



Encuesta Laboral 2019

Informe metodológico
Encuesta Laboral 2019

INSTITUTO NACIONAL DE ESTADÍSTICAS

2020

Memoria

En memoria de Vicente Olavarría André † (1994 – 2020), Economista de la Universidad de Chile. Analista económico del Instituto Nacional de Estadísticas (INE) de Chile.

Destacado profesional, miembro del equipo técnico la Encuesta laboral, cuyos conocimientos, carácter crítico y reflexivo, vocación, compromiso y dedicación fueron fundamentales en el desarrollo de este producto.

SUBDIRECCIÓN TÉCNICA
SUBDEPARTAMENTO DE ESTADÍSTICAS ESTRUCTURALES DEL TRABAJO
DEPARTAMENTO DE ESTADÍSTICAS DEL TRABAJO
SUBDIRECCIÓN DE OPERACIONES
SUBDEPARTAMENTO DE IMPLEMENTACIÓN DE DISEÑOS

Encuesta Laboral

Encla 2019

Informe metodológico

Levantamiento nacional

Versión N°1

Instituto Nacional de Estadísticas, Chile.

Fecha de elaboración: noviembre de 2020.

El uso de un lenguaje que no discrimine ni marque diferencias entre hombres y mujeres ha sido una preocupación en la elaboración de este documento. Sin embargo, y con el fin de evitar la sobrecarga gráfica que supondría utilizar en castellano “o/a” para marcar la existencia de ambos sexos, se ha optado por utilizar -en la mayor parte de los casos- el masculino genérico, en el entendido de que todas las menciones representan siempre a hombres y mujeres, abarcando claramente ambos sexos.

Contenido

I.	Introducción.....	1
II.	Contexto metodológico	2
II.1.	Antecedentes de producto estadístico	2
II.2.	Objetivos del producto estadístico	3
III.	Marco de referencia	5
III.1.	Marco conceptual.....	5
III.2.	Referencias nacionales e internacionales	5
IV.	DISEÑO ESTADÍSTICO	6
IV.1.	Universo.....	6
IV.2.	Población objetivo	6
IV.3.	Marco estadístico	7
IV.3.1.	Conformación del marco muestral.....	7
IV.3.2.	Estratificación del marco muestral.....	8
IV.4.	Cobertura geográfica.....	9
IV.5.	Unidades estadísticas	11
IV.5.1.	Unidad de muestreo.....	11
IV.5.2.	Unidad de análisis.....	11
IV.6.	Definición de variables.....	11
IV.6.1.	Variables principales.....	11
IV.6.2.	Otras variables de interés.....	13
IV.7.	Nomenclatura utilizada.....	14
IV.8.	Diseño Estadístico.....	14
IV.8.1.	Diseño muestral.....	14
IV.8.2.	Selección muestral	14
IV.8.3.	Factores de expansión.....	21
IV.9.	Medidas de precisión.....	29
IV.9.1.	Varianza del estrato.....	29
IV.9.2.	Error relativo	29
IV.9.3.	Tasa de elegibilidad	30
IV.9.4.	Tasa de respuesta por empresa	30
IV.9.5.	Tasa de contacto	30
IV.9.6.	Tasa de rechazo	30
IV.9.7.	Tasa de cooperación.....	31
IV.9.8.	Tasa de logro objetivo	31
IV.10.	Estimación de parámetros.....	31
V.	METODOLOGÍA DE RECOLECCIÓN DE DATOS.....	33
V.1.	Verificación de la muestra	33
V.2.	Técnica y métodos de recolección de datos	33
V.2.1.	Estrategia de recolección de datos	33
V.2.2.	Digitación.....	37

V.2.3.	Período de recolección de datos	38
V.2.4.	Periodo de referencia	38
V.2.5.	Periodicidad.....	38
V.2.6.	Conformación de los equipos de trabajo	38
V.3.	Diseño del instrumento de captura de la información	41
V.4.	Prueba y evaluación del instrumento de captura de información	44
V.4.1.	Mejoras implementadas a los cuestionarios.....	44
V.4.2.	Prueba piloto operativa.....	52
V.5.	Método de capacitación para la recolección de datos	53
V.5.1.	Descripción general de la metodología de capacitación.....	53
V.5.2.	Materiales para la capacitación.....	55
V.6.	Métodos de supervisión	57
V.6.1.	Observaciones operativas	58
V.6.2.	Observaciones técnicas	58
V.6.3.	Supervisión cuestionarios digitados.....	59
VI.	METODOLOGÍA DE PROCESAMIENTO DE DATOS	59
VI.1.	Clasificación y codificación de datos.....	59
VI.1.1.	Codificación actividad económica	59
VI.1.2.	Codificación y depuración de preguntas abiertas	60
VI.2.	Análisis de consistencia de datos y edición	65
VI.2.1.	Lineamientos generales de la edición	65
VI.2.2.	Malla de validadores	67
VI.2.3.	Grilla de suficiencia.....	70
VI.3.	Procedimientos de edición	72
VI.3.1.	Resolución de errores de flujo	72
VI.3.2.	Resolución de errores de pregunta sin respuesta.....	73
VI.3.3.	Resolución de errores de rango	73
VI.3.4.	Resolución de inconsistencias	74
VI.4.	Imputación y creación de nuevas variables	74
VI.4.1.	Imputación de datos por medio de bases complementarias.....	74
VI.4.2.	Diagnóstico imputación.....	77
VI.4.3.	Creación de nuevas variables	78
VI.5.	Cálculo y ajustes de factores de expansión	80
VI.5.1.	Ajuste por elegibilidad.....	80
VI.5.2.	Suavizamiento del factor de expansión	82
VI.6.	Ejecución de cálculos agregados	82
VII.	METODOLOGÍA DE ANÁLISIS.....	83
VII.1.	Análisis de comportamiento de resultados	83
VII.2.	Comparación con registros administrativos	83
VII.3.	Aplicación del control de divulgación	83
VIII.	ALCANCE DE LA PRESENTACIÓN DE RESULTADOS	84
VIII.1.	Productos a difundir	84
VIII.1.1.	Microdatos	84

VIII.1.2. Metadatos	84
VIII.2. Difusión de resultados	85
VIII.3. Accesibilidad a los datos y marco legal	85
IX. Conceptos Claves	86
X. Bibliografía	88

Índice de tabulados

Tabla 1. Agrupación de actividad económica según CIIU4.CL 2012. Ámbito de estudio Encla 2019 ...	8
Tabla 2. Clasificación de las empresas según el número de trabajadores	9
Tabla 3. Agrupación regional de las empresas Encla 2019.	9
Tabla 4. Comunas excluidas del estudio Encla 2019	10
Tabla 5. Variables principales por cuestionario. Encla 2019.	12
Tabla 6. Otras variables de interés. Encla 2019.	13
Tabla 7. Elementos utilizados para el cálculo del tamaño muestral de Encla 2019	15
Tabla 8. Distribución de los tamaños muestrales de empresas por agrupación regional Encla 2019	18
Tabla 9. Distribución de los tamaños muestrales de empresas por tamaño de empresa según número de trabajadores Encla 2019	19
Tabla 10. Distribución de los tamaños muestrales de empresas por tamaño de empresa según agrupación de actividad económica Encla 2019	19
Tabla 11. Distribución de los tamaños muestrales de empresas por agrupación regional y tamaño de empresa según número de trabajadores Encla 2019	20
Tabla 12. Definición códigos de disposición final AAPOR	23
Tabla 13. Dotación recolectores de datos por región	40
Tabla 14. Dotación personal encargado del trabajo operativo	40
Tabla 15. Objetivos, dimensiones y módulos según cuestionario de la Encla 2019	42
Tabla 16. Variables contenidas en los módulos de cada cuestionario de la Encla 2019	43
Tabla 17. Nivel de modificación de las preguntas de los cuestionarios después de las distintas pruebas y evaluaciones de expertos	44
Tabla 18. Porcentaje de preguntas modificadas según cuestionario	45
Tabla 19. Porcentaje de las preguntas modificadas por cuestionario	45
Tabla 20. Porcentaje de las preguntas modificadas por cuestionario.	47
Tabla 21. Tipos de modificaciones de lenguaje inclusivo de género realizadas a los cuestionarios ..	48
Tabla 22. Porcentaje de las preguntas agregadas de los cuatro cuestionarios	49
Tabla 23. Porcentaje de las preguntas de los cuatro cuestionarios modificadas o agregadas por solicitud de la Dirección del Trabajo	51
Tabla 24. Módulos de contenidos de las capacitaciones	54

Tabla 25. Ejemplo cuestionario sindicatos.....	61
Tabla 26. Pregunta g22. “En su opinión, ¿por qué no hay más trabajadores y trabajadoras en su sindicato?” Antes y después de la codificación de las glosas de texto de g22_77_otro. “Otro. Especifique”	64
Tabla 27. Número de validadores en los cuestionarios de la Encla 2019	68
Tabla 28. Número de requisitos de suficiencia total y parcial en los cuestionarios de la Encla 2019	71
Tabla 29. Variables con requisitos de suficiencia total y parcial en los cuestionarios de la Encla 2019	72
Tabla 30. Distribución de respuestas de pregunta a1, sindicatos antes y después de imputar	75
Tabla 31. Distribución de respuestas de pregunta a1, trabajadores, antes y después de imputar ...	75
Tabla 32. Número de respuestas “no válidas” en preguntas de la base de sindicatos, antes y después de imputar	77
Tabla 33. Número de empresas en la muestra, según agrupación regional	79
Tabla 34. Número de empresas, según tamaño	79
Tabla 35. Número de empresas, según CIU4.CL 2012	80

I. Introducción

La Encuesta Laboral (Encla) es un producto mandatado por la Dirección del Trabajo (DT) del Ministerio del Trabajo y Previsión social, cuyo principal objetivo es identificar las condiciones de trabajo y las relaciones laborales al interior de las empresas regidas por el Código del Trabajo. Constituyendo un instrumento fundamental a la hora de estudiar el mundo del trabajo en la ocupación por la gran cantidad de temas que cubre.

Con el fin de generar un producto estadístico que logre cumplir con altos estándares de calidad, en 2019 se estableció que el INE además de diseñar la muestra, se encargaría de la recolección y procesamiento de la información estadística.

El presente documento busca dar cuenta del proceso de producción estadística involucrado en la recolección de la Encla 2019, para así entregar a las y los usuarios un mayor conocimiento de los datos. Todo proceso estadístico se desarrolla en base a una serie de etapas, decisiones y tareas que deben realizarse a fin de generar datos que cumplan con criterios de calidad adecuados, por lo que documentarlos se hace una parte elemental de la producción.

El presente manual se compone de los siguientes contenidos:

- Los capítulos II y III dan cuenta del contexto metodológico y el marco de referencia, dando a conocer los antecedentes de la encuesta (versiones anteriores), sus objetivos, así como los conceptos sobre los cuales gira, además de experiencias similares tanto nacionales como internacionales.
- El capítulo IV da cuenta del diseño estadístico de la encuesta, dando a conocer el universo del estudio, la población objetivo, la conformación del marco muestral, la cobertura geográfica y las unidades de muestreo y análisis.
- El capítulo V expone la metodología de recolección de datos. Describe los períodos de recolección, los flujos de trabajo, así como el diseño de los instrumentos y las pruebas a las que fueron sometidos.
- En el capítulo VI se aborda la metodología de procesamiento de datos, es decir cómo éstos se trabajan cuando son recepcionados. Esto incluye la clasificación de los datos, los procedimientos para detectar inconsistencias (malla de validadores, grilla técnica), los diversos mecanismos para corregirlos (ediciones, imputaciones) y los cálculos y ajustes de ponderadores.
- El capítulo VII describe los métodos mediante los cuales se analizaron los datos.

- Por último, en el capítulo VIII es presentado un listado de los productos de difusión de la encuesta.

II. Contexto metodológico

II.1. Antecedentes de producto estadístico

Desde el año 1998 la Dirección del Trabajo ha realizado la Encuesta Laboral (Encla) con periodicidad bienal o trienal por medio de distintas instituciones públicas y privadas. Si bien en años anteriores aquella entidad ha suscrito convenios con el Instituto Nacional de Estadísticas para que este elabore el diseño muestral de la encuesta, el año 2019 ambas organizaciones firmaron un novedoso convenio, mediante el cual se estableció que el organismo estadístico sería también el responsable de las fases de diseño, construcción, recolección y procesamiento de los datos. Además, ambas instituciones serán las encargadas de ejecutar las fases de análisis y difusión, potenciando así un instrumento que entrega datos de gran relevancia e interés.

La Encla surgió impulsada por la necesidad de contar con información sobre las relaciones y condiciones laborales en Chile. Para ello, se diseñó un estudio que pretendía cubrir una importante proporción del universo de empresas del país y se estableció una estrategia de recolección que incluyera a los principales actores del mercado laboral.

A lo largo de sus distintas versiones la encuesta ha tenido diversas modificaciones, entre las que destacan la ampliación de sus alcances; desde 2008 se aplica en todas las regiones; y el incremento de su tamaño muestral, que ha aumentado con el objetivo de lograr una mejor representatividad, así como expandir sus resultados al universo de las empresas del país. La versión 2019 constituye un hito relevante, pues se cumplen veinte años desde su primera aplicación¹. Por ello también, y con el objetivo de mejorar el producto estadístico, se introdujeron modificaciones en los cuestionarios, las cuales fueron sometidas a pruebas cognitivas. También se introdujeron cambios en el lenguaje para que éste fuera inclusivo. Además, hubo innovación en la etapa de procesamiento de datos, como la implementación de una grilla técnica.

De esta forma, la Encla 2019 es una instancia única para obtener información acerca de las transformaciones que ocurren en el mundo del trabajo y de las empresas, en cuanto a relaciones laborales y condiciones de empleo² y trabajo, así como también es un punto de partida para desarrollar nuevos estudios debido a la variedad de temas que cubre.

¹Las versiones anteriores fueron realizadas en los años 1998, 1999, 2002, 2004, 2006, 2008, 2011 y 2014. Ver informe de resultados de las distintas versiones de la Encla, en el siguiente link <http://www.dt.gob.cl/portal/1629/w3-propertyvalue-22780.html>. La versión 2017 no se publicó por problemas en el proceso de recolección, por lo que no existen antecedentes disponibles.

² En términos conceptuales, según los estándares de OIT (19° CIET, 2013), el concepto de empleo se relaciona con las ocupaciones que tienen un vínculo entre la unidad económica y el trabajador (trabajadores dependientes).

II.2. Objetivos del producto estadístico

El objetivo general del estudio es diagnosticar el estado y evolución de las condiciones de empleo y trabajo, de las relaciones laborales y de la igualdad de género en las empresas regidas por el Código del Trabajo en Chile.

Los objetivos específicos se detallan a continuación:

- Caracterizar las empresas en relación con su entorno económico

La Encla recoge la mirada que los empleadores, o sus representantes, tienen del entorno económico de las empresas, en cuanto a su posicionamiento frente a otras similares, a la evaluación de su desempeño económico y a sus expectativas para el futuro inmediato.

- Caracterizar las condiciones de empleo en las empresas

En lo que respecta a las condiciones de empleo, la encuesta considera tanto las formas de incorporación de los trabajadores a las empresas, como las posibilidades de desarrollo al interior de ellas. Adicionalmente, la Encla caracteriza las formas de retribución al trabajo, tales como las remuneraciones y el pago de gratificación.

- Caracterizar las condiciones de trabajo en las empresas

En materia de condiciones de trabajo, se describe la manera en la que la empresa organiza el tiempo de trabajo en términos de las jornadas laborales y sistemas de turnos. Además, se identifican los factores que inciden en la salud y seguridad de los trabajadores y se caracterizan los mecanismos que les permiten participar en la gestión organizacional.

- Analizar las relaciones laborales que se dan al interior de las empresas

En esta línea, la Encla se ocupa tanto de las formas de organización y acción colectiva, como de las diversas modalidades que adopta la negociación de las condiciones de empleo y trabajo. Además, analiza las expresiones de colaboración y conflicto entre los trabajadores organizados y la dirección de la empresa.

- Analizar la relación entre condiciones de empleo y trabajo y las relaciones laborales

Junto con caracterizar estos tres componentes de la realidad laboral, la encuesta examina los vínculos recíprocos entre determinadas condiciones de empleo y trabajo y los tipos de relaciones laborales que se dan en las empresas.

- Identificar la situación de las empresas en materia de igualdad de género

En esta materia, la Encla se propone informar acerca de las principales brechas, contrastar la presencia de hombres y mujeres en puestos de decisión e identificar la existencia de políticas de igualdad.

- Describir fenómenos laborales emergentes

Además de describir las características estructurales de la realidad laboral, la encuesta persigue dar cuenta de su dinamismo, evidenciando las estrategias adaptativas de los actores laborales y describiendo fenómenos que, en ocasiones, pueden tensionar el marco normativo vigente.

III. Marco de referencia

III.1. Marco conceptual

La Encla tiene como base las definiciones y directrices promovidas por la Organización Internacional del Trabajo (OIT)³ y las principales normas contenidas en el Código del Trabajo de Chile; base regulatoria de las relaciones laborales entre los trabajadores y los empleadores. La importancia de conocer el estado de las relaciones laborales, es que éstas son “el principal medio de acceso de los trabajadores a los derechos y beneficios asociados con el empleo, en las áreas del trabajo y la seguridad social. La existencia de una relación laboral es la condición necesaria para la aplicación de las leyes de trabajo y seguridad social destinadas a los empleados”⁴.

Respecto a sus principales conceptos, la Encla está organizada sobre la base de tres temas: las condiciones de empleo, las condiciones de trabajo y las relaciones laborales (Dirección del Trabajo, 2015).

Las condiciones de empleo dan cuenta de las modalidades de la inserción de los trabajadores en el mercado laboral. Este tipo de participación de los individuos es clave en las sociedades modernas, pues de ella depende la participación de trabajadores en la distribución del ingreso y el bienestar en general.

Las condiciones de trabajo corresponden al conjunto de aspectos o variables relativos al modo en que se organiza el proceso de trabajo, especialmente la organización de la fuerza de trabajo, los medios y el objeto de trabajo. Estas formas de organización del trabajo permiten conocer los efectos tanto organizativos como subjetivos que producen sobre los trabajadores.

Las relaciones laborales corresponden al conjunto de relaciones que se instauran entre empleadores y trabajadores, en el marco de las normas legales establecidas para regularlas. De este modo, la conflictividad habitual en las relaciones entre los actores laborales puede ser abordada mediante la regulación de los procesos de búsqueda en torno a las condiciones de empleo y trabajo, y conducida por cauces institucionales mediante procedimientos normados.

III.2. Referencias nacionales e internacionales

A nivel nacional existe la experiencia de la Encuesta sobre Relaciones Laborales en la Administración Central del Estado (Enclace), cuyo objetivo es analizar el estado y la evolución periódica de las relaciones laborales al nivel de los servicios de la administración central del Estado, excluyendo subsecretarías, intendencias, gobernación y siete servicios particulares debido a que se rigen por el

³ Éstas están orientadas a los cuatro objetivos estratégicos de la institución: i) promover y cumplir las normas y los principios y derechos fundamentales en el trabajo; ii) crear mayores oportunidades para que mujeres y hombres puedan tener empleos e ingresos dignos; iii) mejorar la cobertura y la eficacia de una seguridad social para todos y iv) fortalecer el tripartismo y el diálogo social (<https://www.ilo.org/global/about-the-ilo/mission-and-objectives/lang-es/index.htm>).

⁴ Ver: http://ilo.org/ifpdial/areas-of-work/labour-law/WCMS_165190/lang-es/index.htm.

Código del Trabajo (Universidad Alberto Hurtado, Servicio Civil, 2017). En sus objetivos es una encuesta muy similar a la Encla, pero aplicada al sector público.

A nivel internacional se encuentra la Encuesta sobre Relaciones Laborales en el Lugar de Trabajo (Workplace Employment Relation Study, WERS) del Reino Unido, que se plantea como principal objetivo mapear las relaciones laborales a lo largo del tiempo. De forma similar a la Encla, recopila datos en una muestra representativa de lugares de trabajo de distintas fuentes: empleadores, representantes de empleados y empleados (Department for Business, Innovation & Skills, 2014). La encuesta ha sido aplicada seis veces desde 1980.

Adicionalmente, se tiene como referencia la Encuesta de Relaciones Profesionales y Negociaciones de Empresa (Relations professionnelles et négociations d'entreprise, REPONSE) de Francia, que busca medir la intensidad y efectividad del diálogo social en la empresa y se realiza cada seis años desde 1993. Se compone de una entrevista con un representante de la gerencia, un representante del personal del establecimiento y otra entrevista a algún trabajador (RESPONSE, 2017).

En cuanto a la experiencia comparada en el ámbito del procesamiento de datos y controles de calidad, se revisó la experiencia de encuestas a empresas similares a la Encla 2019, como las procesadas por el Labour Bureau of Statistics (LBS) de EE. UU., las cuales también cuentan con instrumentos para el control de la calidad de la información estadística similares a los de Encla. Sin embargo, en comparación con ellas, la malla de validación propuesta para Encla 2019 es relativamente extensa, y, por lo tanto, altamente exigente. Además, en el INE la etapa de procesamiento de los datos incluye la aplicación de un segundo control de calidad, llamado grilla de suficiencia, que no está presente en las encuestas del LBS estudiadas.

IV. DISEÑO ESTADÍSTICO

IV.1. Universo

El universo del estudio son todas las empresas regidas por el Código del Trabajo que cuentan con iniciación de actividades en el Servicio de Impuestos Internos (SII), desarrollan su actividad económica dentro de los límites territoriales del país, informan un promedio mensual de cinco o más trabajadores asalariados durante el año 2016 (calculado a partir del Formulario 1887) y su actividad principal corresponde a alguna de las actividades económicas en estudio.

IV.2. Población objetivo

La población objetivo a encuestar de la Novena Encla son las empresas formales vigentes⁵ al año 2017, o sea, que cuentan con iniciación de actividades en el SII; que desarrollan alguna actividad económica en las secciones elegidas para el estudio⁶, según el Clasificador Chileno de Actividades

⁵ Se refiere a las empresas que se encontraban en funcionamiento y registraban movimiento comercial durante el año 2017.

⁶ Ver III.3.2. Estratificación del marco muestral.

Económicas (CIU4.CL 2012); y que durante el año comercial 2016 declararon, según el Formulario N°1887⁷, en promedio mensual cinco o más trabajadores contratados directamente por la empresa⁸ y cuyas relaciones laborales se rigen por el Código del Trabajo.

IV.3. Marco estadístico

El marco muestral es un listado que pretende identificar todos los elementos de la población objetivo. A continuación, se explica la construcción y posterior estratificación del marco muestral de la Encla 2019.

IV.3.1. Conformación del marco muestral

El marco muestral fue elaborado a partir del cruce entre el Directorio Nacional de Empresas del INE, año contable 2017 (DNE 2017), y del Formulario N°1887 del SII, año comercial 2016. La información contenida en los directorios antes mencionados proviene, principalmente, de los registros administrativos del SII y se complementa con información recogida en las encuestas continuas que realiza el INE.

Desde el directorio proveniente del F1887 se obtiene el promedio mensual de trabajadores asalariados en el año de referencia, mientras que del DNE se extraen las variables asociadas al RUT, actividad económica y región. En su conjunto, estas variables se utilizan para la posterior estratificación del marco muestral.

Respecto a la asignación de la actividad económica en el DNE 2017, la Sección de Directorio del INE es la encargada de clasificar masivamente a las empresas y realizar mejoras a determinados sectores económicos, según información de encuestas, verificaciones telefónicas u otras técnicas. Por otro lado, la Sección de Nomenclatura del organismo es la encargada de auditar todas las encuestas que recoge el INE y de revisar las actividades que realizan las unidades legales, de modo que la actividad económica final incorporada en el DNE 2017 se encuentre actualizada y difiera, en ocasiones, de la información inicial proveniente del SII.

Tras realizarse la fusión de ambos directorios, se obtiene un total de 292.604 empresas.

Una vez obtenido el listado de empresas en el paso anterior, se excluyen del marco las empresas que presentan, en promedio mensual, menos de cinco trabajadores contratados directamente, así como también aquellas ubicadas en comunas definidas como de difícil acceso o alto costo de recolección y aquellas que no tienen la región registrada en el DNE 2017. Luego de estas exclusiones, el listado contiene en total 192.848 empresas.

Al mismo tiempo, se excluyen aquellas empresas que se encuentran clasificadas en actividades que pertenecen al estudio, pero su razón social nos indica que pertenecen a municipalidades, corporaciones municipales u otro organismo estatal. Para la identificación de estos casos se utilizaron directorios complementarios tales como: (1) catastro de instituciones públicas, que dispone la

⁷ Sobre rentas del Art. 42 N°1 (sueldos), otros componentes de la remuneración y retención del impuesto único de segunda categoría.

⁸ Se excluye al personal a honorarios.

Sección de Directorio de Empresas del INE; (2) listado de empresas que no se rigen bajo el Código del Trabajo, proporcionado por el Departamento de Estudios de la Dirección del Trabajo.

Finalmente, luego de todas las exclusiones antes mencionadas, **el marco muestral cuenta con un total de 98.829 empresas** las que tenían en promedio mensual cinco o más trabajadores asalariados durante el año 2016.

IV.3.2. Estratificación del marco muestral

Tras la conformación del marco muestral, se procede a la estratificación, que corresponde al proceso de juntar elementos de una población en grupos homogéneos previo a la selección de la muestra. Su propósito es mejorar la precisión estadística de los estimadores agrupando las unidades del marco muestral en clases homogéneas en su interior y que difieran de las características del resto. Los estratos deben ser mutuamente excluyentes, es decir, cada elemento en la población debe ser asignado a un solo estrato. Además, los estratos deben ser exhaustivos colectivamente, es decir, ningún elemento de la población puede quedar excluido.

La estratificación del marco muestral se realizó según la agrupación de la actividad económica de acuerdo con la CIIU4.CL 2012, según tamaño de la empresa y según número de trabajadores y agrupación regional. A continuación, se expone la agrupación de actividad económica utilizada⁹:

Tabla 1. Agrupación de actividad económica según CIIU4.CL 2012. Ámbito de estudio Encla 2019

Agrupación actividad económica	Actividad económica	Descripción de la actividad económica
A	A	Agricultura, ganadería, silvicultura y pesca
B	B	Explotación de minas y canteras
C	C	Industrias manufactureras
D-E	D	Suministro de electricidad, gas, vapor y aire acondicionado
	E	Suministro de agua, evacuación de aguas residuales, gestión de desechos y descontaminación
F	F	Construcción
G	G	Comercio al por mayor y al por menor, reparación de vehículos automotores y motocicletas
H-J	H	Transporte y almacenamiento
	J	Información y comunicaciones
I	I	Actividades de alojamiento y de servicio de comidas
K-L	K	Actividades financieras y de seguros
	L	Actividades inmobiliarias
M-N	M	Actividades profesionales, científicas y técnicas
	N	Actividades de servicios administrativos y de apoyo
P	P	Enseñanza
Q	Q	Actividades de atención de la salud humana y de asistencia social
R-S	R	Actividades artísticas, de entretenimiento y recreativas
	S	Otras actividades de servicios

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

⁹ La estratificación por agrupación de actividad económica fue definida y solicitada por la Dirección del Trabajo.

Respecto al tamaño de la empresa según número de trabajadores se establecen cuatro categorías de clasificación: grande, mediana, pequeña y micro. En la Tabla 2 se observan los límites que definen el tamaño de la empresa según el número de trabajadores.

Tabla 2. Clasificación de las empresas según el número de trabajadores

Tamaño de la empresa según su número de trabajadores	Número de trabajadores	
	Límite inferior	Límite superior
Grande	200	Más
Mediana	50	199
Pequeña	10	49
Micro	5	9

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Finalmente, la estratificación por agrupación regional se muestra en la Tabla 3.

Tabla 3. Agrupación regional de las empresas Encla 2019.

Agrupación Regional	Región
Norte Grande	Arica y Parinacota Tarapacá Antofagasta
Norte Chico	Atacama Coquimbo
Zona Centro	Valparaíso Libertador General Bernardo O'Higgins Metropolitana de Santiago
Centro Sur	Maule Ñuble Biobío
Sur	La Araucanía Los Ríos Los Lagos
Zona Austral	Aysén del General Carlos Ibáñez del Campo Magallanes y La Antártica Chilena

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

IV.4. Cobertura geográfica

La cobertura geográfica de la encuesta está constituida por las 16 regiones del país, excluyendo aquellas empresas cuya casa matriz se encuentra en comunas consideradas como zonas de difícil acceso o que presentan un alto costo de recolección, las cuales se especifican a continuación:

Tabla 4. Comunas excluidas del estudio Encla 2019

Región	Provincia	Comunas
Arica y Parinacota	Parinacota	General Lagos
Tarapacá	Tamarugal	Colchane
Antofagasta	El Loa	Ollagüe
Valparaíso	Isla De Pascua	Isla de Pascua
Valparaíso	Valparaíso	Juan Fernández
Los Lagos	Llanquihue	Cochamó
Los Lagos	Palena	Chaitén
Los Lagos	Palena	Futaleufú
Los Lagos	Palena	Hualaihué
Los Lagos	Palena	Palena
Aysén del General Carlos Ibáñez del Campo	Aisén	Guaitecas
Aysén del General Carlos Ibáñez del Campo	Coihayque	Lago Verde
Aysén del General Carlos Ibáñez del Campo	Capitán Prat	O'Higgins
Aysén del General Carlos Ibáñez del Campo	Capitán Prat	Tortel
Magallanes y de la Antártica Chilena	Antártica Chilena	Antártica
Magallanes y de la Antártica Chilena	Antártica Chilena	Cabo de hornos
Magallanes y de la Antártica Chilena	Magallanes	Laguna Blanca
Magallanes y de la Antártica Chilena	Magallanes	Río Verde
Magallanes y de la Antártica Chilena	Magallanes	San Gregorio
Magallanes y de la Antártica Chilena	Tierra del Fuego	Timaukel
Magallanes y de la Antártica Chilena	Tierra del Fuego	Primavera
Magallanes y de la Antártica Chilena	Última Esperanza	Torres del Paine

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Finalmente, la ubicación geográfica de la empresa queda determinada por la localización regional de la casa matriz, señalada en el DNE.

IV.5. Unidades estadísticas

IV.5.1. Unidad de muestreo

La unidad de muestreo es la **empresa** que cuenta con iniciación de actividades en el SII, desarrolla su actividad económica dentro de los límites territoriales del país, informa un promedio mensual de cinco o más trabajadores asalariados durante el año 2016 (calculado a partir del Formulario 1887) y su actividad principal corresponde a alguna de las actividades económicas en estudio.

IV.5.2. Unidad de análisis

La unidad de análisis es la **empresa** formal que cumple los requisitos para formar parte de la población objetivo.

IV.6. Definición de variables

IV.6.1. Variables principales

Las principales variables de Encla 2019 fueron definidas en torno a dos aristas:

- Variables imprescindibles para estimar el factor de expansión.
- Variables claves para el cumplimiento de los objetivos del estudio, encargadas por la Dirección del Trabajo, previa revisión por parte del INE.

Tabla 5. Variables principales por cuestionario. Encla 2019.

Cuestionario	Variables principales	Definición
Empleadores	A1	Señale las actividades o giros realizados por la empresa en orden de importancia, según ventas (1°, 2° y 3° lugar)
	A2	Indique la combinación de actividades y productos o servicios, según importancia en sus ventas
	F3	¿Actualmente esta empresa tiene sindicatos vigentes?
	F4	¿Durante el año 2018, esta empresa tenía sindicatos?
	H2	¿Actualmente se encuentra vigente algún contrato o convenio colectivo en esta empresa?
Sindicatos	A2	¿Durante el año 2018 la empresa que le paga empleó personal subcontratado y/o suministrado por terceros para realizar alguna labor similar a la del personal contratado directamente por la empresa?
	B4	El reajuste de las remuneraciones obtenido por los trabajadores y las trabajadoras durante el año 2018 fue logrado:
	G2	¿Cuántos hombres y mujeres se encuentran afiliados a su sindicato el día de hoy?
	H1	¿Durante el año 2018 H1, los trabajadores y las trabajadoras de la empresa han llevado a cabo alguna de las siguientes formas de expresión de conflicto?
	I3	¿Su sindicato ha negociado colectivamente desde el año 2015 a la fecha?
Trabajadores	I15	¿Como resultado de la última negociación colectiva efectuada por su sindicato, la empresa realizó algún reajuste de las remuneraciones para los trabajadores y las trabajadoras?
	A2	¿Durante el año 2018 la empresa que le paga empleó personal subcontratado y/o suministrado por terceros para realizar alguna labor similar a la del personal contratado directamente por la empresa?
Autoaplicado	A1	Indique el total de personas contratadas directamente por esta empresa al 30 de noviembre de 2018, según región del país y sexo
	A2	Indique el total de personas contratadas directamente por esta empresa al 30 de noviembre de 2018, en cada una de las siguientes categorías según su nivel de autoridad y sexo
	B1	Indique el total de personas contratadas directamente por esta empresa al 30 de noviembre de 2018, según tipo de contrato y sexo
	B2	Indique el total de personas contratadas directamente por esta empresa al 30 de noviembre de 2018, según tipo de contrato y tramo etario
	B5	Indique el total de personas contratadas a honorarios que trabajaban en esta empresa al 30 de noviembre de 2018, según sexo
	C1	¿Durante el año 2018 esta empresa subcontrató la actividad económica principal a otra(s) empresa(s)?
	C7	¿Durante el año 2018 esta empresa subcontrató otra actividad económica (distinta de la principal) a otras empresas o a terceros?
	C15	¿Durante el año 2018 esta empresa contó con personas suministradas por terceros (empresas de servicios transitorios)?
	D1	Indique el total de personas contratadas directamente por la empresa con jornada completa (de 40 horas o más a la semana) al 30 de noviembre de 2018, según tramo de remuneraciones brutas y sexo
	D5	Indique el total de personas contratadas directamente por la empresa con jornada parcial (30 horas o menos a la semana) al 30 de noviembre de 2018, según tramo de remuneraciones brutas y tramo etario
	E1	¿La empresa tuvo personas extranjeras contratadas directamente al 30 de noviembre de 2018?
	J1	¿Cuántos sindicatos y grupos negociadores vigentes hay en esta empresa?
	J4	Para los instrumentos colectivos vigentes al 30 de noviembre de 2018, indique el número de trabajadores y trabajadoras suscritos, así como a quienes el empleador extendió beneficios, según tipo de instrumento colectivo

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

IV.6.2. Otras variables de interés

Además de las variables principales, se construyeron otras a partir de las variables recolectadas con la Encla 2019, tanto en forma independiente como en relación con otras y con registros administrativos. La importancia de estas variables radica en que sintetizan información de dos cuestionarios, para su construcción se requirieron registros administrativos y/o datos que estaban presentes en las bases de la Encla 2014.

Tabla 6. Otras variables de interés. Encla 2019.

Variable de interés	Definición
Tasa de afiliación sindical	Se dividió el número de afiliados al sindicato entrevistado por el número de trabajadores de la empresa ($g2_3 / NT$) ¹⁰
Tasa de feminización de la empresa	Se calculó dividiendo la variable $a1_17_1$ por la variable NT , ambas provenientes de cuestionario autoaplicado ($a1_17_1 / NT$).
Tasa de feminización sindical	Fue derivada de la división de las variables del cuestionario a sindicatos $g2_1$ (mujeres afiliadas al sindicato al momento en que se hizo la encuesta) con $g2_3$ (hombres y mujeres afiliados al sindicato al momento en que se hizo la encuesta): $g2_1 / g2_3$.
Tasa de extranjerización sindical	Derivada de la división de las variables del cuestionario a sindicatos $g3_2$ (extranjeros afiliados al sindicato al momento en que se hizo la encuesta) con $g2_3$ (hombres y mujeres afiliados al sindicato al momento en que se hizo la encuesta): $g3_2 / g2_3$.
Densidad sindical en la empresa	Derivada de la división del número de trabajadores sindicalizados en la empresa y el total de trabajadores en esta (NT).
Presencia de sindicatos	Para su elaboración se utilizaron los registros administrativos de la DT de sindicatos. Se anidaron los sindicatos según RUT de empresa en dos bases de datos: una de empresas con sindicatos activos y otra de empresas con sindicatos activos o en receso. Los RUT de estas empresas con sindicatos activos o con sindicatos en receso fueron buscados en las bases de empleadores de la Encla 2019

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

¹⁰ 25 casos tenían valores superiores a 1, los que iban desde 1,02 hasta 15. Estos pasaron a ser 1 luego de la edición de los datos.

IV.7. Nomenclatura utilizada

Para facilitar la comparación de la información y de acuerdo con las normas internacionales, se utilizó el Clasificador Chileno de Actividades Económicas (CIIU4.CL 2012), compuesto por 18 sectores económicos que luego fueron reclasificados en 13 agrupaciones económicas. Dichos sectores y agrupaciones se presentan en la Tabla 1.

El Clasificador Chileno de Actividades Económicas CIIU4.CL 2012 se rige por los mismos conceptos y criterios del Clasificador Internacional CIIU Rev.4, elaborado por la División de Estadísticas de la Organización de las Naciones Unidas (ONU).

IV.8. Diseño Estadístico

IV.8.1. Diseño muestral

El diseño muestral para la Encla 2019 es probabilístico y estratificado, los 302 estratos se encuentran definidos por las combinaciones donde existen empresas de las siguientes variables¹¹: agrupación de actividad económica de acuerdo con el CIIU4.CL 2012 (trece agrupaciones), tamaño de la empresa según número de trabajadores (cuatro agrupaciones) y agrupación regional¹² (seis agrupaciones).

Los tamaños muestrales fueron determinados con el objetivo de tener estimaciones precisas para los siguientes niveles de estimación:

1. Agrupación regional
2. Tamaño de empresa, según número de trabajadores.
3. Agrupación de actividad económica.
4. Agrupación regional y tamaño de empresa según número de trabajadores.

IV.8.2. Selección muestral

IV.8.2.1. Cálculo del tamaño muestral

Con el objetivo de obtener una adecuada precisión para la estimación del total del número de trabajadores asalariados durante un mes de referencia, se utilizó como pivote el tamaño de la muestra objetivo correspondiente al diseño muestral de 2017. Sin embargo, considerando los cambios que se presentan en los niveles de estratificación para esta versión de la encuesta fue necesario realizar ajustes adicionales que se describen en este apartado, así como los errores relativos y las fórmulas empleadas para su estimación.

¹¹ Son 302 estratos porque en el marco muestral no hay empresas en todas las combinaciones posibles de agrupación regional, agrupación de actividad económica y tamaño de empresa.

¹² Según la región donde se ubica la casa matriz de la empresa.

Paso 1. Cálculos de los tamaños muestrales objetivo por estrato utilizados para el diseño 2017, los que fueron calculados mediante la siguiente fórmula:

$$n_r^{2017} = \frac{(\sum_{r=1}^R N_r^{2017} \cdot \sigma_r)^2 \cdot \left(Z_{1-\frac{\alpha}{2}}\right)^2}{e_r \cdot \sum_{r=1}^R Y_r + (\sum_{r=1}^H N_r^{2017} \cdot \sigma_r^2) \cdot \left(Z_{1-\frac{\alpha}{2}}\right)^2} \quad (1)$$

Donde:

n_r^{2017} : Tamaño de la muestra objetivo de empresas en la región r .

N_r^{2017} : Número de empresas en el marco muestral de la región r .

σ_r : Desviación estándar del número de trabajadores de la región r .

$Z_{1-\frac{\alpha}{2}}$: Estadístico normal estándar correspondiente a un nivel de confianza del 95%
($Z_{0,975}=1,96$)

e_r : Error relativo de la estimación.

Y_r : Total del número de trabajadores en la región r en el marco muestral.

En la siguiente tabla se detallan los elementos considerados para calcular el tamaño de la muestra correspondiente a la encuesta laboral, en su versión 2019:

Tabla 7. Elementos utilizados para el cálculo del tamaño muestral de Encla 2019¹³

Parámetro	Descripción
Variable de diseño	Número de trabajadores: y
Estimador asociado	Estimador del total del número de trabajadores: \hat{Y}
Parámetro asociado	Total del número de trabajadores: Y
Niveles de estimación para los cuales se establece el error	Nacional
	Agrupación regional
	Tamaño de empresas
	Agrupación de actividad económica
	Cruce de agrupación regional y tamaño de empresas
Errores de muestreo esperados (error relativo máximo admisible)	Nacional: 5%.
	Agrupación regional: 5%
	Tamaño de empresas: 6%
	Agrupación de actividad económica: 11%
Nivel de confianza	Cruce de agrupación regional y tamaño de empresas: 10%.
	95%

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

¹³ Los errores muestrales que se obtuvieron cuando se realizó el diseño muestral se pueden observar con más detalle en el Informe de Diseño Muestral de Encla 2019.

En la Encla 2017 se utilizó la región r y en la Encla 2019 se distribuyó el tamaño obtenido a nivel del estrato agrupación regional (n_h), a través de la afijación proporcional al tamaño, en los estratos de tamaño de empresa según número de trabajadores y a su vez a nivel de la agrupación de actividad económica. La distribución se determinó mediante la siguiente fórmula:

$$n_{hks} = n_h \cdot \left(\frac{N_{hks}}{N_h} \right) \quad (2)$$

Donde:

n_{hks} : Tamaño de la muestra objetivo de empresas en la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s .

n_h : Tamaño de la muestra de empresas en la agrupación regional h .

N_{hks} : Número de empresas en el marco muestral en la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s .

N_h : Número de empresas en el marco muestral de la agrupación regional h .

Luego de la distribución anterior, el tamaño de la muestra se ajusta siguiendo las siguientes reglas

- En caso de que el número de empresas del estrato en el marco muestral (N_{hks}) sea menor o igual a 6, se incluyen todas las empresas del marco muestral en la muestra (censo)¹⁴.
- En caso de que el número de empresas del estrato en el marco muestral (N_{hks}) sea mayor a 6:
 - Si el tamaño de muestra objetivo (n_{hks}) calculado en el paso anterior es mayor al número de empresas del estrato en el marco muestral (N_{hks}), se incluyen todas las empresas del marco muestral en la muestra (censo).
 - Si el tamaño de muestra objetivo (n_{hks}) calculado en el paso 2 es mayor a 6, se mantiene este tamaño de muestra.

Así, el total del tamaño muestral para la agrupación regional corresponde a la suma de los tamaños muestrales obtenidos para estrato conformado por tamaño empresa y agrupación de actividad económica, tal como se expresa en la siguiente fórmula.

$$n_h = \sum_{k=1}^K \sum_{s=1}^S n_{hks} \quad (3)$$

Donde:

n_h : Tamaño de la muestra de empresas en la agrupación regional h .

¹⁴ Basado en recomendaciones metodológicas para la CIS 2016 (European Comission, 2016).

n_{hks} : Tamaño de la muestra de empresas en la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s .

Finalmente, el total del tamaño muestral a nivel nacional corresponde a la suma de los tamaños muestrales obtenidos para cada agrupación regional, expresado en la fórmula (4).

$$n = \sum_{h=1}^L n_h \quad (4)$$

Donde:

n : Tamaño de la muestra de empresas a nivel nacional.

n_h : Tamaño de la muestra de empresas en la agrupación regional h .

Paso 2. Con los tamaños muestrales obtenidos finalmente en el paso 1, se agrupan los tamaños muestrales de las regiones para obtener un tamaño muestral a nivel de la agrupación regional definida. De igual manera, se realiza este procedimiento para obtener los tamaños de la muestra por sectores económicos agrupados para la Encla 2019 a partir de las actividades económicas de la Encla 2017.

Paso 3. Posteriormente con la información de la variable promedio mensual de trabajadores asalariados para el año 2016, contenida en el marco muestral conformado para la Encla 2019, se calculan las varianzas, coeficientes de variación y errores relativos esperados a nivel de cada estrato.

Paso 4. Luego de calculados los errores esperados para la estimación del total de trabajadores asalariados, se revisa que en el cruce de los niveles de estimación, agrupación regional y tamaño de empresa no supere 10%. Asimismo, se asegura que el nivel de agrupación de actividad económica no supere el 11%. En caso de que esto ocurra, se debe aumentar el tamaño de la muestra en los estratos correspondientes.

Paso 5. Una vez definido el tamaño muestral objetivo, se calcula una muestra con sobremuestreo (sm) cuya finalidad es cubrir aquellas empresas que en el proceso de trabajo de campo no son logradas, presentando incidencias como: cerradas, término de giro, quebradas, entre otros. Se utiliza la estrategia de sobremuestreo debido a que permite gestionar la muestra objetivo y unidades adicionales simultáneamente en terreno para cubrir las pérdidas mencionadas.

La sobremuestra se define como un porcentaje específico para cada estrato, según la tasa de no logro de la encuesta anterior. Sin embargo, resultó necesario realizar el cálculo tomando en cuenta una tasa de no respuesta ponderada de las versiones 2014 y del diseño 2017 para cada a nivel de región y tamaño. La ponderación considerada para Encla 2014 es de 80%, mientras que para la 2017 fue un 20%. El sobremuestreo se calcula mediante la siguiente fórmula:

$$n_{12019} = \frac{n_{2019}}{1 - ((0,8 \cdot tnr_{2014}) + (0,2 \cdot tnr_{2017}))} \quad (5)$$

Donde:

n_{12019} : Tamaño de la muestra ajustada¹⁵ para Encla 2019.

n_{2019} : Tamaño de la muestra objetivo para Encla 2019.

tnr_{2014} : Tasa de no logro de Encla 2014.

tnr_{2017} : Tasa de no logro de Encla 2017.

En las siguientes tablas se presenta el tamaño de muestra objetivo y el total de empresas de sobremuestreo. El tamaño muestral objetivo es de 4.068 empresas a nivel nacional distribuidas en 302 estratos compuestos por agrupación regional, tamaño de empresa y agrupación de actividad económica. Respecto a la agrupación regional, el mayor tamaño muestral de empresas se concentra en la zona centro con 1.808 empresas, de igual forma esta zona presenta la mayor cantidad de empresas de sobremuestreo.

Tabla 8. Distribución de los tamaños muestrales de empresas por agrupación regional Encla 2019

Agrupación económica agrupada	Marco muestral	Muestra objetivo	Muestra con sobremuestreo
Total	98.829	4.068	5.800
Norte grande	5.170	457	587
Norte chico	4.378	338	473
Zona centro	63.376	1.808	2.697
Centro sur	14.059	568	777
Sur	10.199	593	831
Zona austral	1.647	304	435

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Por otra parte, al observar la distribución de los tamaños muestrales por tamaño de empresa, para las empresas de tamaño grande se requiere alcanzar un tamaño de 1.580 empresas, siendo éste el tamaño de empresa con mayor tamaño muestral.

¹⁵ Tamaño muestral que incluye sobremuestreo.

Tabla 9. Distribución de los tamaños muestrales de empresas por tamaño de empresa según número de trabajadores Encla 2019

Tamaño empresa	Marco muestral	Muestra objetivo	Muestra con sobremuestreo
Total	98.829	4.068	5.800
Grande	3.020	1.580	2.256
Mediana	9.266	864	1.179
Pequeña	43.410	1.229	1.713
Micro	43.133	395	652

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

En cuanto a la distribución de la muestra objetivo por agrupación de actividad económica, se tiene que la actividad con mayor tamaño muestral solicitado es G, “Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas”, seguida por la actividad P, “Enseñanza”.

Tabla 10. Distribución de los tamaños muestrales de empresas por tamaño de empresa según agrupación de actividad económica Encla 2019.

Agrupación actividad económica	Marco muestral	Muestra objetivo	Muestra con sobremuestreo
Total	98.829	4.068	5.800
A	9.602	194	282
B	724	122	157
C	12.259	304	455
D-E	613	115	140
F	12.409	506	730
G	20.471	742	1.088
H-J	10.093	286	422
I	6.002	274	386
K-L	3.986	223	283
M-N	11.479	457	688
P	5.776	517	725
Q	2.100	195	264
R-S	3.315	133	180

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Tabla 11. Distribución de los tamaños muestrales de empresas por agrupación regional y tamaño de empresa según número de trabajadores Encla 2019

Agrupación regional	Tamaño empresa	Marco muestral	Muestra objetivo	Muestra con sobremuestreo
	Total	98.829	4.068	5.800
Norte grande		5.170	457	587
	Grande	83	83	83
	Mediana	409	140	192
	Pequeña	2.205	177	232
	Micro	2.473	57	80
Norte chico		4.378	338	473
	Grande	62	62	62
	Mediana	361	110	153
	Pequeña	2.022	117	162
	Micro	1.933	49	96
Zona centro		63.376	1.808	2.697
	Grande	2.367	1.104	1.688
	Mediana	6.359	227	291
	Pequeña	27.981	385	566
	Micro	26.669	92	152
Centro sur		14.059	568	777
	Grande	283	175	226
	Mediana	1.246	160	217
	Pequeña	6.182	172	233
	Micro	6.348	61	101
Sur		10.199	593	831
	Grande	204	135	176
	Mediana	758	159	213
	Pequeña	4.318	216	303
	Micro	4.919	83	139
Zona austral		1.647	304	435
	Grande	21	21	21
	Mediana	133	68	113
	Pequeña	702	162	217
	Micro	791	53	84

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Las empresas son seleccionadas aleatoriamente, al interior de cada estrato dado por el cruce de agrupación regional, tamaño de la empresa según número de trabajadores y agrupación de actividad

económica. El método aplicado corresponde a un muestreo sistemático con igual probabilidad de selección, a partir de la implementación de los siguientes pasos:

1. Para asegurar un recorrido completo dentro de cada estrato, se ordenan las unidades muestrales en forma ascendente por agrupación regional, tamaño de empresas y agrupación de actividades económicas; y de forma descendente según la variable número de trabajadores.
2. A través del software estadístico SPSS se utiliza el plan de muestras complejas, seleccionando el muestreo sistemático simple.
3. Se especifica en el plan el tamaño muestral por estrato (incluido el sobremuestreo) requerido para la selección.
4. Por defecto, el programa guarda las probabilidades de inclusión aleatoria, pero de igual forma se solicita al SPSS que almacene las variables de tamaño de la población, tamaño de la muestra y ponderación muestral.
5. El valor de la semilla utilizada para la generación de números aleatorios no se encuentra definido previamente, sino que se solicita un número seleccionado en forma aleatoria por el programa.
6. Una vez extraída la selección aleatoria, se genera un archivo que contenga las unidades seleccionadas en forma aleatoria y sus variables de estratificación.

IV.8.3. Factores de expansión

El factor de expansión asignado a una empresa en una muestra probabilística afecta las magnitudes recabadas en la encuesta, de manera que puedan ser representativas del universo. Este cálculo es, técnicamente, el inverso de la probabilidad de la selección, sin embargo se deben realizar ajustes a dicho factor para incorporar la información proveniente del trabajo de campo, donde se reflejan las incidencias que impiden la recolección total de la muestra seleccionada.

Para las versiones anteriores de la Encla se realizaba el cálculo de dos factores de expansión correspondientes a uno por empresas y otro por número de trabajadores, sin embargo, de acuerdo a la incorporación de nuevos ajustes metodológicos incluidos en esta última versión y al resultado de las sucesivas comparaciones entre los factores de expansión obtenidos con las distintas metodologías, se observa que no existen diferencias significativas al utilizar el factor de expansión de empresas o el de trabajadores. Por lo tanto, para esta versión de la encuesta solo se realiza el cálculo del factor de expansión por empresas, describiendo, en este apartado, el nuevo algoritmo aplicado y sus distintos ajustes.

En este contexto, uno de los estudios realizados se trató acerca de una comparación de la estimación de la variable de interés, total de trabajadores, utilizando los factores de expansión por número de trabajadores y por empresas, según la metodología de cálculo de las versiones anteriores de la encuesta. Esto dio como resultado que siempre los intervalos de confianza y la estimación puntual

de alguno de los dos factores contenía al otro. Para ver más detalladamente estas comparaciones se recomienda revisar el Informe de Diseño Muestral de Encla 2019¹⁶.

Por otra parte, cada año el INE trabaja cuidadosamente para que las fuentes de información con las que se construyen los marcos muestrales sean más robustas y coherentes con la realidad nacional, permitiendo así incorporar ajustes necesarios y actualizar las metodologías de las encuestas, lo que fue aplicado para esta versión de la Encla 2019.

IV.8.3.1. Factor de expansión teórico

El factor de expansión teórico corresponde al inverso de la probabilidad de selección de la empresa i perteneciente al estrato de la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , la cual se calcula mediante la siguiente fórmula:

$$P_{hks_i} = \frac{n_{hks}^{sm}}{N_{hks}} \quad (6)$$

Donde:

- P_{hks_i} : Probabilidad de selección de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- n_{hks}^{sm} : Tamaño de muestra con sobremuestreo ajustado de la Encla 2019, en la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- N_{hks} : Número de empresas en el marco muestral de la Encla 2019, en la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

Por lo tanto, el factor de expansión teórico se obtiene mediante la siguiente fórmula:

$$w_{hks_i} = \frac{1}{P_{hks_i}} \quad (7)$$

Donde:

- w_{hks_i} : Factor de expansión teórico de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- P_{hks_i} : Probabilidad de selección de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

¹⁶ El Informe de Diseño Muestral de Encla 2019 puede encontrarse en el micrositio en Encla, junto con toda la documentación y bases de datos del producto: <https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral/condiciones-de-empleo-y-relaciones-laborales>.

IV.8.3.2. Ajuste por elegibilidad

Tras el trabajo de campo, se clasifican las empresas según su estado de elegibilidad, diferenciando entre unidades “elegibles”, “no elegibles” o “elegibilidad desconocida”.

Previo al cálculo de los factores de expansión, se debe clasificar a las empresas seleccionadas para la muestra según el resultado del proceso de recolección. Para esto, se utilizan los códigos de disposición final de la AAPOR (2016) los cuales además se pueden clasificar según estado de elegibilidad. En la Tabla 12 se presenta la descripción de los códigos de disposición final.

Tabla 12. Definición códigos de disposición final AAPOR

Información de Recolección		Clasificación AAPOR	
Estado Empresa	Estado Encuesta	Estado Elegibilidad	Estados de Disposición Final
Activa	Terminada	Elegible	Entrevista completa
Activa	En análisis	Elegible	Entrevista parcial
Activa	Enviada	Elegible	Entrevista parcial
Activa	Fuera de marco	No elegible	No elegible
Activa	No enviado	Elegible	Elegible no entrevistado
Activa	Recepcionada	Elegible	Entrevista parcial
Activa	Rechazo	Elegible	Elegible no entrevistado
Activa	Incompleta	Elegible	Entrevista parcial
Inactiva	Quebrada	No elegible	No elegible
Inactiva	Sin movimiento	No elegible	No elegible
Inactiva	Término de giro	No elegible	No elegible
Inactiva	Paralizada	No elegible	No elegible
Inactiva	Cambio de RUT	No elegible	No elegible
Inactiva	Cerrada	No elegible	No elegible
Inubicable	Inubicable	Elegibilidad desconocida	Elegibilidad desconocida

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

IV.8.3.3. Ajuste por elegibilidad desconocida

Para realizar el ajuste por elegibilidad desconocida, dentro de cada estrato se distribuye el peso de las empresas con dicho estado entre las empresas con elegibilidad conocida (elegibles y no elegibles), para lo cual es necesario calcular una razón de ajuste mediante la siguiente fórmula:

$$\hat{R}_{hks}^{ed} = \frac{\sum_{i \in \Omega_{hks}} w_{hks_i}}{\sum_{i \in \Omega_{hks}^{ec}} w_{hks_i}} \quad (8)$$

Donde:

- \hat{R}_{hks}^{ed} : Razón de ajuste por elegibilidad desconocida para las empresas pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- w_{hks_i} : Factor de expansión teórico de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

- Ω_{hks} : Conjunto de empresas en la muestra pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- Ω_{hks}^{ec} : Conjunto de empresas en la muestra pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s (elegibles y no elegibles).

A partir de esta etapa, las empresas clasificadas en condición de elegibilidad desconocida no son consideradas en el factor de expansión y se les asigna una ponderación nula, ya que su peso se distribuye en aquellas unidades cuya elegibilidad es conocida (elegibles y no elegibles). Para estas últimas, el factor de expansión ajustado por elegibilidad desconocida corresponde al producto entre la razón de ajuste y el factor de expansión teórico. Por lo tanto, el factor de expansión corregido por elegibilidad desconocida se obtiene mediante la siguiente fórmula:

$$w_{hks_i}^{ed} = \begin{cases} \hat{R}_{hks}^{ed} \cdot w_{hks_i} & , si \quad i \in \Omega_{hks}^{ec} \\ . & , si \quad i \in \Omega_{hks}^{ed} \end{cases} \quad (9)$$

Donde:

- $w_{hks_i}^{ed}$: Factor de expansión ajustado por elegibilidad desconocida de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- \hat{R}_{hks}^{ed} : Razón de ajuste por elegibilidad desconocida para las empresas pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- w_{hks_i} : Factor de expansión teórico de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- Ω_{hks}^{ed} : Conjunto de empresas en la muestra pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s (elegibilidad desconocida).
- Ω_{hks}^{ec} : Conjunto de empresas en la muestra pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s (elegibles y no elegibles).

IV.8.3.4. Ajuste por cambio de estrato y no elegibilidad

Frecuentemente en las encuestas a empresas se observa que la información que contienen las variables de estratificación informada en el marco muestral difiere del dato recolectado en el trabajo de campo, de modo que la empresa no pertenece al estrato en el cual fue seleccionada inicialmente.

Esto se debe, principalmente, al desfase existente entre la información contenida en el marco muestral y el trabajo de campo, lo cual se corrige a través del cálculo de los factores de expansión

realizando un ajuste por no elegibilidad y cambio de estrato¹⁷. Por lo tanto, para realizar este ajuste se identifican tres casos:

1. La empresa i es elegible, fue seleccionada en el estrato l , según la encuesta pertenece al estrato l' .
2. La empresa i es elegible y sigue perteneciendo al estrato en donde fue seleccionada.
3. La empresa i es no elegible.

Este ajuste se expresa en la siguiente fórmula:

$$w_{hks_i}^* = \begin{cases} w_{hks_i}^{ed} & , si \quad i \in \Omega_{hks}^{el} \wedge i \in \Omega_{hks}^{l \rightarrow l'} \\ \frac{N_{hks} \left(1 - \frac{n_{hks}^{ne}}{n_{hks}^{ec}} \right) - \sum_{l \neq l'} \sum_{i=1}^{n_{hks}^{el, l \rightarrow l'}} w_{hks_i}^{ed}}{n_{hks}^{el, l \rightarrow l}} & , si \quad i \in \Omega_{hks}^{el} \wedge i \in \Omega_{hks}^{l \rightarrow l} \\ . & , si \quad i \in \Omega_{hks}^{ne} \end{cases} \quad (10)$$

Donde:

- $w_{hks_i}^*$: Factor de expansión ajustado por cambio de estrato y no elegibilidad para la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de empresa k y agrupación de la actividad económica s .
- $w_{hks_i}^{ed}$: Factor de expansión ajustado por elegibilidad desconocida de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- N_{hks} : Número de empresas en el marco muestral, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- n_{hks}^{ne} : Número de empresas en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , categorizadas como no elegibles.
- n_{hks}^{ec} : Número de empresas en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s catalogadas con elegibilidad conocida (elegibles y no elegibles).
- $n_{hks}^{el, l \rightarrow l}$: Número de empresas en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , catalogadas como elegibles y que han sido seleccionadas en el estrato l y después del trabajo de campo se mantienen en el mismo estrato.
- Ω_{hks}^{el} : Conjunto de empresas en la muestra perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , catalogadas como elegibles.

¹⁷ Para la Encla 2019, se adopta el ajuste realizado por el Instituto Nacional de Estadísticas de España en el diseño muestral de las encuestas dirigidas a empresas. Para más información visitar: <https://www.ine.es/daco/daco43/metoite2016.pdf>

$\Omega_{hks}^{l \rightarrow l}$: Conjunto de empresas en la muestra que fueron seleccionadas en el estrato l y según la encuesta siguen perteneciendo al estrato l .
$n_{hks}^{el, l \rightarrow l'}$: Número de empresas en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , catalogadas como elegibles y que han sido seleccionadas en el estrato l y después del trabajo de campo pertenecen a otro estrato l' .
$\Omega_{hks}^{l \rightarrow l'}$: Conjunto de empresas en la muestra que fueron seleccionadas en el estrato l y según la encuesta pertenecen al estrato l' (cambió de estrato).
Ω_{hks}^{ne}	: Conjunto de empresas en la muestra perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , catalogadas como no elegibles.
.	: Indica valor perdido o <i>missing</i> .

IV.8.3.5. Ajuste por no respuesta

Existen algunas unidades muestrales que, aun siendo elegibles, no proporcionan la información solicitada o lo hacen de forma incompleta, lo que se denomina “no respuesta”. Esto tiene como consecuencia que, al momento de realizar los análisis con el factor de expansión definido hasta el paso anterior, no se estaría dando cuenta de la totalidad de las empresas elegibles de la población.

Tras finalizar el proceso de recolección de datos, se aplican metodologías para corregir o mitigar los efectos en los estimadores. Esto implica realizar un ajuste por no respuesta en los factores de expansión. Para el caso de la Encla, se optó por la estrategia de ajuste por celda, que consiste en distribuir los pesos de las unidades que no responden entre las unidades que sí responden de forma independiente al interior de cada estrato. Para realizar este ajuste, primero se calcula una razón de ajuste por estrato, correspondiente a la suma de los factores de expansión de las unidades elegibles sobre la suma de los factores de expansión de las unidades que respondieron la encuesta, tal como se observa en la siguiente fórmula:

$$\hat{R}_{hks}^{nr} = \frac{\sum_{i \in \Omega_{hks}^{el}} w_{hks_i}^*}{\sum_{i \in \Omega_{hks}^{el, R}} w_{hks_i}^*} \quad (11)$$

Donde:

\hat{R}_{hks}^{nr}	: Razón de ajuste por no respuesta para las empresas pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
$w_{hks_i}^*$: Factor de expansión ajustado por cambio de estrato y no elegibilidad para la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
Ω_{hks}^{el}	: Conjunto de empresas en la muestra pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , catalogadas como elegibles.

$\Omega_{hks}^{el,R}$: Conjunto de empresas en la muestra pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , catalogadas como elegibles y que respondieron la encuesta.

Posteriormente, para las unidades que respondieron la encuesta se multiplica el factor de expansión ajustado por cambio de estrato y no elegibilidad por la razón de ajuste por no respuesta, mientras que para el resto de las unidades se deja un valor perdido o *missing*.

$$w_{hks_i}^{nr} = \begin{cases} \hat{R}_{hks}^{nr} \cdot w_{hks_i}^* & , si \quad i \in \Omega_{hks}^{el,R} \\ . & , si \quad i \in \Omega_{hks}^{el,NR} \end{cases} \quad (12)$$

Donde:

$w_{hks_i}^{nr}$: Factor de expansión ajustado por no respuesta para la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

\hat{R}_{hks}^{nr} : Razón de ajuste por no respuesta para las empresas pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

$w_{hks_i}^*$: Factor de expansión ajustado por cambio de agrupación de actividad económica o agrupación regional y no elegibilidad para la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

$\Omega_{hks}^{el,R}$: Conjunto de empresas en la muestra pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , catalogadas como elegibles y que respondieron la encuesta.

$\Omega_{hks}^{el,NR}$: Conjunto de empresas en la muestra pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s , catalogadas como elegibles y que no respondieron la encuesta.

IV.8.3.6. Suavizamiento del factor de expansión

Después de realizar los ajustes descritos anteriormente, se evalúa la presencia de valores atípicos en los factores de expansión, con el fin de determinar la necesidad de realizar un suavizamiento.

Antes de realizar el suavizamiento, se debe establecer o definir algún criterio o mecanismo para identificar valores atípicos en los factores de expansión. Para ello, se establece como regla que, al interior de cada estrato, todo aquel factor de expansión que se encuentre a dos o más desviaciones estándar respecto al factor de expansión promedio del estrato es considerado atípico. Es decir, el factor de expansión $w_{hks_i}^{nr}$ se considera atípico si está a una distancia superior a dos desviaciones estándares respecto al promedio de su estrato. Si denotamos al factor atípico como A_{hks_i} , entonces A_{hks_i} no pertenece al intervalo I_{hks} , tal como se expresa en la siguiente fórmula:

$$A_{hks_i} \notin I_{hks} = [\bar{w}_{hks}^{nr} - 2 \cdot \sigma_{w_{hks}}^{nr}, \bar{w}_{hks}^{nr} + 2 \cdot \sigma_{w_{hks}}^{nr}] \quad (13)$$

Donde:

- A_{hks_i} : Factor de expansión de la empresa i en la muestra, considerado atípico, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- \bar{w}_{hks}^{nr} : Promedio de los factores de expansión ajustados por no respuesta del conjunto de empresas pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- $\sigma_{w_{hks}}^{nr}$: Desviación estándar de los factores de expansión del conjunto de empresas, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

Una vez identificados los estratos con presencia de valores atípicos, se suavizan todos los factores de expansión al interior de éstos, mientras que en aquellos estratos donde no hay presencia de valores atípicos, se mantiene el factor de expansión anterior. La fórmula para realizar el suavizamiento es la siguiente¹⁸:

$$w_{hks_i}^s = \begin{cases} \frac{(w_{hks_i}^{nr} + \bar{w}_{hks}^{nr})}{2} & , si \quad i \in \{ \Omega_{hks}^{el,R} \supset A_{hks_i} \} \\ w_{hks_i}^{nr} & , si \quad i \in \{ \Omega_{hks}^{el,R} \not\supset A_{hks_i} \} \end{cases} \quad (14)$$

Donde:

- $w_{hks_i}^s$: Factor de expansión suavizado de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- $w_{hks_i}^{nr}$: Factor de expansión ajustado por no respuesta de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- \bar{w}_{hks}^{nr} : Promedio de los factores de expansión ajustados por no respuesta del conjunto de empresas de la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- A_{hks_i} : Factor de expansión de la empresa i en la muestra, considerado atípico, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- $\Omega_{hks}^{el,R}$: Conjunto de empresas en la muestra que son categorizadas como elegibles y respondieron la encuesta, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

La fórmula de suavizamiento utilizada garantiza que el factor de expansión suavizado expanda a la misma población que el factor de expansión ajustado por no respuesta. Asimismo, garantiza mantener el orden de los factores de expansión previo y posterior al suavizamiento, es decir, si antes de realizar el suavizamiento $w_{hks_{i=1}}^{nr} > w_{hks_{i=2}}^{nr}$, una vez realizado el suavizamiento $w_{hks_{i=1}}^s > w_{hks_{i=2}}^s$.

¹⁸ La fórmula utilizada para realizar el suavizamiento del factor de expansión consiste en obtener un promedio simple entre el factor de expansión y el promedio de los factores de expansión de su estrato. Esta fórmula puede ser modificada dependiendo si se desea dar más ponderación al promedio o al valor del factor de expansión de la empresa. En este caso, se entrega igual ponderación al promedio y al factor de expansión.

IV.9. Medidas de precisión

IV.9.1. Varianza del estrato

Para obtener las medidas de precisión se debe estimar la varianza a nivel de estrato, tal como se especifica en la siguiente fórmula:

$$\hat{v}(\hat{t}_{hks}) = \sum_{i=1}^{n_{hks}} \hat{N}_{hks} \cdot (\hat{N}_{hks} - n_{hks}) \cdot \frac{S^2(y)_{hks}}{n_{hks}} \quad (16)$$

Donde:

- $\hat{v}(\hat{t}_{hks})$: Estimación del total de trabajadores contratados por las empresas directamente al 30 de noviembre del 2018, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- \hat{N}_{hks} : Estimación del número de empresas en la población de la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s .
- n_{hks} : Número de empresas en la muestra lograda, pertenecientes a la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s .
- $S^2(y)_{hks}$: Cuasi-varianza del número de trabajadores de la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s , en la muestra.

IV.9.2. Error relativo

Por otro lado, el error relativo de la estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre de 2018 en la Encla 2019, se obtiene a partir de la siguiente fórmula:

$$e(\hat{t}_{hks}^{A1}) = \frac{\sqrt{v(\hat{t}_{hks}^{A1})}}{\hat{t}_{hks}^{A1}} \cdot t_{gl, 1-\frac{\alpha}{2}} \quad (17)$$

Donde:

- $e(\hat{t}_{hks}^{A1})$: Error relativo de la estimación del total de trabajadores contratados por las empresas directamente al 30 de noviembre del 2018, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- \hat{t}_{hks}^{A1} : Estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2018, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

$v(\hat{t}_{hks}^{A1})^{19}$: Varianza de la estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2018, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

$t_{gl, 1-\frac{\alpha}{2}}$: Cuantil de la distribución T-Student correspondiente a gl grados de libertad y un nivel de confianza $(1 - \alpha)$.

IV.9.3. Tasa de elegibilidad

Tras el trabajo de campo, se clasifican las empresas según su estado de elegibilidad, diferenciando entre unidades “elegibles”, “no elegibles” o “elegibilidad desconocida”.

$$Tasa\ de\ respuesta = \frac{Empresas\ elegibles}{Elegibles + No\ elegibles + Elegibilidad\ desconocida} \times 100$$

IV.9.4. Tasa de respuesta por empresa

En base a los estados de disposición final de las empresas fue posible construir los principales indicadores operativos, entre los que destaca la tasa de respuesta por empresa.

La tasa de respuesta es la proporción de encuestas terminadas sobre el total de casos elegibles y de elegibilidad desconocida. La fórmula es la siguiente:

$$Tasa\ de\ respuesta = \frac{Encuestas\ terminadas}{Terminadas + Incompletas + No\ contactadas + Rechazos + Inubicables} \times 100$$

IV.9.5. Tasa de contacto

La tasa de contacto es la proporción de casos de la muestra en los que el encuestador logra algún grado de contacto (hablar o captar la atención del entrevistado), independiente del resultado final de la encuesta. Se calcula a través de la siguiente ecuación:

$$Tasa\ de\ contacto = \frac{Terminadas + Incompletas + Rechazos}{Terminadas + Incompletas + No\ contactadas + Rechazos + Inubicables} \times 100$$

IV.9.6. Tasa de rechazo

Es la proporción de unidades que rechazan participar en el estudio de forma definitiva. Se calcula sobre el total de casos elegibles y con elegibilidad desconocida, como lo señala la siguiente ecuación:

¹⁹ Los cálculos realizados para la varianza de la estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2018 fueron realizados en software SAS Guide.

$$Tasa\ de\ rechazo = \frac{Rechazadas}{Terminadas + Incompletas + No\ contactadas + Rechazos + Inubicables} \times 100$$

IV.9.7. Tasa de cooperación

Este indicador indica la proporción de unidades muestrales exitosamente entrevistadas en base a las unidades elegibles contactadas. Se calcula con la siguiente fórmula:

$$Tasa\ de\ cooperación = \frac{Terminadas}{Terminadas + Incompleta + Rechazos}$$

IV.9.8. Tasa de logro objetivo

Se refiere a la proporción de encuestas terminadas en relación con la muestra objetivo. La fórmula se detalla a continuación:

$$Tasa\ de\ logro\ objetivo = \frac{Terminadas}{Muestra\ objetivo}$$

IV.10. Estimación de parámetros

La estimación estadística corresponde al procedimiento mediante el cual se calculan parámetros de la población a partir de datos muestrales obtenidos. De este modo, para la Encla 2019 se realiza la estimación del total del número de trabajadores, variable que se extrae del formulario autoaplicado, específicamente de la pregunta A1: “Indique el total de personas contratadas directamente por esta empresa” al 30 de noviembre de 2018, según región del país y sexo. Esta estimación se calcula para cada nivel de dominio del estudio para exponer la fórmula de cálculo se expresará en el nivel más desagregado de variables.

La estimación se obtiene mediante la siguiente fórmula:

$$\hat{t}_{hks} = \sum_{i=1}^{n_{hks}} y_{hks_i}^{A1} \cdot w_{hks_i}^S \quad (15)$$

Donde:

\hat{t}_{hks} : Estimación del total de trabajadores contratados por las empresas directamente al 30 de noviembre del 2018, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

- $y_{hks_i}^{A1}$: Número de trabajadores contratados directamente por la empresa i al 30 de noviembre del 2018, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- $w_{hks_i}^s$: Factor de expansión suavizado de la empresa i en la muestra, perteneciente a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .

$$\hat{v}(\hat{t}_{hks}) = \sum_{i=1}^{n_{hks}} \hat{N}_{hks} \cdot (\hat{N}_{hks} - n_{hks}) \cdot \frac{S^2(y)_{hks}}{n_{hks}} \quad (16)$$

Donde:

- $\hat{v}(\hat{t}_{hks})$: Estimación del total de trabajadores contratados por las empresas directamente al 30 de noviembre del 2018, pertenecientes a la agrupación regional h , tamaño de la empresa k y agrupación de actividad económica s .
- \hat{N}_{hks} : Estimación del número de empresas en la población de la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s .
- n_{hks} : Número de empresas en la muestra lograda, pertenecientes a la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s .
- $S^2(y)_{hks}$: Cuasi-varianza del número de trabajadores de la agrupación regional h , tamaño de empresa k y agrupación de actividad económica s , en la muestra.

V. METODOLOGÍA DE RECOLECCIÓN DE DATOS

V.1. Verificación de la muestra

Tras el proceso de selección de la muestra, y con el objetivo de asegurarse de que las unidades consideradas se encontraban activas y cumplían con los requisitos para ser incluidas dentro de la población objetivo del estudio, se realizó la etapa de verificación. Esta fue ejecutada por la Unidad de Verificación perteneciente al Subdepartamento de Marcos Maestros del INE. Las variables consultadas fueron: identificación (RUT, razón social, nombre de fantasía, representante legal), medios de contacto (fonos, correos, páginas web y datos del informante), ubicación geográfica y actividad económica (siguiendo los lineamientos del Clasificador CIIU Rev.4.CL v2012).

La verificación es una etapa fundamental para el funcionamiento eficiente de los proyectos, pues permite actualizar el estado y características de las empresas a fin de disminuir problemas de contacto. Ejemplo, una empresa acepta la llamada del INE, pero el informante indica que el RUT consultado se encuentra en situación de término de giro, sin movimiento o paralizada, por lo que la unidad debe ser reemplazada de la muestra seleccionada.

Para el proyecto Encla, se determinó una muestra objetivo de 4.068 empresas, llegando a una muestra de 5.800 considerando sobremuestreo, lo que implicó que en la etapa de verificación se contactara a 15.500 unidades.

V.2. Técnica y métodos de recolección de datos

V.2.1. Estrategia de recolección de datos

V.2.1.1. Contacto inicial

La fase de recolección de datos comenzó con las funciones de gestión del directorio y el envío de una carta certificada, cuyo propósito era notificar a la empresa que había sido seleccionada para formar parte del estudio. Tras esto, el analista inició el contacto con la empresa vía correo electrónico y/o teléfono, describiendo el estudio, identificando a los informantes responsables de cada formulario y agendando una cita con ellos, para que, posteriormente, el encuestador aplicara la encuesta de manera presencial. Si bien los analistas lograron el objetivo, se detectó que era más eficiente que el encuestador “abriera” las empresas de forma presencial o telefónica, sin previa apertura de los analistas. Lo anterior, considerando que esta información fuera comunicada a los encargados de grupo o el analista a cargo de la unidad, para evitar cruces en la gestión de recolección de la empresa. Esto hizo que el proceso de recolección fuera más fluido.

V.2.1.2. Distribución de la carga de trabajo

La etapa de coordinación de análisis operativo inició con la recolección, mediante la administración del directorio y las unidades a encuestar. Estas unidades fueron asignadas a las personas encargadas de grupo, quienes, a su vez, debían distribuir las homogéneamente entre los analistas, utilizando el Sistema de Gestión de Operaciones (SGO) dispuesto para ello.

Para el caso de los cuestionarios presenciales, tras el envío se inició la asignación de la muestra correspondiente a cada una de las regiones en el SGO. Los encargados regionales fueron los responsables de asignar la carga a los recolectores de sus respectivas regiones. Una vez que estos últimos contaban con su carga, visitaron a las empresas y administraron el formulario correspondiente, consignando todos los datos requeridos tanto en el cuestionario como en la hoja de ruta.

En términos iniciales, la distribución se planificó por zonas censales, para lo que se solicitó a la Unidad de Cartografía del INE la entrega de los antecedentes de distrito-zona y manzana, a fin de distribuir de manera más eficiente y ordenada la carga de trabajo entre el personal a cargo de la recolección.

La carga fue asignada de acuerdo la concentración de la muestra. En la primera fase de la recolección se abordó el 72% del total de la muestra y para la segunda fase se gestionaron el resto de las unidades.

A medidas que avanzaba el trabajo de campo, la distribución de carga se fue organizando en función de las unidades pendientes.

V.2.1.3. Gestión del cuestionario autoaplicado

El cuestionario autoaplicado, a diferencia de los otros tres cuestionarios que son de carácter presencial, fue respondido y gestionado de manera autónoma por un representante de la empresa. Este representante pudo recurrir a diferentes informantes al interior de la misma empresa si así lo creyó pertinente, debido a la extensión y diversidad de temas que abarca este cuestionario. En tanto, no debía enviar una respuesta inmediata.

Luego de que el analista contactara al informante a cargo de responder el cuestionario autoaplicado, el primero debía explicar el objetivo de la encuesta, el tipo de información que esta solicita, los plazos para enviar el formulario completo al INE y el proceso de seguimiento que se llevaría a cabo.

Además, el analista debía acordar con el informante el modo de aplicación de formulario, que podía ser a través de un formulario *web* o por papel²⁰ (según las preferencias del informante), puesto que de consensuar lo segundo, el formulario físico debía ser entregado en la empresa por un encuestador²¹. De lo contrario, el analista debía hacer envío del *link* a través del sistema informático dispuesto para ello.

Una vez notificada la recepción del cuestionario, el analista debía tomar contacto con el informante dentro de los siguientes diez días hábiles. En esta instancia el analista generaba un lazo con el informante y ofrecía asistencia para completar la información.

Una vez que el informante concluía el llenado del formulario electrónico, el analista podía acceder inmediatamente al mismo para proceder a la revisión de los datos ingresados. Cuando la respuesta del formulario era a través de papel, su retiro era coordinado con el encuestador a cargo y se

²⁰ La aplicación de una encuesta a través de un cuestionario digital distribuido a través de la *web* es conocida como encuesta CAWI (*Computer Assisted Web Interviewing*) y tiene la ventaja de poder incorporar los validadores de primer nivel durante la fase de recolección. Por otro lado, las encuestas aplicadas en papel se conocen como PAPI (*Paper Assisted Personal Interviewing*) y una de sus ventajas es que el recolector puede aclarar conceptos al informante.

²¹ En la aplicación de la Encla 2019, solo se recibieron 84 cuestionarios autoaplicados en papel. No se registró información sobre el tiempo de aplicación de este formulario, debido a que al ser autoadministrado y a su extensión, generalmente no es respondido “de una vez”.

gestionaba su envío al nivel central por medio de la unidad de enlace regional, para su digitación y posterior procesamiento.

V.2.1.4. Cuestionarios presenciales

Los cuestionarios a empleadores, sindicatos y trabajadores son estructurados y fueron aplicados por un encuestador mediante el uso de un formulario impreso en papel (PAPI). En paralelo a la recolección del cuestionario autoaplicado, el analista debía conseguir los contactos de los informantes de los cuestionarios presenciales y coordinar citas para su recolección, verificando la siguiente información:

- Existencia de sindicatos.
- Potenciales informantes para los correspondientes formularios.

El cuestionario a empleadores se dirigió al encargado o encargada de las relaciones laborales al interior de cada empresa. Ahora bien, esta función no necesariamente se define en un cargo en específico y, por otro lado, dependiendo del tamaño y complejidad de la empresa, el o los responsables de las relaciones laborales pueden ocupar un cargo diferente en cada unidad. Es por esto que se definió como informante idóneo a aquella persona que ocupe alguno de los siguientes cargos (en orden de prioridad):

1. Gerente de recursos humanos
2. Sub-gerente de recursos humanos
3. Jefe de personal
4. Encargado de personal
5. Dueño de la empresa

El criterio para utilizar el cuestionario a sindicatos o a trabajadores en una determinada empresa fue la existencia de sindicatos en ella: si en la empresa existía uno o más sindicatos al momento de realizar la encuesta, se aplicó el cuestionario a sindicatos a aquel con el mayor número de trabajadores (“el sindicato más representativo”). Por el contrario, si en la empresa no existían sindicatos, se aplicó el cuestionario a un representante del personal de la empresa.

Lo anterior implica que el cuestionario prioritario fue el de sindicatos, en tanto se supone que son estas organizaciones las que tienen la prioridad de la representación de los trabajadores y las trabajadoras al interior de las empresas. En el caso de que existieran tanto sindicatos de empresa, establecimiento o de trabajadores eventuales o transitorios, como sindicatos interempresa en una misma empresa, se priorizó el contacto con el sindicato de empresa. Luego, en el caso de que no existiera ninguno de estos dos tipos de sindicatos se procedió a encuestar a un representante de los trabajadores, utilizando para ello el cuestionario diseñado para trabajadores.

El cuestionario a sindicatos se dirigió al presidente del sindicato más representativo existente en la empresa. Ahora bien, si el presidente no se encontraba disponible, el encuestador o la encuestadora utilizaron el siguiente orden de prioridad:

1. Presidente del sindicato

2. Vicepresidente del sindicato
3. Secretario del sindicato
4. Tesorero del sindicato
5. Otro miembro de la directiva sindical
6. Otro miembro del sindicato

El cuestionario a trabajadores se dirigió, preferentemente, a trabajadores o trabajadoras que tuviesen algún cargo de representación en la empresa. Para este cuestionario se definió como informante idóneo a aquel trabajador o trabajadora que ocupe alguno de los siguientes cargos (en orden de prioridad):

1. Representante de los trabajadores ante el comité paritario de higiene y seguridad
2. Representante de los trabajadores ante el comité bipartito de capacitación
3. Trabajador elegido utilizando una tabla de números aleatorios

La tabla de números aleatorios funcionó de la siguiente manera: los encuestadores solicitaron a la empresa el listado de personas contratadas directamente por la misma, excluyendo a quienes habían sido contratados hace menos de un año, se encontraban haciendo uso de su feriado legal (vacaciones) o con licencia médica y/o eran jefaturas de primer y segundo nivel. Desde esta lista el encuestador seleccionó al informante en base a un número aleatorio proporcionado por el Subdepartamento de Diseño de Marcos y Muestras del INE. Para ajustar el número aleatorio al número de trabajadores del listado entregado por la empresa, el encuestador tuvo que realizar la siguiente operación matemática:

$$X_j = [A_j \times (Y_j - 1)] + 1$$

Donde:

X_j : Valor aleatorio ajustado con la información recogida en el trabajo de campo.

A_j : Número aleatorio generado para la j-ésima empresa.

Y_j : Total de trabajadores de la j-ésima empresa informado durante el trabajo de campo.

En el caso de que el trabajador seleccionado no se encontrara, el encuestador contaba con otros dos números aleatorios con los que podía desarrollar la misma operación.

Para los cuestionarios presenciales, en la Región Metropolitana (RM) el analista debía notificar al encargado de grupo las citas agendadas, de modo que el encuestador recibiera por sistema su carga de trabajo. Para las demás regiones, los encuestadores recibieron directamente un *stock* de cuestionarios presenciales (cuestionarios empleadores y cuestionarios trabajadores o sindicatos), los cuales fueron gestionando a medida que el encargado de grupo asigna empresas a través del Sistema.

El encuestador recolectó los cuestionarios presenciales en las fechas agendadas y en la visita a las dependencias de la empresa corroboró los datos de esta, registrándolos en la hoja de ruta de la visita. En este último instrumento se registran todas las visitas que el encuestador hizo a la empresa, así como también el término de la gestión con la unidad. Luego de aplicadas las encuestas, los recolectores de datos las entregaron al encargado regional o coordinador de terreno según correspondiese, quien realizó el envío de forma física al coordinador de enlace, ubicado en la Región Metropolitana.

Las empresas fueron consideradas como logradas cuando los tres cuestionarios fueron respondidos en su totalidad.

V.2.2. Digitación

La digitación tiene por objetivo realizar una transcripción de la información recogida durante la ejecución de la recolección. Una vez que el coordinador de digitación recibía los formularios desde la unidad de enlace, procedía a asignar por sistema la carga diaria a cada digitador, quienes debían ingresar los datos en la plataforma informática diseñada para ello. Durante la digitación, el sistema indicaba la presencia de algún error en alguna validación de primer nivel, producida tanto por una equivocación al ingresar algún código en el sistema o, en caso contrario, por el registro erróneo por parte del encuestador. Ese es el único error frente al cual el digitador puede hacer alguna modificación en el sistema, en ningún caso el digitador corrige alguna inconsistencia u otra clase de error distinto a los que pudiese cometer el mismo al ingresar los dígitos, ya que estos eran controlados mediante los sistemas de validación.

Luego de digitar la información, se realiza la recolección de los formularios por parte del coordinador de digitación, para posteriormente enviarlos a los encargados de grupo del área de análisis, de forma física y virtual, a través del sistema informático SGO.

La unidad de análisis recibió todos los cuestionarios con la información recolectada en terreno y digitada por los encargados de digitación, independiente de su estado de logro. En esta etapa, el analista implementó los últimos controles de calidad de la información propios del proyecto.

En primer lugar, el analista encargado debía velar porque la información ingresada por el digitador fuera exactamente la misma que la registrada en el cuestionario en papel. En caso contrario, tenía la facultad de realizar modificaciones en el sistema para corregir la información correspondiente. Adicionalmente, debía verificar que la encuesta en papel estuviera correctamente llenada, respetando el flujo (saltos de preguntas filtro) y la completitud de la información.

Una vez que la revisión estaba completa, los analistas accedían al reporte de errores que generaba el sistema luego de la ejecución de la validación llamada *Batch*²² y, con ese insumo, aplicaban los procedimientos de corrección para los errores detectados directamente en el sistema, dejando respaldo de la modificación en el cuestionario en papel. Para los errores que no tenían procedimiento asociado, el analista procedía a contactar vía telefónica al informante de la encuesta para recuperar la información faltante o aquellos datos que generaron inconsistencia.

²² Malla de validación que incluye todos los validadores programados.

Cuando todos los hallazgos fueron solucionados, el formulario se aprobó como correcto y se devolvió al encargado de grupo. Esto permitía a los encargados de grupo realizar supervisiones al trabajo efectuado por los analistas. Si, por el contrario, las inconsistencias no eran solucionadas debido a la imposibilidad de contactar al informante, el formulario continuaba en posesión del analista hasta rescatar los antecedentes necesarios.

Luego de que las encuestas se validaban por el sistema informático, estas eran almacenadas de acuerdo con un criterio predeterminado de orden asignado por el módulo del sistema informático de control, quedando disponibles en caso de que fuese necesaria alguna consulta posterior.

V.2.3. Período de recolección de datos

El período de recolección de Encla 2019 inició la segunda semana de mayo y finalizó la última semana de septiembre de 2019²³.

V.2.4. Período de referencia

El período de referencia de esta versión de la Encla 2019 corresponde al año 2018.

V.2.5. Periodicidad

El proyecto tiene una periodicidad bienal o trienal, según la definición del organismo mandante.

V.2.6. Conformación de los equipos de trabajo

En cuanto a la división del trabajo planificada para el equipo encargado de la recolección, este estaba compuesto de los siguientes cargos:

- Analistas: responsables de contactar a los informantes, iniciar la comunicación con las unidades y agendar entrevistas. También son los encargados de verificar la consistencia y completitud de los datos recolectados, realizando ediciones de acuerdo con los protocolos y procedimientos definidos, en caso de que ello correspondiera. Adicionalmente, los analistas monitorean la recepción de cuestionario autoaplicado y asisten a los informantes en sus dudas y consultas.

Debido a las múltiples funciones desarrolladas por los analistas, durante el proceso de recolección se procedió a dividir el equipo en dos funciones principales: un grupo se enfocó en la verificación de consistencia y completitud de los cuestionarios presenciales, y otro en el seguimiento y asistencia del cuestionario autoaplicado. El agendamiento de entrevistas fue asignado a los recolectores.

- Recolectores: los recolectores aplican los cuestionarios presenciales en terreno, siguiendo los procedimientos, plazos y estándares de calidad definidos por el INE, explicados en las capacitaciones. Además, en el caso de que algún informante solicite responder el

²³ El período de recuperación de datos se extendió hasta noviembre de 2019.

cuestionario autoaplicado en papel, los recolectores se hacen cargo de entregárselo al informante y, posteriormente, retirarlo para llevarlo de regreso al INE.

En la práctica, y a fin de agilizar el proceso de recolección, se determinó que era más eficiente que el encuestador “abriera” las empresas de forma presencial o telefónica, sin previa apertura de los analistas. Lo anterior, considerando que esta información fuera comunicada a los encargados de grupo o al analista a cargo de la unidad, para evitar cruces en la gestión de recolección de la empresa. Esto hizo que el proceso de recolección fuera más fluido.

- Encargados de grupo: responsables de planificar y distribuir de forma equitativa la carga de trabajo, haciendo seguimiento y supervisando el proceso de recolección y análisis de la información.
- Coordinador de trabajo de campo: persona encargada de la organización, monitoreo, acompañamiento de los equipos de terreno.
- Coordinador de digitación: persona encargada de la organización, monitoreo, acompañamiento del personal a cargo de la digitación de encuestas.
- Coordinador de análisis: persona encargada de la organización, monitoreo, acompañamiento del personal a cargo de la gestión del cuestionario autoaplicado y del análisis de los cuestionarios presenciales.
- Supervisor de gabinete: personas encargadas de organizar, monitorear y acompañar el trabajo de los recolectores. En específico, son los encargados de recepcionar las encuestas ya gestionadas y monitorear el avance de los recolectores y digitadores.
- Digitadores: personal responsable de la digitación de los cuestionarios presenciales.

El diseño original de trabajo de campo de este estudio contempló 39 recolectores de datos a nivel nacional. Sin embargo, durante las primeras dos semanas se identificó un avance menor a lo proyectado inicialmente, lo que fue subsanado con la incorporación de personal extra para el proceso de recolección.

Tabla 13. Dotación recolectores de datos por región

Región	Recolectores
Tarapacá	2
Antofagasta	3
Atacama	2
Coquimbo	4
Valparaíso	2
O'Higgins	1
Maule	2
Biobío	4
Araucanía	2
Los Lagos	3
Aysén	2
Magallanes	3
Metropolitana	42
Los Ríos	1
Arica	1
Ñuble	1
Total	76

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Tabla 14. Dotación personal encargado del trabajo operativo

Función	Dotación
Coordinador de terreno	1
Coordinador de digitación	1
Coordinador de análisis	1
Supervisores de gabinete	4
Digitadores	4
Analistas	21
Encargados de grupo	3

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

V.3. Diseño del instrumento de captura de la información

La Encla 2019 se compone de cuatro cuestionarios estructurados: empleadores, autoaplicado, sindicatos y trabajadores. En cada una de las empresas seleccionadas debían ser aplicados tres de los cuatro cuestionarios mencionados anteriormente para que la unidad fuese considerada como lograda.

Los cuestionarios a empleadores, sindicatos y trabajadores son en formato PAPI, mientras que el cuestionario autoaplicado es formato CAWI.

Las variables recolectadas en la Encla 2019 abordan diversas temáticas relacionadas al mundo del trabajo y la empresa que se desprenden de los objetivos y dimensiones de la encuesta. En la Tabla 15 se pueden observar los siete objetivos específicos de la encuesta, su desglose en 21 dimensiones y los módulos relacionados por cuestionario.

Es preciso señalar que la Encla 2019 corresponde a la novena versión del estudio, por lo que el diseño grueso de los cuestionarios fue realizado por la Dirección del Trabajo.

Tabla 15. Objetivos, dimensiones y módulos según cuestionario de la Encla 2019

Objetivo específico	Dimensión	Módulo
Caracterizar las empresas en relación con su entorno económico.	Internacionalización de la empresa	Módulo A Autoaplicado
	Evaluación general de la empresa	Módulo A Empleadores
Caracterizar las condiciones de empleo en las empresas.	Trabajadores externos (subcontrato y suministro)	Módulo A Sindicatos y Trabajadores, Módulo C Autoaplicado
	Trabajadores y contratos temporales	Módulo A Sindicatos y Trabajadores, Módulo B y C Autoaplicado
	Otros contratos en la empresa	Módulo B Autoaplicado
	Remuneraciones e incentivos	Módulo B Sindicatos y Trabajadores, Módulo D Autoaplicado
	Capacitaciones	Módulo E Sindicatos y Trabajadores, Módulo I Autoaplicado
Caracterizar las condiciones de trabajo en las empresas.	Organización del tiempo de trabajo	Módulo F autoaplicado, Módulo C Sindicatos, Empleadores y Trabajadores
	Seguridad y salud	Módulo G Autoaplicado, Módulo D Sindicatos, Empleadores y Trabajadores
	Condiciones ambientales y de trabajo	Módulo D Sindicatos, Empleadores y Trabajadores
Analizar las relaciones laborales que se dan al interior de las empresas.	Niveles de autoridad en la empresa	Módulo A Autoaplicado
	Información / Participación / Consulta / Gestión inclusiva	Módulo F Sindicatos, Trabajadores y Empleadores
	Acción sindical	Módulo G Sindicatos
	Incidencia de conflictos colectivos	Módulo H Sindicatos y Trabajadores, Módulo G Empleadores
	Negociación colectiva	Módulo I Sindicatos y Trabajadores, Módulo H empleadores y Módulo J Autoaplicado
Objetivo específico	Dimensión	Módulo
Analizar la relación entre condiciones de empleo y trabajo y las relaciones laborales.	Mismas dimensiones de objetivos 2.1. y 2.2.	Mismos módulos de objetivos 2.1. Y 2.2.
Identificar la situación de las empresas en materia de igualdad de género.	Igualdad de género al interior de la empresa	Módulo A, B, C, D, E y F Autoaplicado
	Protección a la maternidad y paternidad	Módulo H Autoaplicado
	Igualdad de género en el mundo sindical	Módulo G Sindicatos
Describir fenómenos laborales emergentes.	Trabajadores extranjeros	Módulo A Sindicatos y trabajadores, Módulo E Autoaplicado
	Automatización de la producción	Módulo B Autoaplicado
	Reforma laboral 2017 (equipos de emergencia y servicios mínimos, acuerdos grupo negociador, entre otras)	Módulo I Sindicatos y Trabajadores, Módulo H empleadores y Módulo J Autoaplicado

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Los cuestionarios de la encuesta tienen en total 349 preguntas: 77 corresponden al cuestionario de trabajadores, 119 a sindicatos, 67 a empleadores y 86 al autoaplicado. En la tabla a continuación se presenta el número de variables, según módulo y cuestionario²⁴.

Tabla 16. Variables contenidas en los módulos de cada cuestionario de la Encla 2019

Módulo	Variables por módulo y cuestionario			
	Autoaplicado	Empleadores	Sindicatos	Trabajadores
Identificación	-	14	16	-
Identificación e introducción	24	-	-	15
A. Externalización, temporalidad y extranjerización	-	-	7	7
A. Evaluación general de la empresa	-	27	-	-
A. Aspectos generales	71	-	-	-
B. Remuneraciones e incentivos	-	11	4	2
B. Contratos	120	-	-	-
C. Organización del tiempo de trabajo	-	11	14	14
C. Externalización y temporalidad	25	-	-	-
D. Seguridad y salud	-	30	110	110
D. Remuneraciones	165	-	-	-
E. Capacitación	-	29	16	16
E. Trabajadores extranjeros	91	-	-	-
F. Información / Participación / Consulta	-	-	37	35
F. Relaciones laborales y gestión inclusiva	-	37	-	-
F. Organización del tiempo de trabajo	47	-	-	-
G. Acción sindical y relaciones laborales	-	30	114	27
G. Incidencia de conflictos	-	-	-	-
G. Salud y seguridad	24	-	-	-
H. Incidencia de conflictos	-	-	33	30
H. Negociación colectiva	-	20	-	-
H. Protección a la maternidad y paternidad	66	-	-	-
I. Negociación colectiva	-	-	47	25
I. Capacitación	13	-	-	-
J. Negociación colectiva	18	-	-	-
R. Antecedentes personales del informante	-	5	13	7
Total	664	214	411	288

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

²⁴ No todas las variables contabilizadas están incluidas en las bases de datos, debido a que algunas de ellas contienen información confidencial de los informantes, otras corresponden a glosas que fueron codificadas y un grupo solo tenía relevancia en tanto permitía construir otras variables.

V.4. Prueba y evaluación del instrumento de captura de información

V.4.1. Mejoras implementadas a los cuestionarios

Los cuestionarios de la Encla 2019, desde la prueba piloto hasta la recolección nacional, experimentaron una serie de modificaciones que se describen a continuación²⁵.

Es relevante mencionar la magnitud de estas modificaciones: más de la mitad de las preguntas de cada cuestionario se vieron afectadas. En específico, las mejoras fueron implementadas al 55% de las preguntas del cuestionario a sindicatos, al 72% del cuestionario a trabajadores, al 74% del cuestionario autoaplicado y al 30% del cuestionario a empleadores. Si bien la mayoría de estas modificaciones fueron menores (cambios pequeños en las alternativas y en los fraseos para una mejor comprensión e incorporación de lenguaje inclusivo, entre otras razones), alrededor de un cuarto de las preguntas de cada cuestionario fueron modificadas sustantivamente o son nuevas preguntas: 16% en sindicatos, 25% en trabajadores, 22% en autoaplicado y 27% en empleadores.

Tabla 17. Nivel de modificación de las preguntas de los cuestionarios después de las distintas pruebas y evaluaciones de expertos

Nivel de modificación	Cuestionario sindicatos	Cuestionario trabajadores	Cuestionario autoaplicado	Cuestionario empleadores
Sin modificación	45%	28%	26%	30%
Modificación menor ²⁶	39%	47%	52%	49%
Modificación sustantiva ²⁷	10%	13%	15%	18%
Nueva pregunta	6%	12%	7%	9%
Total	100%	100%	100%	100%

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

V.4.1.1.1. Cambios asociados a la prueba cognitiva

La prueba cognitiva de los cuestionarios de la Encla 2019 fue una de las principales estrategias utilizadas para asegurar la calidad de los instrumentos de recolección de los datos y no había sido utilizada hasta ahora en la encuesta.

El método de las pruebas cognitivas, desarrollado en la década de 1980, se basa en marcos teóricos fundamentados en la psicología social y cognitiva, y busca establecer el proceso cognitivo que los informantes utilizan para interpretar y responder las preguntas de una determinada encuesta. Desde esta perspectiva, el proceso de responder preguntas, ya sea en una encuesta o en cualquier otra situación, puede ser entendido a través de cuatro etapas: comprensión, memoria, estimación y respuesta (Tourangeau, 1984). En cada una de estas etapas, pueden suscitarse diversos tipos de

²⁵ Estas modificaciones se realizaron con el acompañamiento y asistencia técnica de la Organización Internacional del Trabajo, por medio de sus expertos en estadísticas del trabajo.

²⁶ Una modificación menor corresponde a agregar instrucciones, ejemplos o modificar el enunciado para permitir su comprensión por parte del informante.

²⁷ Una modificación sustantiva corresponde a un cambio en las alternativas de respuesta y/o en los objetivos de la pregunta.

errores. Las entrevistas cognitivas intentan evaluar la solidez de las preguntas de la encuesta, examinando su desempeño y rendimiento frente al informante para cada una de las etapas mencionadas.

En general, el objetivo de este método es incentivar a las personas a compartir información que provea de pistas sobre el proceso cognitivo utilizado en la respuesta, lo que sirve como insumo para identificar algunos problemas presentes en el diseño de cuestionarios que eventualmente podrían haberse convertido en errores de respuesta.

En el Tabla 18 se puede observar el porcentaje de preguntas de cada uno de los cuestionarios que fueron modificadas a raíz de las pruebas cognitivas, tras acordarse estos cambios en reuniones con la Dirección del Trabajo (marzo de 2019).


Tabla 18. Porcentaje de preguntas modificadas según cuestionario

Cuestionario autoaplicado	Cuestionario empleadores	Cuestionario sindicatos	Cuestionario trabajadores
7%	6%	14%	22%

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Además, como resultado de la prueba cognitiva, se incorporó al inicio de la mayoría de los módulos (cuando correspondía) una “advertencia”, señalando que las preguntas de tal módulo refieren exclusivamente al personal contratado directamente por la empresa. Por ejemplo:

Ilustración 1. Encabezado módulos cuestionarios Encla 2019



ATENCIÓN: Para las preguntas del Módulo B considere únicamente al personal contratado directamente por la empresa, es decir, excluya al personal subcontratado o suministrado por terceros (empresas de servicios transitorios). Excluya también al personal a honorarios.

V.4.1.1.2. Cambios asociados al taller de expertos

Durante el mes de marzo de 2019, la Dirección del Trabajo desarrolló un taller con expertos en temáticas de migraciones, quienes revisaron las distintas preguntas de los cuestionarios de la Encla 2019 e hicieron sugerencias y recomendaciones. Si bien las preguntas agregadas o modificadas fueron pocas (ver Tabla 19), las recomendaciones de los expertos tuvieron un rol central para mejorar la calidad de los cuestionarios.

Tabla 19. Porcentaje de las preguntas modificadas por cuestionario

Cuestionario autoaplicado	Cuestionario empleadores	Cuestionario sindicatos	Cuestionario trabajadores
0%	6%	3%	5%

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

A continuación, se detallan las preguntas agregadas o modificadas y los cuestionarios a las que pertenecen.

- Preguntas A6 y A7, cuestionario empleadores: expertos en migraciones señalaron a la Dirección del Trabajo que la pregunta A6 incluía alternativas que no eran exhaustivas. Solamente se señalaban como opciones: “Porque son más productivos”, “Porque son menos dados a reclamar”, “Porque están dispuestos a trabajar por menos dinero”, “Porque tienen más iniciativa” y “Otro”. El comité de expertos sugirió agregar cuatro motivos más: “Porque cumplían con los requisitos del cargo”, “Porque hay pocos trabajadores nacionales con las cualificaciones o competencias requeridas para el cargo”, “Porque tienen un mayor dominio de habilidades socioemocionales (mejor trato, mejor servicio al cliente, responsabilidad, compromiso, etc.)” y “La empresa tiene la política de ayudarlas a integrarse a la sociedad chilena”.

Debido a estos cambios en la pregunta A6, la pregunta A7 también tuvo que ser modificada, pues incluía alternativas de la A6 y se confundía con sus objetivos. La pregunta A7 pasó de tener el objetivo de conocer “los motivos para los cuales se contrataron extranjeros” a conocer “las situaciones en las cuales se contrataron extranjeros”.

- Pregunta A8, cuestionario empleadores: la pregunta A8 es una nueva pregunta sugerida por el comité de expertos. En esta, se busca capturar la opinión de los informantes del cuestionario a empleadores sobre los artículos 19 y 20 del Código del Trabajo que no les permiten a las empresas con 25 o más trabajadores exceder el 15% de trabajadores extranjeros en su planilla de trabajadores.
- Preguntas A6 y A7, cuestionarios trabajadores y sindicatos: se cambia “de otra nacionalidad” por “personas extranjeras”. Se concordó, en reunión con la Dirección del Trabajo, definir el concepto de “extranjero” o “nacionalidad extranjera” en el manual del encuestador (básicamente como personas que no tienen la nacionalidad chilena, la cual se puede obtener de cuatro maneras según la legislación chilena²⁸), para que el encuestador estuviese capacitado para responder las dudas de los informantes sobre cuándo considerar o no a un trabajador como extranjero.
- Pregunta D20, cuestionarios sindicatos y trabajadores: la pregunta D20 también fue sugerida por el comité de expertos. Con esta, se busca conocer cómo ha impactado, según los trabajadores y sus representantes, la llegada de personas extranjeras a las empresas que operan en Chile.

El comité sugirió hacer una pregunta sobre capacitaciones específica para migrantes. En una posterior reunión entre el equipo Encla y la Dirección del Trabajo se decidió no agregar dicha pregunta, puesto que excedía los objetivos de la encuesta.

Además, hubo un segundo taller de expertos convocado por la Dirección del Trabajo, en el que participaron representantes de las áreas de estudios de otros servicios públicos, organismos

²⁸ Por territorialidad, por consanguineidad, por carta (residieron por más de cinco años de forma continua en el país y tienen residencia permanente) o por gracia de la Ley (otorgada por el Congreso a quienes han realizado grandes servicios a la república).

internacionales e instituciones privadas relacionadas al mundo del trabajo: Sence, Observatorio Laboral, Observatorio Social del Ministerio de Desarrollo Social y la Familia, Organización Internacional del Trabajo (OIT), Organización Internacional para las Migraciones (OIM), Departamento de Estudios Sofofa; que también hicieron determinadas recomendaciones.

V.4.1.1.3. Cambios asociados al uso de lenguaje inclusivo de género

En base a la revisión de los cuestionarios que hizo la Sección de Estadísticas de Género del Departamento de Estudios Sociales del INE, se le propuso a la Dirección del Trabajo realizar cambios en los enunciados y alternativas de las preguntas de los cuatro cuestionarios en la línea del lenguaje inclusivo de género.

Si bien a la Dirección del Trabajo solo se le presentaron ejemplos del cuestionario autoaplicado y empleadores, debido a restricciones de tiempo, las propuestas de cambio eran genéricas a los cuatro cuestionarios y fueron aceptadas para que se aplicaran a todas las preguntas de todos los cuestionarios, siempre y cuando fuese pertinente y velando por el correcto entendimiento de las preguntas por parte de los informantes. En la Tabla 20 se pueden observar los porcentajes de preguntas de cada uno de los cuestionarios que fueron modificadas a fin de incorporar el lenguaje inclusivo de género.

Tabla 20. Porcentaje de las preguntas modificadas por cuestionario.

Cuestionario autoaplicado	Cuestionario empleadores	Cuestionario sindicatos	Cuestionario trabajadores
34%	13%	29%	51%

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

El lenguaje inclusivo se caracteriza por la ausencia de sexismo, discriminaciones y otros sesgos de género que ocultan o infravaloran la participación y presencia de hombres y mujeres en la vida social en la comunicación oral y escrita (INE, 2015). Para lograr este objetivo las preguntas de los cuestionarios fueron modificadas recurriendo a fórmulas lingüísticas que dieran cuenta de la presencia de ambos sexos (cuando corresponde), sin detrimento de alguno, a menos que ese cambio sacrificara el correcto entendimiento de las preguntas. En términos generales, las correcciones realizadas a los cuestionarios fueron en la línea de cambiar los genéricos masculinos por genéricos reales, por sustantivos abstractos no sexuados y/o por desdoblamientos (ver Tabla 21). El uso de cada cambio varió según como fuese más claro el enunciado y las alternativas de respuesta.

Tabla 21. Tipos de modificaciones de lenguaje inclusivo de género realizadas a los cuestionarios

Modificación	Uso en piloto con sesgo de género	Uso actual sin sesgo de género
Genérico masculino se sustituye por una idea genérica real	El hombre Los trabajadores Trabajadores contratados	El ser humano o la persona El personal Personas contratadas
Genérico masculino se sustituye por un sustantivo abstracto no sexuado (metonímico)	Trabajadores Informante Encuestador El trabajador y el empleador	Personas trabajadoras o el personal Persona informante o encuestada Personal encuestador La persona contratada y quién la emplea
Genérico masculino es desdoblado	Trabajadores Socios	Trabajadores y trabajadoras Socios y socias
Se eliminan el uso de barras (/), guiones (-) y arrobas (@)	Hijos(as) Trabajador/a	hijos o hijas Trabajador y trabajadora

Fuente: elaboración propia en base a categorías y ejemplos de INE²⁹

V.4.1.1.4. Cambios asociados a la revisión del equipo técnico

Tras las modificaciones realizadas a las preguntas de los cuatro cuestionarios, luego del análisis de las pruebas cognitivas, las recomendaciones del comité de expertos en migraciones y la incorporación del lenguaje inclusivo de género, estas aún presentaban oportunidades de mejora. Estos problemas fueron solucionados por el equipo técnico de la Encla 2019 y por algunas solicitudes de la Dirección del Trabajo.

- En los cuestionarios a trabajadores y sindicatos se modificaron sustancialmente las preguntas:

- D3 y G28: no distinguía si las actitudes “hostiles” ocurridas en la empresa venían desde los superiores o de entre los mismos trabajadores.
- D4: consideraba acciones de prevención de riesgos psicosociales confusas y poco exhaustivas, las que fueron corregidas en base al Ministerio de Salud (2013: 22-23).
- F4 y F5: mezclaban en una misma pregunta materias disímiles, como las relacionadas a la administración de la empresa con otras relacionadas a condiciones de trabajo.
- F7 y F8: eran dicotómicas (Sí o No), no pertenecientes a la escala de Likert y no especificaban de qué trataban las afirmaciones.

²⁹ INE et al.

- H1: era de selección múltiple y estaba precedida por una pregunta filtro sobre la existencia de conflictos en la empresa en genérico.
- H4: no contenía la alternativa de que uno o más conflictos ocurrieran tanto dentro como fuera de una negociación colectiva.

-En el cuestionario autoaplicado se modificaron sustancialmente las preguntas:

- B2: se cambiaron los tramos etarios que deben cruzarse por tipo de contrato.
- D4: se agregó detalle de la duración de la jornada completa e información sobre la desagregación solicitada.
- D5: se incluyó detalle de la duración de la jornada completa y se modificaron los tramos etarios.
- E2 y E3: se añadieron otras categorías de respuesta como “otra nacionalidad latinoamericana o caribeña” y se incluyó una fila “total”.
- E6: se adicionó el detalle de la duración de la jornada de trabajo parcial.
- F8: se especifica en instrucciones qué entender por “sistema de turnos rotativos”.
- F9: se acota el objetivo de la pregunta y cambian alternativas de respuesta de “diurno/nocturno” a “día/noche”.
- G1: se elimina la fila de respuesta “enfermedades profesionales” para ser integrada en G2 y así poder desagregar por sexo.
- G2: se agrega la fila de respuesta “enfermedades profesionales”.
- H1: se incluyen más derechos de protección de la maternidad y la paternidad.

Además de modificar las preguntas del cuestionario, el equipo técnico agregó preguntas a los cuestionarios en caso de ser pertinente. En la Tabla 22 se puede observar que en el cuestionario a trabajadores un 12% de las preguntas son nuevas, en el cuestionario a sindicatos un 5% y en el cuestionario autoaplicado un 7%. En el cuestionario a empleadores no hubo nuevas adiciones.

Tabla 22. Porcentaje de las preguntas agregadas de los cuatro cuestionarios

Cuestionario autoaplicado	Cuestionario empleadores	Cuestionario sindicatos	Cuestionario t
3%	0%	5%	12%

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

-En el cuestionario a sindicatos se agregaron las preguntas:

- A2: sobre si la empresa externalizó labores similares a las que realiza su personal contratado directamente.
- G10: sobre acciones de participación política y social realizadas por el sindicato.
- G12: sobre las relaciones de su sindicato con otros de la misma empresa, si es que lo hay
- G23: sobre la sede sindical de su organización
- I17: sobre la actitud de la empresa en relación a la última negociación colectiva.

-En el cuestionario a trabajadores se agregaron las preguntas:

- A2: sobre si la empresa externalizó labores similares a las que realiza su personal contratado directamente
- D6: sobre el nivel más alto en el que se toman las decisiones de capacitación en la empresa (copiada desde sindicatos)
- F4: sobre si la empresa informó y/o consultó sobre materias relacionadas a la administración de la empresa (copiada desde sindicatos)
- F6 y F7: sobre el grado de acuerdo del informante con respecto a una serie de afirmaciones sobre la igualdad de género y el clima laboral en la empresa (copiadas desde sindicatos)
- F8: sobre si la empresa cumple los acuerdos que tiene con los trabajadores y trabajadoras.

-En el cuestionario autoaplicado se agregaron las preguntas:

- F2: pesquisa si la empresa tuvo contratadas personas bajo algún sistema excepcional de distribución de trabajo y descansos.
- H3: pide al informante señalar si durante el año 2018 hubo personas contratadas con hijos o hijas menores de dos años.
- H5: indaga si la empresa le entregó a sus trabajadores y trabajadoras el derecho a sala cuna.

V.4.1.1.5. Cambios asociados a solicitudes de la Dirección del Trabajo

Otras modificaciones que tuvieron las preguntas de los cuestionarios fueron las solicitadas por la Dirección del Trabajo entre los meses de marzo y abril de 2019, período en el que el INE aplicó la prueba piloto. En la Tabla 23 se puede observar el porcentaje de preguntas modificadas o agregadas por solicitud de la Dirección del Trabajo en cada cuestionario.

Tabla 23. Porcentaje de las preguntas de los cuatro cuestionarios modificadas o agregadas por solicitud de la Dirección del Trabajo

Cuestionario autoaplicado	Cuestionario empleadores	Cuestionario sindicatos	Cuestionario trabajadores
6%	4%	2%	3%

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Las preguntas que la Dirección del Trabajo solicitó durante el mes de marzo fueron:

- B10 y B11 del cuestionario autoaplicado sobre automatización en la empresa.
- B1 del cuestionario a empleadores, que es un filtro para la pregunta B2, sobre reajustes de las remuneraciones en comparación con el IPC.
- B3 del cuestionario a empleadores sobre la cobertura y diferenciación de los reajustes otorgados por la empresa.
- E1 del cuestionario a empleadores que es un filtro para el módulo de capacitaciones (presente en sindicatos y trabajadores también).
- I15 del cuestionario a sindicatos e I7 del cuestionario a trabajadores, que son filtros para la pregunta sobre reajustes de las remuneraciones producto de la última negociación colectiva.

Respecto a los cambios más importantes solicitados por la Dirección del Trabajo, más allá de aquellos relativos al formato, destaca el realizado en la pregunta G25 sobre centrales sindicales (se elimina como alternativa la Central Clasista de Trabajadores, en tanto no existe formalmente³⁰).

V.4.1.1.6. Cambios a los cuestionarios por reordenamiento de módulos

Más allá de modificaciones particulares a las preguntas, se intervino el orden de los módulos de los cuatro cuestionarios. El diagnóstico inicial fue que el orden de estos módulos no era lógico por lo que, en ocasiones, podía confundir a los informantes, en tanto pasaba de una temática a otra sin haber una clara conexión. Este problema se solucionó a partir de un reordenamiento de los módulos y del orden de las preguntas que los componen.

El criterio para el reordenamiento de los módulos fue que estos fuesen lo más homogéneos posibles entre cuestionarios y que tuviesen una lógica que facilitara la recolección de los datos. La forma utilizada fue agruparlos según el objetivo de la Encla 2019 que permitían cumplir. Estos objetivos pueden resumirse y agruparse en cuatro: igualdad de género, condiciones de trabajo, condiciones de empleo y relaciones laborales.

³⁰ Tras la codificación de las glosas de texto que respondían a la alternativa “otro” en g25, se crearon tres nuevas categorías que corresponden a tres centrales sindicales no formales. Tras el procesamiento de datos se consideraron a estas centrales como alternativas válidas de respuesta en tanto efectivamente existen sindicatos en estas, e incluso en la muestra hay más sindicatos en algunas de estas que en las centrales formales.

V.4.1.1.7. Cambios asociados a la diagramación

En términos de formato y presentación los cuestionarios también tuvieron importantes modificaciones, las cuales fueron realizadas por personal especializado en esta tarea del INE. Una vez que el equipo técnico de la Encla 2019 tuvo listas todas las modificaciones a los cuestionarios, estos fueron enviados al equipo encargado de la diagramación. Dicho proceso duró aproximadamente dos semanas durante el mes de abril y parte de mayo, y requirió una revisión exhaustiva de todos los cuestionarios.

Además, la portada de los cuestionarios y espacio donde se registran los datos de identificación de la empresa y su informante, fueron modificados tanto en su diagramación como en su contenido (advertencias sobre la confidencialidad de las respuestas de los informantes e identificación del tipo de cuestionario).

V.4.2. Prueba piloto operativa

V.4.2.1. Antecedentes

A fin de contextualizar este apartado, es pertinente indicar que la prueba piloto de la Encla 2019 fue realizada en plazos muy acotados y con el fin de testear fases importantes del proceso de recolección, como es la estrategia operativa y el funcionamiento y utilidad del material de trabajo de campo.

Dado lo planteado en los antecedentes de este apartado, los objetivos de la prueba piloto fueron:

- Probar la organización del equipo operativo para el trabajo de terreno
- Evaluar las funciones designadas para cada rol
- Probar el material de trabajo de campo en cuanto a su utilidad y suficiencia
- Probar la estrategia de aplicación de encuestas en terreno
- Testear el flujo de gestión del proyecto a nivel del sistema informático
- Identificar dificultades en las distintas etapas del proceso a fin de anticipar riesgos y elaborar planes de contingencia adecuados

V.4.2.2. Planificación

La fase de recolección de datos se planificó inicialmente para el mes de febrero de 2019, sin embargo posteriormente fue ajustada, resultando en un plazo de tres semanas, entre el 11 y el 29 de marzo. Además, por indicación de la Dirección del Trabajo, este testeo solo se realizó en la Región Metropolitana.

V.4.2.3. Selección unidades a encuestar

La muestra de la prueba piloto fue obtenida a partir del marco muestral de la Encla 2019, considerando únicamente a las unidades cuya casa matriz está situada en la Región Metropolitana, según el DNE 2017. La Dirección del Trabajo solicitó que esta prueba se aplicara a solamente 30 empresas, para lo cual el Subdepartamento de Diseño de Marcos y Muestras, considerando una sobremuestra, seleccionó un total de 60 empresas.

V.4.2.4. Aplicación prueba piloto

La aplicación de la prueba piloto se realizó en los tiempos y el lugar geográfico planificados. Para ello el equipo encargado de la recolección de los datos utilizó un encargado de grupo, dos analistas y cinco encuestadores.

V.4.2.5. Resultados prueba piloto

La prueba piloto fue utilizada para realizar un conjunto de mejoras al trabajo de campo que se describen resumidamente a continuación. En la prueba piloto de la encuesta hubo problemas con la aplicación de los cuestionarios a trabajadores debido a la inexistencia de un procedimiento definido para seleccionar al informante de los trabajadores, que tuvo como consecuencia que cuando no había un comité paritario de higiene y seguridad o un comité bipartito de capacitación, no se aplicó el cuestionario a un representante de los trabajadores en la empresa. Este hallazgo fomentó la rápida elaboración de un procedimiento para la selección aleatoria del trabajador informante, que tras el inicio del trabajo en terreno fue complementado con un protocolo general de selección de sindicatos y trabajadores.

Además, se observó que algunas preguntas de los cuestionarios, especialmente del autoaplicado, no venían con respuestas válidas o venían en blanco. Esto llevó a reforzar la cantidad y rigidez de los validadores, en tanto estos impedían un uso indiscriminado de la no respuesta, de respuestas fuera de rango y de inconsistencia entre preguntas.

Por último, debido a la corta duración de la prueba piloto, su ejecución sirvió para evaluar el funcionamiento de los formularios en terreno, más que la gestión operativa o trabajo de campo. Sin embargo, en términos operativos, la prueba piloto permitió pulir lo planificado con respecto a la función del analista y, producto de ello, se realizaron ajustes en la definición del perfil para ese cargo, orientándolo a la búsqueda de, entre otras cosas, personas con experiencia en *call center*.

V.5. Método de capacitación para la recolección de datos

V.5.1. Descripción general de la metodología de capacitación

V.5.1.1. Capacitaciones realizadas por los equipos técnico y operativo

Las capacitaciones fueron realizadas por los equipos del INE involucrados en el proyecto y contaron también con la participación de representantes del Departamento de Estudios de la Dirección del Trabajo para aclarar eventuales dudas generales.

Inicialmente, se planificó realizar una sola capacitación al personal encargado de la recolección de los datos de todo el país en la Región Metropolitana durante la última semana de abril y la segunda semana de mayo de 2019. Sin embargo, debido a la necesidad de reforzar al equipo de recolección, también fue necesario efectuar capacitaciones después de iniciada la fase de recolección.

En total se realizaron doce capacitaciones dirigidas a encargados de grupo, analistas y encuestadores. El objetivo de estas capacitaciones fue entregar a los actores mencionados los conocimientos técnicos y operativos necesarios para una adecuada recolección de los datos, y se llevaron a cabo entre los meses de abril y agosto de 2019.

La duración de cada capacitación varió de acuerdo con las necesidades del proyecto, pero el contenido de cada una de estas fue similar. A continuación, se presenta el contenido de los módulos en torno a los cuales se organizaron las capacitaciones:

Tabla 24. Módulos de contenidos de las capacitaciones

Módulo	Objetivo
Bienvenida y presentación	<ul style="list-style-type: none"> - Presentar al equipo de capacitación (los “capacitadores”) - Presentación instituciones organizadoras: Dirección del Trabajo e Instituto Nacional de Estadísticas - Entregar programa, manuales, cuestionarios y otros documentos
Aspectos administrativos	<ul style="list-style-type: none"> - Contexto INE - Panorama general de la institución
Introducción y antecedentes	<ul style="list-style-type: none"> - Contexto Encla 2019 y objetivos - Relevancia de la Encla 2019 y de Encla en general - Breve historia Encla, desde 1998 hasta la actualidad (2019) - Dificultades y desafíos de la presente Encla 2019
Planificación	<ul style="list-style-type: none"> - Cronograma con fechas relevantes - Plazos de las actividades (recolección y análisis)
Los informantes de la Encla 2019 y los cuatro cuestionarios	<ul style="list-style-type: none"> - Unidades de información - Los cuatro cuestionarios (presentación general)
El secreto estadístico y el INE	<ul style="list-style-type: none"> - Ley 17.374 que resguarda el anonimato de los informantes y le otorga al INE la facultad de solicitar información estadística
Llenado de cuestionarios y el rol del encuestador	<ul style="list-style-type: none"> - Encuestador e informante - Significados del No Sabe, No Responde y No Aplica - Llenado de cuestionarios: colores y formas de llenado - Los límites de la interpretación del encuestador ante respuestas parciales (y cómo proceder ante respuestas insatisfactorias)
Conceptos estadísticos	<ul style="list-style-type: none"> - Elementos del diseño muestral: población objetivo, unidad de análisis, niveles de estimación, período de referencia, etc. - Marco muestral: conformación del marco, cobertura temática, geográfica, estratificación, etc.
Conceptos teóricos	<ul style="list-style-type: none"> - Explicación de conceptos importantes en el marco de la encuesta (empresa, condiciones de trabajo, condiciones de empleo, relaciones laborales e igualdad de género)
Resultados generales prueba piloto	<ul style="list-style-type: none"> - Dar a conocer brevemente los resultados generales de la prueba piloto - Solo se realizó en la primera capacitación
Roles y responsabilidades	<ul style="list-style-type: none"> - Organigrama. - Explicación de los roles y responsabilidades del equipo que formará parte del recolección Encla 2019 (analistas, encargados, coordinadores, encuestadores, etc.)
Instrumentos operativos	<ul style="list-style-type: none"> - Explicación de instrumentos que servirán de apoyo tanto para el trabajo que haga el analista en terreno como del encuestador (hoja de ruta, sistema SGOH, formulario de contacto del analista)

Módulo	Objetivo
Explicación cuestionario autoaplicado	<ul style="list-style-type: none"> - Explicación objetivo y llenado del cuestionario autoaplicado (módulos, tipos de preguntas, preguntas fundamentales, preguntas difíciles) - Explicación de aspectos operativos del llenado del cuestionario y contacto al informante
Cuestionarios en terreno: empleadores, sindicatos y trabajadores	<ul style="list-style-type: none"> - Orden de prelación - Módulos del cuestionario - Tipos de pregunta y pregunta fundamentales (solo en las primeras capacitaciones) - Presentación y explicación de cada una de las preguntas - Protocolo selección trabajadores y sindicatos (prioridad del sindicato)
Explicación del flujo	<ul style="list-style-type: none"> - Coordinación de análisis, (distribución de unidades y asignación de cargas de trabajo) - Contacto y apertura (contacto empresas e informantes, asignación de estado en sistema, verificación de información y gestión de cuestionarios) - Recolección de los cuestionarios - Procesamiento (digitación, revisión, validación y vuelta a terreno cuando es necesario)
Validadores	<ul style="list-style-type: none"> - Presentar los tipos de validadores que tiene cada cuestionario y los procedimientos que corresponde realizar frente a la activación de cada uno - Solo se realizó en el caso de las capacitaciones a analistas
Evaluaciones	<ul style="list-style-type: none"> - Evaluaciones grupales e individuales de los conceptos estadísticos y teóricos, y de los cuestionarios de la Encla 2019 - Estas evaluaciones solo se realizaron en las capacitaciones 1, 2 y 3

Fuente: Instituto Nacional de Estadísticas, Encla 2019.

V.5.1.2. Capacitaciones realizadas por el área de nomenclaturas del INE

Además de las capacitaciones efectuadas por los equipos técnico y operativo que buscaron entregar a los encuestadores, analistas y encargados de grupo los conocimientos necesarios para efectuar la recolección de los datos, el Área de Nomenclaturas del INE capacitó al personal de terreno en relación con el Clasificador Chileno de Actividades Económicas (CIIU4.CL 2012), a fin de aplicar de manera correcta las preguntas a1 y a2 del cuestionario a empleadores. Estas preguntas revisten una especial complejidad, puesto que es necesario que la respuesta que entregue el informante sea particularmente precisa y detallada para posteriormente poder realizar una correcta codificación de la actividad económica principal de la empresa. Además, la pregunta es especialmente importante, puesto que con ella se crea la variable “agrupación de actividad económica”, que es una de las variables más utilizadas para conformar el estrato y con la cual se ajusta el factor de expansión.

Así, el objetivo de esta capacitación fue definir y reforzar el protocolo que se debía utilizar para identificar correctamente la actividad económica de las empresas seleccionadas, con el fin de clasificarlas de manera satisfactoria.

V.5.2. Materiales para la capacitación

En las capacitaciones se instruye a las y los futuros encuestadores a utilizar correctamente los elementos requeridos para la recolección de los datos. Para ello se elaboraron cuatro cuestionarios (autoaplicado, sindicatos, trabajadores y empleadores), un manual para el informante del

cuestionario autoaplicado y un tarjetero para facilitar la respuesta a determinadas preguntas de los cuestionarios presenciales. A continuación, se explica en detalle cada uno de los elementos mencionados.

V.5.2.1. Cuestionarios

Como ya se ha mencionado, la Encla 2019 se compone de cuatro cuestionarios: empleadores, autoaplicado, sindicatos y trabajadores. En cada una de las empresas seleccionadas debían ser aplicados tres de los cuatro cuestionarios mencionados anteriormente para que la unidad fuese considerada como lograda.

V.5.2.2. Manual para el informante del cuestionario autoaplicado

El cuestionario autoaplicado de la Encla 2019 se respondió, principalmente, a través de un formulario web desarrollado especialmente para este efecto³¹. Para acceder al formulario web del cuestionario autoaplicado, el informante debía utilizar un link, clave y usuario enviados por el analista a cargo de la empresa a su correo electrónico. Luego, el informante podía descargar y visualizar un instructivo de llenado que explica cómo contestar cada uno de los tipos de preguntas y los conceptos más relevantes de la encuesta.

V.5.2.3. Tarjetero

Los tres cuestionarios presenciales de la Encla 2019, es decir, aquellos dirigidos a empleadores, sindicatos y trabajadores, contenían preguntas cuyas alternativas, debido a su número y/o extensión, podrían haber sido difíciles de recordar para el informante y, por lo tanto, de responder adecuadamente. Para evitar sesgos relacionados a estos problemas es que se elaboraron tarjeteros cuya finalidad era que el informante visualizara las alternativas al momento de elegir su respuesta. El tarjetero elaborado fue único para los tres cuestionarios presenciales y en total contenía 21 tarjetas.

Ilustración 2. Ejemplo Tarjetero

TARJETA Nº 2	
Empleadores B3	
El porcentaje del último reajuste de remuneraciones en esta empresa fue	
01	Igual para todos los trabajadores y trabajadoras
02	Fue mayor para las remuneraciones más bajas y menor para las más altas
03	Fue menor para las remuneraciones más bajas y mayor para las más altas
04	Solo recibieron reajuste las remuneraciones más bajas
05	Solo recibieron reajuste las remuneraciones más altas

Fuente: Instituto Nacional de Estadísticas, Encla 2019.

V.5.2.4. Manual del trabajo de campo

El manual de trabajo de campo está dirigido a los encargados de grupo, analistas y encuestadores. Este manual entrega los lineamientos conceptuales y operativos relacionados a la recolección de la encuesta. Los principales contenidos del documento son los siguientes: objetivos de la encuesta, características del diseño muestral (población objetivo, unidad de información, niveles de estimación, período de referencia, periodicidad, características del marco muestral, estrategia muestral, unidad

³¹ 84 empresas solicitaron el cuestionario autoaplicado en formato PAPI.

de muestreo, cálculo del tamaño muestral, método de selección y distribución del tamaño muestral), los conceptos técnicos centrales para poder aplicar adecuadamente el cuestionario (condiciones de trabajo, condiciones de empleo, relaciones laborales e igualdad de género), la presentación de los instrumentos para la producción de información (cuestionarios, hoja de ruta y sistema de gestión operativa), la definición de los roles y responsabilidades de los actores involucrados en la recolección y procesamiento de la información y el flujo de las actividades de terreno.

V.5.2.5. Manuales cuestionarios presenciales

Para efectos de la capacitación a encuestadores y del trabajo operativo, fueron elaborados manuales para los cuestionarios dirigidos a empleadores, sindicatos y trabajadores, además de un instructivo general. El objetivo de cada uno de estos manuales es entregar a los encuestadores —y también a los analistas y los encargados de grupo— las herramientas necesarias para poder aplicar correctamente el cuestionario y resolver las dudas conceptuales que surgieran al momento de la recolección. En dichos manuales se explica en detalle cada pregunta de los cuestionarios, su objetivo, la forma correcta de fraseo, posibles problemas, entre otros aspectos relevantes.

V.5.2.6. Manual cuestionario autoaplicado para analistas

Dado que el cuestionario autoaplicado debía ser contestado autónomamente por el informante (es decir, no es aplicado por un encuestador) y que el analista era la persona responsable de resolver las dudas del informante, se elaboró un manual dirigido únicamente al analista cuyo fin era detallar la manera correcta de responder a cada una de las preguntas del formulario autoaplicado. Además, en este manual se hace referencia a gran parte los validadores que se encuentran en las preguntas del cuestionario web, explicando su propósito y funcionamiento.

V.6. Métodos de supervisión

Para el caso de Enclavados los controles consistieron en los monitoreos que se describen a continuación, sumado al acompañamiento técnico y operativo ofrecido a través de visitas o salidas realizadas por ambos equipos.

Una vez terminada la gestión, los formularios eran recibidos³² en el sistema SGO. Esto permitió disponer de información actualizada en la base de datos, identificando aquellos formularios que volvieron de terreno y, por lo tanto, constituyó una herramienta fundamental para monitorear la eficiencia de la gestión de los encuestadores, el avance del trabajo en terreno en función de lo proyectado y la oportunidad para identificar aspectos no deseados que arriesgaran el ritmo del trabajo de campo.

Diariamente se enviaba el estado de la muestra³³ al equipo de recolección. Con esta información se revisaba la carga de trabajo, el estado de avance de la gestión a nivel nacional y la productividad por

³² La recepción se refiere a la lectura electrónica del código de barra asignado a un formulario. Al momento de hacer la lectura, el sistema identifica que se ha recibido el formulario completo de una determinada unidad.

³³ El estado de la muestra es una base de datos que se descarga directamente desde el servidor, que contiene todas las empresas de la muestra, detallando las etapas en la que se encuentra cada formulario (recepionado, digitado, analizado) y su correspondiente estado (correcto o incorrecto).

encuestador. Este insumo también sirvió para identificar cuestionarios autoaplicados sin ningún dato e incompletos y para solucionar dudas del informante en cuanto al llenado.

El Subdepartamento de Implementación de Diseños del INE realizó el monitoreo de indicadores generales a modo de identificar situaciones que podrían representar algún obstáculo para el éxito del proyecto, como por ejemplo identificación de empresas del tramo censal no recolectadas, porcentajes de cuestionarios analizados versus digitados, potenciales errores en validaciones y soluciones de incidencias en los sistemas informáticos.

V.6.1. Observaciones operativas

- Coordinación de las observaciones operativas

A fin de identificar que el flujo de recolección fuera también implementado en las regiones distintas a la Región Metropolitana, y en coordinación con la unidad de supervisión del equipo operativo, se realizaron observaciones operativas de supervisión en terreno entre los meses de junio y julio de 2019. Las regiones visitadas fueron las siguientes: Tarapacá, Antofagasta, Atacama, Coquimbo, O'Higgins, Valparaíso y Aysén.

Adicionalmente, en la Región Metropolitana se realizó acompañamiento a aquellos recolectores de datos que presentaban rendimientos bajo el promedio regional.

- Metodología de las observaciones operativas

Las observaciones operativas se realizaron en dos dimensiones, por una parte se realizó la revisión en gabinete de los cuestionarios ya levantados por los recolectores de datos, con la finalidad de asegurar el correcto llenado de la portada y de la hoja de ruta de la encuesta y determinar si se cumplían con los protocolos operativos establecidos para el proyecto. Para ello se aplicó una pauta que posterior a la visita del supervisor del INE central a la región, fue enviada al coordinador del proyecto regional con la finalidad de realizar la retroalimentación correspondiente.

Por otra parte, se acompañó a los recolectores de datos a terreno. Dicha actividad tuvo por objeto observar aspectos tales como la correcta presentación del recolector en las empresas, la utilización adecuada del material de trabajo, entre otros. Adicionalmente, durante el acompañamiento los supervisores aplicaron una pauta, cuyos resultados fueron remitidos a los encargados del proyecto correspondientes.

V.6.2. Observaciones técnicas

Durante los meses de junio y julio de 2019 el equipo técnico de la Encla 2019 realizó observaciones técnicas a lo largo del territorio nacional. Esta actividad consistió en observar cómo los encuestadores aplicaban los distintos cuestionarios y registrar los principales hallazgos a nivel de pregunta, con el objetivo de hacer una retroalimentación al trabajo en terreno desde el punto de vista técnico que permitiese mejorar la captura de la información. Para este efecto se utilizó una planilla llamada

“cuadernillo de campo de observación técnica”, la cual ya había sido utilizada por otros productos del Departamento de Estadísticas del Trabajo del INE³⁴.

V.6.3. Supervisión cuestionarios digitados

Para detectar eventuales errores de digitación, el equipo técnico supervisó el 5,2% de los cuestionarios de trabajadores/sindicatos y el 7% de los cuestionarios a empleadores³⁵. Estos fueron elegidos de manera aleatoria, a partir de cada una de las seis muestras parciales que llegaron desde el terreno entre julio y septiembre de 2019. El objetivo de esta supervisión fue realizar una comparación entre lo digitado en las bases de datos y lo señalado en papel en los cuestionarios, identificando eventuales errores. Esta revisión permitió observar algunas diferencias entre el dato registrado en la encuesta de papel y las bases de datos.

VI. METODOLOGÍA DE PROCESAMIENTO DE DATOS

Una vez recepcionados los datos, estos fueron procesados para su orden, limpieza y consistencia, asegurando que la publicación cumpla con todos los estándares de calidad. Para ello se desarrollaron los procedimientos que se describen a continuación, comenzando por la clasificación y codificación, que apunta a asignar un código a algunos de los datos existentes para su mejor comprensión y utilización. También se corrobora que los datos sean consistentes, es decir que no haya contradicción entre éstos, para lo cual se generan herramientas como una malla de validadores espejo y una grilla de suficiencia. Esto permite tener la certeza de que los instrumentos presentan datos de calidad y que cumplen con los objetivos mismos de la encuesta. También se desarrollaron una serie de procedimientos para la corrección de eventuales errores de segundo orden (coherencia entre variables). Además, del cálculo de factores de expansión necesarios para asegurar la confiabilidad de las estimaciones.

El trabajo de procesamiento de datos es clave para que el producto final que se publique contenga datos de calidad, revisados y corregidos en caso de que sea necesario. El detalle de esta fase es lo que se describe en lo que sigue del capítulo.

VI.1. Clasificación y codificación de datos

VI.1.1. Codificación actividad económica

En el presente apartado se describe cómo se efectuó la codificación de actividad económica con la información recolectada por medio del cuestionario dirigido a empleadores y en qué variable y cuestionario quedó almacenada la actividad económica codificada según el Clasificador Chileno de Actividades Económicas (CIU4.CL 2012).

³⁴ Encuesta de Microemprendimiento, Encuesta Nacional de Empleo, Encuesta Suplementaria de Ingresos, entre otras

³⁵ Porcentaje sobre el total de la muestra objetivo.

La codificación de actividad económica de la Encla 2019 fue realizada, principalmente, en base a la información recolectada por medio de las preguntas a1 y a2 del cuestionario empleadores.

Para codificar la actividad económica, se utilizó el modelo de codificación automática de la unidad de procesamiento de la Encuesta Nacional de Empleo (ENE)³⁶ en conjunto con una auditoría de la Sección de Nomenclaturas.

VI.1.2. Codificación y depuración de preguntas abiertas

La codificación, en el procesamiento de datos estadísticos, es la traducción del material no numérico a material numérico por parte del codificador, siendo una etapa fundamental para analizar estadísticamente los resultados de una encuesta (Groves et al. 2011: 305-306). El acto de codificar resume información al combinar muchas respuestas individuales en el código de una categoría y, como en todo acto de resumen, es necesario determinar qué nivel de resumen utilizar para codificar los datos.

En general, las categorías de respuesta de las diferentes preguntas de la Encla 2019 fueron establecidas y definidas antes de comenzar el trabajo en terreno en el diseño de la investigación, pero en algunas preguntas, ante la falta de antecedentes certeros sobre las eventuales respuestas de los informantes, se dio espacio para que los informantes pudieran complementar las alternativas de respuesta propuestas, eligiendo la alternativa “Otro” y especificando una respuesta (variables de respuesta “semiabierta”). Así también, en una menor cantidad de preguntas no se le sugirió al informante ninguna alternativa de respuesta, quedando la respuesta completamente abierta (variables de respuesta “abierta”).

A continuación, se analiza la extensión del uso de las variables de respuesta abierta y semiabierta; se describe el proceso de codificación de estas variables que utilizó como categorías de codificación tanto las contempladas inicialmente en el diseño de la encuesta como otras elaboradas inductivamente a través del análisis de las respuestas abiertas de los informantes; y se explica el proceso de depuración realizado sobre las variables de respuesta abierta y semiabierta, una vez codificadas sus respuestas de texto entregadas por los informantes. Es decir, se explica el proceso por el cual las respuestas de texto de las variables de respuesta abierta y semiabierta pasaron a ser categorías de respuestas válidas de nuevas variables con respuestas “cerradas”.

VI.1.2.1. *Uso de preguntas abiertas y semiabiertas*

El cuestionario a sindicatos contiene 25 preguntas semiabiertas y tres abiertas; el cuestionario a trabajadores posee trece preguntas semiabiertas y una abierta; el cuestionario a empleadores contiene catorce preguntas semiabiertas³⁷ y el cuestionario autoaplicado solo seis preguntas semiabiertas.

³⁶ Link metodología ENE: https://www.ine.cl/docs/default-source/ocupacion-y-desocupacion/publicaciones-y-anuarios/separatas/tem%C3%A1ticas/separata-tecnica-codificacion-automatica-ene.pdf?sfvrsn=273dd63c_3

³⁷ Contiene otras tres preguntas semiabiertas y seis preguntas abiertas (todas las derivadas de a1 y a2), pero estas fueron codificadas en cuando se codificó la actividad económica (apartado VI.3)

Como se observa, tanto las preguntas abiertas como semiabiertas se concentran en los cuestionarios presenciales, particularmente, en aquel dirigido a sindicatos, y son utilizadas, principalmente, para complementar las alternativas de respuesta de algunas preguntas de respuesta cerrada.

Las preguntas semiabiertas, por ejemplo i4_77_otro, solamente son respondidas por aquellos informantes que en la pregunta original de respuesta cerrada, por ejemplo i4, marcaron la alternativa “Otra razón. Especifique”, código “77” (Tabla 25).

Tabla 25. Ejemplo cuestionario sindicatos

i4	i4_77_otro
1	NA [sin valor o <i>missing</i>]
8	NA [sin valor o <i>missing</i>]
7	NA [sin valor o <i>missing</i>]
85	NA [sin valor o <i>missing</i>]
77	había sindicato (solo se encontraba la presidenta) y muy pocos socios [respuesta abierta]
1	NA [sin valor o <i>missing</i>]

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Los informantes que hacen más uso de preguntas abiertas y semiabiertas son aquellos de los cuestionarios sindicatos y empleadores.

VI.1.2.2. *Categorización de respuestas en preguntas abiertas y semiabiertas*

Como señala Groves et al. (2011), un primer punto de atención a la hora de codificar debe ser el marco categorial utilizado para la clasificación del material textual. Es decir, las “categorías de respuesta” que, como se ha señalado, provienen tanto del diseño de la encuesta como del análisis inductivo realizado de las respuestas de los informantes.

Según los mismos autores, para que el sistema de códigos sea útil a la hora de codificar, estos deben: a) tener asignados un número único; b) poseer una etiqueta designada a describir todas las respuestas asignadas a la categoría; c) contar con exhaustividad total del tratamiento de las respuestas (todas las respuestas deben ser capaces de ser asignadas a una categoría); d) tener mutua exclusividad (ninguna respuesta debe ser asignada a más de una categoría); y e) ser un número único de categorías que se ajusten al propósito del análisis (Groves et al. 2011: 306-307).

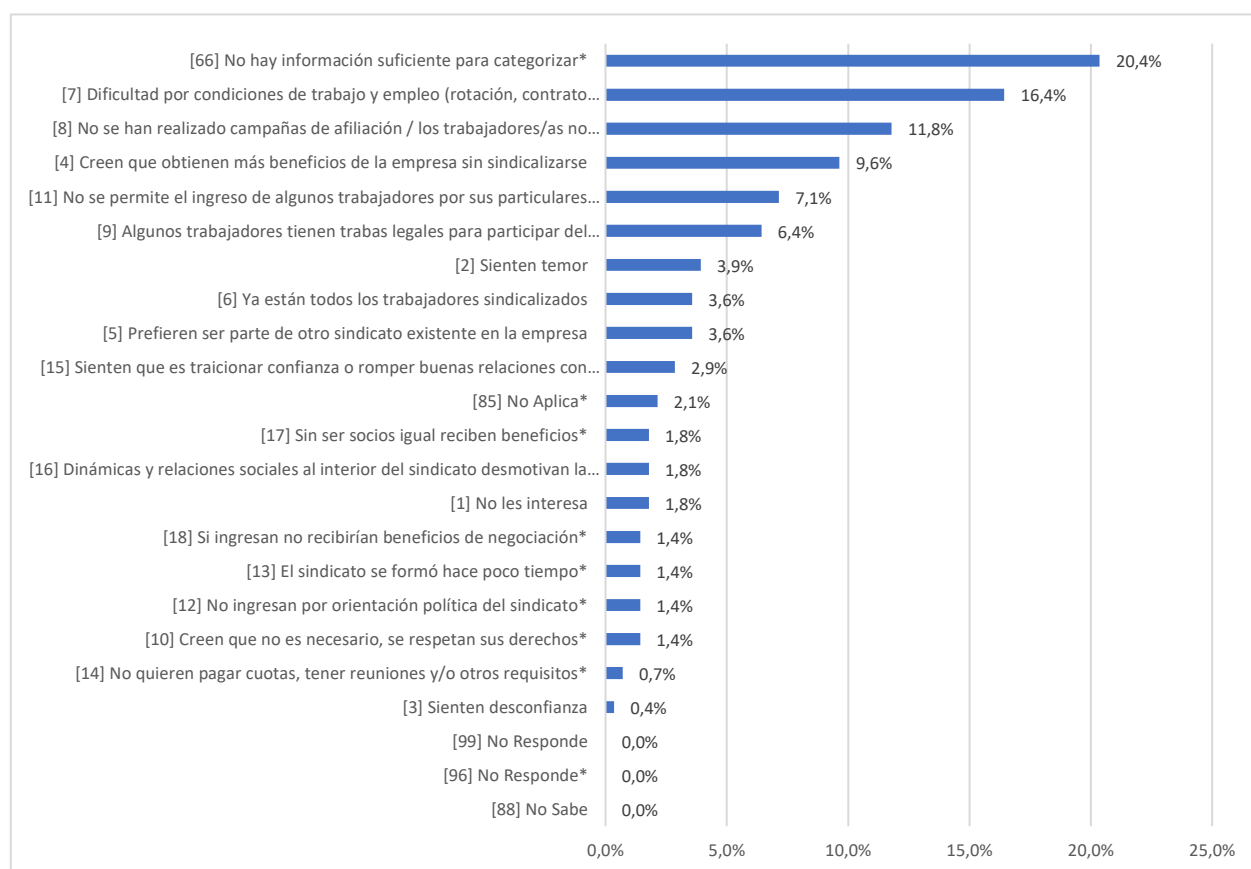
Para realizar la labor de codificar, el codificador debe intentar discernir la intención del informante a través de las palabras registradas y asignarle un solo código de todos los códigos posibles a la respuesta (Groves et al. 2011: 308). Para esto, el codificador debe entender completamente la intención del investigador al definir cada categoría, lo que es función de la calidad de las descripciones de las categorías y la calidad del entrenamiento sobre los conceptos descritos con esas palabras.

La codificación de la Encla 2019 fue realizada por el mismo equipo técnico que participó en la fase de diseño y construcción de la encuesta y, particularmente, en la elaboración de los cuestionarios, por lo que había claridad sobre los objetivos de cada pregunta, sobre la razón de utilizar preguntas de respuesta abierta o semiabierta en algunos módulos de los cuestionarios y sobre la definición de cada categoría utilizada para codificar. Además, este subproceso contó con la revisión y recomendaciones del Departamento de Estudios de la Dirección del Trabajo.

Para el caso de las variables de respuesta semiabierta, se revisaron exhaustivamente cada una de las respuestas registradas como textos, codificando estas respuestas en las categorías originales de las preguntas cuando era posible. Si ninguna de las categorías de las preguntas satisfacía las respuestas abiertas de los informantes, se elaboraron nuevas categorías que permitieran la agrupación de sus respuestas.

Así, por ejemplo, las respuestas de texto de la pregunta g22_77_otro, del cuestionario a sindicatos fueron clasificadas como se muestra en el Gráfico 1 (las categorías de respuesta “nuevas”, es decir, que no se encuentran en el cuestionario, están señaladas con asteriscos).

Gráfico 1. Categorización de respuestas abiertas de la pregunta g22_77_otro. “En su opinión, ¿por qué no hay más trabajadores y trabajadoras en su sindicato?: Otro motivo. Especifique:”. N = 280 (solo quienes marcan “Sí” en g22_77).

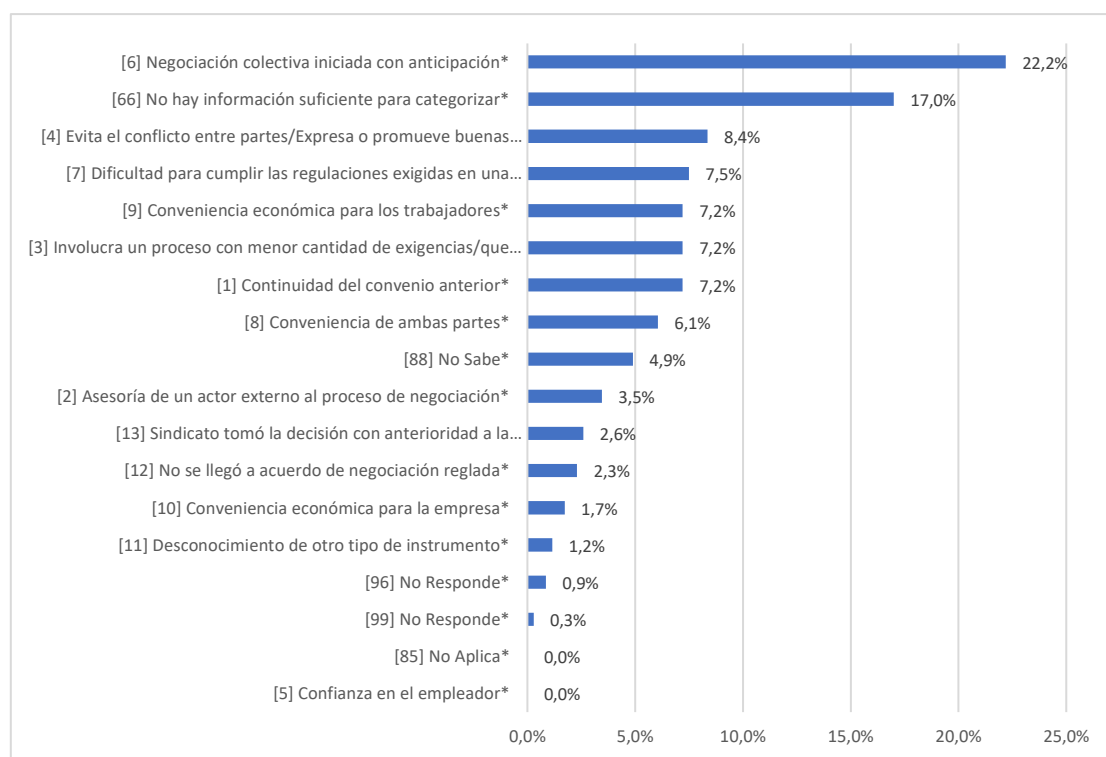


Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

En el Gráfico 1 se puede observar que la categoría que más se utilizó para codificar las respuestas abiertas de los informantes en la pregunta g22_77_otro fue “[66] No hay información suficiente para categorizar” (20,4%). Esta es una categoría nueva (*) que agrupa respuestas no válidas que tenían insuficiente información para ser categorizadas y/o el contenido de la afirmación era tan específico y particular que no podía ser categorizado en las alternativas de respuesta existentes. La segunda categoría que más se utilizó para codificar las respuestas de texto de g22_77_otro fue “[7] Dificultad por condiciones de trabajo y empleo” (16,4%), que es una categoría nueva (*) cuyo contenido no fue pensado como categoría válida de respuesta en el diseño de la encuesta.

En el caso de las preguntas de respuesta abierta que por diseño no tenían ninguna alternativa de respuesta, estas fueron codificadas en base a categorías elaboradas de manera inductiva al revisar las respuestas de texto entregadas por los informantes al responder las preguntas. Así, por ejemplo, las respuestas de preguntas como la i7 del cuestionario a sindicatos fueron categorizadas solamente en base a categorías nuevas, en tanto no había categorías de respuesta definidas previamente (como se observa, todas las alternativas aparecen con un asterisco por ser nuevas).

Gráfico. 2. Categorización de respuestas abiertas de la pregunta i7. “¿Por qué negociaron el instrumento señalado en la pregunta anterior, en lugar de otro tipo de instrumento?”: N = 347 (solo quienes marcan “Convenio colectivo” o “fallo arbitral” en i6).



Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Para categorizar las respuestas textuales de la pregunta i7, la Dirección del Trabajo creó de manera inductiva once categorías, como se puede observar en el Gráfico 2. Las categorías más utilizadas son

“[3] Negociación colectiva iniciada con anticipación” (22,2%) y “[66] No hay información suficiente para categorizar” (17%); todas las demás tienen un uso menor al 10% sobre el total de informantes que respondieron textualmente la pregunta.

VI.1.2.3. *Depuración de las variables de preguntas abiertas y semiabiertas a través de la codificación*

Un grupo de glosas de las preguntas semiabiertas fueron clasificadas en las categorías que la pregunta ya contenía por diseño de investigación. Por ejemplo, en el caso de la pregunta g22_77_otro (Gráfico 1), un 22,9% de las respuestas abiertas (64) fueron categorizadas en las alternativas de respuesta ya existentes de la pregunta g22.

Tabla 26. Pregunta g22. “En su opinión, ¿por qué no hay más trabajadores y trabajadoras en su sindicato?”
Antes y después de la codificación de las glosas de texto de g22_77_otro. “Otro. Especifique”

Precodificación y depuración de g22_77_otro					
Variable	Sí (1)	No (2)	No Sabe (88)	No Responde (99)	Valor perdido
g22_1	433	1067	1	2	28
g22_2	574	927	1	2	27
g22_3	205	1295	1	2	28
g22_4	264	1236	1	2	28
g22_5	130	1370	1	2	28
g22_6	347	1153	1	2	28
g22_77	281	1219	1	2	28
Poscodificación y depuración de g22_77_otro					
Variable	Sí (1)	No (2)	No Sabe (88)	No Responde (99)	Valor perdido
g22_1	437	1063	1	2	28
g22_2	578	923	1	2	27
g22_3	206	1294	1	2	28
g22_4	285	1215	1	2	28
g22_5	139	1361	1	2	28
g22_6	357	1143	1	2	28
g22_77	210	1292	1	0	28

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Como se observa en la Tabla 26, el número de casos que elige la alternativa “Otro. Especifique”, disminuye de 281 a 210 al codificar y sumar a g22 las respuestas abiertas de la variable registradas

en g22_77_otra. Es decir, hay 71 casos que eligieron “Otro” cuando lo correcto era haber seleccionado alguna de las seis alternativas de respuesta ya existentes en g22.

Es importante tener en consideración que las glosas de las preguntas de respuesta abierta y semiabiertas se eliminaron de las cuatro bases de datos por el riesgo que implica mantenerlas para la confidencialidad de los informantes y sus empresas. Sin embargo, se mantienen las variables con las glosas codificadas, las cuales tienen el sufijo “_c”. Por ejemplo: la codificación de g22_otra_77 se llama g22_otra_77_c.

VI.1.2.4. Codificación y nuevas alternativas de respuesta

Las preguntas g25, f6 de sindicatos y f7 de empleadores constituyen excepciones en relación con la manera en la cual se codificaron sus respuestas de texto, puesto que las nuevas categorías codificadas se utilizaron para crear nuevas alternativas de respuesta en la pregunta semiabierta original. Esto llevó a que el número de categorías válidas de estas tres preguntas aumentara tras el proceso de codificación.

Para el caso de la pregunta g25, al momento de mejorar los cuestionarios se tenía conocimiento de la existencia de otras centrales sindicales, pero debido a que carecían de registro formal no fueron incluidas como alternativa de respuesta. Sin embargo, ante la alta frecuencia con que se mencionaron estas centrales no registradas, incluso mayor a las de aquellas que sí lo estaban, se decidió añadirlas como categorías de respuesta. Esto resultó en un incremento de cuatro a siete alternativas en la pregunta g25. Además, hay que tener en consideración que la categoría “No Aplica” (85) de esta pregunta agrupa a decenas de menciones a organizaciones que no eran centrales, sino federaciones, confederaciones y sindicatos.

VI.2. Análisis de consistencia de datos y edición

VI.2.1. Lineamientos generales de la edición

Los subprocesos de depuración, edición e imputación de datos para la Encla 2019 fueron diseñados con el objetivo de velar por la calidad de los datos reportados en las bases de datos finales, con el fin de permitir al investigador o investigadora realizar inferencias estadísticas correctas a partir de ellos (Galván & Medina, 2007). Estos procesos no tienen como objetivo que los parámetros estimados presenten determinadas propiedades estadísticas, excepto en los casos en que estas sean necesarias para asegurar la confiabilidad de la inferencia estadística.

La definición de los objetivos y criterios para la depuración y edición de los datos es una etapa clave de cualquier proceso estadístico. Los métodos específicos de depuración y edición de datos deben ser ajustados a las características de cada producto, tales como su diseño muestral, la periodicidad de las observaciones, sus patrones dominantes de no-respuesta, la existencia o no de información auxiliar, las dimensiones de la base de datos y el uso que se hará de ella (Eltinge & Lurey, 2003).

Los objetivos dan cuenta de los estándares de calidad que se pretenden alcanzar para proveer una base de datos final libre de errores y los criterios establecen cuándo se llevarán a cabo las ediciones para corregir dichos errores.

De modo que los subprocesos de depuración y edición persiguen dos objetivos que suelen resultar contrapuestos: por un lado, buscan corregir la mayor cantidad de observaciones erradas para contar con la mayor cantidad de información válida para posteriores análisis. Y por otro, procuran realizar la menor cantidad de modificaciones a la base de datos, para evitar introducir cualquier tipo de sesgo (Fellegi & Holt, 1976). A continuación, se presentan los objetivos que orientan todos los procedimientos de los subprocesos de depuración y edición de datos de la Encla 2019.

VI.2.1.1. Objetivos de la edición

1. Identificar y cuantificar la no-respuesta, las inconsistencias y los valores atípicos que puedan dar cuenta de errores en las bases de datos.
2. Eliminar de las bases de datos aquellas observaciones que no cumplen con el mínimo estándar de calidad, de acuerdo con los requisitos del marco muestral y las exigencias de los instrumentos de control de calidad: la malla de validadores (MDV) y la grilla de suficiencia (GDS).
3. Corregir los hallazgos previamente identificados, de acuerdo con los criterios establecidos para ello.

VI.2.1.2. Criterios de la edición

A su vez, estos objetivos se encuentran supeditados al conjunto de criterios listados a continuación en orden de importancia. Cabe señalar que la selección de criterios tiene amplios antecedentes en la literatura especializada (Fellegi & Holt, 1976; Sigman, 2001; Cohen, 2003; Graham, Olchowski & Gilreath, 2007) y que su orden fue adecuado a las necesidades específicas de la Encla 2019.

1. Los datos en las bases deben cumplir con los requisitos previamente establecidos en los instrumentos de control de calidad, es decir, las reglas de validación contempladas en la malla de validación y los criterios de suficiencia en la grilla de suficiencia.
2. Se debe editar la menor cantidad posible de datos.
3. Las ediciones de datos siempre deben estar alineadas con los criterios previamente establecidos en la malla de validación, de manera que el procedimiento de edición del dato sea directamente deducible de la regla contenida en dicho instrumento.
4. Se debe priorizar al uso de información consignada dentro del mismo cuestionario para corregir datos erróneos. Concretamente, se deben revisar las respuestas a otras preguntas dentro del mismo cuestionario, así como la información consignada en la hoja de observaciones de los cuestionarios presenciales.
5. Siempre que sea posible, se debe velar por la mantención de la estructura de distribución de los datos y su frecuencia.

6. Siempre que exista la duda razonable de si un dato corresponde o no a un error, se dejará como está. Esto corresponde a un criterio conservador que busca evitar la introducción de sesgo producto de la edición.
7. El número óptimo de ediciones a realizar en cada pregunta debe ser proporcional a la fracción de la muestra total, que presenta respuestas inválidas en dicha pregunta.

VI.2.1.3. Protocolo de edición

De acuerdo con los objetivos y criterios establecidos, los primeros procedimientos a realizar son aquellos que corrigen información que se encuentra errada con certeza y cuyo verdadero valor puede ser deducido simplemente. A estos procesos de edición enfocados en la limpieza de la base de datos les llamaremos *depuraciones*. Algunos ejemplos de estos procedimientos preliminares son:

- Eliminar de la base de datos las unidades que no cumplen con los requisitos mínimos para ser categorizadas en alguno de los estratos definidos por el marco muestral.
- Reemplazar respuestas incorrectamente digitadas por los valores correctos que se encuentran en el cuestionario en papel, o que pueden ser deducidos a partir de información del mismo cuestionario.
- Reemplazar “respuestas perdidas” (*‘missing observation’*) por el código “96” en preguntas categóricas y “-96”, en preguntas numéricas para todos los cuestionarios³⁸.
- Corregir valores en tablas del cuestionario autoaplicado que no se encuentran correctamente cuadradas, pero que contienen información suficiente para deducir el valor de una o más celdas.

VI.2.2. Malla de validadores

La malla de validadores (MDV) corresponde a un conjunto de reglas de consistencia lógica al interior de cada cuestionario. Todas estas reglas de validación fueron programadas en un software estadístico (Stata) y ejecutadas en serie sobre los cuatro cuestionarios que componen la Encla 2019. Este proceso es denominado validación “en lote” o *‘batch’*.

En la MDV encontramos dos tipos de reglas (*‘edits’*) o validaciones: las de primer y de segundo nivel. Separar las reglas de validación en estas dos categorías excluyentes es el procedimiento estándar en procesos de control de calidad de datos (Fallegi & Holt, 1976; Cohen, 2006). Asimismo, corresponde a una práctica habitual de estos procesos al interior del Instituto Nacional de Estadísticas.

³⁸ Esto ocurre, principalmente, por la existencia de erratas y de preguntas de difícil aplicación en los cuestionarios presenciales, lo que tuvo como consecuencia que la base de datos provenientes de la fase de recolección tuviera observaciones perdidas en aquellas preguntas.

- Validaciones de primer nivel: corresponden a las reglas que exigen que cada valor introducido como respuesta a cada pregunta del cuestionario, sea individualmente válido. Esto quiere decir que el valor sea distinto de ‘missing’, que se encuentre dentro del rango definido para él por la pregunta y que respeten el flujo de saltos del cuestionario.
- Validaciones de segundo nivel: corresponden a las reglas que exigen que un conjunto de valores sea consistente entre sí. Estos validadores solo se aplican cuando las preguntas en cuestión han pasado correctamente las validaciones de primer nivel. Por lo tanto, son más exigentes que las anteriores.

En total, la MDV cuenta con 404 condiciones lógicas, reglas o validadores. La Tabla 27 muestra cómo estos se distribuyen entre los cuatro cuestionarios que componen la Encla 2019.

Tabla 27. Número de validadores en los cuestionarios de la Encla 2019

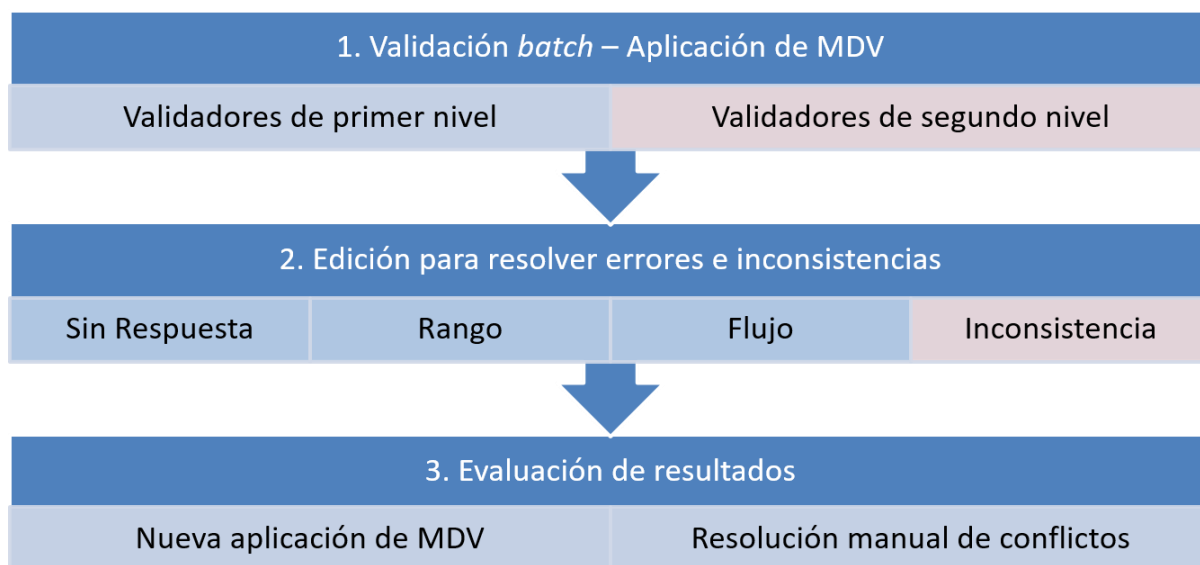
Cuestionario	Validadores de primer nivel	Validadores de segundo nivel
Empleadores	2	23
Sindicatos	12	23
Trabajadores	2	16
Autoaplicado	No Aplica	326

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Como vemos, el cuestionario autoaplicado es el que posee más validadores de segundo nivel. Debido a que se trata de un cuestionario aplicado vía web, los validadores de primer nivel (celda en blanco, rango y filtro) fueron incorporados directamente en la plataforma y no fue necesario incorporarlos en la MDV.

En cuanto a la experiencia comparada en este ámbito, existen encuestas a empresas similares a la Encla 2019 procesadas por el *Labour Bureau of Statistics (LBS)* de EE. UU, las cuales también cuentan con este tipo de instrumentos para el control de la calidad de la información estadística. Sin embargo, en comparación con ellas, la MDV de la Encla 2019 es relativamente extensa, y, por lo tanto, altamente exigente. Además, en el INE la etapa de procesamiento de los datos incluye la aplicación de un segundo control de calidad, llamado grilla de suficiencia.

El siguiente esquema representa los tres pasos principales que estructuran la rutina programada en STATA para validar y editar las bases de datos de la Encla 2019. Corresponde a una rutina semi-automática, ya que combina procesos completamente automatizados con revisiones y resoluciones manuales que exigen la atención y dedicación cuidadosa de un analista.



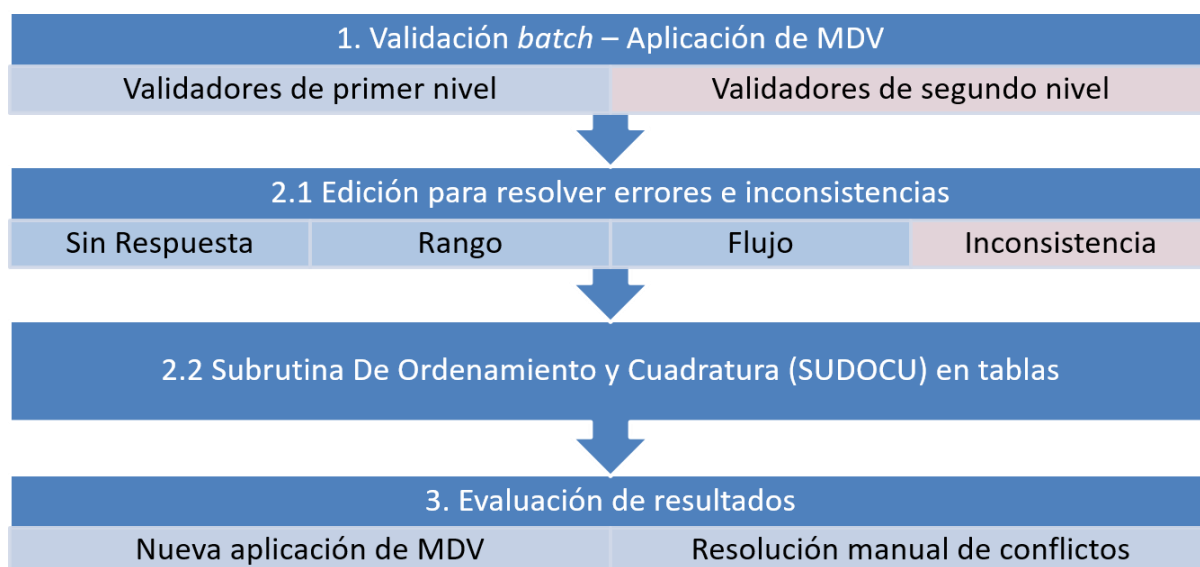
En línea con los objetivos y criterios previamente definidos, subdividimos el proceso de edición en tres pasos. Cuando fue posible, se aplicaron los siguientes métodos de edición de datos para corregir errores e inconsistencias detectados, de acuerdo con los procedimientos habituales del LBS (Cohen, 2006).

- Uso de datos auxiliares
- Uso de otras fuentes de información externas
- Imputación de datos, velando por la conservación parcial o completa de la estructura multivariada

En el caso particular del cuestionario autoaplicado, la edición debió enfrentar un desafío adicional: el **“complejo de tablas”**³⁹, en las cuales se debe desagregar, según diferentes criterios, el mismo número de trabajadores contratados directamente (NT, para abreviar). En algunos casos, dichas tablas se encontraban descuadradas, es decir, indicaban números diferentes que debían ser iguales.

En consecuencia, la rutina programada en STATA consideró un paso adicional para resolver problemas de descuadres o inconsistencias generadas entre números pertenecientes al complejo de tablas. Dicho paso fue llamado *Subrutina De Ordenamiento y Cuadratura* (SUDOCU, para abreviar) y consiste en un algoritmo que corrige los datos consignados en el antes mencionado complejo de tablas, cuando encuentra información suficiente para identificar dichos datos como errores.

³⁹ Traducción del inglés *“table complex”*. Refiere a un conjunto de tablas que se encuentran relacionadas entre sí, por medio de un mismo número. Para la Encla, se trata de subconjuntos y reagrupaciones de un conjunto principal (el número de trabajadores contratados directamente ó NT), los cuales deben ser distribuidos de acuerdo con diferentes criterios de agregación.



Por contrapartida, cuando SUDOCU no encuentra información suficiente para determinar con certeza si un dato es un error o no, la subrutina no realiza modificación alguna. De esta manera, SUDOCU se deriva directamente de los objetivos y criterios definidos previamente para la edición de la Encla 2019. Ella intenta balancear el requisito de corregir información errada e inconsistente en el cuestionario, con el cuidado de mantener intactos los datos originales, siempre que sea posible.

VI.2.3. Grilla de suficiencia

Una vez aplicado el primer control de calidad, así como las correcciones o ediciones pertinentes en cada caso, el flujo continúa hacia un segundo control en dicho sentido, consistente en evaluar que la información ingresada en el cuestionario sea suficiente o, en otras palabras, que cumpla con los llamados “criterios de suficiencia”. Para esto se genera una matriz que recoge las reglas o criterios de suficiencia, llamada “grilla de suficiencia” (GDS).

El objetivo específico de este segundo control de calidad es asegurar que la mayoría de las preguntas de cada cuestionario que pertenecen al grupo de preguntas definidas como principales, contenga información útil, además de válida. En otras palabras, mientras la MDV se asegura de que la información registrada por el informante contenga la menor cantidad posible de errores e inconsistencias, la GDS se preocupa de evaluar si la información es pertinente para el análisis. Por útil o pertinente se entienden aquellas respuestas que son posibles de ser analizadas por el investigador o investigadora que utiliza la base de datos de la Encla 2019. Es decir, la GDS identifica aquellas respuestas que son distintas de “No Sabe” o “No Responde”.

A continuación, detallamos los dos criterios de suficiencia implícitos en la GDS.

- **Suficiencia total:** corresponde al requerimiento de que cada una de las preguntas en esta categoría contengan respuestas distintas de “No Sabe” (88) y “No Responde” (99), en cada

cuestionario.

- **Suficiencia parcial:** corresponde al requerimiento de que una determinada fracción de las preguntas en esta categoría contengan respuestas distintas de “No Sabe” (88) y “No Responde” (99), en cada cuestionario.

En síntesis, en la grilla hay preguntas a las que se les exige “logro total” (100%) y otras que requieren un “logro parcial” (80% - 100%). Todas las preguntas que conforman el primer grupo deben tener respuestas distintas a “No sabe” y “No responde”, de lo contrario, la unidad completa queda identificada como “no lograda”, repercutiendo en que se vuelvan inadmisibles los tres cuestionarios de dicha empresa. A modo de ejemplo, si en el cuestionario de sindicatos un dirigente indica que “No sabe” si en el año 2018 hubo paro, la unidad completa quedará como “no lograda”. En el caso de las preguntas del segundo grupo (80% - 100%), sí se toleran algunos “No sabe” y “No responde”, siempre y cuando éstos no superen el 20% del total de las preguntas del grupo.

La grilla de suficiencia se compone de un grupo de variables, cuya selección consideró los siguientes criterios:

- Variables imprescindibles para estimar los factores de expansión.
- Variables clave del estudio encargadas por la Dirección del Trabajo, previa revisión por parte del INE.

Concretamente, la GDS cuenta con 25 requisitos de suficiencia total y 54 de suficiencia parcial. La Tabla 28 muestra cómo estos se distribuyen entre los cuatro cuestionarios. Por otro lado, la Tabla 29 identifica cuáles variables fueron asociadas a requisitos de suficiencia total o parcial, respectivamente. El detalle de la grilla de suficiencia se presenta en un documento específico adjunto a este informe.

Tabla 28. Número de requisitos de suficiencia total y parcial en los cuestionarios de la Encla 2019

Cuestionario	Suficiencia total	Suficiencia parcial	Total preguntas en GDS
Empleadores	5	12	17
Sindicatos	6	21	27
Trabajadores	1	13	14
Autoaplicado	13	8	21
Total	25	54	80

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

Tabla 29. Variables con requisitos de suficiencia total y parcial en los cuestionarios de la Encla 2019

Cuestionario	Suficiencia total	Suficiencia Parcial
Empleadores	a1, a2, f3, f4, h2	a5, b1, b2, d2, d5, d9, e1, f9, g1, h3, h5, h7
Sindicatos	a2, b4, g2, h1, i3, i15	a1, a3, a7, c7, e1, f2, f8, f9, g3, g4, g11, g16, g17, g28, h4, i2, i6, i16, i18, i22, i23
Trabajadores	a2	a1, a6, a7, c7, d2, e1, e3, f2, f7, f8, g3, g6, h1
Autoaplicado	a1, a2, b1, b2, b5, c1, c7, c15, d1, d5, e1, j1, j4	b10, b11, c11, e2, f1, h1, h4, h5

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

En línea con los objetivos y criterios del proceso de edición, definidos con anterioridad, el proceso de edición de datos se estructuró en dos pasos que presentan insuficiencias totales o parciales, de acuerdo con los resultados de la aplicación de la GDS. Estos pasos son homólogos a los utilizados para resolver inconsistencias.

Cuando fue posible, se aplicaron los siguientes métodos de edición de datos para corregir errores e inconsistencias detectados, de acuerdo con los procedimientos habituales del LBS (Cohen, 2006).

1. Uso de datos auxiliares.
2. Uso de otras fuentes de información externas.
3. Conservación parcial o completa de la estructura multivariada.

Adicionalmente, el equipo técnico realizó recuperaciones telefónicas, retomando contacto con el informante para corregir errores puntuales que hubiesen repercutido en eliminar el cuestionario con problemas del grupo de observaciones terminadas.

VI.3. Procedimientos de edición

VI.3.1. Resolución de errores de flujo

Estos errores ocurren cuando un informante contesta las preguntas de forma tal que no sigue correctamente el flujo del cuestionario, por ejemplo, no realiza un salto cuando debería haberlo hecho. Siguiendo el criterio conservador anteriormente descrito de realizar la menor cantidad de modificaciones posibles, se determinó que la edición más adecuada consistía en anular las respuestas que no debieron haber sido contestadas, de acuerdo con el flujo normal de saltos presentes en el cuestionario. Este constituye el primer paso de la edición destinada a resolver validadores de flujo y nos referimos a ella como etapa VF1.

En una segunda etapa, llamada VF2, los analistas del equipo técnico de la Encla hicieron una revisión exhaustiva de los casos en los que existía más de una pregunta con respuesta al interior de un salto que el informante sí debía tomar, dada su respuesta en la pregunta de filtro. En estos casos, cuando la información en el interior del salto era suficiente y consistente, se consideró que cambiar la pregunta de filtro era menos invasivo que eliminar las respuestas válidas. Por lo tanto, en estos casos se editó la respuesta del informante por aquella consistente con el flujo de respuestas.

VI.3.2. Resolución de errores de pregunta sin respuesta

Estos errores ocurren cuando un informante no contesta una pregunta que debe contestar, lo que implica que el digitador no ingresa un valor, quedando en su lugar un “.” (“missing”) en una pregunta que se encuentra dentro del flujo que el informante debe seguir. Para las preguntas categóricas, el procedimiento general de edición consistió en marcar la no respuesta con un código especialmente reservado para ello, el “96”, análogo al código “99” que asigna el encuestador cuando el informante no responde durante la entrevista. En cambio, para las preguntas numéricas, como aquellas que se encuentran en tablas, el procedimiento general de edición consistió en marcar la no respuesta con el código “-96”, a fin de diferenciarlo del valor 96 que eventualmente podría tomar este tipo de variable.

VI.3.3. Resolución de errores de rango

Los validadores de rango son aquellos que se activan cuando un informante responde una pregunta con un valor que no corresponde. En preguntas categóricas, esto puede ocurrir debido a problemas de la recolección o digitación; en las preguntas abiertas esto puede ocurrir porque el informante no quiere entregar la información correcta, no comprende bien la pregunta, confunde períodos de referencia y conceptos, entre otras hipótesis.

Un ejemplo es la pregunta d3 del cuestionario autoaplicado. Esta pregunta debió ser editada de manera pormenorizada, debido a que el equipo técnico identificó informantes que ingresaron valores por debajo del *salario mínimo legal*, \$288.000, en la remuneración bruta mínima pagada.

En consecuencia, se llevó a cabo una edición que buscó encontrar no el valor correcto, puesto que se determinó que esto no era posible⁴⁰, sino que aquel que fuese consistente con la información sobre remuneraciones ingresada por el informante en D1 y D2. Estas preguntas son una buena fuente de información válida, puesto que fueron previamente editadas durante el proceso de cuadratura de tablas

Concretamente, se validaron los montos ingresados en la celda d3_1_1 cuando se encontraban dentro del intervalo de remuneraciones correspondiente al menor tramo de ingresos (fila), reportado previamente en la pregunta D1, que contara con personas contratadas directamente. Por el contrario, cuando el valor ingresado en d3_1_1 no coincidía con el tramo menor de ingreso reportado

⁴⁰ Los intentos por recuperar telefónicamente esta pregunta no dieron frutos, entre otras razones, porque el informante manifestaba incomodidad con la pregunta. Además, si bien el *formulario 1887* del Servicio de Impuestos Internos o la *Encuesta de Ingresos Suplementarios* del Departamento de Estadísticas del Trabajo contabilizan ingresos por debajo del mínimo legal, estos no reflejan la misma información que pretende levantar la pregunta d3_1_1. Por lo tanto, el equipo técnico determinó prescindir de estas fuentes de información para corregir dicha pregunta.

en D1 o D2, se procedió a evaluar si las siguientes transformaciones del valor ingresado en d3_1_1 sí coincidían:

- a) Se dividió el monto por el número de personas que reciben el mínimo pagado en la empresa (d4_1_3). Evalúa si el informante ingresó la *remuneración bruta mínima mensual TOTAL*, en lugar de per cápita.
- b) Se multiplicó el monto por 4 (número de semanas laborales en el mes de noviembre 2018) y por 40, 41, 42, 43, 44 y 45 (número de horas laborales por semana), alternativamente. Evalúa si el informante ingresó la *remuneración bruta mínima POR HORA*, en lugar de mensual.
- c) Se multiplicó el monto por 20 (número de días laborales en el mes de noviembre 2018). Evalúa si el informante ingresó la *remuneración bruta mínima DIARIA*, en lugar de la mensual.
- d) Se multiplicó el monto por 4 (número de semanas laborales en el mes de noviembre 2018). Evalúa si el informante ingresó la *remuneración bruta mínima SEMANAL*, en lugar de la mensual.

Como se puede ver, cada una de estas posibles transformaciones trae aparejada un supuesto o hipótesis respecto de la manera en que el informante responde la pregunta equivocadamente. Estas hipótesis están ordenadas de la más a la menos probable, de acuerdo con el criterio del equipo técnico y la evidencia levantada en esta pregunta.

VI.3.4. Resolución de inconsistencias

Estos errores ocurren cuando un informante contesta dos o más preguntas de forma contradictoria entre sí. Por ejemplo, entre la pregunta h3 y la pregunta b4_6 del cuestionario a empleadores, habrá una contradicción o se activará un “validador de inconsistencia” cuando un informante marque en la primera pregunta que “No” hubo negociación colectiva en su empresa desde el 2015 a la fecha; y en la segunda diga que su empresa “Sí” incorporó durante el año 2018 la igualdad salarial en negociaciones colectivas.

En el caso de encontrar este tipo de contradicciones, se aplicó el siguiente procedimiento: se tomó como respuesta válida aquella ligada a la pregunta de mayor relevancia, de acuerdo con los criterios de la MDV y GDS. Si la pregunta no se encontraba en esta selección, se tomó como respuesta válida la de la pregunta de cuyas respuestas dependen las respuestas de otras preguntas, como es el caso de las preguntas de salto. De esta forma, al resolver la inconsistencia, se logró minimizar la cantidad de información perdida.

VI.4. Imputación y creación de nuevas variables

VI.4.1. Imputación de datos por medio de bases complementarias

VI.4.1.1. Registros administrativos sobre subcontratación de la DT

Con el fin de obtener un mayor número de respuestas válidas en las preguntas sobre subcontratación en las bases de datos de la Encla 2019, se recurrió a los registros administrativos de la Dirección del

Trabajo sobre la temática. En este proceso se consideraron todos los casos recibidos en las bases de datos desde el terreno, sin importar su estado de análisis.

En específico, se cruzaron tres de las cuatro bases de la Encla (trabajadores, sindicatos y autoaplicado)⁴¹ con los registros de la Dirección del Trabajo.

Primero, se buscó a 21 empresas que en el cuestionario autoaplicado no entregaron respuestas válidas a la pregunta c11. “¿Durante el año 2018, esta empresa fue subcontratada por otras empresas?”. Solamente cinco de estas 21 empresas se encontraba en el registro administrativo de la Dirección del Trabajo, por lo que se realizó la edición correspondiente para esta observación.

En segundo lugar, se buscó en los registros administrativos de la Dirección del Trabajo a 37 empresas sin datos (NA) y a 36 con respuesta “No Sabe” (88) en la pregunta a1. “¿Durante el año 2018, la empresa que le paga empleó personal subcontratado y/o suministrado por terceros?”, del cuestionario a sindicatos. Aplicándose lo que se describe en la tabla a continuación.

Tabla 30. Distribución de respuestas de pregunta a1, sindicatos antes y después de imputar

Variables	Sí (1)	No (2)	No Sabe (88)	NA
a1 (original)	840	628	37	26
a1 (editada)	873	628	19	11

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

En tercer lugar, se buscó en los registros administrativos de la Dirección del Trabajo a 648 empresas sin datos y a 76 con respuesta “No Sabe” (88) en la pregunta a1. “¿Durante el año 2018, la empresa que le paga empleó personal subcontratado y/o suministrado por terceros?”, del cuestionario a trabajadores. De estas empresas, 353 fueron encontradas en los registros administrativos de empresas mandantes la Dirección del Trabajo, las que pasaron de tener valores perdidos u “88” a tener valores “1” (Sí) en a1. Como se observa, las respuestas “No” de la pregunta a1 no han sido alteradas.

Tabla 31. Distribución de respuestas de pregunta a1, trabajadores, antes y después de imputar

Variables	Sí (1)	No (2)	No Sabe (88)	NA
a1 (original)	846	2579	76	648
a1 (editada)	1185	2579	32	353

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

⁴¹ El cuestionario de empleadores no contiene preguntas sobre subcontratación.

VI.4.1.2. *Registros administrativos sobre sindicatos*

Con el fin de obtener un mayor número de respuestas válidas en las preguntas sobre sindicatos en las bases de datos de la Encla 2019 y, además, incluir algunas variables de interés que no forman parte del cuestionario, se recurrió al análisis de los registros administrativos de la Dirección del Trabajo sobre la temática. En específico, se creó una base de datos “ooss” (organizaciones sindicales) con información de la Dirección del Trabajo actualizada hasta noviembre de 2018, la que contenía las siguientes variables:

- Rol sindical único del sindicato (RSU)
- Nombre del sindicato
- Estado del sindicato
- Año de constitución del sindicato
- Número de sindicatos en la empresa en la que se encuentra el sindicato
- RSU del sindicato de empresa, de establecimiento o transitorio más grande de la empresa en la que se encuentra el sindicato (esta variable, junto a la siguiente, permitió crear otra para distinguir si el sindicato encuestado es o no el más grande de la empresa)
- RSU del sindicato de interempresa más grande de la empresa en la que se encuentra el sindicato (esta variable, junto a la anterior, permitió crear otra para distinguir si el sindicato encuestado es o no el más grande de la empresa)
- Región del sindicato
- Federación a la que se encuentra afiliado el sindicato (nombre y RSU)
- Confederación a la que se encuentra afiliado el sindicato (nombre y RSU)
- Central sindical a la que se encuentra afiliado el sindicato (nombre y RSU)
- Tipo de sindicato
- Número de socios del sindicato
- Número de socias del sindicato
- Número total de socios del sindicato

La base “ooss” se cruzó con la base de datos de sindicatos de la Encla 2019 con el fin de que esta última tuviese más información sobre el sindicato y, además, poder realizar una imputación a aquellas preguntas en las que no hubiese respuesta válida (valor perdido, “No sabe” o “No responde”).

Para hacer el *match* entre ambas bases se utilizaron las variables “rut_empresa” y “rsu”, en tanto en una empresa puede haber más de un sindicato y un sindicato puede estar en más de una empresa (sindicatos interempresa).

Tabla 32. Número de respuestas “no válidas”⁴² en preguntas de la base de sindicatos, antes y después de imputar

Tipo de variable	g1	g2_1	g2_2	g2_3	g3_3	g11	g24	g25	g26	g27
Encla 2019 (variable original)	27	40	42	35	40	28	44	88	42	46
Encla 2019 con RRAA (variable editada)	13	13	13	13	13	2	14	47	26	28
Encla 2019 empresa terminadas	0	0	0	0	0	0	0	4	1	2

Fuente: Elaborado por el Instituto Nacional de Estadísticas, Encla 2019.

VI.4.2. Diagnóstico imputación

VI.4.2.1. *Lineamientos generales para analizar la no respuesta*

Todas las encuestas tienen presencia de no respuesta. Esta puede ser de dos tipos: no respuesta a la unidad (donde el individuo no contesta la encuesta) y no respuesta al ítem (cuando no hay respuestas para algunas preguntas del cuestionario). Las razones por la que los informantes no entregan toda la información solicitada son diversas, entre las que destacan el desconocimiento sobre la utilidad de la información recolectada por la encuesta, aburrimiento, desconfianza, no querer responder temas sensibles, entre otros. En este apartado se analizará la no respuesta al ítem.

Para determinar la factibilidad y necesidad de generar datos completos para ciertas preguntas de los cuatro cuestionarios de la Encla 2019, se efectuó un análisis de la no respuesta al ítem con foco en aquellas preguntas incluidas en la grilla de suficiencia.

Si la proporción de datos perdidos es alta, podría haber dos consecuencias: incremento en la varianza de las estimaciones realizadas y sesgo, si los datos no están perdidos al azar.

En términos generales, la no respuesta a las preguntas en grilla de suficiencia de los diferentes cuestionarios presenta valores bajos. Además, para cada uno de los cuestionarios se identificó el patrón de datos perdidos de las preguntas relevantes.

Una parte crucial para tratar la no respuesta es identificar los mecanismos que la motivan. Para realizar esto, es útil verificar si el indicador del patrón de no respuesta está asociado con variables de caracterización (covariables) observadas en las bases de datos. La identificación de que no hay un patrón de no respuesta nos llevaría a observar que la misma se distribuye de forma homogénea entre las diferentes covariables, como pueden ser la rama de actividad, región o tamaño de la empresa.

Un análisis más robusto sobre los mecanismos de la no respuesta consiste en intentar clasificar la misma en tres grandes categorías:

- MAR (Missing at random): asume que la no respuesta depende de valores observados en los datos, pero no de los inobservados. Esto es imposible de testear, ya que requiere datos no disponibles acerca de los valores faltantes.

⁴² Por respuestas “no válidas” considerar celdas vacías (valor perdido, “No Sabe” (88) y “No Responde” (99)).

- MCAR (Missing completely at random): asume que la no respuesta es independiente de valores observados e inobservados, lo cual es más fuerte que el mecanismo MAR y se puede testear solamente con los datos observados.
- MNAR (Missing not at random): asume que la no respuesta depende de valores inobservados.

Cuando el mecanismo MCAR es el que se encuentra presente en los datos faltantes, la probabilidad de tener una no respuesta en una observación particular no depende de la variable faltante y tampoco de covariables. Este mecanismo es el más simple, porque las respuestas de las variables de interés son similares entre los que responden y los que no responden, es decir, los datos estarían perdidos al azar, no generando sesgos en la estimación de los parámetros de interés.

Para testear si la no respuesta sigue un patrón MCAR se utilizó el Test de Little, que puede ocuparse cuando las variables de interés son continuas (Li, C. (2013)). Para las variables de interés categóricas se utilizó el *chi-cuadrado de Pearson* que permite determinar si hay independencia entre la no respuesta de la variable analizada y las covariables.

VI.4.3. Creación de nuevas variables

Este subproceso consiste en la obtención de variables que no son provistas de forma explícita a través del proceso de recolección, pero que son necesarias para los siguientes subprocesos y fases del proceso estadístico (cálculo de los factores de expansión, análisis y difusión de los resultados).

VI.4.3.1. Variables de estratificación

A partir de los cuestionarios autoaplicado y empleadores se derivarán las tres variables de estratificación de la muestra, que permitirán actualizar el estrato al que pertenece cada una de sus unidades. Inicialmente, se utilizó el Formulario N°1887 año 2016 del Servicio de Impuestos Internos (SII) y el Directorio Nacional de Empresas año 2017 (DNE 2017) del INE para elaborar el marco muestral y seleccionar a las empresas. Sin embargo, con los datos obtenidos a través de la recolección, es posible actualizar el estrato de cada empresa, el cual se asigna a cada cuestionario levantado en esa unidad. Las variables a partir de las cuales se construyen dichos estratos son la agrupación regional, el tamaño de la empresa según número de trabajadores y la agrupación de actividad económica. A continuación, se detalla cómo se construyó cada una de estas variables.

VI.4.3.1.1. Agrupación regional

Por medio del cuestionario autoaplicado se obtuvo la variable “casa_matriz_region”, que permite conocer en cuál de las 16 regiones de Chile se ubica la casa matriz de la empresa. Para fines de la estratificación de la muestra, estas 16 regiones fueron categorizadas en las seis agrupaciones regionales que muestra la Tabla 33.

Tabla 33. Número de empresas en la muestra, según agrupación regional

Agrupación regional	Número de empresas en la muestra
Norte grande	351
Norte chico	279
Zona centro	1.766
Centro sur	501
Sur	502
Zona austral	271
Total	3.670

Fuente: Instituto Nacional de Estadísticas, Encla 2019.

VI.4.3.1.2. Tamaño de la empresa

La base de datos del cuestionario autoaplicado contiene un grupo de variables que indican el número de trabajadores contratados directamente por la empresa al 30 de noviembre del año 2018.

La variable “NT” es numérica, siendo su valor mínimo cinco trabajadores, su valor máximo 29.262, su mediana 93 y su media 387 (sin factor de expansión). Esta variable fue categorizada en cuatro agrupaciones, dada la estratificación de la muestra, como se muestra en la Tabla 34.

Tabla 34. Número de empresas, según tamaño

Tamaño empresa	Glosa tamaño	Número de trabajadores asalariados		Número de empresas
		Límite inferior	Límite superior	
Grande	1	200	Más	1.349
Mediana	2	50	199	875
Pequeña	3	10	49	1.146
Micro	4	5	9	300

Fuente: Instituto Nacional de Estadísticas, Encla 2019.

VI.4.3.1.3. Agrupación actividad económica

Las preguntas a1 y a2⁴³ permitieron obtener la actividad económica principal de la empresa, según el Clasificador Chileno de Actividades Económicas CIIU4.CL 2012. Esta actividad fue almacenada en la variable “codigo_actividad_economica” y agrupada en “codigo_actividad_economica_agrup” como se indica en la Tabla 35.

⁴³ Estas tres variables no se presentan en las bases de datos finales en tanto su cruce hacía extremadamente determinable la identidad de las empresas de la muestra.

Tabla 35. Número de empresas, según CIIU4.CL 2012

codigo_actividad_economica_agrup	codigo_actividad_economica	Etiquetas de actividad económica	Número de empresas en la muestra ⁴⁴
A	A	Agricultura, ganadería, silvicultura y pesca	191
B	B	Explotación de minas y canteras	71
C	C	Industrias manufactureras	362
D-E	D-E	Suministro de electricidad, gas, vapor y aire acondicionado; suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación	91
F	F	Construcción	410
G	G	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	658
H-J	H-J	Transporte y almacenamiento; Información y comunicaciones	288
I	I	Actividades de alojamiento y de servicio de comidas	240
K-L	K-L	Actividades financieras y de seguros; Actividades inmobiliarias	170
M-N	M-N	Actividades profesionales, científicas y técnicas; Actividades de servicios administrativos y de apoyo	443
P	P	Enseñanza	432
Q	Q	Actividades de atención de la salud humana y de asistencia social	213
R-S	R	Actividades artísticas, de entretenimiento y recreativas;	101

Fuente: Instituto Nacional de Estadísticas, Encla 2019.

VI.5. Cálculo y ajustes de factores de expansión

En este subproceso se calculan los factores de expansión de la encuesta que hacen posible la expansión de los datos muestrales a la población objetivo. El subproceso de cálculo de los factores de expansión se expone en el apartado IV.8.

VI.5.1. Ajuste por elegibilidad

Tras el trabajo de campo, se clasifican las empresas según su estado de elegibilidad, diferenciando entre unidades “elegibles”, “no elegibles” o “elegibilidad desconocida”. El 86,8% del total de empresas se encuentran clasificadas como elegible, mientras que 9,6% es no elegible y 3,7% corresponde a elegibilidad desconocida.

⁴⁴ Solo se consideraron aquellas empresas de las bases de datos finales (N = 3.861)

La agrupación regional centro sur presenta el mayor porcentaje de empresas elegibles (91,4%), mientras que norte chico es la agrupación regional con menor proporción de empresas elegibles (81%). Por otra parte, las agrupaciones regionales con menor proporción de empresas no elegibles son norte grande y centro sur. Mientras que las agrupaciones regionales sur, norte chico y zona austral presentan mayor proporción de empresas no elegibles, con porcentajes de 15,3%, 12,9% y 12%, respectivamente. En cuanto a las empresas con elegibilidad desconocida, se observa la mayor proporción en el norte grande (7,7%).

Según tamaño, el grande presenta el mayor porcentaje de empresas elegibles, con 91,7%, y en las medianas se encuentra el menor porcentaje de empresas elegibles, con 60,1%. Mientras que la mayor proporción de empresas con elegibilidad desconocida se encuentra en el tamaño micro, correspondiente al 9,7%, al igual que para las empresas no elegibles, con 26,8%.

Se observa que en la actividad industrias manufactureras existe un mayor porcentaje de empresas elegibles (93,0%) con relación al resto de las actividades económicas. Por otro lado, la actividad enseñanza tiene la mayor cantidad de las empresas no elegibles (16,3%) y minería evidencia la proporción más alta de empresas con elegibilidad desconocida (8,9 %).

VI.5.1.1. Ajuste por elegibilidad desconocida

El 9,6% de las unidades es no elegible y 3,7% corresponde a elegibilidad desconocida. A su vez, este último grupo es mayor en el norte grande (7,7%). Para las empresas no elegibles y elegibilidad desconocida, sus valores menores corresponden a la agrupación centro sur (6,5% y 2,1%).

VI.5.1.2. Ajuste por cambio de estrato y no elegibilidad

Analizando cada una de las agrupaciones regionales, se encuentra que para norte grande la mayor proporción de cambio es en la agrupación de actividad económica “B: Explotación de minas y canteras”, con 35,5%, mientras para las agrupaciones regionales norte chico, zona centro y centro sur las agrupaciones de actividad económica con mayores cambios son la “R-S: Actividades artísticas, de entretenimiento y recreativas – Otras actividades de servicio”, con 38,5%, 18,1% y 38,9% respectivamente.

En la agrupación regional sur, la mayor proporción de cambio en la actividad económica es B: Explotación de minas y canteras, con 36,4%.

Para la zona austral se identificó que para el 28,6% de las empresas elegibles que cambió su estrato, la agrupación de actividad económica con mayor porcentaje de cambio es “K-L: Actividades financieras y de seguros – Actividades inmobiliarias”.

Es importante resaltar que al realizar el ajuste por no elegibilidad, se redujo el *stock* poblacional al que se expande antes del ajuste, debido a que se descuenta el peso de las 555 empresas no elegibles. Por lo tanto, la población de empresas después de aplicar este procedimiento es de 79.552, permitiendo corregir los efectos que pudieran causar estas empresas sobre la estimación.

Por otra parte, en cuanto a los cambios de estrato, existe una compensación de empresas dentro de cada uno de ellos, dado que mientras unas empresas se van del estrato, otras llegan al mismo.

VI.5.1.3. Ajuste por no respuesta

Del total de empresas elegibles, 27,1% no respondió la encuesta de forma completa⁴⁵. En cuanto a nivel de agrupaciones regionales, la mayoría de las empresas que no respondieron la encuesta se ubican en el norte grande y centro sur, con una tasa de no respuesta de 28,7% y 28,5% respectivamente, mientras que para la zona centro se encontró el menor porcentaje de no respuesta, 22,9%.

Por otra parte, de las empresas elegibles que no respondieron la encuesta según su tamaño, las de tamaño grande son las que cuentan con mayor porcentaje (29,6%). Al desagregar según agrupación de actividad económica, “I: Actividades de alojamiento y de servicio de comidas” presenta la presenta mayor proporción de empresas que no respondieron a la encuesta, 32,5%.

Al realizar el ajuste por no respuesta, se redujo el marco muestral a 79.437 empresas, debido a que se encontraron estratos sin ninguna empresa lograda, lo que trae como consecuencia no poder repartir los pesos muestrales. Lo anterior ocurrió en 9 estratos que contienen un total de 14 unidades.

VI.5.2. Suavizamiento del factor de expansión

Después de realizar los ajustes descritos anteriormente, se observa la presencia de valores atípicos en los factores de expansión, de modo que surge la necesidad de realizar un suavizamiento.

VI.6. Ejecución de cálculos agregados

Una vez generados los factores de expansión y junto con la base de datos de recolección, se calculan los errores de estimación asociados a la variable de interés. En las siguientes tablas se presenta la estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2018, además del error relativo, según agrupación regional, tamaño de empresa, agrupación de actividad económica y el cruce de agrupación regional y tamaños de empresa según número de trabajadores, respectivamente.

El tamaño muestral logrado finalmente corresponde a 3.670 empresas a nivel nacional con un error relativo de 8,8%, distribuidos en 293 estratos. La cantidad de estratos disminuye con respecto al valor inicial mencionado, dado que no se logró una recolección en cada uno de los estratos iniciales.

Para la interpretación de las estimaciones, se recomienda utilizar con precaución las estimaciones cuyo coeficiente de variación sea mayor a 15% para realizar inferencias poblacionales, ya que podrían llevar a conclusiones poco acertadas. De igual manera, se sugiere acompañar las estimaciones que estén por encima de dicho umbral de aceptación con notas metodológicas que adviertan las restricciones y el error asociado a dicha estimación.

⁴⁵ Incluye a las empresas elegibles clasificadas con alguno de los siguientes estados de disposición final: incompleta, rechazo y no contactada.

VII. METODOLOGÍA DE ANÁLISIS

VII.1. Análisis de comportamiento de resultados

Con el objetivo de explorar los datos a nivel agregado, se realizó un análisis descriptivo de las principales variables pertenecientes a los cuatro cuestionarios en estudio. El objetivo de este ejercicio fue observar el comportamiento de los resultados en términos de revisión y consistencia

VII.2. Comparación con registros administrativos

Con el objetivo de ejecutar un análisis de consistencia, cuando fue posible se contrastaron los resultados de las variables principales de cada cuestionario con información proveniente de los registros administrativos de la Dirección del Trabajo. La consistencia de los resultados con la información contenida en dichas fuentes fue considerada como un indicador de coherencia. Es preciso señalar que este análisis fue realizado durante la ejecución del procesamiento, de manera tal que su resultado contribuyó con la validación de datos.

VII.3. Aplicación del control de divulgación

La Encla 2019 busca ser un aporte tanto a la discusión de políticas públicas relacionadas al mundo del trabajo en Chile, como también un insumo para las y los investigadores y académicos de las distintas realidades laborales del país. Para esto, tanto el INE como la Dirección del Trabajo deben conciliar la transparencia de sus datos, para que estos efectivamente puedan ser utilizados por investigadores y ciudadanos en general, con el anonimato e indeterminación de sus unidades de análisis y de información.

Calibrar estos dos aspectos resulta importante debido a que la normativa nacional los regula por la Ley N° 20.285, que en su artículo 11 establece que “(...) toda persona goza del derecho a acceder a la información que obre en poder de los órganos de la Administración del Estado (...)”, mientras que, por otro lado, se establece el deber de los particulares de responder las encuestas del INE y el derecho de que se mantenga su identidad en el anonimato, mediante las normativas del secreto estadístico (Ley N° 17.374, artículos 29 y 30) y la protección de la vida privada (Ley N° 19.628, artículo 2).

Pero hay también consideraciones éticas y prácticas que llevan a la necesidad de anonimizar efectivamente, como la protección del patrimonio nacional estadístico: “(la anonimización) es importante porque garantiza la calidad de los datos y las tasas de respuesta en futuras encuestas. Si los encuestados sienten que los productores de datos no están protegiendo su privacidad, es posible que no estén dispuestos a participar en futuras encuestas” (Benschop et al. 2019: 12).

A fin de cumplir con los estándares existentes, se siguieron una serie de procedimientos enfocados en garantizar el anonimato de las empresas y de los informantes en las bases de datos. Para esto, se utilizaron métodos del Control Estadístico del Riesgo (Statistical Disclosure Control, SDC) que

permiten conocer la probabilidad de que las identidades de las unidades de las bases de datos sean descubiertas con variables de identificación indirectas, en base a escenarios hipotéticos en los que intrusos cuentan con datos externos que pueden operar como “llaves” (keys variables) para acceder a la identidad de las empresas de la muestra (Benschop et al. 2019).

Para medir el riesgo de divulgación de las bases se identifica el porcentaje de casos únicos en la muestra según distintas medidas. K-anonimidad identifica cuántos casos con combinaciones únicas de variables hay en la muestra (no utiliza factores de expansión): si un caso tiene 3k de anonimidad significa que ese caso es idéntico a otros 2 de la muestra. Si un caso tiene 10k significa que hay otros 9 como él, lo que lo hace común, poco único y, por ende, poco identificable. El riesgo individual consiste en una medida similar, pero que considera al factor de expansión, mientras que el riesgo global suma todos los riesgos individuales. Los estándares según la literatura (Benschop, Machingauta, and Welch 2019; Templ, Meindl, and Kowarik 2020) y en parte los que ha conversado el INE hasta el momento en su mesa de anonimización, son que no más del 5% de la muestra esté sobre los 3k de anonimidad, que no haya casos con riesgos sobre el 1% y que el riesgo global esté bajo el 5%.

VIII. ALCANCE DE LA PRESENTACIÓN DE RESULTADOS

VIII.1. Productos a difundir

Tanto los microdatos como los metadatos son puestos a disposición de los usuarios a través del sitio web del Instituto Nacional de Estadísticas, como del sitio de la Dirección del Trabajo.

VIII.1.1. Microdatos

Bases de dato: como en versiones anteriores, la Novena Encuesta Laboral Encla 2019 cuenta con cuatro bases de datos oficiales. La primera corresponde a la base de datos del cuestionario a sindicatos, la segunda a la base de datos del cuestionario a trabajadores, la tercera a la base de datos del cuestionario a empleadores y la cuarta a la base de datos del cuestionario autoaplicado.

Las bases serán puestas a disposición de los usuarios en los siguientes formatos: .Rdata, .dta, .sav y .csv. La conversión de los formatos de las bases de datos se realizó mediante el software R Studio, específicamente con el paquete haven (Wickham y Miller 2019).

VIII.1.2. Metadatos

Los metadatos entregan información sobre los datos producidos, describiéndolos, caracterizándolos y permitiendo una mejor comprensión de estos.

- Manuales de usuarios de bases de datos: enfocado en orientar el uso de las bases de datos, contiene una descripción general de las bases de datos y de las variables e indicadores en ellas contenidos.
- Manual metodológico: documentación del proceso estadístico del proyecto.
- Ficha técnica: principales características del estudio.
- Formularios ENCLA.
- Informe de diseño muestral

VIII.2. Difusión de resultados

Los productos mencionados en el apartado anterior se encuentran disponibles en el sitio web de la Dirección del Trabajo, así como en la página institucional del INE:

<https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral/condiciones-de-empleo-y-relaciones-laborales>.

VIII.3. Accesibilidad a los datos y marco legal

En base a la normativa vigente, los datos son puestos a disposición de los usuarios tras la aplicación de los procedimientos definidos para la anonimización e indeterminación de las bases de datos, ciñéndose a lo establecido en:

- 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 que 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.
- Artículos 29 ("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) y 30 (“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”) de la Ley 17.374.

IX. Conceptos Claves

- **Comité bipartito de capacitación:** tiene por objetivos acordar y evaluar el o los programas de capacitación ocupacional de la empresa, así como asesorar a la dirección de la misma en materias de capacitación. Toda empresa con más de 15 trabajadores debe constituir un comité bipartito de capacitación, integrado por representantes de la empresa y de los trabajadores.
- **Comité paritario de higiene y seguridad:** es un comité que tiene por objetivo asesorar e instruir a los trabajadores en materias de higiene y seguridad (correcta utilización de los instrumentos de protección, vigilancia del cumplimiento de las medidas de prevención, higiene y seguridad, investigación de las causas de los accidentes del trabajo y enfermedades profesionales que se produzcan en la empresa, entre otros). Todas las empresas con más de 25 trabajadores deben organizar un comité paritario de este tipo, el que estará integrado por representantes de la empresa y de los trabajadores.
- **Conflicto colectivo:** corresponde a un conflicto laboral que involucra a varios trabajadores, quienes actúan colectivamente para enfrentar al empleador. Conflicto laboral corresponde a un conflicto entre trabajador y empleador por causas propias del proceso de trabajo. Puede involucrar a uno o más trabajadores.
- **Contrato colectivo:** es un acuerdo o documento suscrito por los trabajadores y la empresa en el marco de una negociación y en el que se estipulan las condiciones de trabajo y de remuneraciones pactadas por la empresa y los trabajadores. Este instrumento está sometido a un procedimiento reglado, sometido a plazos y la ley establece los mecanismos a los que los trabajadores pueden recurrir en caso de no llegar a acuerdo con el empleador, como es el caso de la huelga.
- **Contrato individual de trabajo:** es una convención por la cual el empleador y el trabajador se obligan recíprocamente, el último a prestar servicios personales bajo dependencia y subordinación del primero; y este a pagar por estos servicios una remuneración determinada. Es consensuado entre trabajador y empleador y debe constar por escrito.

- **Convenio colectivo:** es un documento que tiene los mismos efectos del contrato colectivo pero se diferencia por el hecho de que a él se llega mediante un procedimiento menos regulado que en el caso del contrato colectivo. Por ejemplo, un convenio puede negociarse en cualquier momento, no da derecho a fuero, no da derecho a huelga, etc.
- **Mecanismos de participación:** corresponden a instancias o iniciativas mediante las cuales la empresa permite la participación de los trabajadores en diversos temas (comités de bienestar, comités de cultura, comités de recreación, comités de capacitación, entre otros).
- **Negociación colectiva:** es un procedimiento regulado por la legislación laboral mediante el cual los trabajadores y los empleadores negocian o acuerdan condiciones de trabajo y de remuneraciones por un tiempo determinado. Los trabajadores pueden estar representados por uno o más sindicatos o por grupos de trabajadores (grupo negociador). El resultado de una negociación colectiva es un instrumento colectivo (contrato o convenio).
- **Reglamento interno de higiene y seguridad:** es un reglamento al que están obligadas todas las empresas, independiente del número de trabajadores que tengan, por la Ley 16.744 (Art. 67) sobre seguro de accidentes del trabajo y enfermedades profesionales y cuyo contenido corresponde a normas sobre prevención y seguridad en el trabajo.
- **Reglamento interno de orden, higiene y seguridad:** es un reglamento al que están obligadas las empresas que tengan 10 trabajadores o más, como resultado del artículo 153 del Código del Trabajo. Debe contener las obligaciones y prohibiciones a que deben estar sujetos los trabajadores, en relación con sus labores, permanencia y vida en las dependencias de la respectiva empresa o establecimiento, así como las normas que deben garantizar un ambiente laboral digno y de mutuo respeto entre los trabajadores.
- **Relaciones laborales:** se llama así al conjunto de relaciones que establecen en la empresa los trabajadores y los empleadores, a través de los órganos de representación de ellos (sindicatos, comités, grupos negociadores, etc.).

X. Bibliografía

- AAPOR. (2016). *Standard Definitions. Final dispositions case codes and outcome rates for surveys*. Recuperado el 20 de Noviembre de 2018, de http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf
- Benschop, Thijs, Cathrine Machingauta, y Matthew Welch. 2019. «Statistical Disclosure Control: A Practice Guide».
- Centro UC Encuestas y Estudios Longitudinales (2016). Imputaciones de corte transversal de la VI ronda de la Encuesta de Protección Social.
- Cohen, S. H. (2003). "Editing strategies used by the US Bureau of Labor Statistics in Data Collection over the Internet". *Statistical Journal of the United Nations Economic Commission for Europe*, 20(3, 4), 265-278.
- Cohen, S. (2006). "*Internet collection at the Bureau of Labor Statistics: an option to report data*". *Monthly Lab. Rev.*, 129, 47
- Cochran, W. (1998). *Técnicas de Muestreo*. México D.F.: Compañía Editorial Continental, S.A. de C.V.
- Department for Business, Innovation & Skills. (2014). "*The 2011 Workplace Employment Relations Study. First Findings*". Ver <https://www.gov.uk/government/collections/workplace-employment-relations-study-wers>
- Dirección del Trabajo (2015). ENCLA 2014. Informe de Resultados Octava Encuesta Laboral. https://www.dt.gob.cl/portal/1629/articles-108317_recurso_1.pdf.
- Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., ... & Ritchie, S. (2019). Package 'data.table'. *Extension of 'data.frame'*. <http://ftp.musicbrainz.org/pub/cran/web/packages/data.table/data.table.pdf>
- Dowle, Matt, Arun Srinivasan, y Jan Gorecki. 2019. data.table: Extension of «data.frame».
- Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., ... & Ritchie, S. (2019). Package 'data.table'. *Extension of 'data.frame'*. <http://ftp.musicbrainz.org/pub/cran/web/packages/data.table/data.table.pdf>
- Eltinge, J. L., & Luery, D. M. (2003). "*Imputation in Three Federal Statistical Agencies*". Presentation at the Federal Economic Statistics Advisory Committee

Edwards, Paul K. 1990. *El conflicto en el trabajo: un análisis materialista de las relaciones laborales en la empresa*. Centro de Publicaciones, Ministerio de Trabajo y Seguridad Social.

Fellegi, I. P., & Holt, D. (1976). "A systematic approach to automatic edit and imputation". Journal of the American Statistical Association, 71(353), 17-35.

Galván, M. & Medina, F. (2007). *"Imputación de datos: Teoría y práctica"* (Vol. 54). United Nations Publications.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). "How many imputations are really needed? Some practical clarifications of multiple imputation theory". *Prevention science*, 8(3), 206-213.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *"Survey methodology"* (Vol. 561). John Wiley & Sons.

Instituto Nacional de Estadísticas (2016). *"Manual de Uso de la IIIa Encuesta Longitudinal de Empresas"*

<https://www.economia.gob.cl/wp-content/uploads/2015/07/MANUAL-DE-USO-DE-LA-BASE-DE-DATOS-PDF1.pdf>

Instituto Nacional de Estadísticas (2015). «Uso del lenguaje inclusivo en la producción estadísticas. Una guía práctica desde el enfoque de género.»

Instituto Nacional de Estadística de España. (2016). *Encuesta sobre innovación a las empresas*. Recuperado de <https://www.ine.es/daco/daco43/metoite2016.pdf>

Lohr, S. L. (2000). *Muestreo Diseño y análisis*. México: International Thomson Editores.

Ministerio de Salud. 2013. «Protocolo de vigilancia de Riesgos Psicosociales en el Trabajo».

OIT. "La relación de trabajo". Recuperado de http://ilo.org/ifpdial/areas-of-work/labour-law/WCMS_165190/lang-es/index.htm.

RESPONSE. (2017). *"Relations professionnelles et négociations d'entreprise"*. Ver <https://dares.travail-emploi.gouv.fr/dares-etudes-et-statistiques/enquetes/reponse2017>

Sigman, R. (2001). *"Editing and Imputation in a Standard Economic Processing System"*. In Proceedings of Statistics Canada Symposium 2001.

Templ, M. (2019). *"Paquete estadístico para R-Studio: 'sdcMicro'"* <https://github.com/sdcTools/sdcMicro>

Templ, Matthias, Bernhard Meindl, y Alexander Kowarik. 2020. *sdcmicro: Statistical Disclosure Control Methods for Anonymization of Data and Risk Estimation*.

Tourangeau, Roger. 1984. «Cognitive sciences and survey methods». *Cognitive aspects of survey methodology: Building a bridge between disciplines* 15:73–100.

Universidad Alberto Hurtado; Servicio Civil (2017). “Encuesta anual para el diagnóstico de las relaciones laborales en la Administración Central del Estado (ENCLACE). Informe Resultados. Primera Aplicación.”

<http://biblioteca.digital.gob.cl/bitstream/handle/123456789/806/2017%20Encuesta%20anual%20para%20el%20diagnostico%20de%20las%20relaciones%20laborales%20en%20la%20AC%20del%20Estado%20%28ENCLACE%29.pdf?sequence=1&isAllowed=y>

United Nations Economic Commission for Europe (2013). “General Statistical Business Process Model, Version 5.0”.

https://statswiki.unece.org/download/attachments/185794796/GSBPM%205_0.docx?api=v2.

Wickham, H. (2019). “Paquete estadístico para R-Studio: ‘haven’”.
<https://www.rdocumentation.org/packages/haven>

Wickham, Hadley, Romain François, Lionel Henry, y Kirill Müller. 2019. *dplyr: A Grammar of Data Manipulation*.

Winkler, W. E. (2018). “Cleaning and using administrative lists: Enhanced practices and computational algorithms for record linkage and modeling/editing/imputation.” *Statistics*, 05.

Wickham, Hadley, y Evan Miller. 2019. *Haven: Import and Export «SPSS», «Stata» and «SAS» Files*.