

MANUAL DE USO DE BASE DE DATOS

Santiago, 11 de marzo de 2025

1. ELEMENTOS GENERALES DEL ESTUDIO.

- a. **Nombre del estudio:** Barómetro regional Aysén 2024
- b. **Objetivo:**
 - Obtener información relevante para la reflexión, análisis y diseño de políticas regionales, desde el quehacer investigativo de la Universidad de Aysén, en el marco de la realización del proyecto URY 2395.
- c. **Tipo de Encuesta:** CAPI (Entrevista presencial asistida por software) y CATI (entrevista telefónica asistida por software)
- d. **Universo:** Mayores de 18 años que vivan al menos hace 12 meses en la región.
- e. **Muestra:**
 - Muestra semi-proporcional estratificada para conglomerados provinciales en zonas urbanas y rurales, considerando datos del CENSO 2017.
- f. **Unidad de análisis:** Habitantes de la Región Aysén.
- g. **Trabajo de campo:** Entre el 3 de enero de 2025 y el 21 de enero de 2025.
- h. **Diseño muestral y levantamiento:** EKHOS.
- i. **Encuestas completas:** 465 encuestas válidas, 235 en la Provincia de Coyhaique y 149 en la provincia de Aysén y 81 en el resto de la región.
- j. **Error muestral:**
 - ➡ +/- 4,6% para la muestra global.
 - ➡ +/- 6,4% para el conglomerado de la provincia de Coyhaique.
 - ➡ +/- 8,1% para el conglomerado de la provincia de Aysén
 - ➡ +/- 11,3% para el conglomerado del resto de la región.

2. DESCRIPCIÓN GENERAL DE LA BASE DE DATOS:

- a. **Nombre de la Base de datos:** Barómetro Aysén 2025.03.11
- b. **Diseño muestral:** EKHOS
- c. **Levantamiento:** EKHOS.
- d. **Última Modificación:** 10/03/2025
- e. **Tipo de datos:** Encuesta.

f. Formato de las bases de datos: SPSS (.sav), Stata (.dat), R (.rds) Excel (.csv)

3. DIMENSIONES DEL CUESTIONARIO:

Las dimensiones del cuestionario son 11:

1. **Caracterización del encuestado** (FOLIO hasta OTRA_NAC_COD).
2. **Identidad, pertenencia regional y movilidad** (A1 hasta A3).
3. **Evaluación de la región** (B1 hasta B4_E).
4. **Disposición a la participación en el espacio público** (C1 hasta C2_2_COD).
5. **Percepción del conflicto medioambiental** (D1_A hasta D1_H).
6. **Consumo de medios** (E1_A hasta E2).
7. **Herramientas para el desarrollo social** (F1 hasta F4_H).
8. **Descentralización** (G1_A hasta G3).
9. **Actitudes políticas** (H1 hasta H2).
10. **Preguntas regionales de Aysén** (I1_A hasta I8).
11. **Caracterización de hogar y su Jefe/a** (J1 hasta J7).

4. TRATAMIENTO DE VALORES PERDIDOS¹.

En la presente base de datos se presentan distintos tipos de valores perdidos: a) No aplica, b) No Responde y c) No Sabe / No Responde o NS/NR.

a. No Aplica:

- Se utilizará el código 77 o 777.
- Se aplica en cada celda posterior a un filtro (En caso de que corresponda).

b. No sabe/No Responde:

- Se utilizará el código 98 y 998 para los “No sabe/No responde”.

c. Otros

- Se utilizará el código 44 o 444.

Los formatos de cada uno de estos valores están determinados por la cantidad de categorías de respuesta que contenga una pregunta, es decir, si una pregunta posee atributos que consideran la alternativa 98, se utilizará 998 para “No sabe, No responde”.

Siguiendo con la misma lógica, si existe una variable con categorías mayores o iguales a 77, se debe pasar al siguiente formato (777) para “No aplica”. Un ejemplo de la aplicación del “No aplica”, ocurre en la variable A3 (*Si de usted dependiera, ¿se cambiaría a vivir a otro lugar dentro de la comuna, a*

¹ Los criterios para asignar los códigos a las variables dependen de las encuestas utilizadas, por ejemplo, la Encuesta Social Europea (ESS) utiliza para los no aplica el código 6, 66 y 666. En el caso nacional, la ENE (Encuesta Nacional de Empleo) y las Encuestas Estadounidenses utilizan el valor 7, 77 o 777 para asignar el no aplica. No obstante, existe una tendencia a considerar 8, 88 o 888 a los no sabe y 9, 99 o 999 a los valores asociados a No Responde.

otra comuna de la región, a otra región o a otro país?), si el encuestado dice que no desea cambiarse o vivir en otro lugar, entonces el valor correspondiente para A3 es “No aplica”.

5. CODIFICACION DE VARIABLES

5.1. Preguntas cerradas

La codificación de las preguntas del cuestionario fue trabajada en similitud a lo expresado en distintas encuestas nacionales (EANNA y CASEN) e internacionales (ESS) donde se asocian números enteros consecutivos para identificar cada categoría de respuesta.

- Por ejemplo, la pregunta 17 posee 5 categorías para la pregunta “En general, ¿Cómo evalúa el aporte de la Universidad de Aysén al desarrollo regional en los siguientes aspectos... las cuales corresponden a “Muy bajo”, “Bajo”, “Indiferente”, “Alto” y “Muy alto”, con las categorías de “No sabe/ No responde”, en caso de ser necesarias, donde a las categorías se les asigna valores del 1 al 5 respectivamente, y de 98 para la categoría de no respuesta.

5.2. Preguntas Abiertas

b. Tratamiento de preguntas

Las preguntas de respuestas abiertas y espontáneas son registradas en forma literal y requieren un tratamiento especial, porque las respuestas dadas por las personas representan una amplia gama de respuestas.

EKHOS ocupa una estructura de codificación de tipo árbol, con niveles o ramas principales y secundarias, e incluso en algunos casos, terciarias.

En el caso de la codificación más frecuente, se establece un Libro de Códigos para la pregunta siguiendo una lógica de agrupación conceptual en dos niveles. Esto permite generar una estructura lógica, en la que el primer nivel es un concepto general y el segundo nivel es una especificación de dicho concepto.

c. Patrón de codificación

En esta base de datos, el patrón de las variables distingue:

- Codificación inicial desde el nivel más desagregado que haya sido definido, que en general, es el Nivel 2. A la variable de origen, se le agrega el sufijo “_COD”.
- Recodificación esquemática, siguiendo la agrupación previamente definida con el modelo de árbol, corresponde al Nivel 1. A la variable de origen, se le agrega el sufijo “_RECOD”.

A continuación, se presenta un ejemplo con la respuesta sobre los grupos en los que los entrevistados dicen participar:

Tabla 1: Ejemplo de codificación por categorías de las variables

Nivel 1	Nivel 2	Respuesta Literal
Voluntariado	Sociales	Voluntariado de salud
	Bomberos	Bomberos
Talleres y agrupaciones	Deportivo	Club deportivo
	Artístico	Taller de tejido
	Cultural	Agrupación cultural
Organizaciones sociales	Junta de vecinos	junta de vecinos
	Sindicato	Sindicatos
	Campesino/agua	Comité campesino
	Adulto mayor	Club adulto mayor
Organizaciones gremiales	Gremiales	Asociación gremial de pequeños agricultores

Así, en este ejemplo, en C2_2 se solicita a los encuestados que indiquen en que organizaciones sociales participa (en caso de que haya declarado participar en ellos). Al ser una respuesta abierta y espontánea, fue necesario codificar en un esquema de dos niveles:

- Creando la variable “C2_2_COD”, para codificar inicialmente en forma detallada, desde el nivel 2.
- Luego, una vez codificados todos los casos, se elabora sintaxis para realizar la reagrupación de las categorías del nivel 2 que correspondan a la misma macro categoría del nivel 1. A esta nueva variable (del ejemplo) se nombra C2_2_RECOD.

d. Códigos utilizados

Los códigos utilizados para las variables y respuestas de este apartado son siempre números enteros que se dividieron en dos tipos.

- En aquellas respuestas que la codificación requirió sólo un nivel y que las categorías de respuesta fueran menores o iguales a 10, fueron números enteros comenzando desde el 1 para terminar en el 10 sin contar la categoría “Otros”.
- Por otro lado, cuando hay mayor diversidad y detalle de respuestas, se aplicó una codificación en 2 niveles, se ocupan dos y 3 dígitos, siguiendo un esquema en que el primer dígito o la decena, indica el macro código (nivel 1) y el correlativo asociado a esa decena, el código específico de nivel 2.

Volviendo al ejemplo anterior una codificación que contempla 2 niveles es la variable P4, que se codifica inicialmente como C2_2_COD con 14 categorías, ordenadas esquemáticamente según un concepto general, que permitirá generar una macro

categoría mediante la recodificación de la variable C2_2_COD en C2_2_RECOD, con 5 categorías generales.

5.3. Codificación de Categoría Ocupacional (CIUO 08)

Este proceso consiste en la codificación de las preguntas abiertas asociadas a la categoría ocupacional (variables J4 y J5) utilizando la Clasificación Internacional Unificada de Ocupaciones en su versión 2008, aprobada por el INE.

Para este proceso se estandarizan las respuestas abiertas del cuestionario en respuestas idénticas a las que se encuentran en el cuestionario CASEN 2022 para la respuesta de ocupación. La idea de realizar esta estandarización es que facilita, de modo posterior la codificación desde respuestas de ocupación tipo CASEN a códigos CIUO 08.

En el caso de que haya más de una estandarización afín a una de las respuestas, los codificadores encargados de estandarizar las respuestas tuvieron la instrucción de elegir la opción más adecuada para cada caso, ojalá tan detallada como la respuesta original y en caso de no encontrar esa especificación, la elección debía ser la opción más general respecto a la respuesta original.

Una vez terminada esta “estandarización” de las respuestas a formato CASEN 2022, se realizan 3 procesos separados para llegar a un código CIUO 08 final:

1. En primera instancia se codifican a códigos CIUO las respuestas estandarizadas que solo tienen un posible código CIUO, un ejemplo de este tipo de respuestas es “Abogado”.
2. Las respuestas estandarizadas que tienen más de un posible código CIUO, se mandan a un codificador, que debe zanjar el código usando una interfaz de codificación que entrega información adicional como nivel educacional e ingreso.
3. Las respuestas estandarizadas sin código CIUO/CIUO asociados, se envían a un codificador experto con experiencia INE/CASEN, junto con variables auxiliares y las respuestas abiertas originales con el fin de que codifique esos casos.

Luego de finalizado el proceso se logró codificar 412 de 427 casos del CIUO lo que equivale a un 96,5%. La gran mayoría de los casos que no se lograron codificar ocurren por que la persona no entrega datos de su ocupación.

6. CREACIÓN DE VARIABLE DE GRUPO SOCIOECONÓMICO (GSE)

La creación del GSE se realizó considerando el proceso establecido por la Asociación de Investigadores de Mercado (AIM) en su versión 2023 , el cual considera 4 variables:

1. Ingreso total del Hogar (J7).
2. Cantidad de miembros que componen el Hogar (J6).
3. Ocupación CIUO del jefe de hogar (CIUO08_4N).

4. Nivel educacional del jefe de hogar (J1_1 y J1_2).

El proceso de codificación produce una variable de Grupo Socioeconómica (GSE) que se compone de 7 categorías: AB, C1a, C1b, C2, C3, D y E.

Cabe mencionar que alrededor de 9 casos no tenían valor para alguna de las variables claves usadas en la codificación, por lo que inicialmente no se les pudo asignar un GSE. Dichos casos se fueron a un arbitraje individual de caso 1 a 1 pareando las variables que si tenían valor.

En caso de que aún en este segundo proceso no se pudiera asignar un GSE, se le asigno un valor 98 que tiene el código “No se pudo codificar”.

7. VARIABLES GEOGRÁFICAS

La base de datos contiene variables de tipo demográfico que indican el área o zona donde se realizó la encuesta, estas variables siguen la cartografía censal y son las siguientes:

- DISTRITO: Indica el distrito censal donde se realizó la encuesta.
- ZONA: Indica la zona censal donde se realizó la encuesta, para los casos urbanos y la localidad censal para los casos rurales.
- MANZANA: Indica la manzana censal donde se realizó la encuesta, para los casos urbanos y la entidad en que se realizó la encuesta para los casos rurales.

Además de estos indicadores, la variable “UR_RU” indica si la encuesta se realizó en una zona urbana o rural.

8. USO DEL PONDERADOR

En primera instancia se creó ponderador trivariado con los datos censales de tramos de edad, Conglomerado y Sexo de los encuestados.

Con esta información se elaboró un ponderador para llevar los datos hacia los valores proporcionales que deberían observarse en la muestra si obedeciera la distribución porcentual observada en los datos censales:

De esta forma el factor de ponderación se refleja de la siguiente forma:

$$W_1 = \frac{P_{Npe1}}{P_{npe1}}$$

Donde P_{Npe1} representa la proporción del universo y P_{npe1} representa la proporción de la muestra obtenida, en torno al Tramo de edad, la Macrozona y el Sexo.

A continuación, se presenta la ponderación de cada uno de los segmentos a los que se busca representar.

Tabla N°2: Universo, muestra y ponderación trivariada

Conglomerado	Tramos de edad	Universo		Muestra		Ponderación	
		Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
<i>Coyhaique</i>	18 a 29 años	4,9%	5,1%	4,9%	4,3%	0,99	1,18
	30 a 44 años	9,9%	9,3%	8,8%	8,2%	1,12	1,14
	45 a 59 años	7,4%	7,4%	6,9%	7,3%	1,08	1,01
	60 años o más	7,2%	6,5%	4,9%	5,2%	1,45	1,27
<i>Aysén</i>	18 a 29 años	2,6%	2,8%	3,0%	3,0%	0,85	0,93
	30 a 44 años	4,9%	5,1%	3,7%	5,6%	1,34	0,90
	45 a 59 años	3,8%	4,1%	3,9%	7,3%	0,98	0,56
	60 años o más	3,5%	3,3%	1,7%	3,9%	2,01	0,85
<i>Resto de la región</i>	18 a 29 años	0,9%	1,0%	0,2%	0,4%	4,03	2,30
	30 a 44 años	1,9%	1,9%	3,0%	3,9%	0,63	0,49
	45 a 59 años	1,5%	1,8%	3,9%	3,2%	0,38	0,55
	60 años o más	1,8%	1,7%	1,7%	1,1%	1,03	1,61

En términos del análisis se debe tener en consideración los casos totales se alteran en las salidas descriptivas cuando se está utilizando el ponderador en la base de datos, particularmente en aquellas preguntas con filtro o con menores tasas de respuesta.

Dado que el n resultante con la base ponderada difiere del n recolectado, se sugiere generar una frecuencia de cada variable sin ponderar al momento de presentar los n de cada variable en presentaciones e informes.

9. CONTACTO

Para solicitar el libro de códigos o cualquier información adicional, se recomienda escribir un correo electrónico a estudios@ekhos.cl