RIMISP
Centro Latinoamericano para el Desarrollo Rural

MANUAL METODOLÓGICO

Encuesta de Seguridad Alimentaria y Alimentación

Chile 2022

Índice

Introducción 3
Situación COVID-19 durante fechas de la encuesta de seguridad alimentaria 3
Diseño Metodológico 3
Diseño metodológico de la encuesta 3
Ponderador 5
Clasificación urbano/rural 6
Muestreo 7
Método de muestreo 7
Limpieza de la base de datos. 8
Revisión y limpieza de la base de datos 8

Introducción

La Encuesta de Seguridad Alimentaria y Alimentación (ESAA), enmarcada en el Proyecto Siembra Desarrollo. Pequeña Agricultura y Alimentación resilientes al Covid-19, tiene como propósito recopilar en forma sistemática y de acuerdo con una metodología científica, antecedentes y conocimientos útiles o necesarios para incidir en distintas esferas de la acción pública, que permitan impulsar procesos de desarrollo en países de América Latina.

Específicamente, el proyecto busca impulsar cambios institucionales y de políticas para fortalecer la participación de la agricultura familiar en los sistemas agroalimentarios y avanzar en su sostenibilidad, inclusividad y resiliencia, así como mitigar los efectos negativos sobre la seguridad alimentaria, a partir de la generación de evidencia empírica y el establecimiento de espacios de diálogo de políticas y herramientas de asistencia técnica a gobiernos. En particular:

- Generar evidencia sobre los impactos del Covid-19 en los sistemas agroalimentarios y la agricultura de pequeña y mediana escala en 10 territorios urbano-rurales latinoamericanos.

- Apoyar, a través de asistencia técnica, la implementación de medidas de protección que mitiguen el impacto de la pandemia en la seguridad alimentaria de los grupos más vulnerables, tanto urbano como rurales.

- Promover cambios institucionales y/o de política pública mediante la generación de diálogo basado en evidencia que fortalezcan la participación de la agricultura familiar de pequeña y mediana escala en el sistema agroalimentario de una manera inclusiva, sostenible y con enfoque de género.

Situación COVID-19 durante fechas de la encuesta de seguridad alimentaria

Frente a un contexto internacional y nacional complejo donde la pandemia por COVID-19 ha sido la protagonista hace ya casi 3 años, durante este periodo en Chile se observa una disminución de las medidas de contención de la movilidad, posterior al inicio de la vacunación en diciembre del año 2021. Particularmente para la Región de La Araucanía y Los Lagos se reporta una baja incidencia de casos activos, encontrándose la mayoría de las comunas en Fase de Bajo Impacto Sanitario.

Diseño Metodológico

Diseño metodológico de la encuesta

A continuación se describen los pasos para definir la muestra objetivo.

Universo

De acuerdo a los datos de las proyecciones de población del instituto nacional de estadísticas (INE) basadas en los resultados del Censo de 2017, La población objetivo corresponde a 1.433.385 habitantes de 18 años y más en las regiones de La Araucanía y Los Lagos.

Diseño muestral

Se realizó un muestreo aleatorio estratificado, por región y comuna.

Tamaño muestral

El tamaño muestral fue de 450 casos efectivos por cada región, cuyo error muestral máximo (e.m.m.) es de 4,6% por región, considerando varianza máxima (p=0,5), un nivel de confianza del 95% y un efecto del diseño 1 - Representa el incremento/disminución de la varianza de la estimación basado en un diseño muestral diferente al muestreo aleatorio simple, el que corresponde al tipo de muestreo más básico, en el que se seleccionan directamente desde el universo los individuos que serán consultados, por lo que no existen etapas en la selección. Esto para ambas regiones.-. El error muestral máximo se puede obtener a partir de la aplicación de la siguiente formula:

$$error.muestral = Z_{1-\alpha/2} * \sqrt{\frac{p*(1-p)}{n} * (1-\frac{n}{N}) * deff}$$

donde:

- p : *corresponde a una proporción*
- N : *tamaño del universo*
- n : *tamaño de la muestra*
- Z : *Percentil de la distribución normal*
- $deff$: *efecto del diseño*

La distribución del tamaño de muestra en cada región se ha hecho de manera proporcional a la población que considera cada comuna. En los siguientes cuadros se indica la distribución esperada de la muestra por comuna y por cada región.

Tabla 1: Muestra esperada

La Araucanía	
Comuna	n esperado
Carahue	11
Cholchol	5
Cunco	8
Curarrehue	3
Freire	11
Galvarino	5
Gorbea	7

Los Lagos	
Comuna	n esperado
Calbuco	18
Cochamó	2
Fresia	6
Frutillar	9
Llanquihue	9
Los Muermos	8
Maullín	8

Lautaro	17
Loncoche	10
Melipeuco	3
Nueva Imperial	15
Padre Las Casas	43
Perquenco	3
Pitrufquén	12
Pucón	13
Saavedra	5
Temuco	135
Teodoro Schmidt	7
Toltén	5
Vilcún	13
Villarrica	25
Angol	25
Collipulli	11
Curacautín	8
Ercilla	4
Lonquimay	5
Los Sauces	3
Lumaco	3
Purén	6
Renaico	5
Traiguén	9
Victoria	15
Total	**450**

Puerto Montt	136
Puerto Varas	22
Ancud	23
Castro	26
Chonchi	8
Curaco de Vélez	2
Dalcahue	9
Puqueldón	2
Queilén	3
Quellón	17
Quemchi	5
Quinchao	4
Osorno	83
Puerto Octay	5
Purranque	11
Puyehue	6
Río Negro	7
San Juan de la Costa	4
San Pablo	6
Chaitén	2
Futaleufú	2
Hualaihué	6
Palena	1
Total	**450**

Ponderador

Dado que no se estratificó por Sexo y Edad, esto último con el objeto de mantener la naturaleza probabilística del diseño, fue necesario aplicar una post-estratificación en cada muestra para corregir las distribuciones muestrales haciéndolas similares a las poblacionales en consideración al perfil de los hogares de los participantes, respecto de las variables de edad de los informantes clave, número de integrantes del hogar, zona urbana o rural y región. Dicho factor permite que, la base de datos final represente adecuadamente al universo poblacional estudiado, y se haga la corrección de los sesgos asociados a la sub/sobre representación de ciertos grupos poblacionales, por aspectos asociados a la no respuesta.

El ponderador fue creado considerando la distribución de la población objetivo, utilizando los datos de las Encuesta “Casen en Pandemia 2020”, elaboradas por CASEN. El ponderador para esta encuesta corresponde a la variable del mismo nombre, y debe aplicarse antes del análisis de

los datos.

Clasificación urbano/rural

Se estableció la distinción entre territorio urbano, rural o mixto de acuerdo a su densidad poblacional. Así, se define la siguiente clasificación comunal:

- Las comunas predominantemente rurales se definen como aquellas en que más de un 50% de la población vive en distritos censales de densidad menor a 150 hab./km².
- Las comunas mixtas se definen como aquellas en que entre un 25% y un 49% de la población vive en distritos censales de densidad menor a 150 hab./km².
- Las comunas predominantemente urbanas se definen como aquellas en que menos de un 25% de la población vive en distritos censales de densidad menor a 150 hab./km².

A partir de esta metodología se clasifican las comunas según región. Los resultados se presentan en las tablas 1 y 2.

Tabla 2: Clasificación comunas Región de La Araucanía

Rural	Urbano - Rural	Urbana
CARAHUE	CURACAUTÍN	ANGOL
CHOLCHOL	GORBEA	TEMUCO
COLLIPULLI	LAUTARO	
CUNCO	LONCOCHE	
CURARREHUE	NUEVA IMPERIAL	
ERCILLA	PADRE LAS CASAS	
FREIRE	PITRUFQUÉN	
GALVARINO	RENAICO	
LONQUIMAY	TRAIGUÉN	
LOS SAUCES	VICTORIA	
LUMACO	VILLARRICA	
MELIPEUCO		
PERQUENCO		
PUCÓN		
PURÉN		
SAAVEDRA		
TEODORO SCHMIDT		
TOLTÉN		
VILCÚN		

Tabla 3: Clasificación comunas Región de Los Lagos

Rurales	Urbano - Rural	Urbanas
CHAITÉN	ANCUD	PUERTO MONTT
CHONCHI	CALBUCO	
COCHAMÓ	CASTRO	
CURACO DE VÉLEZ	OSORNO	
DALCAHUE	PUERTO VARAS	
FRESIA	QUELLÓN	
FRUTILLAR		
FUTALEUFÚ		
HUALAIHUÉ		
LLANQUIHUE		
LOS MUERMOS		
MAULLÍN		
PALENA		
PUERTO OCTAY		
PUQUELDÓN		
PURRANQUE		
PUYEHUE		
QUEILÉN		
QUEMCHI		
QUINCHAO		
RÍO NEGRO		
SAN JUAN DE LA COSTA		
SAN PABLO		

Muestreo

Método de muestreo

Las encuestas se llevaron a cabo entre el día Viernes 2 de Diciembre de 2022 y el Miércoles 25 de Enero de 2023, y se logró alcanzar el 100% de la muestra planificada. Se realizaron llamadas en todos los municipios, esto con el fin de cumplir con la distribución de las personas mayores a 18 años según región y comuna. Considerando las proyecciones de población del Censo 2017. Las llamadas fueron de forma aleatoria. Se solicitó que la persona que respondiera la encuesta sea mayor de 18 años y tenga conocimiento sobre la alimentación del hogar. En caso contrario, se le preguntó por otro miembro del hogar con esas características. Si no había alguien disponible en ese momento capaz de contestar la encuesta, se agendaba con el interlocutor otra llamada, a fin

de ubicar a una persona del domicilio que sí pudiera responder las preguntas de la encuesta.

En los casos en que las comunas presentaron 5 o menos casos y que no se pudo lograr el n requerido durante el período de las encuestas, se distribuyó lo restante entre las demás comunas. Vale destacar que gran parte de las comunas tuvieron más encuestas aplicadas que lo especificado por la muestra. Esto se debe a que, en algunos casos, los encuestados ya no vivían en la comuna que indicaba la base de datos de contacto, si no que se habían mudado a otra residencia. En estas situaciones la encuesta se aplicó considerando la comuna de residencia actual de la persona, pero para efectos de la cuota a cumplir en el sistema, el encuestado correspondía aún a su comuna de origen. El entrevistador reportó la situación para que una vez completada la aplicación de la encuesta, tuviera lugar la corrección pertinente.

El resultado del intento de contacto vía telefónica se tipificó con una cierta incidencia (basada en los indicadores AAPOR). Para efetos de la situación donde el encuestado se niega a participar, dicho resultado se cataloga como "Rechazo". Cuando el teléfono no es atendido, el número de teléfono continua en el sistema de llamadas, y se intenta un re-contacto en 3 oportunidades más. Si tras esto el teléfono no es respondido, recibe la categoría de "No contactado / ocupado"

Transcurrido un tiempo límite el protocolo acordado para lograr el n faltante por región, fue trasladar la cuota faltante a ciudades de mayor población. Comparativamente más accesibles a las comunas de menor población y/o rurales, dada su mayor disponibilidad personas y por consiguiente de números de teléfonos que hacen más probable obtener una encuesta efectiva.

En la tabla 4 se incluyen otros detalles relativos al muestreo. Estos aplican para ambas regiones.

Tabla 4: Estadígrafos relevantes

Margen de error	4,6%
Varianza máxima	p=0,5
Nivel de confianza	95%
Efecto estimado del diseño	1

Limpieza de la base de datos.

Revisión y limpieza de la base de datos

El levantamiento de los datos se realizó vía telefónica entre diciembre del año 2022 y enero del año 2023. La Base de datos consta de 900 observaciones y 82 variables. La revisión y limpieza de la Base de Datos se realizó utilizando el lenguaje de programación Python y la interfaz Google Colab. La metodología utilizada se detalla a continuación:

1) Primera inspección a la BBDD: Se observa de forma general el Dataframe, identificando posibles aspectos anómalos.

2) Revisión de los datos faltantes: A partir de la revisión general, se realiza una revisión de los datos faltantes, identificando
3) Revisión de columnas: Se identifica alguna anomalía en las columnas del Dataframe, es decir, columnas con información redundante, con una sola categoría o un solo valor numérico, o bien si se repiten columnas.
4) Identificación de outliers en variables numéricas: Se identifica la presencia de información anómala en variables numéricas, esto a partir de un análisis visual utilizando gráficos de caja.
5) Identificación de errores en variables categóricas: Se identifica la presencia de categorías erróneas en variables categóricas, esto a partir de un análisis visual utilizando gráficos de barra.

La aplicación de la metodología se presenta a continuación:

1) Primera inspección a la Base de datos

Se indica que el Dataframe tiene 900 observaciones y 82 variables, de las cuales 66 son categóricas, 13 numéricas y 3 objetos (ID del estudio, Pueblos originarios y Otra preocupación por la falta de alimentos).

2) Revisión de los datos faltantes:

Mediante un conteo de los datos faltantes para cada una de las columnas del dataframe, se identifica que varias columnas tienen datos faltantes. A partir de la identificación de los datos faltantes, se analiza cada una de las variables para identificar a que se debe la falta de datos, e identificar si deben ser imputados o si la observación debe eliminarse. El detalle se presenta a continuación:

a. Variables sobre pueblos originarios

Se observa que las variables PUEBLOS y PUEBLOS O poseen menos observaciones que el total. Con lo anterior podemos observar que las 177 observaciones de la variable 'PUEBLOS' que nos son datos faltantes, corresponden a las observaciones cuyo valor de la variable 'A3' (¿Se considera Ud. perteneciente a algún pueblo indígena o afrodescendiente?) corresponde a 2 (Sí), es decir, las personas que señalan que pertenecen a alguno de los pueblos originarios.

b. Variables sobre edad de las personas del hogar

A partir del análisis de las variables 'A5', 'A6B' y 'A6' se puede observar que poseen 729 observaciones. Al revisar la pregunta A4 se puede identificar que las observaciones no registradas de A5 A6B y A6 corresponden a las 171 observaciones donde hay solo hay una persona en el hogar, es decir la persona que responde.

c. Variable sexo del jefe de Hogar

Se observa que la variable 'A8' posee datos faltantes. Si observamos la variable 'A7' (¿Es usted el jefe de hogar?), podemos identificar que funciona cómo filtro. Por ende, los datos faltantes en la variable 'A8', corresponden a las personas que indican que si son los jefes de hogar.

d. Variables preocupación sobre acceso a alimentos

Se observa que las variables 'B9_1', 'B9_2', 'B9_3', 'B9_4','B9_5' y 'B9_6' poseen 337 observaciones. Al revisar todas las observaciones que al menos respondieron positivamente una de las preguntas del módulo FIES, se puede ver que corresponden a las 337, es decir que funciona bien el filtro.

e. Variable personas que reciben ingresos

Se observa que las variables 'D2', 'D3', 'D4' y 'D5' poseen 534 observaciones. Se observa que los valores de la pregunta D1_1 corresponde a Si cuando la totalidad del ingreso es por jubilación y No cuando existen otros ingresos. Es decir, esta pregunta funciona como filtro, los datos NaN de las preguntas siguientes corresponden a los hogares cuyos ingresos su totalidad son por jubilación.

f. Variables sobre agricultores

Se observa que las variables 'D7_1', 'D7_2', 'D7_3', 'D7_4', 'D7_5', 'D7_6' y 'D7_7' poseen 118 observaciones. La pregunta D6 funciona cómo filtro para los hogares que no se dedican a la agricultura. Por ende, las observaciones validas corresponden a los hogares que se dedican a la agricultura ('D6' >= 1).

g. Variable ingresos

Se observa que la variable 'D10' posee 892 observaciones. Al analizar los datos de esta pregunta, podemos identificar que existe un 1% de los datos que corresponden a NaN. Esto podría ser explicado por el tipo de pregunta (es sensible preguntar por los ingresos). Al considerar esto, se opta por no hacer cambios al dataframe.

En conclusión, se indica que los datos faltantes, están atribuidos a la aplicación de filtros de otras preguntas. Todos ellos se aplican de forma que no se observan errores ni perdida de datos.

3) Revisión de columnas:

Con el objetivo de identificar alguna anomalía en las columnas del Dataframe, es decir, columnas con información redundante, con una sola categoría o un solo valor numérico, o bien si se repiten columnas, se realiza una revisión del número de categorías de las columnas, revisión de las columnas numéricas y la existencia de columnas repetidas. Se identifica que no existen columnas redundantes, que no aporten información, ni repetidas.

4) Identificación de outliers en variables numéricas:

Mediante una revisión visual de gráficos de cajas se realiza la identificación de outliers. A continuación, se presentan los gráficos para las variables numéricas. Con esto, se determina que no hay valores atipicos en ninguna variable númerica. Considerando esto, no se elimina ningúna observación.

5) Identificación de errores en variables categóricas:

Mediante una revisión de las categorías de las columnas categóricas, se puede identificar que no existen categorías mal tipiadas o con errores.

En conclusión, no se realizó cambios en el Dataframe entregado por la empresa encuestadora, ya que no se pudieron identificar errores en la BBDD.

www.rimisp.org

Chile · Ecuador · Colombia · Centroamérica · México

RIMISP
Centro Latinoamericano
para el Desarrollo Rural