

Casen 2022

Nota uso bases de datos

1. Antecedentes

El presente documento tiene por finalidad proporcionar a los usuarios(as) la información necesaria para utilizar las bases de datos de la Encuesta Nacional de Caracterización Socioeconómica Casen 2022, en atención a la representatividad de la encuesta y la disponibilidad de variables de identificación territorial. De esta manera, la publicación de este documento contiene definiciones básicas a tener en consideración, buscando también potenciar el uso adecuado de la información estadística contenida en esta nueva versión de Casen.

El objeto de estudio de la Encuesta Casen 2022 está definido por los hogares y la población residente en viviendas particulares ocupadas del país, exceptuando aquella población residente en algunas áreas especiales previamente definidas por el Instituto Nacional de Estadísticas (INE)¹.

La encuesta Casen tiene un diseño muestral probabilístico, estratificado y bietápico, con representatividad nacional, por área urbano y rural, y regional. Por lo tanto, el análisis de los resultados de la encuesta debe realizarse con atención a este diseño.

En la publicación de la base de datos de Casen 2022, se ha tomado el resguardo de incorporar la información de desagregación territorial correspondiente al diseño muestral de la encuesta, por lo tanto, se mantienen las variables de identificación territorial área y región. Sin embargo, se ha puesto a disposición de investigadores, para realizar análisis con los resguardos detallados en las siguientes secciones, una base de datos complementaria con variables de identificación provincial y comunal.

Los contenidos de la presente nota están estructurados de la siguiente forma: la sección 2 indica la representatividad estadística de la encuesta y sus implicancias en los factores de expansión, así como también presenta las indicaciones de uso necesarias, en atención al diseño muestral de la encuesta; la sección 3 presenta las indicaciones para realizar el correcto cruce entre las bases de datos publicadas; la sección 4 instruye a los(as) usuarios(as) sobre cómo acceder a las bases de datos de la encuesta y otra documentación de apoyo; y la sección 5 describe, en términos generales, la estructura de las bases de datos a disposición.

2. Diseño, representatividad y factores de expansión

El diseño muestral de la Encuesta Casen 2022 considera como variable de interés la estimación de la tasa de pobreza por ingresos (porcentaje de población en situación de pobreza por ingresos) y ha establecido objetivos de precisión para estimar dicho indicador en los siguientes niveles o dominios de representación:

- (a) Nacional
- (b) Nacional urbano y Nacional rural
- (c) Regional

Dado lo anterior, la Encuesta Casen 2022 es representativa en estos 3 dominios, por lo tanto, para

¹ Estas áreas corresponden a comunas completas o parte de comunas cuyas Unidades Primarias de Muestreo (UPM) fueron clasificadas como áreas especiales en el marco muestral de viviendas MMV 2020. Así, las comunas excluidas son: Ollagüe, Juan Fernández, Isla de Pascua, Cochamó, Chaitén, Futaleufú, Hualaihué, Palena, Guaitecas y O'Higgins. También, se excluyen algunas UPM consideradas como áreas especiales en las comunas de: General Lagos (4 UPM), Colchane (5 UPM), Tortel (1 UPM) y Cabo de Hornos (1 UPM). Finalmente, queda excluida la comuna de la Antártica Chilena dado que no se encuentra disponible en el MMV 2020.

realizar estimaciones representativas se debe utilizar el factor de expansión regional (expr).

En Casen 2022, el factor de expansión regional, calculado por el INE, incorpora una nueva metodología de calibración a totales poblacionales denominado metodología de raking², la cual permite controlar diversas variables simultáneamente y mantiene el diseño muestral ya que disminuye el sesgo que se produce por errores no muestrales, provocados por la no-respuesta y, además, mejora la precisión de las estimaciones. En la metodología aplicada a los factores de expansión de Casen 2022 en el proceso de calibración se utiliza la información proveniente de las Proyecciones de Población, base Censo 2017, ajustadas a la población residente en viviendas particulares ocupadas según sexo y edad y para distintos niveles de desagregación³.

Este factor es pertinente de utilizar para la obtención de resultados de personas, hogares y núcleos expandidos que buscan ser representativos a nivel nacional, regional y por área (urbana y rural). Al emplear dicho factor, los resultados expandidos permiten representar al total de la población del país, en congruencia con la proyección de población del INE que se utilizó como referencia para el diseño muestral de la encuesta (base Censo 2017) y, por otro lado, permite representar las características socioeconómicas que comparten las personas, hogares y núcleos expandidos de la población sobre la cual la muestra busca ser representativa (nacional, área y regiones).

Los resultados que se obtengan a partir de la información de la Encuesta Casen 2022 sin haber aplicado el factor de expansión regional, no deben considerarse estadísticamente representativos de la población correspondiente y, en caso de incluirse en documentos de resultados, deben ser acompañados de una nota que señale explícitamente que se trata de resultados no ponderados o sin expansión.

Por otro lado, Casen 2022 dispone de 2 factores de expansión adicionales⁴, dirigidos principalmente a investigadores(as) que deseen producir sus propias estimaciones a nivel provincial y comunal. Sin embargo, deben ser utilizarlos con precaución en su análisis e interpretación, ya que el diseño de la encuesta no considera representatividad para dichos niveles, por lo que estimaciones directas de la encuesta pueden ser estadísticamente poco fiables o sesgadas.

Los factores de expansión provincial y comunal, a diferencia del factor de expansión regional que utiliza el método de calibración por raking, mantienen la metodología de calibración utilizada en las versiones previas de Casen 2022, el cual utiliza el método de calibración a totales poblacionales a los respectivos niveles territoriales. Esto se debe a que los niveles de precisión de variables usadas para la calibración por raking a nivel regional como sexo y edad, en los niveles provinciales y comunales muestran niveles de precisión muy bajos.

El ajuste de calibración, realizado al factor de expansión “provincial”, se obtiene como la razón entre el total de personas en este grupo poblacional según las estimaciones demográficas del INE y el total de personas residentes en viviendas particulares estimado para cada provincia a partir de la encuesta. Asimismo, el ajuste “comunal”, se obtiene como la razón entre el total de personas en la comuna según las estimaciones demográficas del INE y el total de personas residentes en viviendas particulares estimadas para cada comuna a partir de la encuesta.

Considerando el diseño muestral de la encuesta Casen, sus niveles de representatividad y las

² Para más detalles sobre esta metodología revisar: [Doc INE: Nueva metodología calibración ENE, metodología raking](#)

³ Para más detalle de esta nueva metodología de calibración del factor de expansión, se puede revisar Documento Diseño Muestral Casen 2022, apartado VI. Diseño Factores de expansión el cual será publicado en el mes de agosto de 2022 en el sitio web de Casen 2022, sección Metodologías: <https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2022>

⁴ En esta versión de Casen disponibles en la Base de datos provincia y comuna Casen 2022

limitaciones de precisión a nivel provincial y comunal, el Ministerio no recomienda la realización de estimaciones a nivel provincial y comunal, por lo que para la versión 2022 de la Encuesta Casen, el Ministerio presenta en la base de datos de Casen 2022 sólo la información de región de la vivienda de la encuesta y el factor de expansión regional.

Por otro lado, pone a disposición, con las precauciones mencionadas, una base de datos que contiene como información complementaria a la encuesta Casen 2022, las variables de provincia, comuna y los factores de expansión provincial y comunal, para ser utilizado con los resguardos ya mencionados.

Finalmente, en caso de realizarse estimaciones a nivel provincial o comunal, se recomienda ser explícitos al momento de publicación indicando que no es posible asegurar la representatividad de los resultados obtenidos en base a la siguiente tabla. Por otro lado, las estimaciones en estos niveles deben ser evaluadas, con atención a la calidad y precisión de los datos según se indica en el Manual del Investigador Casen 2022⁵.

Tabla 01: Uso de niveles territoriales y factor de expansión Casen 2022

Factor de expansión	Dominio	Representatividad	Publicación	Metodología Casen 2022
Regional (expr)	Sí, Casen 2022 fue diseñada considerando el territorio a nivel nacional (región, área) para obtener estimaciones confiables.	Sí, usando el factor de expansión regional (expr), los resultados obtenidos tienen representatividad a nivel nacional, nacional urbano, nacional rural y regional.	Para publicar resultados se debe evaluar la calidad y precisión de las estimaciones de acuerdo con lo sugerido por el INE ⁶ .	Metodología raking
Provincial (expp)	No, Casen no considera en su diseño las provincias como dominio para obtener estimaciones confiables.	No, usando factor de expansión provincial (expp), los resultados obtenidos no son representativos.	No recomendable. En el caso de hacerse se debe indicar explícitamente que no son resultados representativos y realizar análisis de calidad sugerido por el INE.	Metodología de totales poblacionales
Comunal (expcc)	No, Casen no considera en su diseño las comunas como dominio para obtener estimaciones confiables.	No, usando factor de expansión comunal (expcc), los resultados obtenidos no son representativos.	No recomendable. En el caso de hacerse se debe indicar explícitamente que no son resultados representativos y realizar análisis de calidad sugerido por el INE.	Metodología de totales poblacionales

Fuente: elaboración propia MDSF

3. Cruce entre ambas Bases de datos

Para poder realizar el cruce de ambas bases de datos, esta se debe realizar con la llave de variables de identificación del hogar y la persona, es decir, “folio” e “id_persona” presente en ambas bases de datos. La identificación de la llave es:

Tabla 02: Variables de cruce de ambas Bases de datos

Variable	Glosa
----------	-------

⁵ Para mayor detalle revisar Manual del Investigador Casen en Pandemia 2020, apartado 4.6. Paso 6: Evaluación de la calidad y precisión de las estimaciones, disponible en: [Manual del Investigador Casen en Pandemia 2020](#)

⁶ Ver nota al pie N°5

folio	Identificación hogar (id_vivienda hogar)
id_persona	Identificación de la persona en el hogar

Fuente: elaboración propia MDSF

4. Acceso a los datos encuesta Casen 2022

Las bases de datos de la encuesta Casen 2022 se encuentran publicadas en la sección “Bases de datos” del portal web del Observatorio Social (observatorio.gob.cl) y del Banco Integrado de Datos del Ministerio de Desarrollo Social y Familia (BIDAT) (bidat.gob.cl), junto con otros documentos y archivos asociados al estudio.

4.1. Acceso a las bases de datos

Bases disponibles en Sección **Bases de datos**:

- Base de datos Casen 2022
- Base de datos provincia y comuna Casen 2022

4.2. Acceso a documentos o archivos de apoyo

Libros de códigos de Bases de datos disponibles en Sección **Bases de datos**:

- Libro de códigos base de datos Casen 2022
- Libro de códigos base de provincia y comuna Casen 2022

Cuestionario disponible en Sección **Cuestionarios**:

- Cuestionario Casen 2022

Documentos metodológicos en Sección **Metodologías**:

- Diseño muestral Casen 2022
- Manual del investigador Casen 2022

5. Descripción de las bases de datos

La base de datos de la encuesta Casen 2022 contiene 72.056 hogares logrados⁷, los que corresponden a 70.751 viviendas y 202.231 personas con información. Los datos proporcionados en la base de datos deben cumplir con los protocolos de anonimización, por lo que se excluye toda la información que haga referencia explícita a las personas encuestadas o su ubicación.

De esta manera, la base de datos Casen 2022 contiene 917 variables las cuales se estructuran de la siguiente manera:

Tabla 03: Estructura de la Base de datos Casen 2022

Tipo de variables	Nº variables
Identificación (hoja de ruta)	17
Factores de expansión y varianza compleja	4
Módulos del Cuestionario	611
Ingresos generados por Cepal	220
Ingresos generados por MDSF	12
Pobreza multidimensional	25

⁷ Se considera un hogar o encuesta lograda aquella encuesta que ha pasado exitosamente por todos los procesos de supervisión y validación.

Adicionales generadas por MDSF	28
Total variables base de datos Casen 2022	917

Fuente: elaboración propia MDSF

Así mismo, la Base de datos provincia y comuna Casen 2022 también contiene 72.056 hogares y 202.231 personas con información de 6 variables las cuales son:

Tabla 04: Variables de la Base de datos de provincia y comuna Casen 2022

Variable	Glosa
folio	Identificación hogar (id_vivienda hogar)
id_persona	Identificación de la persona en el hogar
provincia	Provincia
comuna	Comuna
expp	Factor de expansión provincial
expc	Factor de expansión comunal

Fuente: elaboración propia MDSF