

Nota uso bases de datos Casen 2024

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 2024, en atención a la representatividad de la encuesta, la disponibilidad de variables de identificación territorial y la identificación de las unidades de análisis. 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.

La población objetivo de la Encuesta Casen 2024 está definida 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 2024, 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, adicionalmente, se ha puesto a disposición de personas investigadoras, 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 se estructuran de la siguiente forma: la sección 2 se refiere a al diseño, representatividad estadística de la encuesta y sus implicancias en el uso de los factores de expansión. La sección 3 se refiere a las distintas unidades de análisis de la encuesta. La sección 4 explica cómo realizar el correcto cruce entre las bases de datos publicadas. La sección 5 instruye a los(as) usuarios(as) sobre cómo acceder a las bases de datos de la encuesta y su documentación de apoyo. Y la sección 6 describe, en términos generales, la estructura de las bases de datos publicadas.

2. Diseño muestral, representatividad y factores de expansión

El diseño muestral de la Encuesta Casen 2024 considera como variable de interés la estimación de la tasa de pobreza por ingresos (porcentaje de población en situación de pobreza por ingresos), metodología 2013, 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

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 2021. 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, Colchane, Tortel y Cabo de Hornos. Finalmente, queda excluida la comuna de la Antártica dado que no se encuentra disponible en el MMV 2021.

La Encuesta Casen 2024 permite obtener estimaciones representativas para la variable de interés en cualquiera de estos tres dominios de representación antes indicados, para lo cual se debe utilizar el factor de expansión regional (expr). Al momento de generar estimaciones para otras variables y/o desagregaciones se recomienda evaluar su precisión según se indica en el Manual para la Investigación².

2.1. Factor de expansión regional

En Casen 2024, el factor de expansión regional, calculado por el INE, incorpora la metodología de calibración a totales poblacionales denominada metodología de *raking*³, la cual permite reducir el sesgo asociado a errores no muestrales, provocados por la no respuesta y, además, mejorar la precisión de las estimaciones. La metodología de calibración aplicada a los factores de expansión de Casen 2024 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 desagregación⁴.

El factor de expansión regional debe utilizarse para la obtención de resultados a nivel de personas, hogares y núcleos que buscan ser representativos a nivel nacional, regional y por área (urbana y rural). Al emplear dicho factor, los resultados expandidos permiten representar las características socioeconómicas de las personas, hogares y núcleos de la población sobre la cual la muestra busca ser representativa, en congruencia con la proyección de población del INE utilizada como referencia para el diseño muestral de la encuesta (base Censo 2017).

Si bien, por diseño, la encuesta es representativa a los niveles nacional, regional y por área (urbano y rural), al momento de generar estimaciones se recomienda evaluar su precisión según se indica en el Manual para la Investigación.

Los resultados que se obtengan a partir de la información de la Encuesta Casen 2024 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 “m muestrales”, “no ponderados” o “sin expansión”.

2.2. Factores de expansión provincial y comunal

La Encuesta Casen 2024 dispone de dos factores de expansión adicionales a nivel territorial⁵, dirigidos principalmente a personas investigadoras que busquen producir sus propias estimaciones a nivel provincial y comunal. Sin embargo, ambos factores deben ser utilizados 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 a Casen 2022, que emplea el método de calibración a totales poblacionales para los respectivos niveles territoriales.

² Para más detalle sobre la evaluación de precisión de las estimaciones, revisar el capítulo 4.6 Paso 6: Evaluación de la calidad y precisión de las estimaciones del Manual para la Investigación. Ver: [Manual para la Investigación Casen 2022](#)

³ Para más detalles sobre esta metodología revisar: [Nueva metodología calibración ENE, metodología raking \(INE\)](#)

⁴ Para más detalle de la metodología de calibración del factor de expansión, se puede revisar Documento Diseño Muestral Casen 2022, apartado VI. Diseño Factores de expansión publicado en el sitio web de Casen 2022, sección Metodologías: <https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2022>

⁵ Los factores de expansión de provincia y comuna de Casen 2024 se encuentran disponibles en la base de datos provincia y comuna.

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 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 2024 de la Encuesta Casen, el Ministerio presenta en la base de datos de Casen 2024 solo 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 2024, 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 indicar explícitamente que no es posible asegurar la representatividad de los resultados obtenidos para dichos niveles de desagregación en base a la Tabla 1. Además de evaluar su calidad según los criterios antes señalados.

Tabla 1: Uso de niveles territoriales y factor de expansión Casen 2024

Factor de expansión	Nivel de estimación	Representatividad	Publicación	Metodología Casen 2024
Regional (expr)	Sí, Casen 2024 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 <i>raking</i>
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 (expc)	No, Casen no considera en su diseño las comunas como dominio para obtener estimaciones confiables.	No, usando factor de expansión comunal (expc), 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

⁶ Ver nota al pie N°2

2.3. Factor de expansión de orientación sexual e identidad de género

El módulo de orientación sexual e identidad de género (OSIG) de la Encuesta Casen 2024 permite disponer de información sobre la orientación sexual e identidad de género de personas de 18 años o más, aportando evidencia importante para el diseño y evaluación de políticas públicas.

Con el fin de resguardar la privacidad, las respuestas de este módulo son entregadas directamente por la persona entrevistada, lo que implica que solo se aplican a quienes están presentes en el hogar al momento de realizar la entrevista. El bloque incluye preguntas que indagan, por un lado, sobre la atracción afectiva o sexual que una persona puede tener hacia el sexo opuesto, el mismo sexo, más de un sexo u otro; y por otro, al género con el cual la persona se identifica, considerando las categorías masculino, femenino, transmasculino, transfemenino, no binario u otro.

Dado que no todas las personas tienen la misma probabilidad de responder a estas preguntas, las estimaciones basadas únicamente en quienes las contestan podrían presentar sesgo de selección. En particular, si quienes se abstienen de responder difieren sistemáticamente de quienes sí lo hacen, los resultados reflejarían solo las características de un subconjunto de la población.

Con el objetivo de garantizar que las estimaciones reflejen adecuadamente a toda la población de 18 años o más, en relación con su identidad de género y orientación sexual, se construyó un factor de expansión específico para el módulo OSIG. Este factor permite caracterizar a la población corrigiendo el sesgo generado por la no respuesta al redistribuir el peso de quienes no respondieron en aquellos que sí lo hicieron a partir de sus diferencias observables. De esta manera, las estimaciones resultantes son representativas de la población de 18 años o más, a nivel nacional, nacional por área y regional, al igual que el factor de expansión regional.

3. Unidad de análisis

Las Encuesta Casen 2024 se aplica sobre una muestra de viviendas, dentro de las cuales se busca entrevistar a todos los hogares y se realizan preguntas sobre todos los residentes habituales de cada hogar. Dentro de un hogar puede haber uno o más núcleos familiares dependiendo de las relaciones que existan entre los integrantes del hogar.

A partir de la base de datos de la Encuesta Casen 2024 es posible realizar análisis en distintos niveles: población general, vivienda, hogar y núcleos familiares. En la Tabla 2, a continuación, se especifican las consideraciones que deben tenerse para efectuar cada uno de estos análisis.

Tabla 2: Especificaciones según unidad de análisis Casen 2024

Unidad de análisis	Especificaciones
Población general	No es necesario establecer ningún tipo filtro. Esta opción es la recomendada cuando se busca analizar características individuales de la población encuestada.
Vivienda	La variable "id_vivienda" corresponde al identificador único a nivel de vivienda. Para realizar análisis a nivel de vivienda, es recomendable generar un marcador a partir de dicha variable.
Hogar	La variable "folio" es el identificador único a nivel de hogar. La variable "pco1_a", permite identificar a la jefatura de hogar. De este modo, para realizar análisis a nivel de hogar se recomienda utilizar el filtro "pco1_a = 1"
Núcleo familiar	Las variables "folio" y "nucleo" permiten construir un identificador único a nivel de núcleo familiar. La variable pco2_a, permite identificar a la jefatura de núcleo. De este modo, para realizar análisis a nivel de núcleo se recomienda utilizar el filtro "pco2_a = 1".

Unidad de análisis	Especificaciones
	<p>El Servicio Doméstico Puertas Adentro (SDPA) queda clasificado por defecto en un núcleo al que se le asigna valor 0. Las personas pertenecientes a “núcleo = 0”, no tienen datos para pco2_a.</p> <p>Lo anterior se debe a que esta población no se considera en los indicadores de pobreza por ingresos y multidimensional, ni en el conjunto de medidas de ingreso y estratificación socioeconómica (deciles y quintiles).</p>

4. Cruce entre ambas Bases de datos

Para poder vincular la base de datos de la Encuesta Casen 2024 y la base de datos de provincia y comuna, se debe utilizar como llave la identificación del hogar y la persona, esto es, las variables “folio” e “id_persona”, respectivamente, presentes en ambas bases de datos (ver Tabla 3).

Tabla 3: Variables de cruce de ambas Bases de datos

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

Fuente: elaboración propia MDSF

5. Acceso a los datos encuesta Casen 2024

Las bases de datos de la encuesta Casen 2024 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.

5.1. Acceso a las bases de datos

Bases de datos disponibles en sección “Bases de datos”:

- Base de datos Casen 2024 (formato .dta, .sav y .RData).
- Base de datos provincia y comuna Casen 2024 (formato .dta, .sav y .RData).

5.2. 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 2024
- Libro de códigos base de provincia y comuna Casen 2024

Cuestionario disponible en sección “Cuestionarios”:

- Cuestionario Casen 2024

Documentos metodológicos en sección “Metodología”:

- Diseño muestral Casen 2024
- Manual para la investigación Casen 2024

6. Descripción de las bases de datos

La base de datos de la encuesta Casen 2024 contiene 78.654 hogares logrados⁷, los que corresponden a 77.618 viviendas y 218.367 personas con información. Los datos proporcionados en la base de datos excluyen toda la información que haga referencia explícita a las personas encuestadas o su ubicación con el fin de cumplir con los protocolos de anonimización.

La base de datos Casen 2024 contiene 877 variables estructuradas de la siguiente manera:

Tabla 4: Estructura de la Base de datos Casen 2024

Tipo de variables	N° variables
Identificación (hoja de ruta)	18
Factores de expansión y varianza compleja	4
Módulos del Cuestionario	551
Ingresos generados por Cepal	209
Ingresos generados por MDSF	15
Pobreza multidimensional	41
Adicionales generadas por MDSF	39
Total variables base de datos Casen 2024	877

Fuente: elaboración propia MDSF

La Base de datos de provincia y comuna Casen 2024 también contiene 78.654 hogares y 218.367 personas, pero incluye únicamente las seis variables descritas en la Tabla 5, que se presentan a continuación:

Tabla 5: Variables de la Base de datos de provincia y comuna Casen 2024

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

⁷ Un hogar logrado es aquél que ha sido entrevistado y ha pasado exitosamente por todos los procesos de supervisión y validación.