

Nota uso bases de datos consolidadas Encuesta Casen

1. Antecedentes

Con el objetivo de poner a disposición de la comunidad usuaria información de calidad para la toma de decisiones, la División de Observatorio Social (en adelante, la DOS) de la Subsecretaría de Evaluación Social, responsable de la Encuesta Casen, revisa de manera permanente sus procesos estadísticos, para la mejora de los productos que derivan de ellos.

Durante los últimos años, la Encuesta de Caracterización Socioeconómica Nacional (Casen) ha sido actualizada en la metodología de cálculo de los factores de expansión¹ y de la estimación de las medidas de pobreza por ingresos² y multidimensional³, entre otros aspectos. Como resultado de estos procesos, se han publicado bases de datos complementarias que incorporan los cambios derivados de dichas definiciones metodológicas.

La coexistencia de múltiples versiones de las bases de datos para cada año de la encuesta podría llevar a errores de cálculo en las estimaciones por parte de las personas usuarias de los conjuntos de datos. Para abordar esta situación, la DOS se encuentra desarrollando bases de datos únicas por año para los años 2006 a 2022, que integran las últimas definiciones metodológicas, lo que facilita la comparabilidad de resultados entre versiones y la generación de estimaciones consistentes con los datos oficiales publicados por el Ministerio de Desarrollo Social y Familia.

El presente documento tiene por finalidad proporcionar a las y los usuarios la información necesaria para el adecuado uso de las bases de datos consolidadas de la Encuesta Nacional de Caracterización Socioeconómica Nacional entre los años 2006 y 2022⁴, disponibles en la sección “bases de datos” para cada año de la encuesta; identificando sus principales características y cambios respecto de lo previamente publicado.

El contenido de esta nota se estructura de la siguiente forma: la segunda sección describe las bases de datos consolidadas y explica cómo utilizarlas; la tercera sección presenta los principales cambios con respecto a las bases de datos ya publicadas; y la cuarta sección indica dónde se encuentran las bases de datos consolidadas.

2. Bases de datos consolidadas

2.1. ¿Qué y cómo son las bases de datos consolidadas?

Las bases de datos consolidadas fueron elaboradas a partir de las bases de datos ya publicadas para cada año —disponibles en la sección “Bases históricas”— y las variables asociadas a las metodologías de pobreza 2024 calculadas retrospectivamente, para 2022 en el caso de la pobreza multidimensional, y para el período 2022-2009 en el de la pobreza por ingresos. Sus principales atributos son los siguientes:

- Permiten realizar, con una única base de datos por año, estimaciones consistentes con la metodología de pobreza por ingresos 2024 y 2013, así como con la metodología de cálculo de

¹ Ver [Nota Técnica N°1 de la Encuesta Casen 2022](#)

² Ver [Metodología de Medición de la Pobreza por Ingresos en Chile. Actualización 2024](#)

³ Ver [Metodología de medición de la Pobreza multidimensional en Chile. Actualización 2024](#)

⁴ Esta primera publicación considera las bases de datos consolidadas para los años 2013, 2015, 2017 y 2020. La base de datos consolidada del año 2022 ya se encuentra publicada. Las bases de datos consolidadas para los años 2006, 2009 y 2011 se publicarán en marzo de 2026.

los factores de expansión basada en calibración por *raking*⁵ y en las proyecciones de población del Censo 2017. Es decir, utilizando las metodologías más recientes disponibles.

- Con respecto a la pobreza multidimensional, posibilitan obtener las estimaciones para 5 dimensiones con la metodología 2024 y 2015, y para 4 dimensiones, siempre que los indicadores estén disponibles para cada año.
- Adicionalmente, las bases de datos consolidadas cuentan con cierto nivel de homologación entre los distintos años en cuanto al orden que siguen las variables en la base de datos, el nombre de variables y las etiquetas de variable. Asimismo, incorporan criterios de armonización más específicos para ciertas variables clave de la encuesta, tales como pobreza por ingresos y pobreza multidimensional.

En la Tabla 1, a continuación, se muestra la estructura general que siguen todas las bases de datos consolidadas:

Tabla 1: Estructura general de las bases de datos consolidadas 2013-2022

Grupo variables	Descripción
Identificación	Identifican a las unidades de análisis de encuesta: vivienda, hogar y persona.
Directorio	Variables de identificación geográfica de la vivienda: región, área (urbano o rural), unidad primaria de muestreo, estrato.
Diseño muestral	<ul style="list-style-type: none"> - Factor de expansión regional metodología <i>raking</i> con proyecciones de población 2017. - Factor de expansión de orientación sexual e identidad de género metodología <i>raking</i> con proyecciones de población Censo 2017 (Casen 2022 y 2024). - Factor de expansión orientación sexual e identidad de género con proyecciones de población Censo 2022 (Casen 2015 y 2017). - Factor de expansión de satisfacción con la vida (Casen 2013). - Variables para el cálculo de la varianza compleja.
Hoja de ruta	<ul style="list-style-type: none"> - Fecha de la entrevista. - Variables de entorno.
Cuestionario	<ul style="list-style-type: none"> - Identificación de hogares en la vivienda. - Variables asociadas al cuestionario.
Ingresos Cepal	<ul style="list-style-type: none"> - Variables de ingresos a nivel de persona, hogar y corregidos, elaboradas por Cepal. - Pobreza metodología 2024. - Pobreza metodología 2013.
Ingresos MDSF	<ul style="list-style-type: none"> - Variables de ingresos corregidos a nivel de persona, hogar y per cápita, elaboradas por la DOS. - Deciles y quintiles elaborados por la DOS.
Pobreza Multidimensional	<ul style="list-style-type: none"> - Variables de pobreza multidimensional 5 dimensiones metodología 2024 y 2015. - Variables de pobreza multidimensional 4 dimensiones.
Variables MDSF	Variables auxiliares asociadas al cuestionario elaboradas por la DOS.

Fuente: elaboración propia MDSF

2.2. ¿Cómo se utilizan las bases de datos consolidadas?

La Tabla 2, a continuación, muestra los nombres de variables clave para cada año de la Encuesta Casen entre 2013 y 2022 de manera comparativa con el fin de facilitar su uso.

⁵ La metodología de calibración por *raking* se encuentra implementada en el factor de expansión regional desde 2006 y en el factor de expansión de orientación sexual e identidad de género desde 2022.

Tabla 2: Uso de las bases de datos consolidadas 2013-2022

Grupo variables	Descripción	Casen 2013	Casen 2015	Casen 2017	Casen 2020	Casen 2022
Identificación	Identificador único de la vivienda	id_vivienda	id_vivienda	id_vivienda	id_vivienda	id_vivienda
	Identificador único del hogar	folio	folio	folio	folio	folio
	Identificador de la persona en el hogar	o	o	o	id_persona	id_persona
Directorio	Región (códigos únicos territoriales vigentes en el año de la encuesta)	region	region	region	region	region
	Región (15 regiones)	-	-	region_15 ⁶	-	-
	Región en orden geográfico	region_geo	region_geo	region_geo	region_geo	region_geo
	Región en orden geográfico (15 regiones)	-	-	region_15_geo ⁷	-	-
	Área geográfica (urbano o rural)	area	area	area	area	area
	Unidad primaria de muestreo o segmento de diseño	segmento	-	segmento	segmento y cod_upm	cod_upm
	Estrato de diseño	estrato	-	estrato	estrato	estrato
Diseño muestral	Factores de expansión regional	expr	expr	expr	expr	expr
	Factor de expansión de orientación sexual e identidad de género	-	expr_div	expr_div	-	expr_osig
	Factor de expansión de satisfacción con la vida	expr_r20	-	-		-
	Estratos y conglomerados de varianza	varstrat varunit	varstrat varunit	varstrat varunit	varstrat varunit	varstrat varunit
Hoja de ruta	Fecha de la entrevista y variables de entorno	fecha	fecha_dia fecha_mes fecha_agno	fecha_dia fecha_mes fecha_agno	fecha_entrev	fecha_entrev p1, p2, p3 y p4
Cuestionario	Número de residentes habituales de la vivienda	-	-	-	p6	p9
	Identifica a las viviendas donde hay más de un hogar	-	-	-	p7	p10
	Señala cuántos hogares hay en la vivienda	-	-	-	p8	p11

⁶ La variable “region_15” contiene 15 regiones, según los códigos únicos territoriales utilizados hasta Casen 2015.

⁷ La variable “region_15_geo” contiene 15 regiones, según los códigos únicos territoriales utilizados hasta Casen 2015.

Grupo variables	Descripción	Nombre variable				
		Casen 2013	Casen 2015	Casen 2017	Casen 2020	Casen 2022
	Número de integrantes del hogar	-	tot_per_h	tot_per_h	tot_per_h	tot_per_h
	Identificación jefatura de hogar	pco1	pco1	pco1	pco1	pco1
	Identificación núcleo familiar	nucleo	nucleo	nucleo	nucleo	nucleo_ajustado ⁸
	Identificación jefatura de núcleo	pco2	pco2	pco2	pco2	pco2
Ingresos Cepal	Ingresos persona	La etiqueta de variable solo especifica el tipo de ingreso				
	Ingresos hogar	La etiqueta de variable especifica el tipo de ingreso + “del hogar”				
	Ingresos corregidos persona	La etiqueta de variable especifica el tipo de ingreso + “corregido”				
	Ingresos corregidos del hogar	La etiqueta de variable especifica el tipo de ingreso + “corregido del hogar”				
	Pobreza	pobreza	pobreza	pobreza	pobreza	pobreza
	Pobreza metodología 2013	pobreza_2013	pobreza_2013	pobreza_2013	pobreza_2013	pobreza_2013
Ingresos MDSF	Corrientes de ingreso corregido persona	La etiqueta de variable especifica el tipo de ingreso + “corregido”				
	Corrientes de ingreso corregido hogar	La etiqueta de variable especifica el tipo de ingreso + “corregido del hogar”				
	Corrientes de ingreso corregido p/c	La etiqueta de variable especifica el tipo de ingreso + “per cápita del hogar corregido”				
	Deciles y quintiles	dau qaut dautr qautr	dau qaut dautr qautr	dau qaut dautr qautr	dau qaut dautr qautr	dau qaut dautr qautr
	Pobreza severa	-	-	-	-	pobreza_severa
	Pobreza multidimensional 5 dimensiones	-	-	-	-	pobreza_multi
Pobreza Multidimensional	Pobreza multidimensional 5 dimensiones (metodología 2015)	-	pobreza_multi_2015	pobreza_multi_2015	-	pobreza_multi_2015
	Pobreza multidimensional 4D	pobreza_multi_4d	pobreza_multi_4d	pobreza_multi_4d	-	pobreza_multi_4d
Variables MDSF	Variables auxiliares creadas a partir del cuestionario	tot_hog	tot_hog	disc_wg tot_hog	tot_hog	disc_wg tot_hog

Fuente: elaboración propia MDSF

⁸ En la variable “nucleo_ajustado” de Casen 2022 la jefatura de hogar siempre se encuentra en el nucleo 1. A diferencia de la variable “nucleo”, también presente en 2022, donde no todas las jefaturas de hogar se encuentran en el nucleo 1.

3. Principales cambios con respecto a bases publicadas

La Tabla 3, a continuación, muestra los principales cambios en variables clave implementados en las bases consolidadas en relación con las bases ya publicadas o históricas. Se sugiere considerar estos ajustes antes de utilizar las bases de datos consolidadas.

Tabla 3: Principales cambios con respecto a las bases de datos publicadas

Variable	Ajuste
area	La variable "zona" se renombra como "area" entre 2013 y 2020, quedando como "area" para todos los años.
region y region_geo	La variable "region" corresponde a la región donde se sitúa la vivienda en el orden que siguen los códigos únicos territoriales regionales oficiales en el año de referencia de la encuesta. Se agrega la variable "region_geo", que corresponde a la región donde se sitúa la vivienda, pero en orden geográfico.
expr	La variable "expr" corresponde al factor de expansión regional vigente para todos los años (se eliminan sufijos).
tot_per_h	La variable "tot_per" se renombra y queda como "tot_per_h" para todos los años donde existe e indica el número de integrantes del hogar.
pobreza	Indica situación de pobreza por ingresos metodología 2024 para todos los años. Se homologan las categorías de la variable para todos los años.
pobreza_2013	Indica situación de pobreza por ingresos con la metodología 2013 para todos los años. Se homologan las categorías de la variable para todos los años.
pobreza_multi	Indicador de pobreza multidimensional metodología 2024 (disponible para 2022 y 2024)
pobreza_multi_2015	Indicador de pobreza multidimensional metodología 2015. Se homologan las categorías la variable para todos los años.
educc	La variable "educ", que indicaba el "nivel educacional", se reemplaza por la variable "educc", que corresponde al nivel educacional agregado.
nhog_viv	Variable eliminada por duplicidad con tot_hog (número de hogares declarados en la vivienda)

Fuente: elaboración propia MDSF

4. ¿Dónde se encuentran las bases de datos consolidadas?

Las bases de datos consolidadas de la Encuesta Casen se encuentran publicadas en la sección “Bases de datos” para cada año de la Encuesta en el 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 sus respectivos libros de códigos.

Las bases de datos consolidadas se encuentran disponibles en tres formatos: .dta, .sav y .RData, junto a sus libros de códigos en formato Excel.

Los códigos únicos territoriales oficiales para provincia y comuna, y sus respectivos factores de expansión entre 2013 y 2022, se encuentran disponibles en una base de datos independiente en la sección “Bases de datos”, junto a sus libros de códigos. Para el año 2013 solo existe el factor de expansión de comuna, por lo que esta base se llama “Base de comunas Casen 2013”

Las bases ya publicadas o históricas se encuentran en la sección “Bases históricas” para cada año.