



Subsecretaría de Evaluación Social
División Observatorio Social

**METODOLOGÍA PARA LA CONSTRUCCIÓN DEL UNIVERSO DE POBLACIÓN PARA EL CÁLCULO DE
INDICADORES DE CARACTERIZACIÓN SOCIOECONÓMICA CON REGISTROS ADMINISTRATIVOS**

Índice

1.	INTRODUCCION.....	4
2.	OBJETIVO CONSTRUCCIÓN UNIVERSO EN BASE A REGISTROS ADMINISTRATIVOS.....	4
3.	CONSTRUCCIÓN DE BASES INTERMEDIAS Y FINALES	5
	i. Universo de RUN válidos consolidado por el Ministerio de Desarrollo Social y Familia..	5
	ii. Universo de RUN válidos emitidos por el Registro Civil.	5
	iii. Universo de RUN válidos para cada período (año o semestre), descartando personas fallecidas, con edad superior a 110 años y sin información en la variable sexo	6
	iv. Base final: Base Residentes	7
	v. Base final: Base Residencia Comunal	8
4.	PROCESO #1: DETERMINACIÓN DE RESIDENCIA EN EL PAÍS	9
	a. Base Registro Civil: generación y limpieza de variables intermedias.	9
	b. Bases Adicionales:	10
	i. Generación y limpieza de variables intermedias.	10
	ii. Generación de umbrales de residencia	11
	c. Elaboración de Base Residente: Ratificación de personas que residen en el país	11
	d. Elaboración de Base Residente: Imputación de residencia según criterio de parentesco	12
	e. Validación de residentes	12
5.	PROCESO #2: DETERMINACIÓN DE COMUNA PARA BASE RESIDENTE	17
	a. Homologación de códigos según división política administrativa.....	18
	b. Identificación de comuna para cada base de datos	18
	c. Clasificación de bases de datos según calidad de información comunal	19
	i. Clasificación de Base Comuna: Anual.....	19
	ii. Clasificación de Base Comuna: Primer Semestre.....	19
	iii. Clasificación de Base Comuna: Segundo Semestre.....	19
	d. Elaboración de Base Comuna	20
	e. Elaboración de Base Moda.....	20
	f. Unión a Base Residente y determinación de comuna de residencia para grupo no identificado mediante información de Registro Civil	20
	g. Imputación de la comuna de la madre a los hijos/as residentes de 10 o menos años.....	21
	h. Resultados comunales.....	21

i. Validación resultados comunales	21
6. FUENTES DE DATOS UTILIZADAS	25
Anexo N°1: Indicadores de movimiento usados para identificar condición de residencia en el país	27
Anexo N°2: Clasificación por grupos de base de datos con información comunal para Base Residente 2019	28

1. INTRODUCCION

La construcción de un marco de población es fundamental para la estimación de indicadores, tanto para aquellos que se calculan a partir de encuestas como para los indicadores derivados de registros administrativos. Un marco o universo de población es útil para controlar el sesgo que se puede generar en la dimensión demográfica de las estimaciones. Por ejemplo, el crecimiento de la población producto de un aumento inesperado de la inmigración podría afectar el cálculo de indicadores de ocupación o ingresos.

La construcción del marco poblacional de residentes contempla tres etapas generales que se inician con la información administrativa generada por el Registro Civil e Identificación que incluye además del Rol Único Nacional (RUN) válido, información demográfica básica como las fechas de nacimiento y defunción, además del sexo de las personas. La etapa siguiente consiste en recurrir a las bases administrativas para determinar si las personas se encuentran o no residiendo en el país. Para ello se asume que una persona se encuentra en el país si muestra algún registro o movimiento en un set de catorce fuentes distintas de datos administrativos. La última etapa consiste en la identificación de la comuna de residencia habitual haciendo nuevamente uso de un subconjunto de cuatro fuentes de información administrativas que contienen datos al respecto.

Tras la aplicación de este procedimiento de identificación se procede a validar los resultados comparando con las proyecciones de población respectivas, las distribuciones sexo-edad alcanzadas. Por ejemplo, los residentes identificados en el año 2019 ascienden a 18.591.661 personas, los que, en relación con la población proyectada del mismo año de 19.107.216, representa un 97,3% de cobertura. A partir del total de residentes identificados, se logra establecer una comuna de residencia para 97,8% de dicha población, vale decir, para 18.184.493 personas.

El documento está organizado en dos partes que describen los procedimientos realizados para la determinación de la residencia y la identificación de la comuna de residencia habitual. En la primera parte se detalla el proceso de residencia, las fuentes de información y los umbrales de identificación definidos para determinarla. La segunda parte se refiere al proceso de determinación de la comuna de residencia, en ella se detallan los criterios de calidad utilizados para la clasificación y prelación de sus registros, entre otros aspectos.

2. OBJETIVO CONSTRUCCIÓN UNIVERSO EN BASE A REGISTROS ADMINISTRATIVOS

El objetivo principal del proyecto es la elaboración de una base de datos que contenga información de los individuos que residen en el país para el periodo t , tanto a nivel anual como semestral, entregando información demográfica, así como también información sobre el territorio en donde residen.

Esta base cumple una función de ser el universo válido considerado para el país en el periodo que se encuentre analizando. A partir de esto, la base podrá ser utilizada en conjunto a bases de datos específicas al tema que se quiera indagar, para poder elaborar indicadores de distinta índole, los cuales podrán desagregarse en distintos niveles, según sea necesario.

Uno de los principales requerimientos para elaborar indicadores estadísticos es contar con un universo válido de referencia. En el caso de las encuestas éste se construye por medio de marcos muestrales que se basan en información censal y, en el caso de registros administrativos, éste es factible de construir a partir de información proporcionada por el Registro Civil y otras fuentes administrativas.

Para el logro del objetivo se trabajó con una base de datos proporcionada por parte de Registro Civil a la cual, mediante el complemento de otras bases de datos, se identificó a las personas que para el periodo en específico se encuentren residiendo en el país.

3. CONSTRUCCIÓN DE BASES INTERMEDIAS Y FINALES

Distintos términos que se usarán a lo largo del documento son claves para el diseño e implementación de la presente metodología. Estos se refieren a definiciones del universo de población a generar, los conceptos de movimientos, de residencia y comuna de residencia habitual, así como también a los diferentes nombres que se les asignan a bases intermedias que se utilizan.

i. Universo de RUN válidos consolidado por el Ministerio de Desarrollo Social y Familia

Corresponde al conjunto acumulado de registros de RUN que están a disposición del Ministerio de Desarrollo Social y Familia como resultado de un proceso acumulativo de consultas realizadas al Registro Civil. Para este propósito, la División de Información Social (DIS) de la Subsecretaría de Evaluación Social, confecciona una base de datos que almacena el listado de todos los RUN recolectados a partir del análisis de diferentes bases de datos administrativas que el Ministerio de Desarrollo Social y Familia posee.

Dicho listado es almacenado en una base denominada **“Base Universo MDSF”**, que es luego consultado a Registro Civil y que tiene como resultado la identificación de un universo de RUN válidos.

ii. Universo de RUN válidos emitidos por el Registro Civil.

Corresponde al conjunto de RUN válidos entregados por el Registro Civil como respuesta a la consulta realizada por el Ministerio. Estos RUN además contienen la siguiente información:

- Fecha de nacimiento
- Fecha de defunción
- Profesión
- Comuna
- Estado Civil

Este conjunto resultante es almacenado en una base denominada **“Base Registro Civil”** que representa el universo de RUN entregados por Registro Civil y es la base inicial con la cual se empieza a trabajar para elaborar la base Residente

El Ministerio de Desarrollo Social y Familia no tiene acceso a una base de datos completa con todos los RUN que administra el Registro Civil, sino que, a partir de la información disponible a través de

diferentes bases administrativas de las que dispone, realiza consultas con este servicio, que permiten producir un listado acumulado de RUN válidos a una determinada fecha, tal como se detalla a continuación.

Como se puede apreciar en la Tabla N°1 la cantidad de runes contenidos en las bases consolidadas del MDSF y chequeados con los datos del Registro Civil corresponden a **25.209.160** personas para el periodo 2018-2019 y 25.323.460 para el periodo 2020.

Tabla N°1: Universo validado por Registro Civil para cada año

Base de datos	2018	2019	2020
Universo validado por Registro Civil	25.209.160	25.209.160	25.323.460

Fuente: Elaboración propia usando la Base de Registro Civil.

iii. **Universo de RUN válidos para cada período (año o semestre), descartando personas fallecidas, con edad superior a 110 años y sin información en la variable sexo**

Corresponde a un subconjunto de los RUN a los que tiene acceso el MDSF, el que es segmentado por períodos, años o semestres según corresponda, a fin de descartar: (a) personas fallecidas durante el período en cuestión o cuya edad supera los 110 años¹; (b) personas sin información válida para la variable sexo²; (c) Personas con fecha de nacimiento válida para el periodo (o anterior al 31 de diciembre del periodo t). Este proceso tiene por objeto asegurar que la información usada para construir el universo de población sea confiable.

Este subconjunto se almacena en una base llamada **“Base Universo de RUN válidos”** que representa el primer paso para elaborar la Base de Residentes, y sobre la cual se van a generar indicadores con registros administrativos. El total de personas que se alcanza como resultado de realizar estos descartes corresponde a **20.788.603 runes para el 2018, 20.889.389 para el 2019 y 20.976.572 para el 2020** (ver Tabla N°2).

Tabla N°2: Desglose de RUNs no considerados para la Base Universo de RUN válidos

Base de datos	2018	2019	2020
Universo validado por Registro Civil	25.209.160	25.209.160	25.323.460
Fallecidos hasta el 31 de diciembre t-1	3.806.672	3.914.279	4.025.726
Individuos sin información válida en variable sexo	353	353	355

¹ Existen RUN activos del Registro Civil con edades sobre los 110 años entre los cuales es probable identificar a personas cuyo fallecimiento no ha sido registrado por distintas razones (por ejemplo, en el caso que la persona falleciera en el extranjero y sus descendientes no hubieran realizado trámites para registrar esta defunción el país).

² Se considera a personas que no tienen un sexo registral reportado (sea éste mujer, hombre o indeterminado). Se estima que estos casos no cuentan con información suficiente para ser clasificados al momento de construir universos específicos de población. Para el periodo 2018 y 2019, el porcentaje de individuos sin información válida en la variable de sexo es de 0,001%.

Personas con fecha de nacimiento inválida (o fuera de rango al 31 de junio del periodo t)	362.672	199.161	842.882
Universo de RUNs válidos	20.788.603	20.889.389	20.976.572

Fuente: Elaboración propia usando la Base de Registro Civil.

En principio, la información provista en la Base Universo de RUN válidos permite identificar a todas las personas registradas en Chile (chilenos o extranjeros) y la comuna de residencia declarada ante este servicio. Esta información es fundamental para la definición del universo, pero presenta falencias dado que no permite determinar con certeza:

- i. Si la persona, en el período de análisis, se encuentra residiendo o no en el país.
- ii. Si, entre las personas que residen en Chile, la comuna declarada coincide o no con la comuna de residencia habitual el período de análisis.

Por lo anterior, se deben definir dos procesos:

- i. Definición de la situación de residencia en Chile en el año, en el primer semestre del año y en el segundo semestre del año.
- ii. Definición de comuna de residencia en Chile en el año.

iv. Base final: Base Residentes

Para elaborar esta base, primero se define el supuesto por el que se considerará que una persona reside. Este supuesto se basará en la observación de **movimientos en los registros administrativos**.

Estos movimientos, son los cambios o actualizaciones en los registros administrativos de una persona que dan cuenta de su actividad en el país durante un período de tiempo determinado. Los movimientos en bases de registros administrativos serán utilizados como un proxy para determinar si la persona se encontraba o no residiendo en el país en un año o semestre específico. Los registros administrativos utilizados son llamados **“Bases Adicionales”** que se utilizan para poder determinar la residencia en el país y/o la comuna de residencia. En el anexo N°1 se encuentra el listado de las bases de datos que comprenden este grupo.

La información de los ingresos y salidas al extranjero es registrada sistemáticamente por la Policía de Investigaciones de Chile (PDI) como resultado de su función en el control de las fronteras, estos datos ayudarían a realizar una estimación más precisa de la cantidad de personas residentes en un período determinado. Se está trabajando en conjunto con la PDI para incorporar estos datos administrativos en el futuro y así poder realizar una estimación más precisa de la población residente.

Por lo tanto, se define como un proxy de haber estado residiendo en el país en un año o semestre si durante ese período se observa un movimiento en los registros administrativos del individuo. Este proxy puede derivarse de la existencia de movimientos recogidos en diferentes bases de datos para uno o varios meses, por ejemplo, haber tenido actividad al menos en un mes o día del período. Así, se considera residente durante un periodo de referencia (año o semestre) a una persona que

muestra cambios en las bases de registros administrativos consultadas en al menos un día durante este mismo periodo³

Para el caso de personas menores de 10 años o para sus padres, que no registran movimientos en el período de análisis, se considera un proceso de imputación que se explica posteriormente. Para realizar este proceso se cuenta con una base llamada **“Base Parentesco”**. Esta base contiene dos columnas de run y un campo “tipo” en el cual se indica que el primer run corresponde al padre o la madre del segundo run, además se agrega la edad, sexo y fecha de fallecimiento del run del hijo/a. Esta información es utilizada para imputar, si es necesario, la situación de residencia de la madre/padre o el hijo según corresponda a partir de la situación de residencia de su hijo o madre/padre para el caso de hijos menores de 10 años. Este ejercicio supone que hijos y madres/padres viven juntos en el país o fuera de él, debido a que los niños menores dependen fuertemente de sus padres en todo orden de cuidados.

La medición de todos los indicadores de caracterización socioeconómica debe considerar a los individuos que se encuentran en el país en el año o semestre en cuestión, con el fin de evitar sesgos derivados principalmente de la inclusión de personas que en el año o semestre de análisis fijaron residencia en el extranjero o que aún no ingresaban al país durante dicho período. Para lo anterior se establecen umbrales o condiciones particulares para cada base de datos utilizada (Ver Anexo 1) para determinar si un individuo mostró o no movimiento y por lo tanto si el individuo reside o no en el país durante ese año o semestre.

Realizando estos procedimientos se obtiene una base llamada **“Base Residentes”**. La cantidad de runes en el año 2019, por ejemplo, se reduce a **18.591.661 personas**, cifra que varía respecto de 2018 y respecto de los semestres respectivos.

v. Base final: Base Residencia Comunal

Corresponde a la base que contiene la información de la comuna de residencia más probable para personas que se encuentran residiendo en el país en el año, primer semestre del año o según semestre. Si bien se cuenta con múltiples bases de datos que entregan información sobre comuna de residencia, éstas varían en términos de completitud y actualización. Para determinar qué bases de datos y qué registros dentro de ellas aportan la información más adecuada para asignar cada RUN a su comuna de residencia probable, se evalúan diferentes factores tales como períodos de actualización de la información, análisis de coincidencia entre las distintas fuentes o modas, además del contraste con la información de la comuna de residencia que proviene del Registro Civil.

Para ellos se usan dos bases: la **“Base Comuna”** y la **“Base Moda”**. La Base Comuna contiene información comunal por fuentes de datos y que se cataloga por 3, 4 o 5 grupos ⁴(dependiendo si

³ Se considera residente durante un periodo de referencia (año o semestre) a una persona que muestra cambios en las bases de registros administrativos consultadas en al menos un día durante este mismo periodo. El umbral de un día es el más elemental a suponer, y no excluye que puedan usarse umbrales más exigentes en el futuro. Umbrales más exigentes permitirían aproximar mejor el concepto de residencia habitual utilizado en los Censos, el cual corresponde al lugar donde la persona ha residido o tiene intenciones de residir a lo menos seis meses [<http://www.censo2017.cl/descargas/inmigracion/181123-glosario.pdf>].

⁴ Para ver la clasificación de cada tipo de base ir a la sección 5.c. del documento presente.

es la base anual o semestral) a partir de criterios de periodicidad, tipo de declaración de comuna y temporalidad. La Base Moda contiene la información de los individuos con las respectivas comunas que mostraron más coincidencia a través de las distintas fuentes de datos.

Finalmente, la aplicación de este paso para la identificación de la comuna de residencia reduce nuevamente el universo, por ejemplo, para el año 2019 a un total de **18.262.682 personas o runes** (ver Tabla N°1). El procedimiento genera una brecha de 500 mil runes entre los residentes nacionales y aquellos a los que se pudo establecer su comuna de residencia.

4. PROCESO #1: DETERMINACIÓN DE RESIDENCIA EN EL PAÍS

El proceso de determinación de residencia en el país tiene los siguientes pasos:

a. Base Registro Civil: generación y limpieza de variables intermedias.

1. Una vez que la base se encuentra validada en términos de estructura, se proceden a generar:
 - i. *Variable numérica de sexo*: La variable viene en formato de texto y se traspasa a formato numérico. 1: Hombre, 2: Mujer, 3: Indeterminado.
 - ii. *Variable de mes y año de nacimiento*: Las fechas de nacimiento vienen con seis números juntos que se deben separar en dos: los dos primeros corresponden al mes y los cuatro siguientes al año.
 - iii. *Variable de mes y año de defunción*: Las fechas de defunción vienen con seis números juntos que se deben separar en dos: los dos primeros corresponden al mes y los cuatro siguientes al año. Para el caso de un individuo que no ha fallecido dicho campo se encuentra con valor vacío.
 - iv. *Variable de edad a diciembre del año t-1*: A partir de las variables de año y mes es posible calcular la edad del individuo.
 - v. *Variable de edad a junio del año t*: A partir de las variables de año y mes es posible calcular la edad del individuo.
 - vi. *Fallecidos al 31 de diciembre del año t-1*: A partir de la variable de fecha y mes de defunción se genera una variable dummy que toma valor 1 si la persona al 31 de diciembre del año anterior falleció, y 0 en el caso contrario.
 - vii. *Fallecidos al 30 de junio de del año t*: A partir de la variable de fecha y mes de defunción se genera una variable dummy que toma valor 1 si la persona al 30 de junio del año en curso falleció, y 0 en el caso contrario.
 - viii. *Proxy de inmigrante del año t*: Variable generada con la información entregada en Registro Civil sobre nacionalidad e información a partir de tramos del RUN de los individuos.
 - ix. *Marca universo válido para el año o semestre t*: Variables dummies que toma valor 1 para el grupo de individuos válidos para el año t. Estas variables se crean desde el año 2018 en adelante tanto para la serie anual como semestral. Este grupo se conforma bajo las siguientes condiciones:
 - Edad entre 0 y 110 al 30 de junio de t.

- No haber fallecido al 31 de diciembre de t-1.
 - Tener marca de sexo valida. Es decir, tener un campo valido ya sea femenino, masculino o indeterminado.
2. Una vez generadas las variables necesarias se analizan estas por cada año o semestre t, generando tabulados de sexo y de edad para asegurar que no se detecten saltos atípicos entre períodos. En caso de ocurrir algún salto atípico se revisan las variables creadas que determinan el umbral válido del universo o bien se efectúa consulta a la División de Información Social (DIS) para evaluar consistencia.
3. Ya con la revisión chequeada para todos los períodos se procede a guardar la base por año o semestre con sus variables pertinentes. Estas son:
- i. run_falso⁵.
 - ii. edad al 31 de diciembre de t-1.
 - iii. edad al 30 de junio de t.
 - iv. proxy de inmigrante en t.
 - v. fallecido al 31 de diciembre de t-1.
 - vi. fallecido al 30 de junio de t.
 - vii. marca de universo válido para año o semestre t.

b. Bases Adicionales:

i. Generación y limpieza de variables intermedias.

Tal como se mencionó anteriormente, se requiere de bases de datos adicionales a la información del Registro Civil para la determinación de residencia.

1. Codificación de variables, limpieza de la base de datos. Dado que cada base tiene sus propias características, se generan variables intermedias que prestan utilidad para la construcción del indicador que permite determinar residencia en el país.

En términos generales, las variables intermedias que se crean van ligadas a:

- Transformaciones de variables que están en un formato erróneo (por ejemplo, números en formato string).
- Transposición de la base de datos cuando esta viene acumulada verticalmente (es decir, el individuo se repite en la base de datos) de tal manera que un identificador del individuo muestre toda su información.
- Limpieza de variables que contengan fechas.

⁵ La denominación de “run_falso” corresponde al número run una vez que se ha enmascarado o ha pasado por un algoritmo de anonimización que transforma el valor original o real en uno distinto, pero igualmente único, que permite el trabajo con registros administrativos, pero cautelando la privacidad de los datos personales. Este “run_falso” también puede leerse como “run_enmascarado” o “run_anonimizado”.

Después de haber realizado la limpieza de la base de datos se procede a la generación de una variable dicotómica que indicará residencia del individuo en el país para el año o semestre t . Para la determinación de residencia se trabajará con la generación de umbrales de residencia que se explicarán a continuación.

ii. Generación de umbrales de residencia

El objetivo de definir umbrales es usar el más apropiado según las características de las bases de datos, así como reducir los posibles sesgos por incorporación errónea.

- i. Umbral de residencia de al menos un día: Un individuo es considerado residente si aparece en alguna base de datos en cualquier momento y durante cualquier periodo de tiempo durante ese año o semestre. Este umbral es equivalente a considerar que por lo menos se encuentra 1 día dentro del periodo referencia en el país, independientemente del momento dentro del período en cuestión. Es decir, que muestre actividad⁶, en alguna de las bases de datos consideradas al menos una vez dentro de ese año o semestre.

Cabe notar que no todas las bases permiten realizar la identificación de residencia por semestres, solo aquellas que contienen información mensual o fechas de actualización de la información relevante. Dada esta limitación, se asume que un individuo que mostró movimiento durante el año, vale decir, en una base que no permite la identificación de residencia semestral, se considerará residente, para esta base en particular, tanto en el primer como en el segundo semestre.

- ii. Una vez que se genera dicha variable, se procede a eliminar las variables intermedias generadas para poder guardar la base y dejar únicamente en las bases de datos los siguientes campos:
 - Run falso
 - Indicador de residencia en el año o semestre⁷.
 - Comuna (en caso de venir en la base de datos)
 - Fecha de declaración de comuna (en caso de existir)
 - Fecha de último movimiento (en caso de existir).

c. Elaboración de Base Residente: Ratificación de personas que residen en el país

Ya con las bases revisadas se procede a realizar el proceso de determinación de residencia del individuo en el país.

⁶ Con actividad se hace referencia a que el individuo muestre algún indicio que implique que la misma se encontraba en el país en un determinado momento del año, o bien en un rango de tiempo durante el año. Lo anterior se puede ver a través de fechas de retiros de beneficios, montos, fechas de aplicación de encuesta, entre otros.

⁷ En una primera aproximación se usa el umbral de al menos 1 día del año o del semestre para identificar la residencia. Sin embargo, posteriormente se realizan pruebas con diferentes umbrales para analizar la sensibilidad del universo identificado dependiendo del umbral escogido.

- **Aplicación Umbral de residencia de al menos un día:** Este paso parte con la identificación de los RUNs que muestran movimiento en sus registros período por período, en cualquier momento del período considerado. En el Anexo N°1 se especifica, para cada base de datos a utilizar, el umbral de residencia para el caso de 1 día en el año o semestre.
 - i. Unión de los campos finales de residencia con la base del registro civil. En esta unión se agregan las variables que indican la residencia del individuo y se crea una variable que alude a la base de datos de origen. Por ejemplo, el indicador de residencia de la base RSH se llama “residente_rsh” y así sucesivamente con las demás bases de datos.
 - ii. Todos los individuos identificados como residentes a partir de cada una de las bases de datos administrativas se van guardando en una base de datos denominada “Base Residente”.
 - iii. Se generan 2 insumos: 1) Archivo .dta que guarda a los residentes sin los flag de residencia, es decir, sólo con las variables iniciales del registro civil. Esta sería la BBDD que se utilizará para los siguientes procesos. 2) BBDD con todos los individuos de registro civil no identificados como residentes para tener el respaldo de quienes no fueron identificados como residentes.
 - iv. Una vez construida la base residente se aplica la marca de universo válido que filtra según los criterios de edad y fallecimiento para el año o semestre previamente indicados.

d. Elaboración de Base Residente: Imputación de residencia según criterio de parentesco

A continuación, se hace uso de la base de parentesco elaborada por la DIS, donde se consigna las relaciones de maternidad y paternidad para cada run existente. Las imputaciones consideradas en este paso son de tres tipos:

- i. Imputación de la situación de residencia de la madre a hijos menores de 10 años para los casos que no se posea información de residencia.
- ii. Imputación de la situación de residencia del hijo menor de 10 años a los padres para los casos que no se posea información de residencia: la imputación de la situación de residencia del hijo considera algunas restricciones adicionales: los hijos sin madres no traspasan su residencia al padre, para todas las madres sin pareja se considera la residencia del hijo y solo a los padres de más de 24 años con pareja única que tuvieron su hijo a partir de los 24 años se les imputa la situación residencia del hijo.
- iii. Imputa residencia a todos los niños entre 0 y 1 años, que presentan una comuna de residencia en las bases del Registro Civil.

e. Validación de residentes

Ya con la base ajustada se realiza un proceso de validación con una fuente externa donde se revisa el total de personas, la distribución por sexo y tramos de edad, comparando estas distribuciones con las provenientes de las proyecciones de población del INE para el año de referencia.

A continuación, se muestran las distribuciones totales y por sexo-edad de las bases anuales y semestrales 2018, 2019 y 2020⁸ en comparación con las mismas distribuciones generadas a partir de las proyecciones de población de INE.

En la Tabla N°3 se aprecia que, en el año 2018, la cantidad total de residentes alcanza 18.446.991 versus 18.751.405 de las proyecciones (98,4% de cobertura). Mientras en el año 2019 se identifican 18.591.661 residentes frente a una proyección de 19.107.216 habitantes (97,3% de cobertura). Como se observa en las cifras anteriores, la cantidad total de residentes para ambos años se ubica por debajo de las proyecciones de población.

Esta situación se repite si se comparan las cifras totales desagregadas por sexo. Sin embargo, al desagregar por sexo-edad algunos tramos muestran totales más altos que las proyecciones censales. Es el caso de los tramos de edad [50;59], [70-79] y [80; +] en el año 2018, y en el 2019 ocurre algo similar para los mismos tramos salvo [50;29].

Tabla N°3: Comparación entre residentes y proyecciones de población 2018-2019

2018	Total				Hombre				Mujer			
	N		%		N		%		N		%	
T. Edad	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.
0 a 9 años	2.441.924	2.483.266	13,2	13,2	1.244.233	1.264.644	13,7	13,7	1.197.691	1.218.622	12,8	12,8
10 a 19 años	2.453.902	2.499.286	13,3	13,3	1.252.551	1.271.318	13,8	13,8	1.201.351	1.227.968	12,8	12,9
20 a 29 años	3.024.803	3.038.361	16,4	16,2	1.530.980	1.544.381	16,9	16,7	1.493.823	1.493.980	15,9	15,7
30 a 39 años	2.733.555	2.797.671	14,8	14,9	1.364.717	1.415.645	15,1	15,3	1.368.838	1.382.026	14,6	14,5
40 a 49 años	2.418.591	2.532.795	13,1	13,5	1.183.546	1.259.912	13,1	13,6	1.235.045	1.272.883	13,2	13,4
50 a 59 años	2.309.401	2.296.313	12,5	12,2	1.113.784	1.116.792	12,3	12,1	1.195.617	1.179.521	12,7	12,4
60 a 69 años	1.597.398	1.676.909	8,7	8,9	750.120	788.746	8,3	8,5	847.278	888.163	9,0	9,3
70 a 79 años	954.363	919.501	5,2	4,9	425.673	402.287	4,7	4,4	528.690	517.214	5,6	5,4
80 o más	513.054	507.303	2,8	2,7	189.568	180.759	2,1	2,0	323.486	326.544	3,4	3,4
Total	18.446.991	18.751.405	100	100	9.055.172	9.244.484	100	100	9.391.819	9.506.921	100	100

2019	Total				Hombre				Mujer			
	N		%		N		%		N		%	
T. Edad	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.
0 a 9 años	2.400.363	2.477.648	12,9	13,0	1.223.178	1.261.935	13,5	13,4	1.177.185	1.215.713	12,4	12,6
10 a 19 años	2.457.111	2.509.491	13,2	13,1	1.253.575	1.276.565	13,8	13,5	1.203.536	1.232.926	12,7	12,7
20 a 29 años	3.010.383	3.074.768	16,2	16,1	1.515.663	1.562.572	16,7	16,6	1.494.720	1.512.196	15,7	15,6
30 a 39 años	2.828.983	2.912.114	15,2	15,2	1.405.796	1.474.777	15,5	15,6	1.423.187	1.437.337	15,0	14,8

⁸ Hasta el momento solo se cuenta con información hasta el primer semestre del 2020.

40 a 49 años	2.409.368	2.572.488	13,0	13,5	1.171.819	1.281.692	12,9	13,6	1.237.549	1.290.796	13,0	13,3
50 a 59 años	2.304.444	2.336.167	12,4	12,2	1.105.606	1.137.863	12,2	12,1	1.198.838	1.198.304	12,6	12,4
60 a 69 años	1.658.393	1.742.032	8,9	9,1	775.642	820.100	8,5	8,7	882.751	921.932	9,3	9,5
70 a 79 años	993.075	956.000	5,3	5,0	443.729	420.294	4,9	4,5	549.346	535.706	5,8	5,5
80 o más	529.541	526.508	2,8	2,8	195.247	188.341	2,1	2,0	334.294	338.167	3,5	3,5
Total	18.591.661	19.107.216	100	100	9.090.255	9.424.139	100	100	9.501.406	9.683.077	100	100

Fuente: Elaboración propia usando las Base Residente.

Respecto a la distribución de la población por tramos etarios, la base de residentes conserva una estructura similar a la de las proyecciones de población. En el año 2018, la mayor diferencia en puntos porcentuales se observa en los hombres entre 40 y 49 años (0,6 puntos porcentuales), siendo mayor en las proyecciones de población. Situación similar ocurre en 2019, pero con una diferencia de 0,7 puntos porcentuales.

El comportamiento semestral es análogo al comportamiento de las bases anuales (ver Tabla N°4). La cantidad total de residentes para el primer semestre alcanza 18.311.319 versus 18.751.405 de las proyecciones teniendo así una cobertura de 97,7%. Por otro lado, el segundo semestre tiene 18.119.227 versus 18.751.405 de las proyecciones teniendo una cobertura de 96,6%. Se tiene que ambos semestres presentan coberturas debajo a las proyecciones de población, teniendo que en el segundo semestre es menor la cobertura comparada al primer semestre.

Mirando las cantidades de residentes por sexo la tendencia se repite. Sin embargo, los tramos de edad con una cantidad de residentes mayor a la proyectada corresponden nuevamente a [70-79] y [80; +], durante el primer semestre, pero en el segundo semestre solo se observa una mayor cantidad de residentes en el tramo [70;79]. La máxima diferencia se observa en el grupo de hombres de [40;49] con 0,6 pp. durante el primer semestre y para el mismo grupo con 0,7 pp. en el segundo semestre del 2018.

Tabla N°4: Comparación entre residentes semestrales y proyecciones de población 2018

2018-01	Total				Hombre				Mujer			
	N		%		N		%		N		%	
T. Edad	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.
0 a 9 años	2.429.321	2.483.266	13,3	13,2	1.237.903	1.264.644	13,8	13,7	1.191.418	1.218.622	12,8	12,8
10 a 19 años	2.447.138	2.499.286	13,4	13,3	1.249.295	1.271.318	13,9	13,8	1.197.843	1.227.968	12,8	12,9
20 a 29 años	2.994.404	3.038.361	16,4	16,2	1.514.904	1.544.381	16,9	16,7	1.479.500	1.493.980	15,9	15,7
30 a 39 años	2.705.777	2.797.671	14,8	14,9	1.348.422	1.415.645	15,0	15,3	1.357.355	1.382.026	14,6	14,5
40 a 49 años	2.395.739	2.532.795	13,1	13,5	1.170.742	1.259.912	13,0	13,6	1.224.997	1.272.883	13,1	13,4
50 a 59 años	2.286.690	2.296.313	12,5	12,2	1.103.170	1.116.792	12,3	12,1	1.183.520	1.179.521	12,7	12,4
60 a 69 años	1.587.851	1.676.909	8,7	8,9	745.692	788.746	8,3	8,5	842.159	888.163	9,0	9,3
70 a 79 años	952.348	919.501	5,2	4,9	424.921	402.287	4,7	4,4	527.427	517.214	5,7	5,4
80 o más	512.051	507.303	2,8	2,7	189.170	180.759	2,1	2,0	322.881	326.544	3,5	3,4
Total	18.311.319	18.751.405	100	100	8.984.219	9.244.484	100	100	9.327.100	9.506.921	100	100

2018-02	Total				Hombre				Mujer			
	N		%		N		%		N		%	
T. Edad	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.
0 a 9 años	2.437.094	2.483.266	13,5	13,2	1.241.817	1.264.644	14,1	13,7	1.195.277	1.218.622	12,9	12,8
10 a 19 años	2.445.248	2.499.286	13,5	13,3	1.247.804	1.271.318	14,1	13,8	1.197.444	1.227.968	12,9	12,9
20 a 29 años	2.968.667	3.038.361	16,4	16,2	1.490.264	1.544.381	16,9	16,7	1.478.403	1.493.980	15,9	15,7
30 a 39 años	2.657.360	2.797.671	14,7	14,9	1.311.841	1.415.645	14,8	15,3	1.345.519	1.382.026	14,5	14,5
40 a 49 años	2.343.697	2.532.795	12,9	13,5	1.134.139	1.259.912	12,8	13,6	1.209.558	1.272.883	13,0	13,4
50 a 59 años	2.238.293	2.296.313	12,4	12,2	1.070.598	1.116.792	12,1	12,1	1.167.695	1.179.521	12,6	12,4
60 a 69 años	1.574.984	1.676.909	8,7	8,9	733.667	788.746	8,3	8,5	841.317	888.163	9,1	9,3
70 a 79 años	948.421	919.501	5,2	4,9	421.961	402.287	4,8	4,4	526.460	517.214	5,7	5,4
80 o más	505.463	507.303	2,8	2,7	185.444	180.759	2,1	2,0	320.019	326.544	3,4	3,4
Total	18.119.227	18.751.405	100	100	8.837.535	9.244.484	100	100	9.281.692	9.506.921	100	100

Fuente: Elaboración propia usando la Base Residente.

Por su parte, en el primer semestre del 2019 (ver Tabla N°5) se tiene una cantidad total de residentes de 18.270.574 versus 19.107.216 de las proyecciones, teniendo así una cobertura de 96,1%. Por otro lado, el segundo semestre tiene una cantidad total de residentes de 18.482.499 versus 19.107.216 de las proyecciones, teniendo así una cobertura de 96,7%. Se tiene que ambos semestres presentan coberturas debajo a las proyecciones de población, teniendo que en el segundo semestre aumenta levemente el porcentaje de cobertura.

Mirando las cifras por sexo la tendencia se repite la misma tendencia. Sin embargo, los tramos de edad con una cantidad de residentes mayor a la proyectada corresponden, durante el primer semestre, a [70-79] y [80; +], y durante el segundo semestre solo en el tramo [70;79] la cantidad de residentes es mayor a lo proyectado. Las máximas diferencias en términos porcentuales no superan los 0,8 pp. que se alcanza en el grupo de hombres de [40;49] años.

Tabla N°5: Comparación entre residentes semestrales y proyecciones de población 2019

2019-01	Total				Hombre				Mujer			
	N		%		N		%		N		%	
T. Edad	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.
0 a 9 años	2.370.116	2.477.648	12,9	13,0	1.207.820	1.261.935	13,5	13,4	1.162.296	1.215.713	12,4	12,6
10 a 19 años	2.448.178	2.509.491	13,3	13,1	1.249.080	1.276.565	13,9	13,5	1.199.098	1.232.926	12,7	12,7
20 a 29 años	2.961.234	3.074.768	16,1	16,1	1.487.023	1.562.572	16,6	16,6	1.474.211	1.512.196	15,7	15,6
30 a 39 años	2.779.903	2.912.114	15,1	15,2	1.374.517	1.474.777	15,3	15,6	1.405.386	1.437.337	14,9	14,8
40 a 49 años	2.372.988	2.572.488	12,9	13,5	1.148.550	1.281.692	12,8	13,6	1.224.438	1.290.796	13,0	13,3
50 a 59 años	2.272.836	2.336.167	12,4	12,2	1.087.628	1.137.863	12,1	12,1	1.185.208	1.198.304	12,6	12,4
60 a 69 años	1.645.468	1.742.032	9,0	9,1	767.824	820.100	8,6	8,7	877.644	921.932	9,3	9,5
70 a 79 años	991.192	956.000	5,4	5,0	442.865	420.294	4,9	4,5	548.327	535.706	5,8	5,5
80 o más	528.659	526.508	2,9	2,8	194.840	188.341	2,2	2,0	333.819	338.167	3,5	3,5
Total	18.370.574	19.107.216	100	100	8.960.147	9.424.139	100	100	9.410.427	9.683.077	100	100

2019-02	Total				Hombre				Mujer			
	N		%		N		%		N		%	
T. Edad	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.
0 a 9 años	2.396.271	2.477.648	13,0	13,0	1.221.062	1.261.935	13,5	13,4	1.175.209	1.215.713	12,4	12,6
10 a 19 años	2.454.284	2.509.491	13,3	13,1	1.251.770	1.276.565	13,9	13,5	1.202.514	1.232.926	12,7	12,7
20 a 29 años	2.993.003	3.074.768	16,2	16,1	1.503.160	1.562.572	16,7	16,6	1.489.843	1.512.196	15,7	15,6
30 a 39 años	2.811.494	2.912.114	15,2	15,2	1.392.986	1.474.777	15,4	15,6	1.418.508	1.437.337	15,0	14,8
40 a 49 años	2.394.244	2.572.488	13,0	13,5	1.161.464	1.281.692	12,9	13,6	1.232.780	1.290.796	13,0	13,3
50 a 59 años	2.289.443	2.336.167	12,4	12,2	1.096.295	1.137.863	12,2	12,1	1.193.148	1.198.304	12,6	12,4
60 a 69 años	1.647.409	1.742.032	8,9	9,1	768.589	820.100	8,5	8,7	878.820	921.932	9,3	9,5
70 a 79 años	982.861	956.000	5,3	5,0	437.523	420.294	4,8	4,5	545.338	535.706	5,8	5,5
80 o más	513.490	526.508	2,8	2,8	188.288	188.341	2,1	2,0	325.202	338.167	3,4	3,5
Total	18.482.499	19.107.216	100	100	9.021.137	9.424.139	100	100	9.461.362	9.683.077	100	100

Fuente: Elaboración propia usando la Base Residente.

El último periodo corresponde al primer semestre del 2020 (ver Tabla N°6), en el que se identificaron 18.941.079 residentes, que en relación con las proyecciones de población representa un 97,3% de cobertura. En este semestre hay tres tramos de edad en los que se identifica una mayor cantidad de residentes que la población proyectada: [50;59], [70;79] y [80; +]. La mayor diferencia porcentual entre la distribución de residentes y población proyectada se encuentra en el grupo de hombres [70;79] y es de 0,5 pp.

Tabla N°6: Comparación entre residentes semestrales y proyecciones de población 2020

2020-01	Total				Hombre				Mujer			
	N		%		N		%		N		%	
T. Edad	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.	Residente	Proy.	Resi.	Proy.
0 a 9 años	2.322.366	2.472.769	12,3	12,7	1.183.440	1.259.392	12,7	13,1	1.138.926	1.213.377	11,8	12,3
10 a 19 años	2.463.152	2.525.331	13,0	13,0	1.256.041	1.284.478	13,5	13,4	1.207.111	1.240.853	12,5	12,6
20 a 29 años	2.953.317	3.092.328	15,6	15,9	1.489.144	1.570.365	16,0	16,4	1.464.173	1.521.963	15,2	15,4
30 a 39 años	2.948.644	3.031.451	15,6	15,6	1.477.146	1.535.583	15,9	16,0	1.471.498	1.495.868	15,3	15,2
40 a 49 años	2.485.886	2.618.520	13,1	13,5	1.222.223	1.306.441	13,1	13,6	1.263.663	1.312.079	13,1	13,3
50 a 59 años	2.378.199	2.369.901	12,6	12,2	1.155.210	1.155.899	12,4	12,0	1.222.989	1.214.002	12,7	12,3
60 a 69 años	1.764.719	1.804.002	9,3	9,3	837.493	849.953	9,0	8,9	927.226	954.049	9,6	9,7
70 a 79 años	1.050.718	997.174	5,5	5,1	473.630	440.471	5,1	4,6	577.088	556.703	6,0	5,6
80 o más	574.078	546.834	3,0	2,8	214.188	196.519	2,3	2,0	359.890	350.315	3,7	3,6
Total	18.941.079	19.458.310	100	100	9.308.515	9.599.101	100	100	9.632.564	9.859.209	100	100

Fuente: Elaboración propia usando la Base Residente.

Cabe señalar que las diferencias entre la información que entrega la Base Residentes y las proyecciones de población del INE pueden tener diferentes causas. Entre ellas, se debe considerar el aumento en la migración que ha afectado al país, así como la falta de información que permita identificar a quienes han salido del país en la Base Residente. Con todo, y tal como se demostró anteriormente, las diferencias son pequeñas y permiten tener confianza en que no se producen sesgos importantes en la construcción de las bases residentes.

5. PROCESO #2: DETERMINACIÓN DE COMUNA PARA BASE RESIDENTE

En este apartado se explica el proceso para elaborar el campo de información comunal para las bases residentes.

Resulta relevante notar que, a diferencia de la base anual, en las bases semestrales solamente se trabaja con un campo que indique la región de residencia que se construirá a partir de la variable comunal elaborada. Lo anterior se determina para evitar posibles inconsistencias que se puedan dar con la comuna determinada para la base anual. En específico, al trabajar con información parcial (semestres) respecto a la información comunal que se tiene disponible existe una probabilidad de que se generen variaciones en la comuna declarada de residencia y esta puede ser menos precisa

para los semestres al disponer de menos información. Lo anterior afecta en mayor medida a las comunas de menor población donde puede variar notoriamente en los niveles de cobertura que tenga, lo que finalmente puede entregar una inconsistencia a lo largo de las distintas bases de datos generadas.

Dado que inicialmente no se cuenta con la comuna de residencia del individuo para el año, se realiza un ejercicio de determinación de comuna trabajando con las bases de datos disponibles que tienen el campo de comuna declarado. Dicho criterio de selección de comuna unirá la variable de comuna de cada base de datos de manera agrupada a partir de un conjunto de criterios que clasificará a las bases de datos por nivel de calidad de la declaración de comuna como campo a considerar como residencia.

Antes de determinar las comunas de los individuos es necesario realizar un proceso de homologación de los códigos comunales y una selección de comuna para cada base de datos, particularmente en los casos de bases de datos que presentan múltiples registros de un mismo individuo, y que presentan más de una comuna declarada.

a. Homologación de códigos según división política administrativa

Se revisan las bases de datos que contienen campos de comuna para realizar un análisis de calidad y consistencia, y así evitar un problema a partir de codificación de las comunas antiguas al periodo ya sea debido a que las comunas tienen una codificación antigua o bien el código que se tiene no es válido.

Para lo anterior, se debe identificar si la comuna tiene un código válido, pero que representa una codificación antigua de la comuna. Para este caso se realiza el ajuste del código respectivo, considerando la división político-administrativa vigente al año a la fecha del 30 de junio⁹.

También puede ocurrir que el código no sea válido. En esta última situación, como no se puede determinar cuál era efectivamente la comuna, se opta por dejar como valor missing.

b. Identificación de comuna para cada base de datos

Se trabaja con el grupo de Bases de Datos Adicionales que presentan campo de comuna realizado 2 tratamientos dependiendo de la información que contenga la base de datos:

- Para las bases de datos que contienen solo una observación por individuo, se mantiene esa información comunal.
- Para las bases de datos que presentan múltiples registros para un mismo individuo, como por ejemplo las bases que tienen información mensual, se selecciona la comuna declarada en el período más reciente.

⁹ Esto implica, por ejemplo, que, a raíz de la creación de la Región de Ñuble en el mes de septiembre de 2017, los códigos asociados a cada una de sus comunas deben ser actualizados y homogeneizados al Código Único Territorial vigente a contar del año siguiente.

c. Clasificación de bases de datos según calidad de información comunal

Dentro de las distintas bases de datos recibidas, el tratamiento que se debe dar a la declaración de comuna¹⁰ no es equivalente, principalmente por la actualización de la información de comuna de residencia que no es homogénea entre individuos ni bases.

En base a lo anterior, resulta necesario agrupar la información de comuna de residencia según criterios que entreguen un orden de prioridad al momento de determinar la comuna del individuo. De esta manera, se puede priorizar la información más reciente de cada individuo y no estar usando información antigua respecto a comuna de residencia.

Dicha clasificación varía si se trabaja con la base anual, la base del primer semestre o la base del segundo semestre¹¹. A continuación, se presenta la clasificación para cada caso:

i. Clasificación de Base Comuna: Anual

- **Grupo 1:** comuna de residencia declarada en el año t.
- **Grupo 2:** comuna de residencia declarada en el año t-1.
- **Grupo 3:** comuna de residencia declarada entre el año t-2 y t-5.

Los grupos 1, 2 y 3 son excluyentes unos de otros. Es decir, si un individuo cuenta con información de comuna de residencia procedente de una fuente clasificada en el Grupo 1, no formará parte de los grupos siguientes (aun cuando la persona figure en alguna de las bases de datos pertenecientes a los grupos posteriores), lo mismo ocurre para todos los otros grupos.

ii. Clasificación de Base Comuna: Primer Semestre

- **Grupo 1:** Para los primeros 6 meses del año t, el individuo declara comuna de residencia.
- **Grupo 2:** Para el segundo semestre del año t-1, el individuo declara comuna de residencia.
- **Grupo 3:** Para el primer semestre de t-1, el individuo declara comuna de residencia.
- **Grupo 4:** Para los periodos comprendidos entre el año t-2 y t-5 (anuales), el individuo declara comuna de residencia.

Al igual que en el caso anterior, los grupos son excluyentes unos de otros.

iii. Clasificación de Base Comuna: Segundo Semestre

- **Grupo 1:** Para los últimos 6 meses del año t, el individuo declara comuna de residencia.
- **Grupo 2:** Para los primeros 6 meses del año t, el individuo declara comuna de residencia.
- **Grupo 3:** Para el segundo semestre del año t-1, el individuo declara comuna de residencia.
- **Grupo 4:** Para el primer semestre del año t-1, el individuo declara comuna de residencia.

¹⁰ Es relevante aclarar que no se consideran en el análisis las bases de datos que tienen información comunal pero que no representan la comuna de residencia, sino la comuna en la que se realiza alguna gestión asociada al registro administrativo.

¹¹ El anexo N°2 presenta las bases de datos utilizados para la generación de información comunal para la base residente del año 2019 y 2018, con sus respectivos procesamientos para determinar la temporalidad del registro tanto para el criterio anual como para el criterio semestral.

- **Grupo 5:** Para los periodos comprendidos entre el año t-2 y t-5 (anuales), el individuo declara comuna de residencia.

Al igual que en los otros casos, los grupos son excluyentes unos de otros.

d. Elaboración de Base Comuna

Se elabora una base de datos para cada grupo identificado previamente. Dicha base tendrá los siguientes campos:

- RUN innominado.
- Código comuna.
- Identificador de grupo.

Se genera la base de datos para cada grupo y posteriormente se realiza la exclusión de los RUNs repetidos entre los distintos grupos según nivel de prioridad. Una vez realizado esto se procede a unir las distintas bases en una sola para poder realizar el ejercicio de determinación de comuna. Dicha base se denomina como “Base Comuna”

e. Elaboración de Base Moda

Usando la Base Comuna, se determina la comuna de residencia según la mayor coincidencia en cada grupo (modas). En caso de que un individuo presente 2 modas distintas entonces se optará por seleccionar de manera aleatoria la comuna de residencia. Si presenta 3 o más modas, se opta por no determinar una comuna entre las modas y dejar al individuo sin comuna identificada (valor vacío en la base de datos). Una vez realizado el proceso se guarda la base de datos solamente con los campos de RUN innominado y código de comuna y se le denominará “Base Moda”.

f. Unión a Base Residente y determinación de comuna de residencia para grupo no identificado mediante información de Registro Civil

La Base Moda se une a la base Registro Civil. La base resultante contiene 2 campos de comuna. Por un lado, se encontrará la comuna declarada en la información proveniente de Registro Civil¹², y, por otro, la información comunal resultante del ejercicio de modas. Respecto a estos dos campos se realiza lo siguiente¹³:

¹² La razón por la cual la información comunal de Registro Civil no es utilizada en la Base Comuna previamente se debe a la naturaleza que la variable tiene respecto a las actualizaciones que tiene. Para la variable no se sabe cuándo dicha comuna fue declarada en el tiempo, el único antecedente que se tiene es que es la comuna donde la persona realizó su último trámite del Registro Civil. Por lo tanto, dado que la temporalidad del registro no se puede comprobar se opta por dejarlo como una fase final la información comunal del Registro Civil para indicar la comuna de residencia, dando así mayor prioridad a la información donde sí se sabe la temporalidad de la data.

¹³ Adicionalmente, una vez determinadas las comunas utilizando los 3 grupos mencionados se analiza el posible efecto que tendría, sumar al grupo identificado de Registro Civil las comunas provenientes de las bases de datos para el periodo t-6 hacia atrás, sobre la distribución de la población comunal. Para los años trabajados no se encontraron diferencias importantes.

- i. Para los individuos con comuna identificada desde la Base Moda, se le considerará esa información como la comuna del individuo, independientemente de lo que se indique en la comuna proveniente de Registro Civil.
- ii. Para los individuos sin comuna identificada desde la Base Moda, se les considerará la comuna proveniente de Registro Civil, en caso de estar disponible.

g. Imputación de la comuna de la madre a los hijos/as residentes de 10 o menos años.

Finalmente, una vez elaborado el campo de la información comunal en la Base Residente, se realiza un último ejercicio de imputación a individuos sin comuna. Para este caso, los individuos residentes de 10 o menos años se le imputa la comuna de la madre, siempre y cuando tenga una comuna identificada.

h. Resultados comunales

Los resultados del proceso anterior explicado fueron los siguientes para las bases de la serie completa son los siguientes:

Tabla N°7: Resultados generales identificación campo comunal 2018-2019

Clasificación	2018-1		2018-2		2018		2019-1		2019-2		2019		2020-1	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
BBDD Adicionales	5.571.347	30,4	5.924.488	32,7	7.124.873	38,6	6.417.292	34,9	6.749.848	36,5	6.957.437	37,4	7.355.949	38,8
RC	12.173.387	66,5	11.658.125	64,3	10.862.869	58,9	11.377.835	61,9	11.162.328	60,4	11.063.931	59,5	10.937.467	57,7
Sin Comuna	434.468	2,4	404.411	2,2	324.711	1,8	423.924	2,3	414.807	2,2	407.168	2,2	415.999	2,2
Comuna Madre	132.117	0,7	132.203	0,7	134.538	0,7	151.523	0,8	155.516	0,8	163.125	0,9	231.664	1,2
Total	18.311.319	100	18.119.227	100	18.446.991	100	18.370.574	100	18.482.499	100	18.591.661	100	18.941.079	100

Fuente: Elaboración propia usando la Base Residente.

Sobre estos resultados es posible ver que a medida que avanzan los periodos se está rescatando cada vez más información a partir de las bases de datos adicionales, lo cual resulta ideal considerando que desde dicha fuente de información se tiene certeza respecto a la actualización con respecto a la comuna de residencia.

i. Validación resultados comunales

Se realiza un proceso de validación externa de cobertura comunal con respecto a proyecciones demográficas del INE del año respectivo.

Primero, se analiza la cobertura a nivel regional, tal como se presenta en la Tabla N°8. Respecto a las coberturas resalta que la región con menor nivel de cobertura es la Metropolitana en ambos periodos (94,6% el 2018 y 92,9% el 2019). En segundo lugar, para ambos años Atacama, Aysén y Ñuble presenten una cobertura mayor al 100%. Como último punto, es relevante notar que la cobertura del año 2018 resulta mayor en todas las regiones comparándolo con la cobertura del

2019, donde el caso más notorio es para la región del Ñuble con una variación de 6% entre ambos años.

Tabla N°8: Porcentaje de cobertura regional anual respecto a proyecciones demográficas del INE

Región	2018-anual		2019-anual	
	RES	COB	RES	COB
Arica y Parinacota	239.581	99,00%	241.070	97,60%
Tarapacá	346.319	97,60%	348.158	94,40%
Antofagasta	631.096	97,80%	632.954	94,70%
Atacama	312.296	101,40%	312.809	100,50%
Coquimbo	785.802	97,30%	786.177	95,70%
Valparaíso	1.846.461	96,70%	1.850.366	95,60%
Metropolitana	7.287.472	94,60%	7.351.527	92,90%
O'Higgins	953.750	98,70%	960.325	98,10%
Maule	1.082.259	97,90%	1.089.378	97,40%
Ñuble	520.822	103,30%	489.280	96,30%
Biobío	1.575.534	95,80%	1.580.366	95,50%
La Araucanía	996.078	99,50%	996.813	98,90%
Los Ríos	393.004	98,00%	391.693	97,10%
Los Lagos	872.676	99,50%	873.478	98,80%
Aysén	109.305	103,10%	108.730	101,90%
Magallanes	169.825	97,80%	171.369	97,40%
Total	18.122.280	96,60%	18.184.493	95,20%

Fuente: Elaboración propia usando la Base Residencia Comunal e información del INE.

Por parte de las bases comunales, los resultados de cobertura regional se aprecian en la Tabla N°9 en general se presenta una cobertura menor comparado para las bases anuales (para 2018 y 2019). Se tiene a nivel nacional una cobertura que varía entre 93,9% y 95,3% de individuos con comuna identificada respecto a las proyecciones del INE.

Tabla N°9: Porcentaje de cobertura regional anual respecto a proyecciones demográficas del INE

Región	2018-s1		2018-s2		2019-s1		2019-s2		2020-s1	
	RES	COB	RES	COB	RES	COB	RES	COB	RES	COB
Arica y Parinacota	237.705	98,3	234.100	96,8	237.771	96,2	239.293	96,9	245.947	97,6
Tarapacá	343.850	96,9	338.125	95,3	342.267	92,8	345.793	93,7	354.359	92,6
Antofagasta	625.864	97	615.308	95,4	623.859	93,3	628.127	94	640.465	92,6
Atacama	308.296	100,1	304.812	99	308.565	99,1	310.373	99,7	315.437	100,2
Coquimbo	776.312	96,2	765.695	94,9	775.688	94,4	781.115	95,1	797.825	95,4
Valparaíso	1.821.684	95,4	1.801.302	94,3	1.824.019	94,2	1.835.455	94,8	1.895.379	96,7
Metropolitana	7.194.417	93,4	7.112.608	92,3	7.229.562	91,3	7.285.814	92	7.494.114	92,2
O'Higgins	942.012	97,5	934.060	96,6	947.813	96,8	953.962	97,5	973.792	98,3
Maule	1.067.762	96,6	1.057.791	95,7	1.073.426	95,9	1.081.126	96,6	1.101.756	97,3
Ñuble	484.010	96	520.003	103,1	512.645	100,9	514.278	101,2	526.492	102,9
Biobío	1.560.773	94,9	1.543.441	93,8	1.558.232	94,2	1.565.914	94,6	1.595.837	95,9
La Araucanía	984.515	98,3	974.888	97,4	985.299	97,8	990.806	98,3	1.011.861	99,8

Los Ríos	388.815	97	384.547	95,9	387.212	96	389.399	96,5	396.625	97,7
Los Lagos	863.982	98,5	854.270	97,4	863.275	97,6	868.185	98,2	889.699	99,8
Aysén	107.844	101,7	107.021	100,9	107.413	100,7	107.930	101,2	110.395	102,9
Magallanes	169.010	97,4	166.845	96,1	169.604	96,4	170.122	96,7	175.097	98,2
Total	17.876.851	95,3	17.714.816	94,5	17.946.650	93,9	18.067.692	94,6	18.525.080	95,2

Fuente: Elaboración propia usando la Base Residencia Comunal e información del INE.

El último punto comentado respecto a la variación de coberturas los periodos lo podemos ver reflejado en la Tabla N°10. En dicha tabla se ve que de manera transversal la mayor concentración de comunas se da en el tramo de 90% a 100%, seguido por el tramo de 100% a 110%.

Tabla N°10: Distribución de porcentaje de cobertura respecto a proyecciones INE

Nivel de Cobertura	2018			2019			2020		
	S1	S2	A	S1	S2	A	S1	S2	A
[0%,10%]	0	0	0	0	0	0	0		
]10%,20%]	0	0	0	1	0	0	1		
]20% a 30%]	1	1	1	0	1	1	0		
]30% a 40%]	1	1	1	1	1	1	1		
]40% a 50%]	4	5	2	4	3	2	3		
]50% a 60%]	4	3	5	0	1	3	2		
]60% a 70%]	1	2	2	4	1	2	2		
]70% a 80%]	4	3	4	8	8	6	6		
]80% a 90%]	26	33	15	33	28	20	23		
]90% a 100%]	194	202	160	200	200	203	169		
]100% a 110%]	99	83	142	83	91	96	122		
Mayor a 110%	12	13	14	12	12	12	17		
Total	346	346	346	346	346	346	346		

Fuente: Elaboración propia usando la Base Residencia Comunal e información del INE.

Para evaluar el nivel de cobertura, se determina qué comunas se encuentran subrepresentadas (menos de 75% de cobertura respecto a las proyecciones) y cuáles comunas se encuentran sobrerrepresentadas (más de 110% de cobertura respecto a las proyecciones)¹⁴.

Lo anterior puede influir respecto a que comunas no se consideran para la presentación de resultados de un indicador, dada las diferencias que se tienen con las proyecciones del INE. Sin embargo, que comunas considerar o no dependerán de las coberturas respecto a la población específica de dicho indicador. Para cada año, estas son las comunas con cobertura menor a 75% a nivel nacional¹⁵:

¹⁴ Tanto el límite inferior como el superior son solo referenciales con el objetivo de poder analizar las coberturas comunales.

¹⁵ Dado que para el momento no se posee la base anual para el año 2020 se presenta la información del primer semestre (aun cuando no se llega a nivel comunal en las bases semestrales) a modo de comparación.

Tabla N°11: Comunas con coberturas menores a 75%

2018				2019				2020-1			
NOMBRE	RES	PROY	COB	NOMBRE	RES	PROY	COB	NOMBRE	RES	PROY	COB
Torres del Paine	254	972	26,1%	Torres del Paine	213	998	21,3%	Torres del Paine	188	1.021	18,4%
San Gregorio	228	692	32,9%	San Gregorio	271	686	39,5%	San Gregorio	213	681	31,3%
Timaukel	123	284	43,3%	Timaukel	121	283	42,8%	Timaukel	117	282	41,5%
Colchane	704	1.527	46,1%	Putre	1.238	2.493	49,7%	Putre	1.239	2.515	49,3%
General Lagos	403	799	50,4%	Colchane	781	1.556	50,2%	General Lagos	402	810	49,6%
Putre	1.269	2.467	51,4%	General Lagos	421	805	52,3%	Colchane	816	1.583	51,5%
Río Verde	122	213	57,3%	Camarones	736	1.228	59,9%	Camarones	732	1.233	59,4%
Laguna Blanca	159	270	58,9%	Laguna Blanca	178	266	66,9%	Laguna Blanca	182	264	68,9%
Camarones	723	1.223	59,1%	Río Verde	149	213	70,0%	Río Verde	146	211	69,2%
Ollagüe	190	282	67,4%	María Elena	4.747	6.749	70,3%	María Elena	4.826	6.814	70,8%
María Elena	4.638	6.685	69,4%	Ollagüe	205	284	72,2%	Ollagüe	211	287	73,5%
Sierra Gorda	1.191	1.633	72,9%					Maule	44.849	60.000	74,7%

Fuente: Elaboración propia usando la Base Residencia Comunal e información del INE.

Respecto a las comunas que presentan una cobertura menor al 75% se puede ver que, a excepción de Sierra Gorda el 2018 y Maule el 2019, el grupo de comunas de baja cobertura es transversal en los periodos compuesto por Torres del Paine, San Gregorio, Timaukel, Colchane, General Lagos, Putre, Río Verde, Laguna Blanca, Camarones, Ollagüe y María Elena. De este grupo se pueden desprender 2 puntos. En primer lugar, se tiene que las comunas se caracterizan por ser comunas pequeñas y por otro lado son comunas de los extremos del país (5 de las 11 comunas son de Magallanes y 3 de Arica y Parinacota, 1 de Tarapacá y 3 de Antofagasta).

Por otro lado, las comunas con una cobertura mayor al 110% a nivel nacional son las siguientes:

Tabla N°12: Comunas con coberturas mayores a 110%

2018				2019				2020-1			
NOMBRE	RES	PROY	COB	NOMBRE	RES	PROY	COB	NOMBRE	RES	PROY	COB
Calera de Tango	44.454	27.309	162,8%	Calera de Tango	43.278	27.913	155,0%	Calera de Tango	43.846	28.525	153,7%
San Rafael	12.206	9.689	126,0%	San Rafael	12.098	9.825	123,1%	Hijuelas	23.369	19.099	122,4%
Pencahue	10.401	8.539	121,8%	Pencahue	10.147	8.571	118,4%	Pencahue	10.501	8.601	122,1%
Pirque	34.914	28.799	121,2%	Ñiquén	13.567	11.556	117,4%	San Rafael	12.135	9.959	121,8%
Ñiquén	13.587	11.544	117,7%	Palena	2.099	1.822	115,2%	Palena	2.176	1.827	119,1%
Panquehue	8.785	7.472	117,6%	Pirque	33.812	29.616	114,2%	Ñiquén	13.454	11.567	116,3%

Pelarco	10.217	8.873	115,1%	Pelarco	10.236	8.979	114,0%	Pelarco	10.476	9.083	115,3%
San Pablo	12.130	10.541	115,1%	San Pablo	11.931	10.547	113,1%	San Nicolás	13.932	12.172	114,5%
San Nicolás	13.684	11.959	114,4%	Panquehue	8.492	7.553	112,4%	Pirque	34.735	30.433	114,1%
Palena	2.068	1.819	113,7%	Romeral	17.819	15.946	111,7%	San Pablo	11.931	10.553	113,1%
Río Hurtado	4.957	4.371	113,4%	Alhué	7.995	7.214	110,8%	Pichilemu	20.212	17.882	113,0%
Romeral	17.767	15.721	113,0%	Río Negro	15.799	14.303	110,5%	Panquehue	8.526	7.633	111,7%
Río Negro	15.867	14.335	110,7%					Río Negro	15.883	14.275	111,3%
Huasco	12.034	10.936	110,0%					Romeral	17.960	16.170	111,1%
								Alhué	8.191	7.405	110,6%
								Natales	26.302	23.782	110,6%
								Villarrica	65.177	59.103	110,3%

Fuente: Elaboración propia usando la Base Residencia Comunal e información del INE.

Respecto a las comunas con cobertura sobre 110% vemos que el caso atípico de cobertura es el de Calera de Tango. Por otro lado, si bien en el periodo 2020 estamos viendo un aumento en comunas sobre 110% gran parte de estas que se incorporan son muy cercanos al margen del 110%.

Finalmente, se procede a agregar variables adicionales a la base de datos de residencia provenientes de otras bases de datos, tales como la variable de Calificación Socioeconómica (CSE), indicador de pueblos originarios, indicador de discapacidad y un indicador de región, que es una agrupación del campo comunal. Con esto la base residente se encuentra lista para la generación de indicadores por sector.

6. FUENTES DE DATOS UTILIZADAS

A continuación, se presenta un cuadro con las bases utilizadas para los procesos e indicando si son usadas para la determinación de residencia anual o semestral y/o para la determinación de comuna.

Tabla N°13: Fuentes de datos utilizadas

N°	Bases de Datos Consultadas		Anual	Semestral	Comuna
0.	Registro Civil e Identificación	Base de datos recibida a partir de consulta a Registro Civil	2018-2020	2018-2020	2018 - 2020
1.	Matrícula (Preescolar + Mineduc + Superior)	Base de datos de Matrícula del MINEDUC Base de datos que contiene información de Matrícula de educación superior	2018-2020		2014-2020
2.	RIS Beneficiarios	Base de datos que presenta información respecto a beneficios que reciben individuos a partir de los programas RIS de manera mensualizada	2018-2020	2018-2020	
3.	AFC (Afiliados + Cuentas + Rentas)	Base que contiene información respecto a las Cuentas de AFC y permite hacer nexo entre RUN y número de cuenta Base AFC que contiene las Rentas mensuales según cuenta	2018-2020	2018-2020	2016-2020

4.	RSH	Base de datos del Registro Social de Hogares, se tiene el levantamiento de diciembre del año t	2018-2020		2018-2020
5.	AFP (Cotizaciones)	Base que contiene pagos de cotizaciones y deudas previsionales	2018-2020	2018-2020	
6.	Pensiones (Contributivas + No contributivas)	Base que contiene pagos a pensionados por Vejez, Invalidez y Supervivencia	2018-2020	2018-2020	
7.	Superintendencia de Salud (Cotizantes)	Base de datos de la Superintendencia de Salud conteniendo información de Cotizantes, Coberturas adicionales y licencias.	2018-2020	2018-2020	2016-2020
8.	Superintendencia de Salud (Cargas)	Base de datos de la Superintendencia de Salud conteniendo información de Beneficiarios, Coberturas adicionales y licencias.	2018-2020	2018-2020	
9.	Fonasa (Cotizantes)	Base del Fondo Nacional de Salud que contiene información sobre cotizantes	2018-2020	2018-2020	
10.	Fonasa (Cargas)	Base del Fondo Nacional de Salud que contiene información sobre beneficiarios	2018-2020	2018-2020	
11.	SII	Cada RUN con boleta de honorarios emitidas en el año comercial respectivo	2018-2020		
12.	RCivil-1 (Residencia de Padre/Madre->Hijo 0-10)	Base de datos en la que a cada run se le asocia el de su padre y madre además del sexo, las fechas de nacimiento y defunción.	2018-2020		
13.	RCivil-2 (Residencia Hijo 0-10->Padre/Madre)	Base de datos en la que a cada run se le asocia el de su padre y madre además del sexo, las fechas de nacimiento y defunción.	2018-2020		
14.	RCivil-3 (Hijo 0-1 con comuna -> Hijo residente)	Base de datos en la que a cada run se le asocia el de su padre y madre además del sexo, las fechas de nacimiento y defunción.	2018-2020		

Fuente: Elaboración propia usando la Base Residente e información del INE.

ANEXOS

Anexo N°1: Indicadores de movimiento usados para identificar condición de residencia en el país

N°	Bases de Datos Consultadas	Definición del umbral de movimiento anual o semestral
1.	Matrícula (Preescolar + Mineduc + Superior)	Cada RUN con matrícula en el año.
2.	RIS Beneficiarios	Cada RUN que recibe algún beneficio en el año o semestre.
3.	AFC (Afiliados + Cuentas + Rentas)	Cada RUN que presenta cotizaciones positivas en al menos un mes del año o semestre.
4.	RSH	Cada RUN con entrevista en el año, modificación en año, o modificación de comuna en año.
5.	AFP (Cotizaciones)	Cada RUN que presenta cotizaciones positivas en al menos un mes del año o semestre.
6.	Pensiones (Contributivas + No contributivas)	Cada RUN con montos pagados positivos en algún mes del año o semestre.
7.	Superintendencia de Salud (Cotizantes)	Cada RUN correspondiente a las cotizaciones vigentes, con monto de cotizaciones mensuales del año o semestre son positivas.
8.	Superintendencia de Salud (Cargas)	Cada RUN correspondiente a cargas de cotizantes vigentes, cuyas cotizaciones fueron realizadas en diciembre de año o semestre.
9.	Fonasa (Cotizantes)	Cada RUN con fecha de información en el año, con cotizaciones positivas en algún mes del año o semestre.
10.	Fonasa (Cargas)	Cada RUN correspondiente a cargas de cotizantes identificados como "residentes" en el año o semestre.
11.	SII	Cada RUN con boleta de honorarios emitidas en el año comercial respectivo
12.	RCivil-1 (Residencia de Padre/Madre->Hijo 0-10)	Cada RUN de hijos menores de 10 años con padres identificados como residentes en las bases anteriores.
13.	RCivil-2 (Residencia Hijo 0-10->Padre/Madre)	Cada RUN de madres y padres con hijos menores de 10 años identificados como residentes en las bases anteriores.
14.	RCivil-3 (Hijo 0-1 con comuna -> Hijo residente)	Cada RUN de niños/as de 0 a 1 año que aparezca con comuna en la base del Registro Civil.

Anexo N°2: Clasificación por grupos de base de datos con información comunal para Base Residente 2019

Base de Datos	Criterio utilizado
Base de Registro Civil	Se utiliza la comuna declarada en el último trámite realizado por la persona. Sobre lo anterior no se tiene la fecha de cuando fue realizado
Mineduc Matriculas	<p>Caso Anual: Al tener registros sin fecha, y tampoco teniendo certeza de que en cada año se rectifica la comuna de residencia, se genera una base para cada RUN con la serie de comunas registradas en cada año en la base de matrículas. Para identificar la temporalidad del registro se buscó las series que tienen un cambio de comuna entre un periodo y otro. En caso de encontrarse un cambio, se declara dicho periodo como la temporalidad del registro de comuna, en caso de no poder identificar dicho cambio, no se considera la información comunal para el individuo.</p> <p>Caso Semestral: Dada la naturaleza del registro no es posible separar en 2 semestres. Ante lo anterior se toma como aproximación declarar la comuna anual como comuna del primer semestre.</p>
RSH	<p>Caso Anual: Base de datos con información comunal. Se consideraron los individuos a los cuales se les pudo identificar en qué fecha declararon dicha comuna de residencia. Lo anterior se realizó mediante la fecha en la cual fueron encuestados por primera vez, o con la fecha de cuando cambiaron domicilio. Adicional a lo anterior, también con la serie de meses de comunas declaradas se identificó en las series que tienen un cambio de comuna entre un mes y otro. En caso de encontrarse un cambio, se declara dicho periodo como la temporalidad de registro de comuna, en caso de no poder identificar por ninguno de estos medios, entonces no se considera la información comunal para el individuo.</p> <p>Caso Semestral: En lugar de identificar el periodo según los criterios anteriores, se identifica el semestre respectivo.</p>
AFC	<p>Caso Anual: Al tener registros sin fecha, y tampoco teniendo certeza de que en cada año se rectifica la comuna de residencia, se genera una base para cada RUN con la serie de comunas registradas en cada año en la base de AFC. Para identificar la temporalidad del registro se buscó las series que tienen un cambio de comuna entre un periodo y otro. En caso de encontrarse un cambio, se declara dicho periodo como la temporalidad del registro de comuna, en caso de no poder identificar dicho cambio, no se considera la información comunal para el individuo.</p> <p>Caso Semestral: Dada la naturaleza del registro no es posible separar en 2 semestres. Ante lo anterior se toma como aproximación declarar la comuna anual como comuna del primer semestre.</p>
Superintendencia de Salud – Cotizantes	<p>Base Anual: Base de datos con información comunal mensual, sin embargo, mes a mes no se rectifica que efectivamente sea la comuna de residencia del individuo. A partir de esto se rescataron las comunas de quienes se suscribieron en el periodo t y quienes en un momento del año realizaron un cambio de comuna de residencia (sabiendo de esa manera que para el periodo t la persona reside en dicha comuna). Después se hace el mismo ejercicio para los periodos anteriores para poder determinar la fecha de cambio de comuna.</p> <p>Base Semestral: Dado que la base viene con información mensual la diferencia con lo anterior es que en lugar de registrar el año del cambio de comuna se registra el semestre en el cual se hizo el cambio de comuna.</p>

<p>Superintendencia de Salud – Cargas</p>	<p>Base Anual: Identificando por el cotizante, se le imputo a las respectivas cargas, la comuna de los cotizantes respectivos (que tienen comuna identificada para el periodo). En los casos donde la carga tuvo 2 cotizantes distintos durante el año, se le imputó la comuna del cotizante con más meses con dicho individuo. Se realizó una limpieza de cargas donde presentaban más de un cotizante para un mes, también se limpió para las cargas que presentaban más de dos cotizantes durante el periodo t.</p> <p>Base Semestral: Para este caso la diferencia respecto a la anual es que el pareo entre bases se hace con la resultante de cotizantes semestrales.</p>
---	---