finalmente, tanto para país de origen como de nacionalidad se establece como error glosas que no denoten ningún país (migrante, extranjero).

A partir de la Tabla 3 se observa que la mayoría de las variables de identificación y caracterización tienen bajos porcentajes de error, situándose en valores inferiores al 0,6%. En oposición, la variable con mayor nivel de error corresponde al país de nacionalidad disponible en la base de datos de defunciones del SRCel, en cuyo caso el porcentaje alcanzó 26,3% considerando 3.995 observaciones sobre un total de 15.163. El error de glosa proviene principalmente por tener la palabra “extranjero”, siendo este un país no válido para los fines de uso estadístico en la Estimación.

Cabe señalar que, si bien el porcentaje de error en país de nacionalidad para las defunciones es el más alto de las variables expuestas, no tiene impactos en la base final de análisis para construir el total estimado, pues se trata de una variable de caracterización de la base de datos. De este modo, la variable país de nacionalidad de la base del SRCel no forma parte del filtro en la cuenta de residentes, por lo que se considera la información de país de nacionalidad a partir de la base principal (la del Sermig) en los casos en que la información se integra y no existe país de origen en el SRCel.

Tabla 3. Porcentaje de errores por variable del registro según fuente. Estimación de personas extranjeras 2023

Variable	Mineduc	Minrel	Sermig	SRCel	PDI
Total de casos en estudio	395.220	424.673	10.961.105	15.163	116.367
RUN			0,0	0,0	0,4
Pasaporte		0,0	0,0		
RUN_ALU (IPE)	0,0				
DNI			0,0		
Número del Doc. Identidad				0,1	0,0
COD EXT			0,0		0,0
Nombres	0,0	0,0	0,0	0,0	0,0
Apellidos	0,0	0,0	0,0	0,0	0,0
Sexo	0,0	0,0	0,0		
Fecha de nacimiento	0,2	0,6	0,0	0,0	0,0
País de origen	0,0		0,0		
País de nacionalidad		0,0	0,0	26,3	0,0
Comuna de residencia	0,0		0,0		0,0

Nota: las celdas vacías corresponden a los casos de variables que no aplican en la fuente.

Fuente: INE - Sermig

4.2.3 Porcentaje de observaciones duplicadas (POD)

Es la razón entre cantidad de observaciones duplicadas y el total de observaciones del conjunto de datos, expresada en porcentaje. Se utiliza para saber cuántas observaciones pertenecientes al conjunto de datos se detectan como casos duplicados. Para este indicador se toma como definición de un caso duplicado cuándo todos los valores en dos o más registros son idénticos, es decir, cada campo o columna en los registros tiene el mismo valor.

$$Pod = \frac{\text{Número de observaciones duplicadas de la base A}}{\text{Número total de observaciones de la base A}} \times 100$$

Quedan fuera de la conceptualización de caso duplicado, para la evaluación de la calidad de la base de datos, aquellos en que una unidad se encuentre presente más de una vez en una base de datos (o conjunto de datos) por medio del análisis de una variable identificadora. Dichos casos pueden corresponder a diferentes eventos en los que la unidad participa del acto administrativo, por lo que los valores para ciertas variables van a ser los mismos, pero para otras no. Por ejemplo, una persona usuaria del sistema público de salud, que acudió en diversas oportunidades a realizar exámenes médicos o controles, no corresponde a un caso duplicado en tanto cada prestación médica corresponde a un evento o acto totalmente diferente.

A continuación, se expone el porcentaje de observaciones duplicadas según fuente, considerando a cada institución involucrada en la elaboración de la Estimación:

Tabla 4. Porcentaje de observaciones duplicadas según fuente. Estimación de personas extranjeras 2023

Indicador	Mineduc	Minrel	Sermig	SRCel	PDI
POD	0,0	5,6	0,2	0,0	0,6

Fuente: INE - Sermig

De acuerdo con la Tabla 4, la mayoría de las fuentes tienen porcentajes de observaciones duplicadas inferiores al 5%, con excepción de los registros de visas consulares del Minrel, en donde se encontraron un total de 23.605 observaciones repetidas para un total de 424.673 filas (5,6%). Cabe destacar que estos registros se filtran y finalmente se descartan para el proceso de estimación de personas extranjeras, donde se identifican a las personas de forma única para su posterior análisis estadístico.

4.3 Evaluación del proceso de integración

Este set de indicadores evalúa la calidad de los datos obtenidos desde el procesamiento de los registros administrativos utilizados con fines estadísticos, tanto de los procesos de integración entre bases como potenciales acciones que permitan dar cuenta de la precisión de la información producida.

La estimación de personas extranjeras es un proceso estadístico multifuente, debiendo vincular sus datos a través de identificadores personales, indeterminar a las personas en los resultados finales a través de métodos de anonimización y comparar dichos resultados con otras fuentes. Por un lado, el indicador de la tasa de vinculación entre las bases de datos resulta de gran relevancia porque permite evaluar la similitud o diferencia del origen de los datos de cada base y la calidad de las variables de identificación utilizadas para relacionar las bases de datos. En segundo lugar, el indicador del uso de métodos de anonimización permite asegurar la protección de datos personales de los individuos al momento de difundir los resultados de la estimación. Por último, en la fase de análisis se plantea la comparación con otras fuentes para evaluar el grado de coherencia y dar cuenta de las razones de posibles diferencias. Los indicadores de proceso, a diferencia de la sección anterior y dada su pertinencia, fueron calculados solo para el año de la estimación¹⁸, que en este caso es el 2023.

Es la razón de unidades vinculadas desde un conjunto de datos “a” a otro conjunto de datos base “b”, expresado en porcentaje. Este indicador mide efectivamente el resultado de la vinculación entre dos conjuntos de datos distintos, siendo una medida muy importante de la calidad en el uso estadístico de los registros administrativos. (INE, 2024, pág. 20).

$$PEa \rightarrow b = \frac{\text{Núm. unidades vinculadas de la base A a la base B}}{\text{Núm. total de unidades de la base A que se espera se vinculen a la base B}} \times 100$$

Base A: conjunto de datos que se requiere vincular con B.

Base B: conjunto de datos al que se le vincula la información de A.

Este indicador se divide en cada una de las vinculaciones realizadas durante el preprocesamiento para la deduplicación y consolidación del componente postcensal (Sermig + Minrel + Mineduc) y el procesamiento final, es decir, la vinculación con los registros que sirven para caracterizar o dar la condición de residentes (PDI).

Una baja tasa de vinculación es indicativa de que los conjuntos de datos cubren diferentes partes de la población o responden a distintas poblaciones objetivo; que el proceso de

¹⁸ Información sobre re-estimaciones históricas puede consultarse en el documento metodológico.

vinculación no identifica todas las conexiones que existen entre las unidades en cada conjunto de datos o reflejar problemas de calidad referente a las características de la variable de vinculación, las cuales podrían requerir edición previa. Este indicador está orientado para su uso cuando sea necesario vincular uno o más conjuntos de datos con un registro considerado como base. Idealmente, la calidad del proceso de vinculación dependerá del estado y características de la variable “llave” o de identificación, en cuanto a sus características (INE, 2024, pág. 20).

Pre-procesamiento de base postcensal: vinculación de registros para identificar personas únicas hasta consolidar una base de personas. El proceso de deduplicación con el detalle de los datos iniciales y finales se detalla en el documento metodológico¹⁹.

Respecto de la vinculación, en la estimación de personas extranjeras el componente de registros administrativos comprende principalmente a la base del Sermig, pues es la institución migratoria encargada por Ley de los procesos administrativos asociados a la residencia de los(as) migrantes. Sin embargo, en la práctica las personas migrantes pueden vincularse con el Estado chileno mediante otros actos administrativos, con independencia de una situación migratoria regular. La vinculación con otras bases de datos tiene por objetivo considerar y cubrir a distintos subgrupos poblacionales que podrían en un inicio no estar contenidos en los datos del Sermig.

La primera vinculación se realizó entre la base del Mineduc y la base del Sermig, donde la tasa de vinculación para el año 2023 es de 62,2%:

$$PE_{\text{Mineduc-Sermig}}: \frac{121.971}{196.178} \times 100 = 62,2\%$$

Cabe destacar que como los registros de Mineduc cumplen la función de reportar casos de personas en posible situación irregular y sin trámites en Sermig, resulta esperable que el indicador no llegase a 100%. En este indicador se obtuvo un porcentaje de vinculación de 62,2%, permitiendo depurar los datos y agregar a una parte de la subpoblación de personas extranjeras en situación irregular a la Estimación. Cuando los datos se encuentran vinculados (presentes en ambas bases), se prioriza y quedan asociados con la fuente de información Sermig (en la base de datos dentro del componente regular), mientras que los datos que no se vinculan quedan asociados con la fuente de información Mineduc, agregando personas en el componente irregular.

Posteriormente, se toma la base de datos correspondiente a los permisos de residencia fuera de Chile, cuyas fuentes son Minrel y Sermig²⁰, y se vinculan con datos de Sermig más los casos Mineduc, obteniendo una tasa de vinculación de 48,7%.

¹⁹ Ver documento metodológico en: <https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/demografia-y-migracion>

$$PE_{\text{Minrel-Mineduc-Sermig}}: \frac{82.564}{169.529} \times 100 = 48,7\%$$

Los registros del Minrel entregan información de personas cuyos actos administrativos inicialmente se gestionaron en el extranjero en un consulado o embajada. No obstante, algunas personas realizan trámites en el Sermig sin haber realizado trámites en el Minrel, por lo que una vinculación del 100% tampoco era esperada. Esta vinculación también permite deduplicar individuos para consolidar una base de datos de personas únicas sobre las cuales se espera definir la condición de residencia para ser sumadas o no en la estimación total de personas extranjeras.

Cabe destacar que existen diferencias importantes cuando se trata de registros que informan la situación de personas en situación migratoria regular frente a las irregulares. Dichas diferencias se presentan porque las personas en situación migratoria irregular generalmente no cuentan con RUN o sus datos personales de identificación son escasos (variables usadas como llaves para vincular). También se explican estas diferencias porque la condición de irregularidad está dada por no tener actos administrativos en frontera (por ejemplo, al haber ingresado por pasos clandestinos) o no ser solicitantes de permisos de residencia (no aparecer en bases de Sermig o Minrel).

Procesamiento final: vinculación de registros de la base postcensal de personas únicas con registros de entradas y salidas al país.

Una vez con la base consolidada sin duplicados entre Sermig, Minrel y Mineduc (consolidado postcensal), se envía a la Policía de Investigaciones para completar con información de la fecha del último viaje y rumbo (entrada o salida). A partir de esta caracterización se desprende que se vincula el 76,5% de los datos²¹.

$$PE_{\text{postcensal-PDI}}: \frac{1.195.669}{1.563.363} \times 100 = 76,5\%$$

El porcentaje de vinculación de las bases de datos completas no puede llegar a 100% en la medida que el producto estadístico de la estimación de personas extranjeras considera una parte de la población que está en situación irregular e ingresó al país por un paso no habilitado, es decir, no registra viaje en los controles de frontera regulares de la PDI.

²⁰ El 12 de febrero de 2022 se publicó la actual Ley de Migraciones y Extranjería (Ley 21.325), que establece que los permisos de residencia solicitados en el exterior serán tramitados por Sermig. Por ende, la tramitación de residencias para personas migrantes que se encuentran fuera del país es responsabilidad del Sermig desde mayo de 2022.

²¹ Ver Sección “8.3 Procesamiento de los registros administrativos” del Documento Metodológico, en: <https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/demografia-y-migracion>
En dicha Sección se detallan los procesos de vinculación entre bases de datos en la fase inicial (previo a la cuenta o descuento de residentes).

5 ACCESIBILIDAD Y CLARIDAD

La estimación de personas extranjeras 2023 pone a disposición de los usuarios tanto en la página web del INE²² como la del Sermig²³ un set de archivos que permitan a diversos tipos de usuarios (técnicos especializados, organismos públicos y privados, organismos internacionales y ciudadanía) acceder a la información, ya sea mediante la base de datos de nivel regional, como cuadros estadísticos, informes de resultados o síntesis.

Para garantizar la protección de la información estadística que refiere a las personas extranjeras se utilizan controles de anonimización descritos a continuación.

5.1 Utilización de métodos de Anonimización

La anonimización corresponde a un proceso técnico que consiste en transformar los datos individuales de las unidades de observación, de tal modo que no sea posible identificar sujetos o características individuales de la fuente de información, preservando así las propiedades estadísticas en los resultados. Según lo anterior, este indicador mide que esta actividad se haya realizado para garantizar la privacidad de personas o instituciones (personas naturales y empresas, por ejemplo).

Para este indicador corresponde como resultado el cumplimiento ideal ya que se realiza el proceso de anonimización de las unidades de observación. Esto es así debido a que una vez se identifican las personas para deduplicarlas y vincular las bases de datos, se remueven todas las variables de identificación de la base de datos de trabajo (RUN o cualquier número de identidad, nombres y apellidos). Adicionalmente, en las bases de datos públicas se entrega información a niveles geográficos agregados, siendo en este año el nivel de Región. De este modo, se garantiza la indeterminación de las personas a partir de métodos indirectos gracias a la agrupación de sus características sociodemográficas.

6 COHERENCIA Y COMPARABILIDAD

6.1 Grado de coherencia con otras Fuentes

Es el grado de comparación entre los resultados obtenidos del análisis del registro administrativo con otras fuentes de información. El nivel de coherencia busca conocer en

²² Ver: <https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/demografia-y-migracion>

²³ Ver <https://serviciomigraciones.cl/estudios-migratorios/analisis-sermig/>

qué términos, los resultados provenientes del uso del registro son válidos si los comparamos con fuentes externas al registro utilizado, las que pueden corresponder a otros registros y/o a resultados de encuestas o censos. Este indicador no tiene como propósito que los resultados del registro utilizado y de la fuente externa sean los mismos, sino que se busca garantizar que la interpretación de ambas fuentes de datos sea coherente (INE, 2024, pág. 22).

“Aunque el grado de coherencia es una medida cualitativa y no genera un resultado o valor cuantitativo, es importante resguardar y realizar este tipo de análisis para garantizar y asegurar la validez, sobre todo cuando el uso reviste de alta complejidad” (INE, 2024, pág. 22).

En relación con este indicador puede decirse que el resultado es “cumplimiento ideal: se realiza el análisis con otras fuentes externas y la comparación es coherente con el registro administrativo en términos de tendencias, magnitudes u otros análisis pertinentes” (INE, 2024, pág. 23). En particular, la mayoría de los datos son coherentes y cuando existen diferencias se explican por la complejidad del fenómeno migratorio y las limitaciones de medición que tiene cada fuente de datos.

La estimación de personas extranjeras tiene limitaciones propias del trabajo con una metodología de análisis multifuente, las cuales se encuentran desarrolladas en el documento metodológico²⁴. Otros productos con los que se podrían comparar los resultados de la estimación también tienen limitaciones, como por ejemplo las limitaciones derivadas del diseño muestral en las encuestas de hogares. Específicamente, este tipo de operaciones no incluyen en su diseño la apertura de colectivos como dominios de estudios, esto ya que en general no se cuenta con marcos muestrales que permitan garantizar la presencia de estos, sino que esto estará condicionado por el azar y la presencia de estos en la población total. Asimismo, es posible que la forma en que se mide el fenómeno difiera en alguna medida de la cómo se hace en censos o como se capturan los registros administrativos. Por lo tanto, no es adecuado comparar los resultados específicos obtenidos por cada operación estadística, sino que corresponde sólo revisar si la interpretación de resultados entre ambas es coherente.

Para efectos de este indicador, la comparación se realizó con la Encuesta Nacional de Empleo (ENE) y la Encuesta de Caracterización Socioeconómica Nacional (Casen). En el caso de las encuestas, aunque no tienen un marco o diseño enfocado en medir a la población migrante, dada la incorporación de variables como lugar de nacimiento (residencia de la madre) o nacionalidad, sí es posible cotejar diferencias entre dichas fuentes. No obstante, las

²⁴ Ver capítulo 6 del documento metodológico en: <https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/demografia-y-migracion>

diferencias factibles de cotejar se enfocan en la estructura más que en la población total, nacionalidad o país de origen.

Encuesta Nacional del Empleo (ENE)

Al comparar la estimación de personas extranjeras con la Encuesta Nacional de Empleo (ENE) considerando el trimestre móvil de noviembre-diciembre-enero para el periodo 2018-2023, se observa que la cifra de la estimación se encuentra por sobre el nivel superior del intervalo de confianza para los años 2022 y 2023. En 2020 la estimación estuvo levemente por arriba de la estimación puntual de la ENE, presentando la menor diferencia del periodo con 29.048 personas y una diferencia de 1,9%. La mayor diferencia del período se observa en 2023²⁵ donde la estimación de personas extranjeras presenta 342.292 personas más que la estimación puntual de la encuesta, representando una diferencia de 17,8% (Ver Tabla 5).

Tabla 5. Diferencias absolutas y relativas de la población extranjera estimada según fuente y año, 2018-2023

	2018	2019	2020	2021	2022	2023
Estimación	1.306.859	1.486.810	1.534.813	1.630.812	1.836.137	1.918.583
ENE (Trimestre nov-dic-enero)	1.251.869	1.415.977	1.505.765	1.548.717	1.567.943	1.576.291
Límite inferior del intervalo	1.103.226	1.283.426	1.391.151	1.441.838	1.422.526	1.467.179
Límite superior del intervalo	1.400.512	1.548.528	1.620.379	1.655.596	1.713.360	1.685.403
Dif. Absoluta	54.990	70.833	29.048	82.095	268.194	342.292
Dif. Relativa %	4,2%	4,8%	1,9%	5,0%	14,6%	17,8%

Fuente: Estimación de personas extranjeras INE-Sermig 2023 y ENE para el trimestre móvil noviembre-diciembre-enero de cada año (diciembre mes central).

Adicionalmente, a pesar de que el diseño de la ENE no considera en sus objetivos la apertura por colectivo, se consideró relevante la revisión de la estructura de los principales colectivos que en la encuesta muestran estimaciones confiables y comparar con la estimación INE-Sermig 2023, considerando el orden respecto del aporte al total de personas extranjeras en la encuesta y la estructura más que las estimaciones de totales.

La ENE tiene una mayor concentración de personas venezolanas que representan el 46,8% del total, 8,9 puntos porcentuales (p.p.) mayor que en la Estimación, no obstante, la estimación de migrantes se encuentra dentro del intervalo de confianza de la estimación de la ENE. En la Tabla 6 se presenta el detalle de las comparaciones entre la estimación de personas extranjeras residentes y la ENE.

²⁵ La Encuesta Nacional del Empleo calibra sus factores de expansión a la proyección que se construyó a partir del componente de migración neta de las proyecciones de población base censo 2017. Para más antecedentes ver <https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/documentos-de-trabajo/documentos-de-trabajo/poblacion-extranjera-ene.pdf>

Tabla 6. Diferencias de la población extranjera estimada para 2023 según fuente y país de nacionalidad

País de nacionalidad	INE-Sermig (2023)		Encuesta Nacional del Empleo anual (2023)					Variación p.p. ²
	Total estimado	Distribución % (a)	Total Estimado	Distribución % (b)	Límite inferior del intervalo	Límite superior del intervalo	Calidad de estimación ¹	(a)-(b)
Total	1.918.583	100	1.572.984	100	1.452.805	1.693.163	Confiable	--
Venezuela	728.586	38,0	736.865	46,8	649.189	824.540	Confiable	-8,9
Perú	260.785	13,6	225.431	14,3	200.742	250.120	Confiable	-0,7
Colombia	209.946	10,9	209.194	13,3	186.235	232.154	Confiable	-2,4
Haití	188.131	9,8	75.355	4,8	59.007	91.704	Confiable	5,0
Bolivia	180.266	9,4	144.407	9,2	119.677	169.138	Confiable	0,2
Argentina	83.265	4,3	33.364	2,1	29.291	37.437	Confiable	2,2
Ecuador	50.460	2,6	42.685	2,7	35.620	49.750	Confiable	-0,1
Rep. Dominicana	22.836	1,2	15.730	1	11.630	19.830	Confiable	0,2
Cuba	21.305	1,1	14.805	0,9	11.343	18.266	Confiable	0,2
Brasil	21.272	1,1	11.867	0,8	8.959	14.776	Confiable	0,4
Otro país	136.952	7,1	63.281	4	56.349	70.212	Confiable	3,1
Ignorado	14.779	0,8						

Notas:

1. En base al estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares (INE, 2020).

2. p.p (Puntos porcentuales).

Fuente: INE-Sermig 2023: Estimación de personas extranjeras; INE 2023, Encuesta Nacional de Empleo (ENE) base anual.

Encuesta Casen 2022

Al igual que la ENE, la CASEN no considera en su diseño la apertura de colectivos de personas nacidas en el extranjero²⁶, ni posee ajustes de calibración por población extranjera al momento de ajustar su factor de expansión. No obstante, es relevante poder contrastar la estructura y el aporte al total al menos para aquellos colectivos que presentan estimaciones confiables.

Respecto de la población en la Encuesta Casen 2022 y el valor reestimado de 2022 a partir de la Estimación 2023, se observa una diferencia en el total de cerca de 100 mil personas, la cual se encuentra dentro del intervalo de confianza de la estimación de la encuesta. Por otro lado, la encuesta tiene una mayor concentración de personas venezolanas que representan el 49,2% del total, lo que se traduce en 12 puntos porcentuales mayor que en la Estimación. No

²⁶ Para definir a los inmigrantes internacionales la Casen 2022, al igual que el censo 2017, ocupa la pregunta por país de nacimiento la que se operativiza en el cuestionario así: r1b. Cuando %nombre% nació, ¿en qué comuna o país vivía su madre? Identificando a las personas que responden la categoría 3. "En otro país".

obstante, la Casen presenta una menor cantidad de personas haitianas, con alrededor de 80 mil personas y 4,2 puntos porcentuales menos que lo estimado por INE-Sermig (ver Tabla 7).

Tabla 7. Diferencias de la población extranjera estimada para 2022 según fuente y país de nacimiento

País de nacimiento	INE-Sermig (2022 ¹)		CASEN 2022					Variación p.p. ³
	Total estimado	Distribución (a)	Total Estimado	Distribución (b)	Límite inferior del intervalo	Límite superior del intervalo	Calidad de estimación ²	(a)-(b)
Total	1.836.137	100,0	1.736.691	100,0	1.608.078	1.865.304	Confiable	--
Venezuela	683.273	37,2	854.112	49,2	764.898	943.326	Confiable	-12,0
Perú	257.520	14,0	235.192	13,5	191.949	278.435	Confiable	0,5
Colombia	203.142	11,1	204.998	11,8	176.833	233.163	Confiable	-0,7
Haití	188.633	10,3	105.999	6,1	86.631	125.367	Confiable	4,2
Bolivia	163.905	8,9	118.288	6,8	104.187	132.389	Confiable	2,1
Argentina	80.936	4,4	60.525	3,5	52.590	68.460	Confiable	0,9
Ecuador	47.632	2,6	43.938	2,5	32.519	55.357	Confiable	0,1
Otro país ⁴	211.096	11,5	113.639	6,5	97.158	130.120	Confiable	5,0

Notas:

1. Reestimación realizada en 2023.

2. En base al estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares (INE, 2020).

3. p.p. (Puntos porcentuales).

4. Incluye no responde/país ignorado

Fuente: Casen 2022, Estimación de personas extranjeras INE-Sermig 2023.

Estas diferencias se explican, posiblemente, por las distinciones metodológicas entre fuentes. Por ejemplo, la Casen, al no estar calibrada por población extranjera, está sujeta a la presencia de ésta en la muestra, por lo que no existe un ajuste en la encuesta asociado a personas extranjeras. En el caso de la ENE, que sí está calibrada por población extranjera, dicha calibración no tiene en consideración otras variables de la población extranjera, como su sexo, país de nacionalidad o distribución geográfica, por lo que el diseño de la muestra en conjunto con el proceso de recolección juega un papel importante en los resultados²⁷. Adicionalmente, para la ENE no se publica la información de la población extranjera con

²⁷ Es relevante mencionar que no existe información sistemática y desagregada respecto a la población extranjera que pudiese ayudar a calibrar la encuesta con la periodicidad que esta requiere.

desagregaciones para todas las variables disponibles de la base de datos dado que se analizan criterios de cumplimiento de los estándares de calidad estadística²⁸ del INE.

En ambas encuestas los diseños estadísticos no siempre consideran a la población migrante, pues no es un objetivo principal de las mismas. Adicionalmente, la organización y concentración de los migrantes por colectivo en el territorio podría incurrir que en las manzanas seleccionadas existan eventuales diferencias respecto a la distribución de la población migrante en todo el territorio.

7 CONCEPTOS CLAVES

Integración de datos: *“Este subproceso integra los datos de una o más fuentes. Es en donde los resultados de subprocesos de la fase de ‘Recolección’ se combinan. Los datos que entran pueden provenir de una mezcla de fuentes internas o externas así como una variedad de modalidades de recolección, incluyendo extractos de datos administrativos. El resultado es una serie de datos vinculados. La integración de los datos puede incluir:*

- *La combinación de datos de múltiples fuentes como parte de la creación de estadísticos integrados como cuentas nacionales.*
- *Emparejar o registrar vínculos habituales o de rutina con el objetivo de vincular micro y macrodatos de diversas fuentes.*
- *Priorizar, cuando dos o más fuentes contienen datos de la misma variable, con valores potencialmente distintos”. (UNECE, 2016, pág. 23)*

Integración de datos nivel Micro: *“la integración de datos basada en vinculación de registros o coincidencia estadística de registros de nivel individual utilizando variables de identificación clave.” (UNECE, 2019, pág. 5)*

Integración de datos nivel Macro: *“la combinación de datos basada en datos agregados (estadísticas) de registros a nivel individual”. (UNECE, 2019, pág. 5)*

Deduplicación de registros²⁹: La deduplicación es un proceso, masivo y/o manual, que permite eliminar copias o duplicados de datos o información. Esta técnica se utiliza en

²⁸ Se debe considerar que las desagregaciones de los indicadores o aperturas de los datos pueden traer consigo problemas de precisión estadística o, en casos más críticos, estar sesgadas. Por tanto, se sugiere tomar en cuenta los criterios de calidad estadística que el INE ha definido en el documento [“Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares”](#), con la advertencia de que, a mayor desagregación, el sesgo tiende a aumentar a propósito de la pérdida de unidades muestrales.

²⁹ <https://www.arimetrics.com/glosario-digital/deduplicacion>

diversas áreas, como en las bases de datos, sistemas de almacenamiento e incluso en el correo electrónico.

Al hacer uso de la deduplicación, se consigue reducir el tamaño de los archivos y ahorrar espacio en el disco duro o memoria de almacenamiento. Además, esta técnica permite eliminar errores que se pueden presentar al tener varias copias de un mismo archivo, evitando confusiones e inconvenientes.

El uso que se le da a este proceso en la estimación es la de identificar personas que no se repitan entre sí, tanto al interior de una sola fuente de registro administrativo, como en la integración de distintas fuentes de registro administrativo.

Población residente: *“hace referencia a todas las personas que: a) han vivido en el país de medición durante la mayor parte de los 12 meses anteriores o tienen intención de permanecer en él (o permiso para permanecer en él) durante al menos 6 meses; o b) han vivido en el país de medición durante al menos 12 meses en un año determinado o tienen intención de permanecer en él (o permiso para permanecer en él) durante al menos 12 meses, sin contar las ausencias temporales por vacaciones o motivos laborales”* (Naciones Unidas, 2020, pág. 18) . En la estimación de personas extranjeras 2023 INE-Sermig, se incluyen personas que tienen un estatus migratorio irregular, que no registran una solicitud de permiso de residencia y que llevan residiendo en el país al menos 6 meses durante el año de medición, lo cual se evidencia mediante la fecha de su acto administrativo frente a la respectiva institución, tal es el caso del Sermig, Minrel, Mineduc o PDI³⁰.

³⁰ Si bien el concepto original no distingue sobre el estatus migratorio de las personas, en este concepto hemos incluido la mención a la condición de irregularidad con base a los criterios metodológicos de la estimación 2023, en donde se incorporan casos de personas que tienen un trámite vencido (prórroga de turismo) o registro administrativo (informe policial, empadronamiento biométrico y matrícula escolar), las cuales pueden o no tener información del viaje migratorio en la base de control de fronteras, pero se asume que se mantienen en el país desde la fecha de dicho trámite.

8 ANEXOS

Anexo 1. Insumo para el cálculo de indicadores. Base inicial Mineduc

Institución	Ministerio de Educación	
Fuente de datos	Registros administrativos de la matrícula única Oficial (estudiantes extranjeros)	
Período de tiempo	De 2017 a 2023	
N	395.220	
	Vacíos	Errores
RUN		
Pasaporte		
RUN_ALU (IPE)	0	0
DNI		
Número del Doc. Identidad		
COD EXT		
Nombres	0	0
Apellidos	0	0
Sexo	0	0
Fecha de nacimiento	0	729
País de origen	0	0
País de nacionalidad		
Comuna de residencia	0	0
N° de obs. duplicadas	0	

Fuente: Sermig

Anexo 2. Insumo para el cálculo de indicadores. Base inicial Minrel.

Institución	Ministerio de Relaciones Exteriores	
Fuente de datos	Registros administrativos de las visas consulares	
Período de tiempo	2017 a 2023	
N	424.673	
	Vacíos	Errores
RUN		
Pasaporte	2	46
RUN_ALU (IPE)		
DNI		
Número del Doc. Identidad		
COD EXT		
Nombres	82	31
Apellidos	985	31
Sexo	0	0
Fecha de nacimiento	14	2422
País de origen		
País de nacionalidad	5	0
Comuna de residencia		
N° de obs. duplicadas	23.605	

Fuente: Sermig

Anexo 3. Insumo para el cálculo de indicadores. Base inicial Sermig

Institución	Servicio Nacional de Migraciones	
Fuente de datos	Registros administrativos de los trámites migratorios	
Período de tiempo	1990 a dic 2023	
N	10.961.105	
	Vacíos	Errores
RUN	1.759.699	196
Pasaporte	295.697	0
RUN_ALU (IPE)		
DNI	9.936.748	0
Número del Doc. Identidad		
COD EXT	2.344	0
Nombres	63	6
Apellidos	496	18
Sexo	-	0
Fecha de nacimiento	866	5.357,00
País de origen	14.139	0
País de nacionalidad	13.510	0
Comuna de residencia	400.713	0
N° de obs. duplicadas	19.954	

Fuente: Sermig

Anexo 4. Insumo para el cálculo de indicadores. Base inicial SRCel

Institución	Servicio de Registro Civil e Identificación	
Fuente de datos	Registros administrativos de las defunciones en el país	
Período de tiempo	2018 a 2023	
N	15.163	
	Vacíos	Errores
RUN	3.720	-
Pasaporte		
RUN_ALU (IPE)		
DNI		
Número del Doc. Identidad	11.289	8
COD EXT		
Nombres	-	-
Apellidos	-	-
Sexo		
Fecha de nacimiento	4	-
País de origen		
País de nacionalidad	-	3.995
Comuna de residencia		
N° de obs. duplicadas	0	

Fuente: Sermig

Anexo 5. Insumo para el cálculo de indicadores. Base inicial PDI

Institución	Policía de Investigaciones	
Fuente de datos	Registro de residentes	
Período de tiempo	De 2020 a 2022	
N	116.367	
	Vacíos	Errores
RUN	115.737	415
Pasaporte		
RUN_ALU (IPE)		
DNI		
Número del Doc. Identidad	111.069	1
COD EXT	-	0
Nombres	111.068	0
Apellidos	111.068	0
Sexo		
Fecha de nacimiento	111.068	0
País de origen		
País de nacionalidad	111.068	0
Comuna de residencia	-	0
N° de obs. duplicadas	731	

Fuente: Sermig

9 REFERENCIAS

- INE. (2008). *Manual de usuario de la base de datos censo 2017*. Obtenido de https://www.ine.gob.cl/docs/default-source/censo-de-poblacion-y-vivienda/bbdd/censo-2017/manual-de-usuario/manual_de_usuario_censo_2017_16r.pdf?sfvrsn=38710602_4
- INE. (2018). *Memoria del Censo 2017*. Obtenido de https://www.ine.gob.cl/docs/default-source/censo-de-poblacion-y-vivienda/publicaciones-y-anuarios/2017/memoria-del-censo-2017/libro_memoria_censal_2017_final.pdf?sfvrsn=2f7aa860_6
- INE. (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/buenas-practicas/estandares/estandar/documento/estandar-evaluacion-de-calidad-de-estimaciones.pdf>
- INE. (Marzo de 2024). *Guía de indicadores para la evaluación de la calidad de los registros administrativos* . Obtenido de https://www.ine.gob.cl/docs/default-source/buenas-practicas/directrices-metodologicas/medicion-de-la-calidad/registros-administrativos/gu%C3%ADa-de-indicadores-para-evaluar-calidad-de-los-registros-administrativos.pdf?sfvrsn=3bd3b09a_6