Instituto Nacional de Estadísticas • Chile

METODOLOGÍA

IXEPF

ENCUESTA DE PRESUPUESTOS FAMILIARES

IX ENCUESTA DE PRESUPUESTOS FAMILIARES
Período de información: octubre 2021 – septiembre 2022

Publicación: Octubre 2023
Actualizado a Febrero 2026

Instituto Nacional de Estadísticas
Subdepartamento de Recolección de Estadísticas Sociales
Subdepartamento de Supervisión y Verificación de Estadísticas Sociales
Subdepartamento Diseño de Marcos y Muestras
Subdepartamento de Estadísticas Socioeconómicas

Profesional Responsable
Jefe de Proyecto IX EPF
Francisco Bilbao Quiroga

Edición, Diseño y Diagramación
Aline Rocha Gómez

Diseño Muestral
Iván Touron Romero
Ximena Espinosa Pavez
Marta Cisternas Astudillo

Encargado Nacional de Recolección
Miguel León Muñoz

Encargada de Recolección de Área Metropolitana
Paulina Garrido Mora

Encargada de Recolección de Área Regiones
Muriel Contreras Arriagada

Encargado Nacional de Supervisión
Sebastián Alvarado Hernández

Encargados de Digitación
Claudia Bilbao Muñoz
Rodrigo Pezoa Muñoz

Encargados de Codificación
Ignacio Ríos Medina
Macarena Treupil Marín

Coordinadores Equipo Técnico
Alfredo Heufemann Peña
Camila Mena Carvacho

Analista Socioeconómico Senior
Felipe Pizarro Recabarren

Analistas del Equipo Técnico
Juan Gabriel Carrasco Bastías
Patricio Castillo Díaz
Juan Eduardo Concha Olivares
Nicole Dagnino Fernández
Jonathan González Mejías
Juan Carlos Jimenez Navarro
Paula Lara Martínez
Felipe López Sabalaga
Antonieta Maulén Stark
Catherina Paz Lobos
Javiera Quinteros Hernández
Antonia Riveros Pimentel
Claudia Tejos Candia
Carlos Torres Reyes
Sebastián Valladares Marchant

Dirección: Morandé 801, Santiago de Chile
Teléfono: (56) 23246 3830
Sitio web: www.ine.gob.cl
Correo electrónico: ine@ine.gob.cl
Facebook: @ChileINE
Twitter: @INE_Chile
Instagram: @chile.ine

ÍNDICE.

CONTENIDO

01. Presentación 9

02. Introducción 12

03. Contexto Metodológico

3.1 Antecedentes del producto Estadístico 17

3.2 Objetivos del producto Estadístico 19

04. Marco de Referencia

4.1 Marco Conceptual 22

4.2 Referencias nacionales e internacionales 42

05. Diseño Estadístico

5.1 Población objetivo 50

5.2 Marco estadístico 50

5.3 Diseño muestral 56

5.4 Desarrollo de los factores de expansión 71

5.5 Estimadores y varianza de los estimadores 89

5.6 Clasificadores utilizados 107

06. Metodología de recolección de datos

6.1	Diseño de instrumentos de captura de la información	121
6.2	Prueba y evaluación de los instrumentos de captura de información	136
6.3	Características metodológicas de la recolección oficial de datos	144
6.4	Características organizacionales de equipos participantes de la recolección de datos	188
6.5	Método de capacitación para el personal del proceso de recolección de datos	197
6.6	Sistema Informático	202

07. Metodología de procesamiento de datos

7.1	Métodos de procesamiento de datos	208
7.2	Estructura de trabajo	209
7.3	Flujo de procesamiento	211

08. Metodología de análisis de datos

8.1	Análisis de resultados	247

09. Alcance de la presentación de resultados

9.1	Productos estadísticos a publicar	255
9.2	Difusión de los resultados	257
9.3	Accesibilidad a los datos y marco legal	258

10. Conceptos claves

	Conceptos claves	261

11. Bibliografía 274

12. Anexos 282

Índice tablas

Tabla 1	Comparación entre EPF realizadas por el INE	**45**
Tabla 2	Comunas incluidas en la IX EPF según región	**57**
Tabla 3	Conformación de macrozonas según región	**58**
Tabla 4	Estimación y errores propuestos	**66**
Tabla 5	Distribución Unidades Secundarias de Muestreo (Viviendas) EPF	**67**
Tabla 6	Distribución Unidades Primarias de Muestreo (UPM) EPF	**68**
Tabla 7	Distribución Unidades primarias (UPM) y secundarias de muestreo (Viviendas) EPF	**70**
Tabla 8	Distribución de viviendas seleccionadas según estado de elegibilidad	**74**
Tabla 9	Porcentaje de viviendas clasificadas según elegibilidad y macrozona	**74**
Tabla 10	Clasificación según el logro de la entrevista	**78**
Tabla 11	Indicador de desempeño operativo a nivel de viviendas IX EPF	**78**
Tabla 12	Covariables incluidas en el método *Random Forest*	**80**
Tabla 13	Covariables seleccionadas para el modelo *Propensity Score*	**82**
Tabla 14	Ponderador de no respuesta en cada veintil	**84**
Tabla 15	Estadísticos asociados a cada etapa del factor de expansión	**87**
Tabla 16	Total *Varunit* y Conglomerados según estrato	**103**
Tabla 17	Frecuencia del número de conglomerados que componen los pseudo-conglomerados de varianzas (*Varunit*)	**107**
Tabla 18	Clasificador CIUO 08.CL (2 dígitos).	**108**
Tabla 19	Clasificador CCIF Divisiones	**111**
Tabla 20	Clasificador CCIF 2018.CL Divisiones Residuales	**115**
Tabla 21	Clasificador de Establecimientos	**117**
Tabla 22	Gasto total y gasto *per cápita* del hogar	**138**

Tabla 23	Distribución de la muestra por grupo y región	**141**
Tabla 24	Ingreso disponible promedio (hogar y *per cápita*)	**142**
Tabla 25	Gasto total promedio (hogar y *per cápita*)	**142**
Tabla 26	Comparación de cuestionarios según tecnología de la VIII y IX EPF	**143**
Tabla 27	Servicios del sistema *Survey Solutions* y su uso en la IX EPF	**146**
Tabla 28	Esquema de entrevistas del proceso de recolección	**154**
Tabla 29	Períodos y motivos de planes especiales	**160**
Tabla 30	Tipos de medidas adoptadas en planes especiales	**161**
Tabla 31	Clasificación de códigos de disposición de la unidad según resultado en Hoja de Ruta y Clasificación final según criterios de la grilla técnica	**165**
Tabla 32	Clasificación de viviendas según tipo de código de disposición final	**166**
Tabla 33	Criterios de asignación de unidades a supervisión	**171**
Tabla 34	Indicadores de supervisión	**183**
Tabla 35	Distribución equipo de recolección por región	**192**
Tabla 36	Distribución equipo de supervisión por región	**196**
Tabla 37	Estadística de la no respuesta parcial en Gastos Diarios a nivel de personas que debían contestar el cuestionario	**225**
Tabla 38	Estadística de la no respuesta parcial en Gastos Diarios a nivel de días, considerando personas que debían contestar el cuestionario	**226**
Tabla 39	Comparación de estadística descriptiva entre datos imputados y sin imputar, a nivel muestral	**229**
Tabla 40	Completitud de cuestionarios según clasificación de administrador de gastos y sustentador del hogar	**229**
Tabla 41	Estructura por división del clasificador de CCIF 2018.CL del gasto de consumo final de los hogares pre y post imputación de GD, a nivel expandido (excluye arriendo imputado)	**230**

Tabla 42 Estadística de la no respuesta parcial de ingresos, a nivel de personas **231**

Tabla 43 Resultados muestrales del ingreso bruto promedio por persona según fuentes de ingreso imputadas por método *Hot Deck* (HD) aleatorio **234**

Tabla 44 Estadística de la no respuesta parcial del gasto financiero desagregado por instrumento **237**

Tabla 45 Resultados de la imputación de gastos en servicios financieros **238**

Tabla 46 Resumen del Factor de Ajuste Temporal según instrumento **243**

Índice de gráficos

Gráfico 1 Importancia de las variables para ser incluidas en el modelo *Propensity Score* **81**

Gráfico 2 Ponderador ajustado por no respuesta al interior de cada estrato de muestreo sin suavizamiento **85**

Gráfico 3 Principales indicadores de desempeño operativo (IDO) según los códigos de disposición final **167**

Índice de imágenes

Imagen 1 Línea de tiempo de versiones de la EPF **44**

Imagen 2 Factores de expansión *Wi* ordenados de mayor a menor **86**

Índice de diagramas

Diagrama 1	Esquema de Gasto de Consumo (utilización de los ingresos)	**27**
Diagrama 2	Gasto de los hogares según tipo de consumo	**30**
Diagrama 3	Esquema del ingreso disponible de los hogares	**37**
Diagrama 4	Construcción de Estrato N°1: Región de Tarapacá	**55**
Diagrama 5	Niveles del clasificador CCIF 2018.CL	**112**
Diagrama 6	Instrumentos de recolección de la IX EPF	**122**
Diagrama 7	Fases de recolección en cada submuestra operativa	**148**
Diagrama 8	Etapas del proceso de supervisión	**169**
Diagrama 9	Organigrama equipo de recolección	**189**
Diagrama 10	Organigrama equipo de supervisión	**195**
Diagrama 11	Nomenclatura de tabla de datos	**210**
Diagrama 12	Integración de Hogares	**213**
Diagrama 13	Períodos de referencia Cuestionario Gastos del Recuerdo	**241**

01.

PRESENTACIÓN

01.

PRESENTACIÓN

El Instituto Nacional de Estadísticas (INE) presenta a la ciudadanía la Metodología de la IX Encuesta de Presupuestos Familiares (EPF), cuya recolección se realizó entre octubre de 2021 y septiembre de 2022. Este proceso de producción estadística es la continuación del trabajo que el INE realiza en materia de gastos e ingresos de los hogares desde el año 1956, siendo la EPF uno de los productos que cuenta con mayor trayectoria en la institución.

La primera EPF se realizó en los años 1956–1957, para luego aplicarse con una periodicidad de diez años. Desde la VII EPF, llevada a cabo entre noviembre de 2011 y octubre de 2012, la encuesta se realiza con una periodicidad quinquenal, conforme al compromiso país adquirido con la Organización para la Cooperación y el Desarrollo Económicos (OCDE). La VIII EPF, realizada entre julio de 2016 y junio de 2017, fue la última versión publicada.

La IX EPF mantiene vigentes las recomendaciones conceptuales y metodológicas de diversos organismos internacionales y oficinas estadísticas. En particular, se utilizan las normas internacionales indicadas por la Oficina Estadística de la Unión Europea (Eurostat) en su manual *"Household Budget Surveys in the EU"*, mientras que las definiciones utilizadas se encuentran armonizadas con el Sistema de Cuentas Nacionales, el principal marco conceptual de referencia de la EPF.

Entre los usos de la EPF se encuentran ser el principal insumo para la actualización de los bienes y servicios que componen la canasta del Índice de Precios al Consumidor (IPC) y sus respectivas ponderaciones; su utilización para la conformación de la canasta básica de alimentos (CBA) y el valor umbral de la línea de pobreza por ingresos, la estimación del gasto de bolsillo en salud de los hogares, además del amplio uso por parte de organismos públicos y privados, en la realización de investigaciones socioeconómicas y publicaciones de diversos ámbitos, tales como estudios sobre calidad de vida, nutrición, entre otros.

Para dar cumplimiento al compromiso permanente del INE en cuanto a transparentar su trabajo ante la opinión pública, en las siguientes páginas se ponen a disposición las decisiones metodológicas tomadas en las distintas etapas de la encuesta, desde la organización del proyecto, el trabajo en terreno, el procesamiento de los datos y su análisis final.

En este sentido, la IX EPF incluye mejoras metodológicas entre las que destacan la incorporación de las macrozonas norte, centro y sur como nuevas áreas de estimación, ajustes a la metodología de calibración de los factores de expansión de la encuesta, incorporación de preguntas de caracterización sobre identidad de género, nacionalidad y pueblos originarios, revisiones a los métodos de imputación utilizados, entre otras. Adicionalmente, se destaca la incorporación de dispositivos móviles de captura para la recolección de los datos en terreno.

Finalmente, en nombre del Instituto Nacional de Estadísticas, deseo expresar nuestro agradecimiento a todos quienes colaboraron a lo largo de los cuatro años que requirió la elaboración de la IX EPF, especialmente a los ciudadanos que, con alto sentido cívico, dispusieron de su tiempo para brindar la información solicitada. Asimismo, a las más de 300 personas que trabajaron directamente en este proyecto, a los equipos del INE que brindaron soporte a las distintas áreas del proceso productivo y a quienes integran las Direcciones Regionales del INE. Su trabajo, esfuerzo y colaboración resultaron fundamentales para el logro de los objetivos propuestos.

Ricardo Vicuña Poblete
Director Nacional
Instituto Nacional de Estadísticas

02.

INTRODUCCIÓN

02.

INTRODUCCIÓN

El Instituto Nacional de Estadísticas tiene como misión ser la institución pública de Chile encargada de generar las estadísticas oficiales del país y articular el Sistema Estadístico Nacional, teniendo como propósito disponer de información de calidad para la toma de decisiones a nivel nacional y territorial.

La Encuesta de Presupuestos Familiares (EPF) es un estudio socioeconómico aplicado a hogares, cuyo objetivo es producir información sobre los gastos e ingresos en que estos incurren en un período de tiempo determinado. En 2010, con el ingreso de Chile a la Organización para la Cooperación y el Desarrollo Económicos (OCDE), se estableció el compromiso país de actualizar el Índice de Precios al Consumidor (IPC) con una periodicidad máxima de cinco años. Por ello, la EPF, como principal fuente de información para el IPC, se comenzó a realizar con la misma periodicidad quinquenal.

Esta encuesta requiere de un alto nivel de participación de los hogares que son encuestados, razón por la cual es uno de los procesos de producción estadística más complejos que existen entre las estadísticas de hogares del país. La IX EPF, en especial, involucró un desafío aún mayor, pues se desarrolló durante la emergencia sanitaria producida por la pandemia del COVID-19, lo que implicó implementar protocolos para minimizar el riesgo de contagios, resguardando la salud de quienes recolectaron datos en terreno como de las personas informantes.

Desde su séptima versión, la encuesta ha incorporado recomendaciones conceptuales y metodológicas de distintos organismos internacionales y buenas prácticas de diversas oficinas estadísticas. En su actual versión, la EPF introduce mejoras al proceso de producción, relacionadas con ajustes a los cuestionarios, la introducción de dispositivos móviles en la recolección de datos, nuevos ítems para mejorar la captura de gastos, la optimización del proceso de codificación de datos, además

[1] Acrónimo utilizado para hacer referencia a las siglas en inglés *Generic Statistical Business Process Model.*

[2] Para la revisión de documentos es posible visitar https://www.ine.gob.cl/estadisticas/sociales/ingresos-y-gastos/encuesta-de-presupuestos-familiares

de la estandarización de la organización de su proceso productivo, adaptándolo al estándar internacional Modelo Genérico del Proceso de Producción Estadística (GSBPM[1], por su sigla en inglés). Adicionalmente, desde marzo de 2021, un Comité de Expertos acompañó el proceso de producción de la IX EPF. En distintas sesiones, retroalimentaron los resultados parciales y finales del estudio, con importantes aportes desde sus respectivas áreas. Dicho comité de expertos estuvo conformado por Juan Bravo, Carolina Casas-Cordero, Xavier Mancero y Jorge Selaive quienes a través de sus comentarios y aportes permitieron fortalecer el proceso de producción de la encuesta.

La Metodología de la IX Encuesta de Presupuestos Familiares (EPF) describe el proceso de producción estadística utilizado para recopilar datos representativos de los gastos e ingresos de los hogares, así como las decisiones metodológicas tomadas en los distintos procesos involucrados en su producción.

Adicionalmente, esta metodología forma parte de una serie de documentos de referencia[2] y consulta que el INE pone a disposición de la ciudadanía para quienes deseen conocer instrucciones, procedimientos y decisiones metodológicas del estudio, dando cumplimiento a las orientaciones generales para la documentación de metodologías de la institución.

La Metodología de la IX EPF está enfocada para quienes deseen conocer los métodos y decisiones asociadas a la producción de la última versión de la encuesta. Esta pone a disposición el siguiente contenido:

- Se presentan los antecedentes y objetivos del estudio.

- Se revisan experiencias internacionales respecto a este tipo de encuesta, para luego realizar una presentación de los antecedentes nacionales que alimentan la presente versión y exponer además los objetivos del estudio.

- Se plantea el marco teórico conceptual en el cual se basa la encuesta, abordando los principales conceptos empleados por el estudio, los criterios técnicos utilizados para capturar información del gasto e ingreso de los hogares, la prueba piloto de la encuesta y los instrumentos utilizados para la recolección de datos.

- Se presenta el diseño estadístico de la encuesta, revisando las características del diseño muestral, el desarrollo de los factores de expansión con los correspondientes ajustes en sus ponderaciones, los estimadores y medidas de precisión, además de los clasificadores utilizados por la encuesta.

- Se expone la estructura organizacional de la encuesta y el flujo general asociado al proceso de producción estadística de la IX EPF, basándose en el estándar internacional del Modelo Genérico del Proceso de Producción Estadística (GSBPM).

- Se explica la metodología de recolección de datos, haciendo una revisión del modo de organización, capacitación y metodología de trabajo de los equipos involucrados en esta fase.

- Se aborda la metodología de procesamiento de datos, revisando sus distintos subprocesos y etapas.

- Se presenta la metodología de análisis seguida por la encuesta para la presentación de resultados.

- En el penúltimo capítulo, se reporta el alcance de resultados del estudio, es decir, los productos estadísticos a publicar, su difusión y la accesibilidad a los datos como su marco legal.

- Finalmente, se agrega un glosario asociado a las palabras y expresiones relevantes del documento, además de la bibliografía y anexos del presente documento.

03.

CONTEXTO METODOLÓGICO

03.

CONTEXTO METODOLÓGICO

3.1 Antecedentes del producto estadístico

Las Encuestas de Presupuestos Familiares (EPF), también llamadas encuestas de ingresos y gastos[3], son estudios que permiten identificar principalmente las pautas de consumo, así como el nivel y estructura del gasto de los hogares de una población, complementada con información sobre sus ingresos. La riqueza de información que captura este tipo de encuestas ha permitido diversificar su uso a distintas temáticas. Además de su utilidad histórica en la actualización de la canasta de bienes y servicios que compone el Índice de Precios al Consumidor (IPC), sirve de insumo para la conformación de la canasta básica de alimentos[4] utilizada en la construcción de las líneas de pobreza extrema y pobreza por ingresos, y en otras áreas de estudio como el nivel de vida, tendencias de consumo, patrones alimenticios, gastos de bolsillo en salud, entre otros. Del mismo modo, permite obtener información para la construcción de la cuenta institucional del sector Hogares de las cuentas nacionales elaboradas por el Banco Central de Chile (BCCh, 2017; BCCh, 2022).

A razón de su relevancia y conforme a la misión institucional del INE, la IX EPF requirió la implementación de un proceso de producción estadística que mantiene los estándares de calidad estadística requeridos, siguiendo para ello recomendaciones y lineamientos de organismos internacionales como la Oficina Estadística de la Unión Europea (Eurostat, 2003), la Organización Internacional del Trabajo (OIT, 2003a; OIT, 2003b; OIT, 2003c), de Naciones Unidas (ONU et al., 2016); la Comisión Económica de las Naciones Unidas para Europa (UNECE,

[3] En América Latina y el Caribe es posible identificarlas genéricamente dentro de esta definición. Para más detalles ver "Encuestas de ingresos y gastos de los hogares: experiencias recientes en América Latina y el Caribe - 2021", Documentos de Proyecto, CEPAL e INE (CEPAL/INE, 2021).

[4] Su utilización en la actualización de la canasta básica de alimentos es particularmente relevante, debido a que el indicador de pobreza monetaria basado en la canasta básica de alimentos se ha definido como un indicador a nivel de América Latina y el Caribe para el seguimiento de los Objetivos de Desarrollo Sostenible (ODS), a modo de complemento del indicador global utilizado para este fin (Cepal, 2009).

2016), la Comisión Económica para América Latina y el Caribe (CEPAL/INE, 2021), así como los estándares generados por el INE en la última década.

Conceptualmente, la IX EPF se fundamenta en los lineamientos del Sistema de Cuentas Nacionales (SCN), ocupando un lugar principal el concepto de consumo, y en particular el sector institucional de los Hogares, del cual es representativo la encuesta. Al respecto, la Organización Internacional de Trabajo sostiene:

> *"Los hogares tienen necesidades y deseos personales que se satisfacen directamente mediante el consumo de bienes y servicios resultantes de actividades que son productivas en sentido económico. Estos bienes y servicios se conocen como bienes y servicios de consumo y su valor individual se define como gastos de consumo en ese bien o servicio"* (OIT, 2003c, pág. 30).

De esta forma, la satisfacción de las necesidades de los hogares se puede realizar a través de distintos canales:

- Mediante transacciones en el mercado, utilizando instrumentos monetarios o no.

- Mediante la producción de bienes y servicios al interior de los hogares.

- Mediante transferencias sociales recibidas desde el Estado o instituciones sin fines de lucro que sirven a los hogares (ISFLSH).

La suma de gastos realizados a través de transacciones, producción de los hogares y transferencias sociales recibidas, corresponde al gasto de consumo final efectivo de los hogares. Este concepto abarca los bienes y servicios que tiene a disposición el hogar para la satisfacción de sus necesidades, con independencia de si son adquiridos por los hogares en el mercado, o si el gasto es soportado por otros sectores (gobierno o ISFLSH) (Eurostat, 2003, pág. 22; ONU et al., 2016, pág. 30). Por ello, esta medida sería la adecuada si el objetivo es conocer la condición de vida o el bienestar económico de las personas.

Sin embargo, en la práctica, la medición de tal indicador encuentra dificultades, en particular respecto al cálculo y valorización de transferencias sociales recibidas por los hogares (salud, educación y otros). Por ese motivo, la EPF se orienta a la medición del **gasto de consumo final**[5] de los hogares, que involucra los gastos realizados a través de transacciones en el mercado, utilizando o no instrumentos monetarios como medio de pago (Eurostat, 2003, págs. 24-25)[6]. Esto se adecúa además al uso principal de las encuestas de este tipo, que es la actualización de la canasta de bienes y servicios del IPC, ya que considera los gastos según precios de mercado.

3.2 Objetivos del producto estadístico

Los objetivos de la IX EPF son los siguientes:

- **Objetivo principal:**

Identificar la estructura y características del gasto en consumo final de los hogares urbanos, en las capitales regionales de Chile y sus zonas conurbadas, con un período de referencia de un año.

- **Objetivo secundario:**

Identificar la estructura del ingreso total disponible de los hogares urbanos, en las capitales regionales de Chile y sus zonas conurbadas, con un período de referencia de un año.

A continuación, se describen los principales usos de la Encuesta de Presupuestos Familiares:

- Ser la principal fuente de información para la actualización de la canasta de bienes y servicios que componen el IPC y sus respectivas ponderaciones.

[5] El gasto en consumo final de los hogares es el gasto realizado por los hogares en bienes y servicios que se utilizan para satisfacer directamente las necesidades o carencias de los integrantes del hogar.

[6] Considerado las dificultades prácticas asociadas a la medición del Gasto de Consumo Final Efectivo, Eurostat recomienda la medición del Gasto de Consumo Final para los Estados miembros de la Unión Europea.

• Recopilar información de los tipos de establecimiento donde realizan sus compras los hogares urbanos.

• Ser la principal fuente de información para la actualización de las líneas de pobreza y pobreza extrema para la medición oficial de pobreza por ingresos del país.

• Servir de insumo para la estimación en la compilación de cuentas nacionales (SCN) de diferentes componentes del sector institucional de los Hogares.

• Servir de insumo para la estimación del gasto de bolsillo en salud realizada por el Ministerio de Salud.

• Disponer de información socioeconómica de los hogares urbanos.

• Servir de fuente de información para el desarrollo de estudios de calidad de vida, nutrición, estudios de patrones de consumo, entre otros.

04.

MARCO DE REFERENCIA

04.

MARCO DE REFERENCIA

En este capítulo se desarrolla el marco conceptual que sustenta la realización de la IX EPF, integrado por lineamientos planteados por (Eurostat, 2003), la Organización Internacional del Trabajo (OIT, 2003a; OIT, 2003b; OIT, 2003c) y el Sistema de Cuentas Nacionales 2008 (ONU et al., 2016).

El seguimiento de conceptos y buenas prácticas internacionales para este tipo de estudios permite, por una parte, la estandarización del proceso de producción estadística y, por otra, favorecer una mayor comparabilidad a nivel internacional y en el tiempo.

4.1 Marco conceptual

El presente capítulo describe las bases del marco conceptual de la IX EPF, entre las que se encuentran las definiciones que entregarán lineamientos para la construcción de los instrumentos de captura, clasificadores, implementación, procesamiento y análisis de los datos obtenidos.

4.1.1 Concepto de hogar y persona de referencia

El hogar corresponde a la principal unidad de análisis de la IX EPF. Los organismos internacionales definen hogar como una o más personas, con o sin lazos de parentesco, que comparten domicilio y satisfacen necesidades de alimentación y otros bienes esenciales a través de un presupuesto común (ONU, 2008; CEPAL, 2011; CEPAL/INE, 2021). En tanto, Eurostat, (2003) e investigadores como (Hoffmeyer-Zlotnik, 2009) identifican las siguientes dimensiones en el concepto de hogar[7]:

a. Corresidencia: Implica que las personas viven juntas, es decir, comparten una vivienda particular con una dirección registrada.

[7] Adicionalmente, el hogar puede clasificarse de acuerdo con el número de personas que lo integran (unipersonal o pluripersonal), y por la existencia o no de relaciones de parentesco, dependencia doméstica o normas de convivencia entre las personas.

[8] Eurostat (2003) identifica la agrupación de ingresos como una dimensión distinta a la del gasto. Para efectos de esta metodología, se agrupó en el ítem de financiamiento.

b. Financiamiento doméstico: Involucra compartir gastos e ingresos[8], lo que implica mantener un presupuesto común, mayoritariamente de alimentación y alojamiento.

c. Organización doméstica: Supone la mantención y limpieza conjunta de los espacios del hogar, y el compartir comidas.

d. Familia: asume la existencia de una relación legal o de sangre, a través del matrimonio, adopción o tutela afectiva.

Por su parte, el Sistema de Cuentas Nacionales (ONU et al., 2016) define hogar como "un grupo de personas que comparten la misma vivienda, que juntan, total o parcialmente, su ingreso y su riqueza y que consumen colectivamente ciertos tipos de bienes y servicios, sobre todo los relativos a la alimentación y el alojamiento" (ONU et al., 2016, pág. 723). Esta definición especifica que algunas decisiones relativas al consumo del hogar se deben tomar en conjunto y que cada integrante debe tener algún grado de derecho sobre los recursos colectivos.

Complementariamente, CEPAL/INE (2021) enfatiza que la definición de hogar (que incluye los unipersonales) no se limita a designar una unidad de enumeración de personas, sino que tiene el carácter de unidad de consumo colectivo de alimentación, junto a otros bienes y servicios necesarios para el convivir diario Así también, se menciona que, una vez determinado el número de hogares particulares que residen en una misma vivienda, es necesario establecer su conformación, a partir de los criterios de residencia habitual, ausencia temporal y relación económica con el hogar (CEPAL/INE, 2021, pág. 17).

En consideración de estos antecedentes conceptuales, durante 2014-2015 el INE propuso una definición estandarizada del concepto de hogar, donde lo define como "una o más personas que, unidas o no por vínculo de

parentesco, residen en una misma vivienda o parte de ella y se benefician de un mismo presupuesto, principalmente para alimentación" (INE Chile, 2015b, pág. 1). Este concepto considera los siguientes aspectos:

a. Un hogar está constituido por "una o más personas…": puede estar formado por una sola persona o por varias que se benefician de un mismo presupuesto, principalmente para alimentación.

b. "…que unidas o no por vínculos de parentesco…": para constituir un hogar no es necesario que las personas sean parientes, ya que grupos formados por amigos, conocidos, o estudiantes podrían también constituir un hogar si mantienen algún grado de vínculo económico entre sí. Esta consideración permite incorporar así cambios sociales y culturales que han transformado la estructura de los hogares en el país.

c. "…residen en una misma vivienda o parte de ella…": donde se releva la idea de residencia habitual, que distingue si las personas pertenecen a un hogar de la vivienda-dirección consultada o si pertenecen a otra.

d. "…se benefician de un mismo presupuesto…": todas las personas que componen un hogar se benefician de un presupuesto común, sin importar si aportan o no con ingresos al presupuesto (como es el caso de los bebés, niños o cualquier persona que no genere ingresos).

e. "…principalmente para alimentación": considerando que un hogar puede realizar distintos gastos con un mismo presupuesto, se ha establecido que la alimentación es el gasto más generalizado, y permite distinguir con mayor claridad entre grupos con presupuestos diferenciados.

Respecto a los criterios mencionados, se consideraron definiciones operativas para los hogares participantes en la IX EPF:

- Una persona pertenece al hogar si reside habitualmente en la vivienda, es decir, si ha habitado o dormido ahí regularmente al menos cuatro días de la semana en los últimos seis meses.

- Aquellas personas que residen habitualmente en una misma vivienda, pero mantienen presupuestos separados de alimentación, son consideradas hogares diferentes. En consecuencia, una vivienda podrá estar compuesta por más de un hogar, pero un hogar solo puede pertenecer a una vivienda.

- El criterio de presupuesto compartido excluye a pensionistas e inquilinos, así como al personal de servicio doméstico, por cuánto se entiende que, en estos casos, la provisión de alimentos y/o alojamiento es parte de un intercambio económico (ONU et al., 2016).

- En el caso de las personas que se incorporan de forma reciente a un hogar, el criterio de residencia durante los últimos seis meses se sustituye por la intención de permanecer durante un período igual o mayor en el nuevo lugar.

- Además, siguiendo recomendaciones internacionales, se consideran integrantes del hogar a aquellas personas que, al momento de realizar la entrevista, no se encuentran en la vivienda, pero que cumplen alguno de los siguientes requisitos:

    - Personas que durante el período de aplicación de la encuesta se encuentran temporalmente ausentes por motivos laborales y no tienen residencia habitual en otra vivienda particular[9]

    - Personas que por algún motivo residen en viviendas colectivas[10]

    - Estudiantes ausentes que no tienen otra residencia habitual

[9] Se podrían encontrar ausentes por motivos de vacaciones, trabajo, hospitalización u otro motivo. Ejemplos de estos casos son integrantes del hogar en vacaciones, choferes de buses o vendedores viajeros, entre otros.

[10] Las viviendas colectivas corresponden a hoteles, residenciales, hospitales, cuarteles y cárceles, entre otros.

Con el fin de caracterizar los resultados de la encuesta, se solicitó a los hogares identificar a la persona que cumple el rol de sustentadora principal, definida como aquella persona que más aporta al presupuesto común, no siendo necesariamente quien gana más dinero.

En la versión anterior de la encuesta, se utilizó tanto el concepto de persona sustentadora principal como el de jefatura de hogar, para identificar a la persona de referencia respecto a la cual se caracterizaron los hogares. En la IX EPF se optó por preguntar únicamente por la persona sustentadora principal[11] debido a los lineamientos establecidos en el Sistema de Cuentas Nacionales 2008 (ONU et al., 2016), así como las recomendaciones de Eurostat (2003).

Al respecto, el Sistema de Cuentas Nacionales (SCN) plantea que la identificación de una persona de referencia es necesaria para la agrupación de los hogares en subsectores según las características de esa persona, en razón de lo cual es relevante que sea identificada en función de su importancia económica y no necesariamente según otros criterios, como pueden ser la edad o el prestigio y que están asociadas al reconocimiento de un "jefe de hogar" (ONU et al., 2016, pág. 156). Complementariamente, Eurostat plantea que características de una persona de referencia se utilizan a menudo para clasificar los hogares de la muestra para la presentación de los resultados y para la ponderación de las estimaciones de la encuesta. Así también, plantea que identificar a esa persona es relevante para definir a un informante idóneo para la entrega de cierto tipo de datos (Eurostat, 2003, pág. 20).

Además de lo anterior, se plantea que para este tipo de encuestas es importante identificar y aplicar una definición consistente de persona de referencia que pueda ser utilizada en la clasificación y análisis de información del hogar, mencionando que reconocer como tal a la persona adulta (según criterio de 15 años o más) que contribuye más al ingreso del hogar permite establecer una conexión clara entre el ingreso del hogar y sus patrones de gasto. Por el contrario, se plantea que dejar la elección de una persona de referencia al propio hogar, sin definir criterios, deja mucho espacio a la arbitrariedad, dañando con ello la comparabilidad posterior de los datos Eurostat (2003, pág. 21), y Eurostat (2020b, pág. 31).

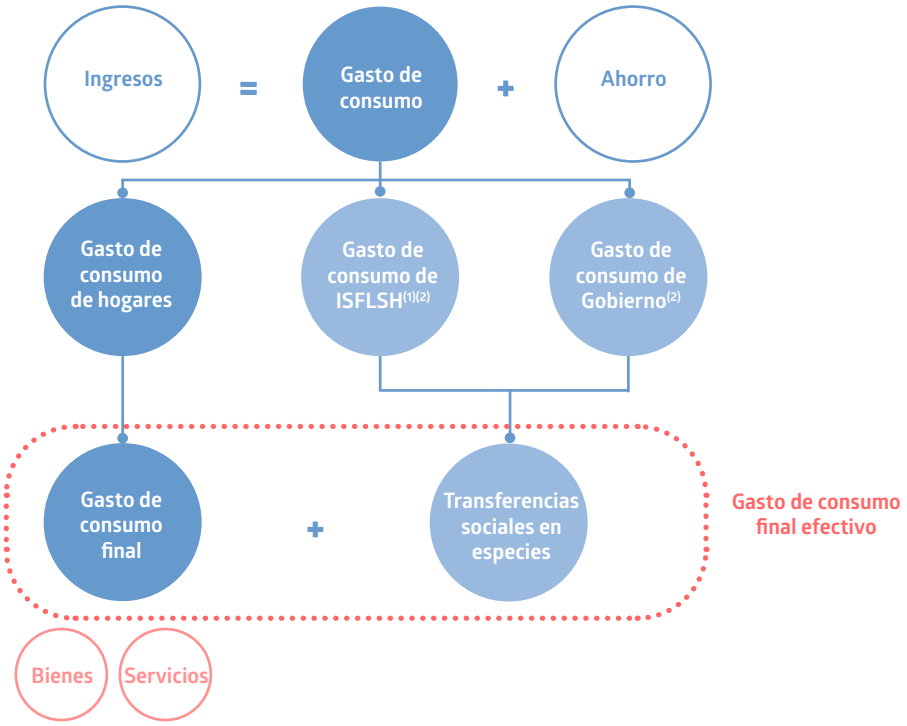
La definición de la *persona de referencia* permitió organizar el registro del resto de los integrantes del hogar, ordenando las relaciones entre sus integrantes en relación a la persona identificada según las siguientes categorías: esposo/a o conviviente, hija/o de ambos, hija/o solo de sustentador/a, hija/o solo de esposo/a o conviviente, hermano/a, padre/madre, cuñado/a, suegro/a, nuera/yerno, abuelo/a, nieto/a, otra persona pariente, otra persona no pariente, servicio doméstico y parientes de éste, pensionista.

[11] El sustentador principal corresponde al integrante del hogar que más aporta al presupuesto familiar, evaluando los últimos seis meses. Este no corresponde, necesariamente, al integrante que recibe más ingresos.

4.1.2 Criterios de captura del gasto

La IX EPF se centró en el concepto de gasto en consumo final de los hogares, que contabiliza los gastos realizados a través de transacciones en el mercado **(Diagrama 1)**. Esta definición excluye los gastos en activos fijos en forma de viviendas o de objetos valiosos[12], así como también los gastos en consumo intermedio. Estos últimos corresponden a bienes y servicios que los hogares adquieren para el desarrollo de actividades comerciales. Así, solo los bienes y servicios adquiridos por los hogares para actividades de producción de uso propio forman parte del gasto de consumo final. El grupo más importante de actividades de producción de uso propio ejecutadas por los hogares corresponde al desarrollo de tareas para los mismos, tales como limpieza de la vivienda, preparación de alimentos, cuidado de menores, entre otros.

Diagrama 1: Esquema de Gasto de Consumo (utilización de los ingresos)



(1) ISFLSH: Instituciones sin Fines de Lucro que sirven a los hogares.
(2) El esquema considera el gasto de consumo que realizan las ISFLSH y Gobierno destinado a los hogares.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

[12] La adquisición de una vivienda corresponde a formación bruta de capital fijo, mientras que los objetos valiosos conservan o aumentan su valor y no son utilizados de la misma manera que los bienes de consumo (ONU et al., 2016).

Un adecuado registro del gasto en consumo final de bienes y servicios requiere considerar dos aristas fundamentales. La primera consiste en el momento de registro del gasto y la segunda tiene relación con el tipo de bien o servicio adquirido.

Respecto a la primera arista, existen tres criterios para el registro de los gastos de consumo, todos ellos presentes de distinta forma en la medición del gasto en la IX EPF. En la 17° CIET, estos criterios se sintetizan de la siguiente forma (OIT, 2003c):

- **Criterio de adquisición del gasto:** Si se adquieren durante el período de referencia, independientemente de la fecha en que se utilizaron o pagaron en su totalidad.

- **Criterio de uso del bien o servicio:** Si se utilizaron durante el período de referencia, independientemente de la fecha en que se adquirieron o pagaron en su totalidad.

- **Criterio de pago del bien o servicio:** Si se pagaron durante el período de referencia, independientemente de la fecha en que se adquirieron o utilizaron.

Para la medición del gasto, la IX EPF utilizó principalmente el criterio de **gasto de consumo adquirido**[13], concepto que resulta de mejor comprensión para los informantes, y que además es recomendado y utilizado internacionalmente (OIT, 2003c; Eurostat, 2003; OCDE, 2013; ONU et al., 2016). Por lo tanto, el principal momento de registro del gasto en la IX EPF corresponde al momento en que el hogar adquirió el derecho de propiedad sobre un bien o el derecho a la prestación de un servicio, independientemente de si el consumo o pago de ese bien se efectuó durante el período de referencia de la encuesta.

Un ejemplo de este criterio son las compras realizadas a crédito. En ellas, se registró el valor total del bien o servicio, tal como si se hubiera comprado al contado, ya que en el momento de compra se genera un compromiso legal de adquisición. En otras palabras, el gasto en consumo

[13] OIT (2003b), Párrafo 142, página 36: "Para los IPC utilizados como índice para el seguimiento de la inflación o como índice de compensación y para las cuentas nacionales, probablemente sea más indicado el método basado en las adquisiciones".

adquirido corresponde al valor precio contado del gasto que se realiza en bienes y servicios, sin considerar si se han comprado o no en cuotas. Ello es así debido a que, por lo general, la información sobre intereses y otros tipos de cargos asociados al pago a crédito son desconocidos por el informante.

Un segundo criterio fue utilizado para la medición de aquellos gastos que los hogares realizan con frecuencia periódica (mensual, bimestral, trimestral, semestral o anual), donde se utilizó el concepto de **gasto de consumo pagado,** debido a que existen gastos cuyo pago es posterior a la entrega del servicio, tales como el consumo de electricidad y agua del hogar. Finalmente, un tercer criterio excepcional correspondió al **uso de los gastos,** utilizado para facilitar el registro del consumo de bienes y servicios asociado a la realización de viajes.

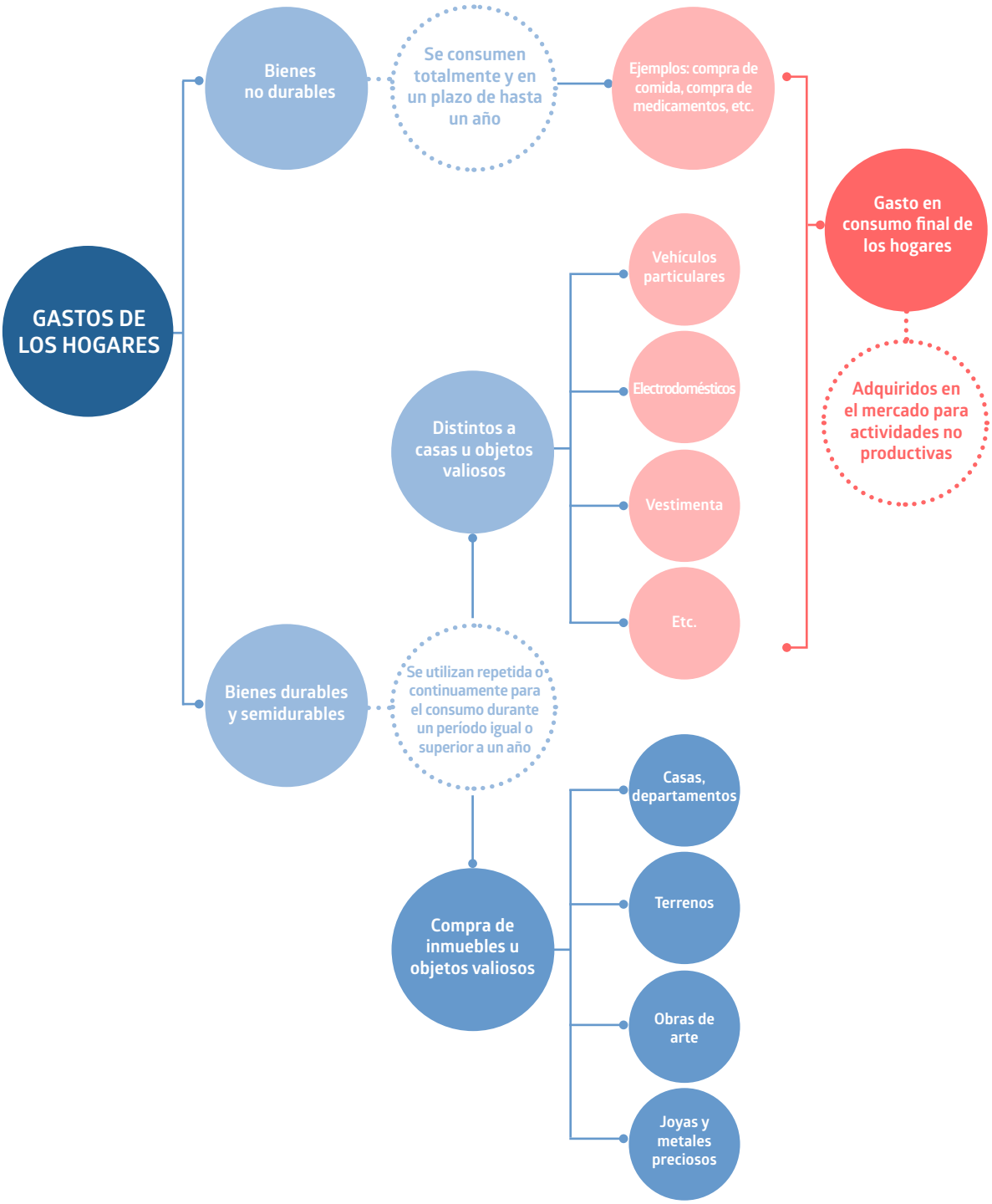
La segunda arista del gasto en consumo final se relaciona con el tipo de consumo. Este refiere a la durabilidad física del bien, es decir, si se puede utilizar solamente una vez para fines de consumo o si puede emplearse de forma repetida durante un plazo determinado (ONU et al., 2016). Según la OIT (2003c), los tipos de bienes utilizados se pueden clasificar en:

- **Bienes durables/semidurables:** Son aquellos que pueden utilizarse numerosas veces en un período superior a un año, sin reducir su capacidad de satisfacer necesidades al hogar.

- **Bienes no durables:** Son los que se consumen total e inmediatamente (y desaparecen) en el plazo de un año.

Por contrapartida, se considera convencionalmente que todos los servicios se consumen inmediatamente en el momento de la prestación de este (OIT, 2003c).

Diagrama 2: Gasto de los hogares según tipo de consumo



Fuente: Sistema de Cuentas Nacionales (ONU, et al., 2016), Eurostat (2003) y CEPAL/INE (2021).

4.1.2.1
Criterios de decisión del gasto en consumo final

Complementariamente, se definieron reglas de decisión para gastos específicos, con el fin de disminuir dudas en terreno que requirieran orientación hacia las personas encuestadoras y facilitar el logro de estimaciones insesgadas. Los criterios definidos fueron:

- Si una vivienda cambió el contrato de algún servicio pagado periódicamente (electricidad, agua, entre otros), dicho gasto fue excluido o incluido de las estimaciones dependiendo si durante el período de referencia se consumió o no dicho servicio.

- En servicios prestados en forma continua, pero con distintas opciones temporales de pago (mensual, trimestral o anual), el gasto se registró en el momento del gasto, capturando además información referida a la periodicidad con el fin de realizar el ajuste temporal correspondiente. Ejemplos de este tipo de casos son las primas de seguro, mensualidades universitarias, entre otros.

- En servicios cuya naturaleza no permite determinar su valor sino hasta concluir la prestación, el gasto se registró en el momento del cobro (por ejemplo, el servicio de luz, donde el gasto se realiza después del consumo).

4.1.2.2
Gastos considerados como parte del consumo final de los hogares

Algunas partidas de gasto consideraron adaptaciones según criterios operativos en la IX EPF, para poder ser incluidas dentro del consumo final de los hogares. Estas son:

- Contrataciones de servicios y compras de bienes durables[14] y semidurables[15], que responden al criterio de adquisiciones (gasto en consumo adquirido) y que se registraron en aquellos hogares que

[14] Se define como "Durables" aquellos bienes que pueden utilizarse para el consumo durante un período superior a un año y que tengan un valor relativamente alto, así como los que pueden utilizarse durante tres años o más.

[15] Se define como "Semidurables" aquellos bienes que pueden utilizarse para el consumo por más de un año, pero menos de tres, y que no tengan un valor relativamente alto.

declararon un compromiso de pago adquirido y el derecho de propiedad correspondiente, siempre que estuviesen dirigidos a consumo final y no a actividades productivas. Esta categoría consideró el registro de automóviles, refrigeradores y televisores, entre otros, adquiridos en el momento en que son puestos a disposición del hogar, independiente de su vida útil.

- Pagos realizados a la administración pública por servicios como certificados y pasaportes, licencias de conducir, entre otros.

- Impuestos incluidos en el precio final de los bienes y servicios.

- Arriendo pagado por los hogares.

- Gasto en servicio doméstico, donde el hogar tiene el rol de empleador del personal doméstico (asalariado).

- Materiales y mano de obra adquiridos para realizar pequeñas reparaciones en viviendas. Por el contrario, gastos destinados a grandes reparaciones no se incluyeron como gasto de consumo de los hogares, ya que se consideran inversión. Reparaciones y mantenimiento de bienes durables de los hogares también se registraron como gastos de consumo final, no así las grandes reparaciones[16].

- Inscripciones a clubes o sociedades, ya que generan un derecho a utilizar sus instalaciones.

- Los regalos comprados por el hogar para otros hogares.

- Finalmente, se incluyeron como gasto en consumo final ciertas imputaciones, tales como:

  - El valor de los bienes procedentes del auto suministro[17].

[16] Las grandes reparaciones se consideran grandes mejoras (reconstrucciones, renovaciones o ampliaciones), que constituyen formación bruta de capital fijo, por lo que no se incluyeron en el registro, que está destinado a la estimación del gasto en consumo final.

[17] En el cuestionario de Gastos Diarios (GD) es posible obtener información sobre el autosuministro de los hogares que posean algún tipo de negocio perteneciente a los integrantes del hogar.

- El valor de los bienes y servicios recibidos como retribución por el trabajo realizado por integrantes del hogar (pago en especies).

- Los arriendos imputados a los propietarios de viviendas o a los que las utilizan gratuitamente[18].

4.1.2.3

Gastos no considerados como parte del consumo final de los hogares

Partidas de gasto que no forman parte del gasto en consumo final de los hogares y que por ello fueron excluidas del registro y estimaciones de la encuesta son:

- Compras de viviendas y grandes reparaciones, que se consideran inversiones[19]. Se excluyó también la adquisición de objetos valiosos[20] y activos financieros (bonos, acciones, etc.).

- Gasto asociado a regalos recibidos desde otro hogar.

- Gastos asociados a seguros de vida, que se consideran inversión.

- Algunos impuestos y tasas pagadas a la administración pública: licencias de propiedad o uso de vehículos, embarcaciones o aeronaves; así como licencias de caza, tiro al blanco o pesca, consideradas como impuestos.

- Intereses y cargos administrativos por retraso en pagos, ya sea por compras a crédito o por pago atrasado en cuentas del hogar (moras).

- Valorización de la utilización de servicios propios de la función estatal, como la administración pública, defensa, justicia, ordenamiento jurídico, entre otros.

- Valorización de servicios domésticos y personales producidos y consumidos dentro del mismo hogar o entre hogares, en forma no remunerada. Sin embargo, los bienes y servicios adquiridos por los hogares para la realización de las actividades que desarrolla el servicio doméstico sí se incluyen dentro de su gasto de consumo final del hogar.

- Bienes de consumo intermedio de los hogares. Como se mencionó anteriormente, estos bienes se utilizan para actividades comerciales, por lo tanto, no forman parte del gasto en consumo final del hogar.

[18] Estos forman parte del sistema de producción que consideran las Cuentas Nacionales para el sector Hogares y se incluyen por constituir el principal bien de inversión de estas unidades económicas. Según señala CEPAL/INE (2021), su medición permite [...] distinguir la situación de bienestar de los hogares que ocupan su propia vivienda frente a la de aquellos que deben recurrir a arrendar una vivienda a un tercero" (pág. 26).

[19] Grandes reparaciones de viviendas en arriendo, cuando son realizadas por sus propietarios, son consideradas bien de consumo intermedio.

[20] Bienes durables de alto valor monetario que no se deterioran con el paso del tiempo se excluyen del IPC, ya que no son utilizados con fines de consumo y son adquiridos principalmente como depósito de valor. Entre estos se pueden considerar las obras de arte, piedras y metales preciosos, así como joyas elaboradas con materiales de ese tipo.

4.1.2.4
Gastos Especiales

En el siguiente apartado, se presentan los criterios utilizados en la IX EPF para el registro de gastos especiales.

a. Gastos en servicio doméstico

La remuneración al servicio doméstico puede incluir pagos en dinero y/o en especies. Estos últimos normalmente comprenden alimentación y alojamiento de la persona empleada[21].

Sin embargo, por convención, la valorización de los servicios domésticos considera solamente la remuneración que, en su condición de asalariados, se paga en dinero y/o especies distintas a la alimentación y vivienda (ONU et al., 2016), ya que la inclusión de alimentos y vivienda pagados al servicio doméstico implicaría una reclasificación de estos gastos del hogar según la finalidad.

b. Gastos en primas de seguros

Se clasifican en dos tipos: seguros de vida y otros tipos de seguros (seguros relacionados con la vivienda, la salud, el transporte y otros riesgos, como seguros frente a terceros o el seguro obligatorio de accidentes personales). En concordancia con la recomendación de organismos internacionales (Eurostat, 2013), la IX EPF excluye de la estimación del gasto del hogar aquellos montos correspondientes al pago de primas de seguros de vida, debido a que estos se consideran una inversión. También, se excluyeron las primas de seguros de ahorro y aquellos relacionados directamente a la actividad laboral.

c. Donaciones recibidas en especie

Las transferencias sociales provenientes de sectores de la economía como el gubernamental y el de instituciones sin fines de lucro que sirven a los hogares son de difícil medición, debido a que los hogares suelen

[21] Que en ocasiones puede incluir el alojamiento de su propia familia.

desconocer los montos asociados a dichas transferencias. Por lo tanto, para efectos prácticos, se considera que el gasto en consumo final es realizado por quien adquiere los bienes para donarlo[22]. Los hogares reciben estos bienes y servicios como transferencias sociales, que se consideran dentro del concepto de consumo efectivo del hogar, según lo ilustrado en el **Diagrama 1.** Respecto a los regalos entre hogares, desde un criterio de gasto en consumo adquirido, el hogar que recibe el bien debería registrar el gasto, debido a que tiene el derecho de propiedad sobre este. Sin embargo, en virtud de un criterio operativo se registra en el hogar que realiza el gasto, porque es donde se conoce en detalle el costo asociado[23] (OIT, 2003c).

[22] Desde esa lógica tanto gobierno como las instituciones sin fines de lucro que sirven a los hogares son quienes debiesen registrar los bienes donados como gasto en consumo final.

[23] Esto se realiza así, porque el hogar que realiza el gasto conoce en detalle el costo asociado a este.

d. Gastos de los hogares propietarios de empresas no constituidas como sociedad

Los gastos que los hogares realizan con fines productivos no se consideran gasto de consumo de los hogares, sino consumo intermedio (ONU et al., 2016). Aunque conceptualmente la diferencia es clara, en la práctica no siempre resulta fácil distinguir cuando un mismo bien o servicio puede utilizarse con fines empresariales y/o de consumo final en el hogar, de tal manera que conocer el bien o servicio específico no basta para la identificación de su tipo.

En ese contexto, la IX EPF enfatizó este aspecto durante la capacitación de las personas encuestadoras para que, frente a casos de difícil resolución durante la captura en terreno, se solicitara a la persona informante la desagregación entre gastos productivos y de consumo final del hogar, con el fin de registrar estos últimos en la encuesta.

4.1.3 Lineamientos conceptuales y metodológicos para capturar los ingresos

Los ingresos de un hogar representan los recursos que tienen a disposición para el consumo y ahorro en un período determinado de tiempo. Su medición es importante, porque permite acercarse a una noción del bienestar económico de un hogar, considerado uno de los indicadores más importantes y accesibles del bienestar, representado por los recursos económicos disponibles para satisfacer sus necesidades (OIT, 2003c).

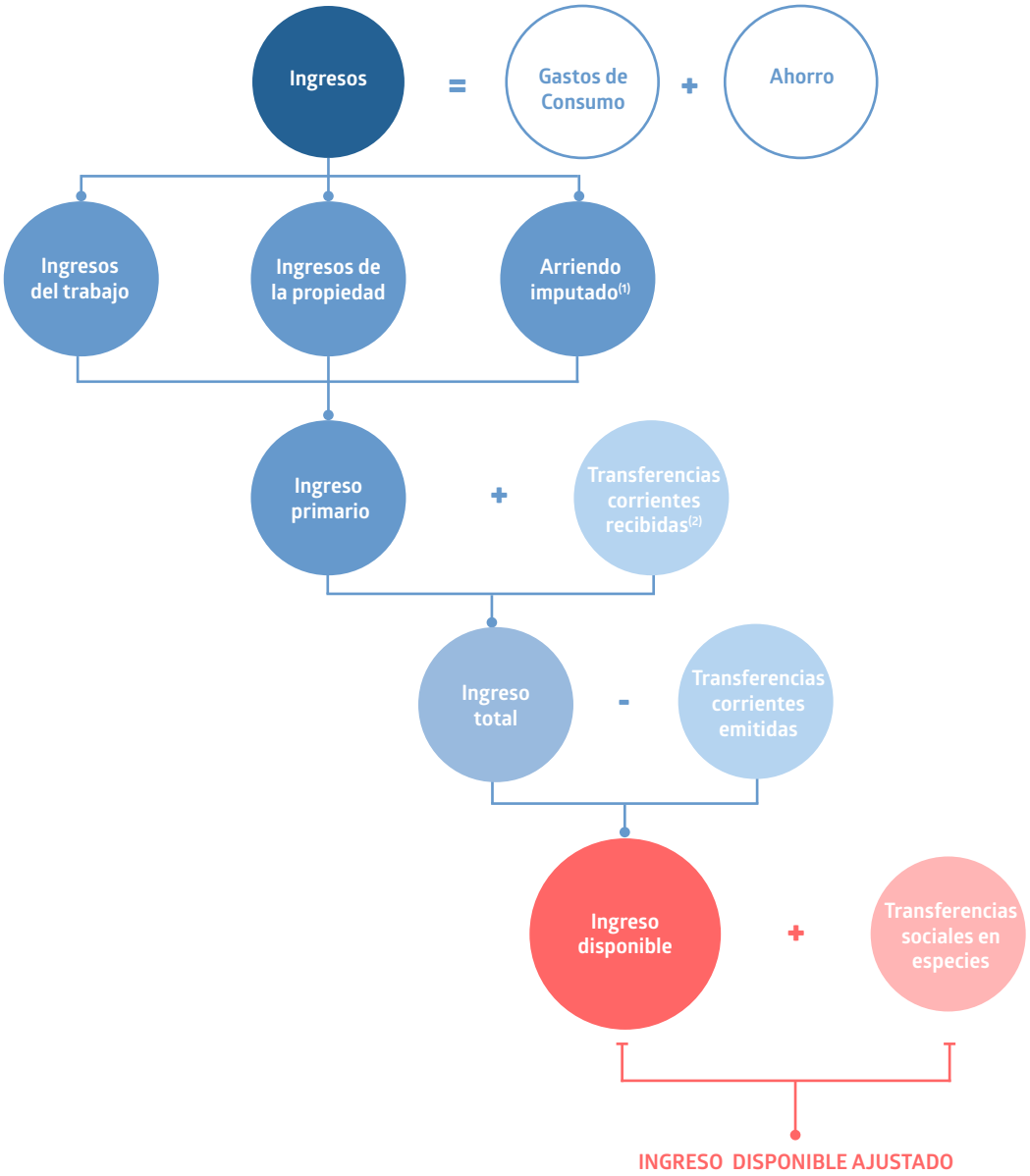
Complementariamente, la información de ingresos proporcionada por los hogares permite cumplir el objetivo secundario de la encuesta, que apunta a identificar la estructura del ingreso total disponible de los hogares. La estimación del ingreso que perciben los hogares está condicionada por las partidas consideradas en su construcción.

El ingreso que reciben los hogares tiene diferentes definiciones, que dependerá de las partidas que se consideren. Para efectos de esta encuesta se define el ingreso total del hogar como el conjunto de ingresos, monetarios o en especie (bienes y servicios) que son percibidos por el hogar o sus integrantes en un período determinado (OCDE, 2013). Es decir, tal como se muestra en el **Diagrama 3** el ingreso total de un hogar corresponde a la suma de los ingresos del trabajo[24], ingresos de la propiedad, ingresos derivados de la auto provisión de bienes producidos por el hogar[25] y las transferencias corrientes recibidas.

[24] El registro de ingresos del trabajo distingue entre trabajadores dependientes porque perciben ingreso bruto (ingreso que recibiría una persona que trabaja de manera dependiente si no se realizaran descuentos legales y otros descuentos acordados con quien lo emplea a la emplea) e independientes que perciben ingreso neto (considera los ingresos generados o ganados por la realización de una actividad independiente o negocio por cuenta propia menos los costos asociados a su realización).

[25] Es el valor neto de los bienes y servicios producidos para el consumo del propio hogar o para trueque (por ejemplo, en la agricultura de subsistencia). Además, considera el valor del arriendo imputado de los hogares y la estimación del valor de la mano de obra del servicio doméstico.

Diagrama 3: Esquema del ingreso disponible de los hogares



(1) El esquema de Cuentas Nacionales para los ingresos incorpora en esta partida los ingresos provenientes de los servicios domésticos no remunerados y del consumo de bienes durables los cuales son excluidos de la medición de esta encuesta considerándose solo el arriendo imputado.
(2) En el caso chileno, las jubilaciones poseen componentes asociados a la seguridad social, capitalización individual o una combinación de ambos. En la IX EPF no es posible diferenciar entre estos componentes, por lo que se incluyen todas en las trasferencias corrientes recibidas.

Fuente: Eurostat (2003) y OCDE (2013).

El ingreso es una variable de flujo que repercute en la riqueza o patrimonio neto de los hogares, que contribuye al gasto en consumo final. El Sistema de Cuentas Nacionales define el ingreso disponible como "el monto máximo que un hogar puede permitirse gastar en bienes o servicios de consumo durante el período, sin reducir su patrimonio neto real, es decir, financiar los gastos sin disminuir sus activos financieros o no financieros, ni aumentando su deuda"[26] (ONU et al., 2016).

En cuanto a la medición de ingresos en estadísticas nacionales, la Encuesta Suplementaria de Ingresos (ESI) es un módulo complementario de la Encuesta Nacional de Empleo (ENE) y se recolecta cada año en el trimestre octubre-diciembre. Su objetivo es caracterizar los ingresos laborales de las personas que son clasificadas como ocupadas, tanto de la ocupación principal (trabajo asalariado y trabajo independiente) como de otras ocupaciones (INE Chile, 2023c).

Al igual que la ESI y la EPF, la Encuesta de Caracterización Socioeconómica Nacional (CASEN), del Ministerio de Desarrollo Social y Familia (MDSF), también utiliza el Sistema de Cuentas Nacionales como marco conceptual. Entre el conjunto de objetivos de CASEN, adquiere relevancia la medición de las diferentes categorías de ingreso para conocer la situación socioeconómica de los hogares del país, identificar la distribución de esos ingresos y estimar pobreza (MDSF, 2023)[27].

La IX EPF, acorde a sus objetivos y siguiendo recomendaciones internacionales dirigidas a este tipo de estudios (OIT, 2003c; OCDE, 2013; ONU et al., 2016), midió el ingreso disponible del hogar, que se obtiene de la suma de los ingresos primarios y las transferencias corrientes recibidas[28] menos las contribuciones sociales y las transferencias corrientes emitidas. El ingreso disponible del hogar se compone de las partidas que se describen a continuación.

[26] El Sistema de Cuentas Nacionales tiene como referencia el período contable. En la IX EPF, el período utilizado en el registro de ingresos fue el mes anterior, ya que, en general, dicha referencia permite reflejar los pagos de sueldos y salarios, cuentas del hogar y créditos, entre otros.

[27] CASEN realiza la medición oficial de pobreza a partir del concepto de ingreso disponible del hogar, que se calcula restando al ingreso total o corriente del hogar (ingreso primario o autónomo + transferencias corrientes percibidas) los impuestos directos, las contribuciones a la seguridad social y las transferencias corrientes a otros hogares (MDSF, 2023).

[28] Una transferencia es definida como un traspaso de recursos (monetarios o en especies) que no supone una retribución y que por ende no implica un intercambio. Las transferencias corrientes corresponden a transferencias monetarias y no incluyen las transferencias sociales en especies, no obstante, si considera las transferencias en especie desde otros hogares.

1. Ingresos del trabajo:
Generados a partir de una actividad económica laboral. Los que incluyen:

a. Ingresos del trabajo dependiente:
Asociados al trabajo dependiente, abarcando sueldos y salarios, comisiones, propinas, bonificaciones o bonos en efectivo, indemnización por despido y contribuciones de seguridad social.

b. Ingresos del trabajo independiente:
Percibidos por quienes trabajan como empleadores o cuenta propia y corresponden a la diferencia entre las ganancias de la unidad económica y el costo de los insumos.

c. Otros ingresos del trabajo:
Asociados a segundas actividades económicas, como también los ingresos de personas que se declararon desocupadas o inactivas durante la semana de referencia, pero tuvieron ingresos del trabajo durante el mes anterior[29].

2. Ingresos de la propiedad:
Se definen como el flujo de ingresos netos obtenidos de la tenencia de activos financieros. Incluye el retorno de dichos activos (interés ganado de los depósitos), recibos de ingresos por inversiones, anualidades y otros pagos regulares de seguros de vida y fondos privados de pensiones (excluidos de la seguridad social), además de las rentas netas recibidas por el arriendo de activos no financieros, tales como recursos naturales, tierras, edificios, equipos, patentes o derechos de autor.

3. Ingresos por arriendo imputado:
Valor estimado de la prestación del servicio de alojamiento en las viviendas ocupadas por sus propietarios[30], o aquellos que ocupan una vivienda cedida por un familiar, el Estado o herencia.

[29] En la IX EPF, la clasificación de Otros Ingresos del Trabajo incluyó a las personas que cambiaron de trabajo durante el mes de referencia, es decir, que declararon un trabajo en el Registro de Personas en el Hogar distinto al declarado en el módulo de Ingresos. Así también, incluyó a las personas que se declararon desocupadas o inactivas durante la semana de referencia al aplicar la encuesta y que, sin embargo, recibieron ingresos asociados al trabajo durante el mes de referencia.

[30] El arriendo imputado es parte de un concepto más amplio conocido como ingresos de auto provisión de servicios producidos por el hogar, que incluye además el valor de los servicios domésticos no remunerados y el valor de los servicios aportados por el consumo de bienes durables. Sin embargo, los ingresos provenientes de los servicios domésticos no remunerados y del consumo de bienes durables son excluidos de esta encuesta y han sido tradicionalmente excluidos de la medición de encuestas y Cuentas Nacionales (Camelo, 2001).

4. Transferencias corrientes recibidas:
Flujos monetarios recibidos por los hogares, que no generan obligación de reciprocidad respecto a quienes las emiten. Incluyen los ingresos por concepto de jubilaciones[31], pensiones, otros beneficios en efectivo asociados a seguridad social o empleo, beneficios de asistencia social en efectivo del Gobierno, transferencias desde otros hogares (dentro y fuera del país) y transferencias en efectivo recibidas desde instituciones sin fines de lucro que sirven a los hogares. Este tipo de transferencias excluye las transferencias sociales en especies que los Gobiernos y las instituciones sin fines de lucro emiten a estos.

5. Transferencias corrientes emitidas:
Flujos monetarios emitidos por los hogares que incluyen impuesto a la renta, contribuciones a la seguridad social, pago por pensiones alimenticias, transferencias a hogares (dentro y fuera del país) y transferencias a instituciones sin fines de lucro que sirven a los hogares.

Los ingresos del trabajo **(1)**, de la propiedad **(2)** y arriendo imputado **(3)** contribuyen al ingreso primario del hogar. Luego, el ingreso total del hogar se obtiene de la suma del ingreso primario y las transferencias corrientes recibidas **(4)**.

A partir de la definición del ingreso total del hogar, se calcula el ingreso disponible del hogar, el que se obtiene al restar del ingreso total los impuestos, los descuentos legales (cotizaciones previsionales y de salud) y las transferencias corrientes emitidas.

Por último, el ingreso disponible ajustado del hogar se construye sumando el ingreso disponible del hogar y las transferencias sociales en especie. Estas últimas corresponden a las transferencias emitidas por el Gobierno e instituciones sin fines de lucro que sirven a los hogares en forma gratuita o a precios menores a los de mercado (por ejemplo, copago en educación), alimentos, vivienda, servicios de salud, enseñanza, cultura y recreación (Camelo, 2001). En la IX EPF, se incluyeron en la medición algunos tipos de transferencias en especies, sin embargo, no fue posible construir el componente en su totalidad.

[31] Las jubilaciones poseen componentes de seguridad social, de capitalización individual (cuentas de ahorro individuales) o una combinación de ambos. En la IX EPF, no fue posible diferenciar entre estos, por lo que se incluyen todos en las transferencias corrientes recibidas.

Una vez definido el concepto de ingreso a medir, la IX EPF definió al **ingreso devengado** como criterio de captura, lo que significa que se registraron los ingresos generados por la actividad económica en el período de referencia (mes anterior), sin importar si estos fueron efectivamente recibidos durante ese tiempo. En otras palabras, ingreso devengado refiere a los ingresos que el informante recibió o debió haber recibido durante el mes anterior.

En el **Diagrama 3** se presenta el concepto de ingreso disponible ajustado mencionado anteriormente, considerado un indicador más cercano al concepto de bienestar de los hogares (OCDE, 2013). Dada la dificultad de medición del ingreso disponible ajustado, la encuesta utiliza el concepto de ingreso disponible como contrapartida del concepto de gasto de consumo final medido por el estudio. Si bien la IX EPF incluyó una pregunta respecto a transferencias sociales en especie recibidas por los hogares, no realizó una consulta en detalle de estas, ya que su objetivo secundario corresponde a identificar la estructura del ingreso disponible.

Si en próximas versiones el objetivo de la encuesta migrara hacia una medición del bienestar económico de los hogares a través del ingreso disponible ajustado, también se debería modificar la forma de medir el consumo, considerando el gasto en consumo final efectivo, que incluye las transferencias sociales recibidas por los hogares (ONU et al., 2016).

[32] En el apartado 5.6 se desarrollan con mayor detalle los clasificadores utilizados.

4.1.4 Clasificadores utilizados

En la IX EPF se utilizaron los siguientes clasificadores[32]:

- **Clasificación Internacional Uniforme de Ocupaciones (CIUO):**

Se empleó a un nivel de dos dígitos[33], los que describen los subgrupos principales. Se utilizó la versión CIUO-08.CL, que corresponde a la adaptación nacional de la clasificación internacional del año 2008. Esta adaptación fue realizada por el INE entre los años 2016 y 2017, y publicada en 2018 (INE Chile, 2018a).

- **Clasificación del Consumo Individual por Finalidades (CCIF):**

Se empleó para clasificar funcionalmente el SCN y así estandarizar los bienes y servicios capturados por la encuesta. Se utilizó hasta el quinto nivel de desagregación, que describe el producto[34]. Se utilizó la adaptación chilena, definida como "Clasificador Chileno del Consumo Individual por Finalidades, CCIF 2018.CL". La adaptación fue realizada por el INE con el apoyo de instituciones externas, como fue

[33] Es importante destacar que, en casos particulares dentro del subproceso de imputación, se utiliza la codificación solamente a un dígito. Esta situación corresponde a una flexibilización de los niveles, principalmente para optimizar la precisión en la imputación de datos.

[34] La clasificación CCIF cuenta con cinco niveles de desagregación: división (primer nivel), grupo (segundo nivel), clase (tercer nivel), subclase (cuarto nivel) y producto (quinto nivel).

el caso de la división 6 del clasificador, correspondiente a salud, en la que el Ministerio de Salud colaboró como revisor, dando como resultado la inclusión y exclusión de ciertos casos específicos en el clasificador adaptado.

- **Clasificador de Establecimientos INE:**

Corresponde a un clasificador interno, empleado para estandarizar los establecimientos que proveen bienes y servicios. Cuenta con 29 categorías y es utilizado como fuente de información para la recolección de precios por parte de los equipos de IPC.

- **Clasificadores de Establecimientos Educacionales:**

El uso dado a este clasificador fue de estandarizar los establecimientos de educación básica, media y superior. Corresponde al conjunto de listados de establecimientos reconocidos por el Ministerio de Educación (MINEDUC). Para los establecimientos de nivel de enseñanza básica y media se utiliza el Rol de Base de Datos (RBD) del Ministerio de Educación.

- **Clasificación Internacional Normalizada de la Educación (CINE):**

En particular la nomenclatura CINE-A, incluido en la versión 2011 del clasificador y que refiere a la clasificación del logro educativo, esto es, el programa educativo más avanzado concluido exitosamente, que suele ser validado por una certificación reconocida por el sistema educativo nacional (Instituto de Estadística de la UNESCO, 2013).

4.2 Referencias nacionales e internacionales

A continuación, se presentan los antecedentes históricos de IX Encuesta de Presupuestos Familiares, dando cuenta de sus principales cambios a través de los años. Adicionalmente, se mencionan las principales referencias internacionales.

4.2.1 Antecedentes históricos de la EPF

La EPF es uno de los productos estadísticos del INE con mayor trayectoria y relevancia, al ser insumo principal para la actualización de la canasta de bienes y servicios que construyen el IPC en el país y sus ponderadores.

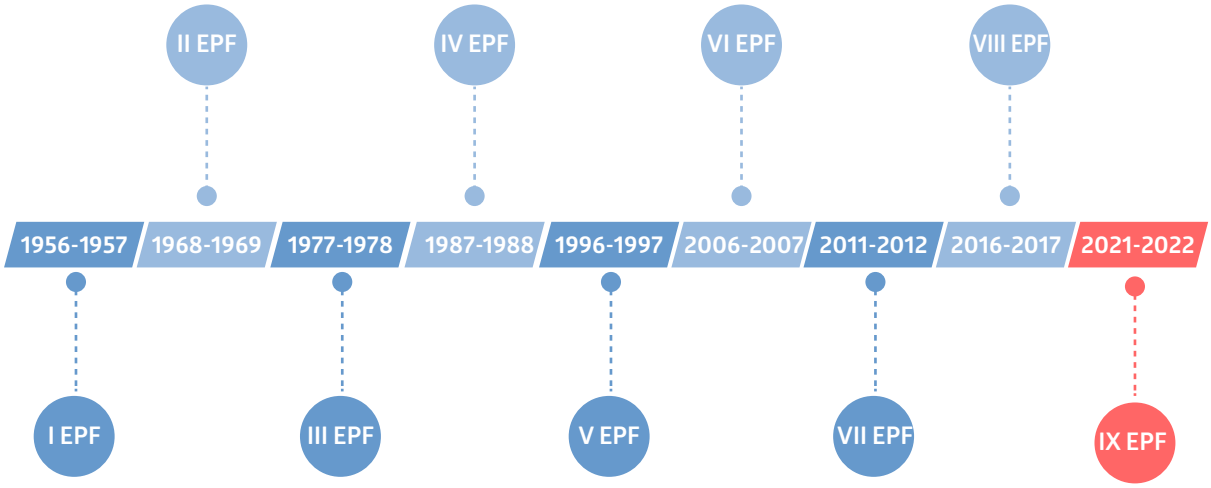
A lo largo del tiempo, las encuestas han variado en sus objetivos, en su metodología de recopilación de los datos y en el tratamiento de estos. El primer estudio sobre el consumo en el Gran Santiago se realizó en el año 1928, con el fin de construir un índice de precios al consumidor. Casi tres décadas después, entre noviembre 1956 y octubre 1957 se realizó la primera Encuesta de Presupuestos Familiares que llevó por título "Estudio Socio-Económico de las Familias de Empleados y Obreros del Gran Santiago", permitiendo realizar el cálculo del IPC de 1957.

Desde entonces, la encuesta tuvo una periodicidad de aproximadamente diez años. Tras el ingreso de Chile a la Organización para la Cooperación y el Desarrollo Económicos (OCDE) en 2010, la periodicidad de la EPF cambia para alinearse con los compromisos de la organización[35], por lo que pasa a ser realizada cada cinco años. Luego del ingreso de Chile a la OCDE, se han realizado dos versiones de la encuesta, siendo la IX EPF la tercera versión desde este cambio.

A continuación, la línea de tiempo detalla los períodos de recolección de datos de cada una de las versiones de la encuesta a la fecha.

[35] El país adquiere el compromiso de actualizar cada cinco años la canasta utilizada para construir el Índice de Precios al Consumidor, con el fin de reflejar oportunamente cambios en los patrones de consumo que se dan con el tiempo.

Imagen 1: Línea de tiempo de versiones de la EPF



Fuente: Instituto Nacional de Estadísticas (INE).

Además de la periodicidad, distintos aspectos característicos de la encuesta han cambiado a lo largo del tiempo. En la siguiente tabla es posible ver un resumen de estos cambios y comparar entre versiones de la encuesta.

Tabla 1: Comparación entre EPF realizadas por el INE

Aspecto	IX EPF	VIII EPF	VII EPF	VI EPF	V EPF	IV EPF	III EPF	II EPF	I EPF
Ámbito geográfico	Capitales regionales y todas sus zonas conurbadas (79 comunas)	Capitales regionales y alguna de sus zonas conurbadas (61 comunas de la VII EPF, además de Chillán y Chillan Viejo, de la nueva Región de Ñuble).	Capitales regionales y algunas de sus zonas conurbadas (60 comunas de la VI EPF, además de Alto Hospicio).	Gran Santiago y ciudades capitales regionales y alguna de sus zonas conurbadas (60 comunas).	Gran Santiago (desde 1992, incluye Padre Hurtado), Iquique y Punta Arenas.	Gran Santiago	Santiago	Hogares particulares de todo el país, procesándo se solo Santiago	Santiago, Providencia, Ñuñoa, San Miguel, Quinta Normal, Renca, Conchalí, Barrancas, La Cisterna, La Granja y Las Condes.
Período de trabajo de recolección	octubre 2021 - septiembre 2022	julio 2016 - junio 2017	noviembre 2011 - octubre 2012	noviembre 2006 - octubre 2007	agosto 1996 - julio 1997	diciembre 1987 - noviembre 1988	diciembre 1977 - noviembre 1978	septiembre 1968 - agosto 1969	noviembre 1956 - octubre 1957
Período de registro	Una quincena	Una quincena	Una quincena	Una quincena	Una quincena	3 meses	3 meses	3 meses	3 meses
Período de recuerdo	3, 6 y 12 meses	3, 6 y 12 meses	3, 6 y 12 meses	3, 6 y 12 meses	1 mes	No hubo	No hubo	No hubo	No hubo
Criterio de medición gasto	Consumo adquirido (para gastos periódicos, consumo pagado)	Consumo adquirido (para gastos periódicos, consumo pagado)	Consumo adquirido (para gastos periódicos, consumo pagado)	Consumo adquirido (excepcional-mente, consumo pagado)	Consumo adquirido	Consumo adquirido	Consumo pagado	Consumo pagado	Consumo pagado
Concepto de gasto	Gasto de consumo final	Gasto de consumo final	Gasto de consumo final	Gasto de consumo final efectivo					
Concepto de ingreso	Ingreso disponible	Ingreso disponible	Ingreso disponible	Ingreso corriente disponible					

Fuente: Instituto Nacional de Estadísticas (INE).

4.2.2 Pautas internacionales para encuestas de presupuesto familiar y la IX EPF

La evolución metodológica de este tipo de encuestas ha sido implementada por las distintas recomendaciones que ha entregado la experiencia internacional en los acercamientos a este tipo de recolección de datos (Eurostat, 2003; CEPAL/INE, 2021). En específico, gran parte del trabajo de la IX EPF se alinea a las recomendaciones del manual de Eurostat *Household Budget Surveys in the EU* (Eurostat, 2003), el que entrega recomendaciones específicas y pautas para desarrollar las encuestas de presupuestos familiares, tomando como base la experiencia de las encuestas de ingresos y gastos realizadas en los países de la Unión Europea.

Por otro lado, el Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (BID, Banco Mundial & CEPAL, 2004b) y el informe Encuestas de Ingresos y Gastos de los Hogares (CEPAL/INE, 2021) entregan lineamientos específicos para las encuestas de este tipo en América Latina y han sido aplicados a lo largo del diseño y la implementación de la IX EPF.

En específico, el informe Encuestas de Ingresos y Gastos de los Hogares recopila las experiencias más recientes en América Latina y el Caribe sobre este tipo de encuestas, y sirve como referencia para esta versión de la EPF, al identificar aquellos elementos comunes que tiene la experiencia latinoamericana en aspectos clave de la encuesta, como el diseño metodológico, el procesamiento, la evaluación de cuestionarios, entre otros.

4.2.3 Experiencias internacionales en encuestas de hogares

En América Latina y el Caribe, las encuestas de ingresos y gastos de los hogares tienen como objetivo principal ser insumo para la actualización del índice de precios al consumidor (CEPAL/INE, 2021) y suelen tener una periodicidad de diez años, la que se reduce a cinco años o menos en el caso de aquellas naciones integrantes de la OCDE. Países

pertenecientes a la Unión Europea, como Alemania, España, Inglaterra e Italia cuentan con encuestas anuales continuas, al igual que otros países como Estados Unidos y Canadá.

Los aspectos metodológicos que caracterizan las encuestas de presupuestos familiares varían entre países, lo que supone diferencias al momento de analizar las experiencias internacionales en encuestas de este tipo. Algunos de los elementos característicos de las encuestas son los siguientes:

- **Período de presencia en el hogar:**

Las encuestas de ingresos y gastos de los hogares utilizan distintos períodos de referencia y formularios que son aplicados en más de una entrevista al hogar seleccionado[36].

La duración del período de cooperación demandado a los hogares por parte del estudio varía entre países, encontrándose mayoritariamente países que utilizan registros diarios de sus gastos durante un período de 7 o 14 días. Dentro del primer grupo se encuentran las encuestas de Argentina, Brasil, Costa Rica, España, Italia, México, Paraguay, Perú y Uruguay. En el caso de Australia, Canadá, Colombia, EE. UU, Portugal y Reino Unido consideran un período de 14 días[37].

Como antecedente de importancia, la Encuesta de Presupuestos Familiares de España se había realizado en un período de 14 días. Sin embargo, para lograr una mayor agilidad en la obtención de información, se tomó la decisión de ajustar la periodicidad a partir del año 2022, cambiando su período de recolección de datos en el hogar desde 14 días a 7 días (INE España, 2023).

La experiencia internacional da cuenta que no hay consenso respecto al período óptimo de presencia en el hogar. La utilización de dos semanas de registro busca mitigar el efecto de semanas atípicas, mientras que el uso de una semana de registro intenta disminuir los días sin registro que genera la fatiga de las personas integrantes del estudio en una segunda semana.

[36] En el caso chileno, para la IX EPF se utilizó una duración de -aproximadamente- una quincena, período en el que las personas integrantes del hogar entregan información sobre ingresos y gastos en los distintos instrumentos de la encuesta.

[37] En el caso de Colombia y Canadá, modifican su período de colaboración a 7 días si se trata de zonas rurales o provincias.

- **Períodos de recuerdo:**

Además de los gastos diarios que se buscan registrar durante los períodos de presencia en el hogar, existen otros tipos de gastos que son consultados considerando períodos de reporte más extensos, ya sea por su estacionalidad, durabilidad u alto costo. Para capturar este consumo de bienes o productos (asociados con algunos servicios específicos y con bienes semidurables y durables), las encuestas de gastos e ingresos de hogares introducen períodos de referencia o de recuerdo.

Si bien no existe acuerdo en la experiencia internacional[38] respecto a cuáles son los períodos de recuerdo óptimos, los más recurrentes corresponden a 3, 6 y 12 meses previos al momento de realización de la entrevista, períodos que fueron utilizados en la IX EPF. La literatura sugiere que, mientras más importante sea el gasto desde el punto de vista del presupuesto de los hogares, más sencillo es de recordar, por lo que aquellos gastos más altos se pueden consultar en períodos de recuerdo más amplios (U.S. BLS., 2004; U.S. BLS., 2006; Bradburn, 2010).

[38] Dado que no existe consenso acerca de un período de recuerdo óptimo, la elección del período adecuado dependerá de una serie de factores que varían de una región a otra y están relacionados con las particularidades económicas, culturales y de consumo de cada región. Los períodos de recuerdo de 3, 6 y 12 meses son recurrentes en las encuestas de presupuesto familiar, ya que ofrecen un equilibrio entre la captura de variaciones significativas en los patrones de gasto y la viabilidad práctica de la recopilación de datos, en especial de aquellos bienes de mayor durabilidad y menor frecuencia en el gasto. Algunos de los países que consideran esos períodos de referencia son Francia, Reino Unido, Bolivia, Costa Rica y Perú. Otros países como España y México varían sus períodos entre una semana, mes, trimestre o semestre dependiendo del tipo del producto registrado, y países como Dinamarca e Irlanda, solo manejan un período retrospectivo de 12 meses.

05.

DISEÑO ESTADÍSTICO

05.

DISEÑO ESTADÍSTICO

En este capítulo se describe el diseño estadístico de la IX EPF, desde los elementos que conforman el diseño muestral, y que dan origen a las viviendas seleccionadas, hasta el desarrollo del factor de expansión, detallando cada una de sus etapas, junto con los estimadores de los indicadores principales de ingreso y gastos obtenidos en la encuesta y sus respectivas medidas de precisión.

5.1 Población objetivo

La población objetivo la constituyen los hogares y las personas que residen habitualmente en las viviendas particulares ocupadas de las áreas urbanas de las capitales regionales del país y sus zonas conurbadas.

5.2 Marco estadístico

El marco estadístico de la EPF es un componente relevante para garantizar la validez y la confiabilidad de los resultados obtenidos. Proporciona las pautas necesarias para el diseño de la encuesta, la selección de la muestra y la estimación de parámetros poblacionales. El marco estadístico que se describe en el presente apartado busca garantizar que los datos recopilados sean representativos y generalizables a la población.

5.2.1 Marco Muestral

Un marco muestral se define como la lista o los procedimientos que permiten identificar a todos los elementos de una población objetivo (Groves, 2004, pág. 68). Al respecto, que este sea completo y preciso es requisito fundamental para que los procesos de diseño, selección muestral, recolección y análisis tengan calidad acorde a los lineamientos de las distintas oficinas de estadísticas. Así también, es necesario contar con marcos muestrales actualizados para no afectar la confiabilidad de las estimaciones y de los productos estadísticos. Si bien los marcos muestrales se desactualizan con el paso del tiempo, es necesario generar

estrategias para mantener procesos continuos de actualización para minimizar eventuales sesgos que se pudiesen generar en el marco. Así, el marco muestral utilizado para la selección de la muestra de la IX EPF corresponde al Marco Muestral de Vivienda 2020 (MMV 2020) conformado a partir de la información del Censo de Población y Vivienda del año 2017, con actualización al año 2020[39].

[39] El MMV 2020 corresponde a una actualización del MMV 2017 que puede incluir nuevas conformaciones de UPM o bien la eliminación de otras. Las UPM son conglomerados homogéneos en relación con el número de viviendas que los conforman. Para mayor información revisar el Documento metodológico Encuesta Nacional de Empleo (ENE) 2020 (INE Chile, 2021).

El MMV 2020 es completo y preciso en su información respecto de todo el territorio nacional y de las unidades presentes en él, debido a la necesidad de contar con marcos de muestreo que posibiliten mejorar la eficiencia y eficacia a nivel técnico y operativo de los estudios realizados por el INE.

En línea con lo anterior, con el fin de contar con un marco muestral que minimice problemas asociados a la desactualización, el INE ha definido realizar actualizaciones anuales, de forma de acotar eventuales problemas de completitud y precisión que presentan los marcos muestrales en general, resguardando también, la comparabilidad y los atributos del marco entre cada período.

5.2.2 Construcción del Marco Muestral

El Marco Muestral de Viviendas (MMV 2020) está constituido por unidades primarias de muestreo (UPM) que corresponden a áreas geográficas homogéneas, en términos del número de viviendas que las conforman.

En la conformación de las UPM se consideran viviendas particulares con moradores presentes o ausentes, excluyendo viviendas de temporada y colectivas. Esto permite contar con unidades estables en el tiempo, disminuir la pérdida de unidades en el proceso de recolección de datos y reducir el riesgo de errores no muestrales.

Cada UPM está categorizada según una estratificación geográfica y socioeconómica.

5.2.3 Estratificación geográfica

La estratificación geográfica está dada por la división político-administrativa, según el ordenamiento región-provincia-comuna. Al interior de las comunas existe la desagregación censal, dando origen a las áreas urbanas y rurales.

La definición de urbano y rural viene dada por las especificaciones del Censo de Población y Vivienda 2017, según la siguiente descripción:

- **Entidad urbana:**
  Asentamiento humano con continuidad y concentración de construcciones en un amanzanamiento regular con población mayor a 2.000 habitantes o, entre 1.001 y 2.000 habitantes, donde menos de 50% de la población que declara haber trabajado, se dedica a actividades primarias.

- **Entidad rural:**
  Asentamiento humano concentrado o disperso que posee 1.000 o menos habitantes o, entre 1.001 y 2.000 habitantes, donde más de 50% de la población que declara haber trabajado, se dedica a actividades primarias.

Sin embargo, al ser la EPF una operación estadística enfocada en el área urbana, el marco *ad-hoc* se reduce, exclusivamente a las áreas urbanas de las 79 comunas elegidas para el desarrollo de la encuesta.

La conformación de las UPM en el área urbana considera un tamaño medio de 200 viviendas, con un intervalo de 20%, es decir, la cantidad de viviendas por UPM se encuentra en el rango [160 - 240] y, en el área rural, un tamaño medio de 90 viviendas, con un intervalo de 20%, quedando su cantidad de viviendas por UPM en el intervalo [70 - 110] con algunas excepciones.

5.2.4 Estratificación socioeconómica

La estratificación por nivel socioeconómico se realiza a nivel de UPM y se basa en un análisis de diversas variables recopiladas en el Censo 2017. De acuerdo con el documento "Estratificación socioeconómica del marco muestral de viviendas 2017" (Guerrero et al., 2019) la clasificación de las UPM según el nivel socioeconómico permite aumentar la eficiencia de la inferencia en las encuestas de hogares que utilizan este marco. En línea con lo anterior, el principal objetivo de realizar la estratificación es inducir mayor eficiencia y precisión en las estimaciones de las estadísticas oficiales, además de servir como fuente de información para encuestas de hogares que requieran mayor análisis de carácter socioeconómico.

Al estratificar se espera que las UPM dentro de un mismo estrato sean similares (disminución de la intra-varianza) y diferentes a las de otro estrato (aumento de la inter-varianza).

A partir de los análisis realizados a las variables recopiladas en el Censo y considerando diversos algoritmos multivariados de clasificación (*Principal Component Analysis* (PCA), K-medias, y *Principal Component Analysis by means of Alternating Least Squares* (PRINCALS)) se obtuvieron 180 vectores de estratificación, que consideraban entre 3 y 5 niveles socioeconómicos que, para evaluar su desempeño se calculó el efecto de diseño de 24 indicadores, eligiendo aquella estratificación que permitiera reducir mayormente la varianza. Los indicadores seleccionados fueron:

1. Tasa de ocupación por sexo
2. Tasa de desocupación por sexo
3. Tasa de inactividad por sexo
4. Porcentaje de extranjeros
5. Porcentaje de personas dentro de la fuerza de trabajo primaria respecto del total de la población económicamente activa (PEA) por sexo
6. Porcentaje de hogares unipersonales
7. Porcentaje de población femenina ocupada por rama
8. Porcentaje de población masculina ocupada por rama
9. Porcentaje de personas según nivel educacional por sexo

Luego de la evaluación, el método utilizado para la estratificación del MMV 2020 es el de estratificación óptima de la primera componente principal del PCA, que considera los siguientes indicadores:

- Porcentaje de personas en la educación superior
- Tasa de ocupación
- Porcentaje de viviendas con índice de materialidad alto
- Indicador sobre total de hijos nacidos vivos

Para estos indicadores, según el método elegido, se obtuvo como resultado una variable de estratificación con tres categorías, denominada nivel socioeconómico (NSE) considerando, además las distribuciones de las UPM según el cruce de comuna y área, ya que la baja prevalencia de UPM en regiones extremas, particularmente en el área rural, hace que sea más conveniente considerar solo 3 estratos.

5.2.5 Construcción de estratos de muestreo

El marco muestral utilizado en la IX EPF se obtiene a partir del Marco Muestral de Viviendas (MMV 2020), en el que, exclusivamente para esta encuesta se construyeron estratos de muestreo según distintos criterios de estratificación.

La conformación de los estratos de muestreo considera la unión de la comuna con el nivel socioeconómico, asegurando que el estrato cuente con una cantidad mínima de UPM. Esta cantidad mínima tiene relación con el trabajo operativo de la encuesta que debe considerar una recolección de datos de al menos, 2 UPM[40]. Por tanto, en cada estrato de muestreo se seleccionaron al menos 24 UPM con determinadas excepciones.

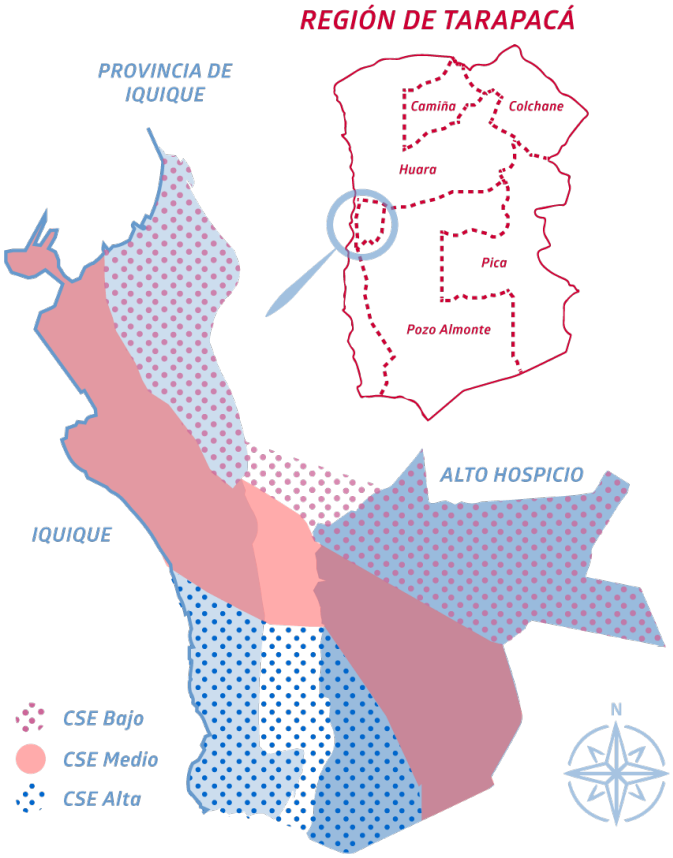
Bajo esta condición, el tamaño de los estratos en el marco muestral, en términos de la cantidad de UPM que los conforman debe considerar que la fracción de muestreo sea máximo 25%.

Cumpliendo con esa condición, los estratos, en la mayoría de las regiones, quedan conformados por comuna-nivel socioeconómico. En caso de no cumplir esta condición, se fusionan (colapsan) comunas, dentro de la misma región y según el mismo nivel socioeconómico o, en regiones donde existe sólo una comuna, se colapsan niveles socioeconómicos hasta lograr la condición establecida. El **Diagrama 4** presenta un ejemplo de la estratificación en la región de Tarapacá según los criterios señalados.

[40] Con ciertas excepciones según cantidad de UPM disponibles por estrato en el marco muestral.

Diagrama 4: Construcción de Estrato N°1: Región de Tarapacá

Región	Nombre comuna	CSE Estrato(1)	UPM	UPM menos 96
1	Iquique	Bajo	24	0
1	Alto Hospicio	Bajo	24	0
1	Iquique	Medio	181	181
1	Alto Hospicio	Medio	63	0
1	Iquique	Alto	115	115
1	Alto Hospicio	Alto	8	0



(1) CSE: Condición socioeconómica.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Esta conformación se mantuvo para todas las regiones que incluyen 1 o 2 comunas en la encuesta. Sin embargo, en el caso de las regiones de Valparaíso, Metropolitana, Biobío y La Araucanía se realizaron procedimientos adicionales para determinar los estratos de muestreo, debido a que en estas regiones la cobertura geográfica está definida por 3 comunas o más. El procedimiento se realizó mediante un análisis de componentes principales y un modelo de *clusterización*. Para el análisis de componentes principales se utilizaron variables determinadas extraídas desde la encuesta CASEN 2017, tales como: tasa de ocupación, porcentaje de personas con educación superior, índice de materialidad de la vivienda, total de hijos nacidos, indicador de hijos nacidos, ingreso disponible *per cápita* del hogar y variables de colindancia geográfica.

A partir de lo anterior, en el marco muestral de la EPF se crean 79 estratos de muestreo conformados por un mínimo de 96 UPM cada uno, con excepción de las regiones de O'Higgins y Aysén, cuyos estratos muestrales quedaron conformados por 92 UPM.

Como se mencionó en párrafos precedentes, la cantidad de UPM a seleccionar en cada estrato de muestreo, debe considerar un mínimo de 24 UPM, salvo algunas excepciones. Según lo anterior, la cantidad de UPM seleccionadas por estrato considera entre 15 y 104 UPM en la muestra.

5.3 Diseño Muestral

El diseño muestral de la encuesta es un diseño probabilístico, bietápico y estratificado. Dentro de cada estrato, la selección de la muestra se realiza en dos etapas:

- **Unidades primarias de muestreo (UPM)** seleccionadas al interior de cada estrato de muestreo.
- **Unidades secundarias de muestreo (viviendas)** seleccionadas al interior de cada unidad de primera etapa.

A partir del diseño muestral se obtiene un tamaño y una distribución en términos de UPM y viviendas que permiten obtener estimaciones de interés con grados de precisión consistentes con estándares nacionales e internacionales en los niveles de estimación definidos. El diseño de la IX EPF permite obtener estimaciones representativas a nivel de Macrozonas (Norte, Centro, Sur y Gran Santiago); en contraste con la VIII EPF que, según su diseño permitía obtener estimaciones representativas para "Gran Santiago" y "resto de capitales regionales".

5.3.1 Cobertura Geográfica

La cobertura geográfica de la IX EPF comprende 79 comunas del país[41], que representan las capitales regionales y sus zonas conurbadas. La **Tabla 2** presenta, por macrozona y región, las comunas presentes en la encuesta.

[41] Respecto a la VIII EPF fueron incluidas 16 comunas. Las razones de esta inclusión obedecen a la revisión de las áreas urbanas funcionales asociadas a las dieciséis capitales regionales del país. Esto es concordante con la noción de "zonas conurbanas" indicada en distintas secciones del documento. Para más información revisar documento "Metodología de Áreas Urbanas Funcionales".

Tabla 2: Comunas incluidas en la IX EPF según región

Macrozona	Región	Comunas incluidas[1]
Norte	Arica y Parinacota	Arica
	Tarapacá	Iquique y Alto Hospicio
	Antofagasta	Antofagasta
	Atacama	Copiapó y Tierra Amarilla
	Coquimbo	La Serena y Coquimbo
Gran Santiago	Metropolitana	Calera de Tango, Cerrillos, Cerro Navia, Conchalí, El Bosque, Estación Central, Huechuraba, Independencia, La Cisterna, La Florida, La Granja, La Pintana, La Reina, Lampa, Las Condes, Lo Barnechea, Lo Espejo, Lo Prado, Macul, Maipú, Ñuñoa, Padre Hurtado, Pedro Aguirre Cerda, Peñaflor, Peñalolén, Providencia, Pudahuel, Puente Alto, Quilicura, Quinta Normal, Recoleta, Renca, San Bernardo, San Joaquín, San Miguel, San Ramón, Santiago, Talagante y Vitacura
Centro	Valparaíso	Valparaíso, Viña del Mar, Quilpué, Concón y Villa Alemana
	O'Higgins	Graneros, Machalí, Olivar, Requínoa y Rancagua
	Maule	Maule y Talca
	Ñuble	Chillán y Chillán Viejo
	Biobío	Concepción, Chiguayante, Coronel, Hualpén, Hualqui, Lota, Penco, San Pedro de la Paz, Talcahuano y Tomé
Sur	Araucanía	Temuco, Padre Las Casas y Vilcún
	Los Ríos	Valdivia
	Los Lagos	Puerto Montt y Puerto Varas
	Aysén	Coyhaique
	Magallanes	Punta Arenas

(1) Actualizado a febrero de 2026. Se incorpora la comuna de Tierra Amarilla.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.3.2 Niveles de Estimación

Se entiende por niveles de estimación o dominios de estudio aquellas desagregaciones geográficas, sociodemográficas o de otro tipo, para las cuales se desea obtener estimaciones del parámetro de interés con medidas de fiabilidad adecuados para garantizar representatividad estadística. La IX EPF permite obtener estimaciones con suficiente grado de precisión del parámetro de interés a nivel nacional y para las macrozonas conformadas según se observa en la **Tabla 3**:

Tabla 3: Conformación de macrozonas según región

Macrozona	Región
Norte	Arica y Parinacota
	Tarapacá
	Antofagasta
	Atacama
	Coquimbo
Gran Santiago	Metropolitana
Centro	Valparaíso
	O'Higgins
	Maule
	Ñuble
	Biobío
Sur	La Araucanía
	Los Ríos
	Los Lagos
	Aysén
	Magallanes

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.3.3 Unidades de muestreo

La IX EPF al poseer un diseño muestral bietápico presentas las siguientes unidades de muestreo:

- **Unidades primarias de muestreo (UPM):** seleccionadas en forma sistemática y con probabilidad proporcional al tamaño, en términos de viviendas particulares, al interior de cada estrato de muestreo.
- **Unidades secundarias de muestreo (USM):** corresponden a viviendas particulares ocupadas; seleccionadas en forma sistemática y con igual probabilidad al interior de cada UPM seleccionada.

5.3.4 Unidad de información

Las unidades de información corresponden a hogares y personas residentes en viviendas particulares seleccionadas para la encuesta.

5.3.5 Unidad de análisis

Las unidades de análisis son hogares y personas pertenecientes a las viviendas particulares que han sido seleccionadas y que contestan la encuesta.

5.3.6 Definición de variable de interés

La variable de interés en términos del diseño muestral IX EPF corresponde al gasto *per cápita* promedio mensual de los hogares.

5.3.7 Cálculo del tamaño muestral

El cálculo del tamaño muestral para la IX EPF prevé una serie de procedimientos que corresponden, en principio, al paso desde un diseño bajo muestreo aleatorio simple a un diseño complejo, seguido de otros dos ajustes, determinados por la finitud de la población y la tasa de no logro de la versión anterior de la encuesta, obteniendo así el tamaño con sobremuestreo.

5.3.8 Consideraciones para el cálculo del tamaño muestral inicial

Para la obtención del cálculo del tamaño muestral en la IX EPF, se establecieron los siguientes criterios:

- **Marco Muestral de Viviendas (MMV):** El diseño muestral de la IX EPF se encuentra alineado con los estándares vigentes de diseño de muestras del INE. Esto implica la utilización del nuevo Marco Muestral de Viviendas (MMV 2020) generado a partir del Censo de Población y Viviendas del año 2017 y, actualizado al año 2020, conformado por unidades primarias de muestreo (UPM) homogéneas, en términos del número de viviendas que las componen y estratificadas en forma geográfica y según condición socioeconómica, como se indica en los apartados anteriores.

- **Representatividad:** Se requiere contar con un tamaño muestral que permita obtener estimaciones representativas para ciertos niveles de desagregación. Para la IX EPF dicha desagregación corresponde a las macrozonas.

- **Precisión estadística:** El tamaño muestral objetivo debe asegurar la obtención estimaciones del parámetro de interés (gasto *per cápita* promedio mensual del hogar) con un error relativo máximo de 5,0% para las macrozonas definidas.

- **Sobremuestreo:** El tamaño calculado, debe considerar un porcentaje de muestra adicional para mitigar posibles problemas asociados al no logro de la encuesta, ya sea por condiciones de elegibilidad de las viviendas o específicamente, por rechazos al momento de la entrevista.

En los párrafos siguientes se describe, en detalle, cada uno de los nueve pasos realizados para definir la muestra con el total de viviendas a encuestar, en principio, a nivel de macrozona, luego a nivel de región y finalmente, a nivel de estratos de muestreo de la IX EPF (comuna-CSE).

El primer paso corresponde a obtener las estimaciones del parámetro de interés y sus estadísticos asociados a partir de los datos recolectados en la VIII EPF.

Sean:

g_{mz} Gasto *per cápita* promedio mensual del hogar en la macrozona mz, en la VIII EPF.

$n \cdot \overline{m} = m_{mz}$ Número de viviendas encuestadas en la macrozona mz, en la VIII EPF.

M_{mz} Número de viviendas en el marco muestral EPF para la macrozona mz.

$Deff\ (g)_{mz}$ Efecto del diseño asociado al gasto *per cápita* promedio mensual del hogar (g) en la macrozona mz en la VIII EPF. Por definición el efecto del diseño es $Deff\ (g) = \sigma^2\ (g)\ MC / \sigma^2 (g)\ MAS$, donde $\sigma^2\ (g)\ MC$ es la varianza bajo un diseño complejo y $\sigma^2\ (g)\ MAS$ es la varianza bajo un muestreo aleatorio simple.

$SE\ (g)_{mz}$ Error estándar de la estimación del gasto *per cápita* promedio mensual del hogar (g) en la macrozona mz en la VIII EPF. Corresponde a la raíz cuadrada de la varianza de la estimación.

$S(g)^2_{mz}$ Cuasivarianza poblacional del gasto *per cápita* promedio mensual del hogar (g) en la macrozona mz en la VIII EPF.

$t^{v}_{1-\alpha/2}$ Cuantil de una distribución t de Student con nivel de confianza $1-\alpha$ (para efectos de la simulación se utiliza un valor fijo $T=2$).

Paso 1

Se obtienen estimaciones, a nivel de cada macrozona del gasto *per cápita* promedio mensual del hogar y sus estadísticos asociados, error estándar y efecto del diseño, usando los resultados del trabajo de recolección de datos de la VIII EPF. La cuasi-varianza poblacional $S(g)^2_{mz}$ se deriva a partir del error estándar obtenido, utilizando la siguiente fórmula:

(1)
$$S(g)^2_{mz} = \frac{n \cdot \overline{m}}{Deff\,(g)_{mz}} \cdot SE(g)^2_{mz}$$

Paso 2

Se obtienen, a partir del error estándar, los errores absolutos por macrozona (d_{0mz}) y luego, los errores relativos (e_{0mz}) según una aproximación del cuantil de una distribución t de Student con nivel de confianza de 95% que, para efectos de la simulación, tiene un valor correspondiente a t=2. Con esto y la cuasivarianza del gasto *per cápita* promedio mensual del hogar en (1) se obtienen los tamaños muestrales iniciales a nivel de cada macrozona, considerando un muestreo aleatorio simple monoetápico m_{0mz} como se muestra en la fórmula (2):

(2)
$$m_{0mz} = \frac{t^{v}_{1-\alpha/2}{}^{2} \cdot S(g)^2{}_{mz}}{d^2_{0mz}}$$

Paso 3

Se calcula el tamaño muestral considerando el efecto de un diseño en dos etapas. Este efecto representa el ajuste al tamaño obtenido en el Paso 2 debido al cambio que experimenta la varianza al pasar de un muestreo monoetápico a uno bietápico. Los valores que se utilizaron como efectos de diseño consideran, por un lado, el comportamiento, en términos de la correlación interclase del nuevo marco MMV 2020 y, por otro, los efectos obtenidos en la recolección de datos de la VIII EPF, considerando como valor máximo 2,5.

La fórmula utilizada para el cálculo del tamaño muestral m_{1r} incorporando el efecto del diseño se presenta a continuación:

(3)
$$m_{1mz} = m_{0mz} \cdot Deff\,(g)_{mz}$$

Donde m_{0mz} representa el número de viviendas estimadas a encuestar en la macrozona mz bajo un muestreo aleatorio simple y m_{1mz} representa el número de viviendas a encuestar en la macrozona mz considerando el efecto de diseño, es decir, bajo un diseño complejo.

Paso 4

A partir del tamaño bajo muestreo aleatorio simple y de la cuasivarianza poblacional obtenida en el Paso 1, se obtiene el error absoluto asociado según la siguiente fórmula:

(4)
$$d_{1mz} = 2 \cdot \sqrt{S(g)^2_{mz}} \,/\, \sqrt{m_{0mz}}$$

Nota: *El error relativo asociado se obtiene al dividir el error absoluto por la estimación del gasto per cápita promedio mensual del hogar.*

Paso 5

Luego, al tamaño m_{1mz} se aplica un ajuste por finitud de la población $(1 - m_{1mz}/M_{mz})$ obteniéndose el tamaño m_{2mz}. La fórmula utilizada se presenta a continuación:

(5)
$$m_{2mz} = \frac{m_{1mz}}{1 + m_{1mz}/M_{mz}}$$

Donde m_{2mz} representa el total de viviendas a encuestar en la macrozona mz considerando un diseño muestral bietápico que incorpora el efecto del diseño y el ajuste por finitud y M_{mz} corresponde al total de viviendas de la población en la macrozona mz, según el marco de muestreo utilizado.

Una vez obtenido el tamaño ajustado por el efecto del diseño a nivel de macrozona, se distribuye en los estratos que lo constituyen en forma proporcional según el número de viviendas por estrato que contiene el Marco Muestral de Viviendas MMV 2020.

Paso 6

Definido el total de unidades a encuestar es posible que, en el trabajo de recolección, no se logren algunas unidades de la muestra por diversas razones, tales como: rechazos, viviendas con moradores ausentes, edificaciones que no forman parte de la muestra objetivo, entre otras. Con la finalidad de salvaguardar la precisión de la estimación de parámetros de interés para la IX EPF, se aplica un factor de ajuste a la fórmula indicada en el paso anterior (5) según la siguiente secuencia: Primero, aplicando las tasas de no respuesta de la VIII Encuesta de Presupuestos Familiares y, luego, las tasas de no elegibilidad de la Encuesta Nacional del Empleo (ENE)[42], obteniendo el tamaño con sobremuestreo.

Utilizar la tasa de no elegibilidad de la ENE como ajuste en el cálculo del tamaño muestral permite, no sólo contar con una aproximación de la cantidad de viviendas que no responderán la encuesta sino que además,

[42] Se determinó utilizar un promedio simple de los meses enero y febrero 2020 de la Encuesta Nacional de Empleo, acotado a las mismas áreas geográficas de la IX EPF, debido a que entregaba información más actualizada de elegibilidad de las unidades que la VIII EPF.

considera el hecho que, en el trabajo de recolección de datos, es posible que en algunas unidades no se conozca su estado de elegibilidad, quedando clasificadas como "elegibilidad desconocida" o bien, que sean edificaciones que no pertenecen al universo de estudio (edificaciones de otros usos) quedando clasificadas como "no elegibles". En presencia de estos casos, la precisión de la estimación del parámetro de interés no debiera verse afectado puesto que, el tamaño con sobremuestreo ya considera una pérdida de unidades por estos motivos.

La fórmula que define el tamaño muestral final para la IX EPF, que incorpora los ajustes conjuntos de efecto del diseño, finitud y tasa de no logro para cada estrato de muestreo (comuna-CSE) está dada por:

Etapa 1:

Tamaño con sobremuestreo considera la tasa de no respuesta de la VIII EPF:

(6.1)
$$m_{2.1\,e} = \frac{m_{2e}}{1 - tnl_{c.1}}$$

Donde:

m_{2e} Número de viviendas objetivo en el estrato de muestreo e

$tnl_{c.1}$ Tasas de no respuesta de la VIII Encuesta de Presupuestos Familiares.

$m_{2.1\,e}$ Número de viviendas con sobremuestreo considerando la tasa de no respuesta de la VIII EPF.

Etapa 2:

Tamaño con sobremuestreo Final: *considera el ajuste 6.1 y el ajuste de la tasa de no elegibilidad de la Encuesta Nacional de Empleo*

(6.2)
$$m_{3e} = \frac{m_{2.1e}}{1 - tnl_c}$$

Donde:

m_{3e} Número de viviendas a encuestar en el estrato de muestreo e (comuna-CSE).

tnl_c Tasa de no elegibilidad obtenida desde la Encuesta Nacional del empleo.

Paso 7

Una vez definido el tamaño con sobremuestreo, se debe ajustar para que, en cada UPM seleccionada se elija la misma cantidad de viviendas (8, en principio). Luego, obtenidos los tamaños ajustados por estrato, se calculan, por agregación, los tamaños con sobremuestreo regionales y por macrozonas.

Finalmente, el escenario elegido corresponde a aquel con representatividad a nivel de macrozonas (Norte, Centro, Sur y Gran Santiago) que considera un tamaño muestral de 29.231 unidades (con sobremuestreo). En la **Tabla 4** se observan los tamaños muestrales obtenidos y los errores esperados para cada una de las macrozonas.

Tabla 4: Estimación y errores propuestos

Macrozona	Estimación de la media Pi (IIIV EPF)	Propuesta error relativo macrozona <=5,5%							
		Error absoluto o propuesto	Error relativo o propuesto	Deff propuesto	Tamaño bajo MAS (m0)	Tamaño ajustado por Deff (m1)	Tamaño ajustado por finitud (m2)	Tamaño con Sobremues-treo	Tamaño UPM
Nivel País (Gasto Promedio *per cápita*)	399.060	10.436	2,60%	2,5	6.782	16.955	16.914	29.231	3.622
Norte	342.803	18.854	5,50%	2,5	1.177	2.944	2.934	4.907	583
Gran Santiago	444.992	20.077	4,50%	2,5	2.461	6.152	6.144	12.112	1.514
Centro	347.472	14.005	4,00%	2,5	2.028	5.069	5.057	7.880	985
Sur	320.015	17.601	5,50%	2,5	1.116	2.790	2.779	4.332	540

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

En la **Tabla 5** se detalla la cantidad de viviendas a encuestar por región en la IX EPF.

Tabla 5: Distribución Unidades Secundarias de Muestreo (Viviendas) EPF

Región	Viviendas a encuestar
Total Nacional	29.231
Arica y Parinacota	640
Tarapacá	782
Antofagasta	1.184
Atacama	552
Coquimbo	1.749
Valparaíso	2.800
Metropolitana de Santiago	12.112
Libertador General Bernardo O'Higgins	976
Maule	1.144
Biobío	2.304
Ñuble	656
La Araucanía	1.736
Los Lagos	1.400
Los Ríos	656
Aysén Del General Carlos Ibáñez Del Campo	280
Magallanes y de La Antártica Chilena	260

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.3.9 Distribución del número de UPM por submuestra mensual

A partir del número de viviendas a encuestar por estrato, se obtiene el número de unidades primarias de muestreo (UPM). Luego, cada UPM es asignada a una submuestra que indica el mes de recolección de datos primero, a nivel regional y luego, por estratos muestreo. En la **Tabla 6** se observa la cantidad de UPM total y por submuestra a nivel nacional y regional:

Tabla 6: Distribución Unidades Primarias de Muestreo (UPM) EPF

REGIÓN	Sub M1	Sub M2	Sub M3	Sub M4	Sub M5	Sub M6	Sub M7	Sub M8	Sub M9	Sub M10	Sub M11	Sub M12	Total general
Total Nacional	302	302	302	302	302	301	302	302	302	302	302	302	3.622
Arica y Parinacota	7	6	6	6	6	7	7	7	7	7	7	7	80
Tarapacá	8	8	8	8	8	6	7	7	7	7	7	7	88
Antofagasta	12	12	12	12	12	13	13	13	13	12	12	12	148
Atacama	6	6	6	6	6	6	5	5	5	6	6	6	69
Coquimbo	16	16	16	16	16	16	17	17	17	17	17	17	198
Valparaíso	30	30	29	29	29	29	29	29	29	29	29	29	350
Metropolitana de Santiago	126	126	126	126	126	127	127	126	126	126	126	126	1.514
Libertador General Bernardo O'Higgins	10	10	11	11	10	10	10	10	10	10	10	10	122
Maule	12	12	12	11	12	12	12	12	12	12	12	12	143
Biobío	24	24	24	24	24	24	24	24	24	24	24	24	288
Ñuble	6	7	7	7	7	7	7	7	7	7	7	6	82
La Araucanía	18	18	18	19	18	18	18	18	18	18	18	18	217
Los Lagos	14	14	14	14	15	15	15	15	15	15	15	14	175
Los Ríos	7	7	7	7	7	6	6	7	7	7	7	7	82
Aysén del General Carlos Ibáñez del Campo	3	3	3	3	3	3	3	3	3	3	2	3	35
Magallanes y de la Antártica Chilena	3	3	3	3	3	2	2	2	2	2	3	3	31

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.3.10 Selección de unidades muestrales

La selección de unidades es un paso relevante en la EPF, ya que determina qué hogares serán incluidos en la muestra y, por ende, influirá en la representatividad de los resultados.

5.3.10.1 Selección de unidades primarias de muestreo (UPM)

La selección de UPM se realizó al interior de cada estrato de muestreo, en forma sistemática con probabilidad proporcional al tamaño del marco muestral de selección, según el número de viviendas que las conformaban.

5.3.10.2
Selección de unidades secundarias de muestreo (viviendas particulares ocupadas)

Las viviendas fueron seleccionadas al interior de cada UPM de forma sistemática y con igual probabilidad.

El algoritmo de selección sistemático de viviendas, que fue aplicado también en otras encuestas, se estandarizó para que, independientemente de quien aplique el algoritmo y bajo las mismas condiciones de ordenamiento previo, se seleccionen las mismas viviendas. Este procedimiento es análogo a introducir una semilla de aleatorización para selecciones aleatorias, las cuales arrojan las mismas selecciones. El procedimiento se describe a continuación.

Paso 1

Considerando solo las viviendas particulares con moradores presentes o ausentes, se ordenaron los elementos de forma ascendente según: región, comuna, vector orden (que identifica el orden de la UPM dentro del estrato de muestreo), distrito censal, código de zona localidad, código de manzana, número de orden de la edificación y número de orden de la vivienda.

Paso 2

Sea m_{ei} el número de viviendas a encuestar y M'_{ei} el número de viviendas particulares con moradores presentes o ausentes en la UPM i perteneciente al estrato e. Se calculó el período Q según la ecuación (7):

$$Q = \frac{M'_{ei}}{m_{ei}} \tag{7}$$

Paso 3

Se calcula el arranque "A" o primera vivienda seleccionada, como un número aleatorio entre 1 y el período Q. Con el fin de asegurar la réplica de los cálculos, se establece una semilla aleatoria fija y única para la IX EPF que corresponde a "12020".

Paso 4

Al arranque "A" se le va sumando sucesivamente el período, para obtener distintos valores que, al redondearlos van generando sucesivas selecciones:

$$A,\ (A+Q),\ (A+2\cdot Q),\ldots,A+(m_i-1)\cdot Q$$

La primera selección, A, ya es un número entero, la segunda es el redondeo de $(A+Q)$, la tercera es el redondeo de $(A+2\cdot Q)$ y, así sucesivamente hasta la m_i -ésima selección, dada por el redondeo de $A+(m_i-1)$.

El número de unidades seleccionadas (primera y segunda etapa) se muestra en la siguiente tabla:

Tabla 7: Distribución Unidades primarias (UPM) y secundarias de muestreo (Viviendas) EPF

Región	UPM	Viviendas
Total Nacional	3.622	29.231
Arica y Parinacota	80	640
Tarapacá	88	782
Antofagasta	148	1.184
Atacama	69	552
Coquimbo	198	1.749
Valparaíso	350	2.800
Metropolitana de Santiago	1.514	12.112
Libertador General Bernardo O'Higgins	122	976
Maule	143	1.144
Biobío	288	2.304
Ñuble	82	656
La Araucanía	217	1.736
Los Lagos	175	1.400
Los Ríos	82	656
Aysén del General Carlos Ibáñez del Campo	35	280
Magallanes y de la Antártica Chilena	31	260

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.4 Desarrollo de los factores de expansión

Como se mencionó en capítulos anteriores, el diseño muestral de la IX EPF es probabilístico, estratificado y bietápico. La primera etapa corresponde a la selección de unidades primarias de muestreo (UPM) al interior de cada estrato. La segunda etapa corresponde a la selección de unidades secundarias de muestreo (viviendas) al interior de cada UPM seleccionada en la primera etapa.

En el trabajo de recolección de datos de la IX EPF se gestionaron 29.231 viviendas, de las cuales, se logró entrevistar 14.961, que contienen 15.134 hogares y 44.688 personas a nivel nacional. Las viviendas que no se lograron entrevistar, a partir de la información proporcionada por la Hoja de Ruta, son clasificadas en 3 grandes grupos: elegibles, no elegibles y de elegibilidad desconocida; información útil para realizar los ajustes al factor de expansión en sus distintas etapas.

En este apartado se describe la forma de cálculo del factor de expansión desarrollado para cada una de las viviendas gestionadas en la encuesta, considerando las siguientes etapas para su conformación:

- **Cálculo de factor teórico o factor de selección de unidades:** Da cuenta de la probabilidad de selección de las UPM, así como también, de las viviendas al interior de cada UPM seleccionada.
- **Ajuste por elegibilidad desconocida y no elegibilidad.**
- **Ajuste por no respuesta.**
- **Suavizamiento.**
- **Calibración a las proyecciones de población estimadas al mes de marzo 2022[43].**

El detalle de cada una de estas etapas se describe a continuación.

5.4.1 Factor teórico

El factor teórico se interpreta como la cantidad de unidades en la población que representa una unidad en la muestra.

Corresponde al inverso de las probabilidades de selección de las unidades muestrales. De acuerdo con el diseño muestral y al tratarse de un muestreo bietápico, este factor se obtuvo a partir de la selección de unidades de primera y segunda etapa.

[43] Se determina como mes de referencia al mes de marzo 2022 debido a que corresponde al mes central del período de recolección oficial IX EPF (octubre 2021 a septiembre 2022).

5.4.2 Ponderador de unidades de primera etapa

El ponderador de unidades de primera etapa corresponde al inverso de la probabilidad de incluir cada UPM en la muestra en el estrato h, su forma de cálculo se observa en la ecuación (1):

$$w_{hi} = \frac{1}{P_h(i)} = \frac{M_h}{n_h \cdot M_{hi}} \quad (1)$$

Donde:

h Corresponde al estrato de muestreo conformado por comuna y clasificación socioeconómica.

M_h Corresponde al total de viviendas en el estrato h según MMV 2020.

n_h Corresponde al total de unidades de primera etapa seleccionadas en cada estrato h.

M_{hi} Corresponde a la cantidad de viviendas que contiene cada unidad de primera etapa i (UPM) seleccionada en el MMV 2020.

5.4.3 Ponderador de unidades de segunda etapa

Corresponde al inverso de la probabilidad de incluir cada vivienda particular ocupada dentro de la UPM seleccionada en la muestra. Su forma de cálculo se observa en la ecuación (2):

$$w_{hij} = \frac{1}{P_{hi}(j)} = \frac{M'_{hi}}{m_{hi}} \quad (2)$$

Donde:

M'_{hi} Corresponde al número actualizado de viviendas particulares ocupadas según el proceso de enumeración o verificación de directorios de las unidades de primera etapa i seleccionadas en el estrato h.

m_{hi} Corresponde al total de viviendas a encuestar en cada unidad de primera etapa i seleccionadas en el estrato h.

A partir de estos dos ponderadores se construye el factor de expansión teórico o de selección de unidades para cada vivienda de la muestra, tal como se expresa en la ecuación (3):

$$W_{hij} = w_{hi} \cdot w_{hij} = \frac{1}{P_h(i)} \cdot \frac{1}{P_{hi}(j)} = \frac{M_h}{n_h \cdot M_{hi}} \cdot \frac{M'_{hi}}{m_{hi}} \quad (3)$$

Donde W_{hij} corresponde al factor de expansión asignado a la ***j-ésima*** vivienda seleccionada al interior de la ***i-ésima*** UPM perteneciente al estrato de muestreo ***h***.

5.4.4 Ajuste por elegibilidad

La población objetivo de la IX EPF está constituida por las personas y hogares que residen habitualmente en las viviendas particulares ocupadas de las 79 comunas presentes en la encuesta. Sin embargo, pese a los esfuerzos realizados por mantener marcos muestrales válidos para esta población, siempre existe la posibilidad de que existan errores de inclusión (o exclusión) en el marco. Por ejemplo, cambios en el uso de la edificación u otro tipo de situaciones que ocurren entre el proceso de actualización de unidades muestrales de segunda etapa y el momento de la entrevista, las cuales afectan la elegibilidad de las viviendas seleccionadas.

Por esta razón, es importante aplicar un ajuste al factor de selección de las viviendas con el objetivo de considerar el hecho que, luego del trabajo de recolección de datos, una proporción de estas es clasificada como no elegible y, en otros casos simplemente, se desconoce su estado de elegibilidad.

Según lo anterior, todas las viviendas seleccionadas se clasifican en tres grandes grupos:

- **Elegibles:**
  Corresponden a edificaciones identificadas como viviendas particulares ocupadas (incluye tanto viviendas entrevistadas como no entrevistadas).

- **No Elegibles:**
Pertenecen a un grupo de edificaciones identificadas como negocios, viviendas colectivas, viviendas deshabilitadas, viviendas de veraneo, viviendas destruidas, entre otras.

- **Elegibilidad desconocida:**
Edificaciones cuyo estado no se pudo determinar. Este es el caso, por ejemplo, de viviendas que no pudieron ser localizadas, unidades a las que no se pudo acceder, entre otras.

La **Tabla 8** muestra la clasificación de las viviendas según su estado de elegibilidad obtenida luego del trabajo de recolección de datos de la IX EPF:

Tabla 8: Distribución de viviendas seleccionadas según estado de elegibilidad

Elegibilidad	Frecuencia	Porcentaje
Total	29.231	100%
Elegible	26.882	91,96%
Elegibilidad desconocida	238	0,81%
No Elegible	2.111	7,22%

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

La **Tabla 9** muestra el porcentaje de viviendas clasificadas según su estado de elegibilidad, tras el trabajo de recolección de datos, al interior de cada macrozona.

Tabla 9: Porcentaje de viviendas clasificadas según elegibilidad y macrozona

Macrozonas	Elegibilidad Desconocida (%)	Elegible (%)	No Elegible (%)
Norte	0,35%	90,20%	9,46%
Gran Santiago	1,32%	93,25%	5,42%
Centro	0,58%	91,59%	7,83%
Sur	0,35%	91,04%	8,61%
Total	0,81%	91,96%	7,22%

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

El factor de expansión teórico o de selección de unidades, tiene valores válidos para todas las viviendas de la muestra, es decir para viviendas elegibles, no elegibles y de elegibilidad desconocida. En las siguientes secciones de este capítulo se describen los distintos ajustes aplicados al factor teórico según la clasificación obtenida para cada vivienda.

5.4.5 Ajuste por elegibilidad desconocida

En el proceso de actualización de directorios de unidades de primera etapa, todas las viviendas poseen un estado conocido de elegibilidad. No obstante, la muestra de viviendas seleccionadas incluye solo viviendas particulares ocupadas (elegibles). Luego del trabajo de recolección de datos dicho estado no se pudo comprobar en algunas unidades, por lo que quedan clasificadas como de elegibilidad desconocida.

El procedimiento que se aplica al factor de expansión para este tipo de viviendas es distribuir proporcionalmente, al interior de cada estrato de muestreo, el peso que se les asignó en la etapa previa a la recolección de datos, en las unidades que poseen elegibilidad conocida, con la finalidad de no subestimar el número de viviendas de la población.

El ponderador de los pesos a redistribuir se calculó como la razón entre dos estimaciones obtenidas con base en el factor teórico (W_{hij}). En el numerador se ubica la estimación del total de viviendas de la población y en el denominador la estimación total de las viviendas con elegibilidad conocida. Esto se observa en la ecuación (4):

(4)
$$R_{h,know} = \frac{\sum_{i \in \Omega_h} \sum_{j \in \Theta_i} W_{hij}}{\sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,know}} W_{hij}}$$

Donde:

Ω_h Es el conjunto de unidades de primera etapa (UPM) seleccionadas en el estrato h.

Θ_i Es el conjunto de viviendas seleccionadas al interior de cada unidad de primera etapa.

$\Theta_{i,know}$ Es el conjunto de viviendas seleccionadas y clasificadas como elegibilidad conocida (en esta categoría son consideradas las viviendas elegibles y no elegibles).

W_{hij} Es el factor teórico que da cuenta de la selección de la vivienda $\boldsymbol{j}$ en la unidad de primera etapa $\boldsymbol{i}$ seleccionada en el estrato $\boldsymbol{h}$.

El factor que da cuenta de la selección de viviendas, ajustado por elegibilidad desconocida, $\boldsymbol{W'_{hij}}$, está dado por la ecuación (5):

$$W'_{hij} = W_{hij} \cdot R_{h,know} \tag{5}$$

Finalmente, el factor de expansión asignado a las viviendas clasificadas con elegibilidad desconocida es eliminado y, la suma total del factor de expansión asignado a las viviendas clasificadas con elegibilidad conocida, es equivalente a la suma del factor de expansión original.

5.4.6 Ajuste por no elegibilidad

En la **Tabla 10** se observa que de un total de 29.231 viviendas seleccionadas para la IX EPF, 2.111 resultaron no elegibles. A partir de este momento tales viviendas son descartadas de la muestra pues no forman parte de la población objetivo de este estudio. Esto se implementa asignando a las viviendas no elegibles un valor blanco (*missing value)* en el factor ajustado por elegibilidad desconocida $\boldsymbol{(W'_{hij})}$.

5.4.7 Ajuste por no respuesta

Aun cuando en esta etapa se consideran solo viviendas particulares ocupadas (elegibles) es posible que no se logre conseguir respuesta en la totalidad de las viviendas de la muestra, ya sea porque los residentes no desean participar en el estudio, o bien, porque no han estado presentes al momento de la entrevista. En estos casos, es necesario realizar un ajuste adicional al factor, considerando distribuir el peso de las viviendas donde no se logró contacto en aquellas donde sí se logró.

Los ajustes de no respuesta aplicados al factor de expansión se realizan considerando información proveniente de la recolección de datos de la encuesta u otras fuentes externas, y tienen como objetivo corregir las posibles fallas cuando no se logra la muestra completa de viviendas elegibles.

5.4.8 Consideraciones

Según lo descrito en el apartado anterior, a partir del trabajo de recolección de datos, es posible visualizar diferentes estados de las viviendas en relación con su elegibilidad (elegibles, no elegibles y elegibilidad desconocida). Estos estados se obtienen según el código de disposición final registrado en la vivienda al momento de la entrevista.

Una vivienda elegible se divide entre entrevistada completa, entrevista parcial y no entrevistada. Las entrevistadas parciales corresponden a aquellas donde el informante llega hasta la mitad del cuestionario y luego, por varias razones, nunca lo completa, pero aun así proporcionan suficiente información como para ser consideradas parciales en vez de interrumpidas o *break off*, que se entiende como un tipo de rechazo (AAPOR, 2023, pág. 10).Las no entrevistadas son viviendas donde existe ausencia completa de información, es decir, cuando existen rechazos iniciales o interrupciones, no contactos por ausencia de moradores o barreras que impiden llegar a las viviendas y otras situaciones de no contacto como el idioma, salud, viajes, entre otras circunstancias de terreno.

La **Tabla 10** muestra la clasificación de las viviendas según su estado de logro de entrevista.

Tabla 10: Clasificación según el logro de la entrevista

Clasificación Entrevista	Sigla	Frecuencia	Porcentaje
Elegibles	-	26.882	100%
Entrevistadas	E	14.961	55,7%
Break off	BO	3.092	11,5%
Rechazo Inicial	RI	4.635	17,2%
No contactadas	NC	3.321	12,4%
Otras no entrevistadas	O	873	3,3%

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

La **Tabla 11** muestra una serie de indicadores respecto al desempeño operativo de cada vivienda.

Tabla 11: Indicador de desempeño operativo a nivel de viviendas IX EPF

Indicador de desempeño operativo	Cálculo del indicador(1)	(%)
Desactualización del marco muestral	$\frac{(ED+NE)}{(E+BO+RI+NC+O+ED+NE)}$	8,0%
Tasa de contacto	$\frac{(E+BO+RI+O)}{(E+BO+RI+NC+O+ED)}$	86,9%
Tasa de cooperación	$\frac{E}{(E+BO+RI+O)}$	63,5%
Tasa de rechazo	$\frac{(BO+RI)}{(E+BO+RI+NC+O+ED)}$	28,5%
Tasa de respuesta	$\frac{E}{(E+BO+RI+NC+O+ED)}$	55,2%

(1) Las siglas ED y NE corresponden a las clasificaciones de "Elegibilidad desconocida" y "No elegibilidad" respectivamente.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

A partir de esta información, surge la necesidad de aplicar un ajuste a los factores de expansión obtenidos en la etapa previa, con el objetivo de lograr que las unidades que no respondieron sean representadas por las que sí, previendo no introducir sesgos debido a la posibilidad de que exista relación entre la no respuesta y la variable de interés. De un total de 29.231 viviendas seleccionadas resultaron 26.882 viviendas elegibles que representan 91,9%. De éstas, 14.961 viviendas contestaron la encuesta, lo que representa el 55,7% de la muestra de viviendas seleccionadas. De lo anterior se desprende que la tasa de no respuesta corresponde a 44,4% de las viviendas.

El método a implementar para compensar la falta de respuesta es el método de estratificación mediante "*Propensity Score*". De acuerdo, a lo indicado por Valliant (Valliant, Dever, & Kreuter, 2013) este método consiste en modelar la probabilidad de respuesta en una encuesta como la realización de un proceso de variables latentes $(R_i^* = x_i^T \beta + u_i)$ es decir, un conjunto de covariables que inciden en la "motivación" (R^*) de participar o no en la encuesta. Así, mediante un conjunto de covariables conocidas para quienes responden y quienes no responden, se busca estimar la probabilidad de responder en la encuesta $(P(R_i^* > \theta))$ y posteriormente, ordenar estas probabilidades subdividiéndolas en 20 partes iguales llamadas veintiles.

Para seleccionar las covariables que más aportan al modelo de regresión logística *(Propensity Score)* se utiliza el método de bosques aleatorios, popularmente conocido como *Random Forest* (Breiman, 2001). A partir de este método, se conforman conjuntos de árboles de decisión individuales, en que cada uno se prueba extrayendo una muestra aleatoria desde la recolección de datos mediante *bootstrapping* (o remuestreo). Esto implica que cada árbol se prueba con conjuntos de datos ligeramente distintos. En cada árbol individual, las observaciones se van distribuyendo según distintas bifurcaciones (nodos) generando la estructura del árbol hasta alcanzar un nodo terminal. La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo. Luego, obtenidas las covariables con el método *Random Forest,* se conforman distintos modelos *logit* a partir de distintas combinaciones de estas covariables.

Las covariables que son incluidas en el modelo miden, por un lado, el estado y entorno de la vivienda y, por otro, variables características del propio encuestador[44] en relación con la posibilidad de lograr la entrevista. Ejemplos de esto son: variables que indican si una vivienda tiene acceso directo o no, o si un encuestador tiene o no experiencia en encuesta de hogares, lo que podría influir en la respuesta de la vivienda, entre otros.

[44] La información de las personas encuestadoras fue recogida por medio de un formulario de caracterización laboral y social, aplicado a todo el personal de operaciones de la IX EPF, entre los años 2021 y 2022.

Las posibles variables (covariables) que se incluyeron en el modelo para predecir la no respuesta son:

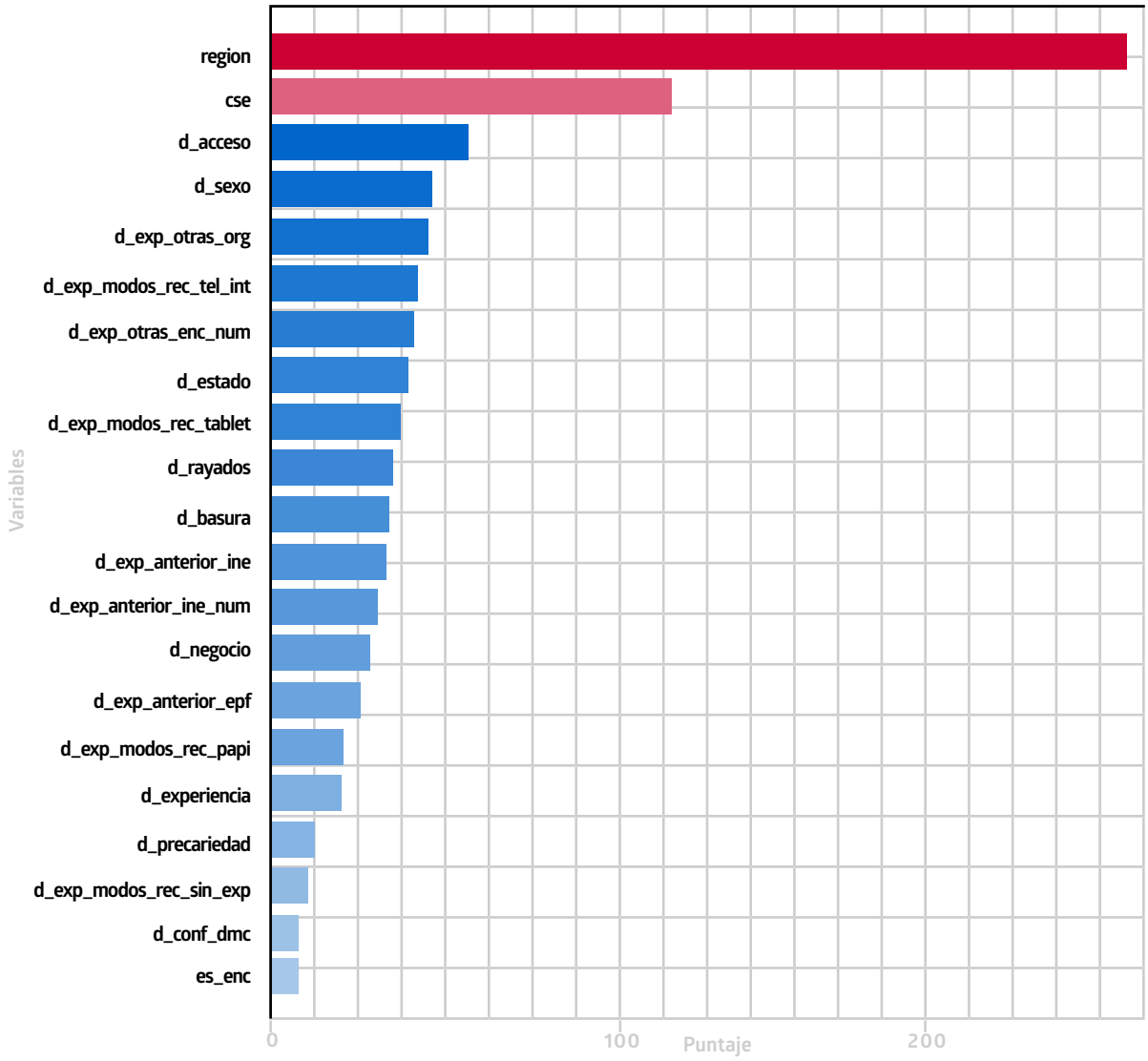
Tabla 12: Covariables incluidas en el método *Random Forest*

Variables de estado y entorno de la vivienda
Acceso a la vivienda
Estado de la vivienda
Basura en el entorno
Rayados en el entorno
Tipo de vivienda
Negocio adosado a la vivienda
Variables de la persona encuestadora
Sexo
Años de experiencia
Experiencia anterior en el INE
Experiencia anterior en EPF
Experiencia anterior en otras organizaciones
Confianza para manejar la Tablet
Experiencia en recolección de datos utilizando cuestionarios de papel
Experiencia en recolección de datos utilizando Tablet
Experiencia en recolección de datos utilizando teléfono inteligente
Sin experiencia en recolección de datos

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Así, a través del método de *Random Forest*, es posible observar la importancia[45] de las variables que, eventualmente serán incluidas en el modelo *Propensity Score* y, posteriormente se seleccionan las mejores candidatas para ingresarlas al modelo. Esto se observa en el **Gráfico 1**:

Gráfico 1: Importancia de las variables para ser incluidas en el modelo *Propensity Score*



Nota: Definición de variables disponibles en el Anexo 2.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

[45] El *Random Forest* tiende a seleccionar variables con alta cardinalidad en la parte superior, dándole más importancia que a otras, en función de la ganancia de información. En la práctica, es una prueba de que está sobre ajustado.

Este gráfico permite determinar las variables que se incorporarán a las pruebas de análisis que, para la IX EPF fueron consideradas las primeras diez con mayor importancia para ser incluidas en los modelos *Propensity Score*, además de revisar las covariables incluidas en los modelos seleccionados en el cálculo de los factores de expansión en etapas de recolección de datos parciales de la encuesta (primer, segundo y tercer trimestre acumulados) con el fin de obtener el mejor modelo considerando un puntaje asignado a cada uno de ellos según los criterios de AIC, BIC, R^2 [46].

[46] AIC y BIC: Criterios de información de Akaike y Schwarz. Medidas de entropía relativa que consideran la bondad de ajuste y penalizan la complejidad del modelo. Un modelo con menor AIC o BIC utiliza de mejor forma la información, mientras que R^2 corresponde a una aproximación de la medida de modelos lineales sobre el porcentaje de la varianza explicada por el modelo.

Luego del análisis se concluye que, si bien, el mejor modelo incluye 14 covariables, por principio de parsimonia, se elige aquel que, con sólo 6 permite predecir la no respuesta sin perder significativamente la eficiencia de dicha predicción. No obstante, a este modelo se decide incorporar una variable adicional que, en el mejor modelo según el método, representa la mayor importancia entre las 14 covariables que lo conforman. Esta corresponde a "Experiencia en recolección de datos utilizando Tablet".

Tabla 13: Covariables seleccionadas para el modelo *Propensity Score*

Variables seleccionadas
Región
Condición socioeconómica (CSE)
Acceso a la vivienda
Sexo
Experiencia en otras encuestas (mayor a 2 años)
Experiencia anterior en otras organizaciones
Experiencia en recolección de datos utilizando Tablet

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Seleccionadas las covariables que participarán en el modelo *Propensity Score,* se calcula la probabilidad de responder de cada una de las viviendas de la encuesta; luego, se ordenan las unidades de mayor a menor según estas probabilidades estimadas y se crean estratos o "celdas de ajustes", que serán los niveles sobre los cuales se realizan los ajustes de no respuesta, según la agrupación de viviendas que poseen características similares asociadas a la respuesta de la encuesta. En este caso, se generan 20 celdas de ajustes o veintiles, con igual cantidad de viviendas, ordenadas por las probabilidades antes mencionadas.

Una vez creadas estas 20 celdas de ajuste, se procede a estimar el factor de viviendas que componen la IX EPF, ajustado a la falta de respuesta. Este ajuste se observa en la ecuación (6):

$$w_{hij}^{NR} = \hat{R}_v \cdot W'_{hij} \tag{6}$$

Donde, $\hat{R}_v$ corresponde a :

$$\hat{R}_v = \begin{cases} \frac{\sum_{j \in \Theta_{eleg,v}} W'_{hij}}{\sum_{j \in \Theta_{R,v}} W'_{hij}} & \textbf{\textit{, Si la vivienda j es elegible y responde la encuesta}} \\ \cdot & \textbf{\textit{, Si la vivienda j es elegible pero no responde}} \end{cases}$$

Con:

$\Theta_{eleg,v}$	Conjunto de viviendas seleccionadas clasificadas como elegibles dentro del veintil $\boldsymbol{v}$.
$\Theta_{R,v}$	Conjunto de viviendas clasificadas como elegibles que responden la encuesta en el veintil $\boldsymbol{v}$.
W'_{hij}	Factor teórico ajustado por elegibilidad.
$\boldsymbol{v}$	Subíndice de la celda de ajuste para aplicar corrección por falta de respuesta (corresponde al veintil $\boldsymbol{v}$=1,...,20).
$\cdot$	Valor perdido o *missing value*.

En la **Tabla 14** se observa el promedio por veintil del factor ajustado por no respuesta, así como también, la cantidad de viviendas elegibles, la cantidad de viviendas que responden la encuesta y la estimación de la razón entre ambas.

Tabla 14: Ponderador de no respuesta en cada veintil

Veintiles	Ponderador de no respuesta promedio Whij_nr	Viviendas elegibles	Viviendas que responden (R)	Razón $(\hat{R}_v)$
Total general	237,86	26.882	14.961	
1	330,44	1.345	520	2,52
2	280,46	1.345	600	2,29
3	277,95	1.344	627	2,13
4	374,68	1.344	583	2,22
5	353,11	1.344	662	1,96
6	275,41	1.344	636	2,09
7	308,47	1.344	714	1,82
8	305,54	1.344	648	2,07
9	365,27	1.344	692	1,95
10	365,27	1.344	681	1,91
11	283,50	1.344	747	1,75
12	344,86	1.344	731	1,78
13	260,23	1.344	742	1,86
14	157,51	1.344	815	1,63
15	165,22	1.344	830	1,66
16	127,50	1.344	820	1,66
17	113,86	1.344	901	1,49
18	118,14	1.344	943	1,43
19	152,36	1.344	993	1,34
20	118,44	1.344	1.076	1,25

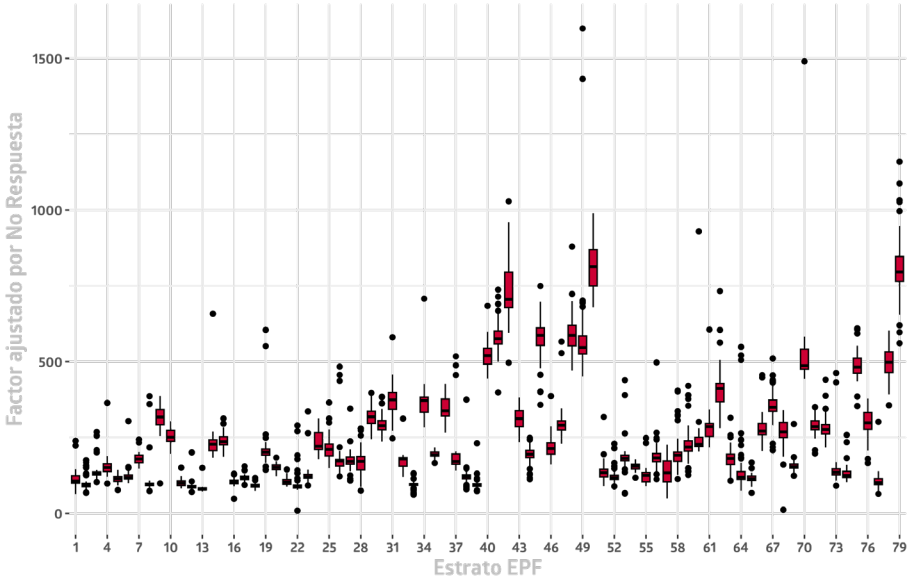
Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.4.9 Suavizamiento del factor de expansión

En la construcción del ponderador ajustado por no respuesta se podrían obtener factores con valores inusualmente alejados de la media (datos atípicos) que, combinado con valores altos (o bajos) de alguna variable de interés, podrían afectar las estimaciones de los parámetros. Por esta razón en las encuestas a hogares normalmente se utiliza un método para aminorar este efecto. Dicho método se denomina "suavizamiento"[47].

En el **Gráfico 2** se observa el comportamiento de los factores de expansión a nivel de estrato donde es posible visualizar una cierta cantidad de valores atípicos:

Gráfico 2: Ponderador ajustado por no respuesta al interior de cada estrato de muestreo sin suavizamiento



Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Según el **Gráfico 2**, la presencia de datos atípicos sugiere realizar suavizamiento para obtener valores más cercanos a la media y disminuir la dispersión al interior de cada estrato. El suavizamiento realizado al factor de expansión ajustado por no respuesta para la IX EPF, corresponde a un método mixto (MM) conformado por el método de contracción a la media (CM) y el método R-K **(ver Anexo 1)**.

[47] https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/documentos-de-trabajo/documentos-de-trabajo/impacto-covid-19-ene.pdf?sfvrsn=1645cfb0_6

5.4.9.1
Método de contracción a la media (CM)

Este método contrae los factores de expansión, acercándolos a su valor promedio dentro de cada estrato $\boldsymbol{h}$, haciendo que la distribución de los factores sea más homogénea. El factor suavizado o contraído a la media (o promedio) viene dado por la ecuación (7):

$$w_i^S = \beta \cdot w_i + (1-\beta) \cdot \bar{w}; 0 \leq \beta \leq 1 \tag{7}$$

Donde β es un parámetro real que permite controlar la intensidad de la contracción, $\boldsymbol{w_i}$ es el $\boldsymbol{i\text{-}ésimo}$ factor de expansión y $\overline{\boldsymbol{w}}$ el promedio de los factores a nivel de estrato basado en la existencia de valores atípicos o no.

5.4.9.2
Método R-K

Este método se caracteriza por dar tratamiento diferenciado a los conjuntos de datos atípicos y no atípicos, generando un aumento proporcional en los factores no atípicos, mientras que origina una disminución proporcional de los factores atípicos. Según lo anterior, este método, por un lado, disminuye en una proporción $\boldsymbol{r}$ los valores de los factores de expansión atípicos[48] y, por otro, aumenta en una proporción $\boldsymbol{k}$ los valores de los factores de expansión no atípicos. La **Imagen 2** muestra dicho procedimiento.

Imagen 2: Factores de expansión W_i ordenados de mayor a menor

Factores		Aumento o disminución	
W_1, W_2, ⋮, W_n	Atípicos	$W_1 \cdot (1\text{-}r)$, $W_2 \cdot (1\text{-}r)$, ⋮, $W_n \cdot (1\text{-}r)$	$t_1 = \left(\sum_{i=1}^{n} W_i\right) - \left(\sum_{i=1}^{n} W_i\right) \cdot r$
W_1, W_2, ⋮, W_m	No Atípicos	$W_1 \cdot (1+k)$, $W_2 \cdot (1+k)$, ⋮, $W_m \cdot (1+k)$	$t_2 = \left(\sum_{j=1}^{m} W_j\right) - \left(\sum_{j=1}^{m} W_j\right) \cdot k$

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

[48] En total, se truncaron o suavizaron 790 valores atípicos, con ambos métodos.

5.4.9.3
Método mixto

Para aplicar el método mixto se separan los factores de expansión en dos subconjuntos, factores no atípicos y factores atípicos. Para el conjunto de los factores no atípicos se aplica el método de contracción a la media considerando β relativamente cerca de la unidad (β=0,8). Para el conjunto de los factores atípicos, se considera β=0,2 para que, en caso de tener un solo valor atípico al interior de un estrato, el método CM no tenga efecto sobre este valor, conservando el valor del factor atípico. Posterior a este cálculo, se aplica el método R-K, en donde se juntan ambas distribuciones de factores (no atípicos y atípicos) considerando valores del multiplicador cercano a la unidad (λ=0,99) para conservar una pequeña distancia entre el mayor de los factores no atípicos, y el menor de los factores atípicos, después del suavizamiento.

Este es el último ajuste que se aplica a los factores de expansión a nivel de vivienda. La **Tabla 15** muestra los estadísticos asociados (recuento de viviendas, suma total y valores mínimos y máximos) para cada una de las etapas de ajuste al factor de expansión.

Tabla 15: Estadísticos asociados a cada etapa del factor de expansión

Etapas del Factor de expansión IX EPF	Suma	Recuento	Mínimo	Máximo	Mediana	Promedio	Desviación estándar
1. Teórico (Whij)	3.810.240	29.231	5,29	797,38	99,47	130,35	86,36
2. Ajustado por elegibilidad desconocida (W'hij)	3.810.240	28.993	5,32	818,58	99,71	131,42	87,23
3. Ajustado por no elegibilidad (W'hij)	3.558.582	26.882	5,32	818,58	100,44	132,38	87,85
4. Ajustado por no respuesta (Whij_nr)	3.558.582	14.961	8,93	1.598,75	175,08	237,86	176,62
5. Ajustado por no respuesta suavizado (W_is)	3.558.582	14.961	67,07	1.490,22	175,45	237,86	174,43
6. Ajustado por no respuesta suavizado (W_is_rk)	3.558.582	14.961	67,07	1.057,38	175,58	237,86	172,39

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.4.10 Calibración de los factores de expansión

En las etapas previas de este documento se consignó un factor de expansión referido a la probabilidad de seleccionar una vivienda al interior de cada estrato. No obstante, a razón de contar con estimaciones de la población actualizadas, se utilizan *stocks* poblacionales provenientes de fuentes externas, como lo son las Proyecciones de Población a la fecha correspondiente a la mitad del período de recolección de datos de la encuesta.

La calibración es un ajuste que se realiza a los factores de expansión a nivel de personas. Para esto, se utiliza el método *Raking*[49] que permite controlar por diversas variables simultáneamente y mantener el diseño muestral original de la encuesta.

5.4.11 Factor de expansión de personas

Para obtener un factor de expansión para cada una de las personas que participan en la IX EPF, se asigna a cada una de ellas el factor de expansión suavizado correspondiente a la vivienda a la que pertenece. La suma de dicho factor a nivel nacional alcanza 10.579.690 personas, cifra estimada antes de realizar el proceso de calibración que, luego de este procedimiento, debe coincidir con el *stock* de personas totales provenientes de las proyecciones de población definidas para el período de recolección de datos de la encuesta.

5.4.12 Calibración a proyecciones de población para el total de personas

Como se mencionó en el capítulo precedente, en el proceso de calibración se utiliza información auxiliar proveniente de las proyecciones de población, según variables o marginales de calibración de interés. El trabajo de recolección de datos de la IX EPF ocurre entre los meses de octubre de 2021 y septiembre de 2022, por lo tanto, la fecha de referencia de las proyecciones corresponde al 30 de marzo de 2022 (mitad del período de recolección de datos). La información auxiliar se utiliza como un vector de totales poblacionales, llamados marginales de calibración. Una vez suavizado el factor, se procede a revisar los coeficientes de variación de las marginales a utilizar, donde se espera obtener valores menores a 5% para todas sus categorías. Se revisaron las siguientes marginales:

- **A nivel regional:** Población total por región.
- **A nivel de sexo :** Población total por sexo.
- **A nivel de edad:** Población total según 3 tramos etarios: [0-14], [15-59] y [60+].
- **A nivel de región y sexo :** Población total por región y sexo.

[49] https://www.ine.gob.cl/docs/default-source/documentos-de-trabajo/nueva-metodolog%C3%ADa-de-calibraci%C3%B3n-de-la-ene-m%C3%A9todo-de-calibraci%C3%B3n-raking.pdf?sfvrsn=bfed7e52_2

Luego, se concluye que las marginales a utilizar en la calibración son:

- **A nivel de edad:** Población total según 3 tramos etarios: [0-14], [15-59] y [60+].

- **A nivel de región y sexo:** Población total por región y sexo.

Una vez calibrados los factores de expansión se obtiene una estimación de 13.063.825 personas a nivel nacional, cifra coincidente con las proyecciones de población.

5.5 Estimadores y varianza de los estimadores

Los estimadores son herramientas estadísticas que permiten hacer inferencias sobre características de la población basándose en los datos de la muestra. En la EPF, los estimadores son esenciales para generar resultados representativos y generalizables sobre los patrones económicos de los hogares, extrayendo conclusiones que son aplicables a toda la población.

5.5.1 Estimador del total de gastos e ingresos de los hogares

La estimación del total del gasto (o ingreso) de los hogares se obtiene multiplicando el valor del gasto (o ingreso) en cada hogar por su factor de expansión, y luego sumando sobre todos los hogares de la muestra por nivel de estimación. Esto es:

(8)

$$\widehat{G}_{c}^{n_e} = \sum_{h}\sum_{i}\sum_{j}\sum_{l} W_{hijl}^{cal} \cdot g_{c\,hijl} \quad \forall\, l \in \{n_e\}$$

Donde:

$\widehat{G}_{c}^{n_e}$	Es el gasto total de los hogares en el nivel de estimación n_e en la categoría de gasto tipo c.
$g_{c\,hijl}$	Es el valor del gasto tipo c medido en el hogar l de la vivienda j de la UPM i perteneciente al estrato h.
W_{hijl}^{cal}	Es el factor de expansión calibrado del hogar l de la vivienda j de la UPM i perteneciente al estrato h.
$\{n_e\}$	Es el conjunto de todos los hogares que pertenecen al nivel de estimación n_e.
C	Corresponde a la categoría del gasto.

5.5.2 Estimador de gastos o ingresos promedio de los hogares

La estimación del gasto (o ingreso) promedio por hogar en la categoría C, en el nivel de estimación n_e se obtiene dividiendo la estimación del gasto o ingreso total en la categoría de gasto C por la estimación de hogares en el nivel de estimación n_e, considerando incluso aquellos hogares que no incurren en gastos (o ingresos) de tipo C. Esto es:

(9)
$$\hat{\bar{G}}_{c}^{n_e} = \frac{\hat{G}_{c}^{n_e}}{\hat{H}^{n_e}}$$

Siendo:

(10)
$$\hat{H}^{n_e} = \sum_{h}\sum_{i}\sum_{j}\sum_{l} W_{hijl}^{cal} \cdot hog_{hijl}^{ne} \qquad \forall\, l \in \{n_e\}$$

Donde:

$\hat{H}^{n_e}$ Es la estimación del total de hogares en el nivel de estimación n_e

hog_{hijl}^{ne} Variable dicotómica (tipo *dummy*) asociada al hogar l de la vivienda j de la UPM i perteneciente al estrato h en el nivel de estimación n_e. Se define de la siguiente forma:

$$hog_{hijl}^{ne} = \begin{Bmatrix} 1 & \forall\ l\ \in \{n_e\} \\ 0 & \forall\ l\ \notin \{n_e\} \end{Bmatrix}$$

Alternativamente, se puede calcular el promedio de gasto por hogar considerando solo a los hogares que realizan el gasto tipo C en el nivel de estimación n_e, en cuyo caso, el denominador de la ecuación (9) se expresa como:

$$\hat{H}_{c}^{n_e} = \sum_{h}\sum_{i}\sum_{j}\sum_{l} W_{hijl}^{cal} \cdot hog_{c\,hijl}^{ne} \forall \quad l\ \in \{n_e\}$$

Donde:

$\widehat{H}^{n_e}$ Es la estimación del total de hogares que realizan gasto tipo C en el nivel de estimación n_e.

$hog^{ne}_{c\,hijl}$ Variable *dummy* asociada al hogar l de la vivienda j de la UPM i perteneciente al estrato h en el nivel de estimación n_e que realiza gasto tipo C. Se define de la siguiente manera:

$$hog^{ne}_{hijl} = \begin{cases} 1 & \forall\ l\ \in \{n_e \cap c\} \\ 0 & \forall\ l\ \notin \{n_e \cap c\} \end{cases}$$

5.5.3 Estimador de proporciones

La estimación de la proporción del gasto (o ingreso) por hogar en la categoría C en relación con el gasto (o ingreso) total por hogar se obtiene dividiendo la estimación del gasto (o ingreso) en la categoría C, por la estimación del gasto (o ingreso) total en el nivel de estimación n_e. Su forma de cálculo se observa en la ecuación (11):

(11)
$$\widehat{\tilde{G}}^{n_e}_{c} = \frac{\widehat{G}^{n_e}_{c}}{\widehat{G}^{n_e}}$$

Siendo:

$$\widehat{G}^{n_e} = \sum_h \sum_i \sum_j \sum_l W^{cal}_{hijl} \cdot g_{hijl} \qquad \forall\ l\ \in \{n_e\}$$

y:

$\widehat{G}^{n_e}$ Es la estimación del gasto (o ingreso) total en el nivel de estimación n_e.

g_{hijl} Corresponde al valor del gasto (o ingreso) medido en el hogar l de la vivienda j de la UPM i perteneciente al estrato h.

5.5.4 Estimador de gastos o ingresos *per cápita* de los hogares

La estimación del gasto (o ingreso) *per cápita* de los hogares se obtiene multiplicando el valor del gasto (o ingreso) *per cápita* de cada hogar por su factor de expansión, y luego sumando sobre todos los hogares de la muestra en el nivel de estimación n_e Su forma de cálculo se observa en la ecuación (12):

(12) $$\widehat{G}_{c}^{n_e PC} = \sum_{h}\sum_{i}\sum_{j}\sum_{l} W_{hijl}^{cal} \cdot g_{c\,hijl}^{ne^{PC}} \qquad \forall \quad l \in \{n_e\}$$

Donde:

$\widehat{G}_{c}^{n_e PC}$ Es la estimación del gasto total (o ingreso) *per cápita* en la categoría c en el nivel de estimación n_e.

$g_{c\,hijl}^{ne^{PC}}$ Es el gasto *per cápita* en la categoría c medido en el hogar l de la vivienda j de la UPM i perteneciente al estrato h en el nivel de estimación n_e, que se calcula dividiendo el gasto en la categoría c medido en el hogar por la cantidad total de personas del hogar.

5.5.5 Estimador de gastos e ingresos promedio *per cápita* de los hogares

La estimación del gasto promedio *per cápita* por hogar se obtiene dividiendo la estimación del gasto (o ingreso) total *per cápita* en la categoría c en el nivel de estimación n_e, por la estimación de hogares en el nivel de estimación n_e considerando, incluso aquellos que no incurren en gasto en la categoría c. Esto es:

(13) $$\widehat{\widehat{G}}_{c}^{n_e PC} = \frac{\widehat{G}_{c}^{n_e PC}}{\widehat{H}^{n_e}}$$

Al igual que en el caso anterior, en el numerador, se obtiene el promedio de gastos por hogar considerando solo a los hogares que realizan el gasto en la categoría c en el nivel de estimación n_e:

(14) $$\widehat{\widehat{G}}_{c}^{n_e PC} = \frac{\widehat{G}_{c}^{n_e PC}}{\widehat{H}_{c}^{n_e}}$$

5.5.6 Medidas de precisión

Actualmente los objetivos analíticos de las encuestas van mucho más allá de la generación de cuadros de resumen básicos o totales de los parámetros de interés. Cada vez es más frecuente que, además de obtener estimaciones de los parámetros, se requiera plantear hipótesis, comprobar o construir modelos y, hasta realizar evaluaciones de impacto. Es por esto que cobra relevancia el hecho de que los resultados incluyan medidas adecuadas de precisión o exactitud de las estimaciones derivadas de los datos de las encuestas.

Una de las medidas claves de precisión en las encuestas por muestreo es la estimación de la varianza, un indicador de la variabilidad que se produce al elegir una muestra, en lugar de confeccionar una lista de toda la población. Otras formas de medir el error de muestreo son a través del error estándar, el coeficiente de variación y el efecto de diseño (ONU, 2009); sin embargo, cabe mencionar que guardan entre sí relación algebraica, por lo que es posible deducir la expresión de cualquiera de ellas a partir de las otras, mediante la aplicación de operaciones matemáticas.

5.5.7 Varianza del estimador

Para el cálculo de la varianza de las estimaciones de interés en la IX EPF se utiliza el factor de expansión calibrado w_{hij}^{cal}, es decir, el factor que incluye todos los ajustes realizados a los ponderadores de selección (descritos en el apartado anterior). Los *softwares* estadísticos como SAS, SPSS, Stata y R actualmente poseen módulos para el cálculo de estimaciones, varianzas y coeficientes de variación, incluyendo los efectos de diseño cuando se trata de diseños muestrales complejo. Sin embargo, para una correcta estimación de la varianza asociada a una estimación de algún parámetro de interés, es necesario cerciorarse de que los estratos de muestreo y los conglomerados cumplan ciertas características. Este tema es abordado en apartado 5.5.9.

5.5.8 Estimación de la varianza

- ***Varianza del total estimado del gasto o ingreso en la categoría de gasto c***

La varianza de un estimador de totales se obtiene mediante un estimador de razón:

$$(\hat{\mathrm{R}}_{ch}^{n_e} = \frac{\hat{\mathrm{G}}_{ch}^{n_e}}{\hat{\mathrm{H}}_{h}^{n_e}})$$

sumando todos los hogares que realizan el gasto (o ingreso) en la categoría c, en el estrato h y área de estimación n_e. La forma de cálculo se observa en la ecuación (15):

(15)

$$\widehat{V}(\widehat{G}_{c}^{n_e}) = \sum_{i} \frac{n^{n_e}}{(n^{n_e}-1)} \cdot [W_{hij}^{cal} \cdot (g_{c\,hi}^{n_e} - \widehat{R}_{c\,h}^{n_e} \cdot hog_{hi}^{n_e})]^2, \qquad \forall\ i\ \in \{n_e\}$$

Donde:

$\widehat{G}_{c}^{n_e}$ Es la estimación del gasto (o ingreso) en la categoría $\boldsymbol{c}$ en el nivel de estimación $\boldsymbol{n_e}$.

$(\hat{\mathrm{R}}_{ch}^{n_e} = \frac{\hat{\mathrm{G}}_{ch}^{n_e}}{\hat{\mathrm{H}}_{h}^{n_e}})$ Es la razón entre la estimación del gasto en la categoría $\boldsymbol{c}$ y la estimación de hogares en el nivel de estimación $\boldsymbol{n_e}$, calculado a partir de la agregación de los estratos $\boldsymbol{h}$.

$hog_{hi}^{n_e}$ Es el número de hogares en el nivel de estimación $\boldsymbol{n_e}$, calculado a partir de la agregación de las UPM $\boldsymbol{i}$ pertenecientes a los estratos $\boldsymbol{h}$.

- ***Varianza del promedio estimado del gasto o ingreso en la categoría de gasto c***

Considerando el promedio estimado del gasto como la razón entre el gasto total estimado en la categoría de gasto en la categoría $\boldsymbol{c}$ del estrato $\boldsymbol{h}$ en el nivel de estimación $\boldsymbol{n_e}$, dividido por el número total de hogares estimado en el estrato $\boldsymbol{h}$ en el nivel de estimación $\boldsymbol{n_e}$, como se observa en la ecuación (16):

(16)

$$\hat{\mathrm{R}}_{ch}^{n_e} = \frac{\hat{\mathrm{G}}_{ch}^{n_e}}{\hat{\mathrm{H}}_{h}^{n_e}}$$

La varianza de la razón para cada nivel de estimación $\boldsymbol{n_e}$ se obtiene a partir de la ecuación (17):

(17)

$$\widehat{V}(\widehat{R}_{ch}^{n_e}) = \sum_{i} \frac{1}{(\widehat{H}_{h}^{n_e})^2} \frac{n^{n_e}}{(n^{n_e}-1)} \cdot [W_{hij}^{cal} \cdot (g_{c\,hi}^{n_e} - \widehat{R}_{c\,hi}^{n_e} \cdot hog_{hi}^{n_e})]^2, \qquad \forall\ i\ \in \{n_e\}$$

- ***Varianza de la proporción estimada del gasto o ingreso en la categoría de gasto en la categoría c respecto del total estimado del gasto (o ingreso) total***

Considerando la proporción estimada del gasto como la razón (o tasa) entre el gasto estimado en la categoría de gasto en la categoría $\boldsymbol{c}$ en el nivel de estimación $\boldsymbol{n_e}$ dividido por el gasto total estimado:

$$(\hat{R}_{ch}^{n_e\ prop} = \frac{\hat{\mathrm{G}}_{ch}^{n_e}}{\hat{\mathrm{H}}_{h}^{n_e}})$$

La varianza de la razón se obtiene a partir de la ecuación (18):

(18)

$$\widehat{V}(\widehat{R}_{ch}^{n_e prop}) = \sum_i \frac{1}{(\widehat{H}_h^{n_e})^2} . \frac{n^{n_e}}{(n^{n_e}-1)} . [W_{hij}^{cal} . (g_{c\,hi}^{n_e} - \widehat{R}_{c\,hi}^{n_e prop} . hog_{hi}^{n_e})]^2, \qquad \forall\ i\ \in \{n_e\}$$

- ***Coeficiente de variación e intervalos de confianza***

El coeficiente de variación es una medida de la variabilidad relativa que se expresa como el cociente entre el error estándar $\hat{\sigma}_{\hat{\theta}}$ y la estimación $\hat{\theta}$ según la ecuación (19):

(19)

$$CV_{\hat{\theta}} = CV(\hat{\theta}) = \frac{\widehat{\sigma}_{\hat{\theta}}}{\hat{\theta}} = \frac{\sqrt{\widehat{V}(\hat{\theta})}}{\hat{\theta}}$$

Donde:

$\hat{\theta}$	Es la estimación del parámetro de interés.
$\hat{\sigma}_{\hat{\theta}}$	Es el error estándar de la estimación del parámetro de interés.
$\hat{V}(\hat{\theta})$	Es la varianza de la estimación del parámetro de interés.

En la inferencia estadística, se quiere estimar parámetros poblacionales a partir de datos de muestras observadas.

El **intervalo de confianza** es un rango de valores determinados, a partir de los valores observados en la muestra. Sus límites, inferior y superior, se determinan de forma tal que la probabilidad de que el parámetro poblacional se ubique entre ellos sea de ***(1–α)%***. La probabilidad asociada al intervalo es también denominada como "nivel de confianza" del intervalo.

El intervalo de confianza se determina mediante la ecuación (20):

(20)

$$\theta \in [\hat{\theta} \pm t^{v}_{1-\alpha/2}]_{1-\alpha} \Longleftrightarrow [\hat{\theta} - t^{v}_{1-\alpha/2} \cdot \widehat{\sigma}_{\hat{\theta}} \,;\, \hat{\theta} + t^{v}_{1-\alpha/2} \cdot \widehat{\sigma}_{\hat{\theta}}\,]_{1-\alpha}$$

Siendo $t^{v}_{1-\alpha/2}$ el percentil ***(1–α/2)*** de la distribución t-Student con $\boldsymbol{v}$ grados de libertad[50].

Los grados de libertad ***(gl)*** se definen como el número de observaciones independientes que son libres de variar (número de individuos en los datos) menos el número de parámetros estimados (número de relaciones impuestas a los datos). Los ***gl*** constituyen uno de los elementos a considerar para poder emplear las distribuciones de frecuencia en las labores de inferencia. Más aún, dan cuenta de la precisión con la que se están estimando los parámetros a través de la muestra, ocurriendo que, mayores ***gl*** conducen a estimaciones más precisas. El recuento del tamaño muestral se realiza dependiendo del tipo de estimador. Para proporciones y razones entre 0 y 1, se realiza el conteo sobre las unidades de análisis del denominador; mientras que para el resto de las estimaciones se hace sobre el total de unidades de análisis que nutren la

[50] En la práctica, es dificultoso determinar en forma precisa los grados de libertad para la estimación de la varianza bajo una muestra compleja; sin embargo, actualmente los paquetes estadísticos para el análisis de datos bajo muestras complejas emplean la regla de cálculo de grados de libertad fijo, o como lo llaman Vaillant, Dever & Kreuter (2013) "la regla de oro", que consiste en obtener los grados de libertad mediante la resta entre el total de UPM menos el total estratos en el nivel de estimación requerido.

estimación[51]. El cálculo de los ***gl***, en el contexto de muestra compleja, se obtiene al descontar el total de estratos con observaciones en la subpoblación evaluada del total de UPM con observaciones en la subpoblación evaluada[52].

El **error absoluto** de la estimación corresponde a la máxima diferencia que se puede esperar entre el parámetro estimado y el verdadero parámetro poblacional con una probabilidad de $(1-\alpha)$ • 100, y se obtiene mediante la ecuación (21):

$$E_{abs}(\hat{\theta}) = t^{v}_{1-\alpha/2} \cdot \widehat{\sigma}_{\hat{\theta}} \tag{21}$$

El **error relativo** corresponde a la máxima variación porcentual que se puede esperar entre el parámetro estimado y el verdadero parámetro poblacional con una probabilidad del $(1-\alpha)$ • 100%, y se obtiene por el cociente entre al error absoluto dividido por la misma estimación según la ecuación (22):

(22)

$$E_{rel}(\hat{\theta}) = t^{v}_{1-\alpha/2} \cdot \widehat{\sigma}_{\hat{\theta}}/\hat{\theta} = t^{v}_{1-\alpha/2} \cdot CV_{\hat{\theta}} \; = \; E_{abs}(\hat{\theta})/\hat{\theta}$$

Dados los tamaños muestrales de la IX EPF (considerando como unidad de análisis a los hogares) el que una estimación sea o no aceptable dependerá del coeficiente de variación obtenido y de la desagregación del nivel de estimación, ya que al analizar un subgrupo de la muestra[53] se debe considerar la menor cantidad de observaciones disponibles. Por tanto, es relevante considerar la cantidad de unidades con las que se busca realizar alguna inferencia estadística.

[51] Departamento de Metodología e innovación Estadística (2020). Estándar para la evaluación de la calidad de las estimaciones en encuesta de hogares.

[52] Al igual que en el caso del tamaño muestral, dichos conteos se realizan dependiendo del tipo de estimador. Para proporciones y razones entre 0 y 1, se realiza el conteo sobre las unidades de análisis del denominador; mientras que para el resto de las estimaciones se forman sobre el total de unidades de análisis que nutren la estimación.

[53] Por ejemplo, niveles funcionales de la clasificación de productos o cruces de características de los integrantes del hogar.

Para interpretar el error muestral, usualmente se selecciona el coeficiente de variación asociado a una estimación, para lo cual se recomienda considerar, a modo de referencia, los siguientes rangos:

- ***cv*** menor o igual a 15% indica que la estimación es fiable.
- ***cv*** mayor a 15% y menor o igual a 30% indica que la estimación es poco confiable, por lo que debe ocuparse con precaución.
- ***cv*** mayor a 30%, indica una estimación no fiable, por lo que no se recomienda su uso[54].

Aun cuando el coeficiente de variación presente rangos adecuados en términos de precisión, las estimaciones podrían derivar en errores de interpretación por parte de personas usuarias de las bases de datos, debido a la disparidad de los tamaños muestrales (número de viviendas) que sustenta el cálculo de la estimación desagregada. Lo anterior podría implicar detectar estimaciones de gastos en productos con coeficientes de variación aceptables, pero calculadas con base en un número reducido de hogares (observaciones) lo que no permitiría dirimir si el valor obtenido en este estimador es o no fiable.

En cuanto al número de observaciones, según el *"Estándar para la evaluación de la calidad en las estimaciones de encuestas de hogares"* (INE Chile, 2020). se recomienda que las estimaciones evaluadas consideren, al menos, contar con 60 unidades para considerar fiable dicha estimación. En efecto, las estimaciones de parámetros de interés con una cantidad pequeña de observaciones presentan, en principio, una alta dispersión en la estimación de las medidas de error, lo que dificulta su interpretación como indicador de calidad[55].

[54] Para más información ver documentación referida: Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares.(INE Chile, 2020)

[55] La referencia del tutorial del paquete estadístico en el *software* R "Calidad" se encuentra en el siguiente link: https://inesscc.github.io/calidad/articles/tutorial.html.

[56] Para más información ver documentación referida: Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares. (INE Chile, 2020).

Por consiguiente, análisis conjunto del número de conglomerados con información de la variable de interés y del coeficiente de variación efectivo, al nivel de desagregación evaluado, debe ser considerado para definir el uso que se le puede dar a las estimaciones:

- Las estimaciones fiables corresponden a aquellas que cumplen con todos los criterios de calidad propuestos, por tanto, pueden ser reconocidas como estimaciones confiables y precisas de un parámetro poblacional.

- Las estimaciones poco fiables, son aquellas que cumplen los criterios del tamaño muestral y grados de libertad, pero que no se ajustan a los requerimientos establecidos para el ***ee*** o ***cv***, por lo cual, se recomienda utilizar esta estimación con precaución para interpretar características poblacionales, ya que podrían llevar a conclusiones poco acertadas. Más en detalle, pertenecen a esta categoría aquellas estimaciones evaluadas con el ***ee*** que no cumplan con el umbral de aceptación y también aquellas cuyo ***cv*** sea mayor a 15% y menor o igual a 30%.

- Las estimaciones no fiables corresponden a aquellas que no cumplen con alguno de los dos primeros criterios establecidos para el tamaño muestral o los ***gl***. También se incluyen dentro de esta categoría a aquellas estimaciones que no son proporciones ni razones entre 0 y 1 cuyo ***cv*** supere 30%. Para estos casos se debe consignar que no se recomienda el uso de estas estimaciones[56].

5.5.9 Variables que identifican el diseño muestral complejo

Como se mencionó en apartados precedentes, el diseño muestral de la IX EPF es probabilístico, estratificado y bietápico, es decir, un diseño muestral complejo.

En un diseño muestral estratificado, la estimación de la varianza total se obtiene mediante la suma de las varianzas estimadas en cada estrato y la varianza de cada estrato se estima a partir de la variabilidad de la característica de interés entre los conglomerados que los conforman.

En la IX EPF, la mayoría de los estratos muestrales están compuestos por la combinación de las variables Comuna y Clasificación Socioeconómica[57]. Por su parte, los conglomerados se identifican por medio de las UPM. En total, para el trabajo de recolección de datos se consideraron 79 estratos de muestreo conformados por 3.622 conglomerados.

En un diseño muestral estratificado para que la estimación de la varianza sea factible, se deben seleccionar y lograr como mínimo dos conglomerados en cada estrato muestral. Adicionalmente, para evitar estimaciones de varianza volátiles, se recomienda seleccionar un número adecuado de unidades muestrales de segunda etapa (viviendas) al interior de cada conglomerado.

En la práctica y, por razones ajenas al diseño, en ocasiones podría ocurrir que algún estrato muestral lograra respuesta en un solo conglomerado. Ante esta situación, se pueden realizar las siguientes acciones: (i) mantener esta información, lo que implica que no es posible estimar varianzas; (ii) asumir que estos conglomerados fueron seleccionados "con certeza", en cuyo caso los estratos no contribuyen a la estimación de la varianza total y se les asigna varianza 0, generando una subestimación

[57] También existen estratos que juntan comunas en niveles socioeconómicos, al interior de una misma región; así como también estratos conformados por una sola comuna y sus tres niveles socioeconómicos.

de la varianza; y (iii) crear unidades que agrupen estratos unitarios junto a otros estratos (denominadas pseudo-estratos de varianza, *varstrat* o pseudo-estratos) con el objetivo de contar con una estimación de la varianza más certera.

De acuerdo con lo anterior, y en busca del objetivo de aplicar los más altos estándares de calidad estadística, se opta por la última opción, con la finalidad de poner a disposición las variables que permitan identificar las características del diseño complejo de la encuesta y producir estimaciones de varianza consistentes y estables (Valliant, Dever, & Kreuter, 2013).

Adicionalmente, también es frecuente que los conglomerados no contengan un número estable de unidades requeridas para la estimación de la varianza, lo que hace necesaria la agrupación de varios conglomerados, al interior de un mismo estrato (original o pseudo-estrato). Esto se traduce en la creación de una nueva variable como resultado del agrupamiento de dichos conglomerados, denominada pseudo-conglomerados de varianza, *varunit* o pseudo-conglomerados.

En el contexto de la IX EPF, por diseño, cada estrato de muestreo tiene como mínimo 24 UPM seleccionadas y dentro de cada UPM originalmente, se seleccionaron 8 viviendas. Luego, por motivos de desactualización de marco donde, en algunas UPM no fue posible seleccionar las 8 viviendas requeridas, se decidió seleccionar, en algunos casos, el total de viviendas elegibles que contenían las UPM (en general consistió en 4 viviendas) y, aumentar en otras UPM el número de viviendas a seleccionar dentro del estrato de muestreo en cuestión, seleccionando 10, 12 o 16 viviendas.

Luego del trabajo de recolección de datos, es posible que se logren menos viviendas por UPM de las que se planificaron y más aún, que dentro de una UPM no se logre ninguna vivienda, lo que indicaría que se pierde una UPM completa al interior de un estrato de muestreo. El resultado del trabajo de recolección de la IX EPF indica que, por estrato de muestreo se lograron, al menos, 15 UPM, lo que indica que no es necesario colapsar estratos para formar *varstrat* o estratos de varianza[58].

Con respecto al logro al interior de las UPM (viviendas) se observa que existen 163 UPM donde sólo se logró 1 vivienda, lo que produce inestabilidad en la estimación de la varianza. En estos casos se decide colapsar las UPM unitarias con UPM cercanas geográficamente y dentro del mismo estrato de muestreo (original o pseudo-estrato) y submuestra operativa de recolección de datos, con algunas excepciones.

[58] Para la estimación de la varianza se requiere que cada estrato contenga, al menos, 2 UPM logradas en la muestra.

En el párrafo siguiente se describe la metodología para la generación de conglomerados de varianza (pseudo-conglomerados o *varunit*). Previo, se ordena la base por: Submuestra – Estrato muestral – Nivel Socioeconómico – Código de comuna – Orden UPM, y se identifica la presencia de UPM unitarias. En el caso de que existan UPM unitarias, se colapsan con otras UPM según los siguientes criterios:

1. UPM lo más cercana posible (dada por el vector Orden UPM) dentro de la agregación:

   Submuestra – Estrato muestral – Nivel Socioeconómico – Código de comuna

2. Si no es posible colapsarla con otra dentro de la misma comuna, se elige la UPM en la comuna más próxima geográficamente a la UPM unitaria según la siguiente agregación:

   Submuestra – Estrato muestral – Nivel Socioeconómico

3. Si no se puede dentro del mismo Nivel Socioeconómico, se elige la UPM más próxima a la UPM unitaria según la siguiente agregación:

   Submuestra – Estrato muestral

En cada una de estas etapas se debe tener consideración respecto a no sobrepasar el número original de viviendas a encuestar, dado por el diseño muestral y, si existen 2 UPM o más, igual de cercanas geográficamente, se elige la que tenga menor número de viviendas logradas.

La **Tabla 16** muestra la cantidad de UPM y pseudo-conglomerados de varianzas *(varunit)* al interior de cada estrato de muestreo.

Tabla 16: Total *Varunit* y Conglomerados según estrato

Región	Nro. Estrato	Comuna	Nivel Socioeconómico	Conglomerados	*Varunit*
Total				*3.555*	*3.399*
1	1	Alto Hospicio - Iquique	Bajo	28	28
2	2	Antofagasta	Bajo	48	47
3	3	Copiapó - Tierra Amarilla	Bajo	31	31
4	4	Coquimbo - La Serena	Bajo	49	47
5	5	Valparaíso	Bajo	50	46
5	6	Concón - Quilpué - Villa Alemana- Viña del Mar	Bajo	46	46
6	7	Graneros - Machalí- Olivar - Rancagua - Requínoa	Bajo	28	28
7	8	Maule - Talca	Bajo	34	34
8	9	Chiguayante- Concepción- Hualpén Hualqui - Penco - Talcahuano - Tomé	Bajo	35	33
8	10	Coronel - Lota - San Pedro de la Paz	Bajo	35	33
9	11	Padre las Casas - Temuco - Vilcún	Bajo	46	42
10	12	Puerto Montt - Puerto Varas	Bajo	42	41
11	13	Coyhaique	Bajo-medio	35	35
12	14	Punta Arenas	Bajo-medio-alto	31	31
13	15	Cerrillos - Cerro Navia- Estación Central- Lo Espejo - Lo Prado- Maipú- Padre Hurtado- Pedro Aguirre Cerda- Peñaflor- Pudahuel- Quinta Normal	Bajo	45	44
13	16	Huechuraba- Independencia - La Reina - Las Condes- Lo Barnechea- Macul- Ñuñoa- Peñalolén- Providencia- Recoleta- Santiago - Vitacura	Bajo	48	45
13	17	Conchalí - Lampa - Quilicura - Renca	Bajo	41	40

Región	Nro. Estrato	Comuna	Nivel Socioeconómico	Conglomerados	*Varunit*
13	18	La Cisterna - La Granja- San Joaquín- San Miguel- San Ramón	Bajo	47	46
13	19	La Florida - La Pintana - Puente Alto	Bajo	41	41
13	20	Calera de Tango- El Bosque - San Bernardo- Talagante	Bajo	45	41
14	21	Valdivia	Bajo -Alto	38	38
15	22	Arica	Bajo	40	40
16	23	Chillán - Chillán Viejo	Bajo- Alto	41	41
1	24	Alto Hospicio - Iquique	Medio	30	30
2	25	Antofagasta	Medio	51	50
3	26	Copiapó - Tierra Amarilla	Medio-Alto	38	37
4	27	La Serena	Medio	52	51
4	28	Coquimbo	Medio	47	46
5	29	Concón - Valparaíso - Viña del Mar	Medio	89	84
6	30	Quilpué - Villa Alemana	Medio	55	54
6	31	Rancagua	Medio	38	36
6	32	Graneros - Machalí - Olivar - Requínoa	Medio	27	26
7	33	Maule - Talca	Medio	93	92
8	34	Coronel - Concepción	Medio	37	35
8	35	Chiguayante - Hualqui- Lota - Talcahuano	Medio	29	25
8	36	Penco - San Pedro de la Paz - Talcahuano	Medio	36	33
8	37	Hualpén - Tomé	Medio	35	30
9	38	Padre las Casas - Temuco - Vilcún	Medio	104	101
10	39	Puerto Montt - Puerto Varas	Medio	88	87

Región	Nro. Estrato	Comuna	Nivel Socioeconómico	Conglomerados	*Varunit*
13	40	Cerrillos - Cerro Navia- Estación Central - Lo Prado	Medio	42	41
13	41	Lo Espejo - Pedro Aguirre Cerda - Pudahuel - Quinta Normal	Medio	45	41
13	42	Maipú - Padre Hurtado- Peñaflor	Medio	45	42
13	43	La Reina - Las Condes- Lo Barnechea - Macul- Ñuñoa - Providencia- Vitacura	Medio	53	45
13	44	Peñalolén	Medio	48	45
13	45	Conchalí - Huechuraba- Lampa - Quilicura - Renca	Medio	45	43
13	46	Independencia - Recoleta	Medio	49	46
13	47	La Cisterna - San Joaquín - San Miguel	Medio	43	42
13	48	La Florida - La Granja - San Ramón	Medio	45	45
13	49	Calera de Tango- El Bosque - San Bernardo- Talagante	Medio	44	44
13	50	La Pintana - Puente Alto	Medio	41	41
13	51	Santiago	Medio	49	45
14	52	Valdivia	Medio	44	44
15	53	Arica	Medio-Alto	40	40
16	54	Chillán - Chillán Viejo	Medio	41	41
1	55	Alto Hospicio - Iquique	Alto	30	29
2	56	Antofagasta	Alto	47	45
4	57	Coquimbo - La Serena	Alto	43	40
5	58	Quilpué - Valparaíso - Villa Alemana	Alto	48	48
5	59	Concón - Viña del Mar	Alto	59	53

Región	Nro. Estrato	Comuna	Nivel Socioeconómico	Conglomerados	*Varunit*
6	60	Graneros - Machalí - Olivar - Requínoa	Alto	27	24
7	61	Maule - Talca	Alto	15	14
8	62	Chiguayante - Concepción - Hualpén- Hualqui - Penco - Talcahuano - Tomé	Alto	35	31
8	63	Coronel - Lota - San Pedro de la Paz	Alto	34	32
9	64	Padre las Casas - Temuco - Vilcún	Alto	63	60
10	65	Puerto Montt - Puerto Varas	Alto	45	45
13	66	La Cisterna - La Granja- San Joaquín - San Miguel - San Ramón	Alto	44	40
13	67	El Bosque - La Florida- La Pintana - San Bernardo	Alto	46	45
13	68	Calera de Tango - Puente Alto- Talagante	Alto	41	40
13	69	Cerrillos - Cerro Navia- Estación Central - Lo Espejo - Lo Prado - Pedro Aguirre Cerda	Alto	45	44
13	70	Maipú - Padre Hurtado- Peñaflor - Pudahuel - Quinta Normal	Alto	43	41
13	71	La Reina - Lo Barnechea - Peñalolén	Alto	50	47
13	72	Conchalí - Huechuraba- Lampa - Quilicura - Renca	Alto	45	43
13	73	Independencia - Recoleta	Alto	50	49
13	74	Macul	Alto	43	40
13	75	Ñuñoa	Alto	43	40
13	76	Providencia	Alto	57	49
13	77	Vitacura	Alto	57	50
13	78	Las Condes	Alto	48	40
13	79	Santiago	Alto	50	50

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Tabla 17: Frecuencia del número de conglomerados que componen los pseudo-conglomerados de varianzas *(Varunit)*

Nro. Conglomerados	Frecuencia (n° de *Varunit*)	Porcentaje (%)	Frecuencia Acumulada	Porcentaje acumulado
1	3.245	91,4%	3245	91,4%
2	304	8,6%	3549	99,9%
3	2	0,06%	3551	100,00%

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.6 Clasificadores utilizados

Los clasificadores utilizados en la EPF contribuyen a la coherencia y la uniformidad en la recopilación, el análisis y la interpretación de los datos. Al utilizar códigos, términos y categorías específicas, el uso de clasificadores facilita la comparabilidad de los resultados a lo largo del tiempo y entre diferentes estudios y poblaciones.

Los clasificadores estadísticos en muchos casos son generados por organizaciones internacionales, algunas de las organizaciones son, Organización Internacional del Trabajo (OIT, 2012) y la Organización de las Naciones Unidas (ONU, 2018). La misión del INE es proveer la adaptación, actualización y divulgación de los clasificadores, es por ello que se crea un instrumento para establecer una clasificación nacional y facilitar comparaciones internacionales de las ocupaciones habituales de la población económicamente activa de Chile, permitiendo también presentar los datos de una manera clara y sencilla.

5.6.1 Clasificación Internacional Uniforme de Ocupaciones (CIUO-08.CL)

El Clasificador CIUO-08.CL es una adaptación nacional de la Clasificación Internacional Uniforme de Ocupaciones (OIT, 2007), tiene una estructura jerárquica, se compone de diez grandes grupos, y dentro de cada grupo hay subgrupos principales en las que se clasifican las ocupaciones, estos se establecen según similitud de tareas y especialización de las competencias requeridas. La codificación de ocupaciones de la IX EPF se realizó en base al clasificador CIUO-08.CL. En el caso particular de la EPF, la clasificación utilizada es a dos dígitos, por ejemplo, un enfermero, se codificó en el código 22, Profesionales de la Salud.

Tabla 18: Clasificador CIUO 08.CL (2 dígitos)

Gran grupo	Subgrupo Principal	Glosa gran grupo y subgrupos principales
1		Directores, gerentes y administradores
	11	Miembros del Poder Ejecutivo y Legislativo, personal directivo de la administración pública y de otras organizaciones sociales y/o Políticas, directores ejecutivos y gerentes generales.
	12	Directores, gerentes administrativos y de servicios comerciales.
	13	Directores, gerentes y administradores de producción y operaciones.
	14	Directores, gerentes y administradores de hoteles, restaurantes, comercios y otros servicios.
2		Profesionales, científicos e intelectuales
	21	Profesionales de las ciencias y de la ingeniería.
	22	Profesionales de la salud.
	23	Profesionales de la educación.
	24	Profesionales de los negocios y la administración.
	25	Profesionales de tecnología de la información y las comunicaciones.
	26	Profesionales de derecho, ciencias sociales y culturales.
3		Técnicos y profesionales de nivel medio
	31	Técnicos de las ciencias y la ingeniería.
	32	Técnicos de la salud.
	33	Técnicos en operaciones financieras y administrativas.
	34	Técnicos de servicios jurídicos, sociales, deportivos y culturales.
	35	Técnicos de la tecnología de la información y las comunicaciones.
	36	Técnicos en educación.
4		Personal de apoyo administrativo
	41	Oficinistas.
	42	Empleados en trato directo con el público.
	43	Auxiliares y ayudantes de registros contables y encargados del registro de materiales.
	44	Otro personal de apoyo administrativo.

Gran grupo	Subgrupo Principal	Glosa gran grupo y subgrupos principales
5		Trabajadores de los servicios y vendedores de comercios y mercados
	51	Trabajadores de los servicios a las personas.
	52	Vendedores.
	53	Trabajadores de los cuidados personales.
	54	Personal de los servicios de protección y seguridad.
6		Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros
	61	Agricultores y trabajadores calificados de explotaciones agropecuarias cuya producción se destina al mercado.
	62	Trabajadores forestales calificados, pescadores y cazadores cuya producción se destina al mercado.
	63	Trabajadores agropecuarios, pescadores, cazadores y recolectores de subsistencia.
7		Artesanos y operarios de oficios
	71	Operarios de la construcción (no incluye electricistas)
	72	Operarios de la metalurgia y operarios de máquinas de herramientas; mecánicos de vehículos, maquinarias, aviones y bicicletas.
	73	Artesanos y operarios de las artes gráficas.
	74	Trabajadores especializados en electricidad y electrónica.
	75	Operarios de procesamiento de alimentos, de la confección, ebanistas y otros oficios.
8		Operarios de instalaciones máquinas y ensambladores
	81	Operadores de instalaciones fijas y máquinas.
	82	Ensambladores.
	83	Conductores de vehículos y operadores de equipos pesados móviles.

Gran grupo	Subgrupo Principal	Glosa gran grupo y subgrupos principales
9		Ocupaciones elementales
	91	Auxiliares de aseo y trabajadores de casa particular.
	92	Obreros agropecuarios, pesqueros y forestales.
	93	Obreros de la minería, la construcción, la industria manufacturera y el transporte.
	94	Cocineros de comidas rápida y ayudantes de cocina.
	95	Trabajadores ambulantes de servicios y vendedores ambulantes (excluyendo comida de consumo inmediato).
	96	Recolectores de desechos y otras ocupaciones elementales.
0		Ocupaciones de las fuerzas armadas
	01	Oficiales de las fuerzas armadas.
	02	Suboficiales de las fuerzas armadas.
	03	Otros miembros de las fuerzas armadas.

Fuente: Instituto Nacional de Estadísticas (INE), CIUO-08.CL.

5.6.2 Clasificación del Consumo Individual por Finalidades (CCIF 2018.CL)

La CCIF es un sistema de clasificación funcional del Sistema de Cuentas Nacionales (SCN), el cual es utilizado a nivel internacional y se encuentra a cargo de las Naciones Unidas (ONU, 2018). La CCIF busca la asignación de códigos estandarizados en bienes y servicios de consumo final de los hogares, instituciones sin fines de lucro que sirven a los hogares (ISFLSH) y el Gobierno. Esta clasificación es utilizada mayormente por las encuestas de ingresos y gastos, Cuentas Nacionales y en el IPC.

El clasificador de la CCIF consta de 15 Divisiones, que se orientan a estructurar el consumo de bienes y servicios de los hogares en Chile. Para efectos de la EPF y su adaptación nacional, se consideraron las

Divisiones 01 a la 13, que van desde la División 1; Alimentos y bebidas no alcohólicas, hasta la división 13; Cuidado Personal, asistencia social, bienes y servicios diversos[59]. Esta adaptación da como resultado el clasificador CCIF 2018.CL (INE Chile, 2023d).

Tabla 19: Clasificador CCIF Divisiones

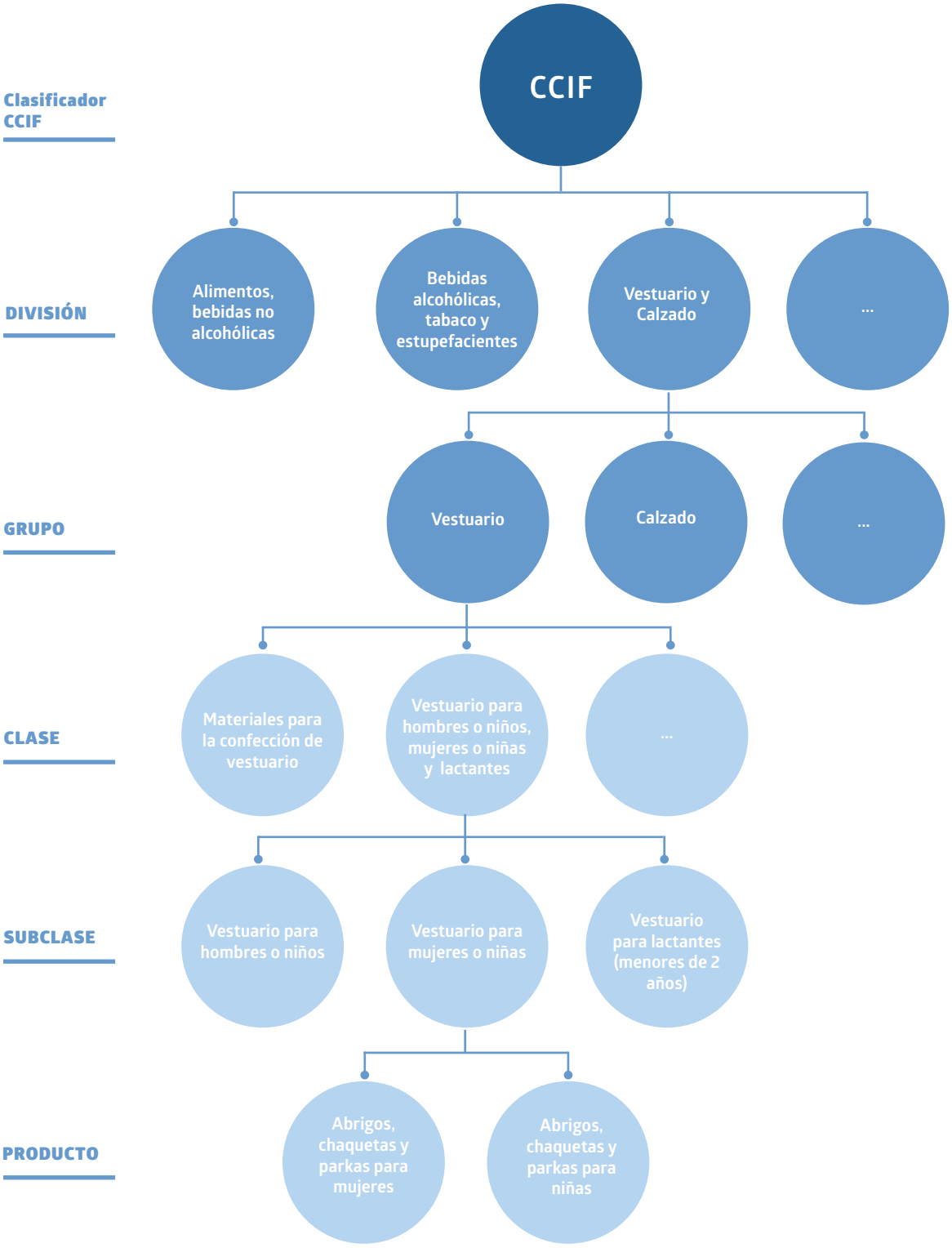
Divisiones	Glosa divisiones
División 01	Alimentos y bebidas no alcohólicas
División 02	Bebidas alcohólicas, tabaco y estupefacientes
División 03	Vestuario y calzado
División 04	Vivienda, agua, electricidad, gas y otros combustibles
División 05	Muebles, equipamiento para el hogar y mantenimiento rutinario de la vivienda
División 06	Salud
División 07	Transporte
División 08	Información y comunicación
División 09	Recreación, deportes y cultura
División 10	Servicios de educación
División 11	Servicios de restaurantes y alojamiento
División 12	Seguros y servicios financieros
División 13	Cuidado personal, asistencia social y bienes y servicios diversos

Fuente: Instituto Nacional de Estadísticas (INE), CCIF 2018.CL

El mayor nivel de agregación de la CCIF son las divisiones (13), las que se forman de grupos. Los grupos, a su vez, están formados por clases y estas se componen por subclases. Dichas subclases son el resultado de la agregación de los productos. Los niveles de división, grupo, clase y subclase permiten comparaciones internacionales. Por otra parte, el nivel de producto responden a la adaptación realizada por cada país. A continuación, se muestra un ejemplo del funcionamiento del clasificador con el siguiente diagrama.

[59] El motivo por el cual la EPF considera para su clasificación las primeras 13 de las 15 divisiones se debe a que las divisiones 14 y 15 clasifican el gasto en consumo individual de instituciones sin fines de lucro y del gobierno, respectivamente, lo que no es parte de los objetivos del estudio.

Diagrama 5: Niveles del clasificador CCIF 2018.CL



Fuente: Instituto Nacional de Estadísticas (INE), CCIF 2018.CL.

El clasificador distingue, además, a nivel de clase, entre bienes y servicios.

Adaptación del clasificador CCIF

En la CCIF de la VII y VIII EPF se utilizó una adaptación de la clasificación internacional CCIF 1999 (COICOP según su nombre en inglés *Classification of Individual Consumption According to Purpose*, de la Comisión de Estadísticas de las Naciones Unidas), la cual tuvo relativamente poca variación entre la VII y la VIII EPF. No obstante, para la IX EPF se utilizó la CCIF 2018.CL (INE Chile, 2023d), significando un cambio sustancial y un importante trabajo de adaptación.

Considerando ajustes en el clasificador, más las adaptaciones para la versión nacional, los mayores cambios entre la CCIF utilizada para la VIII versión y la IX EPF son:

- La creación de una División 13: En la VIII versión la División 12 de "Bienes y servicios diversos" engloba la actual División 12 de gastos en "Seguros y servicios financieros" y la División 13 de: "Cuidado personal, asistencia social, bienes y servicios diversos".

- En la División 6 de "Salud", se eliminó la diferenciación de codificación según dependencia administrativa del establecimiento de salud.
- Una gran cantidad de productos de tecnología e información fueron trasladados de la División 09 "Recreación y cultura" (VIII CCIF) a la División 08 (Información y Comunicación): entre este tipo de productos se encuentran los computadores y otras herramientas de información.
- En la División 10 de "Servicios de Educación" se eliminó la diferencia de clasificación del servicio según dependencia administrativa del establecimiento y el ciclo de enseñanza.

- La División 11 sobre "Servicios de Restaurantes y Alojamiento", en la VIII versión se diferenciaba entre los "servicios de restaurante por *delivery*" o para "consumo en el local", mientras que en la IX EPF la clasificación es entre Restaurantes con "servicio completo" o "servicio limitado".
- En esta nueva versión CCIF 2018.CL, los platos preparados se han trasladado de la División 11 a la División 01 (Alimentos y bebidas no alcohólicas) debido a la creciente disponibilidad de este tipo de comidas[60].

[60] Los platos preparados se venden, por lo general, en supermercados, donde la comida en ocasiones suele ser preparada o cocinada.

La codificación de bienes y servicios se realizó sobre registros provenientes de Gastos diarios (GD), Gastos del Recuerdo (GR), Gastos periódicos del Hogar (GH), de Ingresos por transferencias emitidas (IN), Gastos en Educación (GE) y del módulo de Gastos en Servicios Financieros y Seguros (GSF).

Los códigos residuales, si bien no están contemplados en la CCIF de Naciones Unidas, se han utilizado en la adaptación nacional de las versiones de la EPF. La mayor parte de estos códigos se incorporan a la transitoria[61] División 16 que responde a "Gastos no desglosados de diferentes divisiones y otros bienes y servicios no clasificados previamente" **(Tabla 20)**. Este criterio responde a una decisión operativa frente a casos de registros deficientes en las glosas a codificar, principalmente para productos no desglosados, por ejemplo: "tomate + lechuga + papel higiénico + cajetilla de cigarros". También se contemplaron códigos para gastos agregados correspondientes a divisiones distintas, pero que son afines o complementarias. Por último, si no corresponde a divisiones distintas y no se obtiene una clasificación residual de divisiones afines o complementarias, el gasto se asignó al código residual de cada clase.

Finalmente, los registros de los productos codificados en la División 16, se distribuyen según el peso de la división en el gasto total, salvo aquellos registros que, si bien fueron registrados en la encuesta, no corresponden a gasto de consumo final de los hogares y, por lo tanto, quedan fuera del marco conceptual del estudio.

[61] Es de importancia destacar que esta división 16 es utilizada solo de forma transitoria durante la fase de procesamiento. Por esta razón, no forma parte de la adaptación nacional del clasificador.

Tabla 20: Clasificador CCIF 2018.CL Divisiones Residuales

Divisiones	Código	Glosa divisiones
División 01	16.1.1.01.01	Gastos no desglosados en alimentos y bebidas no alcohólicas
División 02	16.1.1.01.02	Gastos no desglosados en bebidas alcohólicas, tabaco y estupefacientes
División 03	16.1.1.01.03	Gastos no desglosados en vestuario y calzado
División 04	16.1.1.01.04	Gastos no desglosados en vivienda, agua, electricidad, gas y otros combustibles
División 05	16.1.1.01.05	Gastos no desglosados en muebles, equipamiento para el hogar y mantenimiento rutinario de la vivienda
División 06	16.1.1.01.06	Gastos no desglosados en salud
División 07	16.1.1.01.07	Gastos no desglosados en transporte
División 08	16.1.1.01.08	Gastos no desglosados en información y comunicación
División 09	16.1.1.01.09	Gastos no desglosados en recreación, deportes y cultura
División 10	16.1.1.01.10	Gastos no desglosados en servicios de educación
División 11	16.1.1.01.11	Gastos no desglosados en servicios de restaurantes y alojamiento
División 12	16.1.1.01.12	Gastos no desglosados en seguros y servicios financieros
División 13	16.1.1.01.13	Gastos no desglosados cuidado personal, asistencia social y bienes y servicios diversos
División 1 y 2	16.1.1.01.14	Gastos no desglosados correspondientes a las divisiones 1 y 2
División 1 y 11	16.1.1.01.15	Gastos no desglosados correspondientes a las divisiones 1 y 11
División 1 y 5	16.1.1.01.16	Gastos no desglosados correspondientes a las divisiones 1 y 5
División 1 y 13	16.1.1.01.17	Gastos no desglosados correspondientes a las divisiones 1 y 13
División 2 y 11	16.1.1.01.18	Gastos no desglosados correspondientes a las divisiones 2 y 11
Divisiones 5 y 13	16.1.1.01.19	Gastos no desglosados correspondientes a las divisiones 5 y 13
Divisiones 8 y 9	16.1.1.01.20	Gastos no desglosados correspondientes a las divisiones 8 y 9
Cuotas por participación en actividades recreativas	16.1.1.01.21	Gastos no desglosados correspondientes a pagos de cuotas por participación en actividades recreativas (cuotas para: asados, fiestas, celebraciones, paseos, etc.)
Donaciones (*)	16.1.1.01.22	Gastos en donaciones a instituciones sin fines de lucro (bomberos, iglesias, fundaciones, etc.)

Divisiones	Código	Glosa divisiones
Limosnas (*)	16.1.1.01.23	Gastos en limosnas
Compra de divisas (*)	16.1.1.01.24	Cambio de dinero (insumo para calcular el costo)
Descuentos (*)	16.1.1.01.25	Descuentos en bienes y servicios n.c.p.
Automóvil dado en parte de pago (*)	16.1.1.01.26	Automóviles dados en parte de pago
Productos no desglosados de distintas divisiones n.c.p.	16.1.1.01.27	Gastos en bienes y servicios no desglosados n.c.p.
Bonificaciones en Salud (*)	16.1.1.01.28	Gastos en copagos, bonificaciones en salud
Propinas (**)	16.1.1.01.29	Gastos en propinas
División 4 y 5	16.1.1.01.30	Gastos no desglosados correspondientes a las divisiones 4 y 5
División 9 y 11	16.1.1.01.31	Gastos no desglosados correspondientes a las divisiones 9 y 11
Divisiones 3, 9, 11 y 13	16.1.1.01.32	Gastos no desglosados correspondientes a las divisiones 3, 9, 11 y 13
Divisiones 6 y 13	16.1.1.01.33	Gastos no desglosados correspondientes a las divisiones 6 y 13
Divisiones 1, 2, 5, y 9	16.1.1.01.34	Gastos no desglosados correspondientes a las divisiones 1, 2, 5, y 9

(*) Estos productos utilizados en la división residual no corresponden a gastos de consumo final de los hogares, por tanto, no son incorporados en las bases de datos publicadas por la encuesta.

(**) Las propinas posteriormente se codifican de acuerdo con gasto asociado en la División 11.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.6.3 Clasificador de establecimientos

El clasificador de establecimientos corresponde a un clasificador INE generado de forma ad hoc para la encuesta. El clasificador sistematiza diversos puntos de venta de bienes o servicios, considerando 29 categorías de establecimientos y donde son 3 residuales. El primer residual corresponde a "Otros establecimientos" y agrupa establecimientos que no se pudieron clasificar en las categorías anteriores, pero que son identificables en su tipo, mientras que el segundo "Establecimientos imposibles de identificar en su tipo" denota los establecimientos que no cuentan con suficiente información para ser clasificados, en esa misma dirección está el residual de salud "atención de salud pública imposible de identificar en su tipo" (código 16).

La codificación de establecimientos responde a un uso interno dentro de la encuesta, relacionado con obtener información respecto al tipo de establecimiento donde los hogares realizan sus gastos en bienes y servicios.

Tabla 21: Clasificador de Establecimientos

Código	Glosa teórica del grupo
01	Supermercados
02	Almacén tradicional
03	Ferias libres
04	Restaurantes y bares
05	Distribuidoras - mayoristas
06	Vegas - mercados
07	Comercio ambulante
08	Multitiendas
09	Ferreterías y multiferreterías
10	Farmacias
11	Clínicas
12	Centros médicos y consultas particulares
13	Atención pública de salud primaria
14	Atención pública de salud secundaria
15	Atención pública de salud terciaria
16	Atención de salud pública imposible de identificar en su tipo
17	Internet
18	Extranjero

Código	Glosa teórica del grupo
19	Tienda especializada [Indicar el tipo específico]
20	Feria artesanal- Feria de las pulgas
21	Kioscos
22	Máquina Autoservicio
23	Outlet- Bodegas
24	Servicio Profesional- Oficio Independiente
25	Tienda
26	Tienda multipropósito
27	Vendedor catálogo
88	Otros Establecimientos
99	Establecimiento Imposible de Identificar en su tipo

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

5.6.4 Clasificador de establecimientos educacionales (RBD)

El clasificador de establecimientos educacionales utilizado por la encuesta corresponde al listado de establecimientos de educación preescolar, básica y media reconocidos por el Ministerio de Educación.

En específico, para el caso de los establecimientos que imparten educación básica y media, todos ellos poseen un rol único de identificación denominado Rol Base de Datos (RBD), el cual se compone de un código de hasta 5 dígitos, junto con un dígito verificador del mismo. Este código es el símil al RUT en los establecimientos educacionales y se encuentra estandarizado a nivel país. Para la IX EPF se trabajó con el listado RBD del Ministerio de Educación del año 2020.

06.

METODOLOGÍA DE RECOLECCIÓN DE DATOS

06.

METODOLOGÍA DE RECOLECCIÓN DE DATOS

El presente capítulo corresponde a la Metodología de Recolección utilizada para la IX versión de la Encuesta de Presupuestos Familiares (EPF). El objetivo principal es describir los instrumentos empleados para capturar información y los lineamientos que guiaron la fase de recolección de la información requerida para el cumplimiento de los objetivos del estudio.

El diseño de la metodología de recolección se basó principalmente en la experiencia obtenida de sus versiones anteriores, la VII EPF (2011-2012) y VIII EPF (2016-2017), las que presentaron leves diferencias. Los ajustes realizados provienen de la observación de buenas prácticas internacionales en los procesos de recolección de datos, del estudio de literatura especializada, del intercambio de experiencias internacionales con otras encuestas de gastos e ingresos, del análisis de la no respuesta total de la VIII EPF, la retroalimentación entregada por el personal del trabajo de recolección de la anterior versión de la encuesta, a través de procesos de evaluación internos, y los aprendizajes logrados gracias a dos estudios piloto realizados en miras a la IX EPF, el primero realizado entre noviembre y diciembre del 2018 y el segundo entre octubre y diciembre de 2020.

El capítulo consta de seis secciones que dan cuenta de distintas decisiones metodológicas y organizacionales relativas a la recolección de datos. El capítulo inicia con la descripción de los instrumentos de recolección, especificando contenidos y períodos de referencia de la información solicitada. Posteriormente, se señalan las pruebas pilotos ejecutadas para evaluar los instrumentos y tomar decisiones respecto de sus contenidos y tecnología. En tercer lugar, se describen las características metodológicas adoptadas en el proceso, dando cuenta de técnicas de producción de información, tecnología para recopilar datos, flujo de proceso asociado a las etapas de ejecución de la recolección, supervisión y digitación de datos, además de los indicadores empleados para monitorear las etapas más críticas del proceso. Posteriormente, se describen aspectos de diseño

organizacional de los equipos involucrados, para luego explicar el método de capacitación de ese personal. Finalmente, se exponen las principales características y funcionalidades del sistema informático que dio soporte a las distintas etapas del proceso de recolección.

6.1 Diseño de instrumentos de captura de la información

La presente sección describe los instrumentos utilizados en la IX EPF para capturar la información de los hogares y recopilar paradatos del proceso de recolección. Específicamente, las siguientes secciones describen el contenido de cada instrumento.

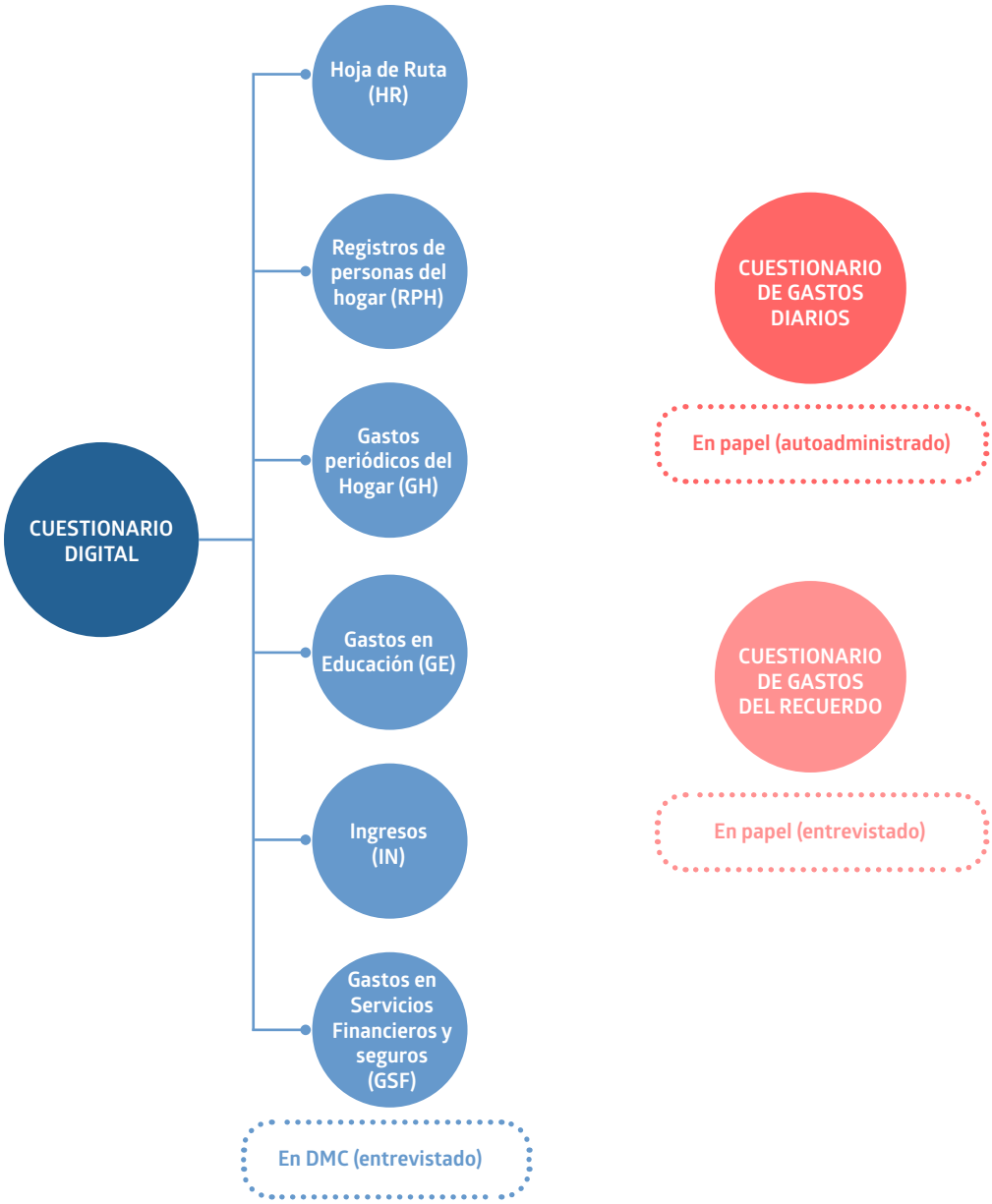
6.1.1 Instrumentos para la captura de información de los hogares

La IX EPF utilizó tres cuestionarios estructurados, compuestos por 996 preguntas[62], principalmente cerradas y pre-codificadas. Las preguntas abiertas refieren a montos y a descripciones de bienes o servicios adquiridos, de transferencias en especies recibidas y emitidas, además de las ocupaciones, todas las cuales son clasificadas y codificadas posteriormente, a partir de los clasificadores indicados en el apartado 5.6.

Debido a las dimensiones de la encuesta, los instrumentos de la EPF se agrupan tanto por la tecnología utilizada, como por la temática para recopilar información de los hogares. En el **Diagrama 6** los instrumentos se presentan según la tecnología, es decir, cuenta con un cuestionario digital implementado en DMC y dos cuestionarios en formato papel. En las siguientes subsecciones se describen en detalle cada uno.

[62] El total de preguntas aplicables en cada entrevista estuvo condicionado a las características del hogar y presencia de distintos tipos de gasto o ingreso. De las 750 preguntas, 537 contaban con al menos una condición habilitante, que debía cumplirse para ser desplegada en la entrevista.

Diagrama 6: Instrumentos de recolección de la IX EPF



Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Temáticamente, los instrumentos de captura de información de la IX EPF se clasificaron según los siguientes contenidos[63]: 1) un módulo que consulta sobre características sociodemográficas del hogar, 2) tres módulos que capturan gasto con el cuestionario digital más 2 cuestionarios que capturan gastos en el formato de cuestionario en papel y 3) un módulo que registra los ingresos del hogar[64]. Cada cuestionario y módulo que tiene como objetivo medir el gasto, se diferencian entre sí, fundamentalmente, por el tipo de bienes y servicios que incluye cada uno y por el período de referencia respecto al gasto que es consultado.

En los apartados siguientes se revisan los objetivos y características de cada módulo y cuestionario, junto con las poblaciones de referencia para los que se captura la información.

6.1.1.1
Registro de Personas en el Hogar (RPH)

El primer módulo de captura de información es el Registro de Personas en el Hogar (RPH). Este módulo busca caracterizar sociodemográficamente a las personas integrantes de los hogares, residentes en las capitales regionales y sus zonas conurbadas.

Estas características permiten no solo analizar los patrones de gasto y estructura del ingreso disponible de distintos tipos de hogares, sino que también posibilitan los ejercicios de imputación y del cálculo de factores de expansión.

La información de cada hogar se captura a través de las secciones que se describen a continuación, la mayoría de estas permite establecer enlaces o conexiones con el resto de los instrumentos:

[63] La Hoja de Ruta, integrada al cuestionario digital, no está descrita en esta sección debido a que su objetivo corresponde al de un instrumento de control y medición del esfuerzo operativo. Este módulo se describe en la siguiente sección 6.1.2.

[64] Cabe señalar que el módulo de Ingresos también registra una porción pequeña del gasto de los hogares. Sin embargo, no es mencionado entre los instrumentos que miden gasto, debido a que su objetivo fundamental es medir ingresos.

- **Características de las personas integrantes del hogar:**

Captura información demográfica de quienes integran el hogar y establece los lazos de parentesco que estas tienen con quien ejerce el rol de persona sustentadora principal del hogar[65]. Una de las características importantes a capturar en esta sección es la edad de las distintas personas que integran el hogar, ya que esta información permite definir quién debe responder el cuestionario de Gastos Diarios (6.1.1.2), instrumento base para la captura de gastos.

- **Educación:**

Además de capturar información básica sobre curso y nivel más alto alcanzado por cada integrante del hogar, recoge información sobre la existencia de estudios actuales, formales o complementarios. Asimismo, para quienes estudian en educación formal, caracteriza la dependencia del establecimiento educativo. Esta información descriptiva permite la validación y edición de los gastos por conceptos de educación, registrados en el módulo de Gastos en Educación (6.1.1.4).

- **Condición de actividad económica:**

Registra información respecto a la situación laboral de una persona de 15 años y más, en el marco de medición de la fuerza de trabajo[66]. Adicionalmente, caracteriza la ocupación principal de las personas que se declaran ocupadas e identifica los tipos de ingreso percibidos. La caracterización de la actividad económica permite determinar qué secciones del módulo de Ingresos (6.1.1.7) se deben consultar para cada integrante del hogar, disminuyendo el número de preguntas a realizar y la probabilidad de rechazo del módulo de Ingresos.

- **Salud y previsión:**

Recopila información sobre la afiliación y cotización a sistemas de salud y previsión de las personas integrantes del hogar de 18 años y más, identificando adicionalmente el sistema al que pertenecen. En el caso de cotizar en un sistema previsional, la caracterización de este permite la estimación del descuento obligatorio a las personas que se declaran como asalariados. Contar con esta información permite la construcción del ingreso disponible del hogar.

- **Vivienda principal[67]:**

Captura información sobre la materialidad de la vivienda, junto con la información sobre la tenencia de la vivienda principal. Además, registra información necesaria para el cálculo del arriendo imputado[68].

[65] La persona sustentadora principal corresponde al integrante del hogar que más aporta al presupuesto familiar, según los últimos seis meses. Este no corresponde, necesariamente, al integrante que recibe más ingresos.

[66] La secuencia de preguntas se realiza siguiendo los criterios de la OIT, implementados también en la actual Encuesta Nacional de Empleo (ENE) del INE.

[67] Se refiere a la vivienda utilizada como residencia habitual, es decir, aquella donde la persona habita o duerme al menos cuatro días a la semana, durante los últimos seis meses.

[68] El arriendo imputado corresponde a los beneficios del hogar por habitar en la vivienda (propia, arrendada a un valor inferior al de mercado, cedida, gratuitamente u otras condiciones). Como los beneficios no son monetarios, se realiza una estimación, por parte del propietario, arrendatario o beneficiario, para acercarnos a las ganancias que ha percibido el hogar por el hecho de habitar la vivienda.

- **Vivienda secundaria:**

Indaga sobre la tenencia de vivienda secundaria, la que se define por cuatro características: 1) es una vivienda propia (pagada, pagándose, herencia o sucesión compartida), 2) el hogar la tuvo a disposición el mes anterior, 3) el hogar la tuvo a disposición al menos tres meses durante el último año y 4) su uso principal es la recreación o el descanso. Esta información se utiliza para determinar si debe aplicarse o no la sección de preguntas del módulo Gastos periódicos del Hogar, asociadas a los gastos generados por dicha vivienda[69].

[69] Para definir la vivienda secundaria, se establecen todas estas condiciones, ya que se busca diferenciarla de viviendas que el hogar arrienda con el fin de generar ingreso y en dichos casos los gastos asociados a esa vivienda corresponderían a gastos de consumo intermedio.

[70] Es decir, la persona informante declara que no tiene una condición que represente un impedimento para responder el cuestionario. Las condiciones que impiden la participación pueden ser de carácter temporal o permanente (ceguera, sordera, dificultad de desplazamiento, entre otros).

[71] Las quincenas operativas pueden ser de 14, 15 o 16 días dependiendo del mes en la cual se encuentran.

6.1.1.2
Cuestionario de Gastos Diarios (GD)

El cuestionario de Gastos Diarios (GD) captura el gasto en bienes y servicios que realiza cada integrante del hogar de 15 años o más que se encuentran aptas física y/o mental para contestar[70]. Este instrumento, dispuesto en formato papel, es el único autoadministrado en la encuesta y es entregado a las personas informantes para que lo utilicen como una libreta en la cual, durante todos los días de la quincena operativa[71], deben anotar de forma detallada todos los gastos en consumo final que han realizado.

La disposición del cuestionario en papel corresponde a dos páginas por día. En la primera página se solicita a la persona informante pegar las boletas de los gastos que han realizado en el día y detallar si estas no cuentan con suficiente información para su traspaso. Al término de la primera página se encuentra un cuadro exclusivo para el llenado de la persona entrevistadora cuyo objetivo es registrar la información para las boletas de grandes cadenas de supermercado. La información

de boletas será usada posteriormente para la rectificación y comprobación de valores de gasto correspondientes a Alimentos y Bebidas. La segunda página incluye tanto información del día y fecha como 15 líneas de registro para las ocho columnas que deben ser contestadas por la persona informante. Adicionalmente, al término de la página, se encuentra un cuadro de llenado exclusivo para la persona entrevistadora en la que se deberá dar cuenta sobre la situación de registro del día, estas situaciones son: 1) día sin gasto; 2) día sin registro[72].

Los gastos deben ser registrados sin importar si estos se realizan para beneficio propio, de otra persona integrante del hogar o como regalo a otro hogar. Otro objetivo de este cuestionario es obtener información sobre el autosuministro de los hogares que posean algún tipo de negocio perteneciente a los integrantes del hogar. De esta forma, si un bien fue adquirido a través del retiro de productos desde un negocio, este se debe registrar en el cuestionario de GD y el monto debe ser contabilizado al valor venta o precio de mercado.

6.1.1.3
Módulo de Gastos periódicos del Hogar (GH)

En el módulo de Gastos periódicos del Hogar (GH) se registran todos los gastos en consumo final que realiza el hogar y que presentan algún tipo de regularidad (mensual, trimestral, anual, etc.).

El ejemplo más claro son los gastos en servicios básicos de una vivienda, tales como el gasto en agua, electricidad, entre otros. Para la captura de los gastos periódicos se utiliza el criterio de consumo pagado (revisado en el punto 4.1.2), debido a que se consulta por lo que el hogar pagó o debería haber pagado por el consumo del bien y servicio durante el mes anterior (período de referencia). De este modo, no se contabiliza como gasto el cobro de consumos anteriores a dicho período o los intereses en mora asociados a esa deuda.

[72] Este recuadro es muy importante para diferenciar entre: un día que no tiene gasto porque la persona lo declaró explícitamente. De un día sin registro el que corresponde a una incógnita, hay ausencia de información, pero se desconoce si es porque la persona olvidó anotar el gasto de ese día o simplemente no realizó ningún gasto.

El módulo de GH se divide en las siguientes secciones:

- **Gastos periódicos relativos a la vivienda principal:**
Registra los gastos periódicos relacionados con el funcionamiento de la vivienda, los que normalmente presentan una frecuencia mensual de pago. Debido a lo anterior, se consulta al hogar por el valor del gasto asociado al consumo del mes anterior a la aplicación de la encuesta.

- **Gastos periódicos relativos a la vivienda secundaria:**
Captura los gastos periódicos relativos a la/s vivienda/s secundaria/s de aquellos hogares que declararon poseerla/s en RPH.

- **Gastos periódicos en telecomunicaciones del hogar:**
Captura los gastos periódicos relativos al consumo de servicios de telecomunicaciones, tales como televisión, telefonía e internet, servicios de transmisión en línea, entre otros.

- **Gastos periódicos en aseo municipal:**
Captura los gastos periódicos relativos al pago de aseo municipal.

- **Gastos relativos al uso de vehículos con fines domésticos:**
Captura gastos periódicos vinculados al uso de vehículos, tales como combustible, estacionamiento y peaje, entre otros.

- **Gastos en servicio doméstico contratados por el hogar:**
Captura gastos periódicos vinculados a la contratación de servicios domésticos, tales como limpieza de la vivienda, cuidado de niños, mantenimiento de jardín, entre otros.

- **Otros gastos periódicos del hogar:**
Recopila información de otros gastos pagados de manera periódica por el hogar, tales como suscripciones a diarios, revistas, gimnasio, entre otros.

6.1.1.4
Módulo de Gastos en Educación (GE)

El objetivo del módulo de Gastos en Educación es recopilar todos los gastos relacionados con la educación de las personas del hogar que se encuentran estudiando actualmente, tanto en educación formal como complementaria, se considera información a nivel de persona y de hogar.

El módulo GE se divide en las siguientes secciones:

- **Gastos en educación formal:**
Permite registrar gastos de quienes integran el hogar que asisten a educación formal en un año académico. Considera pagos por concepto de matrícula, aranceles, cuotas de centro de padres, cuotas de incorporación, exámenes de admisión, actividades escolares, además de derechos y servicios en educación superior. Complementariamente, esta sección captura información relativa a becas y créditos como alternativas de financiamiento para la educación, cabe señalar que esta información no constituye gasto en consumo final del hogar.

- **Gastos en educación no formal:**
Permite registrar gastos relativos a formas de educación que pueden ser tanto alternativas como complementarias a la educación formal. Se consulta por gastos de matrícula y arancel en preuniversitarios, cursos de idiomas, tutorías o clases particulares, seminarios o congresos, capacitaciones o certificaciones, además de otro tipo de cursos no incluidos en las categorías anteriores (talleres, clases particulares o colectivas, entre otros).

- **Sección de otros gastos relativos en educación:**
Permite registrar gastos anuales y mensuales realizados debido a la asistencia de alguna persona integrante del hogar a la educación formal en sus etapas preescolar o escolar. Considera un registro de gastos por concepto de vestimenta, calzado, útiles escolares, textos de estudio, transporte escolar, entre otros.

6.1.1.5
Cuestionario de Gastos del Recuerdo (GR)

El cuestionario de Gastos del Recuerdo (GR) captura aquellos gastos en bienes y servicios que se realizan con menor frecuencia, pero que pueden representar un monto importante desde el punto de vista del presupuesto del hogar. Los gastos en bienes y servicios se capturan mediante 21 tablas (o secciones) distribuidas en tres períodos acotados de recuerdo: desde el momento de la entrevista hasta 3, 6 y 12 meses previos al mes en que se aplica la encuesta en el hogar[73].

Para cada período de recuerdo se capturan gastos en productos específicos, los que son definidos según la capacidad de recuerdo del hogar. En primera instancia, los períodos establecidos en el cuestionario se corroboraron o ajustaron de acuerdo con los resultados de las pruebas del cuestionario, el análisis de encuestas internacionales y a través del estudio de la frecuencia de compra según los datos analizados de la VIII EPF.

Las tablas del cuestionario de Gastos del Recuerdo agrupan bienes y servicios para facilitar el recuerdo de la persona y, a su vez, acotar temporalmente este recuerdo en los tres períodos antes descritos. Los períodos del recuerdo y sus respectivas tablas son descritos a continuación.

- **Período de recuerdo 3 meses**

•**Tabla 1:** Mantención y reparación de la vivienda (servicios y materiales)

•**Tabla 2:** Consultas médicas, cirugías ambulatorias, dentistas y otros profesionales de la salud

•**Tabla 3:** Exámenes médicos

•**Tabla 4:** Artículos de corrección médica

•**Tabla 5:** Libros y espectáculos

•**Tabla 6:** Vestuario

[73] Cabe aclarar que los períodos de recuerdo más amplios contienen a los más acotados. Para más detalle revisar el Manual de Recolección de datos Tomo II (INE Chile, 2023c).

•**Tabla 7:** Calzado

•**Tabla 8:** Repuestos, accesorios y mantenimiento para vehículos

•**Tabla 9:** Artículos de telefonía

•**Tabla 10:** Maletas, bolsos, carteras, joyas y relojes

•**Tabla 11:** Equipos de deporte, campamento e instrumentos musicales

•**Tabla 12:** Artefactos eléctricos pequeños para el hogar y herramientas (manuales y eléctricas)

● **Período de recuerdo 6 meses**

•**Tabla 13:** Muebles

•**Tabla 14:** Computación

•**Tabla 15:** Artefactos electrónicos

•**Tabla 16:** Grandes artefactos para el hogar (eléctricos o no)

● **Período de recuerdo 12 meses**

•**Tabla 17:** Viajes realizados

•**Tabla 18:** Vehículos nuevos

•**Tabla 19:** Vehículos usados

•**Tabla 20:** Hospitalizaciones

•**Tabla 21:** Otros gastos en período de recuerdo no considerados con anterioridad

6.1.1.6
Módulo de Gastos en Servicios Financieros y Seguros (GSF)

El módulo de Gastos en Servicios Financieros y Seguros (GSF) busca capturar el gasto asociado al uso y/o tenencia de instrumentos financieros. Los gastos financieros corresponden a intereses, gastos en administración, comisiones, seguros asociados y no asociados a dichos instrumentos.

El módulo GSF se divide en las siguientes secciones:

•Créditos Hipotecarios (ch)
•Tarjetas de crédito de Establecimientos comerciales (te)
•Crédito de Consumo y crédito Automotriz (cca)
•Cuentas corrientes y líneas de crédito (ctt)
•Cuentas vistas y cuentas RUT (cv)
•Tarjetas de crédito bancarias (tcb)
•Créditos para educación superior (ces)
•Seguros (se)

Cada sección, a excepción de las secciones Cuenta vista y cuenta RUT y Seguros, se divide en dos grupos de preguntas. El primer grupo, **zona de caracterización** de la sección, busca identificar la tenencia del instrumento financiero y contrastar la información en términos de monto solicitado o utilizado, plazo, valor de la cuota, institución emisora, entre otros. El segundo grupo, **zona de gasto** de la sección, captura los gastos en intereses, gastos de administración, comisiones y seguros asociados. No obstante, para

responder la zona de gasto es indispensable que el hogar disponga de los comprobantes con la información del instrumento.

Por otra parte, en las secciones Cuenta vista y cuenta RUT y Seguros, se capturan las características de cada instrumento y para declarar el gasto financiero no requiere que la persona informante disponga del comprobante, por lo tanto, el gasto se registra según el recuerdo de la persona.

6.1.1.7
Módulo de Ingresos (IN)

El módulo de Ingresos (IN) responde al objetivo secundario de la EPF, es decir, permite identificar la estructura del ingreso disponible de los hogares urbanos, en las capitales regionales del país y sus zonas conurbadas. En este se registra información respecto a los ingresos monetarios y no monetarios de todas las personas integrantes del hogar de 15 años o más[74]. El período de referencia corresponde al mes anterior al de la entrevista y se registran los ingresos generados según el criterio de ingreso devengado, es decir, sin importar si estos realmente se recibieron durante ese mes.

Los distintos componentes del ingreso se registran en las siguientes secciones:

- **Ingresos del trabajo dependiente (ta):**
Captura los ingresos provenientes del trabajo dependiente[75] correspondiente a la actividad principal, estos incluyen los ingresos generados a partir de personas asalariadas o a honorarios. Se busca información desde el salario bruto hasta los descuentos legales o voluntarios. Adicionalmente, se capturan remuneraciones en especies y/o servicios que puedan ser recibidas.

- **Ingresos del trabajo independiente (ti):**
Recopila información de los ingresos generados a partir de negocios por cuenta propia o por profesionales independientes que dirigen y organizan su propia actividad.

[74] No es parte de los objetivos de la encuesta capturar ingresos asociados al trabajo infantil.

[75] Entendido como aquel que realiza sus actividades laborales bajo subordinación y dependencia, es decir, tiene jefatura, cumple un horario, tiene continuidad en el trabajo, periodicidad de la remuneración, entre otros.

- **Otros ingresos del trabajo (ot):**

Captura ingresos derivados por dos tipos de fuentes: ingresos provenientes de segundas actividades e ingresos derivados de una actividad laboral distinta a la principal, que previamente fue declarada en el RPH para la semana de referencia.

- **Ingresos por jubilaciones y/o pensiones de vejez (ju):**

Registra ingresos de jubilaciones y/o pensiones de vejez provenientes de ahorros previsionales y, en algunas situaciones, aportes fiscales (pilar básico solidario y pensiones provenientes de cajas de ahorro fiscales, entre otras).

- **Ingresos por otras pensiones y transferencias recibidas (tr):**

Captura ingresos que no son generados por una actividad económica. Incluye pensiones alimenticias, pensiones de invalidez o discapacidad, montepíos, pensiones asistenciales, entre otras. Además, registra transferencias (monetarias y en especies) recibidas desde otros hogares, dentro o fuera del país.

- **Otros ingresos de carácter no habitual (nh):**

Recopila información sobre ingresos significativos en el presupuesto del hogar, pero que no son periódicos o habituales, tales como herencias, loterías, indemnizaciones, entre otros. El objetivo de esta sección es capturar una fuente de ingresos no corrientes[76], sin embargo, estos no son considerados para el cálculo del ingreso disponible del hogar.

- **Ingresos de la propiedad y ganancias por tenencia de instrumentos financieros (ip):**

Captura ingresos derivados de la tenencia de propiedades (muebles o inmuebles) y de instrumentos financieros, como acciones, cuentas de ahorro, depósitos a plazo, entre otros.

[76] Este tipo de ingresos puede explicar una estructura de gasto anómala de ese hogar en particular durante el período de referencia.

- **Transferencias emitidas (ite):**
Registra información sobre el total de transferencias periódicas destinadas a otros hogares e instituciones sin fines de lucro, pueden ser en dinero y/o en especies.

6.1.1.8
Períodos de referencia de los instrumentos de captura de datos

Los instrumentos de recolección de datos tienen distintos períodos de referencia. No obstante, para llevar los datos a una misma unidad temporal, los gastos e ingresos se deben ajustar temporalmente. En el caso de la EPF, la información se mensualiza, es decir, se llevan a un "mes tipo".

A continuación, se detallan los distintos períodos de referencia según la información que se desea obtener:

- **Cuestionario de Gastos Diarios:**
Tiene como período de referencia la quincena de aplicación de la encuesta.

- **Módulo de Gastos periódicos del Hogar:**
Posee un período de referencia distinto dependiendo de la sección en la que fue registrado el gasto. En la mayoría de las secciones se captura el gasto del mes anterior y para el resto de las secciones existe diversas opciones de periodicidad para que el hogar las declare.

- **Módulo de Gastos en Educación:**
Tiene un período de referencia de 12 meses considerando los gastos asociados al año académico actual[77].

- **Cuestionario de Gastos del Recuerdo:**
Puede tener un período de referencia de 3, 6 y 12 meses dependiendo del tipo de gasto registrado. Todos los períodos de referencia terminan el día anterior al inicio de la submuestra operativa y dependiendo del tipo de gasto declarado, cambia la cantidad de meses del período.

[77] Es uno de los períodos de referencia más complejos debido al corte del recuerdo del año académico, para más detalle revisar el capítulo de Gastos en Educación del Manual de recolección de datos Tomo II (INE Chile, 2023c).

- **Módulo de Ingresos:**
Tienen como período de referencia el mes anterior a la quincena de aplicación de la encuesta.

- **Módulo de Gastos en Servicios Financieros y Seguros:**
Tiene como período de referencia el mes anterior a la quincena de aplicación de la encuesta.

6.1.2 Instrumento de captura de paradatos de la recolección (Hoja de Ruta)

Para dar trazabilidad y caracterizar el proceso de recolección de la encuesta, además de los módulos temáticos, la IX EPF incorporó en el cuestionario digital un módulo complementario, que permitiese el registro de las actividades más importantes contempladas por la metodología de recolección. Este módulo, denominado Hoja de Ruta (HR), es parte de los estándares institucionales de calidad aplicables a encuestas de hogares (INE Chile, 2019) y tuvo como objetivo caracterizar el resultado de la encuesta en cada contacto realizado con las viviendas y hogares de la muestra[78], propiciando un conocimiento transversal sobre la calidad de los datos obtenidos.

La Hoja de Ruta estuvo compuesta por siete secciones, que fueron completadas por las personas encuestadoras a medida que avanzaba la ejecución de la recolección de datos en cada vivienda y hogar contenido en la muestra. Las secciones y sus contenidos se listan a continuación.

- **Entorno de la vivienda (ev):**
Registró información sobre las características de la vivienda (tipo de acceso y estado físico) y del entorno cercano a ella (presencia de basura y vandalismo/rayados), a través de preguntas de observación completadas al momento de ejecutar el primer intento de contacto.

[78] En principio, se registró una Hoja de Ruta para cada unidad muestral de la encuesta. No obstante, en aquellos casos en que se detectaron viviendas con más de un hogar residente, se solicitó registrar la información de la cantidad de hogares detectados y, posteriormente, se habilitó un nuevo cuestionario digital vinculado al resto de cuestionarios asociados a la misma unidad muestral a través de un código identificador único (folio). De esta forma, fue posible generar Hojas de Rutas separadas, y, en consecuencia, recolectar la información de cada hogar de manera separada.

- **Registro de contacto (rc):**

Capturó información de cada intento de contacto, permitiendo identificar responsable, localización GPS, fecha, hora y resultado del intento. La clasificación de resultados fue realizada a partir de una clasificación estándar a nivel institucional para entrevistas presenciales a hogares, compuesta por 23 códigos, que derivan en la clasificación de seis categorías de respuesta[79].

- **Control de Gastos Diarios (control GD):**

Capturó información de elegibilidad para la entrega del cuestionario de GD a integrantes del hogar y atributos de respuesta a la recepción de esos cuestionarios. Permitió monitorear la cantidad de cuestionarios entregados y retirados, además de un conteo referencial de días con registro en cada uno. Por último, posibilitó la asociación informática del identificador de cuestionario en papel al respectivo hogar entrevistado.

- **Control de Gastos del recuerdo (control GR):**

Registró información descriptiva de la entrevista del cuestionario GR (fecha y duración de la entrevista). Esta sección, permitió la asociación informática de cada cuestionario en papel con el hogar entrevistado junto con el monitoreo de la ejecución de la entrevista de este formulario.

- **Registro de RUN de integrantes del hogar (RUN):**

Capturó Rol Único Nacional de cada integrante del hogar de 15 años o más, a partir de la consulta a informante idóneo del hogar.

- **Agenda de citas con la persona informante (ac):**

Buscó facilitar la organización y gestión de entrevistas pactadas con el hogar, especificando fecha, hora y lugar de citas acordadas con las personas informantes.

- **Resultado de la encuesta (re):**

Registró el resultado del proceso completo de recolección en cada hogar en que se inició un proceso de entrevista.
Entre los usos de la Hoja de Ruta, el más relevante fue la caracterización

[79] Las categorías de resultado son: No elegibles, Elegibilidad desconocida, Elegibles-Entrevistada, Elegibles-Rechazadas, Elegibles-No entrevistadas por otros motivos, Elegibles-No Contactadas.

del resultado final del trabajo de recolección en cada unidad muestral y hogar incluido en el estudio. Ello constituye un insumo relevante, tanto para medir el error de no respuesta total como para realizar los ajustes a los factores de expansión de la encuesta.

6.2 Prueba y evaluación de los instrumentos de captura de información

El instrumento de captura de información es la vía por la cual se obtienen datos directos de los hogares participantes en la EPF. La prueba y evaluación de este instrumento es esencial para garantizar que las preguntas sean claras, comprensibles y capaces de recopilar información precisa sobre gastos e ingresos.

La prueba del instrumento permite identificar problemas potenciales antes de la implementación completa de la encuesta. Esto incluye ambigüedades en las preguntas, posibles sesgos en las respuestas y dificultades que puedan enfrentar quienes encuestan al proporcionar información.

Para la definición de la tecnología de recolección y ajustes a los distintos cuestionarios se realizaron dos pruebas piloto, en 2017 y 2020, cuya aplicación permitió generar evidencia para las decisiones que se implementaron en la IX EPF. A continuación, se definen los principales aspectos de cada una de las pruebas, sus principales antecedentes, objetivos y resultados.

6.2.1 Prueba Piloto 2017 de Dispositivos Móviles

En el año 2017 se ejecutó una prueba piloto cuyo objetivo fue "analizar el efecto tecnológico asociado a realizar un cambio en el formato de recolección de datos de la Encuesta de Presupuestos Familiares (EPF), desde entrevistas realizadas en papel a aquellas ejecutadas a través de dispositivos móviles de captura de información" (INE Chile, 2018, pág. 4). El efecto de esta intervención se observó en las variables principales del estudio, es decir, las estimaciones de gasto total promedio mensual e ingreso disponible promedio mensual de los hogares.
Adicionalmente, puesto que el objetivo de la prueba piloto 2017 fue

evaluar el efecto del cambio tecnológico en la EPF, diferentes aspectos de la metodología de recolección de la encuesta debieron ser adaptados para este cometido. Las principales adaptaciones corresponden a la disminución de las submuestras operativas de 15 a 7 días y la modificación del esquema de visitas[80].

Dado este diseño de prueba piloto los principales hallazgos obtenidos fueron los siguientes.

En cuanto al objetivo de comparar el efecto tecnológico en el análisis de resultados de las principales estimaciones de la EPF, no se encontraron diferencias estadísticamente significativas en las estimaciones del ingreso disponible del hogar ni del ingreso disponible *per cápita* (INE Chile, 2018, pág. 85).

No obstante, se observaron diferencias significativas en la estimación del gasto total *per cápita*. En este indicador, los hogares del grupo tratamiento (CAPI) presentaron un menor gasto en comparación con los resultados obtenidos en la VIII EPF, siendo menores en $-69.118. Este impacto se atribuyó principalmente a la captura del gasto en los Gastos del Recuerdo el cual fue diseñado con múltiples tablas compuestas con preguntas abiertas. Esta decisión ocasionó un aumento en los tiempos de aplicación, lo que a su vez contribuyó a la omisión de datos por parte de las personas informantes provocando una sub-declaración del gasto en este cuestionario.

[80] Debido a la exclusión de la Libreta de Gastos Individuales (LGI) y Libreta de Gastos en Servicios Financieros y Seguros (LSF) de la prueba piloto 2017.

Tabla 22: Gasto total y gasto *per cápita* del hogar

Variable	PAPI	CAPI	C-P	p-valor
1 Gasto total del hogar, VIII EPF	642.569	711.010	68.441	0,382
2 Gasto total del hogar, prueba Piloto	604.058	582.332	-21.726	0,727
3 Cambio en el promedio de gasto	-38.510	-128.678	-90.168	0,111
4 Gasto *per cápita* del hogar, VIII EPF	246.511	332.949	86.439	0,020*
5 Gasto *per cápita* del hogar, prueba Piloto	248.301	265.622	17.321	0,572
6 Cambio en el promedio de gasto *per cápita*	248.301	265.622	17.321	0,572
Muestra	193	167	360	

Nota: PAPI refiere de Paper And Pencil Interview y corresponde al grupo control de la prueba piloto 2017, mientras que CAPI es Computer Assisted Personal Interview y es el grupo tratamiento.
** p < 0,05; ** p < 0,01*

Fuente: Instituto Nacional de Estadísticas (INE), Prueba Piloto 2017 de Dispositivos Móviles en la Encuesta de Presupuestos Familiares (EPF).

Tampoco se encontraron diferencias estadísticamente significativas en términos de errores de procesamiento (INE Chile, 2018, pág. 78), porcentaje de hogares con valores inválidos, ni respecto a los resultados del análisis de no respuesta al ítem, por libreta y por persona, ni en el análisis de tiempos de duración de las entrevistas entre los grupos experimentales. Por el contrario, se encontraron diferencias estadísticamente significativas en las tasas de cooperación y respuesta, las cuales favorecieron al grupo control.

Por último, respecto a la evaluación cualitativa de la recolección con la aplicación EPF Móvil (INE Chile, 2018, pág. 101), las personas recolectoras evaluaron de manera positiva la funcionalidad entregada por la aplicación en términos de la recuperación de información guardada, la claridad de los registros, la disminución de los errores de registro y, en síntesis, la percepción general de una mejora en la calidad de los datos recolectados.

En conclusión, los resultados de la prueba piloto realizada el año 2017 (INE Chile, 2018, pág. 128) dan cuenta de importantes ventajas en el uso de dispositivos móviles en términos de la mejora en la calidad del dato, eficiencia en el flujo de la entrevista, mejoras en el procesamiento y análisis de los datos, y la disminución de los costos asociados a la aplicación de la encuesta EPF mediante estos.

Sin embargo, los resultados también dejan en evidencia las principales dificultades del cambio tecnológico en el modo de recolección, las cuales están vinculadas principalmente al diseño, la programación y la interfaz de los cuestionarios digitales, enfatizando la necesidad de dotar de mayor flexibilidad a la aplicación del cuestionario y la adaptación de la Libreta de Gastos del Recuerdo, en la que se propuso la alternativa de incorporar listas de establecimientos o productos pre-codificadas en vez de preguntas abiertas. Finalmente, se recomienda continuar con la realización de pruebas piloto para aprovechar las potencialidades de la tecnología en la aplicación de la encuesta, la adecuación del diseño de los cuestionarios, aumento del tamaño de la prueba, entre otros.

6.2.2 Prueba Piloto IX EPF

Los antecedentes mencionados sobre la incorporación de dispositivos móviles junto con los resultados de la prueba piloto y las observaciones realizadas por los usuarios, es que en el año 2020 se realiza una nueva prueba piloto para el cambio tecnológico y prueba de cuestionarios: la prueba piloto IX EPF. En esta prueba, a diferencia de la realizada en 2017, la aplicación seleccionada para la construcción del cuestionario fue en *Survey Solutions* (SuSo), *software* desarrollado por el grupo de datos del Banco Mundial.

Se diseñó un estudio piloto para ser aplicado entre octubre y diciembre de 2020, con el objetivo de evaluar el efecto de los ajustes a los cuestionarios y metodología de recolección para la IX Encuesta de Presupuestos Familiares, en la captura y procesamiento de información de la encuesta.

Inicialmente, se planificó la evaluación del cambio tecnológico a través de entrevistas presenciales, en que se aplicarían los cuestionarios (digital y en papel). Sin embargo, frente a la pandemia del COVID-19 fue necesario adecuar esta evaluación, generando condiciones que se ajustaran a las restricciones de movilidad y requisitos de distanciamiento en las interacciones.

Específicamente, estas condiciones implicaron rediseñar la metodología de

recolección junto con los instrumentos, adoptando un esquema de recolección que contempló cinco entrevistas, tres entrevistas presenciales y dos telefónicas. Este cambio buscó disminuir la exposición al contagio de las personas informantes y recolectoras durante las entrevistas presenciales y reducir la cantidad de traslados del equipo de recolección. Esto se planificó considerando un escenario donde la pandemia se mantenía constante en los siguientes años.

Los ajustes implicaron una modificación a los protocolos y flujos de trabajo en el proceso de recolección con el fin de asegurar la comparabilidad entre ambos grupos experimentales (recolección del grupo control y grupo tratamiento[81]). Si bien se mantuvo la estructura de tres fases de la VIII EPF, preparación, aplicación, cierre y recuperación, se incorporaron modificaciones en algunas actividades de cada fase adaptándolas a las características de la modalidad mixta.

En suma, se buscaba medir el efecto de los ajustes en las principales estimaciones que resultan de la encuesta, los errores no muestrales, la cantidad de preguntas y evaluar, también, el efecto de estos cambios desde la perspectiva de la persona encuestadora.

Para esta prueba piloto se utilizó el marco muestral de viviendas 2020 y se seleccionaron cuatro regiones para su realización, las cuales permitieron aplicar la prueba en distintas zonas geográficas del país y que también existiese diversidad respecto a los resultados observados en sus indicadores operativos obtenidos en la VIII EPF.

La muestra seleccionada para la prueba no buscó representatividad de la población, en términos de inferencia o validez externa. Por el contrario, a través de un diseño experimental por muestra dividida y aleatorizado se pretende controlar la validez interna del estudio. Es decir, la validez interna determina hasta qué punto "el investigador puede atribuir la variación observada en la variable dependiente a la presencia de la variable independiente" (Tejedor, 1981).

[81] El grupo control correspondía a todos los cuestionarios en papel de la VIII EPF y el grupo tratamiento a la aplicación de un cuestionario digital más un cuestionario autoadministrado en papel (Gastos Diarios). La adaptación del modo de recolección (presencial y telefónica) se aplicó a ambos grupos experimentales.

Por lo anterior, y dado que el objetivo del estudio fue evaluar las diferencias entre el grupo de control y tratamiento para generar propuestas de modificación a la IX EPF, se realizaron ejercicios de análisis de potencia estadística. Estos ejercicios permiten determinar el tamaño muestral necesario para abordar el estudio de manera preliminar.

La siguiente tabla resume la distribución de la muestra por grupo y región.

Tabla 23: Distribución de la muestra por grupo y región

Región	Comuna	Control (VIII EPF)	Tratamiento (PP IX EPF)	Total
Tarapacá	Iquique	113	112	225
Metropolitana	La Granja, Las Condes, Macul, Santiago	150	150	300
Araucanía	Temuco	135	135	270
Los Ríos	Valdivia	112	113	225
Total grupo		510	510	1.020

Fuente: Instituto Nacional de Estadísticas (INE), Prueba Piloto 2020 de la IX Encuesta de Presupuestos Familiares (EPF).

Se analizó el efecto que generaba aplicar todos los cuestionarios de la encuesta en dispositivos móviles, a excepción de Gastos Diarios, en los distintos indicadores operativos, principales resultados y errores no muestrales. Los principales hallazgos de la segunda prueba piloto fueron:

a. Mayor proporción de inválidos en el grupo de control (cuestionarios VIII EPF) en los módulos de Registro de Personas del Hogar, Gastos en Educación, Gastos periódicos del Hogar y Gastos en Servicios Financieros y Seguros. Respecto a Gastos del Recuerdo, la mayor proporción de inválidos se encontró en el grupo de tratamiento (cuestionario digital).

b. Adicionalmente, se observó una pérdida de información de los hogares desde la primera entrevista a la quinta, afectando principalmente a los módulos de Ingresos y Gastos en Servicios Financieros y Seguros.

c. Respecto a los indicadores principales del estudio, no se observaron diferencias estadísticamente significativas (al 95% de confianza) en las mediciones del gasto promedio mensual del hogar y *per cápita* y del ingreso disponible promedio mensual del hogar y *per cápita*. La **Tabla 24** resume los resultados para el ingreso disponible, total hogar y *per cápita*. Las columnas Control y Tratamiento dan cuenta del grupo experimental, mientras que la columna C-T corresponde a la diferencia promedio entre estos grupos.

Tabla 24: Ingreso disponible promedio (hogar y *per cápita*)

Estimaciones[1]	Grupo Control	Grupo Tratamiento	C-T	Valor-p
Ingreso disponible hogar	**821.692**	**711.574**	**110.118**	**0,16**
Ingreso disponible *per cápita* del hogar	**369.084**	**294.760**	**74.324**	**0,09**
Tamaño muestra	**258**	**199**	**457**	

(1) Todas las estimaciones están expresadas en pesos corrientes, no consideran valores imputados y excluyen arriendo imputado.

Fuente: Instituto Nacional de Estadísticas (INE), Prueba Piloto 2020 de la IX Encuesta de Presupuestos Familiares (EPF).

Respecto al gasto promedio mensual, en la **Tabla 25** se observa que, de la misma forma que el ingreso, no hay diferencias estadísticamente significativas, a un 95% de confianza, entre los grupos de control y tratamiento.

Tabla 25: Gasto total promedio (hogar y *per cápita*)

Estimaciones[1]	Grupo Control	Grupo Tratamiento	C-T	Valor-p
Gasto promedio del hogar	**822.040**	**775.695**	**46.345**	**0,50**
Gasto promedio *per cápita* del hogar	**350.243**	**319.421**	**30.823**	**0,35**
Tamaño muestra	**258**	**199**	**457**	

(1) Todas las estimaciones están expresadas en pesos corrientes, no consideran valores imputados y excluyen arriendo imputado.

Fuente: Instituto Nacional de Estadísticas (INE), Prueba Piloto 2020 de la IX Encuesta de Presupuestos Familiares (EPF).

Según los resultados obtenidos[82] es que se decide mantener la aplicación en papel de Gastos del Recuerdo además del cuestionario autoadministrado de Gastos Diarios. En la **Tabla 26** es posible ver los cambios de cuestionario y el formato final de aplicación en la IX EPF en comparación a la VIII versión del estudio.

Tabla 26: Comparación de cuestionarios según tecnología de la VIII y IX EPF

VIII EPF		IX EPF	
Cuestionario	Tecnología	Cuestionario	Tecnología
Registro de Personas en el Hogar (RPH)	Papel	Registro de Personas en el Hogar (RPH)	Digital para dispositivo móvil
Libreta de Gastos Individuales (LGI)	Papel	Cuestionario de Gastos Diarios (GD)	Papel
Libreta de Gastos del Hogar (LGH)	Papel	Gastos periódicos del Hogar (GH)	Digital para dispositivo móvil
		Gastos en Educación (GE)	Digital para dispositivo móvil
Libreta de Gastos del Recuerdo (LGR)	Papel	Gastos del Recuerdo (GR)	Papel
Libreta de Gastos en Servicios Financieros y Seguros (LSF)	Papel	Gastos en Servicios Financieros y Seguros (GSF)	Digital para dispositivo móvil
Libreta de Ingresos (LI)	Papel	Ingresos (IN)	Digital para dispositivo móvil
Hoja de Ruta (HR)	Papel	Hoja de Ruta (HR)	Digital para dispositivo móvil

Fuente: Instituto Nacional de Estadísticas (INE), VIII y IX Encuesta de Presupuestos Familiares (EPF).

[82] Las decisiones asociadas a los resultados obtenidos de la Prueba Piloto 2020 fueron presentadas y discutidas con el Comité de Expertos que acompañó el proceso de producción estadística de la IX EPF.

6.3 Características metodológicas de la recolección oficial de datos

Las técnicas y métodos de recolección de datos constituyen el puente que conecta los objetivos de la EPF con la realidad de los hogares. Son el medio a través del cual se obtiene información directa y detallada sobre los gastos e ingresos, permitiendo una comprensión precisa del presupuesto de los hogares.

6.3.1 Período de recolección y distribución operativa de la muestra

El proceso de recolección de datos de la IX Encuesta de Presupuestos Familiares tuvo lugar entre el 01 de octubre del año 2021 y el 30 de septiembre del 2022. Durante este período, se buscó entrevistar a todos los hogares residentes en 29.231 viviendas seleccionadas en el diseño muestral, en el transcurso de un año continuo.

Cada submuestra mensual fue subdividida en dos grupos de tamaño similar de viviendas, llamados submuestras operativas, las cuales se asociaron a un período específico de colaboración, correspondiente a una de las quincenas[83] del mes. De este modo, durante la recolección oficial del estudio se entrevistó durante 24 submuestras operativas. La asignación de las viviendas a una submuestra operativa se realizó de forma sistemática, de modo que las distintas UPM tuviesen presencia en ambas quincenas.

6.3.2 Técnica de producción de información

El trabajo de recolección capturó información sociodemográfica, de ingresos y gastos de los hogares mediante la combinación de entrevistas presenciales y un cuestionario en papel autoadministrado. Las

[83] El concepto de quincena es referencial, en tanto su duración dependerá de los días contenidos en el mes de trabajo de recolección. Así, en la IX EPF al igual que en su VIII versión, existieron "quincenas" de 14 (febrero), 15 y 16 días de duración. De este modo, la quincena es equivalente al período de colaboración de los hogares en el cuestionario de GD. La quincena de colaboración asignada podía ser modificada, en casos excepcionales y por razones de fuerza mayor de las personas informantes o necesidades del proyecto. No obstante, las viviendas solo podían ser entrevistadas dentro del mismo mes al que estaban asignadas originalmente. Por ejemplo, si una vivienda fue asignada a la segunda quincena de noviembre de 2021, sólo podía cambiar su período de colaboración a la primera quincena de ese mismo mes.

entrevistas estaban organizadas en distintos momentos de la quincena y contemplaron un orden según la temática consultada.

La aplicación de la encuesta en cada hogar requirió de la autorización verbal por parte de una persona integrante del hogar mayor de edad (18 años o más). Para el resto de las entrevistas, las personas informantes idóneas[84] podían variar según cada entrevista o cuestionario, manteniendo como condición que conociera la información necesaria para contestar sobre diversos temas que se solicitan en los cuestionarios.

Por su parte, el cuestionario de Gastos Diarios que es autoadministrado[85], fue el único en que se exigió informante directo. Dicho cuestionario se intentó entregar a cada integrante de 15 años o más del hogar que estuviese en condiciones físicas y/o mentales para responder.

Estos cuestionarios eran entregados al inicio de la quincena y retirados al final del período de colaboración. El llenado de estos cuestionarios fue asistido por quienes encuestaron durante la quincena, brindando apoyo tanto para comprender las preguntas, forma de llenado y, en general, para aclarar cualquier duda que las personas informantes pudieran tener al respecto.

6.3.3 Tecnología de producción de información

La actual versión de la EPF contempló dos tecnologías de captura de información: dos cuestionarios en papel y un cuestionario digital implementado en dispositivo móvil tipo Tablet[86]. De este modo, mientras el cuestionario digital EPF fue recolectado mediante entrevistas presenciales asistidas por computadora (modo CAPI[87]), los cuestionarios de Gastos del Recuerdo y de Gastos Diarios fueron respondidos con papel y lápiz (modo PAPI[88]).

[84] Si bien la autorización verbal era entregada por una persona mayor de edad, el resto de los cuestionarios podía ser respondido por una persona de 15 años o más.

[85] Esto significa que la propia persona informante directa tiene que dar cuenta de sus gastos, a diferencia del resto de los cuestionarios en los que se entrevista de forma presencial por una persona encuestadora.

[86] Tal como se describió en la sección 6.1.1.

[87] Sigla derivada de la expresión en inglés *"Computer Assisted Personal Interview"*.

[88] Sigla derivada de la expresión en inglés *"Paper Assisted Personal Interview"*.

Para el caso del cuestionario digital, su diseño y construcción fue desarrollado con la aplicación *Survey Solutions* (SuSo), software diseñado por el Banco Mundial para el desarrollo de encuestas. Este cuenta con tres conjuntos de servicios distintos, entre ellos, *Designer, Headquarters e Interview,* los que ofrecen una interfaz intuitiva tanto para la creación, diseño e implementación de la encuesta. En la **Tabla 27** se describen, uno a uno, los conjuntos de servicios utilizados, considerando sus principales características y usos en la IX EPF.

Tabla 27: Servicios del sistema *Survey Solutions* y su uso en la IX EPF

Servicios utilizados	Características principales	Usos para ix epf
Designer	**Servicio web destinado al diseño y construcción de cuestionarios digitales. Admite formatos de pregunta de texto, numéricas, categorías selección única y múltiple, ubicación GPS, fecha y hora, entre otras. Permite la creación de estructuras anidadas (módulos, secciones, preguntas) para la organización de preguntas, la configuración de secuencias dependientes de las respuestas ingresadas (condiciones de habilitación), así como la incorporación de alertas y detección de errores de respuesta basadas en reglas de validación (de atributos del contenido de la respuesta y/o su relación con otras respuestas).**	**Construcción iterativa del cuestionario digital por parte de profesionales del equipo técnico de la encuesta. Permitió testear tempranamente distintas estructuras y secuencias de preguntas, para resguardar una mejor usabilidad durante las entrevistas.**
Headquarters	**Servicio web con funcionalidades para la administración de usuarios, gestión de muestra y monitoreo de indicadores clave del proceso de recolección. Contempla distintas funciones de administración de usuarios, perfiles, asignación de cargas de trabajo y recolección mediante formularios web.**	**Aprovechamiento de funcionalidades de administración y gestión, mediadas por el sistema informático desarrollado por el INE, denominado Sistema de Gestión Operativa (SGO), el que las integró(1) y complementó con otras funcionalidades requeridas para la organización y supervisión del proceso.**
Interviewer	**Corresponde a una aplicación que se descarga en el dispositivo móvil de captura, la que permitía la captura de información bajo modalidad CAPI.**	**Esta ofertaba diversas funcionalidades para la gestión de las encuestas como, por ejemplo, monitoreo de carga asignada a las personas encuestadoras, según estado (asignadas, iniciadas, completadas), ejecución de cada una de las entrevistas, y la sincronización de datos con plataforma *Headquarters,* para su revisión.**

(1) El SGO se conectaba a Headquarters mediante una API, esto facilita la integración con el sistema de Survey Solutions, en la asignación y seguimiento de encuestas, lo que permitió la administración, monitoreo y gestión de la coordinación central.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

El cambio tecnológico de papel a cuestionario digital hizo posible mejoras relevantes, en comparación a la anterior versión del estudio, en cuanto a la eficiencia del proceso de producción, el resguardo de calidad de la información recolectada y la detección oportuna de dificultades durante la ejecución de la recolección.

En lo referido a la eficiencia del proceso, el uso de dispositivos móviles para la captura de datos permitió reducir la cantidad de formularios a digitar (en comparación con la anterior versión del estudio), lo que permitió requerir una menor dotación para cumplir esta función. En cuanto al resguardo de calidad de los datos, la disposición de 250 validadores automatizados en los módulos del cuestionario digital mejoró la oportunidad de corrección de información, permitió detectar omisiones y aclarar errores o inconsistencias de información en el momento de entrevista.

A la vez, respecto a la oportunidad en el seguimiento de avance del proceso de recolección, la captura de datos del módulo Hoja de Ruta mediante dispositivos móviles aceleró el ingreso de información de avance de la recolección a los sistemas informáticos, flujo de datos que en otras versiones del estudio dependía de la digitación de esos instrumentos, con el consecuente desfase entre la ejecución de actividades y su reflejo en los indicadores de desempeño operativo. De esta forma, el cambio tecnológico permitió al equipo de coordinación de la encuesta contar con la información de manera más oportuna e inmediata para el monitoreo del avance de la gestión en cada submuestra, característica fundamental para la acción oportuna frente a contingencias o resultados insatisfactorios.

6.3.4 Ejecución de la recolección de una submuestra operativa

Las labores involucradas en la ejecución de la recolección de la IX EPF se organizaron de modo similar a su anterior versión, es decir, las actividades a desarrollar en cada submuestra operativa fueron organizadas en tres fases[89] consecutivas:

[89] Las fases que aquí se identifican asociadas al trabajo de recolección no deben confundirse con la mención a las fases que componen el modelo de producción estadística GSBPM.

- **Preparación**, previa al inicio de la quincena y destinada a generar las condiciones para el contacto de las personas informantes.

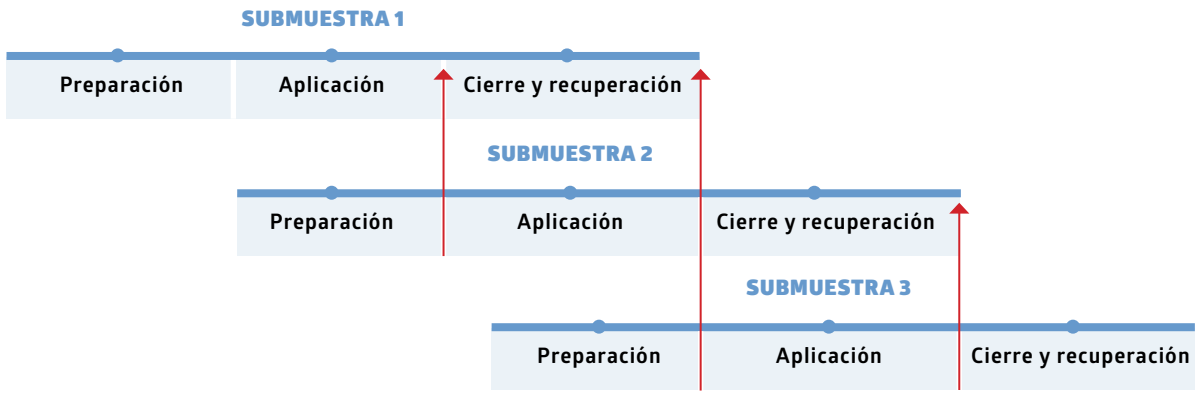
- **Aplicación**, correspondiente al período de colaboración de las personas informantes y destinada a aplicar los distintos cuestionarios.

- **Cierre y recuperación**, posterior a la quincena, en ella se realiza la revisión y recuperación de información faltante.

Cada fase de la ejecución de la recolección contempló una serie de actividades y tareas, orientadas a maximizar los niveles de contacto y cooperación de los hogares, al mismo tiempo que buscaban resguardar la calidad de los datos producidos. De este modo, si bien el período de colaboración de los hogares se extendió durante 15 días aproximadamente, el trabajo completo de la submuestra operativa se extendió por, al menos, 58 días continuos[90].

Los equipos de recolección de datos a nivel nacional realizaron sus actividades de manera simultánea durante las tres fases de recolección en cada una de las submuestras. Esto implicó que existiera una superposición de cada submuestra operativa, puesto que las quincenas se trabajaron de manera consecutiva, sin brechas entre una y otra durante el año calendario completo. Esta superposición de fases es posible observarla en el **Diagrama 7.**

Diagrama 7: Fases de recolección en cada submuestra operativa



Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

[90] Tal como se explicará en las siguientes páginas, la duración del trabajo de una submuestra podía variar según la quincena del mes, los plazos de cierre con los que contaban las regiones y la existencia de planes especiales que modificaban los plazos de las actividades. De este modo, una submuestra operativa podía involucrar hasta 60 días.

A continuación, se presenta una síntesis de las actividades contempladas por cada fase de trabajo, así como sus características más relevantes.

6.3.4.1

Fase de preparación

Su objetivo principal fue lograr el consentimiento de los hogares que habitaban las viviendas seleccionadas en la muestra, de modo que aceptasen participar de la IX EPF durante la quincena.

La fase completa de preparación duró 35 días[91], desde que se recibió e ingresó la submuestra al Sistema de Gestión Operativa (SGO) hasta que se culminaron (idealmente) las tareas del contacto inicial con las unidades muestrales. Sin embargo, la persona encuestadora solo fue responsable de esta última actividad, que tuvo una duración aproximada de 7 días. Las tareas que debía realizar la persona encuestadora incluyeron trabajo en gabinete y terreno. Esta fase incluyó las siguientes actividades:

1. Precarga de la submuestra al sistema informático:
Confirmar que se encontrase cargada la información de las UPM y viviendas seleccionadas para la submuestra en el SGO de la encuesta. Posteriormente, la información podía ser conocida por los distintos grupos de recolección de datos. En específico, entre las tareas involucradas se registró:

a. Revisión de la consistencia del directorio respecto de viviendas seleccionadas.
b. Asignación de submuestra operativa.
c. El ingreso de datos del directorio en el SGO.
d. Asignación de UPM en cascada a cada perfil involucrado en la recolección de datos. Esta fue responsabilidad de la coordinación nacional del proceso de recolección y fue realizada mensualmente.

[91] La fase de preparación contó con algunas actividades realizadas a nivel de submuestra mensual, por ejemplo, la identificación de las unidades muestrales incluidas en el directorio fue realizada para toda la muestra del mes por la persona encargada de grupo. Por lo tanto, fueron 35 días antes del inicio de la aplicación de la encuesta para la segunda quincena y para la primera quincena fueron 20 días aproximadamente.

2. Organización previa del trabajo en terreno:
Requerida para proveer a los grupos de recolección el material necesario para el trabajo de la submuestra operativa, contemplaba la distribución, revisión y organización de los materiales necesarios para realizar el trabajo[92] de una submuestra operativa. Su principal responsable fue la persona encargada de grupo.

3. Reconocimiento de UPM en gabinete:
Requerida/destinada para conocer el emplazamiento, las características y el entorno de las UPM a encuestar en la submuestra. Implicó el reconocimiento de las viviendas asignadas a cada UPM para la submuestra mensual y operativa. Para evitar contratiempos, se recomendó realizar este reconocimiento 30 días antes del inicio del mes en que se entrevistaría la submuestra. Incluyó la difusión del estudio en el sector a encuestar, mediante el pegado de afiches[93] en espacios de alta concurrencia. Esta actividad fue realizada por la persona encargada de grupo con una periodicidad mensual.

4. Identificación de unidades incluidas en el directorio:
Esta actividad permitió verificar directamente la existencia de las unidades seleccionadas en la submuestra. Fue ejecutada por las personas encargadas de grupo, quienes debieron identificar todas las unidades incorporadas en la submuestra y chequear si correspondían a viviendas particulares ocupadas. Si bien esta actividad fue principalmente responsabilidad de la persona encargada de grupo, fue indispensable que la persona encuestadora reiterara estas labores antes de iniciar los contactos en las viviendas, ya que podían surgir dificultades no detectadas previamente.

Dentro de las tareas de la persona encargada de grupo, se incluyó la entrega de una carta institucional dirigida a la jefatura de hogar (carta de apertura[94]), en la que se informaba que la vivienda había sido seleccionada para participar del estudio. Esta carta fue entregada al menos dos semanas antes de que la persona encuestadora acudiera a la vivienda, con el fin de propiciar la colaboración de las personas informantes. En caso de ser necesario, esta carta de apertura se volvió a entregar a algún integrante del hogar, al momento del contacto inicial.

En caso de detectar viviendas seleccionadas al interior de comunidades de acceso controlado (edificios, condominios, etc.), esta actividad incorporó el contacto y entrega de cartas institucionales a administradores y/o representantes de estas comunidades, informando el total de viviendas en que debía aplicarse la encuesta y el período en que se realizaría el proceso.

[92] Consultar el Manual de Recolección de datos Tomo I (INE Chile, 2023a).

[93] Los afiches tenían por objetivo informar la presencia de los encuestadores en el sector, indicando el mes en que estarían aplicando encuestas en ese territorio. Un ejemplo de los afiches distribuidos en el territorio se puede observar en el Anexo 3.

[94] Se puede observar un ejemplo en el Anexo 4.

Como punto de innovación para la IX EPF, se incorporó un formulario de Reconocimiento de UPM[95], que permitía el registro de la gestión de esta actividad, con el fin de entregar reportes del trabajo en terreno a las personas encargadas regionales[96] ya que este instrumento permitía hacer seguimiento y codificar esta etapa.

5. Distribución y asignación de cargas de trabajo:
Implicó definir las viviendas de las que se haría responsable cada persona encuestadora, considerando las características del sector a encuestar y localización territorial. Este proceso debió realizarse, a más tardar, ocho días antes del inicio de la quincena e involucraba asignación informática de cargas de trabajo y entrega de directorios, cartografía, material de difusión y cuestionarios a los encuestadores. El responsable de esta actividad era la persona encargada de grupo.

6. Contacto inicial con la unidad muestral:
Consistía en la visita de las personas encuestadoras a las viviendas asignadas en la muestra, para persuadir a los hogares residentes a que participaran del estudio. Esta actividad se inició con al menos siete días de anticipación al primer día de la quincena y su principal responsable era la persona encuestadora. Durante esta actividad también estaba permitido concertar citas de apertura con las personas informantes para otro día cercano al inicio de la quincena de colaboración[97].

Si durante el contacto inicial se lograba motivar la cooperación de las personas informantes, estaba permitido realizar la entrevista del RPH, también conocida como apertura de vivienda. Adicionalmente, los dos días previos al inicio de la quincena, estaba permitida la entrega de GD a las personas que tuvieran 15 años o más y fuesen integrantes del hogar.

Ante viviendas con personas moradoras ausentes, quienes encuestaban debieron realizar al menos siete intentos de contacto, en distintos días y horarios, incluyendo días de fines de semana, antes de desistir de generar un contacto con el hogar. A su vez, cuando no era posible conseguir la participación de las personas informantes en la encuesta debido a rechazos, la vivienda era reasignada a otra persona encuestadora para reiniciar los intentos de contacto y el logro de la entrevista (esta acción será explicada con mayor detalle en el siguiente apartado).

Dada la normativa vigente, el INE no está facultado para entregar incentivos

[95] Formulario de Reconocimiento de UPM, también se encontraba en formato digital, no obstante, constituyó un instrumento separado del cuestionario EPF digital.

[96] El organigrama será revisado en la sección 6.4.1.

[97] Tal como se explicará en el esquema de entrevistas, la apertura solo podía realizarse hasta el quinto día de la quincena de colaboración.

económicos u obsequios a las personas informantes[98], por lo cual las labores de persuasión fueron realizadas utilizando como medio una segunda carta informativa (idéntica a la entregada previamente), más un díptico informativo[99], sumado a la identificación de quiénes formaron parte del equipo de recolección como personas funcionarias del Instituto Nacional de Estadísticas (mediante credencial y vestuario institucional), en conjunto con los argumentos y justificación del estudio sugeridos en capacitaciones y en el Tomo I del Manual de Recolección de Datos (INE Chile, 2023a). Se debe considerar que la actual versión de la EPF tuvo la complejidad de ser recolectada en período de pandemia, por lo que el proyecto consideró para cada equipo de recolección un kit de elementos de protección personal, para la realización de sus actividades en terreno y en los hogares[100].

6.3.4.2
Fase de aplicación

La fase de aplicación tuvo como objetivo producir, junto con los hogares residentes en las viviendas seleccionadas, la información necesaria para cumplir con los objetivos del proyecto. Esta fase, tuvo como tiempo de duración 15, 16 o 17 días, es decir, un día más que el período de colaboración del cuestionario de GD (quincena[101]), a fin de aumentar la probabilidad de registro de los Gastos Diarios del último día de referencia.

Para aquellos casos en que se logró la participación de los hogares, el esquema fue organizado en cuatro entrevistas durante la quincena de colaboración del hogar, con el fin de evitar el agotamiento de las personas informantes, dar apoyo en el registro del cuestionario de GD y

[98] Esta situación difiere respecto a lo que realizan otras Oficinas Estadísticas, en que la entrega de incentivos monetarios ha logrado mejorar los indicadores de desempeño operativo. Al respecto, es ilustrativo revisar la experiencia del INE de España donde a partir del año 2007 introdujeron incentivos a los hogares participantes a través de la entrega de tarjetas de regalo (INE España, 2016).

[99] Díptico informativo, presentado en Anexo 5.

[100] Este kit de elementos de protección personal (EPP), consideraba alcohol gel, mascarilla, toallas desinfectantes, protector facial.

[101] Dado que las viviendas se asignaban a una quincena de recolección de información, estas podían contener 14, 15 o 16 días dependiendo si el mes contaba con 28, 30 o 31 días.

asegurar la oportunidad de recuperar información faltante o imprecisa en los primeros instrumentos aplicados.

Cada una de estas entrevistas tenía períodos de realización predefinidos por un calendario de referencia[102]. El calendario se estructuró según los plazos fijados en el esquema de entrevistas, que se expone en la **Tabla 28.** La secuencia de las entrevistas fue organizada en un orden que permitiese consultar desde la información más general (información sociodemográfica), pasando por la diversidad de gastos realizados por el hogar, hasta llegar a la información más sensible (ingresos y gastos financieros). De este modo, se buscó disminuir el impacto negativo de consultar información cuya privacidad es especialmente protegida por las personas, como es el caso de los ingresos o deudas.

Cada cuestionario se recolectó en una entrevista específica. En la primera entrevista, por ejemplo, el cuestionario GD era entregado y se explicaba cómo registrar la información, para posteriormente hacer seguimiento de su llenado durante las entrevistas siguientes. El esquema de entrevista fue construido de tal forma que cada cuestionario debía realizarse en un momento específico, recuperando información necesaria en el siguiente contacto.

Las entrevistas contaron con un plazo de aproximadamente cuatro días para ser realizadas, lo que buscaba generar un equilibrio entre la necesidad de llevar un avance ordenado en el trabajo de recolección y la flexibilidad requerida para adaptarse a la disponibilidad de tiempo de las personas informantes. A medida que se realizaban las entrevistas, esta gestión era registrada en la Hoja de Ruta que, tal como se señaló en el apartado 6.1.2, se encontraba integrada y disponible en el cuestionario digital.

Previamente a cada entrevista se realizó un trabajo de gabinete, proceso que consistió en la preparación del material de entrevista, registrando la información muestral de los cuestionarios en papel a entrevistar (GR) o entregar (GD). El propósito fue chequear, a partir de la información capturada en RPH, algunos filtros de cada cuestionario, para identificar

[102] Un calendario de referencia consistía en la publicación de fechas de inicio y término del plazo para realizar cada actividad del trabajo de recolección en una submuestra.

qué preguntas correspondía aplicar a cada persona del hogar y verificando la existencia de preguntas con necesidad de aclaración o corrección. El trabajo de gabinete posterior a cada entrevista consistió en la revisión de la información capturada y registro de observaciones explicativas.

Tabla 28: Esquema de entrevistas del proceso de recolección

Preparación	Aplicación				Cierre
	Entrevista 1	Entrevista 2	Entrevista 3	Entrevista 4	
Mes previo al inicio de la quincena	*Desde 2 días antes hasta día 5 de la quincena*	*Desde día 6 hasta día 9 de la quincena*	*Desde día 7 hasta penúltimo día de la quincena*	*Día siguiente al fin de la quincena*	*6 días siguientes al fin de la quincena*
Precarga de muestra en el SGO	Identificación de hogares en la vivienda	Recuperación RPH	Recuperación GH	Recuperación IN	Última revisión
Asignación de cargas de trabajo	Entrevista RPH	Entrevista GH	Recuperación de GR	Recuperación de GSF	Recuperación de datos
Reconocimientos Unidades muestrales	Entrega y explicación de GD	Entrevista GR	Entrevista IN	Seguimiento y retiro de GD	Orden de material
Entrega de cartas de apertura		Entrega Hoja Auxiliar GR(1)	Entrevista GSF		Despacho a Digitación
Contacto inicial		Seguimiento de GD	Retiro de Hoja Auxiliar GR		
			Seguimiento de GD		

(1) La Hoja Auxiliar (era entregada en la segunda entrevista y retirada en la tercera) fue considerada como un apoyo para las personas informantes en relación con el cuestionario de Gastos del Recuerdo, y la facilitación de su llenado. Su contenido consiste en una tabla tipo donde era posible declarar productos o servicios adquiridos por el hogar, registrando cantidad, descripción, valor total, identificación de compra por internet, tipo de establecimiento y fecha de compra (mes/año). Este instrumento se dejaba en el hogar para facilitar la declaración de gastos consultados en el cuestionario de GR por parte de personas integrantes que no participaron de aquella entrevista.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

La información producida fue revisada por las personas encargadas de grupo tras cada entrevista, a fin de supervisar el avance de la fase de aplicación y detectar posibles imprecisiones u omisiones en los cuestionarios. Estas revisiones fueron registradas en el Formulario de Correcciones (instrumento digital alojado en el Sistema de Gestión Operativa) y, a partir de estos datos, las personas encargadas de grupo informaban por escrito a encuestadoras y encuestadores respecto de las instrucciones de corrección y solicitudes de recuperación de datos, las que debían ser ejecutadas en la siguiente entrevista al hogar.

En síntesis, las actividades desarrolladas durante esta fase fueron las siguientes.

- **Identificación de hogares en la vivienda:**
Consistió en la reiteración de la actividad de identificación de unidades incluidas en el directorio y en la identificación de los hogares residentes en la vivienda seleccionada. Esta actividad fue responsabilidad de las personas encuestadoras, quienes se apoyaron en preguntas presentes en el RPH. Cabe aclarar que la identificación de hogares es parte de un protocolo institucional interno que busca resguardar la calidad de los procesos de recolección de datos en las muestras de viviendas[103].

- **Primera entrevista al hogar:**
Implicó la aplicación del RPH y la entrega del cuestionario de GD, el que se entregó a cada persona integrante del hogar, que tuviera 15 años o más y que estuvieracon condiciones físicas y/o mentales aptas para responder aquel instrumento. Esta entrevista marca el inicio[104] de la colaboración con las personas informantes y podría realizarse, a más tardar, el día 5 de la quincena.

Al finalizar la entrevista, entre la persona informante y encuestadora, se acordaba la fecha y hora del siguiente encuentro, y se indicaba por parte de quien encuestaba la necesidad de disponer de comprobantes asociados a los gastos periódicos del hogar (cuentas de agua, luz, gas, mensualidades de establecimientos educativos, gastos comunes, entre otros) para la siguiente entrevista, con el fin de facilitar la recolección de la información asociada a los cuestionarios que se aplicaron en la siguiente oportunidad.

- **Revisión y reasignación de cargas de trabajo:**
Este es un proceso que involucró un análisis y evaluación, por parte de las personas encargadas de grupo, considerando la cantidad de hogares efectivamente aperturados por cada persona encuestadora, y de esta forma realizar

[103] Al respecto revisar en la metodología, el concepto de hogar utilizado por la encuesta.

[104] Es en esta entrevista donde se explica al informante sobre una posible visita de un supervisor o supervisora durante el período. Para mayor información consultar el Manual de Recolección de Datos Tomo I (INE Chile, 2023a).

una reasignación de viviendas[105] entre el equipo de recolección de un mismo grupo. Esto se realizaba en caso de producirse desequilibrios importantes en las cargas de trabajo o rechazos a participar del estudio.

La reasignación de hogares con rechazo a otra persona encuestadora respondió a la necesidad de reiniciar la interacción con los hogares, evitando posibles efectos negativos derivados de la predisposición de personas informantes y/o encuestadoras tras un primer encuentro con resultados negativos. De este modo, la reasignación permitía que un nuevo integrante del equipo de recolección, con características y estilos de persuasión distintos a la persona inicial, realizara nuevos intentos de contacto, buscando adaptarse de mejor modo a las características de la persona informante.

Para revertir situaciones de rechazo, también se emplearon formatos de cartas institucionales personalizadas, en las cuales se solicitaba reconsiderar la decisión de no colaborar con la encuesta, aludiendo a la relevancia de los resultados de la EPF para el IPC y su impacto en las decisiones de política pública. Solo cuando se producía un segundo rechazo se permitía abandonar los esfuerzos de entrevista en un hogar.

- **Segunda entrevista al hogar:**

Involucraba la recuperación de información pendiente del RPH, la aplicación del módulo de Gastos periódicos del Hogar (GH), Gastos en Educación (GE), el cuestionario de Gastos del Recuerdo (GR), y el seguimiento al cuestionario autoadministrado de Gastos Diarios (GD).

El seguimiento al cuestionario de GD consideraba la revisión, día por día, de los gastos registrados, la verificación de la existencia de gastos en los días en que no se produjeron registros y el apoyo en el registro de gastos no consignados previamente. Luego de esta verificación, la gestión se registraba en Hoja de Ruta (HR) en el cuestionario digital, desde donde la persona encargada de grupo realizó el seguimiento a GD.

Adicionalmente, al final de esta entrevista se entregaba la Hoja Auxiliar de GR, formulario destinado a complementar la información declarada en Gastos del Recuerdo, a partir de un autoreporte de gastos comprendidos por ese cuestionario que no hubiesen sido declarados durante la entrevista. De este modo, la Hoja Auxiliar buscaba dar oportunidad a las personas integrantes del hogar ausentes en la entrevista para que registraran los gastos realizados en los últimos tres, seis y doce

[105] La reasignación se realizaba a nivel de viviendas y no de hogares para facilitar el vínculo con los informantes y optimizar los desplazamientos de los encuestadores. Esto quiere decir que, si en una vivienda existían dos hogares, al realizar la reasignación de las cargas, ambos hogares eran entrevistados por el mismo encuestador.

meses consultados por ese cuestionario.
Por último, se fijaba la fecha y horario de la siguiente cita, para la cual se solicitaba la disposición de comprobantes de ingreso (liquidaciones de sueldo, pago de pensiones, entre otros) y de gastos en servicios financieros y seguros (recibos de créditos, cuentas bancarias, etc.), para hacer más expedita la siguiente entrevista[106].

- **Tercera entrevista al hogar:**

Comprendía la recuperación de información faltante en los módulos de Gastos periódicos del Hogar (GH) y Gastos en Educación (GE), además del retiro de la Hoja Auxiliar y recuperación de datos faltantes del cuestionario de GR. Luego de ello, involucraba la aplicación de los módulos de Ingresos (IN) y Gastos en Servicios Financieros y Seguros (GSF). Por último, de la misma forma que en la segunda entrevista, se realizó la revisión y seguimiento al cuestionario de GD, para luego registrar la gestión en la Hoja de Ruta como fue mencionado con anterioridad.

- **Cuarta entrevista al hogar y retiro de GD:**

Incluía la recuperación de datos faltantes en cualquiera de los cuestionarios antes entrevistados. Su principal actividad consistía en la revisión de los datos registrados en el cuestionario GD, complementando descripciones de gasto u otros datos que lo requirieran. Solo se permitía realizar esta entrevista con posterioridad al último día de la quincena.

Al retirar los cuestionarios GD las personas encuestadoras informaban al hogar la posibilidad de que fuera posible volver a contactarles, telefónica o presencialmente[107], para precisar alguno de los datos aportados al estudio.

[106] Si bien se solicitaba y era deseable que las personas informantes pudiesen contar con esta información como apoyo para responder a los módulos, no era condición necesaria para realizar las siguientes entrevistas, ya que se recurría a la declaración de las personas informantes (recordatorio) si no contaban con comprobantes.

[107] Para realizar las recuperaciones, también fue permitido el uso de correo electrónico institucional como medio de comunicación con las personas informantes, no obstante, para permitir su uso se exigía la comunicación telefónica previa con las mismas. Más detalles respecto de este procedimiento pueden conocerse en el Manual de Recolección de Datos Tomo I (INE Chile, 2023a).

6.3.4.3

Fase de cierre y recuperación

Su finalidad fue evaluar y mejorar la calidad del dato producido, para el posterior envío de los materiales a digitación. Tuvo una duración de entre seis y diez días[108], a partir del fin de la fase de aplicación. Las actividades desarrolladas en esta fase eran las siguientes:

1. Revisión previa al cierre de los cuestionarios:
Implicó la revisión, por parte del equipo de personas encuestadoras y encargadas de grupo, de todos los cuestionarios y módulos aplicados a los hogares entrevistados, verificando la consistencia de la información solicitada y el cumplimiento de las instrucciones de corrección. Junto a lo anterior, esta actividad incluía traspasar los gastos de las boletas que habían sido pegadas en GD a sus respectivos días.

En caso de existir información faltante o inconsistente, las personas encargadas de grupo podían solicitar la recuperación de los datos, mediante llamado telefónico o entrevista al hogar. Quienes encuestaron podían intentar la recuperación telefónica en cinco oportunidades. Si no tenían éxito tras esos intentos, necesariamente debían visitar al hogar para recuperar la información.

2. Entrevista de recuperación:
Consistía en un contacto adicional con el hogar, con el fin de recuperar información faltante. Su realización era obligatoria en aquellos casos en que se omitiera la respuesta en más de tres preguntas o en alguna pregunta filtro.

Este proceso exigía realizar al menos tres intentos de contacto en la vivienda. Si, pese a esos esfuerzos, no era posible recuperar la información, debían agregar una observación al cuestionario explicando la situación de la información faltante.

[108] Se otorgaba un plazo mayor de cierre a grupos de recolección con mayor concentración muestral, con el propósito de facilitar el orden en el proceso y otorgar a los encargados de grupo un tiempo suficiente para la revisión final del material.

3. Envío de encuestas a digitación:

Constituía la última actividad realizada en el trabajo de una submuestra operativa e incluía la organización de los cuestionarios aplicados en papel y la elaboración de guías de despacho por parte de los grupos de recolección para la entrega del material en cascada ascendente a las personas encargadas regionales y de área. Estas recibían el material, verificaban la completitud del despacho[109], confirmaban su recepción en el sistema informático y, finalmente, entregaban el material al equipo de digitación.

Respecto a esta actividad, cabe aclarar que el Sistema de Gestión Operativa (SGO) impedía la emisión de guías de despacho en aquellos casos que, de acuerdo con los registros del Formulario de Correcciones, mantuvieran errores sin corregir. Este control informático tenía por objetivo procurar que las personas encargadas de grupo ingresaran una observación que explicara el motivo por el cual no fue posible recuperar la información faltante.

Todos los cuestionarios en papel aplicados en una submuestra operativa eran incluidos en un solo despacho, con el fin de evitar posibles pérdidas de material. Habitualmente, los despachos solo contenían los dos cuestionarios que se recolectaban en papel (GR y GD). No obstante, el proceso de recolección contempló planes de contingencia, en que era permitida la entrevista del resto de instrumentos en tecnología papel y lápiz, datos que posteriormente eran ingresados por las personas encuestadoras antes del cierre de la submuestra. En estos casos, los cuestionarios de contingencia en papel también se adjuntaban en la guía de despacho[110]. Asimismo, la entrega del material al equipo de digitación procuraba la entrega de submuestras operativas completas a nivel nacional, resguardando que todo el material estuviese físicamente en un mismo lugar, para mayor orden en la administración de las encuestas aplicadas.

[109] La verificación de la completitud del despacho se realizaba a través de la lectura de los códigos de barra presentes en cada cuestionario, con lo cual se comprobaba que todos los cuestionarios aplicados en un hogar se hubiesen incluido en el despacho.

[110] Es importante señalar que los cuestionarios de contingencia no ingresaban al proceso de digitación habitual, pero sí se registraba su uso en planillas de gestión, que permitían llevar un control de la cantidad de cuestionarios utilizados, en qué región y el motivo de la incidencia que gatilla el uso.

6.3.4.4

Planes especiales implementados durante la recolección

Con el fin de garantizar un proceso detallado y estándar, se implementaron estrategias de organización del trabajo de recolección, teniendo en cuenta las posibles contingencias que pueden surgir durante el trabajo en terreno. Sin embargo, la presencia de eventos nacionales importantes, tales como las fiestas patrias o períodos de vacaciones, requirió de ajustes en el trabajo realizado en algunas submuestras, lo que llevo la implementación de planes especiales.

Estos planes especiales fueron desarrollados por la coordinación nacional de recolección, en colaboración con el equipo de gestión central de la encuesta. Su principal objetivo fue salvaguardar la calidad de los datos y los niveles de cooperación de las personas informantes, y así dar una respuesta estándar ante situaciones de carácter nacional que afectaban el normal desarrollo de la recolección de una o varias submuestras. Los planes especiales adoptados y sus motivos se presentan en la **Tabla 29** y **Tabla 30.**

Tabla 29: Períodos y motivos de planes especiales

Plan especial	Submuestras Operativas Involucradas	Mes calendario	Motivo
Verano 2021-2022	06-07-08-09-10	diciembre 2021 - febrero 2022	Potencial disminución de colaboración de informantes por fiestas de fin de año (navidad y año nuevo) e inicio de período estival, comúnmente utilizado para vacaciones.
Invierno 2022	18-19-20-21	junio – agosto 2022	Potencial disminución de colaboración de informantes por período de vacaciones escolares de invierno.
Septiembre 2022	22-23-24	agosto – septiembre 2022	Aminorar el impacto del plebiscito constitucional y la posibilidad de ausencia de informantes, debido a feriados de fiestas patrias y semana de vacaciones en algunos establecimientos educacionales.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Se entregaron informativos a nivel nacional en cada uno de estos hitos para informar sobre los ajustes propuestos para hacer frente al período de recolección correspondiente. La implementación de los planes era respaldada por las personas encargadas regionales, quienes intensificaban su monitoreo durante estos períodos y contribuían a resolver cualquier situación que afectaran la recolección en caso de contingencia. El tipo de medidas de ajuste adoptadas en los planes especiales se presentan en la **Tabla 30.**

Tabla 30: Tipos de medidas adoptadas en planes especiales

Ajuste	Descripción	Submuestras operativa	Plan especial
Contactar las viviendas de ambas quincenas antes del inicio del mes	Realización de los contactos iniciales de ambas submuestras operativas del mes durante el período de preparación de la primera quincena. Se consultó a los hogares entrevistados sobre viajes u otra actividad planificada que obstaculizaran la aplicación de la encuesta[1].	06 - 07 - 08 - 09 - 10 - 18 - 19 - 20 - 21 - 22 - 23 - 24	Verano 2021-2022 Invierno 2022 septiembre 2022
Solicitar cambios de submuestra operativa para los hogares que se ausentarán durante la quincena asignada	Ante la presencia de un viaje u otras actividades se solicitaba el cambio de submuestra operativa de la vivienda. Los cambios debían ser autorizados antes de la aplicación de la encuesta, es decir, como máximo el día 8 del mes. Los cambios solo podrán ser en el mismo mes.	06 - 07 - 08 - 09 - 10 - 18 - 19 - 20 - 21 - 22 - 23 - 24	Verano 2021-2022 Invierno 2022 septiembre 2022
Emplear cartas especiales para revertir rechazos	Junto a los instructivos se enviaron las cartas para revertir rechazos iniciales (durante apertura) y rechazos durante la aplicación de la encuesta. Las personas encargadas de grupo debían personalizar los datos de las cartas en los casos de rechazo y gestionar la impresión de la carta. Las cartas debían ir firmadas por la jefatura de operaciones o en el caso de la RM por el encargado nacional de recolección. Las cartas buscaron una interacción más personalizada con los hogares, que a la vez enfatizaron la confidencialidad y anonimato de la información entregada. La carta para rechazos iniciales insiste en la obligación de entregar información al INE, mientras que la carta para rechazos durante la aplicación enfatiza la importancia del hogar seleccionado en la representatividad de los datos.	06 - 07 - 08 - 09 - 10 - 18 - 19 - 20 - 21 - 22 - 23 - 24	Verano 2021-2022 Invierno 2022 septiembre 2022

Ajuste	Descripción	Submuestras operativa	Plan especial
Gestiones adicionales para revertir rechazos	Adicional a las cartas de rechazo iniciales o rechazos de hogares aperturados, la persona encargada de grupo realizaba gestiones tales como visitas en terreno a las personas informantes de las unidades seleccionadas, envíos de correo electrónico y/o llamados telefónicos. Dependiendo de la cantidad de unidades rechazadas se solicitó apoyo a supervisión para revertir rechazos.	06 - 07 - 08 - 09 - 10 - 18 - 19 - 20 - 21 - 22 - 23 - 24	Verano 2021-2022 Invierno 2022 Septiembre 2022
Flexibilizar la cantidad de entrevistas, uniendo la tercera y cuarta entrevista	En casos particulares y por medio de una evaluación, se flexibilizó la cantidad de entrevistas realizadas a los hogares, uniendo las entrevistas 3 y 4 del esquema. Esto se podía realizar solo en los casos en que la apertura se realizó en las fechas de ampliación del plazo de la apertura o si fue solicitado por el informante. Se revisaba la cantidad de días con registros declarada por las personas encuestadoras en el seguimiento de GD para aceptar unir las entrevistas. Se estableció como requisito mínimo, contar con al menos 5 días de registros declarados en la revisión de la segunda entrevista. No se autorizó agrupar otras entrevistas distintas a las señaladas.	06 - 07 - 08 - 09 - 10 - 18 - 19 - 20 - 21 - 22 - 23 - 24	Verano 2021-2022 Invierno 2022 Septiembre 2022
Ampliación de plazo de apertura	Se autoriza la realización de aperturas de viviendas de la submuestra hasta el séptimo día de la quincena. La recomendación de la coordinación nacional fue intentar aperturar todas las viviendas al quinto día de la quincena, por lo que el plazo del séptimo día fue para casos excepcionales. Se solicitó intentar la recuperación de gastos previos al día de entrega de gastos diarios o al menos intentar recuperar dos días anteriores a la primera entrevista.	07 - 08 - 09 - 10 - 20 - 21 – 23 - 24	Verano 2021-2022 Invierno 2022 Septiembre 2022
Adelantamiento de la cuarta entrevista	Se autoriza el retiro de GD en viviendas días antes del cierre de la submuestra. El objetivo es ajustarse a la disponibilidad de informantes, en casos donde el hogar no podrá colaborar.	06 - 07 - 08 - 09 - 10 - 18 - 19 - 20 - 21 - 22 - 23	Verano 2021-2022 Invierno 2022 Septiembre 2022
Ampliación de plazo para cuarta entrevista	Se contó con un plazo máximo de siete días hábiles para la cuarta entrevista de la submuestra, la que consideró la recuperación de datos pendientes de los módulos de Ingresos y Gastos en Servicios Financieros, junto con el retiro de GD. El objetivo es ajustarse a los desplazamientos de los informantes y descomprimir los últimos días de la apertura de la siguiente submuestra operativa.	06 - 19	Verano 2021-2022 Invierno 2022

Ajuste	Descripción	Submuestras operativa	Plan especial
Ampliación de plazo para despacho de la submuestra operativa	Dada la estrategia anterior, el mayor plazo otorgado para la cuarta entrevista implicó dar tres días más de plazo para el despacho de la submuestra al equipo de digitación.	06 - 18	Verano 2021-2022 Invierno 2022
Adelantar entrega del cuestionario de Gastos Diarios	Se autorizó la entrega de GD en viviendas de la submuestra desde cinco días antes del inicio de la quincena. A quienes planificaron entregar el cuestionario de Gastos Diarios con anticipación, se les solicitó confirmar con los informantes el inicio del registro de sus gastos diarios. Además, se coordinó la segunda entrevista para los primeros días establecidos en el calendario, referente a la segunda entrevista, esto con el fin de hacer seguimiento al llenado de Gastos Diarios.	07	Verano 2021-2022

(1) Ante esta instrucción, se solicitaba a las personas encuestadoras tener cuidado en el modo de preguntar, evitando generar desconfianza en los informantes.
Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Es necesario mencionar que los ajustes señalados buscaron adaptarse a aquellos casos que escapaban de la normalidad en donde se presentaron mayores dificultades de contacto y cooperación, por lo que no fueron aplicadas a la totalidad de las viviendas pertenecientes a las submuestras citadas en la **Tabla 30.**

6.3.5 Indicadores de desempeño operativo

A partir de la información capturada en la sección de Registro de Contacto de la Hoja de Ruta, durante todo el proceso de recolección fue posible clasificar el resultado de la encuesta en cada unidad de la muestra. Esta información fue utilizada para calcular indicadores de desempeño operativo, los que corresponden a estadísticos que permitieron la evaluación y seguimiento del proceso de recolección en un momento determinado.

Para el cálculo de los indicadores se distinguió entre una clasificación inicial, realizada en la apertura del hogar de una submuestra operativa, y una clasificación preliminar, realizada al finalizar la fase de cierre de una submuestra[111]. La distinción entre la clasificación inicial y preliminar radicaba en la definición de unidades entrevistadas y el tipo de código considerado (códigos de resultado de la entrevista o código de disposición de la vivienda).

La clasificación inicial se realizó considerando el mejor resultado del contacto con la vivienda y estableció como entrevistadas aquellas unidades que decidieron participar en el estudio (apertura del hogar) y que, por lo tanto, se hubiese aplicado el RPH. A diferencia de lo anterior, en la clasificación preliminar se consideró el código de disposición de la vivienda y se definió como entrevistadas las unidades que tuviese aplicado el RPH y que, además, contaran con al menos un cuestionario de Gastos Diarios retirado y con días de registro válidos. Los grupos de clasificación de códigos de disposición, según tipo de resultado de la entrevista y momento de evaluación se presentan en la **Tabla 31.**

[111] Se establece que la clasificación final del resultado de la encuesta solo es posible tras finalizar el trabajo de recolección y procesamiento de datos, una vez realizada la evaluación de la calidad de los datos a través de la aplicación de la grilla técnica de mínimos de calidad y cantidad.

Tabla 31: Clasificación de códigos de disposición de la unidad según resultado en Hoja de Ruta y Clasificación final según criterios de la grilla técnica

Categoría según pertenencia a población objetivo	Categoría según nivel de contacto	Clasificación	Sigla	Código de Hoja de Ruta		Clasificación final según criterios de grilla técnica[1]
				Inicial *(durante trabajo de submuestra operativa)*	Preliminar *(tras cierre de submuestra operativa)*	
Elegible	Contactadas	Entrevista completa	E	11	11 - con al menos una GD	E-Cumple con los criterios de calidad
	Rechazadas	Entrevista interrumpida (*break off*)	BO	-	11- sin GD	BO- No cumple con los criterios de calidad
		Rechazo inicial	RI	21	21	RI
	Otras no entrevistadas por otros motivos	Otras no entrevistadas	O	41-42-43-44-49	41-42-43-49	O
	No contactada	No contactadas	NC	31-32-33-34	31-32-33-34	NC
Elegibilidad desconocida			ED	52-53-54-59	52-53-54-59	ED
No elegible			NE	61-62	61-62	NE

(1) La evaluación de la grilla técnica de mínimos de calidad es una etapa que se aborda en la fase de procesamiento y se describe en la sección 7.3.3 Revisar y validar.

Nota: Si bien el código 44 "Concertó cita con el informante" no es parte de los Códigos de Disposición Final, fue considerado como no contacto en la clasificación inicial para permitir el cálculo de los indicadores de desempeño operativo sobre el total de las unidades muestrales.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Este modo de clasificación de códigos, según tipo de resultado y momento en que se realizaba la evaluación, resultó de utilidad para disponer de indicadores de desempeño operativo que permitieron diagnosticar detalladamente el avance de las labores de una submuestra en cada región y tomar decisiones oportunas para mejorar los resultados de la encuesta. Las categorías de disposición final se presentan en la siguiente tabla.

Tabla 32: Clasificación de viviendas según tipo de código de disposición final

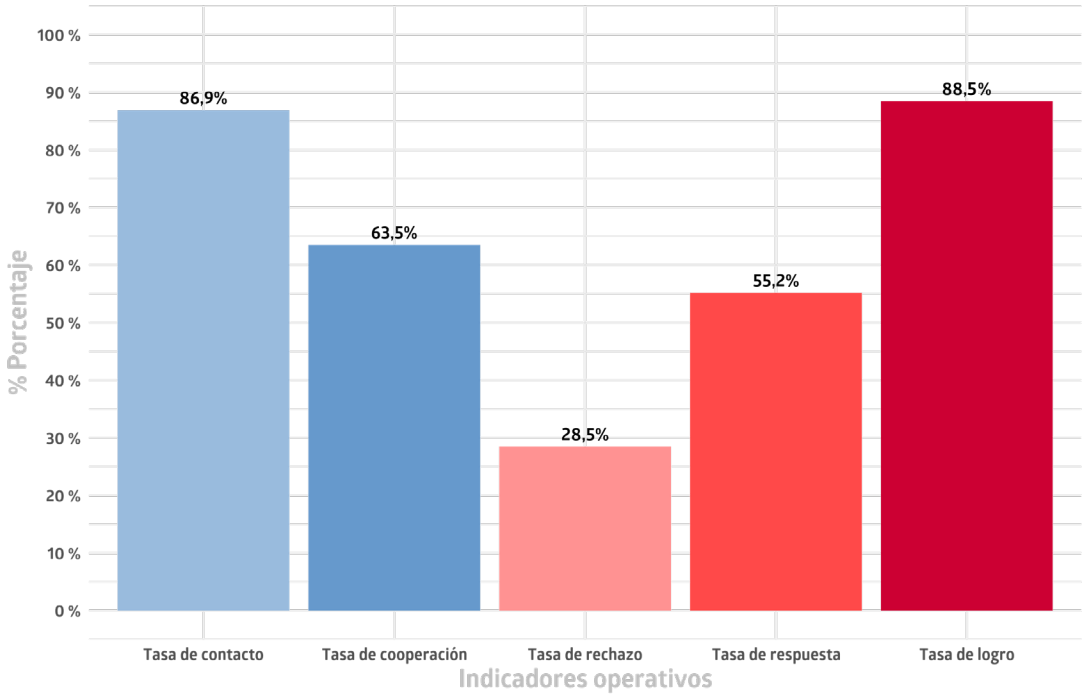
Clasificación de CDF	Cantidad de unidades	Porcentaje del total
Entrevistadas (E)	14.961	51,2%
Entrevista interrumpida[1] (BO)	3.092	10,6%
Rechazo Inicial (RI)	4.635	15,9%
Otros no contactados (O)	873	3%
No contactados (NC)	3.321	11,4%
Elegibilidad desconocida (ED)	238	0,8%
No elegible (NE)	2.111	7,2%
Total	29.231	100%

(1) Tanto el break off (BO) como el rechazo inicial (RI) corresponden a la misma categoría de Rechazo (R), se presentan de forma desagregada para efectos del estudio.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Finalmente, la clasificación de las unidades de la muestra permitió resumir la información disponible en tasas o indicadores de desempeño operativo para hacer seguimiento del estudio **(Gráfico 3).**

Gráfico 3: Principales indicadores de desempeño operativo (IDO) según los códigos de disposición final



Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

6.3.6 Supervisión de la recolección

Durante la recolección de datos de la operación estadística IX EPF, se implementaron medidas de aseguramiento de la calidad que permitieron controlar y reducir los errores no muestrales, contribuyendo a evitar sesgos que pudieran interferir en la interpretación de los datos del estudio.

Existen principalmente dos tipos de errores que afectan la precisión y confiabilidad de una encuesta: los que provienen del diseño muestral y los ajenos al muestreo. En el caso de los últimos, se producen por factores ajenos al diseño muestral que afectan la precisión y confiabilidad de las estimaciones. Pueden tener más de una fuente y es posible clasificarlos como errores de cobertura, errores de no respuesta, errores de medición y errores de procesamiento. Los errores de medición están originados por una inadecuación en las respuestas recogidas por el instrumento de captura de datos, y se pueden clasificar en cuatro factores: errores producidos por las personas encuestadoras, por aquellas que informan, errores debidos al cuestionario y errores producidos por el método de recolección de datos (Díaz de Rada, 2004).

Los procedimientos de la etapa de supervisión de la IX EPF tuvieron como objetivo disminuir los errores de medición, mediante su identificación en la ejecución de la recolección de la encuesta, y la retroalimentación a las personas encuestadoras con respecto a las desviaciones producidas en el método de la recolección de datos.

Como característica principal, los equipos de supervisión del INE se configuran de forma independiente de los equipos de recolección de datos, es decir, el personal contratado para ejecutar la etapa de supervisión pertenece a un equipo de trabajo separado en funciones y dependencia de aquel que ejecuta la recolección de datos.

Esta distancia funcional y organizacional está orientada a una mayor independencia en las tareas de control de calidad de la encuesta, buscando producir mayores incentivos en la identificación y verificación de inconsistencias, como también identificar oportunidades de mejora en los procedimientos de recolección ejecutados. Este control de calidad de las encuestas se adiciona al realizado por las personas Encargadas

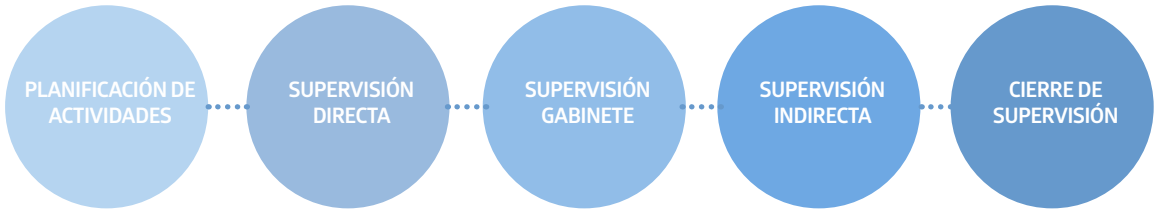
de Grupo de recolección, quienes, además de gestionar al equipo de personas recolectoras, debían revisar diariamente las encuestas completadas, identificando errores en los cuestionarios y finalmente, dar retroalimentación al personal de recolección.

Los objetivos que se definieron para el equipo de supervisión fueron los siguientes:

- Verificar el cumplimiento de los protocolos según la metodología y procesos definidos para cada actividad de la recolección de datos.
- Identificar situaciones periódicas y de excepción en la aplicación de la encuesta, buscar soluciones y estandarizar criterios para su mejora continua.
- Generar estrategias de comunicación para las personas recolectoras con los hogares, detectando y corrigiendo eventuales desviaciones de los objetivos de la encuesta.
- Retroalimentar a los equipos de recolección en temas que se identifiquen a partir de los hallazgos y que deben ser reforzados.
- Apoyar los procesos de capacitación necesarios para la inducción de nuevas personas integrantes del equipo de trabajo de recolección, durante el proceso de aplicación de la encuesta.

El proceso de supervisión se dividió en las siguientes etapas durante su ejecución: Planificación de actividades, supervisión directa, supervisión de gabinete, supervisión indirecta y cierre de supervisión (retroalimentación), donde cada una de ellas se detallan a continuación.

Diagrama 8: Etapas del proceso de supervisión



Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

6.3.6.1

Planificación de actividades de supervisión

La planificación proporcionó operatividad a la etapa de supervisión, realizando actividades de coordinación de cargas de trabajo y coordinación con los equipos de recolección para la correcta ejecución de las actividades de supervisión.

La coordinación operativa del proyecto a través de las personas encargadas nacionales de recolección y supervisión respectivamente, desarrollaron una calendarización del proceso de recolección y supervisión, con el fin de adecuar las actividades de trabajo en la recolección de datos y facilitar las instancias de control de calidad de la encuesta.

La persona encargada nacional de supervisión mediante correo electrónico notificó la calendarización del proceso de supervisión a las personas encargadas de área de supervisión, las cuales debieron revisar la planificación y estimar las cargas de trabajo de sus áreas de responsabilidad. Para ello, debían acceder al sistema de gestión operativa (SGO) y revisar los indicadores de desempeño operativo del equipo de recolección de datos. Esto, con el fin de identificar unidades y personas recolectoras que requirieron una supervisión con base en un análisis y criterios establecidos por el equipo de supervisión de acuerdo a los resultados de desempeño operativo.

Tabla 33: Criterios de asignación de unidades a supervisión

Tipos de Supervisión	Criterios	Categorías	Estado	Código de visita
Directa Apertura	Aplicar a Encuestadores(as) nuevos	Elegibles	Unidades asignadas sin contacto con personas informantes	No Aplica
	Encuestadores(as) con Tasa de rechazo sobre 40%			
	Al menos una vez por encuestador(a)			
	Selección Aleatoria			
Directa Aplicación	Folios con módulos con mayor % de errores y/o inconsistencias indicadas por el Encargado(a) de Grupo (EG)	Elegibles	Unidades aperturadas	11
	Solicitud de reforzamiento del EG de un módulo específico			
	Al menos una vez por encuestador(a)			
	Selección Aleatoria			
Indirecta	Folios con módulos con mayor % de errores y/o inconsistencias indicadas por el EG	Elegibles	Unidades aperturadas + 2 módulos adicionales distintos al RPH o Encuestas completadas aprobadas por Encargado(a) de Grupo (Retiro MGD)	11
	Solicitud directa del Encargado(a) de Grupo (EG)			
	Sospecha de falseamiento			
	Encuestadores(as) con Tasa de respuesta superior al 50%			
	Selección Aleatoria			
Hoja de Ruta	Encuestadores(as) con Tasa de respuesta superior al 50%	Elegibles, No elegibles, Elegibilidad Desconocida	Unidades gestionadas	11, 21, 31, 32, 33, 34, 41, 42, 43, 44, 49, 52, 53, 54, 59, 61, 62, 63, 64, 65, 66, 67 y 69
	Selección Aleatoria			

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

El porcentaje y frecuencia de unidades a supervisar, varió según la experiencia de aquellos que entrevistan y su antigüedad en el trabajo de recolección de datos. Cuando los entrevistadores no tenían experiencia, o acababan de incorporarse al instituto, convenía sostener reuniones frecuentes al inicio de la recolección (cada uno o dos días) y revisar las rutas con el fin de detectar rápidamente los errores y evitar que algunos de estos se convirtieran en hábitos[112].

De acuerdo con las recomendaciones internacionales y los estándares definidos en otras operaciones estadísticas ejecutadas por el INE, se estableció para la IX EPF una cobertura de supervisión como mínimo de un 20% de los hogares entrevistados de la muestra. Este porcentaje se desagregó en supervisiones directas en terreno con las personas encuestadoras y la supervisión indirecta asociada una re-entrevista[113] con las personas informantes para la verificación de la correcta aplicación de la encuesta. Adicionalmente se definió una meta de supervisión del 20% de los hogares que no lograron ser contactados y un 20% de las unidades no elegibles del total de la muestra.

6.3.6.2
Supervisión Directa

La supervisión directa fue el procedimiento donde se evaluó el cumplimiento de los protocolos de apertura y procedimientos de aplicación, a partir de la observación del trabajo realizado por la persona recolectora. Consistió en acompañar en el momento de la entrevista, a cada una de las personas encuestadoras en terreno, con la finalidad de evaluar la correcta aplicación de los protocolos de trabajo, el registro de los módulos de la encuesta y la interacción con las personas informantes. Debía realizarse, como mínimo, una vez a cada persona encuestadora durante el período de recolección de datos, de acuerdo con lo solicitado por la persona encargada del área de supervisión.

[112] La guía ESOMAR para la armonización de las normas sobre el trabajo de recolección recomienda, revisar completamente los tres primeros trabajos de los entrevistadores nuevos o sin experiencia.

[113] Es una medición repetida de una misma unidad en una encuesta que se aplica mediante entrevistas que permiten volver a formular las mismas preguntas de la entrevista inicial o una selección de ellas.

La supervisión directa contó con dos instrumentos: el "Formulario de Supervisión Directa - Apertura" y el "Formulario de Supervisión Directa – Aplicación". El formulario utilizado dependía de la entrevista a supervisar, el primero, fue utilizado en la primera entrevista con el hogar, mientras que el segundo fue utilizado en segunda y/o tercera entrevista.

El "Formulario de Supervisión Directa - Apertura" estuvo compuesto por: 1) Registro de contacto con el hogar, que permitía monitorear la gestión en terreno realizada por la persona supervisora; 2) Apertura, este apartado consistió en observar la presentación de la encuesta a la persona informante y las indagaciones realizadas para identificar la vivienda y hogares en la unidad seleccionada; 3) Llenado RPH; este consideraba dos aspectos a evaluar en cada una de las secciones del módulo, la lectura de las preguntas y si se entregaban detalles en caso de ser necesario; 4) Antecedentes generales de la entrevista, estaba compuesto por una selección de preguntas que evaluó diferentes aspectos de la aplicación y 5) Resultado de supervisión y observaciones, este permitió clasificar el estado de la supervisión y registrar observaciones cuando estimara la persona supervisora.

En cuanto al "Formulario de Supervisión Directa – Aplicación", también contó con 1) Registro de contacto con el hogar cumpliendo el mismo objetivo mencionado en Directa Apertura y adicionalmente el formulario contaba con la posibilidad de habilitar los módulos y cuestionarios que se aplicaban en segunda y tercera entrevista que eran: 2) Llenado de GH; 3) Llenado de GE; 4) Llenado de GR; 5) Llenado de IN; 6) Llenado de GSF, en cada uno de estos se evaluó la lectura de preguntas y la entrega de detalles por la persona encuestadora cuando era necesario; como también contenía un apartado de 7) Antecedentes generales de la entrevista y 8) Resultado de supervisión y observaciones.

La supervisión Directa de Apertura y Aplicación se realizaron de manera continua en cada submuestra operativa, para cada una de ellas estaban establecidos los criterios de asignación de unidades a supervisar. La supervisión directa, se realizó con énfasis en aquellas personas recolectoras sin experiencia en la recolección de datos y de acuerdo a los resultados operativos obtenidos. Los acompañamientos fueron definidos con anticipación, previa coordinación con el equipo de recolección.

Con respecto a la aplicación de la supervisión directa, en cualquiera de sus opciones (segunda o tercera entrevista), la persona supervisora debió mantener un rol de observación, incluso en aquellos casos en que la persona recolectora cometiera un error u omisión en la aplicación. Para los casos en que se identificaban oportunidades de mejora durante la aplicación, se contempló un espacio de retroalimentación a encuestadoras y encuestadores ya finalizada la entrevista y sin presencia de quien informara. De esta manera se resguardaba que la persona encuestadora no fuera corregida frente a la persona informante. Sin embargo, se consideraron cuatro excepciones donde la persona supervisora pudo suspender su rol de observación:

- Cuando la persona supervisora observó que la persona encuestadora se equivocaba en la identificación de la unidad muestral a encuestar.
- Cuando la persona encuestadora planteó directamente alguna pregunta a quien supervisó respecto al protocolo o aplicación de instrumentos. En esta situación, la persona que supervisa debió responder de manera precisa a partir de la información disponible o, en caso de desconocer la respuesta, plantear la posibilidad de ser respondida en la siguiente entrevista.
- Cuando la persona supervisora observó que, de no realizar alguna intervención entre quien encuesta y la persona informante, el hogar rechazaría la aplicación de la encuesta. Estas intervenciones debieron siempre tener una lógica de mediación para facilitar la apertura o continuidad de la colaboración del hogar.
- Cuando la persona encuestadora alteró de forma evidente las tareas programadas para ese contacto, adelantando u omitiendo labores establecidas por el esquema de entrevistas, sin que ello haya sido informado y justificado previamente con su encargado.

El rol de observación culminaba junto con la entrevista al hogar y su interacción con la persona informante. Era obligatorio al finalizar la entrevista de supervisión directa de apertura o aplicación, la persona supervisora retroalimentara verbalmente a la persona encuestadora respecto a sus fortalezas y aspectos de mejora del proceso. Adicionalmente, quien supervisaba, debía revisar la completitud de su formulario en gabinete y sincronizar su dispositivo móvil de captura.

Utilizando el Formulario de supervisión como insumo, la persona supervisora generó un Informe de Supervisión Directa que tuvo como objetivo el registro de la observación de la aplicación de la entrevista realizada por la persona recolectora. Este informe se estructuro en tres momentos distintos de la entrevista, la apertura del hogar, conocimientos de los módulos del cuestionario y antecedentes generales, en donde se identificaron fortalezas, debilidades y oportunidades de mejora. Concluido el informe, este se envió por correo electrónico a la persona responsable en el equipo de recolección con la finalidad de que los aspectos

identificados fueran retroalimentados y corregidos en las siguientes submuestras operativas. Además de lo anterior, este informe fue utilizado como insumo para las retroalimentaciones realizadas al cierre de cada submuestra.

6.3.6.3 Supervisión de Gabinete

Consistió en la revisión de cuestionarios de Gastos Diarios (GD) y Gastos del Recuerdo (GR) posterior a la aplicación, a través del Formulario de Correcciones del sistema de gestión operativa (SGO). El objetivo de este tipo de supervisión fue identificar errores, inconsistencias, omisiones de información, entre otros, propiciando su resolución durante la aplicación y cierre de cada submuestra operativa.

Desde el equipo de coordinación del proyecto se estableció que la revisión realizada por las personas supervisoras debía ser de un 25% de los cuestionarios GD y GR, en cada una de las regiones del país.

Para el trabajo de revisión de la supervisión de gabinete, las personas supervisoras contaron con el Manual de Recolección de Datos tomo I y II; la Guía de Revisión para personas encargadas grupo y supervisoras IX EPF, documento que especificó la categorización que debía aplicarse a los distintos errores identificados y el procedimiento de registro en el Formulario de Correcciones; así como el documento Formas de registro de productos y servicios IX EPF, el que describió el detalle requerido para el registro de diferentes bienes y servicios que debían ser clasificados posteriormente por el equipo de codificación.

La revisión de los cuestionarios mencionados se desarrolló en dos instancias, el cuestionario de GR en fase de aplicación posterior a la segunda entrevista y el o los cuestionarios de GD en fase de cierre y recuperación posterior a la cuarta entrevista, esto de acuerdo con lo establecido por la calendarización del proyecto.

Para realizar la entrega de estos cuestionarios desde el equipo de recolección al equipo de supervisión, se implementó el protocolo de entrega de materiales, que describió el flujo de entrega y devolución de cuestionarios y establecía la interacción entre ambos equipos durante este proceso. Para ello se instauraron tres flujos de coordinación, debido a las diferencias en las estructuras organizacionales de los equipos de cada región.

Recepción de cuestionarios

El equipo de recolección realizaba la entrega de los cuestionarios que se encontraron aplicados y la fecha de entrega tanto para GR como GD, era establecida por la calendarización de la IX EPF el que consideraba en promedio 3 días de revisión para cada cuestionario. La persona encargada de área en región metropolitana y las supervisoras en el resto de las regiones, eran las responsables de revisar el material entregado por el equipo de recolección y firmar conforme la Planilla de Entrega de Materiales. Esta planilla era provista por el equipo de supervisión y también debía ser firmada por las personas Encargadas de grupo de cada dirección regional, exceptuando las regiones Metropolitana, Valparaíso y Biobío en la cuales fue firmada por una persona Encargada Regional.

Revisión de cuestionarios

Una vez que se dispuso del material correspondiente, las personas supervisoras comenzaron la supervisión de gabinete. El registro e identificación de hallazgos se realizó en dos soportes: sobre los cuestionarios y en el Formulario de Correcciones del sistema informático, el cual permitió consignar, agrupar y dar cuenta de la correcta revisión. En los cuestionarios, la persona supervisora debió marcar la detección de algún error, inconsistencias u omisiones de información marcando con lápiz de color rojo y con un círculo cada campo de pregunta en la que se detectó el hallazgo, además, demarcar con un asterisco el número de línea en que se identificó.

En el sistema de gestión operativo el equipo de supervisión tuvo a disposición un formulario de correcciones, en el cual la persona supervisora identificó al hogar mediante los filtros disponibles. Una vez identificado el hogar, se seleccionó el cuestionario, sección y pregunta del hallazgo, el número de línea de RPH al que correspondía, la descripción del error (según el listado disponible en el formulario) y se agregaron las observaciones que describieran de mejor manera la corrección. Solo era posible realizar correcciones cuando los manuales, el flujo de aplicación, la información de la zona de consultas y/o la coherencia interna de los instrumentos, entregaran respaldo suficiente para hacerlo. Se destaca, además, que todo hallazgo fue identificado en el Formulario de Correcciones.

Una vez llenado el Formulario de Correcciones, se generó un reporte de correcciones el cual contenía cada uno de los hallazgos ingresados al sistema por la persona supervisora con las observaciones a revisar, fecha y hora de revisión y los nombres de las personas responsables. Este reporte, fue impreso y entregado junto con los cuestionarios a la persona encuestadora para dar solución a los hallazgos.

● Devolución de cuestionarios

Posterior a la revisión, y ya adjunto el Reporte de Correcciones a los cuestionarios supervisados, estos fueron devueltos al equipo de recolección, dejando registro de ello en la Planilla de Entrega de Materiales[114] firmando conforme la devolución ambos equipos responsables.

6.3.6.4
Supervisión Indirecta

La supervisión indirecta es la medida de control que permite garantizar que la información entregada por las personas encuestadoras sea fidedigna, no manipulada, de calidad y no se incurra en falsificación de datos. Esta actividad consistió en una verificación de los datos obtenidos en terreno por las personas encuestadoras, principalmente de encuestas logradas. En este sentido, quienes ejecutaron la supervisión indirecta debieron contactarse con las viviendas, en ausencia de la persona encuestadora, para corroborar la composición del hogar, los protocolos de aplicación y las respuestas registradas. Esta verificación de datos se realizó mediante una supervisión vía contacto telefónico o supervisión indirecta de la hoja de ruta efectuada en terreno.

Para cumplir esta actividad, tanto en modalidad telefónica como presencial, las personas encargadas de área asignaron bajo criterios previamente establecidos, las unidades a supervisar para cada uno de los integrantes del equipo de supervisión.

Para el caso de las supervisiones telefónicas se verificó que hubiese al menos dos contactos telefónicos por cada hogar y que la información de las unidades a supervisar se encontrara precargada en los módulos de gabinete del formulario de supervisión indirecta.

[114] Documento utilizado para formalizar y registrar la entrega de cuestionarios en do de se identifican a las partes involucradas.

Las supervisiones indirectas telefónicas se aplicaron a través del Formulario de Supervisión Indirecta en dos diferentes fases de la submuestra operativa:

- **En fase de aplicación:**
  Considera los módulos de apertura y registro de personas en el hogar (RPH), más dos módulos o cuestionarios adicionales (entre Gastos Diarios, Gastos en Educación, Gastos Periódicos del Hogar, Gastos del Recuerdo, Ingresos y Gastos en Servicios Financieros y Seguros).

- **En fase de cierre:**
  Se aplicó íntegramente el formulario de supervisión indirecta. Para el estudio, esta es la supervisión que se priorizó, ya que constituía una selección de preguntas esenciales para el logro de la encuesta.

Para la selección de la persona informante, el equipo de supervisión utilizó los mismos criterios establecidos por recolección, es decir, un integrante del hogar, de 18 años o más, que se encuentre en condiciones físicas y/o mentales para responder. Podía no ser necesariamente la persona que respondió cada módulo, pero sí alguien que se encontraba en condición de chequear las respuestas.

En caso de que existiera sospecha de falseamiento de datos o evidencia de alguna mala práctica, la persona que realizaba la supervisión indirecta tenía la obligación de intentar corroborar los datos con la misma persona informante del módulo, mediante supervisión indirecta o de hoja de ruta, según sea el caso. Adicionalmente, debía supervisar de manera exhaustiva el material producido por dicho encuestadora o encuestador, incluyendo cuestionarios en papel y submuestras anteriores. Si la sospecha es confirmada, la persona supervisora informaba inmediatamente la situación a la persona encargada de área de supervisión, quien procede a la activación del "Protocolo de Falsificación"[116].

El formulario de supervisión indirecta se estructuró en nueve secciones, entre las que se encontraban "Datos de Contacto", "Apertura", "Registro de Personas del Hogar (RPH)", "Gastos periódicos del Hogar (GH)", "Gastos en Educación (GE)", "Ingresos (IN)", "Gastos del Recuerdo (GR)", "Gastos en Servicios Financieros y Seguros (GSF)" y "Resultado de Supervisión y observaciones" y adicionalmente una subsección de "Control de cuestionarios de Gastos Diarios". Para el correcto registro de cada sección del instrumento, la información de las variables a verificar del cuestionario de recolección se precargaba automáticamente en el formulario de supervisión, en un módulo interno denominado "Registro en Gabinete". Esta automatización, permitió corroborar e integrar las respuestas del cuestionario digital de recolección y compararlas con la

[115] Protocolo estandarizado con procedimientos para casos de detección de inconsistencias, irregularidades o falseamiento de datos, que se presenten durante el proceso de recolección de datos.

información que proporcionaba la persona informante de supervisión en el módulo "Registro de re-entrevista", evitando de esta manera, posibles errores en la edición de la respuesta original de las entrevistas realizadas por las personas encuestadoras.

Las secciones correspondientes a RPH, GE, GH, GR, IN y GSF, consistían en una verificación de la aplicación de las principales preguntas de cada módulo. Para esto, era necesario que la persona supervisora aplicara nuevamente estas preguntas, registrara las respuestas entregadas por las personas informantes, y completara la pregunta "resultado" de acuerdo a la información proporcionada por las personas entrevistadas. Es importante señalar que la forma de llenado de esta pregunta estaba determinada por los criterios del protocolo de conciliación de la supervisión indirecta, pudiendo presentarse dos situaciones: coincidencia entre las personas informantes de recolección y supervisión; y no coincidencia. Este documento, permitió categorizar las diferencias con el propósito de encontrar mejoras de información, identificar errores o solicitar verificación de información.

Las categorías de clasificación para estas respuestas fueron las siguientes:

1. **Coincide:**
   Registra respuestas coincidentes entre ambas entrevistas. Considera tanto coincidencias exactas como valores que, debido a su rango, sean determinados como coincidentes en el protocolo de conciliación.

2. **Mejora de información:**
   Registra respuestas donde se evidencie un dato de mejor calidad, sea por las características de la persona informante (presentaba mayor idoneidad en la reentrevista) o porque la misma persona informante rectifica información registrada anteriormente.

3. **Error de registro:**
   Registra respuestas con diferencias entre el registro original y el de la reentrevista.

4. **Verifique información:**
   Registra respuestas en las que no hay claridad de si la respuesta es coincidente o no. Estos casos se deben a que la persona informante declare duda en responder y que el protocolo de conciliación así lo estipule.

5. **No disponible:**
   Registra una respuesta distinta de la entregada por la persona informante original, a la respuesta entregada por el informante que respondió el cuestionario de supervisión, y quien informa desde supervisión considera no disponer de información suficiente para responder, es decir, hay una imposibilidad de evaluar la coincidencia de la información.

6. **Otro:**
   Cualquier otra situación no contenida en los códigos anteriores. Es obligatorio complementar con el campo de observaciones.

Una vez realizada la supervisión, en el caso que se detectaran diferencias, la persona supervisora debió registrar los hallazgos y/o diferencias detectadas en el formulario de correcciones, luego de clasificarlas en las categorías mencionadas anteriormente. Una vez indicados los errores detectados la persona encargada de grupo, se encarga de realizar el cierre del proceso de supervisión indirecta.

La información proporcionada por el formulario de supervisión indirecta permitía identificar aspectos a mejorar, lo cual se utilizaba como un insumo para los procesos de retroalimentación llevados a cabo en la etapa de cierre de la supervisión.

• Supervisión de Hoja de Ruta

En el ámbito de la supervisión indirecta, también se implementó un tipo de supervisión de Hoja de Ruta. Esta consistió en verificar la gestión realizada por la persona encuestadora durante la submuestra operativa. Específicamente buscó confirmar los registros de contactos realizados y el código de disposición seleccionado; como también confirmar la agenda de cita de la persona encuestadora acordada con la informante. Esta supervisión se debía realizar sin presencia de la persona encuestadora.

El Formulario de Supervisión de Hoja de Ruta fue utilizado para realizar este tipo de supervisión y estaba compuesto por las secciones Registro de contacto de Supervisor, que permitía monitorear la gestión en terreno realizada por la persona supervisora; Registro de contacto con el hogar (RC); Agenda de citas con el informante (AC), ambas correspondían a la gestión realizada por la persona encuestadora en la unidad seleccionada y era precargada en el Formulario de Supervisión de Hoja de Ruta y por último, la sección Resultado de supervisión y Observaciones, la cual permitía clasificar en Supervisión Lograda, Rechazo y No supervisada por otros motivos.

La supervisión de Hoja de Ruta, era posible realizarla durante toda la submuestra operativa. Los criterios de asignación consistían en aplicar esta supervisión a personas recolectoras con Tasa de Respuesta inferior al 50% y de forma aleatoria, especialmente en unidades de los grupos "No elegibles", "Elegibilidad Desconocida" y "No entrevistadas por otros Motivos". Sin embargo, también era posible realizar este tipo de supervisión cuando existían solicitudes focalizadas por parte del equipo de recolección y en situaciones en que la persona supervisora, a través del trabajo de supervisión previo, consideraba necesario verificar los datos de Hoja de Ruta, debido a identificación de registros erróneos y sospecha de malas prácticas.

Concluida la supervisión de Hoja de Ruta, la persona supervisora en trabajo de gabinete debía registrar las diferencias halladas durante supervisión en el Formulario de Correcciones del SGO y posteriormente enviar un aviso del hallazgo con número de identificador único del hogar por correo electrónico a la persona responsable en el equipo de recolección con la finalidad de corregir y/o complementar la información.

6.3.6.5 Cierre de supervisión

Una vez finalizado el proceso de supervisión, las personas supervisoras realizaron una revisión y análisis de la información contenida en los informes de supervisión directa, y en los aspectos identificados en las supervisiones de gabinete e indirecta. Para ello se realizaban las siguientes tareas:

- Preparar una minuta de retroalimentación identificando los principales hallazgos del proceso de supervisión.
- Reunirse con el equipo de recolección para realizar la entrega de los resultados (informe de supervisión directa) y minuta de retroalimentación.
- Retroalimentar el trabajo supervisado al equipo de recolección, en donde iba dirigida a las personas encargadas de grupo. En esta instancia se abordaban las temáticas identificadas como oportunidades de mejora, fortalezas de las personas recolectoras y recomendaciones de buenas prácticas.
- Enviar la minuta de retroalimentación a la encargada o encargado de área de supervisión con copia a la persona encargada del grupo.

Las personas encargadas de área de supervisión derivaron esta información al responsable de coordinar el proceso de supervisión nacional, para que se realizara el análisis de estos resultados, complementándolo con los análisis de las bases de datos generadas a través del formulario de correcciones del sistema informático. Esta información era proporcionada al responsable de la coordinación nacional de la recolección mediante presentaciones que se realizaban en la reportabilidad de la coordinación operativa central.

6.3.7 Indicadores de supervisión

Los indicadores de supervisión se utilizaron durante todo el proceso de recolección oficial de datos, lo que permitió realizar el seguimiento de la etapa de supervisión de la encuesta de forma oportuna, generando información relevante de errores no muestrales para la evaluación y ajustes del proceso de control de calidad. Estos fueron construidos mediante la información disponible en los estados de logro de la hoja de ruta de supervisión, la cual utilizaba los códigos adaptados al estándar indicado por la (AAPOR, 2023).

La Hoja de Ruta de supervisión se incorporó en cada formulario digital y tuvo como objetivo caracterizar la gestión de cada supervisión, registrando las actividades más importantes realizadas por las personas que ejecutaban el control de calidad de la encuesta, midiendo el logro de su trabajo.

Entre los usos de este instrumento, lo más relevante fue el registro de intentos de contactos y la caracterización de resultados final de la gestión de supervisión en cada unidad muestral. Esto constituyó el insumo principal para la construcción de indicadores de gestión, el monitoreo del rendimiento y productividad en la ejecución del proceso de supervisión de la encuesta.

Tabla 34: Indicadores de supervisión

Indicador	Descripción	Fórmula de cálculo
Tasa de supervisión en unidades entrevistadas (TSE)	Este indicador corresponde al número total de unidades entrevistadas que fueron supervisadas logradas de manera directa e indirecta con clasificación de logradas por región, aprobadas por encargado(a) de área de supervisión, sobre el total de unidades logradas de manera inicial (CDI) en recolección aprobadas por encargado(a) de grupo por región, en un período de tiempo determinado.	$TSE = \frac{\text{N° total de unidades "E" supervisadas de manera directa e indirecta según tipología por región, aprobadas en un periodo t}}{\text{N° total de unidades "E" de manera inicial en recolección por región, aprobadas en un periodo t}}$
Tasa de supervisión en unidades no entrevistadas (TSNE)	Este indicador corresponde al número total de unidades supervisadas de manera directa e indirecta en unidades no entrevistadas con clasificación de logradas por región, aprobadas por encargado(a) de área de supervisión, sobre el total de unidades no entrevistadas de manera preliminar (CD) en recolección aprobadas por encargado(a) de grupo por región, en un período de tiempo determinado. **Unidades No Entrevistadas:** *Break off* (BO) + Rechazadas (R) + No contactadas (NC) + No entrevistadas por otros motivos (O) + elegibilidad desconocida (ED) + no elegibles (NE).	$TSNE = \frac{\text{N° total de unidades supervisadas de manera directa e indirecta en unidades no entrevistadas según tipología por región, aprobadas en un periodo t}}{\text{N° total de unidades no entrevistadas de manera preliminar en recolección por región, aprobadas en un periodo t}}$
Tasa de supervisión indirecta en unidades entrevistadas (TSI)	Este indicador corresponde al número total de unidades supervisadas de manera indirecta con clasificación de logradas por región, aprobadas por encargado(a) de área de supervisión, sobre el total de unidades Entrevistadas de manera inicial (CDI) en recolección aprobadas por encargado(a) de grupo por región, en un período de tiempo determinado.	$TSI = \frac{\text{N° total de unidades supervisadas de manera indirecta según tipología por región, aprobadas en un periodo t}}{\text{N° total de unidades "E" de manera inicial en recolección por región, aprobadas en un periodo t}}$
Tasa de supervisión directa en personas recolectoras (TSD)	Este indicador corresponde al número total personas recolectoras supervisadas de manera directa en apertura o aplicación, aprobadas por encargado(a) de área de supervisión, sobre el total de personas recolectoras contratadas por el proyecto, en un período de tiempo determinado.	$TSD = \frac{\text{N° total de recolectores supervisados de manera directa por región, aprobadas en un periodo t}}{\text{N° total de recolectores contratados en recolección por región, aprobadas en un periodo t}}$

Indicador	Descripción	Fórmula de cálculo
Tasa de supervisión en unidades no elegibles (TSNEL)	Este indicador corresponde al número total de unidades no elegibles supervisadas en tipología hoja de ruta con clasificación de logradas por región, aprobadas por encargado(a) de área de supervisión, sobre el total de unidades no elegibles de manera inicial (CDI) en recolección aprobadas por encargado(a) de grupo por región, en un período de tiempo determinado.	$TSNEL = \frac{\text{N° total de unidades NE supervisadas en tipología hoja de ruta por región, aprobadas en un periodo } t}{\text{N° total de unidades NE de manera inicial en recolección por región, aprobadas en un periodo } t}$

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

6.3.8 Digitación de datos recolectados en papel

El proceso de digitación es el punto de inflexión en el cual los datos capturados en papel se transforman en registros digitales y pasan a ser parte de las bases de datos que luego ingresarán a la fase de procesamiento. Esta actividad se llevó a cabo a través de labores que se describen a continuación.

6.3.8.1 Recepción de material

La etapa de recepción de material estuvo a cargo de quien coordinó los equipos, junto a las personas encargadas de grupo de digitación. Este proceso constó de dos pasos que permitieron la revisión y posterior digitación de las encuestas.

El material de trabajo de recolección llegó al equipo de digitación, a través de las personas encargadas de grupo quienes separaron por regiones la submuestra operativa. Luego, se procedió a revisar que la información registrada en la guía de despacho coincidiera con los cuestionarios proporcionados (números de folios, cantidad de cuestionarios GR y GD y número de las UPM).

Una vez revisados los cuestionarios de la guía de despacho, se debía verificar que dicha información coincidiera con la que se encontraba registrada en el SGO. Si se detectaba errores en la recepción de información, se notificaba la incidencia al equipo de recolección, a través de las personas encargadas regionales o encargadas de grupo, con el fin de corregir los errores y proceder a rectificar la recepción de material.

6.3.8.2
Asignación de cargas de trabajo

La recepción del material permitió la asignación de cuestionarios a las personas digitadoras, considerando la cantidad de cuestionarios totales a digitar de Gastos Diarios (GD) y Gastos del Recuerdo (GR).

En promedio, la carga de trabajo de cada persona fue de 157 cuestionarios y 88.508 campos por submuestra operativa, lo que se consideró como un parámetro aproximado para asignar las cargas de trabajo.

Se debe mencionar que cada guía de despacho fue asignada a una persona digitadora, esto debido a que los cuestionarios asociados a un hogar estaban contenidos en una sola guía de despacho. Una vez que las personas digitadoras avanzaban en su labor, se revisaron las cargas de trabajo del equipo, con el fin de realizar eventuales ajustes y redistribuciones de cargas.

Las cargas de trabajo eran asignadas de forma paulatina, buscando mantener un equilibrio entre el personal de digitación y procurando que ninguna persona integrante del equipo se quedara sin carga de trabajo. Los ajustes en las cargas de trabajo buscaron optimizar los tiempos de digitación, resguardando que los cuestionarios de toda la submuestra operativa terminaran de digitarse antes de finalizar el plazo contemplado para la entrega al equipo técnico y que no se solapara con la siguiente submuestra.

La etapa de asignación de cargas de trabajo estuvo a cargo de la persona coordinadora del equipo, junto a quienes estaban a cargo de los grupos de digitación.

6.3.8.3
Ingreso de información en el sistema informático

Para ingresar los datos registrados de los cuestionarios GR y GD, cada persona digitadora tuvo acceso al sistema informático de la IX EPF (SGO), en el que debían registrar su carga de trabajo. Cada perfil del equipo podía acceder a las distintas funciones del sistema, según sus tareas que desarrollaron, personas coordinadoras, encargadas de grupo y digitadoras, estas últimas tenían acceso únicamente a la sección de ingreso de los cuestionarios de GR y GD solo para aquellos hogares que les fueron asignados.

Quien digitaba debía ingresar en el SGO de la IX EPF toda la información contenida en los cuestionarios de GR y GD de la forma más fehaciente posible. Por lo cual se les orientó, considerando instrucciones detalladas de cómo ingresar la información en el sistema utilizando los códigos establecidos.

6.3.8.4
Revisión de errores de digitación

Durante el proceso de digitación se detectaron algunos problemas asociados al registro de información. Estos problemas fueron informados por la coordinación de digitación a los equipos correspondientes con el fin de lograr una resolución. Los errores detectados se asociaron a los siguientes motivos:

- **Errores de tipo informático:**

Son aquellos errores asociados a las caídas de los servidores, interrupciones fortuitas del servicio de internet o al funcionamiento de ciertas validaciones del sistema de digitación, como a la inconsistencia de lo registrado por la persona encuestadora en el RPH con lo que era vinculado al sistema de digitación de la encuesta.

- **Errores de trabajo de recolección:**

Estos errores correspondían a deficiencias u omisiones de información generadas en la fase de recolección de datos. Este tipo de errores estaban asociados principalmente al no traspaso de las boletas a las tablas de GD.

En cada submuestra operativa se realizó un proceso de revisión de los datos digitados, para esto se consideró una revisión del 10% de los cuestionarios digitados[116]. Para realizarla, se seleccionó una guía de despacho con el mayor número de cuestionarios e información ingresada (esta información se logró gracias a la planilla llamada Gastos efectivos), comparando los registros ingresados en el sistema con aquellos contenidos en los cuestionarios en papel.

Durante el proceso de digitación de cada submuestra operativa se realizó un proceso el cual se denominó Revisión cruzada, la cual consistió en revisar la información registrada en el sistema y contrastarla con la registrada en los cuestionarios. Para ello, una persona digitadora debía revisar una guía de despacho completa con sus respectivos cuestionarios que había sido digitado por otro integrante del equipo. El propósito fue registrar los errores e inconsistencias en el sistema, indicando que no correspondía con la información registrada en los cuestionarios. Una vez finalizada la revisión cruzada se le devolvía la guía a quien había digitado inicialmente la información, quien era la encargada de corregir la información en el sistema.

6.3.8.5
Orden y despacho del material

Una vez finalizada la digitación y revisada cada submuestra operativa, se procedía a realizar la etapa de orden y despacho del material. Durante este proceso se verificaba que todos los cuestionarios estuviesen efectivamente digitados y finalizados, para ello, se utilizó un reporte de asignación de folios asignados, entregado por el SGO.

[116] Los datos registrados están procesados con la información obtenida desde la submuestra operativa número 8 a la 24

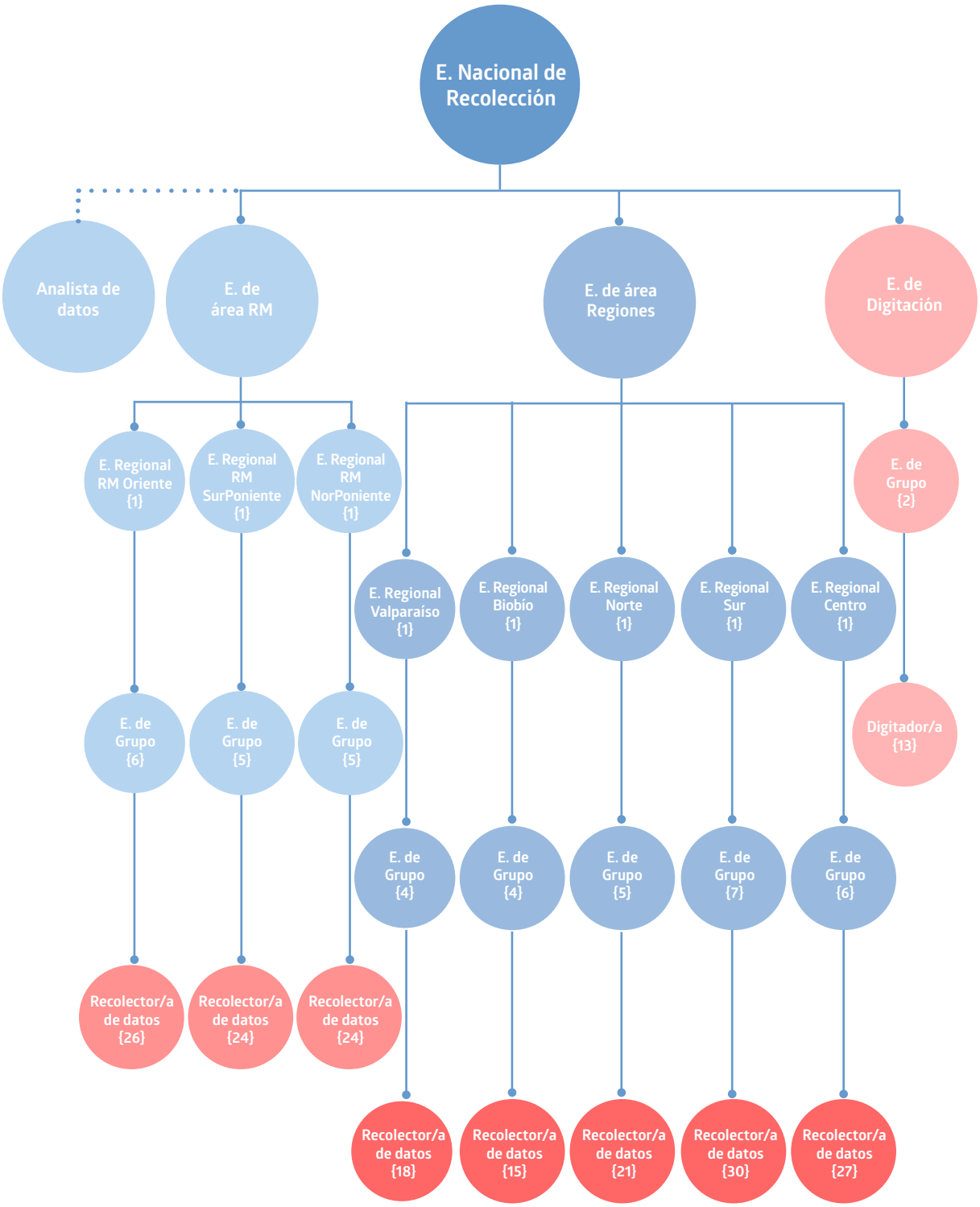
En esta etapa se enviaba el material a almacenaje, el que quedó a disposición del equipo técnico, debido a que eventualmente se podía volver a los cuestionarios físicos desde el subproceso de revisión y edición de datos. La tarea de ordenar y despachar el material fue realizada por las personas encargados de grupo junto a quien coordinaba los equipos de digitación.

6.4 Características organizacionales de equipos participantes de la recolección de datos

6.4.1 Organigrama del equipo de recolección

El equipo de recolección tuvo un alcance nacional y estuvo compuesto por 255 personas, entre personas encuestadoras, encargadas de grupo, encargadas regionales, encargadas de área, coordinador nacional operativo, una persona analista de datos y 16 personas en el equipo de digitación. La estructura del equipo de recolección se presenta en el **Diagrama 9.**

Diagrama 9: Organigrama equipo de recolección



Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

En cuanto a la responsabilidad relativa a la ejecución de la recolección, el equipo mantuvo dos niveles en su organización, coordinación central de recolección y equipos de recolección. Los que se describen en los siguientes párrafos.

En primera instancia la coordinación central de recolección[117], la que estaba compuesta por las personas encargadas regionales, de área y nacional, quienes eran las responsables de la elaboración y capacitación de todo lo relacionado con el trabajo de recolección, así como del monitoreo de los indicadores de desempeño operativo[118] entregado por la persona analista de datos y las propuestas de estrategias para mejorar los mismo.

Esta coordinación se dividió en dos áreas, Región Metropolitana y otras Regiones. Cada una de ellas tuvo una persona encargada de área, la cual trabajaba con el apoyo de las personas encargadas regionales quienes eran responsables de una zona específica al interior de un área en cuánto a la organización, acompañamiento y monitoreo de los equipos de recolección respectivos.

En segundo lugar, los equipos de recolección, coordinados por una persona encargada de grupo, estaban compuestos por entre tres y seis personas encuestadoras, dependiendo la cantidad de viviendas asignadas en cada región. Estos eran los responsables de la difusión en terreno de la EPF y la producción de datos en los hogares seleccionados.

La IX EPF, a diferencia de su versión anterior, incorporó al equipo de digitación como parte de la estructura de recolección[119]. El equipo estuvo compuesto por 13 personas digitadoras, 2 encargadas de grupo de digitación y una coordinadora, que ejecutaron sus funciones de manera centralizada, desde la Región Metropolitana. Esto permitió, por una parte, aprovechar economías de escala respecto a la distribución del material a digitar y, por otra, contar con una comunicación directa y continua de retroalimentación con los equipos de trabajo de recolección y equipo técnico.

[117] Se definen estas áreas para generar un seguimiento más detallado de los resultados operativos y respuesta oportuna ante potenciales riesgos en sus niveles de respuesta.

[118] A cargo de la persona analista de datos operativos.

[119] Esta decisión es consecuencia de la adopción del GSBPM, que identifica la digitación como parte del subproceso Cerrar la recolección (UNECE, 2019).

6.4.2 Distribución del personal de recolección

De acuerdo con la distribución de la muestra en cada región del país, se realizó la estimación del personal de trabajo para el proceso de recolección de datos de la EPF. A su vez, a cada grupo se le asignó una persona encuestadora adicional ante posibles eventualidades de rotación de personal o ausencias laborales (licencias médicas, vacaciones, etc.).

Por otra parte, para determinar el número de personas encargadas de grupo se consideró la cantidad necesaria de encuestadores y encuestadoras en cada una de las regiones, estableciendo un mínimo de tres y un máximo de seis por persona encargada. Finalmente, se designaron personas encargadas regionales, para coordinar y dar seguimiento de los resultados de sus zonas específicas. La cantidad de personas del equipo de recolección en cada región se presenta en la **Tabla 35.**

A nivel nacional el criterio de asignación de equipos fue similar a la VIII EPF, donde se establece el principio de distribución geográfica de acuerdo con la residencia de las personas encuestadoras, experiencia tanto de trabajo como de los territorios, entre otros. Para el caso de la Región Metropolitana, las zonas de las personas encargadas regionales fueron definidas territorialmente de modo que estuviesen compuestas por comunas con características sociodemográficas similares, de este modo, se conformaron las zonas oriente, norponiente y sur[120], organización que permitió que las personas encuestadoras y encargadas de grupo se adaptaran a los estilos de persuasión más efectivos en cada territorio y a realizar las entrevistas considerando las características y horarios de la población informante.

La preparación de la recolección EPF en la Región Metropolitana contó con diversas instancias de apoyo, para preparar las estrategias que se deben emplear en distintos territorios. Ente las que destacan las reuniones con las personas encargadas de la región en la versión anterior, las que permitieron generar lineamientos de gestión de la muestra a nivel territorial. También, existió una mesa técnica transversal a los equipos de recolección de estadísticas sociales del Instituto Nacional de Estadísticas, en la que se revisaron las principales incidencias en terreno detectadas en la región para el último año y se debatieron recomendaciones para la distribución y asignación de cargas de trabajo, lo que permitió generar los ajustes necesarios anticipadamente al inicio del proceso.

[120] La conformación original de estas zonas tuvo ajustes durante el proceso de recolección, producto del efecto de la rotación de personal en algunos períodos. Ante esas situaciones, se realizó una redistribución de comunas y/o UPM a nivel de directorio, lo que permitió transferir cargas de trabajo a equipos zonales con mayor presencia de personas recolectoras.

Tabla 35: Distribución equipo de recolección por región

Región	Encuesta-dor (a)	E. de Grupo	E. Regional	E. de Área	E. Nacional	Analista de datos	E. digitación	EG de digitación	Digitado-res	Total personal
Arica y Parinacota	4	1								5
Tarapacá	5	1								6
Antofagasta	8	2								10
Atacama	4	1								5
Coquimbo	10	2								12
Valparaíso	18	4	1							23
Metropolitana	74	16	6(1)	2	1	1	1	2	13	116
O´Higgins	6	1								7
Maule	7	2								9
Ñuble	4	1								5
Biobío	15	4	1							20
Araucanía	11	2								13
Los Ríos	4	1								5
Los Lagos	9	2								11
Aysén	3	1								4
Magallanes	3	1								4
Nacional	*185*	*42*	*8*	*2*	*1*	*1*	*1*	*2*	*13*	*255*

(1) En la Región Metropolitana incluye encargados regionales de las zonas norte, centro y sur del área de regiones.

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

6.4.3 Perfiles de cargo, selección y condiciones contractuales del equipo de recolección

Después de establecer la estructura de roles necesarios para el proceso de recolección, se desarrollaron los perfiles de cargo que delinearon las características deseadas en términos de nivel educativo, experiencia laboral y conocimientos informáticos para cada uno de los perfiles, según las tareas a desempeñar.

El proceso de selección de los equipos de recolección contempló las etapas de evaluación curricular, evaluación psicolaboral, evaluación técnica y entrevista con las jefaturas. Dado el contexto sanitario del Covid-19, todas estas etapas fueron realizados de manera telemática.

La evaluación curricular y psicolaboral de las personas postulantes estuvo a cargo del equipo de desarrollo de personas del INE. La evaluación psicolaboral se centró en la medición de variables de conocimientos generales y rasgos de personalidad, mediante pruebas individuales en línea. Por otra parte, la evaluación técnica, realizada para todos los perfiles telemáticamente, buscó evaluar las competencias necesarias para el desempeño del cargo. En las entrevistas finales, ejecutadas por personal de coordinación de operaciones sociales del INE, se evaluó la habilidad comunicativa de quienes postularon y se indagó detalles respeto de su experiencia laboral.

Las personas del equipo de recolección fueron contratadas en cascada, iniciando por los cargos de coordinación nacional, hasta llegar a los perfiles que componían los grupos de recolección. Para estos últimos, el período de contratación comprendió la capacitación y la ejecución de la recolección oficial de la IX EPF, incluyendo la fase de cierre de la última submuestra[121].

[121] El período de contratación de las personas encargados regionales fue de un año nueve meses, mientras que el de las personas encargadas de grupo fue de un año cuatro meses, cabe destacar que el equipo en cuestión de la Región Metropolitana tuvo una extensión de contratos hasta diciembre del 2022, para apoyar las recuperaciones y la digitación, por lo que tuvieron un período de un año seis meses de contrato. Las personas encuestadoras permanecieron en el proyecto durante un año tres meses. En cuanto a las personas encargadas de área, tuvieron una continuidad hasta el año 2023.

El equipo de personas encuestadoras fue contratado en calidad de honorarios, con una remuneración fija mensual, debiendo cumplir con las unidades asignadas por la coordinación, dedicando distintos días y rangos de horarios establecidos previamente para la apertura y el seguimiento de las viviendas, realizando siete intentos de contacto en los distintos horarios donde al menos dos de ellos correspondieran a días de fin de semana y/o festivos.

En cuanto a los horarios de las personas encuestadoras para la realización de funciones en terreno y gabinete, estos fueron flexibles y de común acuerdo entre personas encargadas de grupo y encuestadoras, para tener una mayor adaptación a los horarios disponibles de las personas informantes y ejecutar las distintas entrevistas durante la recolección de datos respetando el esquema de entrevistas descrito en la sección 6.3.4.2.

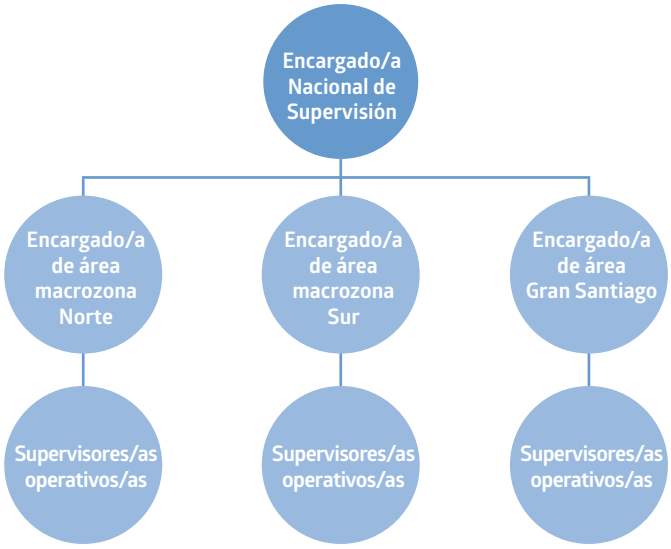
En cuanto a las personas encargadas de grupo, al igual que las personas encargadas regionales, de área y nacional, debían cumplir con jornadas de trabajo de 44 horas semanales, distribuidas de lunes a viernes, en los mismos horarios vigentes para el resto del INE. Estos horarios fueron utilizados para la comunicación entre la coordinación nacional de recolección, las direcciones regionales y el resto de los equipos institucionales.

6.4.4 Organigrama del equipo de supervisión

El equipo de supervisión de la IX Encuesta de Presupuestos Familiares estuvo compuesto por 31 personas: una persona responsable de coordinar la implementación y ejecución del proceso de supervisión a nivel nacional, tres personas encargadas de área, responsables de la macrozona[122] norte, sur y el Gran Santiago, respectivamente; y 27 supervisores operativos, encargados directamente del trabajo de control de calidad de la encuesta.

[122] Las áreas geográficas delimitadas para la supervisión no corresponden a las mismas definidas para las áreas de estimación de la encuesta. Por lo tanto, en el equipo de supervisión la macrozona norte incluyó desde Arica y Parinacota hasta la región de O' Higgins, excluyendo la Región Metropolitana que tenía su propio encargado de área, mientras que la macrozona sur incluyó desde la región de Maule a Magallanes.

Diagrama 10: Organigrama equipo de supervisión



Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

El equipo de supervisoras y supervisores operativos se distribuyó en cada una de las regiones del país, considerando la cantidad de personas recolectoras por región y de acuerdo al análisis presupuestario que permitiera asegurar las coberturas establecidas para los procesos de supervisión sobre la muestra teórica de la encuesta.

La asignación de cargas de trabajo era realizada por las personas encargadas de área de supervisión, quienes fueron responsables de la selección de unidades a supervisar, el seguimiento operativo de las actividades de supervisión y apoyo metodológico en las tareas de control de calidad.

Tabla 36: Distribución equipo de supervisión por región

Regiones	Supervisor(a) operativo(a)	Encargado(a) de área	Encargado(a) Nacional
Nacional	*27*	*3*	*1*
Arica y Parinacota	1		
Tarapacá	1		
Antofagasta	1		
Atacama	1		
Coquimbo	1		
Valparaíso	3		
Metropolitana de Santiago	8	3	1
O'Higgins	1		
Maule	1		
Biobío	3		
Ñuble	1		
La Araucanía	1		
Los Lagos	1		
Los Ríos	1		
Aysén	1		
Magallanes	1		

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

6.5 Método de capacitación para el personal del proceso de recolección de datos

La recolección de datos en la EPF exige a quienes encuestan la profunda comprensión sobre los conceptos, los procedimientos y la importancia de cada una de las preguntas. El método de capacitación corresponde al mecanismo que permite a todo el personal asimilar estos aspectos críticos y aplicarlos con precisión en la ejecución de sus labores.

El método de capacitación utilizado buscó la mejor forma de asegurar que el personal participante del proceso de recolección (ejecución de la recolección, supervisión y digitación) comprendiera completamente los objetivos de la encuesta y los conceptos relacionados con los gastos e ingresos de los hogares.

6.5.1 Capacitación del equipo de recolección

El proceso enfocó sus objetivos en capacitar personas encuestadoras y encargadas competentes en los procedimientos de recolección de datos, desde la interacción inicial con los hogares hasta la correcta captura de la información en los cuestionarios del estudio.

El proceso de capacitación del equipo de recolección se realizó de forma *online* debido a la contingencia sanitaria del Covid-19. Durante este proceso se buscó entregar a las personas los conocimientos de acuerdo con su perfil y las herramientas necesarias para asegurar la calidad de los datos recolectados.

La capacitación durante todo el período de recolección de datos se dividió en dos etapas:

- **Capacitación inicial:**
Consistió en una inducción *online*[123] respecto del proceso productivo de la encuesta, conceptos básicos, la metodología del trabajo de recolección y formas de aplicar los cuestionarios de la encuesta. Se desarrolló mediante cursos expositivos y talleres prácticos, a través de los cuales

[123] Esta inducción incorporó elementos de carácter administrativo, como temas contractuales, temas de seguridad, sindicales, etc.

se explicaba detalladamente las labores de terreno y gabinete (oficina) que cada persona debía realizar según su rol.

Al inicio del proyecto se realizó la capacitación para las personas encargadas de grupo y las que fueron impartidas por encargadas de área y regionales. El curso *online* contemplaba una semana durante la cual revisaron las funciones específicas de su cargo y el uso de las herramientas de trabajo disponibles.

Durante la siguiente semana, ingresaron a la institución las personas encuestadoras cuya capacitación *online* tuvo una duración de tres semanas. En esta también participaron las personas encargadas de grupo con el fin de que conocieran los conceptos y metodología de la encuesta. Durante esta capacitación se revisaron los principales contenidos del estudio, el detalle de cada uno de los cuestionarios a aplicar, la división del esquema de entrevistas y se aclararon las dudas asociadas al flujo de la encuesta. Además, se incorporó información atingente a las labores impartidas por el equipo de supervisión. Esta capacitación fue impartida por los integrantes del equipo de coordinación de recolección (personas encargadas regionales, de área y nacional), el equipo de supervisión y el equipo técnico. Debido a que la capacitación se realizó como un curso online, se generaron 12 salas en la plataforma *Teams*, donde se dividieron los equipos por regiones, generando un número equitativo de personas por sala donde se fueron presentando los distintos instrumentos de la encuesta. Para estas presentaciones se generaron duplas donde en cada sala estaba presente, una persona del equipo técnico y otra del equipo de supervisión o recolección para ir presentando los distintos contenidos referentes a la EPF.

- **Capacitación continua:**

Debido a la rotación de personas encuestadoras, se realizó capacitaciones durante todo el período de recolección de datos, porque las personas ingresaban en tiempos distintos, según la necesidad de los equipos de cada región. Para esto, se realizaba una capacitación online con una duración de dos semanas, la que a través de videos grabados de la capacitación inicial se instruía la metodología y los contenidos de la encuesta. Además, se generaron espacios para aclarar dudas donde

las personas encargadas regionales, quienes estaban a cargo de estás capacitaciones en conjunto con el equipo técnico, guiaban la capacitación y cada día se iba sumando la persona encargada del cuestionario o módulo de la unidad técnica para revisar y aclarar dudas de las personas encuestadoras que participaban de esta instancia de aprendizaje.

6.5.2 Capacitación del equipo de supervisión

El proceso de capacitación del equipo de personas supervisoras se desarrolló en dos etapas: (i) la inducción a la metodología y cuestionarios de la encuesta y (ii) la inducción en aspectos específicos del trabajo de supervisión operativo. Ambas etapas se implementaron en una capacitación telemática, para el total de personas supervisoras6, ésta tuvo una duración de tres semanas.

- **Inducción a la metodología y cuestionarios de la encuesta:**

Esta etapa tuvo una duración de 10 días cronológicos y las temáticas estaban organizadas inicialmente en torno a elementos de inducción institucional y prevención de riesgos, seguidas de una descripción general de la encuesta, conceptos básicos de la EPF, descripción de las fases de preparación, aplicación y cierre de recolección de datos, para luego abordar los instrumentos asociados a la correcta aplicación de la encuesta. Junto a esto, y según la información expuesta, se realizó una serie de talleres prácticos, resolviendo además las consultas y dudas realizadas en el proceso de capacitación de las personas recolectoras, con fines de retroalimentación.

- **Inducción al trabajo de supervisión:**

Esta etapa tuvo una duración de cuatro días y medio, y principalmente se organizó en una sala de capacitación, la cual se implementó a través de la aplicación *Teams*. Las temáticas en que se centraron los módulos de esta capacitación fueron composición y metodología de la supervisión, ámbitos y tipos de supervisión, protocolos de supervisión y uso de sistema de gestión operativa.

Respecto al módulo de composición y metodología se presentó la estructura del equipo y su funcionamiento a nivel nacional, además de los fundamentos teóricos del diseño de la supervisión y sus métodos de aplicación.

En el módulo de ámbitos y tipos de supervisión, se expusieron los tipos de supervisión utilizados en la IX EPF y se profundizaron en sus aspectos operativos. Adicionalmente estos módulos tenían asociados talleres prácticos y uso de dispositivos móviles.

La inducción sobre protocolos consistió en la presentación de los flujos de actividades a realizar en los casos de falsificación, conciliación de respuestas y entregas de materiales para revisión del equipo de supervisión.

Finalmente, en el módulo de sistema de gestión se presentó y explicó el uso de las secciones de supervisión del SGO, indicadores de desempeño operativo e indicadores de supervisión.

El material utilizado para esta inducción fue el Manual de recolección para supervisores y supervisoras, Protocolo de falsificación, Protocolo de conciliación, Protocolo de coordinación de entrega de materiales y la Guía de revisión para personas encargadas de grupo y supervisoras.

6.5.3 Capacitación del equipo de digitación

La capacitación del equipo de digitación consistió en una inducción inicial de dos días de duración, en la que se abarcaron los siguientes temas:

- Características generales de la encuesta con sus objetivos y usos.
- Proceso utilizado en la producción de la encuesta.
- Inducción y funcionamiento del sistema informático para ingresar la información de los cuestionarios en papel (GR y GD) en el SGO de la IX EPF.

Para realizar esta capacitación se realizaron presentaciones, las que se complementaron con un manual de digitación como apoyo a la inducción, debido a que se realizó de forma virtual por la pandemia.

La capacitación inicial se adaptó a un solo día, en aquellos casos en

los que la persona digitadora ingresó durante el proceso de ejecución de la encuesta[124], esta inducción se combinó con la entrega de material de apoyo y el manual de digitación. Adicionalmente, para las personas nuevas, existió una fase de adaptación, que implicó una menor carga de trabajo en las primeras semanas de la digitación. Junto a esto, se realizó un proceso de acompañamiento por parte de una persona encargada de grupo y digitadora con mayor experiencia, las que fueron guiando el trabajo de las personas nuevas y resolviendo dudas.

Las capacitaciones realizadas a las nuevas encargadas de grupo de digitación, que ingresaron durante la ejecución de la encuesta, también se adaptó un proceso de inducción cuya duración fue de un día. El propósito fue entregar conocimientos básicos y necesarios que permitieran desarrollar sus laborales, de la misma forma que la persona digitadora, se implementó una semana de integración en la que fue acompañada por la persona coordinadora y encargada de grupo de mayor experiencia.

El proceso de capacitación continua se generó cuando se detectaron errores o inconsistencias en la aplicación de instrucciones o criterios de digitación, se realizó una retroalimentación (individual o grupal dependiendo de la temática a tratar) en la que se reforzaban o entregaban nuevos protocolos para mejorar el proceso de digitación de encuestas. Además, cada vez que se terminaba de digitar una submuestra operativa, se realizaba una reunión de equipo para reforzar los criterios que se debían utilizar, aclarar dudas y/o tratar temas administrativos.

[124] Debido a la alta rotación de personal durante la ejecución de la encuesta.

6.6 Sistema Informático

La gestión del proceso de recolección se realizó a través de un sistema informático desarrollado por el INE para la EPF, denominado Sistema de Gestión Operativa (SGO). Tal como fue explicado previamente (apartado 6.3.3), este sistema estaba integrado con la plataforma *Headquarters,* de *Survey Solutions.*

Las funciones del SGO fueron creadas a partir del sistema operativo de la VIII EPF, se realizaron cambios y adaptaciones derivadas del cambio tecnológico de la recolección mediante la incorporación de dispositivo móvil y la necesaria integración con *Survey Solutions,* de tal forma que se mantuvo aquellas funcionalidades asociadas al ingreso de cuestionarios en papel, para hacer factible la digitación de datos de Gastos Diarios y Gastos del Recuerdo.

6.6.1 Funcionalidades asociadas a la ejecución de la recolección

El sistema informático cumplió un rol fundamental en las labores de monitoreo realizadas por la coordinación nacional de recolección, facilitando el acceso a información actualizada. Las funciones más relevantes asociadas al sistema informático fueron las siguientes:

- **Gestión muestral:**
Luego del ingreso de las bases de datos de las viviendas seleccionadas, el sistema permitía la asignación de las cargas de trabajo, habilitaba la creación de las llaves de entrevistas en los dispositivos móviles de captura, tener acceso a los módulos digitales, revisar la gestión de las personas encuestadoras a través de la hoja de ruta, permitía la creación de nuevos hogares en las viviendas seleccionadas y la creación e impresión de las guías de despacho.

- **Revisión de cuestionarios:**
El sistema permitió el ingreso al cuestionario digital, en el que podía ser revisado por la persona encargada de grupo, quien tenía la posibilidad de incorporar observaciones y comentarios[125], y así indicar la presencia de errores, omisiones o inconsistencias de información de cada módulo. También contaban con el Formulario de Correcciones, en que se ingresaban los errores e inconsistencias de los cuestionarios en papel. En este registro se debía clasificar el tipo de error, a través de la identificación de la pregunta y el número identificador de la persona del hogar (número de línea) en la que se presentaba el problema. En ambos casos, se podían ingresar instrucciones específicas de corrección; para los módulos digitales se ingresaban como comentarios en cada una de las preguntas y para los cuestionarios en papel se podía imprimir el formulario de correcciones y así entregar las instrucciones de corrección y de recuperación de datos por escrito.

[125] Cabe aclarar que las y los encargados de grupo no podían editar la información registrada por la persona recolectora, siendo esta ultima la única con la facultad de hacer ediciones durante la recolección de datos.

- **Consultas y solicitudes de autorización de medidas especiales:**

El sistema permitía orientar a los equipos de recolección frente a dificultades detectadas durante la etapa de recolección. Las solicitudes podrían estar referidas a problemas de identificación de unidades muestrales, requerimientos de actualización de información domiciliaria, modificación de plazos de entrevistas, modificaciones de números de visitas solicitadas por el hogar y/o la modificación de la submuestra operativa asignada.

Estas solicitudes eran analizadas por el equipo de coordinación nacional de recolección, el cual sugería posibles estrategias de solución ante cada problemática, autorizando o rechazando cada solicitud. La importancia de este módulo es que se generaba un registro de los ajustes realizados y los motivos asociados a ellos, y así poder contar con más información respecto del contexto en que se ejecuta la etapa de recolección.

- **Descarga de bases:**

Con esta información entregada por el SGO y SuSo la analista de datos lograba general el cálculo de los indicadores de desempeño operativo, reportando así durante dos veces al día un informe con los resultados iniciales y preliminares a nivel nacional, y desagregado a nivel de región, personas encargadas de grupo y encuestadoras. Este informe presentaba estos datos a nivel de submuestra operativa y acumuladas, es decir, con el conjunto de submuestras operativas cerradas hasta la fecha de corte de los datos.

Debido a la implementación de un módulo digital en la IX EPF, generando un cambio en la recolección se debió implementar la tecnología necesaria para aquello, lo que incorporó el uso de un dispositivo móvil de captura (DMC) o Tablet donde se instaló la aplicación Interviewer de SuSo para la recolección de los datos. Gracias a estos las personas encuestadoras podían ingresar la información recogida en terreno con los hogares participantes y luego de un proceso de sincronización de los datos estos podían ser visualizados

- **Consultas y solicitudes de autorización de medidas especiales:**

El sistema permitía orientar a los equipos de recolección frente a dificultades detectadas durante la etapa de recolección. Las solicitudes podrían estar referidas a problemas de identificación de unidades muestrales, requerimientos de actualización de información domiciliaria, modificación de plazos de entrevistas, modificaciones de números de visitas solicitadas por el hogar y/o la modificación de la submuestra operativa asignada.

Estas solicitudes eran analizadas por el equipo de coordinación nacional de recolección, el cual sugería posibles estrategias de solución ante cada problemática, autorizando o rechazando cada solicitud. La importancia de este módulo es que se generaba un registro de los ajustes realizados y los motivos asociados a ellos, y así poder contar con más información respecto del contexto en que se ejecuta la etapa de recolección.

- **Descarga de bases:**

Con esta información entregada por el SGO y SuSo la analista de datos lograba general el cálculo de los indicadores de desempeño operativo, reportando así durante dos veces al día un informe con los resultados iniciales y preliminares a nivel nacional, y desagregado a nivel de región, personas encargadas de grupo y encuestadoras. Este informe presentaba estos datos a nivel de submuestra operativa y acumuladas, es decir, con el conjunto de submuestras operativas cerradas hasta la fecha de corte de los datos.

Debido a la implementación del cuestionario digital en la IX EPF, se generó un cambio en la recolección debido a la implementación de la tecnología necesaria para aquello, lo que incorporó el uso de un dispositivo móvil de captura (DMC o Tablet) donde se instaló la aplicación *Interviewer* de *Survey Solutions* para la recolección de los datos. Gracias a esta aplicación las personas encuestadoras podían ingresar la información recolectada en terreno con los hogares seleccionados y luego de un proceso de sincronización de los datos estos podían ser visualizados por los encargados y encargadas de grupo en el SGO y así dar fluidez y seguimiento a las respectivas etapas del complejo proceso de recolección de datos.

6.6.2 Funcionalidades asociadas a la supervisión de la recolección

La implementación y ejecución del proceso de supervisión requirió de herramientas que permitieran registrar y realizar el seguimiento de la gestión operativa de las distintas actividades durante la recolección de datos. Para ello, fueron desarrollados módulos de supervisión incorporados en el sistema de gestión operativa (SGO), los que tenían como objetivo facilitar el acceso a la información de las unidades a supervisar y poner en disposición los datos relacionados con la gestión operativa de la encuesta. Los principales módulos utilizados en la supervisión son:

- **Gestión de unidades a supervisar:**

Cuando el directorio fue cargado en el sistema por la persona encargada nacional de recolección, los módulos de supervisión asociados a (i) asignar área de trabajo y (ii) asignación de carga de supervisión, permitieron identificar todas las unidades a gestionar. Este último posibilitó la visualización de todas las unidades disponibles por los distintos perfiles del equipo de supervisión mediante filtros asociados a la submuestra; región, folios disponibles, folios asignados y tipo de supervisión. En el caso del filtro asociado a los folios asignados, el sistema permitió desagregar por personas encuestadoras, encargadas de grupo, supervisoras y tipo de unidad, ya sea vivienda u hogar, lo que facilitó la administración de la carga de trabajo.

- **Administración de usuarios:**

Consistió en un módulo que permitió la creación, gestión y edición de los perfiles del equipo de supervisión y la integración con los usuarios utilizados en la plataforma de *Survey Solutions*[126], mediante la gestión web y los aplicativos en los dispositivos móviles utilizados por las personas supervisoras.

- **Formulario de correcciones**:

Fue un módulo utilizado por las personas supervisoras y encargadas de grupo, la cual permitió el registro de las revisiones de los cuestionarios de Gastos del Recuerdo y Gastos Diarios. Específicamente, se realizó el registro de errores, posibles inconsistencias u omisiones de información, identificando el cuestionario, módulo y la pregunta donde se detectó el problema. Además, se clasificó el tipo de error, esto aseguró que las personas supervisoras ingresaran instrucciones detalladas para su corrección. El equipo de supervisión generó los reportes de correcciones por cada hogar que fuera supervisado, con el propósito de entregar estos reportes a los equipos de recolección y proceder a su corrección o recuperación de datos.

[126] Usuarios de sistema son las personas que utilizan la plataforma SuSo para llevar a cabo las tareas de supervisión, ya sea como "entrevistadores" o "supervisores".

Además, esta herramienta fue utilizada como el canal de comunicación oficial para informar la detección de hallazgos en los procesos de supervisión indirecta (detallada en la sección 6.3.6.4). Todas las diferencias identificadas durante este tipo de supervisión luego de su conciliación fueron registradas e informadas por este módulo del sistema de gestión operativo.

6.6.3 Funcionalidades asociadas a la digitación de cuestionarios en papel

A continuación, se describen los módulos más utilizados del sistema informático de la IX EPF en el proceso de digitación.

- **Folios asignados para digitación:**

Registro de todos los hogares de la submuestra seleccionada con sus respectivos cuestionarios. Este módulo permitió consultar información referente a los hogares asignados para digitar, identificando a la persona digitadora responsable y aquellos hogares que debían ser asignados.

- **Estado de cuestionario:**

Dicho reporte entregaba un listado con la información de todos los cuestionarios de cada hogar y el estado en que se encontraba la digitación. A partir de la revisión de este módulo se controlaba que las personas digitadoras fuesen digitando su material según los plazos establecidos. Al finalizar cada una de las submuestras, se verificaba que todos los cuestionarios estuviesen digitados y finalizados.

- **Campos digitados:**

Visualizaba la información de los cuestionarios asignados y la cantidad de cuestionarios digitados en cada submuestra por la persona digitadora. Este reporte sirvió para revisar el avance de cada una, comprobando la cantidad de hogares y cuestionarios asignados con la cantidad exacta de registros digitados. Sobre la base de esta información se realizaba la revisión de las cargas de trabajo y eventuales reasignaciones de cargas entre el equipo de personas digitadoras.

07.

METODOLOGÍA DE PROCESAMIENTO DE DATOS

07.

METODOLOGÍA DE PROCESAMIENTO DE DATOS

Se define como procesamiento, la operación que convierte los datos brutos (información que no ha sido organizada y/o preparada) en información adecuada para su análisis, agregación y generación de informes de resultado. En este apartado, se desarrollan todos los subprocesos llevados a cabo en la fase de procesamiento de datos.

Es importante destacar que la IX EPF, por primera vez, realizó su recolección de datos utilizando Dispositivos Móviles de Captura (DMC) en terreno, para el llenado de sus Módulos de Hoja de Ruta (HR), Registro de Personas en el Hogar (RPH), Gastos Financieros y Seguros (GSF), Ingresos (IN), Gastos en Educación (GE) y Gastos periódicos del Hogar (GH). Los únicos módulos impresos y encuestados a través de cuestionarios en papel fueron Gastos Diarios (GD) y Gastos del Recuerdo (GR).

Como métodos de recolección, la IX EPF utilizó dos sistemas informáticos; *Survey Solutions* (SuSo) que fue utilizado como plataforma de diseño y construcción, además de ser utilizado como servidor, y aplicación Android, para el soporte del cuestionario de recolección en DMC. Mientras que la gestión de trabajo de recolección de datos y soporte del proceso de digitación de cuestionarios físicos (GD y GR) estuvieron a cargo del sistema de gestión operativa (SGO).

7.1 Métodos de procesamiento de datos

El procesamiento de datos fue desarrollado por el equipo de analistas socioeconómicos de la IX EPF. Este trabajo utilizó como insumo principal todos los datos recolectados en terreno, tanto los capturados a través de DMC, como aquellos obtenidos mediante los módulos en papel, los cuales son digitados con posterioridad.

La fase de procesamiento se desarrolló bajo el alero del modelo genérico internacional GSBPM[127] y responde a un flujo de trabajo, el cual abarcó los siguientes ocho subprocesos:

[127] El GSBPM entrega un marco estándar que permite modernizar los procesos de producción estadística, desde el diseño de la encuesta hasta la publicación de los resultados. De todas formas, el GSBPM debe ser entendido como un modelo flexible, el cual ha sido adecuado a las necesidades y objetivos de la encuesta.

1. Integración
2. Clasificar y codificar
3. Revisar y validar
4. Editar e imputar
5. Derivar nuevas variables y unidades
6. Calcular ponderadores
7. Calcular agregaciones
8. Finalizar los archivos de datos

7.2 Estructura de trabajo

Estos ocho subprocesos se rigen por una estructura de trabajo que da paso a la ejecución de un flujo establecido, a fin de generar datos y archivos, los que sirvieron como fuente para el desarrollo de cada uno de los subprocesos y sus etapas. La estructura, que se divide en entregas por submuestra operativa y submuestra acumulada (muestra completa)[128], es la siguiente:

Entregas por submuestra:

- **Base Bruta (recolectada en DMC y papel)**

  •Descarga de base de datos desde el Servidor de recolección DMC

  •Descarga de cuestionarios recolectados en papel

- **Tablas integradas**

  •Adaptar la estructura de base de datos a tabla plana por módulo

  •Asignar códigos especiales debidos a flujo de cuestionario y valores perdidos

- **Tablas codificadas**

  •Asignación de códigos CIUO-08.CL, CCIF 2018.CL, Clasificador de establecimientos, Rol Base de Datos (RBD) de establecimientos de educación

[128] Entregas de submuestras acumuladas según la rendición comprometida a IPC. Estas rendiciones consistieron en el envío de tres submuestras mensuales acumuladas por trimestre y seis por semestre.

- **Tablas validadas**

  •Revisar el rango de las variables, flujos, saltos, entre otros

Entregas por muestra completa:

- **Tablas Editadas e Imputadas**

  •Editar o depurar los datos inconsistentes o con valores sin respuesta (que no se imputen)

  •Imputación de Gastos e Ingresos

- **Base Final**

  •Tabla de Gastos

  •Tabla de Personas

Como se mencionó anteriormente, luego de cada subproceso se generó un archivo de datos el cual fue nombrado haciendo referencia al módulo de la encuesta que lo generó, al subproceso al que pertenece y a la submuestra operativa desde la cual se están procesando los datos (ver **Diagrama 11**). Estos archivos se presentan como tablas intermedias, que corresponden a hitos que expresan estados diferentes de la base de datos durante el procesamiento, y tablas auxiliares, correspondientes a datos parciales que fueron usados como entrada o salida de acuerdo con una necesidad acotada y especifica. En la siguiente sección del documento se presentará en detalle cada subproceso y se utilizará la misma nomenclatura que se utilizó para nombrar estos archivos para describir cada uno de estos.

La nomenclatura para definir el nombre de cada una de las tablas generadas a lo largo del procesamiento fue la siguiente:

Diagrama 11: Nomenclatura de tablas de datos

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

7.3 Flujo de procesamiento

Si bien los subprocesos se describen de manera secuencial, es decir, existe un orden paso a paso de todos los subprocesos, en la práctica el flujo de trabajo fue iterativo. Es decir, a pesar de no seguir un orden lineal, fueron desarrollados respetando y cumpliendo la realización de cada uno de ellos. Existe una serie de instancias en las que se combinaron, tanto temporalmente como en términos de flujo de procesos, tareas que pertenecen a distintas etapas e incluso diferentes fases del proceso. El flujo de procesamiento podía retroceder a subprocesos o actividades que fueron previas de manera lineal, para trasladarse hacia etapas más avanzadas del procesamiento. Por ejemplo, se presentaron instancias en que, a pesar de encontrarse procesos ejecutados, estos se debieron volver a generar[129] para el desarrollo posterior de hitos más avanzados dentro del procesamiento **(ver anexo 7)**.

Los subprocesos descritos a continuación, dan cuenta de qué actividad se realizó en cada uno. Dentro de estos subprocesos ocurren distintas actividades, las que avanzaron mediante iteraciones cuyo producto son las tablas intermedias. Los subprocesos junto a sus etapas se describen a continuación:

7.3.1 Integración

El subproceso de integración tuvo como principal objetivo unificar y verificar los datos recolectados, comprobando la calidad y validez de cada uno de ellos. Las etapas desarrolladas en este subproceso fueron las siguientes:

1. Descarga de tablas:
Para dar inicio al subproceso de integración de datos, el encargado de operaciones notificó a la coordinación técnica del cierre de la recolección de una submuestra operativa. Luego de esto, la coordinación técnica descargó las tablas del cuestionario digital desde el sistema de dispositivos móviles, mientras que los datos de los cuestionarios en papel fueron descargados desde el sistema de gestión operativo, tras el aviso del cierre del proceso de digitación.

[129] Por ejemplo, por ajustes a la información el flujo de procesamiento considera un retorno a etapas previas.

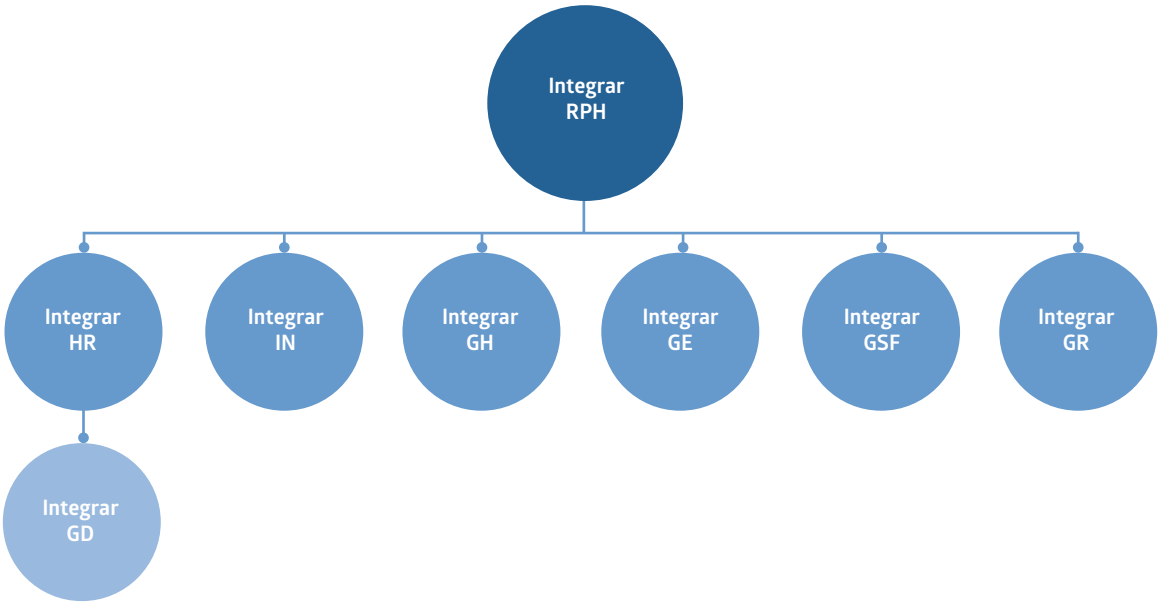
2. Verificación existencia de tablas y posibles errores:
El equipo de analistas socioeconómicos verificó la completitud de información en las tablas, reportando a la coordinación cuando se identificaron posibles errores, los cuales además podían tener dos tipos de error:

- Error de descarga
- Error de cuestionario

Al tratarse de un error de descarga, estos fueron enviados a los analistas informáticos quienes notificaron vía correo la solución del problema. Si se trató de un error de cuestionario, se envió propuesta de la corrección al error por parte de la coordinación o el analista socioeconómico con el fin de realizar nuevamente la descarga.

3. Integrar tablas:
Se realizó un chequeo de pareo de hogares y personas, si este no fue correcto, se generaron ajustes y revisiones con Hoja de Ruta y/o RPH. También se produjeron modificaciones con los módulos de recolección en papel GR y GD para consolidar y unificar la información del pareo. De ser correcto el chequeo, se procedió a integrar las distintas tablas **(ver Diagrama 12).**

Diagrama 12: Integración de Hogares



Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

El proceso de pareo de Hogares y personas consistió en la verificación de que los registros existentes en Hoja de Ruta, RPH y GD fueran consistentes, para así asegurar el correcto procesamiento posterior de registros y la inclusión de los distintos hogares e individuos en cada submuestra operativa. Dentro de los errores más frecuentes observados en el proceso de pareo se encuentran:

- **Error en la descarga de cuestionarios:**
  Según el tipo de error diagnosticado en esta etapa se generaron soluciones en el módulo correspondiente. El error en la descarga de cuestionario ocurrió cuando la descarga se realizó de manera anticipada y, por lo tanto, no estaba finalizado el proceso de digitación final. Esto se solucionó realizando una segunda descarga.

- **Error en la carga de cuestionarios:**
  La causa común para el error en la carga de cuestionarios fue no finalizar el proceso de carga desde digitación, provocando que no todos los cuestionarios fueran cargados.

- **Inconsistencias entre RPH y GD:**
  Ocurre cuando el hogar o integrantes del hogar solo se encontraban en uno de los módulos (RPH o GD), debido a clasificaciones incorrectas de la condición de finalización, es decir, el hogar se categorizó como *break off*, cuando en realidad presentó registros suficientes, o se categorizó como entrevistado cuando no cumplía los mínimos de calidad de la encuesta. Esto significó que variables como el código de disposición final (CDF) y el indicador de retiro de GD presentaran inconsistencias en el pareo. La solución fue corregir las variables mencionadas para que así coincidieran los registros en los módulos.

- **Inconsistencias entre HR y GD:**
  Una de las inconsistencias más comunes entre los módulos HR y GD ocurrió entre la edad de la persona que responde GD y los códigos de asignación y retiro del módulo HR. Ante estas diferencias se consultó con el módulo RPH con el fin de corroborar la edad y realizar las ediciones pertinentes. Adicionalmente, con el fin de corroborar el retiro efectivo de GD se verificó su existencia con el equipo de digitación para armonizar los códigos de asignación como los de retiro presentes en HR. Finalmente, se realizaron ediciones en el Código de Disposición Final (CDF) si este no era consistente con la información presente en GD,

  Luego de llevar a cabo la integración de tablas, fueron separados los registros a codificar de cada uno de los módulos involucrados según el clasificador correspondiente. Terminado el proceso, se notificó la existencia en cada módulo de las tablas para su codificación. Adicionalmente generó la integración de observaciones provenientes de recolección y las tablas a codificar para su uso complementario en el proceso de codificación.

7.3.2 Clasificar y codificar

Las etapas del subproceso de clasificación y codificación son las siguientes:

- **Recepción y revisión de base de datos provenientes del equipo técnico:**
  El objetivo de esta etapa fue proveer al equipo de codificación con las tablas necesarias para realizar la codificación de los registros. En este proceso la persona encargada de Codificación y Revisión recibió y revisó las tablas provenientes del equipo técnico, verificando la cantidad de registros y la existencia de las variables entregadas necesarias para realizar el proceso de codificación. Cuando una tabla no cumplió con los estándares requeridos fue devuelta al equipo técnico para su modificación. Así también, el equipo técnico generó una tabla con la codificación automatizada de registros, la que fue revisada por el equipo de codificación, quienes agregaron sugerencias en caso de ser necesarias.

- **Codificación:**
  El objetivo de esta etapa fue realizar la codificación de los registros asignados en cada tabla. En primer lugar, se codifican todos los registros de ocupaciones, establecimientos educacionales, establecimientos comerciales y bienes y servicios consignados en la encuesta, utilizando para ello los clasificadores CIUO 08.CL, Clasificador de establecimientos educacionales (RBD) del Ministerio de Educación, Clasificador de establecimientos INE, y CCIF 2018.CL, respectivamente.

  En caso de encontrar registros complejos que dificultaron la codificación, la persona codificadora debió recurrir a la persona encargada de codificación y/o a las personas revisoras de codificación para realizar las consultas pertinentes. Así también, para los gastos registrados de manera deficiente o incompleta, se buscó información adicional en fuentes complementarias[130].

  Una vez codificada la totalidad de los registros asignados y resueltas las dudas sobre los registros complejos, la persona codificadora entregó las tablas codificadas., Finalmente, las personas codificadoras verificaron la consistencia entre la carga entregada (al inicio del proceso) y la recibida en esta instancia.

- **Revisión cruzada entre pares (post codificación):**
  El objetivo de esta etapa fue disminuir la cantidad de errores en el proceso de codificación. Para ello se intercambiaron las tablas codificadas entre pares y se procedió a realizar una revisión del trabajo realizado por otra persona codificadora. En caso de hallar errores, estos se consultaron entre el quipo y se tomó una decisión conjunta.

- **Consolidación de tablas y envío a equipo técnico:**
  El objetivo de esta etapa fue enviar el material codificado al equipo técnico.

[130] Como recurso complementario de los gastos se recurrió a búsqueda en internet se utilizó para resguardar la correcta identificación del producto, servicio o establecimiento, y así reducir los casos en que se utilizaran códigos residuales debido a marcas o establecimientos poco conocidos. Los hallazgos recurrentes se introducían en los instrumentos de apoyo a la codificación.

Una vez finalizados los procesos de codificación y revisión cruzada, la persona encargada de codificación consolidó la tabla de registros que fue entregada al equipo técnico.

Cuando existieron dudas sobre algún registro y no se podía resolver con la información disponible, se envió a revisar el módulo físico para verificar si en ella existía alguna información adicional que permitiera codificar.

- **Recodificación de tablas ex post:**

El objetivo de esta etapa fue depurar la información de las tablas codificadas para disminuir la cantidad de errores no muestrales producidos por el proceso de codificación.

Luego de entregadas las tablas al equipo técnico y una vez que este había avanzado en el proceso de análisis de datos, se generó una instancia en la que se solicitaba la revisión de un conjunto acotado de registros. Esto se produjo debido a la detección de errores, inconsistencias u omisiones en los registros codificados. También se realizaron recodificaciones de algunos productos por necesidad de los equipos de índice de precios al consumidor IPC (Departamento de Estadísticas de Precios).

La persona encargada de codificación tuvo la tarea de recibir estas tablas con los registros seleccionados anteriormente y distribuir la carga de recodificación. Posteriormente, se procedió a recodificar los registros, corrigiendo los errores detectados. La persona encargada de codificación fue la encargada de consolidar los registros recodificados para su envío al equipo técnico.

7.3.3 Revisar y validar

El subproceso de revisión y validación[131] implicó la creación de la tabla de registros ya integrada con la codificación y su posterior validación para dar cumplimiento con los criterios de calidad de los datos.

Las etapas del subproceso de revisión y validación son las siguientes:

1. Validación de tabla de registros:
En cada etapa del procesamiento se generó una malla de validación de los registros con el fin de detectar errores e inconsistencias entre estos. Adicionalmente, este proceso es funcional para vigilar el correcto avance en la calidad de los datos.

2. Primera aplicación grilla técnica 1
Esta etapa también formó parte de la validación, en ella se evaluó la suficiencia de la información provista por cada uno de los hogares, a través de la asignación de un puntaje. En caso de no cumplir con el puntaje establecido, el hogar fue excluido de la base de personas y gastos final. Aquellos hogares que no pasaron la grilla técnica, pero se mantuvieron cercanos al rango del puntaje aceptado, fueron incluidos en la fase de recuperación que permitió obtener los datos truncados si existía disponibilidad de ellos. Si esta información era válida, los hogares pudieron superar el puntaje mínimo establecido y así pasar a formar parte de la base final.

3. Conversión de unidad monetaria:
Aquellos valores de ingresos o gastos registrados con unidad monetaria distinta al peso chileno fueron convertidos a moneda nacional al tipo de cambio registrado el último día hábil de la submuestra operativa correspondiente.

4. Detección de atípicos:
Se identificaron aquellos gastos anómalos descartando potenciales errores de medición, mediante la aplicación de técnicas estadísticas de detección de atípicos y revisión de casos según criterio experto.

5. Detección de duplicados y traspasos:
Los tipos de gasto capturados por cada módulo se encuentran ligados a la periodicidad con que se consumen determinados bienes y servicios, esto se refleja en el ajuste temporal para mensualizar los gastos reportados por los hogares, esto puede ser revisado en detalle en el punto 7.3.5 de este mismo documento. En la práctica, las personas informantes pueden registrar sus gastos en módulos donde no es posible aplicar un ajuste temporal adecuado para el producto. En el caso anterior, el gasto debe ser traspasado a un módulo que sea adecuado para el ajuste temporal de este.

[131] El subproceso de Revisar y Validar, procede de manera secuencial, pero sin seguir un orden que limite avanzar y regresar a etapas anteriores para su desarrollo. Así mismo, trabaja conjuntamente desarrollando hitos intermedios, con el subproceso de Editar e Imputar (ver Anexo 6).
Se señala que durante este subproceso fueron excluidos de la base de datos aquellos hogares que no cumplieron con los estándares mínimos de calidad. Los criterios definidos como estándares mínimos de calidad responden a la denominada *grilla de calidad*.

A partir de la tabla de gastos codificados, y de manera centralizada, se identifican los gastos a traspasar entre módulos o que deben ser eliminados dado que está duplicados. Para lo primero se compara el módulo efectivo (módulo donde fue registrado el gasto) con los módulos de registro (módulos donde se puede aplicar un factor de ajuste temporal correcto con relación al tipo de gasto declarado). Cuando un gasto no se encuentra registrado en uno de los módulos de registro (propio de los gastos de su mismo tipo) es etiquetado como gasto a traspasar al módulo mandante (módulo con el factor de ajuste temporal más adecuado para tal producto). Para la detección de duplicados se detectan aquellos registros donde en un mismo hogar se duplica el gasto en un producto en dos o más módulos de la encuesta, es decir, se evalúa que el producto y el valor total sean iguales, si esto ocurre entonces se evalúa la exclusión del gasto de aquel módulo que no sea mandante.

6. Segunda aplicación de grilla técnica 2:

Se evaluó en una segunda instancia la suficiencia de información asociada a cada hogar. El objetivo de aplicar nuevamente la grilla técnica fue identificar si aquellos hogares que se encontraron cercanos al punto de corte en la primera etapa lograron superar dicho puntaje mínimo de calidad después de los procesos de recuperación y edición de datos. La decisión final en cuanto al puntaje de corte fue que este debe ser mayor o igual a 71 puntos.

7.3.3.1
Grilla técnica de calidad

La grilla técnica de calidad corresponde al método utilizado para evaluar los mínimos de calidad y cantidad de la información entregada por cada uno de los hogares que participaron en la encuesta. La evaluación se realizó estableciendo ponderaciones para cada módulo y sus preguntas prioritarias.

El propósito de la grilla técnica en la IX EPF, fue cuantificar el nivel de no respuesta al ítem y la suficiencia en el cuestionario de Gastos Diarios para clasificar a un hogar como entrevistado o rechazo *(break off)*.

La importancia atribuida a la correcta clasificación de los hogares radica en que las principales estimaciones de la encuesta deben ser calculados en base a los hogares clasificados como entrevistados. Por lo tanto, la grilla permitió excluir del análisis a los hogares que presentaron un importante nivel de subdeclaración de información, con la finalidad de reducir el sesgo en las estimaciones finales para gastos e ingresos[132].

Los procedimientos de la grilla técnica consistieron en:

- La asignación de puntaje a cada hogar (considerando todos los módulos), el cual podía ser valorado de 1 a 100 puntos, en los que se evaluaron ciertos módulos y preguntas aisladas de estos.

- Se establecieron distintos criterios de evaluación y ponderaciones para los módulos y preguntas seleccionadas, estos criterios traducen la completitud y consistencia de la información proporcionada por los hogares.

- Bajo el criterio de asignación de puntajes, la grilla estableció un puntaje de corte igual a 71 puntos de exigencia, si el hogar tiene mayor o igual puntaje es considerado como un hogar entrevistado y parte del estudio.

[132] El documento "Reclasificación de la no respuesta: distinción entre la no respuesta al ítem y la no respuesta a la unidad " describe los criterios utilizados en la VIII Encuesta de Presupuestos Familiares (EPF) para distinguir entre un hogar entrevistado y aquel que debe ser reclasificado como *break off*.

- Los hogares que no alcanzaron puntaje de 71 puntos fueron clasificados como *break off* y fueron excluidos del cálculo de las estimaciones finales de la encuesta.

- El cuestionario de GD captura un alto porcentaje del gasto de los hogares[133]. Debido a lo anterior, a este módulo se le asignó el puntaje más alto en cuanto a los criterios evaluados por la grilla técnica, verificando que todos los integrantes elegibles del hogar hubiesen respondido un mínimo de cuatro días de registro, con al menos un fin de semana incorporado dentro de este registro. Adicionalmente, del total de integrantes del hogar que debían contestar GD, se distinguió a aquel clasificado como administrador de gastos del hogar respecto de aquellos integrantes elegibles, ajustando la evaluación para que el puntaje de GD de un administrador de gastos ponderase el doble que el puntaje de un de otro tipo de integrante al interior de cada hogar. Esta distinción se realizó considerando que no todos los integrantes aportan de igual manera al gasto del hogar.

Durante el procesamiento, la grilla técnica fue aplicada en dos oportunidades durante el proceso de validación, con el objetivo de focalizar el trabajo de recuperación en aquellos hogares en donde se estimó que era posible alcanzar el mínimo de 71 puntos. Asimismo, dirimir entre unidades entrevistadas parcialmente y unidades que, por su nivel de no respuesta al ítem, debiesen ser clasificadas como entrevistas interrumpidas o *break off*.

La primera aplicación de la grilla generó los siguientes niveles de clasificación de los hogares:

- **Supera los criterios de cantidad y calidad:**
El hogar cumple satisfactoriamente con el mínimo de los criterios establecidos para ser incluido en el análisis.

- **No supera los criterios de cantidad y calidad, pero permanece en el proceso:**
El hogar no cumple de manera satisfactoria con el mínimo de criterios

[133] En la IX EPF aproximadamente un 40% del total del gasto capturado corresponde a información proveniente de este cuestionario.

establecidos para ser incluido en el análisis. No obstante, el puntaje de la grilla se acerca al puntaje de corte, por lo cual es factible recuperar datos faltantes en el proceso de revisión y recuperación de la encuesta.

- **No supera los criterios de cantidad y calidad:**
El hogar no cumple de manera satisfactoria con el mínimo de criterios establecidos para ser incluido en el análisis. La cantidad de datos faltantes implica no incluir a los hogares en los siguientes procesos y, posteriormente, en los datos finales.

La segunda aplicación de la grilla se llevó a cabo una vez finalizados los subprocesos de recuperación y edición de datos. De esta forma, al existir errores de digitación o información recuperada, se evaluó a los hogares con los datos corregidos. El objetivo de esta aplicación fue determinar el código de disposición final de los hogares contactados.

7.3.4 Editar e imputar

El subproceso de edición se llevó a cabo con la finalidad de mejorar la calidad de los datos y dar paso a la imputación de datos faltantes. A continuación, se detalla el desarrollo de ambas fases de este subproceso llevado de forma independiente.

7.3.4.1 Edición

El desarrollo de este subproceso inicia con las ediciones manuales que requieren una revisión detallada de cada analista socioeconómico. Luego, se ejecutan ediciones automatizadas con rutinas de código, que permiten la corrección inmediata de errores que se pueden asignar a un patrón establecido. Con el devenir del procesamiento, se dio paso a nuevas ediciones en conjunto con las iteraciones incluidas (detalladas más adelante), en conjunto con el subproceso de revisión y validación. Este subproceso estuvo compuesto de seis etapas:

1. Primera edición manual : Ediciones surgidas de la revisión de observaciones o módulos físicos (GD y GR), para casos en lo que fue factible solucionar el error de manera automática.

2. Segunda edición manual y automática : Edición posterior al segundo hito en el subproceso de validación. En la cual se procedió a editar manualmente cuando fuere necesario para dar paso a la edición automática en aquellos casos que se requirió. Es en esta etapa del subproceso de Edición, en donde se une la edición manual 2 junto con la codificación correspondiente al mismo hito.

3. Edición de gastos duplicados y traspasos: Debido a la extensión de la EPF y la exigencia que existe sobre las personas informantes, existe siempre la posibilidad de que se declare más de una vez un mismo gasto (por ejemplo, el pago de una matrícula escolar registrada en GE y GD). Es por ello, que esta etapa tiene como objetivo principal identificar mediante el uso de ciertos criterios, aquellos casos en que se duplica el registro del gasto. Al finalizar esta etapa, es posible identificar y editar los registros duplicados para eliminar esta característica, dejando solo una observación en el módulo óptimo para su registro.

4. Edición de atípicos: Una vez identificados los registros atípicos de cada módulo se procedió a editarlos de manera manual.
A continuación, se describe el procedimiento de detección de atípicos realizado en cada uno de los módulos:

a. Gastos Diarios: IQR (rango intercuartílico) y método de Chebyshev[134], sumado a revisiones adicionales para productos en división 1 y 2 del clasificador CCIF.

b. Ingresos: Se utilizó el método Percentil 99 sobre las submuestras operativas acumuladas.

c. Gastos del Recuerdo: Se utilizó el método Chebyshev.

d. Gastos en Servicios Financieros y Seguros: Se utilizó el método IQR (rango intercuartílico).

[134] El teorema de Chebyshev establece que para cualquier número k mayor que 1, al menos 1 – 1 / k 2 de los valores de los datos en cualquier distribución de forma se encuentran dentro de k desviaciones estándar de la media.

e. Gastos del Hogar: Se utilizó el método IQR (rango intercuartílico).

f. Gastos en Educación: Se utilizó método IQR (rango intercuartílico).

Una vez detectados los valores atípicos *(outliers),* según modulo y característica del dato[135], fueron consolidados en un único archivo, para ser sometidos al subproceso de recuperación descrito en el apartado 6.1.3.3. Ello quiere decir que los datos recuperados se validan, para luego ser incorporados en las distintas tablas por módulo.

5. Edición de recuperación: Luego del subproceso de recuperación, no todos los datos con error lograron ser editados, lo que hizo necesario llevar a cabo un procedimiento para que dichos datos fueran validados. Enfocado en ello, el objetivo de esta etapa fue que todos los datos tuviesen un valor válido, a excepción de aquellos casos que debieron ser recuperados. Lo anterior implicó que todos los datos inválidos que no lograron modificarse en el subproceso de recuperación recibieron algún tratamiento automatizado definido previamente. Algunos ejemplos de ello son: el llenado de campos vacíos con valor -88 (no responde), corrección de fechas en GR para generar consistencia con el período de referencia, corrección de gastos de la vivienda secundaria en función de lo declarado en RPH y modificación en datos para el módulo de GH para no superar la periodicidad anual, entre otros.

6. Editada final 6: Como último hito de la validación final, se realizó una recodificación de los hogares que obtuvieron mejora en cuanto a calidad de los datos entregados en primera instancia luego de pasar por el proceso de recuperación, además de ediciones automáticas y manuales para los datos que no lograron mejorar su calidad. Por ejemplo, aquellos gastos poco detallados que debieron ser recodificados con código CCIF 00.0.0.00.00 quedando fuera.

[135] Esto quiere decir que no necesariamente todos los valores atípicos se enviaron a recuperación. Por ejemplo, los gastos atípicos provenientes del cuestionario de Gastos Diarios, no fueron elegibles para enviar a recuperar, debido a las características de estos gastos y la distancia temporal entre el período de captura de información con la recuperación.

7.3.4.2
Imputación

Para este procedimiento y al igual que con la detección de gastos atípicos, se utilizó la tabla validada más avanzada disponible por submuestra. Esta tabla es la Validada 6, la que se utilizó de forma acumulada con las 24 submuestras operativas procesadas.

Luego de acumular las submuestras, se procedió a identificar aquellos registros para imputar. La imputación abarcó tanto a los datos faltantes como a los valores atípicos aislados.

A continuación, se dan a conocer en detalle los métodos de imputación utilizados en la IX EPF:

1. Imputación Cuestionario de Gastos Diarios

Al igual que en la VIII EPF, se decidió imputar aquellos cuestionarios que tienen información incompleta (poseen al menos un día sin registro), como también aquellos cuestionarios rechazados (cuestionarios sin ningún día de registro, rechazados directa o indirectamente o no devueltos por la persona informante). Este procedimiento busca hacerse cargo del aumento en la no respuesta parcial en GD, siendo importante señalar su amplia utilización, no solamente en la VIII EPF, sino también en otras encuestas de gastos e ingresos de países como Australia (Australian Bureau of Statistics, 2015-2016), Argentina (Instituto Nacional de Estadística y Censos (INDEC), ciudad Autónoma de Buenos Aires, 2019), Brasil (Instituto Brasileiro de Geografia e Estatística (IBGE), 2011), España (Instituto Nacional de Estadística de España, 2022), Reino Unido (Great Britain and Northern Ireland, Office for National Statistics, ONS., 2017) y Sudáfrica (Statistics South Africa, 2011).

Una de las conclusiones del documento "Métodos de Imputación VIII EPF: Gastos diarios e ingresos de la actividad laboral y jubilaciones" (INE Chile, 2019) estableció que la imputación de cuestionarios completos mediante *hot deck* es beneficiosa en términos de reducción de sesgos inducidos por la no respuesta parcial, es robusto hacia subgrupos poblacionales con mayor propensión a no colaborar y, finalmente, es parsimonioso en la difusión de resultados y replicación de resultados. Sin embargo, antes de profundizar en la metodología de imputación en GD y sus características, es necesario cuantificar la magnitud de la no respuesta parcial.

Tabla 37: Estadística de la no respuesta parcial en Gastos Diarios a nivel de personas que debían contestar el cuestionario

Cuestionarios GD	Cantidad	Porcentaje
Cuestionarios GD esperados (1)	35.033	100%
Cuestionarios GD contestados (completos y parciales)	26.940	76,9 %
Diferencia (no contestados)	8.093	23,1%

(1) Para este cálculo, se restan aquellos códigos asociados a informantes que no debían contestar este instrumento (informantes no elegibles).

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Se define como persona informante esperada aquella que es integrante del hogar, tiene 15 años o más y se encuentra en condiciones físico y/o mentales para responder[136]. Por lo tanto, la no respuesta parcial se contabiliza a partir de aquella persona informante esperada que entrega un cuestionario sin ningún día de registro, no acepta contestar GD o no entrega el cuestionario físico a la persona entrevistadora al cerrar la recolección en el hogar.

En la **Tabla 37**, se observa que la no respuesta parcial es de un 23,1% del total de los cuestionarios esperados, mientras que a nivel de días **(Tabla 38)** la no respuesta parcial aumenta a un 28,2%, considerando los días sin registro tanto de los cuestionarios con información parcial como de aquellos rechazados. Según los datos recolectados en la VIII EPF[137], GD captura cerca del 42% del gasto total de los hogares, siendo el cuestionario con mayor peso relativo en la Encuesta de Presupuestos Familiares. Lo anterior resalta la importancia de abordar la no respuesta mediante procesos de imputación de datos faltantes.

[136] Adicionalmente, se establece como requisito que la persona esté 7 días o más de forma continua en el hogar.

[137] En la VIII versión de la EPF, el cuestionario de GD se denominaba Libreta de Gastos Individuales (LGI).

Tabla 38: Estadística de la no respuesta parcial en Gastos Diarios a nivel de días, considerando personas que debían contestar el cuestionario

Categorías	N° de días	Porcentaje
Días con registro[1]	382.387	71,7
Días con gastos	205.726	38,6
Días sin gastos	176.661	33,1
Días sin registro	150.621	28,3
Total de días	533.008	100

(1) Los días con registros corresponden a la suma de los "Días con gastos" y "Días sin gasto".

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Tal como se introdujo anteriormente, el método seleccionado para realizar la imputación de datos faltantes en la IX EPF fue mediante *hot deck*. Este método consiste en reemplazar las observaciones faltantes *(missing)* con observaciones provenientes de la misma base de datos a través de un criterio definido. Esta selección se puede realizar de diversas maneras; secuencial, vecinos cercanos y aleatorio (Lohr, 2021).

En una selección secuencial, se sustituye la información perdida de una unidad observacional por la última leída por el computador dado un orden establecido para la tabla de datos; mientras que la selección basada en vecino cercano define una función de distancia entre observaciones e imputa el valor de una unidad respondiente a la unidad con observación perdida que tenga la distancia más cercana. Finalmente, la selección aleatoria consiste en generar un conjunto *(pool)* de donantes, basados en un grupo de covariables, donde se selecciona uno de estos al azar y su(s) valor(es) se imputan a la unidad observacional con datos perdidos.

En la IX EPF, la imputación de datos faltantes consistió en dos etapas que se detallan a continuación:

I. Imputación de cuestionarios incompletos:

a. Se sustituye un día sin registro por otro con registro *autodonado* de la misma persona. La condición para esto es que el día donado debe corresponder al mismo día de la semana. Este proceso continúa siempre que exista un día de la semana símil con registro.

b. Si el proceso anterior dejó algún día sin registro por falta de día símil con registro, se procede a buscar un día proveniente de algún cuestionario completo de otra persona. La búsqueda del donante se realiza mediante *hot deck* aleatorio con donantes de características similares.

II. Imputación de cuestionarios rechazados:

Se busca un donante desde un cuestionario completo, el receptor recibe todos los registros provenientes del donante. La búsqueda se realiza a través de *hot deck* aleatorio con donantes de características similares.

Siguiendo los procedimientos establecidos en la VIII EPF (INE Chile, 2019), la creación de grupos para realizar la donación de registros mediante *hot deck* está basada en la selección de variables que cumplen tres requisitos:

- **Presentar una correlación significativa y relevante con el gasto:**

Dado que la variable de interés en la IX EPF es el gasto del hogar y, por consiguiente, su estructura, es necesario que las variables utilizadas para construir los grupos tuviesen una correlación alta y significativa con el gasto total por cuestionario. Esto permitió seleccionar variables para la construcción de grupos que fueron significativas y acataron el principio de parsimonia.

- **Explicar la probabilidad a responder GD:**

Cuando la probabilidad a responder el cuestionario GD difiere en las personas informantes, produciendo sesgo en las estimaciones finales.

Esta situación se genera porque hay información valiosa de diversas características que no se captura. En el caso de la IX EPF es posible perfilar grupos en la muestra que tienen distinta probabilidad de respuesta según sus características socioeconómicas, generando subrepresentación y, por ende, sesgo en la estimación final. Para generar los grupos de imputación se consideraron estos perfiles construidos bajo un modelo probabilístico no lineal.

- **Estar dentro de las variables utilizadas en el diseño muestral:**

El mecanismo de imputación debe considerar las probabilidades de selección de cada unidad. Es deseable que una unidad imputada tenga las mismas (o similares) probabilidades de selección que el donante, así al ser extrapolados sus resultados mediante el factor de expansión no se generen sesgos en las estimaciones. Por lo tanto, controlar por las variables del diseño muestral es una medida necesaria.

Encontrar el donante óptimo suele ser una tarea difícil de lograr, porque para que esto ocurra donante y receptor debe compartir características iguales. Sin embargo, la situación óptima pocas veces sucede debido a que la muestra no es lo suficientemente grande para obtener la cantidad de donantes necesaria. Por ello se tuvo que crear un esquema de flexibilización en la búsqueda de donantes, esto consistió en establecer niveles de exigencia jerárquicos, donde se van eliminando variables de forma secuencial hasta que todos los receptores consiguen un donador. La eliminación de las variables estuvo basada en los criterios de selección y, adicionalmente, en la conservación de variables geográficas y temporales para mantener la consistencia en los datos imputados respecto a los datos recabados.

En la **Tabla 39** se exhiben estadísticas de resumen para el proceso de imputación de GD. Se observa que el gasto promedio por cuestionario antes de la imputación por no respuesta es menor en $2.381, donde el gasto promedio por cuestionario aumenta de $240.752 a $243.133. Esto es debido a que el procedimiento de imputación suplió la falta de respuesta de todos los días sin registro para los cuestionarios incompletos y rechazados.

Se observa que los cuestionarios completos antes de la imputación presentan el mayor gasto promedio por cuestionario dentro de las clasificaciones presentadas, esto es porque tanto administradores de gasto y sustentadores del hogar presentan las mayores tasas de completitud en la muestra. En la **Tabla 40** se presentan las estadísticas muestrales de completitud para la condición de sustentador del hogar y administrador del gasto, como es posible apreciar aquellos administradores y sustentadores concentran el 43,8% de los cuestionarios completos, mientras que, para aquellas personas con la condición única de administradora de gastos, la cifra alcanza un 28,6% de los cuestionarios completos.

Tabla 39: Comparación de estadística descriptiva entre datos imputados y sin imputar, a nivel muestral

Comparación cuestionarios de GD individuales	N° Obs.	Promedio	Mínimo	Máximo	SD	CV	p25	p50	p75
Cuestionarios GD completos y parciales sin ajustes	26.940	240.752	0	3.815.405	270.585	1,1	50.427	160.225	339.738
Sólo cuestionarios de GD completos	21.475	252.484	0	3.815.405	280.123	1,1	53.796	171.950	356.915
Cuestionarios de GD post imputación	35.033	243.133	0	3.815.405	278.384	1,1	47.164	171.950	343.791

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Tabla 40: Completitud de cuestionarios según clasificación de administrador de gastos y sustentador del hogar

Clasificación	Cantidad de Cuestionarios	Porcentaje de Cuestionarios
Administrador y sustentador	9.414	43,8%
Solo administrador	6.145	28,6%
Solo sustentador	992	4,6%
Otro	4.924	22,9%
Total	21.475	100,0%

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

En la **Tabla 41** se muestran los efectos de la imputación en GD sobre las estimaciones del gasto promedio por hogar. Después de los procedimientos de imputación en GD, el gasto promedio por hogar estimado aumenta en $145.877. Adicionalmente, las divisiones CCIF 1, 2 y 11 presentan un mayor aumento porcentual relativo en términos de gasto promedio por hogar. Lo anterior es debido a que estas divisiones presentan la mayor cantidad de registros en GD.

Tabla 41: Estructura por división del clasificador de CCIF del gasto de consumo final de los hogares pre y post imputación de GD, a nivel expandido (excluye arriendo imputado)

División CCIF	Glosa	Total de hogares antes del ajuste		Total de hogares después del ajuste	
		Gasto promedio por hogar	% del gasto respecto al total	Gasto promedio por hogar	% del gasto respecto al total
1	Alimentos y bebidas no alcohólicas	232.994	17,8	307.947	21,2
2	Bebidas alcohólicas, tabaco y estupefacientes	16.009	1,2	21.431	1,5
3	Vestuario y calzado	40.488	3,1	41.227	2,8
4	Vivienda, agua, electricidad y otros combustibles	228.630	17,5	232.879	16,0
5	Muebles, equipamiento para el hogar y mantenimiento rutinario de la vivienda	81.620	6,3	87.882	6,1
6	Salud	107.683	8,3	115.283	7,9
7	Transporte	206.846	15,8	217.905	15,0
8	Información y comunicación	96.402	7,4	96.907	6,7
9	Recreación, deportes y cultura	55.835	4,3	64.399	4,4
10	Servicios de educación	57.782	4,4	57.791	4,0
11	Servicios de restaurantes y alojamiento	71.075	5,4	88.262	6,1
12	Seguros y servicios financieros	65.817	5,0	65.821	4,5
13	Cuidado personal, asistencia social y bienes y servicios diversos	44.726	3,4	54.046	3,7
Total		1.305.905	100	1.451.782	100

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

2. Imputación de Ingresos

Para la imputación de ingresos, al igual que en la VIII EPF, el Método utilizado fue *hot deck*. La corrección de la no respuesta se concentra en las principales fuentes de ingreso, tales como los ingresos laborales derivados de la actividad económica principal, y los provenientes de jubilaciones o pensiones de vejez. Las tasas de no respuesta parcial según las diferentes fuentes se pueden observar **Tabla 42.**

Tabla 42: Estadística de la no respuesta parcial de ingresos, a nivel de personas

Categoría en la Ocupación principal	N° obs con Respuesta	% Obs con Datos Faltantes	% Obs con Datos Faltantes	N total
Dependientes	13.542	1.498	10,0%	15.040
Asalariados	12.934	1.445	10,0%	14.379
Honorarios	608	53	8,0%	661
Independientes	4.460	737	14,2%	5.197
Profesionales	599	92	13,3%	691
Cuenta propistas	3.861	645	14,3%	4.506
Jubilados	6.034	303	4,8%	6.337

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Las personas consideradas como trabajadoras dependientes corresponden a las que declaran una relación de subordinación y dependencia en su relación laboral, sin implicar necesariamente la existencia de un contrato laboral. Lo anterior se materializa en una regularidad en el pago de su salario, tenencia de jefatura y cumplimiento de horario, entre otros.

En la imputación de ingresos, los trabajadores dependientes se clasificaron en dos grupos independientes, asalariados y trabajadores a honorarios. Esta separación surge debido a que cada grupo posee características diferentes, las cuales fueron consideradas al momento de imputar los datos faltantes. Estas características radican en el pago de cotizaciones de salud y previsión, donde los honorarios realizan con montos proporcionales distintos y con temporalidades que pueden diferir respecto a los trabajadores asalariados.

Para la elección de variables se utilizaron los procedimientos abordados en la VIII EPF (INE Chile, 2019), estas consisten en:

a. Variables teóricas:

Estas variables son consideradas como determinantes del ingreso según diversos trabajos académicos. En lo que respecta a la selección, las variables escolaridad, edad, edad al cuadrado y sexo, están fundamentadas por el modelo de capital humano, teoría del ciclo de vida, y más recientemente, el modelo de señalización (Weiss, 1995) (Mincer, 1974). El modelo de capital humano y señalización ha mostrado evidencia de la correlación entre el ingreso y la escolaridad. Así mismo, se ha constatado la importancia de utilizar la variable edad para aproximarse a la experiencia laboral y, la edad al cuadrado para suavizar los retornos de la actividad laboral que van disminuyendo con el tiempo. Además, ha sido confirmada y estudiada la diferencia salarial entre hombre y mujeres (brecha de género).

Además, se ha considerado la incorporación de variables como la clasificación socioeconómica de la UPM (CSE) y las variables geográficas consideradas en el diseño muestral de la encuesta. Las probabilidades de selección de una UPM están condicionadas por estas variables, lo que hace relevante incluirlas como forma de control (Andridge & Little, 2009).

En el contexto de las jubilaciones, la evidencia empírica revela una conexión estrecha entre la edad y el monto ahorrado (no registrado en la encuesta), lo cual justifica la inclusión de esta variable. Además, se ha observado que las jubilaciones están relacionadas con la trayectoria laboral y el sexo de la persona, se observa que las mujeres tienen menos períodos de cotización y se jubilan a una edad más temprana, por lo tanto, tendrán pensiones menores (CEPAL, 2016).

b. Variables empíricas:

Se utilizaron variables generadas en la IX EPF que presentan una alta correlación con los ingresos y una significancia como variables explicativas en modelos de ingresos. Las cuales se utilizaron en la VIII EPF, a partir de la realización de evaluaciones de diversas variables que podrían estar relacionadas con los ingresos, derivando en la selección de las siguientes: Clasificador Internacional Uniforme de Ocupaciones

(CIUO)[138], Clasificación Internacional de la Situación en el Empleo (CISE), Sustentador principal y presencia de menores de 6 o 15 años en el hogar. En el caso de las personas que reciben ingresos por jubilación y/o pensiones de vejez, también se incluyó como variable el sistema previsional al que estaba afiliada cada persona.

A partir de las variables seleccionadas, se crearon conjuntos de individuos mediante la formación de *clústeres* de imputación. Debido a que la búsqueda de individuos similares en función de todas las variables definidas puede ser exigente, se implementaron reglas para establecer un orden de flexibilización de las variables utilizadas en la formación de los grupos. Para llevar a cabo este proceso, se tuvieron en cuenta tres elementos importantes que también ayudaron a priorizar la construcción de los niveles de exigencia de los grupos. Para este fin, se crearon matrices de exigencia;

- Variables correlacionadas con el ingreso.
- Variables que explican la estructura del ingreso.
- Variables relacionadas con la propensión a responder.

La determinación del orden en el cual las variables serían relajadas y agrupadas en las matrices de exigencia se realizó considerando su nivel de correlación con los ingresos, su criterio de inclusión basado en variables teóricas y empíricas, así como el criterio de eliminación de variables. Para construir la matriz de exigencia, las variables se agruparon y se relajaron gradualmente hasta que se eliminaron aquellas variables que no estaban altamente correlacionadas. Una vez que esto sucedió, se volvieron a aplicar las variables que se habían definido previamente en el proceso de toma de decisiones correspondiente a la primera fase de este subproceso.

Antes de generar la tabla mencionada, se definieron un listado de criterios que fueron alojados en una tabla que contiene todos los puntajes ya definidos. El método utilizado es similar a lo implementado en la VIII EPF, sin embargo, el planteamiento de la selección aleatoria y el criterio de flexibilización de variables fueron actualizados según las pruebas realizadas para la IX EPF.

En la **Tabla 43**, se pueden evidenciar los principales puntos de la distribución antes y después de imputación por *hot deck*, en cada fuente de ingresos que fue imputada. En general, el rango intercuartil y la mediana presentan leves cambios con la imputación. Del mismo modo, los coeficientes de variación se mantienen estables después del proceso de imputación realizado.

Se observa que los comportamientos de los estadísticos descriptivos estudiados no presentan grandes variaciones entre los datos observados e imputados. Esto se

[138] CIUO, estandariza ocupaciones como variable categórica; además, cuenta con un componente teórico, ya que existe evidencia de la relación del puesto de trabajo y el ingreso.

debe, en parte, a que se está imputando dentro de rangos aceptables de no respuesta, tal como indica la Tabla 42 (Medina & Galván, 2007).

Finalmente, en la **Tabla 43** se presentan los resultados de la imputación de ingresos, en los que se compara la estructura del ingreso total promedio de los grupos de ingreso con y sin imputación.

Al comparar la estructura del ingreso total promedio de las personas, se observa que la imputación realizada, lo incrementa en un 8,2%. Específicamente la imputación produce un aumento de un 10,27% en los ingresos de ocupación, y de un 5,5% en los ingresos por jubilaciones o pensiones de vejez. No obstante, se estima que la imputación en estas partidas no altera la estructura en grandes proporciones, considerando que el peso relativo de cada partida es similar al presente en la estructura del ingreso sin imputar.

Tabla 43: Resultados muestrales del ingreso bruto promedio por persona según fuentes de ingreso imputadas por método *Hot Deck* (HD) aleatorio

Por fuente de ingreso	N° Obs.	Promedio	Mínimo	Máximo	SD	CV	p25	p50	p75
Asalariados	11.586	$988.034	$0	$21.000.000	$1.175.082	1,2	$400.000	$620.000	$1.150.000
Asalariados HD	14.379	$965.269	$0	$21.000.000	$1.120.249	1,2	$407.000	$616.295	$1.100.711
Honorarios	548	$979.474	$0	$13.000.000	$1.186.385	1,2	$414.500	$677.677	$1.178.750
Honorarios HD	690	$937.041	$0	$13.000.000	$1.099.396	1,2	$410.064	$658.125	$1.100.000
Cuenta propistas	3.812	$490.612	$0	$20.000.000	$825.527	1,7	$110.750	$300.000	$500.000
Cuenta propistas HD	4.506	$488.610	$0	$20.000.000	$791.783	1,6	$120.000	$300.000	$500.000
Profesionales	580	$991.836	$0	$10.000.000	$1.277.516	1,3	$300.000	$520.000	$1.120.000
Profesionales HD	691	$983.541	$0	$10.000.000	$1.218.158	1,2	$300.000	$550.000	$1.200.000
Jubilados	5.303	$271.158	$0	$4.200.000	$306.734	1,1	$150.000	$190.000	$310.020
Jubilados HD	6.337	$277.004	$0	$4.200.000	$299.228	1,1	$160.000	$193.917	$320.000

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

3. Imputación de la estructura del gasto en Servicios Financieros y Seguros

Según se explicó en el apartado dedicado a la descripción de los instrumentos (4.2.4), el módulo de Gastos en Servicios Financieros y Seguros (GSF), se compone de 8 secciones. Seis de las ocho se encuentran relacionadas, al menos indirectamente, al financiamiento con crédito y se subdividen en dos componentes:

- **Caracterización del crédito:**
  Monto solicitado, plazo, cuota e institución que otorga el instrumento.

- **Zona de captura de los gastos en servicios financieros:**
  Intereses, gastos de administración, comisiones y seguros asociados a los instrumentos financieros.

Para responder el primer conjunto de preguntas (zona de caracterización de cada sección), es opcional que las personas informantes entreguen información complementaria. Mientras que, para responder el segundo conjunto de preguntas (zona de gasto de las secciones), sí se solicita a los informantes contar con comprobantes.

Cuando no se accedió a la información complementaria para capturar los gastos en servicios financieros (zona de gasto), y se contó con la caracterización del crédito, que permitió la estimación del gasto en servicios financieros según estos componentes, tal como monto, cuota y plazo, siempre y cuando se haya cumplido el criterio de consistencia, en que el valor a pagar sea mayor que el monto solicitado.

La evaluación del riesgo de crédito, disposición a contratar un servicio de seguros y posibilidad de que exista un componente de gasto asociado al instrumento, se evalúan con un conjunto de características tanto del instrumento (tipo de institución, períodos pagados), individuales de la persona titular del instrumento financiero (educación, laborares, previsión, salud, etc.), la persona informante del módulo, del hogar (residencia, número de personas, tramos etarios, entre otras), de la vivienda (materiales, vecindario, etc.) y variables de control.

Según las características mencionadas para predecir el gasto, se aplicó el método penalizado *Lasso* (Tibshirani, 1996). Esta técnica permite seleccionar variables y modelos que evitan el problema de sobreajuste y cuentan con la bondad de controlar la multicolinealidad entre variables. Además, por las características del módulo y la sensibilidad de los datos capturados, se presentó el problema de alta dimensionalidad, $(k>n)$ debido a que el número de posibles variables bordeó las trecientas y en ciertos casos superó el número de observaciones, a pesar de esto, *Lasso* permite evaluar empíricamente la inclusión de características en alta dimensionalidad. También, el método se

extiende para evaluar las probabilidades de tenencia de gastos mediante un *logit Lasso.* Finalmente, los datos de donantes se dividieron entre sets de entrenamiento y test para validar la predicción dentro del conjunto total, simultáneamente, las múltiples estimaciones se contrastaron con validación cruzada y se buscó una penalización que permitiera un equilibrio entre parsimonia y predicción. Con estas estimaciones, fue posible la desagregación de los componentes de los gastos en servicios financieros para las estimaciones:

a. Gastos e interés
b. Comisiones
c. Gastos de administración
d. Seguros asociados

La estimación de los gastos en servicios financieros en la IX EPF, se realizó solo para créditos hipotecarios, créditos de consumo y automotrices y cuentas vista, RUT/visa. En los dos primeros casos se realizó por la importancia del gasto que representa para un hogar y en cuenta vista, por motivos de no respuesta al gasto y el uso frecuente del instrumento.

Para llevar a cabo la imputación del gasto en los créditos, se utiliza la estructura presente en los registros donantes, que fueron aquellos créditos hipotecarios y créditos de consumo y automotrices que contaron con datos para donar a los registros sin gastos. Estos donantes requerían que dispusieran de toda la información respecto al monto, cuota, plazo y zona de gasto contestada, mientras que, en cuenta vista solo es necesario contar con el valor de la comisión.

En la IX EPF la caracterización de los instrumentos financieros fue de relevancia, en el 65,8% de créditos hipotecarios y en el 70% de créditos de consumo y automotriz, se contó con una caracterización consistente, pero no se capturó información desagregada de los gastos. En estos casos, se predijo la proporción de cada gasto financiero respecto a la cuota. Se priorizó la predicción de esta proporción en vez de la estimación directa del gasto, dado que los gastos dependen del monto, cuota, plazo, riesgo y tiempo pagado, por ende, es posible que los gastos difieran en la media, pero deberían estar dentro de rangos similares. La

distribución de los gastos se puede observar en la **Tabla 44.** Si bien, el porcentaje de créditos con información faltante respecto a la estructura es relativamente alto, se cuenta con información consistente para estimar el gasto e imputar la desagregación.

Tabla 44: Estadística de la no respuesta parcial del gasto financiero desagregado por instrumento

Tipo de Crédito	Instrumentos con comprobante o gasto		Datos faltantes		N Teórico
	n	%	n	%	
Crédito hipotecario	281	9,6%	2.632	90,4%	2.913
Crédito de consumo y automotriz	247	6,6%	3.469	93,4%	3.716
Cuenta vista, RUT/visa	13.209	79,9%	3.330	20,1%	16.539

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Uno de los principales elementos a destacar, es el cambio de la estructura posterior a la imputación **(Tabla 45)**. Se modificó el peso relativo de los gastos en administración e intereses para créditos hipotecarios, esto ocurrió debido a la antigüedad de los créditos imputados, ya que, si el crédito fue solicitado previo a un año, se imputaron gastos en administración a la desagregación del gasto financiero total del crédito. Por otra parte, es posible el cambio de la proporción del gasto en intereses dada la antigüedad y la amortización calculada según las características de los créditos hipotecarios. De este modo, la imputación de la estructura del gasto financiero se distribuyó entre interés, gastos en administración, comisiones y seguros asociados, que dependen de las características de los instrumentos, de titulares y del hogar. A continuación, se presenta la estructura del gasto financiero total, con y sin imputación, de acuerdo con su desagregación.

Tabla 45: Resultados de la imputación de gastos en servicios financieros

Desglose del gasto	Gasto financiero del crédito	
	Observado	Posterior a imputación
Crédito hipotecario	100%	100%
Interés	83,3%	79,8%
Administración	1,6%	2,1%
Comisiones	0,9%	1,2%
Seguros asociados	14,3%	17,0%
Crédito de consumo y automotriz	100%	100%
Interés	91,4%	91,3%
Administración	1,9%	0,9%
Comisiones	0,3%	0,4%
Seguros asociados	6,5%	7,4%

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

4. Imputación de la estructura del gasto en Gastos del Hogar y Gastos del Recuerdo

El modelo de imputación por el cual se rigieron los módulos Gastos del Hogar y Gastos del Recuerdo utilizó la media o mediana del gasto. Estos se calcularon empleando registros cercanos en distintos niveles de desagregación. Las variables y condiciones específicas por módulo fueron las siguientes:

a. Cuestionario de Gastos del Recuerdo:

Variables utilizadas:

- El establecimiento en el cual se realizó el gasto
- La clasificación socioeconómica de la unidad primaria de muestreo en la cual se encuentra la vivienda encuestada
- La región en la que se encuentra la vivienda
- El producto en qué se realizó el gasto

Con la información de estas variables clave se generó una estimación con el objetivo de obtener valores consistentes con las las características del gasto que se buscaba imputar. Lo anterior hace referencia al hecho de que en algunos casos no fue posible imputar los valores en el nivel de desagregación más detallado, dado que no existieron suficientes registros para realizar la estimación. En el caso de que no existieran suficientes registros se realizó la estimación en un nivel menos desagregado dejando fuera una variable clave para la estimación. Además, para determinar el uso de la media o la mediana para la estimación de la imputación se evaluó el comportamiento estadístico en terminos de distribución.

b. Módulo Gastos periódicos del Hogar:

Variables utilizadas:

- El producto en que se realizó el gasto
- La clasificación socioeconómica de la unidad primaria de muestreo en la cual se encuentra la vivienda encuestada
- La región en la que se encuentra la vivienda
- El mes en que se realizó el gasto

En este módulo se sometieron al proceso de imputación gastos particularmente altos (determinados como inválidos). En estos casos, seleccionados por criterio normativo, se determinó imputar con la media del código CCIF, en el mismo mes y región en que se encontró el gasto.

Adicionalmente, para ciertos casos de arriendos imputados donde la información entre GH y RPH era inconsistente se asignó el promedio entre estas observaciones, si uno de los dos tuvo información faltante se le asignó el valor registrado en el otro módulo. Para el resto de los arriendos imputados y los de vivienda secundaria, se les imputó la media de los otros arriendos imputados, de acuerdo con los siguientes criterios sucesivos:

- UPM (Unidad Primaria de Muestreo) + CSE (clasificación socioeconómica de la UPM)
- Comuna + CSE (clasificación socioeconómica de la UPM)
- Comuna
- Región + CSE (clasificación socioeconómica de la UPM)

Es decir, se intentó imputar con la media de los registros con la misma UPM y CSE. Para los registros que no contaron con datos para cumplir el criterio, se pasó a la siguiente etapa, y así sucesivamente.

7.3.5 Derivar nuevas variables y unidades

Desde la coordinación se generó la tabla de factores de ajuste temporal, la cual fue enviada al equipo de analistas socioeconómicos, quienes hicieron los ajustes de gasto según factor de ajuste temporal. Luego de la realización de ajustes se generó la tabla preliminar, además de las tablas estructuradas por submuestra, dando cierre al subproceso con el envío de la notificación correspondiente a los coordinadores de sección, quienes realizaron la revisión final de los insumos.

7.3.5.1

Ajuste temporal (ajuste al mes de referencia)

Debido a los distintos períodos de referencia[139] utilizados en la recolección, los gastos se deben ajustar temporalmente para llevar todos los gastos a la misma unidad temporal. En el caso de la EPF los gastos se mensualizan, es decir, se calculan a la escala de un mes tipo. Para comprender la manera en la que se lleva a cabo el ajuste temporal, además de tener en consideración el período de colaboración de los hogares, es útil pensar en el período de recuerdo del cuestionario de Gastos del Recuerdo (GR). Ello da como resultado un total de 48 quincenas: 24 de trabajo de recolección de datos y 24 de recuerdo.

[139] Gastos diarios, 15 días; gastos del recuerdo, 3, 6 o 12 meses hacia atrás; gastos en servicios financieros y seguros, el mes anterior; gastos periódicos del hogar, mensuales o con otras frecuencias como trimestrales, semestrales, etc.

A modo de ejemplo, en el **Diagrama 13** las quincenas 25 y 26 corresponden al primer mes de trabajo de recolección de datos oficial (octubre de 2021), mientras que las submuestras 1 a la 24 dan cuenta del período de recuerdo del cuestionario GR el cual corresponde a 12 meses desde el día anterior al inicio de la submuestra operativa. Análogamente, las quincenas 27 y 28 representan al segundo mes de recolección (noviembre de 2021), mientras que las submuestras 3 a la 26 corresponden al período de recuerdo de un año.

Diagrama 13: Períodos de referencia Cuestionario Gastos del Recuerdo

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Para la primera submuestra operativa del trabajo de recolección de datos (octubre de 2021), se tienen los siguientes períodos de recuerdo:

- 3 meses: quincena 19 – 24
- 6 meses: quincena 13 – 24
- 12 meses: quincena 1 – 24

Para la segunda submuestra del trabajo de recolección de datos (noviembre de 2021) se tienen los siguientes períodos de recuerdo:

- 3 meses: quincena 19 – 25
- 6 meses: quincena 13 – 25
- 12 meses: quincena 1 – 25

I. Ajuste temporal del Cuestionario de Gastos Diarios (GD)

En el caso de GD, el factor de ajuste temporal es igual al total de días del mes de trabajo de recolección de datos, dividido por el número de días de la quincena, lo cual da como resultado un número cercano a 2.

$$Ajuste = \frac{n°días\ del\ mes}{n°días\ de\ la\ quincena}$$

El objetivo de esto es que la quincena de trabajo de recolección captura la mitad del gasto que un hogar realiza en un mes. Por ende, es necesario multiplicar dicho gasto por un factor cercano a 2.

II. Ajuste temporal del Módulo de Gastos periódicos del Hogar (GH)

El Módulo Gastos del Hogar contiene preguntas que implícitamente contemplan una periodicidad mensual y otras en las que es posible que los informantes registren una periodicidad específica. Ello genera dos ajustes temporales diferentes:

- **Gastos consultados de manera mensual:** no se requiere ajuste
- **Gastos con otras periodicidades:** valor del gasto, dividido por el número de meses que abarca la boleta.

$$Ajuste = \frac{valor\ del\ gasto}{n°meses\ que\ abarca\ la\ boleta}$$

III. Ajuste temporal del Módulo de Gastos en Educación (GE)

En el Módulo de Gastos en Educación el ajuste temporal se rige por temporalidades que pueden derivar en tres ajustes diferentes:

- **Gastos en educación mensual:** no requiere ajuste
- **Gastos anuales en educación:** suma del gasto realizado en el año, dividido por 12

$$Ajuste = \frac{valor\ del\ gasto\ n°\ de\ veces\ que\ realiza\ el\ gasto}{12}$$

- **Gastos semestrales en educación:**

$$Ajuste = \frac{valor\ del\ gasto\ n°\ de\ veces\ que\ realiza\ el\ gasto}{6}$$

IV. Ajuste temporal del cuestionario Gastos del Recuerdo (GR)

El ajuste temporal del cuestionario de Gastos del Recuerdo (GR) es igual al número de días del mes de trabajo de recolección, dividido por el número de días del período de recuerdo. El resultado se utiliza como un multiplicador de los gastos.

$$Ajuste = \frac{n°\ días\ del\ mes}{n°\ días\ de\ recuerdo}$$

Teniendo en consideración los períodos de recuerdo mostrados anteriormente (ver Diagrama 14), a modo de ejemplo, se presentan los ajustes temporales para la primera submuestra operativa:

3 meses: el factor es igual a 0,34 ya que el numerador sería 31 que corresponde al número de días que tiene octubre, dividido por la suma de los días comprendidos entre las quincenas 19 a la 24, que da como resultado 91.

6 meses: el factor es igual a 0,17, ya que el numerador se mantiene en 31, pero ahora es dividido por la suma de los días comprendidos entre las quincenas 13 a la 24, que lo que da como resultado 181.

12 meses: el factor es igual a 0,085, siendo el numerador 31, y dividido por la suma de los días comprendidos entre las quincenas 1 a la 24, que da como resultado 365.

De acuerdo con esta metodología de cálculo, se obtienen ajustes cercanos al equivalente de dividir los gastos por valores aproximados a **3**, **6** y **12**. Cabe señalar que, para la segunda quincena de cada mes, la suma de días es mayor que para la primera, en uno o dos días más aproximadamente. El efecto de ello es que el gasto se divide por un número mayor, lo cual compensa el hecho de que en la segunda quincena las personas registren más gastos, como resultado de tener un período de recuerdo más largo (las primeras dos semanas del mes).

En resumen, el factor de ajuste temporal depende tanto del módulo en el que se registró el gasto como del producto clasificado por la CCIF. La siguiente tabla ilustra los períodos a los cuales se ajusta el gasto declarado por los hogares para mensualizarlo.

Tabla 46: Resumen del Factor de Ajuste Temporal según instrumento

Módulo	Temporalidad (factor de ajuste temporal)					
HR	-					
RPH	-					
GD						Quincena
GH					Mes pasado	
GE	Año escolar					
GR	12 Meses	6 Meses	3 Meses			
GSF					Mes pasado	
IN (TE)					Mes pasado	

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

7.3.6 Calcular ponderadores

En tanto la coordinación verificó la acumulación de tablas estructuradas por módulo y submuestra, el equipo de analistas socioeconómicos generó la tabla de caracterización de encuestadores, para llegar así a la generación de la base insumo para el factor de expansión. Luego, se procedió a enviar al equipo de Marcos y Muestras, quienes fueron los responsables de elaborar y generar el factor de expansión.

Este subproceso se dio por finalizado una vez que el equipo de Marcos y Muestras generó el factor de expansión, dando paso a su revisión final, que en caso de presentar observaciones envió de regreso al equipo de coordinación para generar nuevamente el insumo para su elaboración. De no contar con comentarios u observaciones, el equipo de Marcos y Muestras hizo envío final del factor de expansión.

7.3.7 Calcular agregaciones

Este subproceso consta de tres etapas. En primer lugar, se realizan análisis de cantidad de hogares en el estudio, consistencia de los hogares y revisión de gastos e ingresos atípicos finales

En segundo lugar, se incorporó el factor de expansión a los datos, para dar paso al cálculo de agregaciones (gasto promedio por hogar y gasto promedio por división, entre otros). Este cálculo se llevó a cabo de forma paralela entre dos softwares, R y STATA, con el objetivo de corroborar la consistencia de este proceso.

Respecto a la tercera etapa y final, en ella se llevó a cabo la validación de los resultados obtenidos, si estas presentaron reparos y observaciones fueron debidamente revisadas por la coordinación.

7.3.8 Finalizar los archivos de datos

Luego de la consolidación definitiva de las bases finales, se realizó el junto con asegurar las prácticas de seguridad estandarizadas para el resguardo de los datos. Así mismo, se determinó la ruta de cada una de las entregas correspondientes a la base final de procesamiento, lo que incluyó ratificar que se hayan generado todos los insumos complementarios.

08.

METODOLOGÍA DE ANÁLISIS DE DATOS

08.

METODOLOGÍA DE ANÁLISIS DE DATOS

8.1 Análisis de resultados

La fase de análisis de resultados busca interpretar, explicar y validar mediante fuentes internas y externas los resultados oficiales obtenidos de los datos que fueron procesados previamente.

Dado que las fases del procesamiento ya descritas, pueden ser iterativas y paralelas (UNECE, 2016), las actividades de análisis de datos comienzan antes de que la última submuestra de recolección termine. Los subprocesos involucrados en el análisis de resultados se describen en los siguientes apartados.

8.1.1 Preparar borrador de resultados

El equipo de analistas socioeconómicos en conjunto con la coordinación técnica inició la fase de revisión de los datos, corroborando que estos correspondieran al período de referencia analizado. Así mismo, el equipo técnico revisó que cumpliesen con la estructura y cantidad de datos a nivel de variables y unidades reportadas, así como con la completitud de campos y variables. Cuando alguna de las revisiones no cumplió con los criterios requeridos, se volvió a los subprocesos previos ("Calcular agregaciones" o "Finalizar los archivos de datos") según correspondía, indicando la causa del incumplimiento de los requisitos para iterar el proceso y esclarecer los resultados.

8.1.1.1 Revisar base de datos

Se revisó la base de gastos y la base de personas, incluyendo el cálculo de medidas, promedios ponderados, medianas, máximos, mínimos y otros estadísticos, según los lineamientos establecidos en el subproceso "Diseñar el procesamiento y análisis".

La validación final de estas bases le correspondió a la coordinación de analistas socioeconómicos.

8.1.1.2
Elaborar productos estadísticos

Después de la obtención de las bases de gasto y personas, se inició el proceso de elaboración de productos estadísticos: análisis detallado por módulo, indicadores operativos y socioeconómicos, presentación de resultados y tablas de resultados. Estos productos fueron validados tanto con fuentes internas como externas, si esta etapa no fue exitosa se iteró hacia el procesamiento específicamente a la revisión y edición de los datos en cuestión.

8.1.2 Validar los resultados

Luego de la elaboración de los productos estadísticos, se procedió a la validación de estos siguiendo recomendaciones institucionales e internacionales. Adicionalmente, se consideraron productos estadísticos de temáticas relacionadas a la medición de gastos e ingresos, como también eventos y datos relacionados con la realidad económica y social que intervinieron en el período de la operación estadística en terreno.

8.1.2.1
Validar indicadores

En esta etapa se analizaron los indicadores operativos y socioeconómicos, junto con sus coeficientes de variación (precisión estadística), errores muestrales y no muestrales. Siguiendo estándares institucionales e internacionales, con el objetivo de hallar discrepancias y justificar las causas de estas.

La revisión de indicadores operativos se realizó para la totalidad de los dominios geográficos del estudio: a) Total Capitales Regionales, b) Gran Santiago y c) Macrozonas (norte, centro y sur). Se buscó la consistencia con la trayectoria histórica de los indicadores, y en caso contrario, se buscaron las causas de esta divergencia.

Para los indicadores socioeconómicos y tablas de resultados se analizaron los indicadores de precisión estadística con el fin de corroborar su

calidad y posterior publicación . Este proceso se realizó siguiendo las normativas derivadas del "Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares" (INE Chile, 2020), como también recomendaciones de organismos internacionales (Gutiérrez, 2020).

8.1.2.2
Comparar estadísticas

El proceso de comparaciones estadísticas se realizó considerandos fuentes internas (institucionales) como externas. En el caso de las fuentes internas se utilizaron operaciones estadísticas institucionales que abordaron temáticas relacionadas a los gastos e ingresos, lo que permitió comparaciones directas para diversas estimaciones e indicadores. A continuación, se listan los pasos seguidos para estas comparaciones:

- Identificación de fuentes internas utilizadas.

- Realización del contraste de resultados con información proveniente de versiones anteriores de la encuesta (VII EPF y VIII EPF), también de otras operaciones estadísticas que abordaron las mismas variables de interés en gastos e ingresos en períodos temporales similares. Una fuente interna de gran importancia para la validación y contraste de los indicadores de ingreso corresponde a la Encuesta Suplementaria de Ingresos (ESI), tanto en su versión 2021 como 2022[140].

- Buscar consistencia con los resultados con el fin de generar insumos para la validación. En el caso de las comparaciones de la IX EPF con su versión anterior -VIII EPF (2016-2017)- se contrastó la estructura del gasto e ingreso de ambos períodos.

Terminado el proceso de comparación interna, se procedió a la comparación de datos con fuentes externas. La identificación de fuentes externas utilizadas, las cuales corresponde a:

[140] Se contrataron los ingresos con ambas versiones de la ESI debido a que ambas presentan una comparabilidad tanto con el primer período de la EPF como con el último período. Considerando que la IX EPF comenzó su recolección en octubre 2021 y finalizó en septiembre 2022, fue de mucha importancia contrastar sus indicadores de ingresos según cercanía temporal con cada versión de la ESI.

- **Cuentas Nacionales:**

Dado que el marco conceptual de la EPF es el Sistema de Cuentas Nacionales, es una fuente de comparación importante para la encuesta. Para realizar la comparación fue necesario evaluar las diferencias de los resultados de la EPF respecto a la distribución de gasto por división entregada por Cuentas Nacionales y explorar sus posibles causas[141].

- **Encuesta Casen 2022:**

En el caso de esta fuente de comparación se puso especial atención a posibles diferencias en el ingreso promedio por hogar. Además, se compararon los ingresos según su fuente: trabajo asalariado, independiente, ingresos de la propiedad, etc. Al respecto, no se observaron grandes diferencias a nivel de estimaciones agregadas.

- **Encuestas de ingresos y gastos de otros países:**

Se realizó la comparación de la estructura de gastos considerando dos aspectos: a) tener un nivel de PIB *per cápita* similar a Chile y b) utilización de la CCIF como sistema de clasificación de gastos.

- **Otras fuentes externas:**

Utilizadas de acuerdo con temáticas específicas de la encuesta, tales como: precios minoristas informados por la Oficina de Estudios y Políticas Agrarias (ODEPA); información del Servicios de impuestos externos (SII), de la Superintendencia de Pensiones (SP), de la Comisión para el Mercado Financiero (CMF), entre otros[142].

Luego del contraste de resultados con información proveniente de fuentes externas se corroboró la consistencia de ciertas variables con los resultados de la comparación externa. En caso de discrepancia se continuó con la etapa de "Investigar las fuentes de discrepancia" como parte del flujo.

En casos que no arrojaron diferencias se continuó con la actividad "Disponer estadísticas", debido a que se logró realizar el contraste a niveles desagregados logrando su validación.

[141] No obstante, al momento del análisis no fue posible obtener una estructura del gasto en consumo final para el mismo período del estudio. Por lo que los análisis consideraron la diferencia de los períodos de referencia comparados.

[142] Para conocer mayor detalle de todas las comparaciones, consideraciones metodológicas y resultados, favor revisar "Informe de Principales Resultados IX EPF".

En el caso de identificación de discrepancias se procedió a:

- Búsqueda de diferencias metodológicas con las fuentes contrastadas (por ejemplo: cobertura, tipo de instrumento, período de captura y modalidad de recolección).

- Revisión del comportamiento histórico de las series, buscando algún tipo de comportamiento atribuible a diferencias estacionales.

- Generación de informe explicativo que explicase la justificación de la discrepancia y la forma en que se trató la diferencia.

8.1.3 Interpretar y explicar los resultados

Este subproceso tiene por objetivo generar un análisis de los resultados, para interpretar y explicar los resultados estadísticos obtenidos y determinar en qué medida reflejan las necesidades planteadas y los objetivos propuestos al comienzo de la producción estadística. Consiste en las siguientes actividades:

- Revisión en completitud de los productos estadísticos provenientes del subproceso "Validar los resultados".

- En caso de cumplir con los criterios de calidad establecidos en el proceso "Diseño y planificación", prosiguió la ejecución del análisis estadístico según lo establecido en los lineamientos metodológicos de análisis de cada operación estadística.

- En caso de no cumplir con los criterios de completitud, coherencia y consistencia, definidos por la operación estadística en el proceso de "Diseño y planificación", se volvió al subproceso "Validar los resultados" para subsanar las observaciones.

Luego de la obtención de resultados y productos estadísticos finales se hizo envío a la jefatura de proyecto, cuya labor consistió en visar los resultados, interpretarlos y comenzar el proceso de control de divulgación.

8.1.4 Aplicar control a la divulgación

Este subproceso tiene por objetivo garantizar que los datos y metadatos a difundir cumplan con el marco normativo de confidencialidad acorde a las políticas y normas establecidas por el INE. El subproceso de control a la divulgación de microdatos se compone de cinco actividades y se debe tener en cuenta que no responde a un proceso lineal, es decir, puede entenderse como un subproceso iterativo.

1. Realizar definiciones previas al control a la divulgación estadística (SDC[143], por su sigla en inglés). El propósito de esta etapa es establecer los requerimientos necesarios para iniciar el subproceso de control a la divulgación, que incluye la revisión de insumos, análisis de las necesidades de los usuarios y características estadísticas prioritarias, y la determinación de la necesidad de protección de confidencialidad. Esto último, está estrechamente relacionado con la interpretación de las leyes y normas sobre el secreto estadístico en Chile.

- Los procedimientos descritos para esta etapa aplican para todas las operaciones estadísticas y productos relacionados cuya recolección de información y/o publicación sea realizado por el INE (muestras, censos, procesos de múltiples fuentes y registros estadísticos) que darán a conocer información al público general u otros usuarios.

- En esta etapa se debe revisar que no existan restricciones legales que impidan la publicación de los microdatos.

- Por otra parte, si el conjunto de datos no posee variables sensibles o variables de identificación (directa o indirecta), se pasa a la siguiente etapa del flujo de Análisis de resultados: Generar reportes y liberar datos[144].

[143] En inglés, *Statistical Disclosure Control*.

[144] Las exclusiones generales responden a los requerimientos de información del Banco Central de Chile (BCCh). Además, se excluyen las operaciones estadísticas cuyos conjuntos de microdatos no son recolectados ni publicados por el INE.

2. Preparar y explorar datos originales: el objetivo principal de esta etapa es preparar los datos originales, explorar características y su estructura.

3. Medir y evaluar riesgos: cálculo de medidas de riesgo sobre los datos previos a publicar, basado en estos resultados y determinar si el archivo de microdatos es seguro para su publicación.

4. Seleccionar, aplicar y evaluar: selección de métodos apropiados para el control a la divulgación estadística y lograr el nivel de riesgo definido junto con la preservación de características definidas.

5. Generar reportes y/o liberar datos.

8.1.5 Finalizar productos

Realizados todos los procedimientos mencionados en las etapas anteriores, es posible aprobar la finalización del producto final de la IX Encuesta de Presupuestos Familiares con el fin de ponerlo a disposición de la ciudadanía.

09.

ALCANCE DE LA PRESENTACIÓN DE RESULTADOS

ALCANCE DE LA PRESENTACIÓN DE RESULTADOS

9.1 Productos estadísticos a publicar

La publicación considera los siguientes productos estadísticos:

a. Base de datos de personas:
Corresponde a la información sociodemográfica y de ingresos de las personas, cada fila corresponde a una persona perteneciente a un hogar. Cada persona, hogar y vivienda se encuentran desprovistas de datos de identificación, con el fin de resguardar garantías de confidencialidad y anonimato que son descritas posteriormente en este documento. Esta base de datos contiene variables agrupadas según las siguientes temáticas:

- Variables sociodemográficas
- Variables educativas
- Variables sobre la condición de actividad económica, ingresos y situación sobre la afiliación de las personas en los sistemas de salud y previsión
- Variables de gasto
- Variables sobre la vivienda principal y secundaria
- Variables que identifican si el hogar realiza gastos asociados a la vivienda principal y secundaria
- Factor de expansión y variables utilizadas para especificar el diseño muestral complejo

b. Base de gastos:
Contiene información de todos los módulos de la encuesta, donde cada fila corresponde al gasto mensualizado registrado por los hogares en el período de referencia, codificado según productos de la CCIF utilizada por la encuesta. Esta base de datos contiene variables tales como:
- **Macrozona:** área de estimación de la encuesta
- **Folio:** identificador único del hogar que permite vincular la base de gastos con la base de personas

- Factor de expansión y variables utilizadas para especificar el diseño muestral complejo
- **CCIF:** código del clasificador de gastos en bienes y servicios adaptado para la IX EPF
- Niveles de clasificación de la CCIF (división, grupo, clase, subclase y producto)
- Gasto mensualizado de los hogares según tipo de producto
- Glosa teórica de la CCIF adaptada para la IX EPF

c. Base de cantidades:

Corresponde a una base donde cada fila corresponde a un gasto con identificación única clasificado según la nomenclatura CCIF utilizada por la IX EPF. Las divisiones consideradas en la base de cantidades son la 01 de Alimentos y bebidas no alcohólicas, 02 Bebidas alcohólicas, tabaco y estupefacientes y 11 Servicios de Restaurantes y alojamiento.

d. Informe de Resultados de la IX EPF:

Documento que entrega las principales estimaciones obtenidas por la encuesta a nivel de gasto e ingresos y desagregaciones según macrozona, división de la CCIF, nivel educacional y categoría ocupacional. Adicionalmente, presenta un análisis comparativo con versiones anteriores de la encuesta y fuentes de información nacionales y extranjeras.

e. Síntesis de Resultados de la IX EPF:

Documento que entrega las principales estimaciones obtenidas por la encuesta y con mayor énfasis en la comunicación gráfica de sus resultados.

f. Tabulados oficiales de la IX EPF:

Corresponden a los cuadros estadísticos que presentan con mayor detalle las estimaciones poblacionales, gastos e ingresos obtenidas por la encuesta. El formato presentado son tablas en Excel que considera un número elevado de desagregaciones y estimaciones.

9.2 Difusión de los resultados

La difusión de los resultados puede distinguir dos tipos de usuarios:

- **Internos:**
  A nivel institucional, los resultados aportan con información relevante para la actualización de los bienes y servicios que conforman la canasta del IPC y sus respectivos ponderadores, además de entregar información relevante respecto a los lugares de adquisición de los distintos bienes y servicios de parte de la ciudadanía.

- **Externos:**
  Organismos públicos que requieren información de la encuesta para fines específicos y la población general, que puede disponer de los microdatos de la encuesta (base de personas y de gastos), Informe de Resultados, Tabulados y otros documentos de interés para los fines que requieran.

Complementariamente, la difusión de los resultados del estudio considera la publicación de distintos documentos con acceso libre a través del portal institucional:

- **Informe de Calidad IX EPF:**
  Entrega información del conjunto de acciones de calidad implementadas en la producción de la encuesta para cumplir con los requerimientos asociados al código de buenas prácticas para las estadísticas de la institución.

- **Manual de codificación IX EPF:**
  Entrega información del proceso de codificación realizado según los clasificadores de interés de la encuesta.

- **Manual de uso de base de datos IX EPF:**
  Orienta a las personas usuarias de la encuesta acerca del uso de las distintas bases de datos publicadas y la metodología de construcción de los principales indicadores de la IX EPF, de tal manera que puedan en principio replicar las estimaciones y luego profundizar en la obtención de resultados de acuerdo con sus propios intereses de investigación.

- **Cuestionarios IX EPF:**
  Incluyen tanto los módulos aplicados a través de la plataforma *Survey Solutions* elaborada por el Banco Mundial como aquellos aplicados en papel, para disposición de personas investigadoras y público en general.

- **Clasificación de Consumo Individual por Finalidades IX EPF:**
  Contiene los metadatos que enmarcan la utilización de la CCIF 2018.CL en la IX EPF.

9.3 Accesibilidad a los datos y marco legal

La publicación de microdatos y documentos oficiales de la encuesta tiene como requisito la anonimización previa de los datos recolectados, así como la indeterminación de estos. La IX EPF, como el resto de productos estadísticos elaborados en el INE, tiene como fundamento la Política de Protección de Datos Personales institucional en el aseguramiento de la confidencialidad de la información producida. En ella, se establece que los datos producidos por el organismo no pueden hacer referencia expresa a personas que participen en sus estudios. Por ese motivo, las bases definitivas excluyeron variables de identificación de las unidades de análisis que participaron en el estudio (personas, hogares, viviendas) o variables que aumentaran el riesgo de identificación de las mismas. Adicionalmente, las bases de datos se estructuran de acuerdo con las áreas de estimación determinadas en el diseño muestral.

Las normativas vigentes, que enmarcan el quehacer de la IX EPF e INE en general sobre esta materia son las siguientes:

- Compromiso de confidencialidad y reserva de datos protegidos por el secreto estadístico e información sensible y/o sujeta a la ley 19.628 sobre la protección de datos personales.

- Ley 19.628 de protección de vida privada y datos.

- Ley 20.285 regula el principio de transparencia de la función pública y el derecho de acceso a la información de los órganos de la Administración del Estado.

- Artículo 29° de la Ley 17.374- El Instituto Nacional de Estadísticas, los organismos fiscales, semifiscales y Empresas del Estado, y cada uno de sus respectivos funcionarios, no podrán divulgar los hechos que se refieren a personas o entidades determinadas de que hayan tomado conocimiento en el desempeño de sus actividades.

- El estricto mantenimiento de estas reservas constituye el "Secreto Estadístico". Su infracción por cualquier persona sujeta a esta obligación hará incurrir en el delito previsto y penado por el artículo 247, del Código Penal, debiendo en todo caso aplicarse pena corporal.

- Artículo 30° de la Ley 17.374- Los datos estadísticos no podrán ser publicados o difundidos con referencia expresa a las personas o entidades a quienes directa o indirectamente se refieran, si mediare prohibición del o los afectados" (Extractos de la Ley Orgánica 17.374 de creación del Instituto Nacional de Estadísticas, publicada el 10-12-1970).

- Ley N° 17.336 sobre Propiedad Intelectual, en su inciso primero del Artículo 88: "El Estado, los Municipios, las Corporaciones oficiales, las Instituciones semifiscales o autónomas y las demás personas jurídicas estatales serán titulares del derecho de autor respecto de las obras producidas por sus funcionarios en el desempeño de sus cargos".

- La difusión de información y accesibilidad a los datos involucra la tensión entre dos principios. En primer lugar, uno referido a la importancia de la libre difusión de los datos. Al respecto Naciones Unidas menciona que las estadísticas oficiales constituyen un elemento indispensable en el sistema de información de una sociedad democrática, de manera tal que los organismos nacionales en la materia deben facilitar que la población ejerza su derecho a información mediante un acceso imparcial a las estadísticas producidas. En segundo lugar, otro referido a la necesidad de asegurar

esta información, para prevenir su mal uso y mantener la confianza de las poblaciones a las que sirven.

Es posible señalar, con relación a la seguridad de datos, que el INE cuenta con un Código Nacional de Buenas Prácticas Estadísticas que en su principio 4 sobre confidencialidad establece que el "INE y los demás miembros del Sistema Estadístico Nacional (SEN) deben garantizar la protección y confidencialidad de la información con la que se producen las estadísticas oficiales, así como evitar la identificación de las fuentes" (INE, 2015a, pág. 8).

Además de la exclusión de las variables de identificación directa, la labor del Instituto Nacional de Estadísticas y de la IX EPF en particular consideró dentro del proceso de difusión de información la evaluación de métodos de Control a la Divulgación Estadística (SDC según su sigla en inglés: *Statistical Disclosure Control*), también llamados de "anonimización", que buscan establecer un equilibrio entre los principios de libre acceso a la información y aseguramiento de la confidencialidad de la información, entendiendo que existe una tensión o trade-off entre la protección de la seguridad de los datos y la utilidad de su libre difusión.

El objetivo de un proceso SDC bien implementado es encontrar el punto óptimo en el que la utilidad para los usuarios finales se maximice a un nivel de riesgo aceptable.

Mayores detalles al respecto formarán parte de un documento específico sobre el proceso de anonimización de la IX EPF.

10.

CONCEPTOS CLAVE

10.

CONCEPTOS CLAVE

Adaptación de un clasificador: Se habla de adaptación cuando un clasificador internacional (o de referencia) se modifica a nivel de estructura, para que sus categorías queden acorde a la realidad nacional (o a alguna realidad específica).

Adopción de un clasificador: Se habla de adopción cuando se utiliza directamente el clasificador internacional (o de referencia), sin realizar modificaciones a nivel de estructura.

Anonimización: Concepto equivalente a control a la divulgación estadística (SDC por sus siglas en inglés). Es un proceso técnico que consiste en transformar los datos individuales de las unidades de observación, de tal modo que no sea posible identificar sujetos o características individuales de la fuente de información, preservando así las propiedades estadísticas en los resultados (INE Chile, 2022).

Bien o servicio de consumo: Corresponde a un bien o servicio utilizado por los hogares para la satisfacción directa de sus necesidades.

Clasificación del Consumo Individual por Finalidades (CCIF): La CCIF es una clasificación funcional del Sistema de Cuentas Nacionales (SCN) que tiene por objetivo analizar el propósito o finalidad de las transacciones o los gastos que realizan ciertos sectores institucionales (Naciones Unidas, 2018). La clasificación estandariza los bienes y servicios capturados en la IX EPF y asegura la comparabilidad nacional e internacional con otros estudios o productos estadísticos hasta nivel de subclase. La adaptación nacional (CCIF 2018.CL) tiene una estructura jerárquica y su sistema de cifrado es numérico, con un nivel de desagregación de cinco niveles, en que los niveles de división, grupo, clase y subclase se adoptaron según el clasificador internacional y a nivel de producto fue adaptado a la realidad nacional.

Clasificador: Es una estructura jerárquica que agrupa los sucesos de un fenómeno según determinados criterios, estableciendo categorías de clasificación únicas e irrepetibles. Además, poseen atributos definitorios que son similares entre ellos (internamente homogéneos), pero que, al mismo tiempo, permiten diferenciarlos de las demás categorías (Externamente heterogéneos).

Clasificador de establecimientos: Clasificador INE que permite realizar la identificación del tipo de establecimiento de acuerdo su utilidad y/o servicio entregado. Dentro de su estructura es posible identificar 29 tipos de establecimientos, donde 3 son códigos residuales.

Clasificador Internacional Uniforme de Ocupaciones (CIUO): La clasificación CIUO es una herramienta para organizar los empleos en una serie de grupos definidos claramente en función de las tareas realizadas, clasificación que pertenece a la familia internacional de las clasificaciones económicas y sociales. A nivel nacional, la adaptación CIUO 08.CL es el estándar oficial para el procesamiento y análisis de las ocupaciones, que permite organizarlas según el tipo de tareas realizadas en un puesto de trabajo y las competencias requeridas para ello, dependiendo del nivel y la especialización de estas competencias.

Codificación: Proceso que permite convertir información verbal o textual en códigos que representan categorías dentro de un esquema de clasificación estadística, para facilitar el procesamiento de datos (INE Chile, 2022). En la generación de estos códigos se utilizan criterios de clasificación para estandarizar la codificación de información y facilitar su posterior análisis. Además, es un proceso analítico que implica un criterio experto de analistas y codificadores especialistas en que cada código representa una categoría preestablecida, que cumple con ser exhaustiva y excluyente entre sí.

Codificación Automatizada: Corresponde a la asignación de códigos a través de procesos Informáticos estandarizados, en base a información presente en los cuestionarios.

Criterios de Clasificación: Ante la necesidad de establecer una correcta clasificación, apuntando a un proceso lo más acertado posible, aludiendo a que muchas veces las descripciones o respuestas dadas por los informantes pueden resultar un tanto difusas, debido al desconocimiento de conceptos esenciales para la posterior clasificación, se declara, los criterios de clasificación proveen una herramienta altamente importante, es por ello, que para facilitar el análisis posterior existe una toma de decisiones basada en el criterio experto de los analistas y codificadores especialistas.

Diseño Muestral: Corresponde a la elección del tipo de muestreo y tamaño muestral a utilizar. Ambos son elegidos considerando el cumplimiento de los objetivos de la encuesta. En el caso de la IX EPF se ha utilizado un diseño muestral probabilístico, estratificado y bietápico.

Encuesta de Presupuestos Familiares (EPF): Encuesta socioeconómica que busca identificar y caracterizar la estructura del gasto de consumo final de los hogares urbanos en las capitales regionales de Chile y sus zonas conurbadas, con un período de referencia de un año.

Entrevistas interrumpidas *(break off)*: Corresponde a una subcategoría del rechazo en la que se establece comunicación con los hogares y estos entregan parte de la información requerida, no obstante, los datos reportados no cumplen con los criterios de suficiencia para considerar la unidad como entrevistada.

Error de estimación: La diferencia entre un valor estimado y el verdadero valor de un parámetro o, a veces de un valor a predecir (OCDE, 2007), en relación con la confiabilidad de la estimación.

Error muestral: Proviene del hecho de entrevistar solo una parte de la población y no a la población completa. Por tanto, todas las encuestas tienen un error muestral.

Error no muestral: Este tipo de error agrupa a todas las fuentes que se originan debido a los problemas prácticos que implican realizar una encuesta. En esta categoría se incluyen errores de cobertura (ej.: información incompleta del proceso de enumeración de las viviendas), errores de medición (ej.: preguntas formuladas de manera ambigua) y errores de procesamiento (ej.: error al digitar o codificar un dato), entre otros. Todos estos errores afectan de alguna forma la precisión y la confiabilidad de las estimaciones que se realizan a partir de la encuesta.

Estimación: Se refiere a la inferencia sobre el valor numérico de valores de población desconocidos a partir de datos incompletos, como una muestra (OCDE, 2007).

Estimación Puntual: Es una de las dos bases principales de estimación en el análisis estadístico. La estimación de puntos intenta dar el mejor valor estimado individual de un parámetro, en comparación con la estimación de intervalo, que procede especificando un rango de valores.

Estimador: Corresponde a una regla, función o método para estimar un parámetro de una población (OCDE, 2007).

Factor de expansión: Se interpreta como la cantidad de unidades en la población que representa una unidad en la muestra. De acuerdo con el diseño muestral, el factor de expansión se calcula en función del total de UPM y viviendas seleccionadas en cada estrato, así como también del número de UPM totales en el estrato (según información reportada por el Marco Muestral de Viviendas) y del número de viviendas particulares enumeradas.

Gasto en consumo adquirido: Acto mediante el cual se adquiere el derecho de propiedad sobre un bien o el derecho a la prestación de un servicio, independiente de si el consumo de ese bien se efectúa durante un período de referencia de la encuesta.

Gasto en consumo final: Gasto en bienes y servicios realizados por los hogares. Se excluyen los gastos en activos fijos en forma de viviendas o de objetos valiosos, así como los gastos en consumo intermedio, así como las transferencias en especies adquiridas para los hogares por el Gobierno o las ISFLSH. Solo aquellos bienes y servicios adquiridos para actividades no comerciales forman parte de su gasto de consumo final.

Gasto en consumo intermedio: Corresponden a bienes y servicios que los hogares adquieren para el desarrollo de sus actividades productivas.

Grilla Técnica de mínimos de calidad: Procedimiento de validación para evaluar la cantidad y calidad de los datos reportados por cada hogar, asignando puntajes según la presencia de información prioritaria para el cumplimiento de los objetivos del estudio. La grilla de mínimos de calidad asigna a cada hogar un puntaje entre 1 y 100 puntos, en los que se evalúan los módulos específicos y preguntas de cada cuestionario de la encuesta.

Hogar: Corresponde a una o más personas, con o sin lazos de parentesco, que comparten domicilio y satisfacen sus necesidades de alimentación y otros bienes esenciales a través de un presupuesto común.

Imputación de datos: Método utilizado para completar los datos perdidos de una encuesta, mediante un procedimiento estadístico (estrategia para lidiar con la no respuesta al ítem).

Indicador: Variable o factor cuantitativo o cualitativo que proporciona un medio sencillo y fiable para medir logros, reflejar los cambios vinculados con una intervención o ayudar a evaluar los resultados de un organismo de desarrollo.

Indicadores de desempeño operativo: Estadísticos para el monitoreo del trabajo de recolección de datos. Permite la evaluación del avance y son construidos a partir de la información disponible en los módulos Registro de Contacto en el Hogar y Control de Libretas de Gastos individuales de la Hoja de Ruta.

Índice de Precios al Consumidor (IPC): Indicador económico que mide la variación de precios de una canasta de bienes y servicios representativa de los gastos de los hogares urbanos, cuya cobertura geográfica corresponde a todas las capitales regionales y sus zonas conurbadas dentro de la frontera del país.

Informante idóneo: Persona apropiada para proporcionar la información que requiere una operación estadística, pues tiene la capacidad, edad y conocimientos para responder sobre el o los temas de estudio. Esta persona puede responder por sí misma o por los integrantes del hogar en el caso de las encuestas de hogares (INE Chile, 2022).

Ingreso disponible del hogar (incluye arriendo imputado): Corresponde a la sumatoria de todas las fuentes de ingreso del hogar que fueron percibidos en el mes anterior a la aplicación de la encuesta. Se excluyen las transferencias emitidas (se restan a los ingresos totales del hogar) e incluye el arriendo imputado para aquellos hogares propietarios de su vivienda o que usufructúan de ella y el arriendo imputado estimado por la vivienda secundaria (para la vivienda secundaria) Esta variable se encuentra expresada en pesos chilenos.

Ingreso primario: Corresponde a la suma de los ingresos del trabajo, de la propiedad y de auto provisión de servicios producidos por el hogar.

Ingreso total del hogar (incluye arriendo imputado): Corresponde a la sumatoria de todas las fuentes de ingreso de todos los integrantes del hogar, percibidas en el mes anterior a la aplicación de la encuesta. Se calcula como la suma entre el ingreso primario y las transferencias corrientes recibidas (excluyendo las transferencias recibidas en especies).

Ingresos de auto provisión de servicios producidos por el hogar: Aquellos ingresos asociados a la producción doméstica de servicios para el autoconsumo del hogar. Incluye el valor de la prestación del servicio de alojamiento en las viviendas ocupadas por sus propietarios (arriendo imputado), el valor de los servicios domésticos no remunerados y el de los servicios aportados por el consumo de bienes durables. Los ingresos provenientes de los servicios domésticos no remunerados y del consumo de bienes durables fueron excluidos de esta encuesta y han sido tradicionalmente excluidos de la medición de Cuentas Nacionales (Camelo, 2001).

Ingresos de la propiedad: Corresponde al flujo de ingresos netos que se obtienen de la tenencia de activos financieros. Incluye el retorno de dichos activos (interés ganado de los depósitos), recibidos de ingresos por inversiones, anualidades y otros pagos regulares de seguros de vida y fondos privados de pensiones (excluidos de la seguridad social) además de las rentas netas recibidas por el arriendo de activos no financieros, tales como recursos naturales, tierras, edificios, equipos, patentes o derechos de autor.

Ingresos del trabajo: Ingresos generados por una actividad económica laboral. Estos incluyen ingresos del trabajo asalariado, ingresos del trabajo independiente y otros ingresos del trabajo.

Marco de muestreo: Información cartográfica y del número de viviendas del cual es posible seleccionar una muestra. En el caso de la IX EPF se utiliza el Marco Muestral de UPM del INE.

Medición de resultados: Sistema para evaluar los resultados de las intervenciones para el desarrollo en función de los objetivos declarados.

Metadato estadístico: Son datos que entregan información necesaria para el uso e interpretación adecuada de las estadísticas por parte de las personas usuarias. Los metadatos describen los datos producidos por medio de documentación de contenidos relacionados, por ejemplo, con la metodología; el trabajo de recolección de datos; el procesamiento; análisis y la calidad; entre otros, de una operación estadística particular (INE Chile, 2022).

Modelo Genérico del proceso Estadístico (*Generic Statistical Business Process Model* – GSBPM): Modelo genérico internacional asociado a procesos de producción estadística. El GSBPM describe y define el conjunto de procesos necesarios para la producción de estadísticas oficiales, y según la Comisión Económica de las Naciones Unidas para Europa el GSBPM "provee un marco estándar y una terminología armonizada para ayudar a las organizaciones estadísticas a modernizar sus procesos de producción estadística, así como a compartir métodos componentes. Además, el GSBPM puede ser usado para integrar estándares en datos y metadatos, como modelo para la documentación de procesos, para la armonización de infraestructuras de cómputo estadístico y para proveer un marco de referencia para mejoras y evaluaciones del proceso de calidad" (UNECE, 2019, pág. 4).

Muestra con reemplazo: Tipo de muestreo que se utiliza para solucionar el problema de rechazo de la encuesta y que considera la selección de viviendas de reemplazo para ser utilizadas solo cuando alguna unidad rechaza la encuesta.

No respuesta: Es una forma de ausencia de información presente en la mayoría de las encuestas, que puede derivar en la no obtención de la medida de una o más variables que se encuentren en estudio en una encuesta. Es posible distinguir entre la no respuesta total y la no respuesta parcial, según participación, producida cuando existe ausencia de datos para una o más de las variables consultadas en una unidad muestral.

No respuesta parcial: Respuesta que tiene lugar cuando las personas participan de las entrevistas, no obstante, existen ciertas preguntas o temas que no pueden o no están dispuestos a contestar.

No respuesta total: Unidades que rechazan participar en el estudio (ver definición de rechazo).

Operación estadística: Aplicación de un conjunto interrelacionado de procesos, que comprende la detección de necesidades, diseño, construcción, recolección de datos, procesamiento, análisis, difusión y evaluación, lo cual conduce a la obtención de resultados estadísticos sobre un tema de estudio (INE Chile, 2022).

Período de referencia: Corresponde al período de tiempo acotado, respecto del cual se solicita a los entrevistados que informen sobre actividades, hechos o experiencias de interés.

Población objetivo EPF: Hogares (una o más personas) que residen habitualmente en las viviendas particulares ocupadas de las capitales regionales del país (incluyendo sus zonas conurbadas).

Porcentaje: Un porcentaje es un tipo especial de proporción donde la razón se multiplica por una constante, 100, de modo que la relación se expresa por 100 (OCDE, 2007).

Precisión: Corresponde a la aproximación de cálculos o estimaciones a los valores exactos o verdaderos que las estadísticas tenían la intención de medir (OCDE, 2007). Es por ello que cuanta más precisión se exige a la estimación de un parámetro, menor deberá ser la dispersión de las estimaciones de los indicadores principales de la encuesta (Vivanco, 2005).

Proceso estadístico: Conjunto interrelacionado de procesos, actividades y tareas cuyo fin es la generación de resultados estadísticos. Comprende la detección de necesidades, el diseño, construcción, recolección de datos, procesamiento, análisis y evaluación (INE Chile, 2022).

Producción estadística: Conjunto de subprocesos de transformación de insumos obtenidos a partir de la recolección, almacenamiento, procesamiento y análisis necesarios para la obtención de productos estadísticos (INE Chile, 2022).

Producto estadístico: Resultados físicos o digitales de una operación estadística, en general corresponden a publicaciones de información mediante la presentación de datos y metadatos, que buscan satisfacer las necesidades de la comunidad estadística (INE Chile, 2022).

Quintil de Ingresos: Es un valor de posición de la variable de ingresos (del hogar o per cápita). Los cinco valores o quintiles de ingreso se determinan al ordenar todos los datos de acuerdo con sus valores observados en la variable de ingresos, para luego separa las unidades estadísticas en cinco grupos de igual tamaño. Cada grupo representa el 20% de los valores observados de la variable ingreso de la muestra. Así los quintiles son los valores umbrales que separan estos grupos, y por extensión pasaron a designar los grupos que separan. El primer quintil se refiere al 20% de las unidades con menores ingresos observados, el segundo quintil se refiere al 20% de las unidades siguientes en la escala de ingresos, y el quinto quintil, al 20% de las unidades con mayores ingresos. De esa manera, los quintiles conforman una transformación de la variable ingresos de una variable de razón a una ordinal en cinco variables. Es importante señalar que la quintilización de los ingresos no constituye en sí misma una variable de estratificación socioeconómica, ya que para lograr una estratificación socioeconómica se deberían incorporar al análisis otras dimensiones.

Rechazo: Corresponde a aquellos hogares seleccionados para participar en el estudio, pero que declinan entregar información o bien entregan datos que no cumplen con los criterios necesarios para considerar aquellas unidades como entrevistadas.

Recuperación de datos: Etapa dentro del procesamiento que tiene como objetivo mejorar la calidad de los datos provenientes de las fases de digitación y trabajo de recolección de datos, mediante trabajo de recolección, mediante la revisión física de libretas, lectura de observaciones y recuperación en terreno.

Sobremuestreo: Se refiere a la inclusión en la muestra total de una mayor cantidad de unidades estadísticas que la cantidad definida como muestra objetivo o teórica. Se utiliza para prevenir el impacto de la no respuesta en un estudio probabilístico y así lograr mayor consistencia entre muestra objetivo y la muestra efectiva.

Submuestra: Período definido para dividir la muestra objetivo o teórica del estudio.

Submuestra Operativa: Es un subconjunto de unidades de muestreo seleccionadas para ser encuestadas durante una quincena. Su referencia es en términos generales, ya que el rango de días que contempla cada submuestra operativa fluctúa entre 14, 15 y 16.

Supervisión directa: Corresponde al acompañamiento de un supervisor a un encuestador en la aplicación de los instrumentos, con el objeto de observar en detalle la interacción de este con el entrevistado.

Supervisión indirecta: Se refiere al contacto con los informantes sin la presencia del encuestador, y sin aviso previo a él, en el que se realiza una entrevista de los aspectos principales de la encuesta que hayan sido aplicados hasta ese momento.

Tasa: Se refiere a la ocurrencia de eventos en un intervalo de tiempo específico (OCDE,2007).

Tasa de respuesta: Indica la proporción del total de unidades elegibles o de elegibilidad desconocida (incluido el sobre muestreo) donde se obtuvo datos suficientes para realizar cálculos oficiales. Esta tasa se obtiene dividiendo el número de hogares con información suficiente por la suma de las unidades elegibles y de elegibilidad desconocida.

Transferencias corrientes emitidas: Son flujos monetarios emitidos por los hogares que incluyen impuesto a la renta, contribuciones a la seguridad social, pago por pensiones alimenticias, transferencias a hogares (dentro y fuera del país), y transferencias a instituciones sin fines de lucro que sirven a los hogares.

Transferencias corrientes recibidas: Son flujos monetarios recibidos por los hogares que no generan relaciones a reciprocidad respecto a quienes las emiten. Estas incluyen los ingresos asociados a las jubilaciones, pensiones, otros beneficios en efectivo asociados a la seguridad social o al empleo, beneficios de asistencia social en efectivo del Gobierno, transferencias

desde otros hogares (dentro y fuera del país), transferencias en efectivo recibidas desde instituciones sin fines de lucro que sirven a los hogares. Este tipo de transferencias excluye las transferencias sociales en especies que los Gobiernos y las instituciones sin fines de lucro emiten a los hogares.

Validación: Verificación automática utilizada para garantizar que los datos ingresados sean razonables y factibles.

Validación de datos: Subproceso de la fase del procesamiento de datos, que comprueba que la información contenida en la base de datos provenientes de la digitación cumpla con ciertos criterios que aseguren la calidad de los datos.

Imputación de datos: Método utilizado para completar los datos perdidos de una encuesta, mediante un procedimiento estadístico (estrategia para lidiar con la no respuesta al ítem).

Vivienda: Toda edificación construida, convertida o dispuesta para el alojamiento permanente o temporal de personas, así como cualquier clase de albergue, fijo o móvil, ocupado como lugar de residencia.

11.

BIBLIOGRAFÍA

11.

BIBLIOGRAFÍA

AAPOR. (2023). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (10th Edition ed.). Recuperado el 02 de marzo de 2023, de https://aapor.org/wpcontent/uploads/2023/05/Standards-Definitions-10th-edition.pdf

Andridge, R., & Little, R. (2009). The Use of Sample Weights in Hot Deck Imputation. *J Off Stat*, 25(1), 21-36.

Australian Bureau of Statistics. (2015-2016). *Encuesta de Gastos de los Hogares, Australia (2015-2016). Methodologies, Partial responde and imputation*. Obtenido de https://www.abs.gov.au/methodologies/household-expenditure-survey-australia-summary-resultsmethodology/2015-16#quality-declaration.

BCCh. (2017). *Cuentas Nacionales de Chile: Compilación de Referencia 2013*. Banco Central de Chile, Santiago. Obtenido de https://www.bcentral.cl/documents/33528/2561751/Cuentas_Nacionales_de_Chile_Compilaci%C3%B3n_de_Referencia_2013.pdf/e173f1e6-f3c3-71dc-56f0-40c4907869fc?t=1595966088593

BCCh. (2022). *Cuentas Nacionales de Chile: Compilación de Referencia 2018*. Banco Central de Chile, Santiago. Obtenido de https://www.bcentral.cl/documents/33528/3434019/CompilacionReferencia2018.pdf/fa19a03f-5ca2-e6da-f367-ee2fd5ee6fb5?t=1695905318723

BID, Banco Mundial, CEPAL. (2004). *Plataforma del Proyecto MECOVI. Obtenido de Banco Interamericano de Desarrollo (BID), el Banco Mundial y la Comisión Económica para América Latina y el Caribe (CEPAL).* Obtenido de https://repositorio.cepal.org/handle/11362/31826

Bradburn, N. (2010). *Recall Period in Consumer Expenditure Surveys Program. NORC/University of Chicago.* Obtenido de https://www.bls.gov/cex/methwrkshp_pap_bradburn.pdf

Budlender, D. (2003). *The Debate about Household Headship. Obtenido de Social Dynamics: A journal of African studies.* Obtenido de http://www.statssa.gov.za/publications/DiscussHouseholdHead/DiscussHouseholdHead.pdf

Camelo, H. (2001). *Ingresos y gastos de consumo de los hogares en el marco del SCN y encuestas de hogares*. Obtenido de Serie: Estudios Estadísticos y Proyectivos, División de Estadísticias y Proyecciones Económicas, CEPAL. Santiago: CEPAL.: https://www.cepal.org/sites/default/files/publication/files/4718/S01010054_es.pdf

CEPAL. (2009). *La medición de los ingresos en la encuesta Casen 2009.* Santiago: División de Estadística y Proyecciones Económicas de la CEPAL.

CEPAL. (2011). *Recomendaciones para la ronda de los censos de la década de 2010 en América Latina. Santiago: CEPAL. CELADE.* Obtenido de http://repositorio.cepal.org/bitstream/handle/11362/5511/1/S1100525_es.pdf

CEPAL. (2016). La brecha de género en jubilaciones y pensiones. Los casos de Argentina, Brasil, Chile y Uruguay. (V. Amarante, M. Colacce, & P. Manzi, Edits.) Obtenido de https://repositorio.cepal.org/bitstream/handle/11362/40650/2/S1600929_es.pdf

CEPAL/INE. (2021). *Encuestas de ingresos y gastos de los hogares: experiencias recientes en América Latina y el Caribe", Documentos de Proyectos* (LC/TS.2021/49), Santiago. Obtenido de http://repositorio.cepal.org/bitstream/handle/11362/46898/S2100114_es.pdf

Díaz de Rada, V. (2004). *Problemas de representatividad en las encuestas con muestreos probabilísticos. Universidad Pública de Navarra. Departamento de Sociología.* Obtenido de https://www.raco.cat/index.php/Papers/article/download/25793/25627

DIESA. (1984). *Improving concepts and methods for statistics and indicators on the situation of women.* Obtenido de Department of International Economic and Social Affairs: https://unstats.un.org/unsd/publication/SeriesF/SeriesF_33E.pdf

Eurostat. (2003). *Eurostat*. Obtenido de Household Budget Surveys in the EU, Methodology and recommendations: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KSBF-03-003-__-N-EN.pdf

Eurostat. (2013). *Compendium of HICP reference documents.* Methodologies and Working papers

Eurostat. (2020a). *European Statistical System handbook for quality and metadata reports. European Union, Luxemburgo.* Obtenido de https://ec.europa.eu/eurostat/documents/3859598/10501168/KS-GQ-19006-EN-N.pdf

Eurostat. (2020b). *Household Budget Survey 2015 Wave EU Quality Report (Version 1).* Obtenido de https://ec.europa.eu/eurostat/documents/54431/1966394/HBS_EU_QualityReport_2015.pdf

Great Britain and Northern Ireland, Office for National Statistics, ONS. (2017). *Living Costs and Food Survey. Living costs and food survey technical report April 2015 to March 2016, Imputing missing values (5051).* Obtenido de https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomean dwealth/methodologies/livingcostsandfoodsurvey#technical-report.

Groves, R. M. (2004). *Survey methodology.*

Guerrero et al. (2019). Estratificación socioeconómica del marco muestral de viviendas 2017.

Gutiérrez, A. M. (2020). Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional. serie *Estudios Estadísticos*(101 (LC/TS.2020/52)).

Hoffmeyer-Zlotnik, J. &. (2009). *Private Household Concepts and their Operationalisation in Cross-National Social Surveys. Manheim:* GESIS-Leibniz-Institute for the Social Sciences. https://www.statd.si/mz/mz6.1/hoff.pdf.

INE Chile. (2015). *Grupo de Trabajo: Estandarización de la Identificación de Hogar (Documento interno). Instituto Nacional de Estadísticas de Chile.* Obtenido de Departamento de Calidad Estadística

INE Chile. (2016). *Jefatura de hogar: usos del concepto, historia, críticas y expresión en los indicadores. Instituto Nacional de Estadísticas de Chile. Santiago: INE.* Obtenido de http://historico.ine.cl/genero/files/estadisticas/pdf/documentos/jefatura_hogar.pdf

INE Chile. (2018a). *CIUO 08.CL Clasificador Chileno de Ocupaciones.* Santiago. Obtenido de https://www.ine.gob.cl/docs/default-source/buenas-practicas/clasificaciones/ciuo/clasificador/ciuo-08-cl.pdf

INE Chile. (2018b). I*nforme metodológico Prueba Piloto 2017 de dispositivos móviles en la Encuesta de Presupuestos Familiares (EPF).* Santiago. Obtenido de https://www.ine.gob.cl/docs/default-source/encuesta-de-presupuestos-familiares/documentos-de-trabajo/viii-epf-(julio-2016---junio-2017)/informe-metodol%C3%B3gico-prueba-piloto-2017-de-dispositivos-m%C3%B3viles-en-la-epf.pdf?sfvrsn=9e0b3c19_6

INE Chile. (2019a). *Métodos de imputación VIII EPF: Gastos diarios e ingresos de la actividad laboral y jubilaciones.* Documento de trabajo, Santiago. Obtenido de https://www.ine.gob.cl/docs/default-source/documentos-de-trabajo/190320-documento-imputacion-viii-epf.pdf?sfvrsn=668a66c0_2.

INE Chile. (2019b). *Orientaciones para la aplicación del cuestionario Hoja de Ruta. Santiago.*

INE Chile. (2020). *Estándar para la evaluación de la calidad de estimaciones en encuestas de hogares.*

INE Chile. (2021). *Documento metodológico Encuesta Nacional de Empleo (ENE) 2020. Obtenido de: https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/metodologia/espanol/metodologia-ene-2020.pdf?sfvrsn=793380e9_14*

INE Chile. (2022). *Glosario de conceptos estadísticos.* Departamento de Metodologías e Innovación Estadística, Santiago. Obtenido de https://www.ine.gob.cl/docs/default-source/buenas-practicas/estandares/estandar/glosario-de-conceptos-estad%C3%ADsticos/glosario-de-conceptos-ine-v_1-0.pdf?sfvrsn=43ff6ab6_2

INE Chile. (2023a). *Manual de recolección de datos IX EPF Tomo I.* Departamento de Estadísticas Demográficas y Sociales, Santiago.

INE Chile. (2023b). *Manual de recolección de datos IX EPF Tomo II.* Departamento de Estadísticas Demográficas y Sociales, Santiago.

INE Chile. (2023c). *Documento metodológico Encuesta Suplementaria de Ingresos (ESI) 2022.* Departamento de Estadísticas del Trabajo, Santiago. Obtenido de https://www.ine.gob.cl/docs/default-source/encuesta-suplementaria-de-ingresos/metodologia/documento-metodol%C3%B3gico/documento-metodol%C3%B3gico-esi-2021.pdf?sfvrsn=8cb12ee9_4

INE Chile. (2023d). Clasificador chileno del consumo individual por finalidades. Obtenido de https://www.ine.gob.cl/docs/default-source/buenas-practicas/clasificaciones/ccif/clasificador/ccif_2018-cl.pdf

INE España. (2022). *Encuesta de Presupuestos Familiares.* Obtenido de Metodología, método de imputación (80-84): https://www.ine.es/metodologia/t25/t2530p458.pdf.

INE España. (2023). Encuesta de Presupuestos Familiares. Obtenido de Metodología: https://www.ine.es/metodologia/t25/t2530p45822.pdf

Instituto Brasileiro de Geografia e Estatística (IBGE). (2011). *Pesquisa de Orçamentos Familiares 2008-2009, Análise do consumo alimentar pessoal No Brasil, Río de Janeiro. Procedimentos de crítica e imputação (35-36).* Obtenido de https://biblioteca.ibge.gov.br/visualizacao/livros/liv50063.pdf.

Instituto de Estadística de la UNESCO. (2013). *Clasificación Internacional Normalizada de la Educación, CINE 2011. Montreal, Canadá.* Obtenido de https://uis.unesco.org/sites/default/files/documents/isced2011-sp.pdf

Instituto Nacional de Estadística y Censos (INDEC). (2019). *Encuesta Nacional de Gastos de los Hogares 2017-2018, Informe de Gastos, Imputación de Gastos e Ingresos (183-198).* Obtenido de https://www.indec.gob.ar/ftp/cuadros/sociedad/engho_2017_2018_informe_gastos.pdf.

Leeuw, E. D. (2008). *International Handbook of Survey Methodology.*

Lohr, S. L. (2021). *Sampling: Design and Analysis* (Vol. Chapman & Hall CRC Texts in Statistical Science). CRC Press.

Lohr, S. L. (s.f.). *Text in Statistical Science, Sampling: Design and Analysis (337).* Obtenido de https://books.google.cl/

MDSF. (2023). *Resumen de resultados Casen 2022: Pobreza por Ingresos, Pobreza Multidimensional y Distribución de los Ingresos.* Observatorio Social. Obtenido de https://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/casen/2022/Presentacion%20de%20resultados%20Casen%202022.pdf

Mincer, J. (1974). Schooling, experience, and earnings. *National Bureau of Economic Research.* Obtenido de http://papers.nber.org/books/minc74-1

Observatorio Social. (2020). *Casen en Pandemia 2020. Medición de Ingresos y de Pobreza por Ingresos en pandemia. Nota técnica n° 2 – octubre 2020. Santiago.* Obtenido de http://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/casen/2020/Nota_tecnica2_Medicion_de_ingresos_y_de_pobreza_por_ingresos_en_pandemia.pdf

OCDE. (2007). *Glossary of Statistical Terms.* New York.

OCDE. (2013). OECD *Framework for Statistics on the Distribution of Household Income, Consumption and Wealth. París: Organización para la Cooperación y el Desarrollo Económico.* Obtenido de http://www.oecd-ilibrary.org/economics/framework-for-statisticson-the-distribution-of-householdincome-consumption-and-wealth_9789264194830-en

OIT. (2003a). *Informe de la Decimoséptima Conferencia Internacional de Estadísticos del Trabajo (CIET).* Obtenido de https://www.ilo.org/wcmsp5/groups/public/---dgreports/--stat/documents/meetingdocument/wcms_087570.pdf

OIT. (2003b). *Informe I: Estadísticas de ingreso y gasto de los hogares. Decimoséptima Conferencia Internacional de Estadísticas del Trabajo.* Obtenido de https://www.ilo.org/public/spanish/bureau/stat/download/mels-r-1.pdf

OIT. (2003c). *Informe II: Estadísticas de ingreso y gasto de los hogares. Decimoséptima Conferencia Internacional de Estadísticas del Trabajo. Ginebra: OIT.* Obtenido de https://www.ilo.org/public/spanish/bureau/stat/download/17thicls/r2hies.pdf

OIT. (2007). Resolución sobre la actualización de la Clasificación Internacional Uniforme de Ocupaciones. Ginebra.

OIT. (2012). *World of Work Report 2012* . Obtenido de https://www.ilo.org/global/publications/books/worldof-work/lang--en/index.htm

ONU. (2008). *Principles and recommendations for population and housing censuses - Revision 2. New York:* ONU. Department of Economic and Social Affairs. Obtenido de https://unstats.un.org/unsd/publication/seriesm/seriesm_67rev2e.pdf

ONU. (2009). *Diseño de muestras para encuestas de hogares. Directrices Prácticas* (Vol. Serie FN°98). Obtenido de https://unstats.un.org/unsd/publication/seriesf/Seriesf_98s.pdf

ONU. (2016). *Principios Fundamentales de las Estadísticas Oficiales*. Obtenido de https://unstats.un.org/unsd/dnss/hb/S-fundamental%20principles_A4-WEB.pdf

ONU et al. (2016). *Sistema de Cuentas Nacionales 2008. New York*. Obtenido de https://unstats.un.org/unsd/nationalaccount/docs/SNA2008Spanish.pdf

ONU. (2018). *Classification of Individual Consumption According to Purpose (COICOP).* División de Estadísticas. Obtenido de https://unstats.un.org/unsd/classifications/unsdclassifications/COICOP_2018_-_pre-edited_white_cover_version_-_2018-12-26.pdf

Rada, V. D. (2015). Manual de trabajo de campo en la encuesta: presencial y telefónica. Madrid: Centro de Investigaciones Sociológicas, 2da Edición. Obtenido de https://dialnet.unirioja.es/servlet/libro?codigo=612448

Statistics South Africa. (2011). *Encuesta de Ingresos y Gastos (2010-2011). Income and expenditure of households 2010/2011, Data editing and imputation (6).* Obtenido de https://microdata.worldbank.org/index.php/catalog/1545/related-materials.

Tejedor, F. (1981). *Validez interna y externa en los diseños experimentales.* Revista española de pedagogía.

Tibshirani, R. (1996). *Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1), 267–288.* Obtenido de http://www.jstor.org/stable/2346178

U.S. BLS. (2004). *Temporal Patterns of Survey Response Rates and Reporting Rates in the U.S. Consumer Expenditure Interview and Other. U.S. Bureau of Labor Statistics; Cho, Moon; Eltinge, John; Steinberg, Barry.* Obtenido de https://www.bls.gov/cex/cesrvymethscho1.pdf

U.S. BLS. (2006). *Analysis of Recall Effect on the Reporting of Expenditures For the Consumer Expenditure Interview Survey. U.S. Bureau of Labor Statistics; Steinberg, Barry; Chopova, Boriana; Tan, Lucilla; Ogden, Jared; Bahizi, Pierre, Washington, DC, United States*. Obtenido de https://www.bls.gov/osmr/pdf/st060150.pdf

UNECE. (2019). *Modelo Genérico del Proceso Estadístico (GSBPM).* Comisión Económica de las Naciones Unidas para Europa. Obtenido de https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples.* New York: Springer.

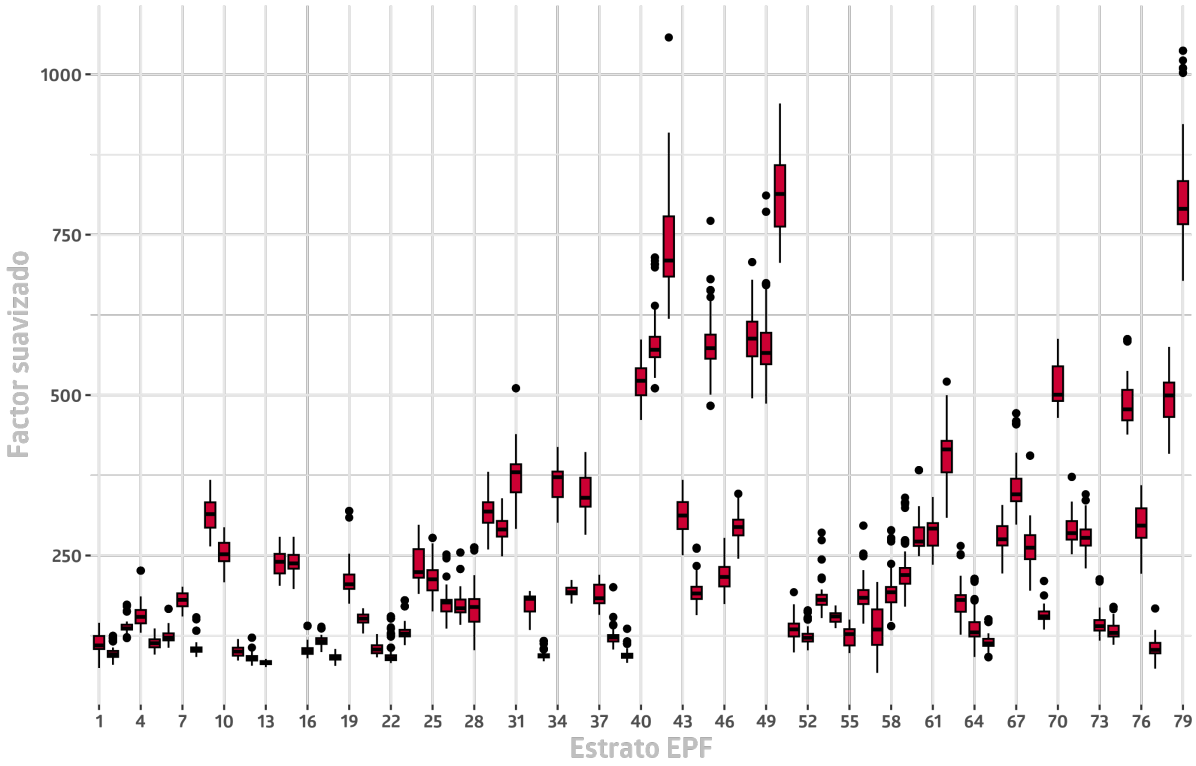
Weiss, A. (1995). *Human Capital vs. Signalling Explanations of Wages. Journal of Economic Perspectives, 9*(4), 133–154. Obtenido de https:// pubs.aeaweb.org/doi/pdfplus/10.1257/jep.9.4.133

12.

ANEXOS

ANEXO

Anexo 1 Ponderador ajustado por no respuesta al interior de cada estrato de muestreo con suavizamiento



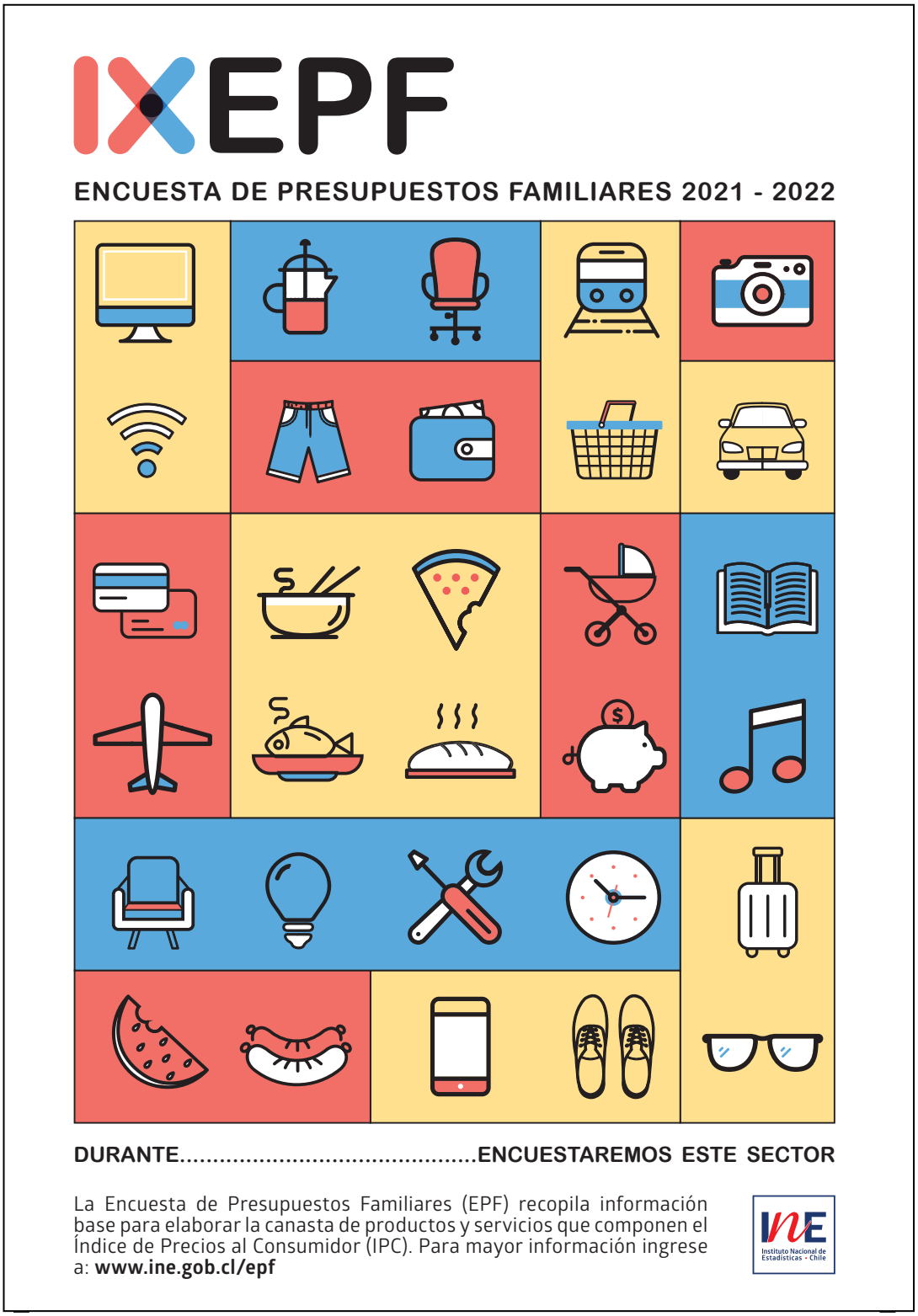
Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Anexo 2 Definición de variables

Variable	Definición
region	Región
cse	Clasificación socioeconómica
d_acceso	Tipo de acceso
d_sexo	Sexo
d_exp_otras_org	¿Cuenta con experiencia en encuestas en otra(s) organizacion(es), ya sean sean públicas o privadas?
d_exp_modos_rec_tel_int	Experiencia en recoleccion de datos utilizando telefono inteligente
d_exp_otras_enc_num	¿En cuántas encuestas ha trabajado?
d_estado	Estado de la vivienda
d_exp_modos_rec_tablet	Experiencia en recoleccion de datos utilizando tablet
d_rayados	Percepción de cantidad de rayados en los muros del entorno de la vivienda
d_basura	Percepción de cantidad de basura en el entorno de la vivienda
d_exp_anterior_ine	¿Trabajó anteriormente en encuestas en el Instituto Nacional de Estadísticas?
d_exp_anterior_ine_num	¿En cuántas encuestas ha trabajado en el INE?
d_negocio	La vivienda ¿posee algún negocio adosado?
d_exp_anterior_epf	¿Trabajó anteriormente en la Encuesta de Presupuestos Familiares o alguno de sus pilotos?
d_exp_modos_rec_papi	Experiencia en recoleccion de datos utilizando cuestionarios de papel
d_experiencia	¿Cuántos años de experiencia tiene como recolector(a) de datos?
d_precariedad	Precariedad de la vivienda
d_exp_modos_rec_sin_exp	Sin experiencia en recoleccion de datos
d_conf_dmc	En una escala de 1 a 7, donde 7 es muy confiado y 1 es nada confiado, ¿qué tan confiado se siente acerca de su capacidad para manejar la tablet durante la encuesta EPF?
es_enc	Persona trabaja como encuestador/a en la IX EPF

Fuente: Instituto Nacional de Estadísticas (INE), IX Encuesta de Presupuestos Familiares (EPF).

Anexo 3 Afiche IX EPF



Anexo 4 Carta de apertura IX EPF

INE
Instituto Nacional de Estadísticas - Chile

IXEPF
Encuesta de Presupuestos Familiares 2021-2022

ine.gob.cl

Su respuesta representa a Chile

Santiago, agosto de 2021

CARTA DE APERTURA
IX Encuesta de Presupuestos Familiares (EPF)

Señor (a)
Jefe (a) de Hogar
Presente

Estimado(a) señor(a):

Tengo el agrado de comunicar a usted que el Instituto Nacional de Estadísticas de Chile (INE) está desarrollando la **IX Encuesta de Presupuestos Familiares (EPF)**, en el marco de la generación de estadísticas oficiales del país.

Esta encuesta busca **identificar la estructura y características del gasto de consumo final de los hogares** urbanos de nuestro país, en las capitales regionales y sus zonas conurbadas. Es decir, queremos conocer cómo gastan sus ingresos los hogares, y saber, por ejemplo, cuánto dinero se destina para alimentación, salud, transporte o entretención.

Esta información es fundamental para disponer de estadísticas actualizadas, que reflejen con precisión la realidad socioeconómica de los hogares del país. **Del mismo modo, es requerida para actualizar la canasta de bienes y servicios del Índice de Precios al Consumidor (IPC) y para actualizar y calcular la línea de pobreza del país**.

En este contexto, pongo en su conocimiento que su vivienda ha sido seleccionada en forma aleatoria para participar en esta encuesta, representando a su vez a otros hogares de su comuna. Todas las respuestas que su hogar suministre serán utilizadas en forma estrictamente confidencial y están protegidas por el Secreto Estadístico, según lo dispuesto en la **Ley Nº 17.374**. Aquella ley faculta al INE para solicitar información a cualquier persona natural o jurídica que habita en el territorio nacional, a la vez que otorga penas de multa y cárcel a quién la divulgue a terceros, ajenos a la institución.

Finalmente, cabe señalar que su respuesta es indispensable para la construcción de estadísticas que reflejen la realidad nacional. Por tanto, agradecemos su valiosa cooperación en esta encuesta y le invitamos a contactarse con nosotros, llamando a nuestros teléfonos de **atención ciudadana 232461010 / 232461018, si surge alguna duda**.

Nuestro sincero agradecimiento,

SANDRA QUIJADA JAVER
Directora Nacional
Instituto Nacional de Estadísticas

INE
Instituto Nacional de Estadísticas - Chile

IX EPF
Encuesta de Presupuestos Familiares 2021-2022

INFORMACIÓN DE IMPORTANCIA

La persona que le entregó esta carta es parte del personal del INE, quien debe identificarse exhibiendo su credencial institucional y estar equipado con elementos de protección personal (mascarilla y protector facial), como parte de las medidas de prevención de contagios por COVID-19. Esta persona será responsable de responder a todas las preguntas que usted o algún integrante de su hogar tenga. Además, coordinará las fechas y horarios en que nuestro equipo visitará su hogar para el desarrollo de la encuesta.

La participación de su hogar en la encuesta durará una quincena, período en que se pedirá a las personas de 15 años o más que integran su hogar que registren todos sus gastos diarios en una libreta. Junto con ello, se solicitará al hogar que responda preguntas sobre gastos de carácter mensual como agua, luz y arriendo, entre otros, así como también sobre otros gastos realizados durante el último año. Por último, se consultará sobre los ingresos de las personas que integran su hogar, información que complementará los datos sobre los gastos realizados.

Para efectuar este trabajo, **un encuestador o encuestadora del INE visitará cuatro veces su hogar durante la quincena**, con la finalidad de apoyar los registros de gastos diarios de su hogar, resolver dudas y completar distintas secciones del cuestionario. Cada una de estas visitas presenciales demorará, en promedio, 30 minutos. En algunos casos, se procederá a recuperar datos puntuales posterior al término de la quincena, a través de llamados telefónicos que complementarán las visitas presenciales.

Es posible que su hogar reciba también la visita o llamado telefónico de una persona supervisora, quien realizará algunas preguntas relacionadas al trabajo efectuado por quien encuestó su hogar, para controlar la calidad del estudio.

Si desea más información sobre esta encuesta, puede escribir un correo a **epf@ine.gob.cl**. Además, puede conocer otras características de esta encuesta en **www.ine.cl** y específicamente en: **https://www.ine.gob.cl/epf**

También puede verificar la identidad del personal del INE, ingresando su rut en https://verificador.ine.cl o leyendo, con un dispositivo móvil, el código QR presente en su credencial. Puede acceder a esta información, así como a otros antecedentes del estudio, mediante los siguientes códigos QR:

Encuesta de Presupuestos Familiares

Verificador de Encuestadores

Muchas gracias por su colaboración.

ine.gob.cl

Anexo 5 Díptico IX EPF

EL ENCUESTADOR DE HOGARES

Para cumplir con el objetivo de la IX EPF se cuenta con un equipo altamente capacitado de encuestadores de hogares. Estas personas son las encargadas de recopilar información coherente y verídica de los distintos tipos de gastos e ingresos en que incurren los hogares seleccionados.

¿CÓMO IDENTIFICO A LOS ENCUESTADORES?

Los encuestadores de hogares contarán con la indumentaria necesaria para ser reconocidos por toda la población: camisa o blusa con el logo del INE, chaqueta o parka con el logo del INE y la credencial que los identifica como personal de la institución.

Para su seguridad y la de todos los encuestados, podrá consultar el RUT del encuestador en el sitio www.ine.gob.cl, en la sección **"Verificación Encuestador"** y comprobar si es o no personal del INE. Además, se podrá acceder al vericador desde un Smartphone con el código QR presente en la credencial de cada encuestador. Para cualquier duda o consulta se puede contactar al correo epf@ine.gob.cl

Su vivienda fue visitada por un encuestador/a del INE el	A las
/ /	: hrs.

Nombre del encuestador/a:
R.U.T. del encuestador/a: -
Nombre de jefe directo:
Teléfono de jefe directo:

¡Su vivienda ha sido elegida!

IX EPF

ENCUESTA DE PRESUPUESTOS FAMILIARES 2021 - 2022

Su vivienda ha sido seleccionada para responder la IX Encuesta de Presupuestos Familiares (EPF) que el INE realiza. Invitamos a usted y a toda su familia a participar de este proceso fundamental para conocer la estructura de los gastos e ingresos de los hogares en Chile.

¿QUÉ ES EL INE?

El Instituto Nacional de Estadísticas (INE) es uno de los organismos públicos con más trayectoria en Chile. Se fundó en 1843 y actualmente tiene como misión producir y difundir las estadísticas oficiales de Chile, proporcionando información confiable y accesible a los usuarios, para la toma de decisiones y un mayor conocimiento de la realidad del país.

¿QUÉ ES LA ENCUESTA DE PRESUPUESTOS FAMILIARES?

La Encuesta de Presupuestos Familiares (EPF) es un estudio que permite conocer las modificaciones en la estructura del gasto de los hogares del país, como también sus nuevas pautas de consumo. Desde 2011 se realiza cada cinco años y su información es la base para elaborar la canasta de productos y servicios que mide el Índice de Precios al Consumidor (IPC). Aquel indicador mide la variación de los precios de diversos productos. La EPF es, además, utilizada como la principal fuente de información para la actualización de las líneas de indigencia y de pobreza del país.

¿CUÁNDO Y DÓNDE SE REALIZARÁ ESTA ENCUESTA?

La IX EPF se desarrollará durante el segundo semestre del 2021 y el 2022. Su cobertura contempla las áreas urbanas de cada una de las capitales regionales del país y sus zonas conurbadas. En esta oportunidad se encuestarán cerca de 28.000 viviendas a nivel nacional.

¿POR CUÁNTO TIEMPO ENCUESTARÁN MI HOGAR?

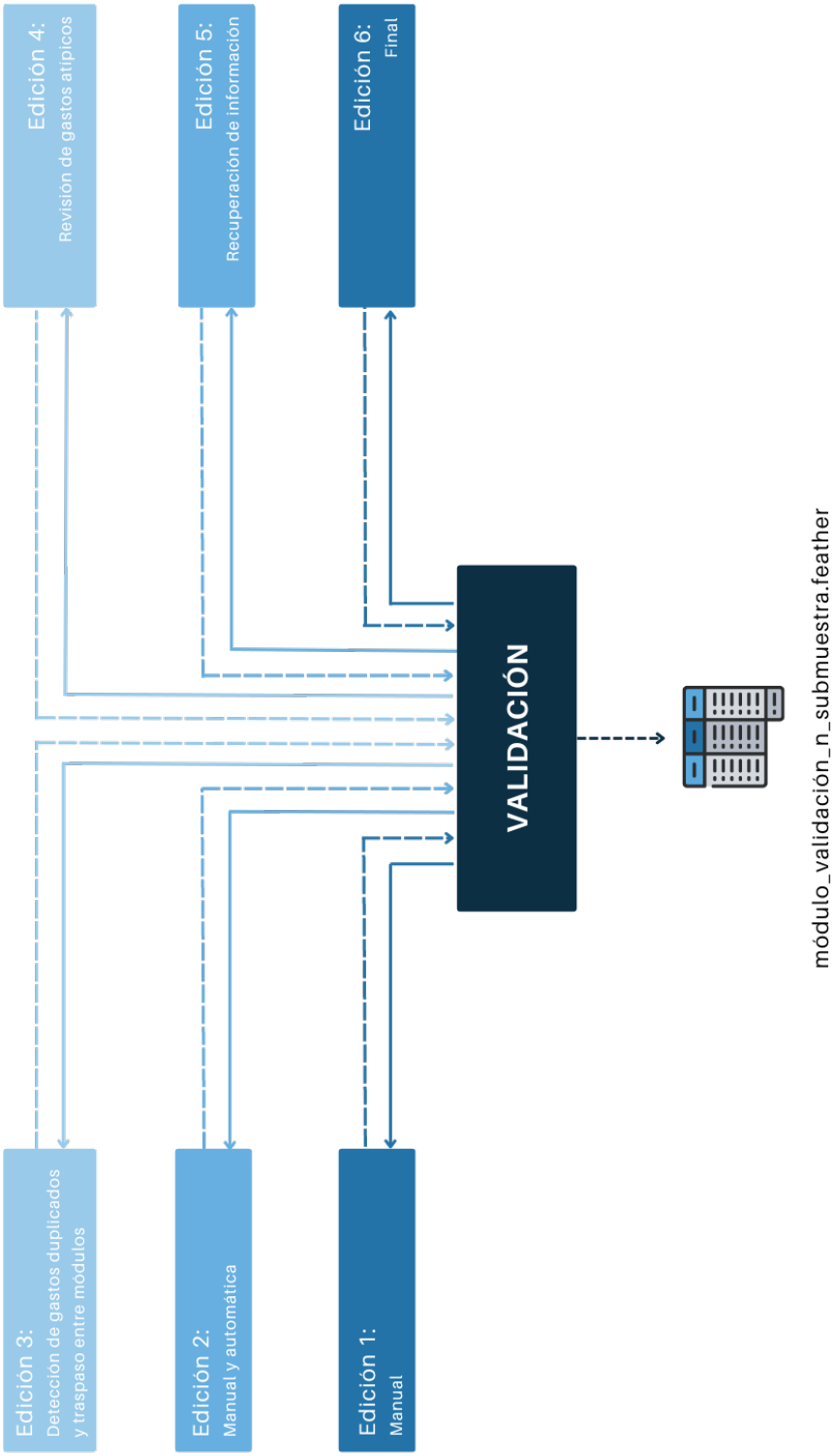
La encuesta tiene una duración de 15 días, período en que los integrantes del hogar deberán entregar información sobre sus gastos e ingresos. Se solicitará al hogar que informe sobre aquellos gastos de carácter mensual, como agua, luz, arriendo, entre otros, como también los gastos realizados en los últimos 3, 6 y 12 meses. Cada integrante del hogar de 15 años o más recibirá un formulario especial donde registrar sus gastos diarios. Para efectuar este trabajo, un encuestador INE visitará cuatro veces su hogar durante estos 15 días, con la finalidad de apoyar en el llenado de las libretas y resolver dudas. Cada visita demorará, en promedio, 30 minutos. Para realizar estas visitas, el encuestador se adaptará a su disponibilidad de tiempo.

¿LA INFORMACIÓN QUE YO ENTREGUE SERÁ PUBLICADA?

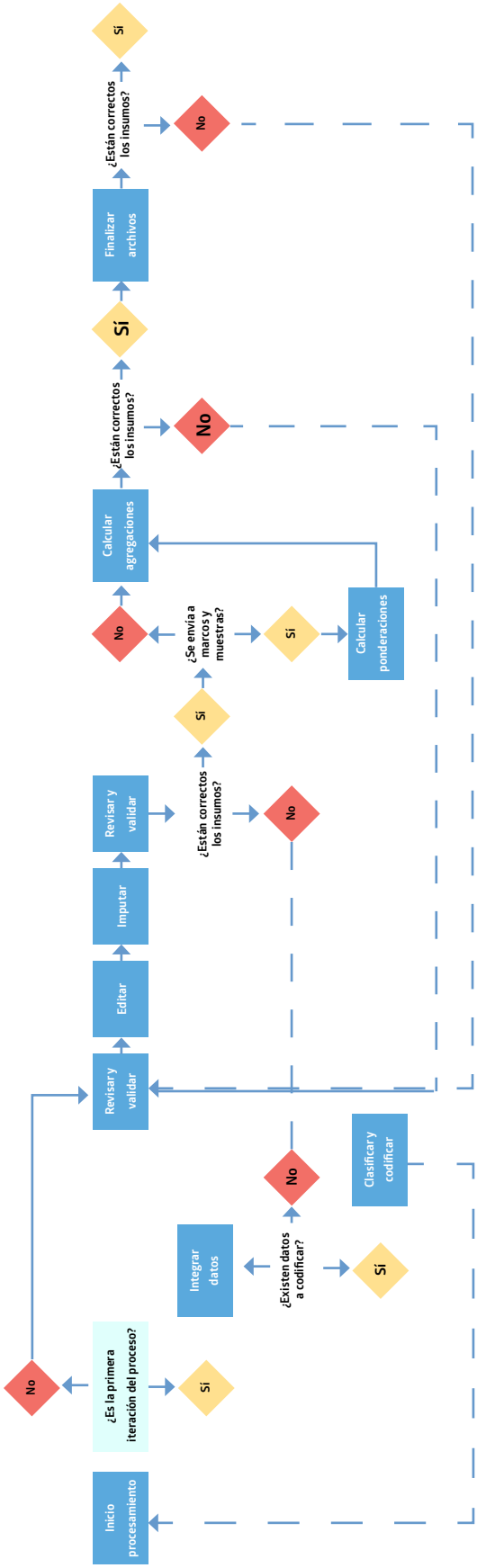
La información entregada será totalmente confidencial y se utilizará exclusivamente para este estudio. Para su tranquilidad, estos antecedentes están protegidos por el Secreto Estadístico, según lo dispuesto en la Ley N° 17.374, misma ley que faculta al INE a solicitar información a todas las personas.

La Encuesta de Presupuestos Familiares se rige por estándares internacionales de alta calidad y que utilizan los países pertenecientes a la Organización para la Cooperación y el Desarrollo Económico (OCDE). Responda con confianza, pues su información se encuentra protegida y resguardada por ley.

Anexo 6 Resumen del Flujo Subproceso de validación



Anexo 7 Resumen del Flujo de procesamiento



Instituto Nacional de Estadísticas • Chile

www.ine.cl