



UNIVERSIDAD DE CHILE
Facultad de Ciencias Sociales
Escuela de Sociología

Impacto del uso de informantes *proxy* en la Encuesta Nacional del Empleo y la Encuesta de Caracterización Socioeconómica Nacional 2015

Memoria para optar al título de Socióloga

Nombre estudiante : Patricia Iglesias Muñoz
Profesor : Rodrigo Asún

18 de octubre de 2018

Agradecimientos

A mi profesor guía, Rodrigo Asún Inostroza, quien recibió con gran entusiasmo mi tema y que estuvo en todo momento disponible para ayudarme en esta tesis. Agradezco su disposición a hacer que este proceso fuera agradable y de grandes aprendizajes, tal como hace 10 años.

A mis compañeros de trabajo que me asistieron técnicamente: Eduardo Toro, Sebastián Palacios, Daniel Pérez y Klaus Lehmann.

A mis amigos de todos los días: Daniela, Paula, María José, Carlos y Berta. Gracias por su eterno cariño, paciencia y palabras de apoyo. La rutina es hermosa con ustedes cerca.

A mi editora y amiga de y para toda la vida, Paula Bell, y a mi asesor personal y confidente Francisco Espinoza. Siempre será viernes a su lado.

A mis grupos de tesis, quienes fueron mi bastón emocional todos estos años: Daniela, David, Javiera y Nicolás, y a los amigos que estuvieron en alguna parte del proceso: Loreto, Rubén, Sebastián y Paloma.

A mi papá y mamá por hacer todo lo posible por asegurarse de que yo fuera a la universidad y por apoyarme al decidir tomar este camino.

A Juan Manuel Cabrera. Gracias por hacerme querer la sociología, por tenerme esa fe gigante que tienes en mí, por hacer que me cuestionara todo en tiempos en que ciertas cosas no se cuestionaban. Sé de dónde salí, y eso va a estar siempre conmigo.

A Nicolás, mi mejor amigo y compañero hace –suerte la mía– tantos años. Gracias por tu apoyo incondicional y tu lindo amor, tu empuje y por impulsarme siempre a ser mejor. El futuro es incierto, pero hasta ahora es de los dos.

Contenido

Presentación.....	2
Pregunta y objetivos de investigación	4
Pregunta	4
Objetivo general.....	4
Objetivos específicos	4
Marco conceptual	5
Antecedentes.....	19
Rol de la persona informante en la calidad de las respuestas	19
Encuesta Nacional del Empleo y Encuesta de Caracterización Socioeconómica 2015....	27
Marco metodológico.....	33
Depuraciones de bases de datos.....	33
Caracterización de los y las informantes	34
Variaciones en respuestas ENE-EME.....	42
No respuesta parcial CASEN.....	54
Análisis	57
Caracterización de los y las informantes.....	57
Caracterización por tipo de informante	71
Variaciones en respuestas ENE-EME.....	81
No respuesta parcial CASEN.....	105
Conclusiones.....	121
Anexo	127
Pruebas t para igualdad de medias y r para medir tamaño del efecto	127
Bibliografía.....	130

Presentación

En Chile, gran parte de las encuestas públicas se aplican de manera presencial en viviendas particulares. Esto implica producir información a partir de encuestar a distintos tipos de personas, requiriendo la definición de criterios técnicos y operativos para su recolección, tales como agrupar a las personas bajo el concepto de hogar o también la definición del tipo de informante a utilizar.

Esta investigación se centrará en el/la informante *proxy* –aquél que responde por sí mismo/a y otras personas-, por su rol preponderante en las encuestas más importantes del país, así como por la falta de investigación en Chile respecto al impacto que tiene su utilización. El énfasis estará en las eventuales diferencias entre el informante *proxy* y el directo/a, correspondiendo el último al que responde por sí mismo.

Existen dos grandes dimensiones a analizar. La primera corresponde a quiénes están respondiendo las encuestas públicas de hogares. Existe documentación que problematiza la posible existencia de una sobrerrepresentación de determinados grupos sociales entre los/las informantes (quienes tienen la característica común de estar en sus viviendas en determinados horarios y están disponibles para contestar), lo que a su vez podría influir en la calidad del dato.

La segunda dimensión refiere a la calidad de la información que los/las informantes manejan y otorgan en las encuestas. En este aspecto, la investigación internacional entrega resultados contradictorios, no existiendo evidencia concluyente respecto a qué tipo de informante es el más adecuado y sobre los temas en los cuales es preferible usar uno sobre otro. En Latinoamérica, y específicamente en Chile, no existen estudios que aborden esta temática desde las Ciencias Sociales.

Así, en esta investigación, se indagará en la utilización del informante *proxy* y en el impacto que esto tiene en dos encuestas importantes a nivel nacional: la Encuesta Nacional de Empleo (ENE) del Instituto Nacional de Estadísticas (INE), y la Encuesta de Caracterización Socioeconómica (CASEN) del Ministerio de Desarrollo Social. La primera hace preguntas complejas respecto a la participación en el mercado laboral, mientras que la segunda ahonda en esferas como trabajo, salud, educación y vivienda.

Dadas las características de ambos instrumentos, se puede profundizar en los impactos que tiene usar informante *proxy* en ambas encuestas (lo que será abordado de manera más extensa en el marco metodológico). En el caso de la ENE, es posible comparar resultados con la Encuesta Nacional de Microemprendimiento (EME), en consideración de que ambas encuestas tienen preguntas en común y de que ésta se aplica a una población cuyo marco muestral se desprende de los/las trabajadores independientes de la ENE. La EME se levanta con informantes directos, es decir, se define por protocolo que la persona a entrevistar es exclusivamente aquella sobre la cual se está preguntando, aun cuando en la ENE haya sido respondida por otro/a integrante del hogar.

Por su parte, la CASEN caracteriza la participación de cada integrante del hogar durante la aplicación de la encuesta, al final de cada módulo. De esa manera, se puede determinar qué módulos fueron respondidos por la persona sobre la cual se está preguntando y en cuáles respondió otra persona. Además, se trata de una encuesta que integra las respuestas “No sabe” y “No responde”¹, por lo que es factible comparar la no respuesta parcial según tipo de informante. De este modo, se pondrá en evaluación un supuesto implícito en ambas encuestas públicas, las cuales sirven para evaluar y diseñar políticas sociales: que la utilización de informantes de este tipo es inocua. Por esto, el objetivo de esta investigación es cuestionar la utilización de informantes *proxy*, ya que si están entregando información errónea se podría generar una distorsión no identificada entre la realidad y lo que observan las instituciones públicas que ejecutan políticas.

¹ Ambas respuestas se definen como “No respuesta parcial”. Sobre esto se ahondará en el marco conceptual.

Pregunta y objetivos de investigación

Pregunta

¿Cómo se ven afectados los resultados de la Encuesta Nacional del Empleo 2010-2017 y la no respuesta parcial de la Encuesta de Caracterización Socioeconómica Nacional 2015, por el uso de informantes *proxy*?

Objetivo general

Indagar en el impacto que tiene utilizar informantes *proxy* en los resultados de la Encuesta Nacional del Empleo y en la no respuesta parcial de la Encuesta de Caracterización Socioeconómica Nacional 2015.

Objetivos específicos

1. Caracterizar a las personas que son informantes *proxy* versus informante directo/a, de la Encuesta Nacional de Empleo 2010-2017 y la Encuesta CASEN 2015.
2. Describir las diferencias en las respuestas de informantes *proxy* y directos/as en la Encuesta Nacional de Empleo y la V Encuesta de Microemprendimiento.
3. Conocer los porcentajes de no respuesta parcial de informantes *proxy* y directo/a en la Encuesta CASEN 2015.

Marco conceptual

Ante todo, esta tesis se presenta como un aporte sobre cómo es posible generar conocimiento sobre la realidad desde la sociología.

Desde un principio, la sociología ha problematizado el acercamiento que se tiene como disciplina hacia el conocimiento de la realidad. Durkheim (2001) ya lo planteaba de esta manera en “Las reglas del método sociológico”, en donde establece que la reflexión, como ejercicio del investigador o investigadora, es anterior a la ciencia, existiendo conceptos preformados (o preconociones) de manera previa a la observación. Durkheim lo caracteriza como la “intrusión del arte en la ciencia” (2001, pág. 55), y permearía a quienes se propusieran crear conocimiento desde la sociología.

A partir de ese relato, existiría una relación inseparable entre el investigador o investigadora y su objeto de observación. Bourdieu, Chamboredon y Passeron lo revisan en “El oficio del sociólogo”, indicando que “(...) el único criterio de esta pertenencia [a la sociología] reside, en realidad, en la aplicación de los principios fundamentales de la teoría del conocimiento sociológico que, en tanto tal, de ningún modo separa a autores a los que todo aleja en el plano de la teoría del sistema social” (2002, pág. 16). También, desde otras disciplinas científicas, Maturana y Varela plantean que todo lo que se dice es resultado de los procesos de observación y, por ende, todo es dicho por un/a observador/a (2003).

A raíz de esto, se deriva que el conocimiento sociológico de la realidad se encuentra “distorsionado” por el observador u observadora, independiente del tipo de metodología que se utilice. Así, los hechos sociales tendrían siempre un carácter subjetivo y serían irreductibles a los métodos de la ciencia (Bourdieu, Chamboredon, & Passeron, 2002), por lo que se difuminaría la aspiración a ser neutral en la observación y generación de conocimiento

Respecto a esto, Donna Haraway posiciona –desde la perspectiva de la epistemología feminista– el concepto de “conocimiento situado”. Éste sostiene que la producción de conocimiento nunca se encuentra fuera de (o sobre) situaciones históricas y sociales particulares, descartando así la pretensión de neutralidad. Además, propone que el conocimiento surgido desde grupos subalternos que tradicionalmente no están asociados a

su producción, supondría una ventaja en tanto diversifica las miradas desde las cuales se observa la realidad social, pudiendo tener una mejor aprehensión de ella (Piazzini, 2014).

Con el afán de que la investigación sociológica otorgue un conocimiento que se acerque a la realidad, Bourdieu *et al* (2002) consideran necesario reflexionar sobre la práctica científica, sobre “la ciencia que se está haciendo. Tal tarea, propiamente epistemológica, consiste en descubrir en la práctica científica misma, amenazada sin cesar por el error, las condiciones en las cuales se puede discernir lo verdadero de lo falso, en el pasaje desde un conocimiento menos verdadero a un conocimiento más verdadero” (pág. 20). En esta línea, se deben “someter las operaciones de la práctica sociológica a la polémica de la razón epistemológica, para definir, y si es posible inculcar, una actitud de vigilancia que encuentre en el completo conocimiento del error y de los mecanismos que lo engendran uno de los medios para superarlo” (pág. 14).

Desde ese planteamiento surge el concepto de *vigilancia epistemológica*, el cual iría en contra de la aplicación mecánica y poco reflexiva de preceptos y metodologías. Sugiere someter el uso de técnicas y conceptos a un “examen sobre las condiciones y los límites de su validez”, en tanto que “toda operación, no importa cuán rutinaria y repetida sea, debe repensarse a sí misma y en función del caso particular” (pág. 16). Este ejercicio sería propio del “‘oficio’ del sociólogo”, en tanto *habitus* que “no es sino la interiorización de los principios de la teoría del conocimiento sociológico” (pág. 16), y a la vez, permitiría “someter las verdades próximas a la ciencia y los métodos que utiliza a una rectificación metódica y permanente” (pág. 14).

Considerando esta discusión teórica, el suponer que las metodologías cuantitativas son objetivas porque ocupan números es erróneo, y es que el debate sobre la realidad interpretada y construida subjetivamente también permea a los métodos cuantitativos (Vidal, 2013). Éstos podrían presentar sesgos en la etapa de diseño de la investigación (definición de preguntas, objetivos, metodologías y elaboración de cuestionarios), en el trabajo de campo o en el análisis de las bases de datos producidas.

Para entender el principal instrumento utilizado por la metodología cuantitativa –las encuestas–, se debe tener claridad sobre qué exactamente. Éstas se definen como un método sistemático para reunir información de ciertas entidades con el propósito de construir

descriptores cuantitativos de los atributos de la población entera de la cual las entidades son miembros (Groves, y otros, 2009). En particular, las encuestas en hogares (como lo son la ENE y la CASEN) son relevantes ya que dan cuenta de indicadores demográficos y socioeconómicos vigentes sobre los hogares y las personas, lo que las convierte en un “elemento indispensable en el análisis de las políticas económicas y sociales, la planificación del desarrollo, la gestión de programas y la toma de decisiones en todos los niveles”, teniendo particular importancia para los países en desarrollo y en transición (Naciones Unidas, 2007, pág. 35)

En función de los objetivos de esta investigación, el foco estará puesto en la decisión metodológica de utilizar informante *proxy*. Esta instrucción se decide de manera previa al trabajo de campo, pero tiene una incidencia relevante al momento de aplicar la encuesta. En ese momento existen dos fuentes de tensión: la persona que aplica la encuesta y la(s) persona(s) que la contestan. Aquí, como en toda relación social, se interponen dos sistemas de prenociones que tratan de ser objetivados en un instrumento de medición. Por tanto, el *habitus* del encuestador/a y del encuestado/a inciden en el resultado que se imprime en el formulario.

En relación con la persona que aplica la encuesta, el efecto que tiene en la cantidad y la calidad de las respuestas se relaciona con distintos factores. Cea (2005) visibiliza estudios que indican que tanto variables sociodemográficas como el género, la edad, el nivel educativo, el estatus socioeconómico, la pertenencia indígena, la religión, así como la “actuación” del encuestador/a y las expectativas que tiene al momento de realizar la entrevista, tienen efectos en la disposición a responder por parte de la persona informante.

Por su parte, respecto a la(s) persona(s) que contesta(n) la encuesta, los sesgos tienen distintas fuentes. Una de ellas es el tipo de informante, el que puede ser directo, es decir, cuando, por decisiones metodológicas previas al levantamiento, debe responder directamente el sujeto sobre el cual se busca información; o *proxy*, definido como la persona que responde sobre otra que es parte de la muestra (The American Association For Public Opinion Research, 2016).

La definición de “*proxy*” es a partir de la figura utilizada en encuestas de hogares (una persona que contesta por otra). En este punto, cabe problematizar el concepto de “idóneo”

en español versus el uso de “*proxy*” en inglés para referirse a la persona que informa, sobre todo teniendo en consideración que tanto la ENE como la CASEN ocupan el primero.

En español, “idóneo” se define como “adecuado y apropiado para algo” (Real Academia Española, 2018). Por otro lado, el significado de “*proxy*” se traduce como “la autoridad de representar a otra persona, especialmente para votar”, “una persona autorizada para actuar en nombre de otra” (Oxford Dictionaries, 2018)², “autoridad o poder de actuar por otro” y más puntualmente “persona autorizada a actuar por otra” (Merriam-Webster, 2018)³.

Dado que el objetivo de esta tesis es problematizar la utilización de este tipo de informante, se optará por llamarle “*proxy*” a lo largo del documento. Además de ser inclusivo en términos de género, resulta más pertinente al no conocer aún la idoneidad o adecuación que tiene ese tipo de informante en las encuestas ENE y CASEN, más allá de estar en la posición de responder por otra persona.

Además, es importante destacar que en ambas encuestas el/la informante siempre es *proxy* (antes idóneo/a): el procedimiento operativo no indica que se debe encuestar a la persona sobre la cual se está preguntando, y de suceder así, no es referida como informante directo/a. Por ende, se entiende que el/la informante *proxy* es aquél que responde por sí mismo/a y/o por las otras personas que componen el hogar.

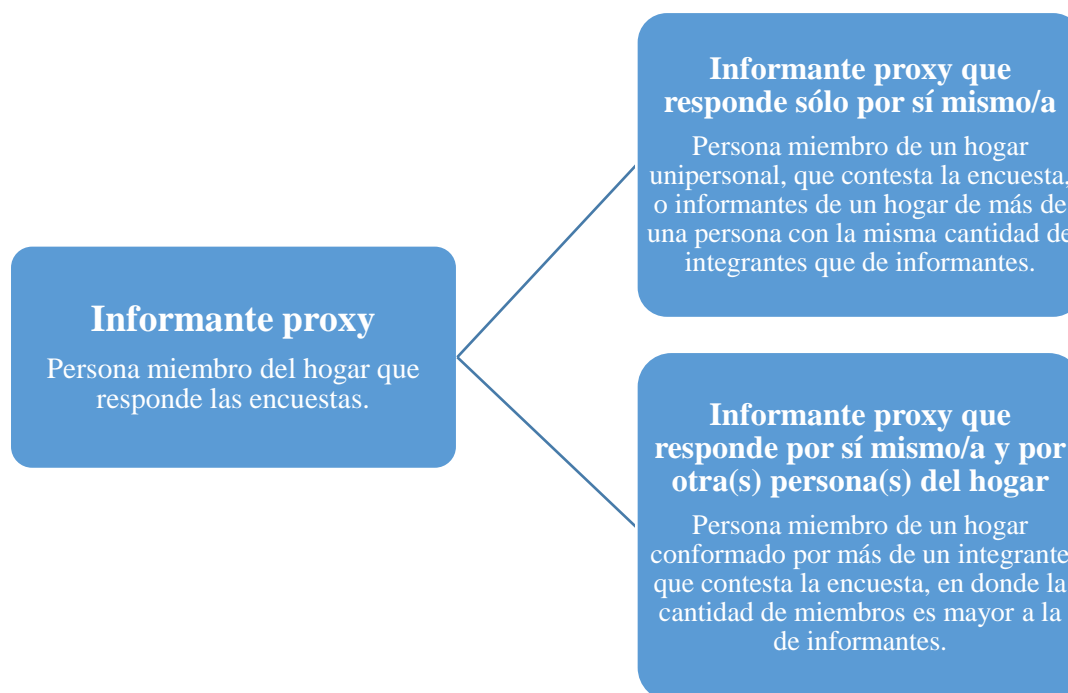
Así, los/las informantes *proxy* se pueden categorizar como:

- a. Informantes *proxy* que responden solo por sí mismos/as.
- b. Informantes *proxy* que responden por sí mismos/as y por otra(s) persona(s) del hogar.

² En inglés: “The authority to represent someone else, especially in voting” y “A person authorized to act on behalf of another”.

³ En inglés: “Authority or power to act for another” y “a person authorized to act for another”.

Figura 1: Categorías del informante proxy



Fuente: Elaboración propia a partir de los procedimientos operativos y definiciones conceptuales de la ENE y CASEN.

Hecha esta distinción, con el fin de facilitar la comprensión del texto y de estar en línea con el lenguaje utilizado en la bibliografía, se denominará “informante directo/a” al informante *proxy* que responde sólo por sí mismo/a.

Para hacer un acercamiento a las implicancias en la calidad de las respuestas de un/una informante *proxy*, es importante considerar los procesos cognitivos existentes al momento de responder preguntas de una encuesta.

De acuerdo con Roger Tourangeau en su texto “*Cognitive Sciences and Survey Methods*”⁴ (1984), en el contexto de la encuesta, la persona informante debe entender la pregunta, recordar los hechos relevantes, hacer un juicio si es que la pregunta lo solicita, y seleccionar una respuesta. Estos cuatro pasos los denomina:

- a. Comprensión: cuando se interpreta la pregunta.
- b. Recuperación: cuando se recupera la información necesaria para responder la pregunta.

⁴ Ciencias Cognitivas y Métodos de Encuesta

- c. Juicio: cuando se combina o resume la información recordada.
- d. Reporte: cuando se formula la respuesta y se pone en el formato requerido.

El autor plantea que, conocidos estos procesos, existe amplia cabida para los errores: las personas informantes pueden entender mal la pregunta o las categorías de respuesta, pueden olvidar o recordar mal información crucial, pueden juzgar mal la información que recuerdan, así como pueden reportar de manera incorrecta sus respuestas. Tourangeau indica además que no necesariamente los cuatro pasos siguen el orden determinado, así como tampoco suceden en su totalidad en todas las ocasiones.

Estos procesos, además, están relacionados con la no respuesta parcial. Como lo definen Beatty y Herrmann (2002), cuando las personas informantes se enfrentan a la pregunta de una encuesta deben decidir entre dos caminos: responder o no responder. La no respuesta parcial resulta cuando resuelven de manera negativa en cualquiera de los dos casos. Así, la no respuesta parcial ocurre cuando falta la respuesta a una pregunta (Groves, y otros, 2009).

Beatty y Herrmann problematizan la no respuesta parcial: desde un punto de vista estadístico, la ausencia de respuesta crea dificultades analíticas debido a que disminuye el tamaño de la muestra efectiva. Por ende, la no respuesta parcial debería ser evitada. Sin embargo, desde una perspectiva psicológica es más complejo plantear si la no respuesta es deseable o no: se busca que las personas informantes provean información que tenga un cierto nivel de precisión cognitiva. Si ese estándar no se cumple, la no respuesta parcial sería deseable.

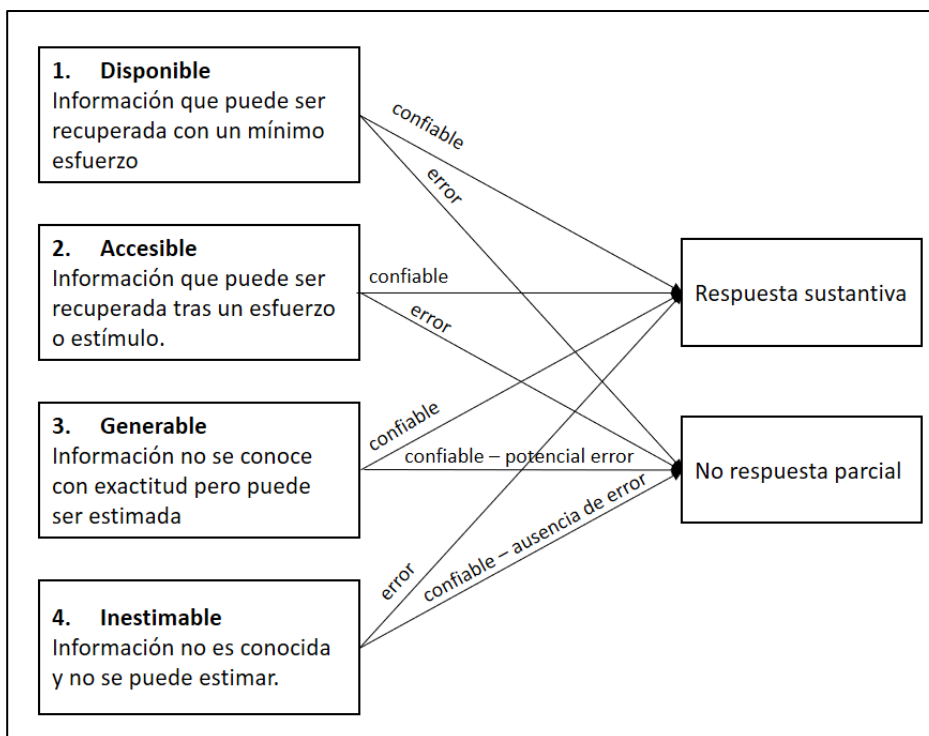
Los autores sostienen que la decisión de responder o no a una pregunta se ve motivada por los siguientes factores (Beatty & Herrmann, 2002):

- a. Estado cognitivo: el conocimiento de las personas informantes se podría clasificar en cuatro estados cognitivos:
 - Disponible: la información solicitada puede ser evocada con un mínimo esfuerzo.
 - Accesible: la información solicitada puede ser evocada tras un esfuerzo o estímulo.

- Generable: la información solicitada no es conocida con exactitud, pero se puede estimar usando otra información de la memoria.
- Inestimable: la información solicitada no es conocida y no se puede estimar.

Esta clasificación y su relación con la no respuesta son visibles en el siguiente esquema:

Figura 2: Procesos de decisión relacionados a la no respuesta parcial, según estado cognitivo



Fuente: Traducción propia de Groves y otros (2009), *Survey Nonresponse*⁵.

Los autores establecen, por ende, que cualquier estado cognitivo puede llevar a una respuesta sustantiva o a una no respuesta parcial, y que, como muestra la figura, se pueden presentar tanto errores de comisión (dar una respuesta sin el conocimiento suficiente) como de omisión (no dar una respuesta cuando existe conocimiento de ella).

- b. Juicios de adecuación: éstos son necesarios cuando una potencial respuesta contiene incertidumbre, es decir, fue estimada o adivinada. Los autores plantean que en los estados cognitivos medios (accesible y generable), al dar una

⁵ No Respuesta en Encuestas

respuesta, puede no ser claro para la persona informante si la información cumple o no los requerimientos de la pregunta, lo que la llevará a responder la pregunta o a dar una no respuesta parcial.

La dificultad de este juicio va a depender del grado de incertidumbre involucrado, de la complejidad de la pregunta y/o de preguntas con guías poco claras en relación al nivel de certeza que requiere una respuesta.

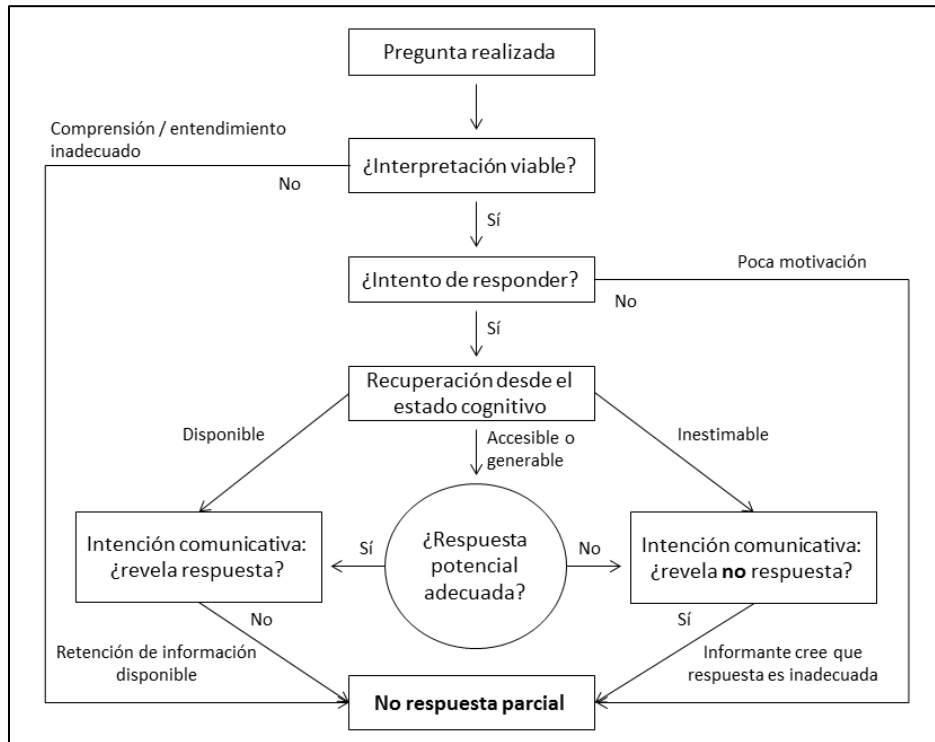
- c. Intención comunicativa: se define como la decisión de qué responder. Las personas informantes pueden decidir no dar una respuesta sustantiva por distintas razones: no quieren esforzarse en recordar y reportar información aun cuando podrían hacerlo, o bien piensan una respuesta, pero deciden no entregarla.

Las personas informantes con una intención comunicativa negativa (es decir, no dan respuesta) pueden decir que no saben, negarse a responder, o dar una respuesta sustantiva que saben que es incorrecta. De esta manera, la intención comunicativa juega un papel determinante tanto en los errores de comisión como en los de omisión.

La intención comunicativa dependerá del interés en el contenido de la encuesta, de la complejidad de las preguntas, de la extensión del instrumento, y del riesgo percibido en responder o no responder.

La forma en la que se articulan el estado cognitivo, los juicios de adecuación y la intención comunicativa, para llegar a la no respuesta parcial, se hace visible en el siguiente esquema:

Figura 3: Modelo de decisión para la no respuesta parcial



Fuente: Traducción propia de Beatty y Herrmann (2002), *To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse*⁶.

El proceso funciona de la siguiente manera (Beatty & Herrmann, 2002):

1. Se inicia cuando se le hace una pregunta a la persona informante, quien debe interpretar de manera viable lo que se le está preguntando. En ese momento surge la primera oportunidad para la no respuesta parcial, si es que la persona no entendió la pregunta. Si la interpreta correctamente, el proceso puede continuar.
2. Luego, la persona informante debe decidir si hacer o no el esfuerzo necesario para responder, lo que se transformaría en la primera decisión relacionada a la intención comunicativa. Si es que no existe motivación o ésta no es suficiente, la persona puede detener el proceso de respuesta.
3. Si continúa, la persona informante debe “rescatar” cualquier información disponible que sea relevante para la pregunta. Cuando la recuperación entrega información de los estados cognitivos extremos (disponible o inestimable), la intención comunicativa será la única determinante de la no respuesta parcial: si está

⁶ Responder o No Responder: Procesos de Decisión Relacionados con la No Respuesta Parcial

disponible, la única explicación para la no respuesta es la intención comunicativa de no-divulgación. Si es inestimable (es decir, no sabe la respuesta), la no respuesta parcial es el resultado más certero. Si la información de la persona informante se ubica en los niveles accesible o generable, debe evaluar la calidad de la eventual respuesta, y en ese caso puede que personas con el mismo nivel de conocimiento decidan opciones diferentes (responder o no responder).

4. Luego del juicio de calidad de la información, las personas informantes continúan con la decisión de la intención comunicativa, lo que puede llevar a una respuesta sustancial o a una no respuesta parcial.

Además, el esquema presentado por Beatty y Herrmann entrega cuatro explicaciones para la no respuesta parcial:

1. Comprensión inadecuada de la pregunta.
2. Baja motivación.
3. Una decisión de retener información aun cuando ésta cumpliría con los objetivos de la pregunta.
4. Convicción de que la no respuesta parcial es la opción más precisa, dada la precisión que solicita la pregunta.

Este modelo de no respuesta parcial, consecuentemente, entrega dos caminos posibles: uno, determinar estrategias para disminuir la no respuesta parcial, y dos, promover la no respuesta parcial cuando la información no cumple con los estándares necesarios. En línea con esto, las personas investigadoras tras una encuesta, tienen cuatro opciones para controlar la no respuesta parcial (Beatty & Herrmann, 2002):

1. No ofrecer opciones de no respuesta parcial y abstenerse de aceptarlas mediante indagación por parte del encuestador o encuestadora (como es el caso de la ENE).
2. No ofrecer opciones de no respuesta parcial (“No sabe”), pero aceptarlas cuando éstas sean reportadas (así opera la encuesta CASEN).
3. Ofrecer explícitamente “No sabe” como una opción.
4. Incentivar a las personas informantes que no respondan preguntas a menos que confíen en la precisión de sus respuestas.

Si bien desde esta tesis no se podrán determinar los motivos de no respuesta parcial (ya que eso implicaría ahondar en otras materias distintas a las que aquí se presentan), sí es importante tener presentes estos esquemas, ya que uno de los factores de comparación entre los distintos tipos de informante será precisamente la tasa de no respuesta parcial en ellos. En función de los resultados, se podría considerar la pertinencia de sugerir una línea investigativa a futuro.

Ya conocidos los procesos cognitivos complejos que tienen lugar al momento de responder una encuesta, es importante profundizar puntualmente en otros sesgos que se presentan en el levantamiento.

Un sesgo importante proveniente de la persona informante –tanto *proxy* como directo(a)- es el de deseabilidad social. Éste ocurre cuando una persona informante responde lo que es aceptado socialmente en lugar de su comportamiento o actitud real, para evitar la vergüenza, ansiedad y angustia que podría traer el revelar preguntas socialmente indeseables (Kaminska & Foulsham, 2013).

Según Tourangeau, este sesgo tiene lugar en la última etapa del proceso cognitivo: la verbalización de la respuesta. Kaminska y Foulsham (2013) lo complementan al establecer que cuando sucede, la persona informante entendió la pregunta, recuperó la información de su memoria de forma adecuada y la integró bien, y por ende se asume que la persona llegó a la respuesta correcta y está consciente de aquello, por lo que el sesgo de deseabilidad social sucede solo cuando, de manera deliberada, se entrega una respuesta incorrecta al final.

Bound, Brown y Mathiowetz (2000), por su parte, lo caracterizan como un proceso dinámico, y que depende del tópico sobre el que se pregunta, del contexto social inmediato, y del ambiente social más amplio en el momento en que se hizo la pregunta. Los autores establecen que este tipo de sesgo se produce a partir de cierto tipo de preguntas, que fomentan patrones de sub-respuesta para comportamientos y actitudes socialmente indeseadas, así como de sobre-respuesta para aquéllas que sí son deseadas. Según los/as autores/as, algunos de los tópicos más proclives son los relacionados a ingresos, participación en elecciones, prácticas sexuales, uso de drogas, discriminación racial, entre otros.

Es importante saber que el sesgo de deseabilidad social no opera de igual manera en toda la población. Según Cea D'Ancona (2005), este sesgo es más habitual en informantes con nivel educativo más alto (al percibir mayor intencionalidad en la pregunta), así como en sociedades hispanas, que se ven más afectadas por las apariencias, mientras que es menor en la población más joven. Como se revisará más adelante, las personas informantes en ambas encuestas tienden a concentrarse en rangos mayores de edad, y al menos un tercio cuenta con educación superior. Eso, en conjunto con las variables culturales indicadas por la autora, hacen que las personas informantes de la ENE y CASEN sean más proclives a cometer este sesgo en las preguntas que consideren sensibles.

También en el levantamiento se pueden generar sesgos relacionados a los conceptos utilizados. Campanelli, Rothgeb y Martin (2005) problematizan el uso de conceptos que no son explicados a las personas informantes, como “trabajo”, “empleo”, “empresa” y “en busca de trabajo”, ya que, si bien son conceptos de uso diario, no existe claridad de que los significados de éstos sean por un lado simples, y por otro que sean compartidos de manera universal por todas las personas. Esto es crítico, ya que los investigadores e investigadoras confían en la comprensión de los significados de estos conceptos que tienen las personas informantes (Campanelli, Rothgeb, & Martin, 2005).

Por otro lado, en el trabajo de campo suele quedar a discreción del personal de levantamiento qué integrante del hogar será informante, existiendo espacio para que sea aquella persona que está en ese momento en la vivienda y tiene disponibilidad para contestar, por sobre aquella que estaría más capacitada para hacerlo. Esto puede inducir al “sesgo de auto-selección”: las personas que están en la vivienda y tienen disponibilidad suelen tener características distintas de aquellos que están ausentes, las cuales a su vez están generalmente correlacionadas con el tipo de información recolectada (Bardasi, Beegle, Dillon, & Serneels, 2011). Así, las diferencias existentes entre informantes *proxy* y el resto de miembros del hogar, reflejan los diferentes atributos que poseen las poblaciones disponibles para ser encuestadas y las que no lo están (Biggs, 1992).

Esto, junto al manejo conceptual por parte de las personas informantes que se mencionó anteriormente, se torna particularmente problemático para temáticas como mercado laboral (que constituye prácticamente la totalidad de preguntas de la ENE y es un módulo completo

de la CASEN), en tanto las personas que se encontrarían en las viviendas en los horarios de levantamiento tenderían a ser mayormente desocupadas o inactivas, y no ocupadas. Esta situación conllevaría a dos problemas: las respuestas sobre las personas ocupadas (que involucran características del empleo, horas de trabajo, ingreso, disposición a trabajar más horas, etcétera) estarían siendo entregadas por personas no ocupadas, lo que podría causar problemas a partir del uso de los conceptos y por el sesgo de auto-selección.

Finalmente, y como plantea Moore (1988), las encuestas de hogares que recolectan datos sobre cada miembro seleccionado en cada hogar de la muestra, requieren una decisión adicional respecto al resto de las encuestas: cuánto esfuerzo gastar reuniendo datos sobre individuos desde cada individuo. En ese contexto, el autor plantea la pregunta sobre en qué medida se sacrifica la calidad cuando se opta por reglas que permiten la respuesta *proxy*.

Los argumentos en general se relacionan a los recursos económicos con los que se cuenta para el levantamiento de encuestas: las que utilizan en mayor medida el informante directo son más caras que aquéllas que tienen una mayor tolerancia al informante *proxy*, y para productos de alto alcance (como serían la ENE y la CASEN), las diferencias en dinero pueden ser importantes. Además de incurrir en mayores gastos, como indican Bardasi, Beegle, Dillon y Serneels (2011), la aplicación de encuestas en informantes directos/as hace el trabajo de campo más laborioso.

Un elemento para aportar al debate, se relaciona con la idoneidad del tipo de informante *proxy* o el directo. Si bien lo intuitivo es atribuir que el/la informante directo/a es quien podría dar las respuestas más cercanas a la realidad, no siempre se puede asumir esa suposición. Biggs (1992) plantea directamente que si dos tipos de informantes proveen respuestas distintas, no se puede afirmar que una es mejor que la otra, sino que esto solo se podría afirmar a través de la corroboración mediante registros administrativos.

Moore (1988), a su vez, plantea que aunque se tendería a confiar más en los/las informantes directos/as, ciertas instancias específicas pueden aportar a que éste no responda la verdad, en particular cuando la temática de la encuesta provoque presiones relacionadas al deber ser de las personas (lo que, a su vez, es una manifestación del sesgo de deseabilidad social).

Bardasi y otros (2011) vuelven a problematizar la supuesta idoneidad del tipo directo de informante por sobre el *proxy*, dando cuenta que no existen evidencias de aquello, en tanto se han identificado razones específicas que justifican las potenciales discrepancias que podrían haber entre ambos tipos de informantes. Al igual que Biggs, recomiendan como única forma de verificación el uso de registros administrativos.

Antecedentes

Los estudios dirigidos al impacto que tienen los distintos tipos de informantes en encuestas relacionadas a las ciencias sociales (y en particular, al mercado laboral), se realizaron principalmente en las décadas de los 70's, 80's y 90's del siglo XX. La mayoría de éstos corresponden a estudios realizados en Estados Unidos, y giran en torno a las diferencias de las respuestas dadas por un(a) informante *proxy* y uno(a) directo(a), encontrando tipos de preguntas que son más sensibles que otras a que estas discrepancias ocurran. Además, identifican las diferencias respecto a la frecuencia de la no respuesta parcial.

Rol de la persona informante en la calidad de las respuestas

En cuanto a la no respuesta, Moore (1988) en “*Self/Proxy Response Status and Survey Response Quality: A Review of the Literature*”⁷, señala que existe evidencia de que el tipo de informante (*proxy* o directo/a) tiene efectos en la completitud de las respuestas. En particular, indica que el auto-reporte –que refiere al informante directo/a que produce mayores tasas de no respuesta de hogares y personas, pero que tienen una menor tasa de no respuesta parcial (comparado con el *proxy*).

Biggs (1992) adhiere a Moore en su trabajo “*Self/Proxy Respondent Rules and Data Quality*”⁸, donde, tras realizar una revisión de distintos estudios, afirma que efectivamente el tipo de informante *proxy* tiene una mayor no respuesta parcial, y menor no respuesta a nivel de persona y hogar.

Además, en referencia a cuáles son los temas más sensibles para la no respuesta parcial en informantes *proxy*, Biggs (1992) identifica que en la *Survey of Income and Program Participation*⁹ (SIPP) de Estados Unidos, la no respuesta parcial se concentra en las variables de características del trabajo principal (como horas por semana y salario por hora).

En la misma línea y de manera anterior, Martin y Butcher (1982) analizaron la *Labour Force Survey*¹⁰ y la *National Dwelling and Housing Survey*¹¹ (ambas de Reino Unido),

⁷ Respuestas Directa y *Proxy* y Calidad de la Respuesta en Encuestas: Una Revisión de la Literatura

⁸ Reglas de Informantes Directos/as y *Proxy* y Calidad del Dato

⁹ Encuesta de Ingresos y Participación en Programas

¹⁰ Encuesta de Fuerza de Trabajo

concluyendo que la tasa de no respuesta parcial (que incluyó los “no sabe” y las respuestas imposibles de codificar) fue mayor en informantes *proxy* que en los casos en que se reportó directamente, específicamente en las preguntas de fuerza de trabajo e ingresos.

En términos de la calidad de la respuesta, tampoco existe una posición única. Se plantean ítems más problemáticos que otros, así como tipos de preguntas que podrían ser más dificultosos según el tipo de informante (Martin & Butcher, 1982).

Biggs, en su estudio “*Self/Proxy Respondent Rules and Data Quality*” (1992), analizó una cantidad considerable de investigaciones dirigidas a este respecto. Su principal conclusión fue que las respuestas *proxy* subestiman la participación en programas de apoyo a los ingresos del gobierno, ofrecen respuestas más consistentes a asuntos sensibles, y tienen una dificultad mayor al reportar detalles y eventos de corta duración. Por otro lado, indica que, como se problematizó anteriormente, la evidencia no es conclusiva sobre las diferencias existentes en la calidad de la información dada según tipo de informante, y que si bien no hay que rechazar el uso de informantes *proxy*, se debe mantener control sobre las respuestas dadas por estos.

Todos los estudios analizados por Biggs ocupan una de estas tres metodologías (1992):

1. Aproximación experimental: se obtienen dos respuestas (una *proxy* y una auto-respuesta) a las mismas preguntas, dentro del mismo hogar.
2. Estudios de validez: se usa información externa más precisa (como registros administrativos), para evaluar de mejor manera la calidad del dato.
3. Acercamiento no experimental de inconsistencia de la respuesta: evalúan las respuestas *proxy* en función del grado de correspondencia entre respuestas auto-reportadas y *proxy*, o por la consistencia de la respuesta en el tiempo según el tipo de informante. Estos estudios no controlan el sesgo de auto selección, y tampoco se utilizan registros administrativos.

En cuanto a los estudios con enfoque experimental y de validez, los resultados fueron los siguientes:

¹¹ Encuesta Nacional de Vivienda y Alojamiento

Tabla 1: Resultados de estudios experimentales y de validez que evalúan la calidad del dato según tipo de informante

Enfoque	Encuesta/Estudio	Características	Resultados
Experimental	<i>Survey of Income and Program Participation (SIPP)</i> . 1984	Muestra pequeña (167 casos), pérdida de 35 casos por una falla en el procesamiento, alta tasa de no respuesta.	En 25% de los casos, el informante <i>proxy</i> identificó mal la información de sueldos.
			Los auto-informantes tendían más a dar una respuesta, especialmente para características del trabajo principal como horas a la semana y pago por hora Los <i>proxy</i> eran más proclives a reportar empleos relacionados con pagos más altos y número de horas trabajadas más alto.
	<i>Interviewing Methods Study</i> ¹² (Reino Unido), 1982	Muestra de 999 hogares, 1.324 respuestas pareadas, 70% eran marido y mujer y sobre ellos se presentaron los resultados.	VARIABLES de empleo como ocupación, momento del pago, horas trabajadas la semana anterior, tenían menos de un 10% de desacuerdo. El mayor desacuerdo (16%) fue en días trabajados la semana pasada.
			Las variables de ingreso ocupadas tuvieron más inconsistencia y una proporción mayor de <i>proxy</i> que no conocían las respuestas. Respuestas de ingreso con bandas de 50 libras tuvieron un 20% de error, mientras que en bandas de 1 libra el error fue de 65%.

¹² Estudio de Métodos de Entrevista

Enfoque	Encuesta/Estudio	Características	Resultados
			El porcentaje de desacuerdo entre el <i>proxy</i> y el directo/a de los ingresos aumenta con el nivel de ingresos.
			Las variables de vivienda tuvieron un mayor grado de correspondencia entre ambos informantes. Sin embargo, el pago de arriendo o hipoteca fueron inconsistente en el 20% (una libra de error por semana) y 12% de proxies no sabían.
Validez	<i>SIPP</i> , 1990	Se usaron registros administrativos para ocho programas gubernamentales para evaluar la calidad de los datos de partición en dos olas del <i>SIPP</i> . Muestras de 5000 casos	Ninguno de los cruces mostró diferencias significativas.

Fuente: Sistematización y traducción propia de “*Self/Proxy Respondent Rules and Data Quality*” (1992) de Brian Biggs

Por otro lado, están los estudios de acercamiento no experimental de inconsistencia de la respuesta:

Tabla 2: Resultados de estudios experimentales que evalúan la calidad del dato según tipo de informante

Encuesta/Estudio	Características	Resultados
-------------------------	------------------------	-------------------

Encuesta/Estudio	Características	Resultados
<i>Labor Force Survey</i> (Canada), 1980-1984	Se realizaron re-entrevistas a un 2% de la muestra de la LFS con diferentes encuestadores con una semana de diferencia, con miembros del hogar que supieran de los temas. Se comparó según tipo de informante con una submuestra en donde la respuesta original no estaba disponible para el encuestador.	No se encontró una relación evidente entre las respuestas de los informantes <i>proxy</i> y la inconsistencia en las respuestas.
		En la mayoría de las preguntas, las tasas de inconsistencia cuando había un cambio de informante (de <i>proxy</i> a directo/a y viceversa), eran marginalmente mayores que en los casos en que no había cambio.
		Al comparar la distribución de las respuestas cuando había un cambio en el tipo de informante, no se encontró un efecto <i>proxy</i> para ninguna variable.
<i>Labor Market Activity Survey</i> ¹³	Se buscó evaluar el impacto del tipo de informante en la incidencia de los trabajos sin confirmar ¹⁴ . El encuestador evaluaba la habilidad de responder del <i>proxy</i> , y de ser necesario se contactaba a la persona sobre la cual se estaba preguntando.	Los empleos no confirmados eran el 2% y 1,6% del total de empleos reportados de informante <i>proxy</i> y no <i>proxy</i> , respectivamente. No se concluyó que los reportes <i>proxy</i> fueron un factor que constituyera problemas.
<i>Income Survey</i>	Se observaron las diferencias	Informantes <i>proxy</i> tenían tasas de no

¹³ Encuesta de Actividad en el Mercado Laboral

¹⁴ Aquellos empleos que fueron reportados como vigentes al final de un período de referencia, pero que no eran subsecuentemente reportados al inicio del siguiente periodo de referencia.

Encuesta/Estudio	Características	Resultados
<i>Development Program</i> ¹⁵ (ISDP), 1980	en las tasas de no respuesta de informantes <i>proxy</i> y auto-informantes.	respuesta más altas para todas las fuentes de ingresos reportadas. Casi un cuarto de los informantes <i>proxy</i> no entregó información sobre el salario por hora, comparado al 5% de no-proxies. Las tasas de no respuesta eran más altas para los ingresos de auto-empleo.
<i>SIPP</i>	Dos estudios distintos buscaron evaluar la tasa de no respuesta parcial según tipo de informante.	El primer estudio (1983) definió que las tasas de no respuestas eran más altas para informantes <i>proxy</i> en variables de salario, ingresos por auto-empleo, y pago por hora. Las tasas fueron de 22% para <i>proxy</i> y 10% para auto-respuesta. El segundo estudio (1990) identificó que la no respuesta parcial en la SIPP era uniformemente más alta para los informantes <i>proxy</i> . Se identificaron diferencias considerables en la no respuesta en 13 de 36 categorías de ingresos, siendo las tasas de no respuesta de los <i>proxy</i> el doble que las de los auto-informantes.
<i>U.S. Current Population Survey</i>	Como parte de los chequeos regulares de calidad del dato,	Las tasas de inconsistencia eran menores (2% a 4%) cuando

¹⁵ Programa de Desarrollo de Encuestas de Ingresos

Encuesta/Estudio	Características	Resultados
<i>(CPS)</i> ¹⁶ , 1986	se reentrevistó a una muestra de 5.000 viviendas, con encuestadores con más experiencias. Para el 20% de esta submuestra, la respuesta original no estaba disponible para el encuestador. Los resultados presentados corresponden a los casos con respuestas distintas en un período de dos años.	informantes <i>proxy</i> no estaban involucrados, y eran mayores (6% a 12%) cuando las respuestas eran dadas por dos informantes <i>proxy</i> diferentes.

Fuente: Sistematización y traducción propia de “Self/Proxy Respondent Rules and Data Quality” (1992) de Brian Biggs

Por su parte, Daniel Hill (1987) en “*Response errors in labor surveys: comparisons of self and proxy*”¹⁷, analizó la Encuesta *SIPP* de Estados Unidos. Acotó el análisis a los hombres entre 25 y 55 años de la muestra, y solo a los casos en que se proveía la totalidad de la información solicitada. Bajo este modelo, comparó los resultados según el tipo de informante que había entregado la información.

En primer lugar, se concentró en cómo se compone el conjunto de informantes *proxy*, identificando que:

- Entre más limitado sea el tiempo que pasa en la vivienda la persona, menores son las probabilidades de que responda por sí misma. Así, un hombre que trabaja 40 horas a la semana –fuera de su vivienda– tendría 6,5% menos de probabilidades de responder por sí mismo, comparado con otros hombres que no trabajan.
- Las personas informantes *proxy* son menos educados que las y los informantes directos.

¹⁶ Encuesta de Población Actual de Estados Unidos

¹⁷ Errores de Respuesta en Encuestas de Empleo: Comparaciones de informantes directo/a y *proxy*.

- Los hombres casados tienen 30% menos probabilidades de reportar por sí mismos, respecto a los no casados.

En cuanto a las diferencias en los resultados, el autor indica que existe un sesgo relacionado al tipo de informante, así como una varianza del error entre los ingresos reportados por informantes *proxy* y mediante auto-reporte. Este último tipo de informante daría cuenta de ingresos 20% más bajos.

La encuesta de empleo del Reino Unido (la *Labor Force Survey*), en su guía del usuario del año 2016 llamada “*LFS Background and Methodology*”¹⁸, acreditan el uso del informante *proxy* basándose en el estudio “*A study of proxy response in the Labour Force Survey*”¹⁹ de Fiona Dawe e Ian Knight (1997), el cual concluye que:

- El nivel de concordancia entre las respuestas dadas por informantes *proxy* y las auto-respuestas fue de 80% para gran parte de las variables claves, habiendo varias con una concordancia mayor al 90%.
- Las variables que requerían menos información directa (como el nivel más alto de estudio), y aquellas que requerían información numérica detallada (como las horas trabajadas e ingresos) mostraron menos coincidencias entre los dos tipos de informantes, lo cual aumentaría las tasas brutas de error.
- La confianza en los datos obtenidos por informante *proxy* para las variables que requerían información numérica precisa (también como las horas trabajadas e ingresos), aumenta a medida que disminuye el nivel de detalle. Cuando la información se usa en bandas o se solicitan promedios, la coincidencia aumenta.
- Además, investigaron cuáles relaciones (entre *proxy* y persona sobre la cual se pregunta) proveen una menor tasa de error en la calidad del dato. Identificaron que las parejas responden mejor variables de ocupación, pero todo lo contrario sucede en las preguntas de nivel de calificación. Además, y aunque ningún tipo de relación puede entregar mejor información por sobre otra, plantean que la

¹⁸ Antecedentes y Metodología de la Encuesta de Fuerza Laboral. Disponible en <https://www.ons.gov.uk/file?uri=/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance/volume12016.pdf>

¹⁹ Un estudio de las respuestas *proxy* en la Encuesta de Fuerza Laboral.

persona *proxy* más capacitada para responder es aquella que se ve más afectada por las acciones de los otros integrantes del hogar, aunque las variables de ingreso son más confiables si se restringen a las parejas.

Por otro lado, Bardasi y otros (2011) indican en su estudio realizado en Tanzania de comparación de información entre distintos tipos de informantes –en el cual se eligieron al azar informantes *proxy*, los cuales debían informar tanto por sí mismos como por otros dos integrantes del hogar–, que la información obtenida por informante *proxy* lleva a una menor tasa de ocupación masculina²⁰ y a un sub-reporte de horas trabajadas por hombres. En este caso, la información sobre ocupación femenina no se vio afectada.

Además, identificaron que las discrepancias entre información auto-reportada y entregada por informante *proxy* disminuyen cuando éste es la pareja de la persona sobre la cual se pregunta y/o a medida que aumenta su nivel educativo.

A modo de síntesis, la evidencia da cuenta de que las principales diferencias según tipo de informante tienen lugar en la no respuesta parcial y en las respuestas numéricas, como ingresos y número de días y horas trabajadas. En cuanto a la calidad de la información en general, si bien se han encontrado diferencias entre informantes *proxy* y directos/as, éstas no han sido estadísticamente significativas y no permiten concluir que existan brechas sustantivas entre ambos tipos de informante para las encuestas revisadas.

Encuesta Nacional del Empleo y Encuesta de Caracterización Socioeconómica 2015

En nuestro país no existen análisis que consideren el impacto que puede tener la utilización del/de la informante *proxy*, lo cual podría ser relevante al momento de elegir y diseñar la metodología de las encuestas públicas.

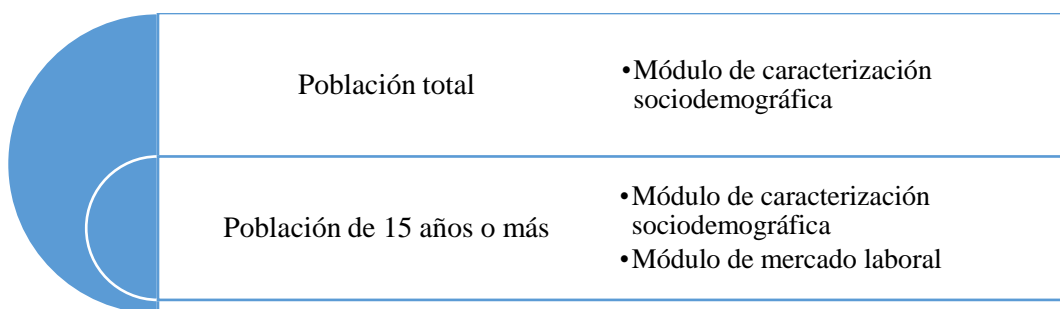
En este caso, interesa acercarse a los impactos del/de la informante *proxy* en la Encuesta Nacional del Empleo (ENE) realizada por el Instituto Nacional de Estadísticas (INE), y en la Encuesta de Caracterización Socioeconómica Nacional (CASEN) ejecutada por el Ministerio de Desarrollo Social en su versión 2015, que corresponden a los principales

²⁰ En particular este hallazgo se relacionó al sub-reporte general de las actividades agrarias, las cuales son realizadas principalmente por hombres (Bardasi, Beegle, Dillon, & Serneels, 2011).

instrumentos que se utilizan como referencia para la elaboración y evaluación de políticas públicas.

El objetivo principal de la ENE es caracterizar a la población de 15 años y más respecto a su situación en el mercado laboral en Chile (Instituto Nacional de Estadísticas, 2018), midiendo la ocupación, desocupación e inactividad, y caracterizando el empleo y su calidad (tipos de jornadas, garantías del trabajador, informalidad, etcétera). Esta encuesta se aplica desde el año 2010 en su versión más reciente, ajustada a los requerimientos internacionales dictados por la Organización Internacional del Trabajo. El cuestionario se divide en dos partes: el módulo de caracterización sociodemográfica para todos/as los integrantes del hogar, y el módulo central con las preguntas sobre mercado laboral para las personas de 15 años o más.

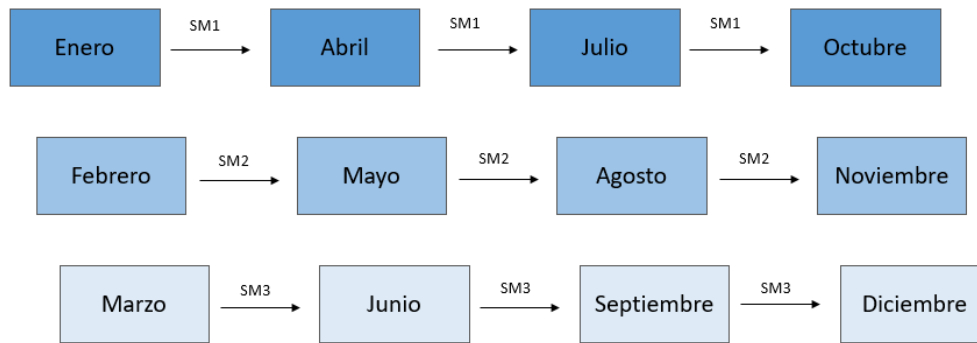
Figura 4: Módulos de la ENE a aplicar para cada población



Fuente: Elaboración propia a partir de Manual conceptual, operativo y de procedimientos ENE (no disponible online).

La ENE además tiene la característica de ser una encuesta que se aplica mediante panel: consiste en tres submuestras (submuestras 1, 2 y 3), que se repiten en cada trimestre del año, como muestra la siguiente figura:

Figura 5: Funcionamiento de las submuestras de la ENE



Fuente: Elaboración propia a partir de Manual Conceptual y Metodológico Diseño Muestral Encuesta Nacional del Empleo (ENE), disponible en <http://www.ine.cl/docs/default-source/laborales/ene/antecedentes-metodologicos/manual-conceptual-y-metodol%C3%B3gico-dise%C3%B1o-muestral.pdf?sfvrsn=6>.

Las tres submuestras son totalmente independiente entre sí, no repitiéndose casos entre ellas. Dentro del funcionamiento de cada submuestra, entre aplicaciones se rota un sexto de su muestra urbana, un noveno de la muestra RAU²¹ y un doceavo de la rural. Así, durante el año, cada submuestra es actualizada para cada aplicación.

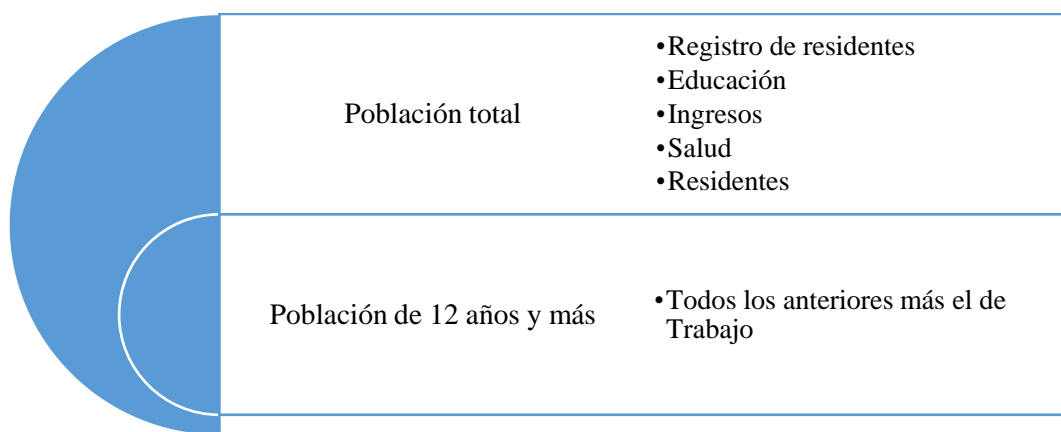
Los resultados de un mes –que son los que se difunden de manera oficial– corresponden al levantamiento de todo el trimestre móvil: el mes en cuestión es el central, el primero el anterior y el tercero el posterior. Así, por ejemplo, si se quieren conocer los datos de octubre, se debe utilizar la base correspondiente al trimestre móvil septiembre-octubre-noviembre (SON).

La CASEN, por su parte, tiene como objetivo conocer de manera periódica el estado de los hogares y la población, con énfasis en quienes se encuentran en situación de pobreza y de otros grupos prioritarios (definido así por la política social), en relación a aspectos sociodemográficos, de educación, salud, vivienda, trabajo e ingresos. De manera más específica, busca estimar el impacto de la pobreza, cómo se distribuyen los ingresos, cuáles son las necesidades de la población, y evaluar las brechas entre segmentos sociales (Ministerio de Desarrollo Social, 2018).

²¹ Resto de Área Urbana: estratos formados por centros urbanos con menos de 40.000 habitantes.

La encuesta, en su versión 2015²², se compone de siete módulos: Registro de residentes (H), Educación (E), Trabajo (O), Ingresos (Y), Salud (S), Residentes (R) y Vivienda (V). El único que tiene filtro de edad es el de Trabajo, el cual se aplica solo para personas de 12 años o más. Los de Registro de residentes, Educación, Ingresos, Salud y Residentes son aplicables a todos/as los miembros del hogar. El de Vivienda realiza preguntas de caracterización de la vivienda y no de los/las integrantes del hogar, por lo que solo requiere ser contestado por una persona.

Figura 6: Módulos de la CASEN 2015 a aplicar para cada población*



*El módulo de Vivienda se excluyó del diagrama ya que no realiza preguntas sobre personas.

Fuente: Elaboración propia a partir de cuestionario CASEN 2015 disponible en http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Cuestionario_CASEN_2015.pdf

La importancia de estas encuestas no solo se remite a la entrega de información necesaria para el diseño, aplicación y evaluación de políticas públicas, sino que también tienen un importante rol en la opinión pública, pues al medir temas sensibles para la población - empleo y pobreza, respectivamente- cumplen un rol fundamental en la evaluación del desempeño del gobierno y las autoridades en cargo.

Ambas encuestas tienen características específicas en su aplicación:

²² Durante la realización de esta tesis se publicaron los resultados y base de datos de la encuesta CASEN 2017. Se descartó el uso de la nueva versión en tanto, de acuerdo a las experiencias de otras encuestas (como la misma ENE), no existe evidencia que indique que podrían cambiar significativamente los datos relacionados a las personas informantes entre ambas mediciones.

Tabla 3: Contrastes entre la ENE y la CASEN 2015

	ENE	CASEN 2015
Periodicidad	Mensual. El módulo de Ingresos se aplica solo en los meses de octubre, noviembre y diciembre de cada año.	Cada dos años. Sólo tres encuestas CASEN (2003, 2006 y 2009) se realizaron cada tres años.
Duración del levantamiento	Cuatro o cinco semanas por mes.	Tres meses (noviembre 2015 a enero 2016).
Tamaño de la muestra	Varía según el mes de aplicación, pero en promedio son 11.959 viviendas por mes. El tamaño de la muestra de hogares variará en función de los hallazgos en el levantamiento.	83.887 hogares.
Uso de panel	Dependiendo del tipo de estrato, las viviendas de la ENE son visitadas de forma continua cada tres meses: en ciudades y grandes centros urbanos son en total seis visitas, en RAU nueve, y en sectores rurales doce. En cada muestra mensual, va rotando un sexto de la muestra urbana, un noveno de la muestra RAU y un doceavo de la rural.	No usa panel en la versión 2015.
Representatividad	Representativa a nivel nacional, regional y capitales regionales, además de algunas ciudades del país.	Representativa a nivel nacional, regional, por zona de residencia (urbana y rural), y en comunas que, sumadas, concentran el 80% o más de las viviendas de cada

	ENE	CASEN 2015
		región (según marco muestral del INE).
Informante proxy	Miembro del hogar de 15 años o más que declare poder responder el cuestionario por él/ella y el resto de los integrantes del hogar.	Jefe/a de hogar. En su ausencia, algún miembro del hogar de 18 años o más.
Registro de informante proxy	Registra a uno o una de sus integrantes como informante proxy. No permite registrar si otras personas participaron de forma parcial en la respuesta del cuestionario.	Al finalizar cada módulo, se puede registrar si la persona sobre la cual se está preguntando estaba presente y respondió, estaba presente y no respondió, o no estaba presente y no respondió.
Registro de no respuesta parcial	No	Sí
Permite elaboración de marco muestral para otras encuestas	Los identificados como empleadores/as o trabajadores/as por cuenta propia en la ENE, conforman el marco muestral de la Encuesta de Micro Emprendimiento.	No

Fuente: Elaboración propia a partir Metodología de Diseño Muestral Encuesta de Caracterización Socioeconómica Nacional, Casen 2015, (http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Metodologia_de_Disenio_Muestral_Casen_2015.pdf), Manual de Trabajo de Campo CASEN 2015 (http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Manual_Casen_2015.pdf), y Manual conceptual, operativo y de procedimientos ENE (no disponible online).

Marco metodológico

Para resolver la pregunta de investigación y sus objetivos, se trabajó con tres bases de datos: Encuesta Nacional del Empleo (desde 2010 a 2017²³), CASEN 2015, y de manera complementaria, la V Encuesta de Microemprendimiento (EME) 2017²⁴ –cuya incorporación y relevancia en este estudio será justificada más adelante–.

Se optó por trabajar con bases de datos ya producidas por dos motivos: en primer lugar, tanto la ENE como la CASEN cuentan con muestras de gran tamaño, lo que resulta favorable al momento de identificar si determinados perfiles de personas tienden a ser informantes por sobre otros. En segundo lugar, si bien habría sido pertinente producir información nueva con énfasis en el tipo de informante, los costos monetarios habrían sido muy altos para realizarlo en el marco de esta tesis.

Cabe destacar que los ejercicios analíticos desarrollados fueron de carácter exploratorio, en tanto no existe evidencia empírica respecto a esta temática en estas encuestas en particular, ni en otras del país.

Todos los procesamientos fueron realizados con los softwares Stata y SPSS.

Depuraciones de bases de datos

Antes de realizar los análisis específicos de esta tesis, fue necesario hacer un trabajo de depuración en las bases de datos.

En el caso de la ENE, fue depurada para que las variables que se utilizaron en el análisis no presentaran inconvenientes. Se trabajó principalmente en la variable “Informante”²⁵, la cual fue duplicada –para no alterar la variable original–, creando una nueva variable llamada “informante_rec”.

²³ La ENE fue reformulada el año 2010, cambiando el cuestionario y los protocolos de levantamiento. Por ende, se utilizará la información disponible desde esa fecha.

²⁴ La base de datos de la encuesta CASEN fue obtenida del sitio <http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/basedatos.php>, mientras que tanto la ENE como la EME fueron solicitadas mediante Transparencia: la base pública de la ENE no incluye la variable que identifica a las personas como informantes ni la que individualiza a las personas, y la EME pública no incorpora esta última.

²⁵ Que identifica a la persona del hogar que actuó principalmente de informante.

Luego de esto, se corrigieron los datos que estaban digitados erróneamente con valores que no correspondían a ninguna de las posibles categorías. En este proceso se identificaron casos digitados con el valor “3”, “4” y “7”, aun cuando las únicas categorías válidas eran “0” para no informantes, y “1” para informantes. Todos los casos estaban presentes en hogares que ya presentaban un informante, por lo que se tomó la decisión de categorizarlos como “No informantes”.

Finalmente, se recodificaron los datos con errores de categorización. En particular entre los/las menores de 15 años que estaban clasificados como “No informantes”, se dejó la casilla vacía²⁶. Por otro lado, quienes tenían 15 años o más y que no estaban clasificados se categorizaron como “No informantes”.

En el caso de CASEN y la EME las bases ya se encontraban depuradas, por lo que solo fue necesario hacer los cambios propios del análisis.

Caracterización de los y las informantes

Se caracterizó a las personas que son informantes de la ENE 2010-2017 y la CASEN 2015 en términos de sexo, edad, nivel educacional, situación en el mercado laboral (ocupación, desocupación o inactividad) y motivos de estar fuera del mercado laboral para quienes señalen ser inactivos/as. Esto se realizó mediante frecuencias simples y cruces de variables.

Esta caracterización se contrastó con la realidad nacional, obtenida de los resultados expandidos a la población de la CASEN 2015, con el fin de conocer si las características de los/las informantes son similares a las de la población, y evaluar si hay grupos sub o sobre representados.

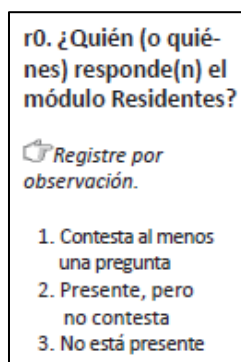
La identificación del/de la informante es diferente para cada una de las encuestas. En el caso de la ENE, la base de datos incluye una variable (“Informante”) que permite identificar quién es la persona informante del hogar que responde tanto por sí mismo/a como por los otros integrantes del hogar. Una de las limitantes es que solo identifica a un/una informante, sin reconocer si hubo participación de otra/s persona/s. De esta manera, si otra persona participó en las respuestas del cuestionario, no queda consignado en ningún

²⁶ El informante idóneo/a debe tener 15 años o más. Por ende, las personas de 14 o menos no son categorizados como “No informante”.


espacio del instrumento. Esto sin duda puede presentar sesgos respecto a quiénes son informantes, pero es el acercamiento más factible que se puede conseguir observando los datos de la ENE.

En cuanto a la CASEN 2015, en cada uno de los módulos se identifica a las personas del hogar que responden alguna pregunta y/o están presentes durante la entrevista. Así, para cada una se pregunta su participación en el cuestionario y se deja establecido en el instrumento. En específico, los códigos asignados a las respuestas son: 1. Contesta al menos una pregunta; 2. Presente, pero no contesta; y 3. No está presente. En los casos de los códigos “2” y “3”, las personas no participan en las respuestas del cuestionario, por lo que se considera que las preguntas relativas a estas personas fueron contestadas por un(a) informante *proxy*. De esta manera, se infiere que todos/as quienes tienen código “1” (Contesta al menos una pregunta) fueron informantes directos/as.

Imagen 1: Ejemplo de pregunta para identificar presencia o ausencia de la persona en CASEN 2015



r0. ¿Quién (o quiénes) responde(n) el módulo Residentes?

 Registre por observación.

1. Contesta al menos una pregunta
2. Presente, pero no contesta
3. No está presente

Fuente: Cuestionario CASEN 2015, disponible en http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Cuestionario_CASEN_2015.pdf

Ya que el registro de la persona informante en la CASEN es por módulo, se trabajó con una variable unificadora para así caracterizar a quienes efectivamente informaron a lo largo de todo el cuestionario y no solo en preguntas o módulos particulares. Esa variable no existe y debió ser creada. En función de esta necesidad, se consideraron informantes quienes en todos los módulos tienen un “1” (Contesta al menos una pregunta)²⁷.

Se excluyeron del análisis los hogares en que ninguna persona haya respondido simultáneamente todos los módulos del cuestionario, es decir:

²⁷ Se excluirá el módulo de Vivienda, el cual no se aplica por persona.

- La información de cada integrante del hogar fue entregada en su totalidad por otra persona –mediante informante *proxy*– (respuestas “2” y “3”).
- La información de cada integrante del hogar fue entregada por informante *proxy* y por informante directo: algunos módulos fueron respondidos por *proxy* y otros los respondió la persona directamente.

Esta decisión se justifica en tanto, en estos casos, es imposible identificar con certeza qué integrante del hogar fue el que entregó mayoritariamente la información, ya que ninguno/a completó un cuestionario en su totalidad. Si bien, en el segundo caso, existen algunos módulos respondidos de manera directa, queda la duda de por qué esa persona que contestó no participó en todos los módulos. Por las características del trabajo en terreno, se puede deber a encuestas realizadas en distintas visitas en que la información pendiente se consultó directamente con la persona en cuestión, pero no es posible determinarlo a partir de los datos. Ante esa incertidumbre, se optó por excluirlo del análisis, considerando además que esto no implicaba una disminución importante del total de casos (el detalle se abordará en el capítulo 1 del análisis).

Por otro lado, dentro de las personas informantes se distinguió entre informantes directos/as y *proxy*.

Para acotar la categoría de directos/as, se utilizaron: a) hogares unipersonales y, b) hogares que tienen la misma cantidad de informantes e integrantes, ya que en esos casos es más probable que cada uno haya informado por sí mismo a que lo hayan hecho por los otros. El criterio de hogares unipersonales es válido tanto para la ENE como para la CASEN, mientras que el de la cantidad de informantes e integrantes es aplicable solo a esta última, dado que la ENE señala a una única persona como informante.

Se consideraron *proxy* los/las informantes del resto de los hogares, es decir, de todos aquellos en que la cantidad de informantes fue menor a la cantidad de integrantes (aun cuando haya más de un/una informante, como sucede en algunos casos de CASEN). En la ENE fueron todos los hogares que no eran unipersonales.

Así, para este capítulo, se clasificará a las personas informantes de acuerdo a si responde solo por sí mismo/a (informante directo/a), o a si responde por sí mismo/a y por otros

integrantes del hogar (informante *proxy*). En este sentido, para clarificar, no son las respuestas las que se clasificarán como directas o entregadas por un *proxy*.

Como se mencionó anteriormente, la caracterización se realizó en base a ciertas variables sociodemográficas: sexo, edad, nivel educacional, situación en la ocupación, y razones de inactividad para quienes corresponda. Para llevar a cabo este proceso fue necesario recodificar ciertas variables:

- Edad: se crearon grupos etarios para dar una interpretación más acabada a la información:
 1. 15 a 17 años para ENE y de 12 a 17 años para CASEN. El límite inferior varía ya que en la primera encuesta el informante debe tener al menos 15 años, mientras que en la CASEN el margen se amplía a los 12 años.
 2. 18 a 29 años
 3. 30 a 44 años
 4. 45 a 59 años
 5. 60 años o más

- Nivel educacional: se formaron cuatro grupos de acuerdo al nivel más alto aprobado. En CASEN sólo fue necesario recodificar los grupos para que calcen en una de las categorías, ya que la base ofrece el nivel terminado. En la ENE, en cambio, la base indica solo el nivel cursado, pero incluye la variable dicotómica “termino_nivel”, en que “1” indica que terminó y “2” que no. Así, se cruzó la variable de nivel (“cine”) con “termino_nivel” para definir los grupos por nivel más alto aprobado:
 0. Sin educación formal: quienes no tienen educación formal o tienen educación básica incompleta.
 1. Básica completa: quienes tienen básica completa, media humanista incompleta o media técnica incompleta.
 2. Media completa: quienes tienen media humanista completa, media técnica completa, técnico nivel superior incompleta o profesional incompleta.
 3. Técnico superior completa: quienes tienen técnico nivel superior completo.

4. Profesional completa: quienes tienen educación profesional completa, postgrado incompleto o postgrado completo.

- Razones de inactividad: la ENE y la CASEN cuentan con un set distinto de razones de inactividad. Éstas fueron homologadas de la siguiente manera:

Tabla 4: Homologación de códigos ENE y CASEN para preguntas por razones de inactividad

Códigos de homologación		Razones ENE		Razones CASEN	
1	Posibilidad de empezar a trabajar pronto	1	Iniciará pronto una actividad por cuenta propia.	1	Consiguió trabajo que empezará pronto o iniciará pronto una actividad por cuenta propia.
		2	Encontró un empleo que empezará pronto.	2	Está esperando resultado de gestiones ya emprendidas.
		13	Espera los resultados de un proceso de selección o que lo llamen.		
2	Se hace cargo de tareas del hogar	3	Por responsabilidades familiares permanentes.	3	No tiene con quién dejar a los niños.
		11	Por responsabilidades familiares de carácter temporal.	4	No tiene con quién dejar a adultos mayores.
				5	No tiene con quién dejar a otro familiar.
				10	Quehaceres del hogar.
3	Estudios	4	Está estudiando o preparando estudios.	11	Estudiante.
4	Tiene otros	5	Es jubilado.	12	Jubilado(a), pensionado(a) o

Códigos de homologación		Razones ENE		Razones CASEN	
	ingresos				montepiado(a).
		6	Es rentista.	13	Tiene otra fuente de ingresos (seguro cesantía, mesadas, rentas, transferencias del Estado, etc.).
		7	Es pensionado(a) o montepiado(a).		
5	Tiene problemas de salud	8	Por motivos de salud permanentes.	6	Está enfermo o tiene una discapacidad.
		10	Por motivos de salud temporales.		
6	Percepción de limitaciones	15	Cree que por su edad no le darán empleo.	7	Piensa que nadie le dará trabajo (porque no cuenta con la capacitación requerida, por su edad, etc.).
		16	Cree que no lo encontrará.		
		18	Cree que ningún empleo o actividad se adapte a su calificación.		
7	Otros	9	Espera la estación de mayor actividad	8	Las reglas, horarios y distancias de los trabajos no le acomodan
		12	Está embarazada	9	Ofrecen sueldos muy bajos
		14	Algún miembro del hogar no se lo permite	14	Se cansó de buscar o cree que no hay trabajo disponible

Códigos de homologación		Razones ENE		Razones CASEN	
		17	Se cansó de buscar	15	Busca cuando lo necesita o tiene trabajo esporádico
		19	Le piden demasiados trámites para iniciar una actividad por cuenta propia	16	No tiene interés en trabajar
		20	No sabe dónde dirigirse ni a quién acudir	17	Otra razón
		21	No quiere, no necesita trabajar		

Fuente: Libro de códigos CASEN (disponible en http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Libro_de_Codigos_Casen_2015.pdf) y Diccionario de códigos ENE (disponible en http://historico.ine.cl/ene/files/pdf/DIC_VAR_BASES_DE_DATOS_EXTERNOS.pdf).

- Situación en el mercado laboral: para determinar si las personas se encontraban ocupadas, desocupadas o inactivas en la ENE, se debió recodificar la variable “cse”:

Tabla 5: Recodificación para identificar la situación en la ocupación

Recodificación		Códigos ENE	
1	Ocupado/a	1	Ocupado tradicional
		2	Ocupado no tradicional
		3	Ocupado ausente
2	Desocupado/a	4	Cesante
		5	Busca trabajo por primera vez
3	Inactivo/a	6	Iniciador
		7	Inactivos que buscaron
		8	Inactivos que estuvieron disponibles
		9	Inactivos que no buscaron ni estuvieron disponibles

Fuente: Manual conceptual, operativo y de procedimientos ENE.

En el caso de CASEN no fue necesario recodificar, ya que la base incluye de manera previa la variable “activ” con las alternativas “1. Ocupados”, “2. Desocupados”. “3. Inactivos” y “99. Blancos”.

Además, se realizó un análisis de correspondencias múltiples (ACM), con el fin de conocer qué categorías de las variables sociodemográficas están a menor distancia de las categorías “Informante” y “No informante”. Esto se calculó a partir de las coordenadas de cada punto (entregadas por el ACM), y luego aplicando la fórmula para distancia entre dos puntos en un plano:

$$\text{Distancia entre dos puntos} = \sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2)}$$

En donde “x” e “y” son las coordenadas de los puntos a los cuales se les medirá la distancia.

Esto, en conjunto con la caracterización realizada anteriormente, ayudó a indicar cuáles son los atributos que se acercan mayormente a ser o no informante, tanto en la ENE como en la CASEN.

Posteriormente, el ejercicio se repitió en ambas encuestas, con la salvedad de que lo que interesaba conocer eran las categorías más asociadas con cada tipo de informante (*proxy* y *directo/a*).

Como explica Morales Jacob (2004), el ACM transforma variables nominales en dimensiones numéricas, las cuales incorporan la asociatividad o relación entre las variables. La visualización de estas dimensiones y las variables se materializa en un mapa perceptual, el cual se compone de ejes, coordenadas y puntos, siendo los ejes las dimensiones y los puntos las categorías de las variables.

A grandes rasgos, el modelo de ACM se evalúa a partir de la inercia, la cual mide la dispersión de los puntos respecto al centroide (Greenacre, 1984). En términos gruesos, la inercia puede ser homologada con la varianza explicada del modelo.

Si bien el ACM conlleva la generación de dimensiones, en esta tesis el foco no estuvo en identificarlas o caracterizarlas, ya que el objetivo era conocer la relación entre las categorías y el ser o no informante. De hecho, la elección de esta técnica estadística se debe

principalmente a que, a diferencia de otros tipos de análisis, el ACM evalúa no solo las relaciones entre variables, sino que además las relaciones entre las categorías de las distintas variables (Morales Jacob, 2004).

Uno de los requisitos del ACM es usar variables nominales u ordinales, lo se cumple en este caso. Por ello, se utilizó la prueba de Phi para determinar previamente si las variables tenían relación con ser o no ser informante.

Se optó por esta prueba ya que, de acuerdo a Merino (Pardo Merino & Ruiz Díaz, 2005), es un estadístico basado en chi-cuadrado, pero que a diferencia de éste, puede cuantificar el grado de relación entre dos variables aislando el efecto del tamaño muestral. En consideración de que tanto la ENE como la CASEN son muestras con decenas de miles de casos, fue necesario realizar un análisis que no estuviera sesgado por sus tamaños. Además, es apropiado para el nivel de medición de las variables en cuestión.

Variaciones en respuestas ENE-EME

Otro aspecto a profundizar es la calidad de las respuestas dadas por los y las informantes *proxy*. A pesar del debate sobre qué tipo de informante es mejor, cabe problematizar en un principio si las respuestas serían diferentes encuestando a un o una informante directo/a sobre un/una *proxy*.

Para esto se trabajó con la V Encuesta de Microemprendimiento 2017²⁸, cuyo marco muestral está conformado por los/las trabajadores/as por cuenta propia y empleadores/as con hasta 10 trabajadores/as incluyéndose (denominados Microemprendedores) que residen en viviendas particulares ocupadas dentro del territorio nacional, y que fueron identificados/as previamente en la ENE 2017 del trimestre marzo-abril-mayo (MAM). La V EME se levantó en mayo, junio, julio y agosto de 2017²⁹.

La particularidad de esta encuesta es que, por protocolo, utiliza exclusivamente informante directo/a, aun cuando antes haya sido respondida mediante informante *proxy* en la ENE.

²⁸ La ENE y la EME son levantadas en terreno por distintos equipos, lo que podría inducir a diferencias entre ambos resultados.

²⁹ Se debe tener en consideración que la ENE y la EME, si bien comparten algunas preguntas, son instrumentos diferentes con distintos objetivos. Además, son levantadas en terreno por distintos equipos. Por ende, se enfatiza el carácter exploratorio de este análisis.

Esto permitiría comparar las respuestas entregadas, y definir si existen diferencias entre informantes cuando la actividad principal –sobre la cual se hacen las preguntas– no ha cambiado entre encuestas.

Para esto, se creó una nueva base de datos con las variables de interés de la ENE y EME a partir de la variable “idrph”, que es común en ambas bases y que individualiza a las personas encuestadas. De esta manera, cada persona que es encuestada tanto en la ENE como en la EME comparte en ambas bases el valor de la variable “idrph”, pudiendo identificar así las respuestas para cada persona en las dos encuestas, lo que se expresó en la nueva base de datos. Además, se incorporó una variable que identificó a quienes también fueron informantes en la ENE.

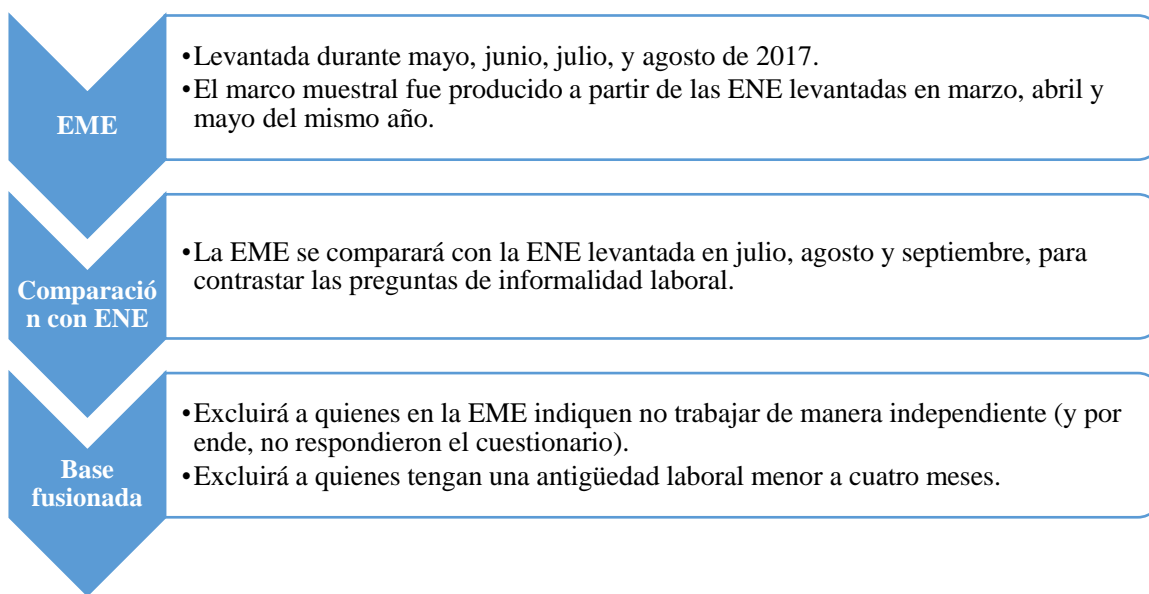
A partir del tipo de informante de la ENE se generaron dos grandes grupos:

1. Grupo control: conformado por quienes en la EME y la ENE respondieron de manera directa.
2. Grupo experimental: conformado por quienes en la EME contestaron de manera directa y en la ENE no fueron informantes, es decir, respondió un/una informante *proxy*.

Es importante mencionar que la comparación de preguntas se hizo entre aquellas de informalidad laboral incorporadas en el cuestionario de la ENE y la EME en julio de 2017 (no estaban con anterioridad). Por ende, no fue posible contrastar la EME con la ENE del período MAM (trimestre marzo-abril-mayo, del cual se obtiene el marco muestral para la EME), sino que se utilizó el trimestre julio-agosto-septiembre (JAS). Esto es posible hacerlo debido al carácter de encuesta panel de la ENE –detallado en los antecedentes–, en que gran parte de quienes respondieron en MAM también lo hacen en JAS. Para asegurar que las personas tuvieran y se refirieran al mismo empleo, se eliminaron quienes indicaron una antigüedad laboral menor al tiempo transcurrido entre ambas encuestas³⁰.

³⁰ Esto se analizará en la encuesta que se haya aplicado de manera posterior, que en general es la ENE.

Figura 7: Esquema de proceso de fusión y comparación EME-ENE



Fuente: Elaboración propia a partir de definición metodológica.

Finalmente, se excluyó del análisis quienes en la EME indicaron que cambiaron de situación ocupacional (es decir, dejaron de ser independientes). Esto se hizo a partir de la revisión de la pregunta a1 de la EME (“Usted dijo haber trabajado como Empleador o Trabajador por Cuenta Propia en _____ de 2017, ¿continúa con esa actividad o con otra?”).

Realizado todo este proceso, se pudo contrastar las respuestas dadas en la EME y la ENE por los distintos tipos de informantes. Se denominará “Momento 0” la medición de la EME (al establecer la línea base) y “Momento 1” a la posterior medición de la ENE.

En función de los momentos de medición y de los grupos de control y experimental, existen cuatro sub-poblaciones a analizar:

Figura 8: Sub-poblaciones de análisis para análisis comparativo EME-ENE

	Momento 0 (EME)	Momento 1 (ENE)
Grupo control	Directo/a	Directo/a
Grupo experimental	Directo/a	<i>Proxy</i>

Fuente: Elaboración propia a partir de definición metodológica

En un primer momento, se compararon los resultados entre el grupo control y experimental en el M0 (es decir, entre informantes directos de la EME), solo de manera descriptiva. No se buscó caracterizar estadísticamente las diferencias entre ellos, en tanto era posible que éstas existieran por no haber criterios de selección para cada uno, más allá de quién informó en el M1. Así, las diferencias podrían suceder debido a, por ejemplo, el tipo de ocupación y su incidencia en que la persona se encontrara en la vivienda, y, por ende, en sus probabilidades de responder de manera directa la ENE.

Luego, se profundizó en las diferencias entre ambos grupos en el M1, para evaluar si las diferencias perduraban. Finalmente, se analizó el porcentaje de respuestas que se mantuvieron idénticas del M0 al M1, en ambos grupos.

Para determinar si las diferencias fueron significativas entre ambos momentos, se creó una nueva variable para cada par de preguntas a analizar, en que “1” indicó respuestas iguales y “0” respuestas diferentes. Quienes respondieron solo en un cuestionario figuraron como valores perdidos.

Esta nueva variable se contrastó, mediante una prueba t para una muestra, con un estadístico de valor “1” (que representaba el 100% de respuestas respondidas de igual manera en ambos momentos). Cuando las diferencias fueron significativas (determinado a partir de los valores t y sus grados de libertad), se profundizó en cuáles categorías presentaban una mayor fuga de respuestas entre mediciones.

Además de esto, se estimó el tamaño del efecto, definido como “el grado en que el fenómeno se presenta en la población” (Cohen, 1988), para determinar el grado de fuerza de la asociación entre las variables y descartar que la significación encontrada se debiera principalmente al tamaño de muestra utilizado en la EME. Para esto, se utilizó la medida de correlación (r) para medir el tamaño del efecto. En este caso el “efecto” fue la mantención o cambio de respuestas entre ambas mediciones, tanto en el grupo control como en el experimental.

Se prefirió su uso sobre otras posibilidades (como d de Cohen, por ejemplo), pues su interpretación es más intuitiva al solo tener valores de 0 a 1, y porque además facilita el meta-análisis (Ledesma, Macbeth, & Cortada de Kohan, 2008). Se calculó a partir de los grados de libertad y el valor de t , y se interpreta de la siguiente manera (University of Colorado Colorado Springs, 2018):

- Tamaño del efecto pequeño: r menor o igual a 0,1.
- Tamaño del efecto mediano: r mayor a 0,1 y menor o igual a 0,243.
- Tamaño del efecto grande: r mayor a 0,243.

Además, otra ventaja de r es que, al elevarla al cuadrado, indica el porcentaje de varianza en la variable dependiente que se explica por la variable independiente (University of Colorado Colorado Springs, 2018).

En concreto, la comparación se realizó sobre las siguientes preguntas compartidas entre la EME y la ENE:

Tabla 6: Preguntas a comparar entre ENE y EME 2017

Preguntas ENE	Preguntas V EME	Requiere recodificación
b1. ¿Cuál es el oficio u ocupación que realizó la semana pasada?, ¿qué tareas realizó en esta ocupación?	d19. ¿Cuál es el oficio, labor u ocupación principal que realiza como trabajador independiente?	No
b14a. ¿A qué se dedica como trabajador por cuenta propia?	d20. ¿A qué se dedica su negocio, empresa o actividad por cuenta	No

Preguntas ENE	Preguntas V EME	Requiere recodificación
	propia?	
i4. La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos (SII) o tiene iniciación de actividades?	e40. En su actual negocio, empresa o actividad por cuenta propia, ¿inició actividades en el Servicio de Impuestos Internos (SII)?	Sí
i5. La empresa, negocio o actividad en la que trabaja, ¿se encuentra registrada como...		
i6. ¿La empresa, negocio o actividad por cuenta propia... (pregunta sobre contabilidad)	e38. ¿Lleva algún tipo de registro de las cuentas de su negocio, empresa o actividad por cuenta propia?	Sí
i7. A través de la contabilidad, ¿se pueden separar los gastos del negocio de los gastos del hogar?	e39. A través de la contabilidad, ¿puede separar los gastos de su negocio, empresa o actividad por cuenta propia de los gastos de su hogar?	No

Fuente: Manual conceptual, operativo y de procedimientos ENE, y Manual Técnico y Operativo del Encuestador EME 2017. La columna de recodificación es de elaboración propia.

Las preguntas b1 y d19 son de respuesta abierta y se clasifican en las siguientes categorías:

1. Miembros del poder ejecutivo y de los cuerpos legislativos y personal directivo de la administración pública y de empresas públicas,
2. Profesionales científicos e intelectuales,
3. Técnicos y profesionales de nivel medio,
4. Empleados de oficina,
5. Trabajadores de los servicios y vendedores de comercios y mercados,
6. Agricultores y trabajadores calificados agropecuarios y pesqueros,
7. Oficiales, operarios y artesanos de artes mecánicas y de otros

oficios, 8. Operadores de instalaciones y máquinas y montadores, 9. Trabajadores no calificados, y 10. Otros No identificados.

También b14 y d20 son de respuesta abierta, y se clasifican en las siguientes categorías: 1. Agricultura, silvicultura y pesca, 2. Explotación de minas. 3. Industrias manufactureras, 4. Suministro de electricidad, gas, vapor y aire acondicionado, 5. Suministro de agua; alcantarillado, gestión de desechos y actividades de saneamiento, 6. Construcción, 7. Comercio al por mayor y al por menor; reparación de los vehículos de motor y de las motocicletas, 8. Transporte y almacenamiento, 9. Alojamiento y servicios de comida, 10. Información y comunicación, 11. Actividades financieras y de seguros, 12. Actividades inmobiliarias, 13. Actividades profesionales, científicas y técnicas, 14. Actividades administrativas y servicios de apoyo, 15. Administración pública y defensa; planes de seguridad social de afiliación obligatoria, 16. Enseñanza, 17. Servicios sociales y relacionados con la salud humana, 18. Artes, entretenimiento y recreación, 19. Otras actividades de servicio, 20. Actividades de los hogares en calidad de empleadores, y 21. Actividades de organizaciones y órganos extraterritoriales.

En estas preguntas la comparación fue sobre las codificaciones y no las glosas. La decisión se basa en el carácter privado de las glosas (no son incluidas en las bases públicas para impedir la identificación de las personas), y por la carga de trabajo que implica comparar estas mismas, sobre todo sin conocer los procesos propios de nomenclatura.

Todas las preguntas a comparar entre la ENE y la EME tienen objetivos similares, como se observa en la siguiente tabla:

Tabla 7: Objetivos de preguntas a comparar entre ENE y EME

Preguntas	Encuesta Nacional del Empleo	Encuesta de Microemprendimiento
Objetivo b1 ENE/d19 EME	Caracterizar con el mayor detalle posible la ocupación en la cual se desempeña el(la) trabajador(a) en la semana de referencia.	Caracterizar la ocupación, oficio y las tareas que realiza el informante para su actividad independiente principal (a la que se dedica la mayor parte del

Preguntas	Encuesta Nacional del Empleo	Encuesta de Microemprendimiento
		tiempo).
Objetivo b14 ENE/d20 EME	Conocer la rama de actividad en que se desempeña el(la) trabajador(a). Variable utilizada para codificar según la Clasificación Internacional de Actividad Económica CIIU Rev4 ³¹ .	Describir en detalle la actividad económica a la que se dedica el negocio, empresa o actividad por cuenta propia del informante seleccionado.
Objetivo i4 e i5 ENE/e40 EME	i4 - Identificar la formalidad de la unidad económica (empresa, negocio, etc.) en la que trabaja o es dueño el trabajador independiente, a través del registro de la unidad económica en el SII. i5 - Identificar la formalidad de la unidad económica (empresa, negocio, etc.) en la que trabaja o es dueño el ocupado independiente a través del tipo	Caracterizar la formalidad del emprendimiento según el tipo de registro que tiene el negocio o empresa ante el Servicio de Impuestos Internos.

³¹ CIIU es la sigla para el Clasificador Industrial Internacional Uniforme de todas las actividades económicas, indicador de Naciones Unidas creado en 1948 que indica la clasificación internacional de referencia de las actividades productivas. El objetivo es “ofrecer un conjunto de categorías de actividades que se pueda utilizar para la reunión y difusión de datos de acuerdo con estas actividades” (Naciones Unidas, 2009). La denominación de Rev4 refiere a la cuarta revisión de este indicador, es decir, a la versión que incluye una nueva etapa de comentarios y correcciones aportadas por expertos y usuarios de todo el mundo. En Chile, el Instituto Nacional de Estadísticas crea una versión del CIIU aplicable a la realidad nacional, llamado Clasificador Chileno de Actividades Económicas.

Preguntas	Encuesta Nacional del Empleo	Encuesta de Microemprendimiento
	de registro en el SII.	
Objetivo i6 ENE/e38 EME	Conocer si la empresa, negocio o unidad económica en la que trabaja o de la que es dueño, cuenta o no con algún tipo de registro contable.	Saber si el negocio o empresa tiene algún grado de formalidad en términos de contabilidad, llevando registros contables de los ingresos y gastos.
Objetivo i7 ENE/e39 EME	Conocer si a través de la contabilidad declarada, se puede separar los gastos del negocio de los gastos del hogar.	Identificar si el informante seleccionado puede separar los gastos del negocio de los gastos del hogar a través del sistema de registro que declaró mantener.

Fuente: Manual conceptual, operativo y de procedimientos ENE, y Manual Técnico y Operativo del Encuestador EME 2017.

Algunas preguntas a comparar no son exactamente iguales y, por ende, las categorías de respuestas fueron recodificadas para que así fueran equivalentes. Las preguntas involucradas, y las recodificaciones, son las siguientes:

1. En preguntas i4 e i5 de la ENE, y e40 de la V EME

Siendo las preguntas de la ENE:

- i4. La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos (SII) o tiene iniciación de actividades?
 1. Sí
 2. No
 88. No sabe
 99. No responde

- (Sólo para quienes responden 1 en i4) i5. La empresa, negocio o actividad en la que trabaja, ¿se encuentra registrada como...
 1. ...trabajador independiente (con boleta de honorarios)?
 2. ...persona natural (el RUT de la empresa es igual al RUT del dueño)?
 3. ...otro tipo de registro (Limitada, E.I.R.L, S.A, Spa)?

88. No sabe

99. No responde

Se unieron i4 e i5 de la ENE en una nueva variable (“i4_e5”), en donde la categoría “1” de i4 fue reemplazada por las categorías “1”, “2” y “3” de i5, manteniendo las categorías “2” original (“No”). Esto se justificó ya que sólo quienes responden “1” en i4 pasan a i5, y permitió tener en una misma variable el tipo de registro de la empresa. Los “88” y “99” de i4 e i5 se codificaron como no respuesta parcial de i4_i5.

Así, la variable i4_i5 de la ENE quedó de la siguiente manera:

- i4_i5. La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos (SII) o tiene iniciación de actividades?
 1. Sí, como trabajador independiente (con boleta de honorarios)
 2. Sí, como persona natural (el RUT de la empresa es igual al RUT del dueño)
 3. Sí, con otro tipo de registro (Limitada, E.I.R.L, S.A, Spa)?
 4. No

88. No sabe

99. No responde

En la EME, se transformó la siguiente variable:

- e40. En su actual negocio, empresa o actividad por cuenta propia, ¿inició actividades en el Servicio de Impuestos Internos (SII)?
 1. Sí, como trabajador independiente (boleta de honorarios)
 2. Sí, como empresa constituida como persona natural
 3. Sí, como Empresa Individual de Responsabilidad Limitadas (EIRL) o Sociedad de Responsabilidad Limitada (Ltda.)

- 4. Sí, como otro tipo de empresa
- 5. No
- 6. No, estoy en proceso

En concreto, se fusionaron las categorías “3” y “4” por un lado, y “5” y “6” por otro, para mantener la misma estructura de la ENE, resultando la siguiente pregunta:

- e40_rec. En su actual negocio, empresa o actividad por cuenta propia, ¿inició actividades en el Servicio de Impuestos Internos (SII)?
 - 1. Sí, como trabajador independiente (boleta de honorarios)
 - 2. Sí, como empresa constituida como persona natural
 - 3. Sí, como otro tipo de empresa
 - 4. No

Entre ambas existe una diferencia: la variable de la ENE incluye “no sabe” y “no responde”, mientras que la de la EME no lo hace. Se mantuvo esta discrepancia en tanto permitió conocer qué categorías de la EME absorbieron la no respuesta parcial de la ENE.

2. En preguntas i6 de la ENE y e38 de la V EME

Siendo i6 de la ENE:

- i6. ¿La empresa, negocio o actividad por cuenta propia...
 - 1. ...acude a los servicios de un contador para llevar la contabilidad completa?
 - 2. ...se encuentra acogido al régimen de contabilidad simplificada?
 - 3. ...sólo cuenta con registros personales de gastos e ingresos?
 - 4. No cuenta con ningún tipo de contabilidad
 - 88. No sabe
 - 99. No responde

Y siendo e38 de la EME:

- e38. ¿Lleva algún tipo de registro de las cuentas de su negocio, empresa o actividad por cuenta propia?

1. Sí, a través de contabilidad completa formal
2. Sí, a través de contabilidad completa con registros personales
3. Sí, a través de contabilidad simplificada
4. Sí, a través de contabilidad parcial con registros personales
5. No, ningún tipo de contabilidad
88. No sabe
99. No responde

Se transformó e38 de la EME, fusionando las categorías “2” y “4”, resultado de la siguiente manera:

- e38_rec. ¿Lleva algún tipo de registro de las cuentas de su negocio, empresa o actividad por cuenta propia?
 1. Sí, a través de contabilidad completa formal
 2. Sí, a través de contabilidad simplificada
 3. Sí, a través de registros personales
 4. No, ningún tipo de contabilidad
 88. No sabe
 99. No responde

Es importante tener dos consideraciones respecto a este punto. En primer lugar, en la ENE la categoría “1” incluye el ejemplo de la figura del contador para ayudar a guiar la respuesta, pero no excluye la contabilidad completa realizada por otra persona. En la misma línea, como explicita el Manual de Trabajo de Campo de la V EME³², no es necesario que un contador la realice, pudiendo ser llevada a cabo por alguna otra persona (incluye al propietario/a de la empresa) con conocimientos de contabilidad. Por ende, podría existir alguna diferencia respecto a esa categoría debido a los fraseos distintos, en caso de que el ejemplo del contador haya distorsionado la percepción del o la informante.

En segundo lugar, la categoría “2” de la EME (“Sí, a través de contabilidad completa con registros personales”) incluye a quienes realizan contabilidad completa pero no la presentan frente al SII, elaborándose ese fraseo para ayudar la respuesta por parte del informante. Sin

³² No disponible online, se debe gestionar consulta con funcionarios de la EME.

embargo, en términos formales, es equivalente a la categoría “3” de la ENE (“Sólo cuenta con registros personales de gastos e ingresos”).

En función de las comparaciones a realizar entre la EME y la ENE, es importante tener en consideración una limitación dada la naturaleza de las preguntas a analizar. En concreto, varias de ellas refieren a temáticas particulares relacionadas con la persona trabajadora en tanto independiente (como inscripción en el SII, formas de contabilidad y separación de gastos), y, por ende, no corresponden a preguntas sobre un tipo de trabajo “tradicional”: dependiente, fijo y estable.

Si bien esto es reflejo de los procesos de transformación que ha tenido el mercado laboral en los últimos años³³, sí puede aumentar la complejidad y, por ende, las posibilidades de diferencia entre informantes *proxy* y directo/a.

No respuesta parcial CASEN

Finalmente, se indagó en la no respuesta parcial de la Encuesta CASEN según tipo de informante, para conocer si ésta tiene una presencia significativamente diferente y relevante en las respuestas dadas por los y las informantes *proxy*, en comparación con las dadas por los informantes directos/as.

Cabe destacar que, salvo las preguntas analizadas en el apartado anterior, la ENE prácticamente no incluye la no respuesta parcial. Esto no significa que durante el levantamiento ésta no exista, sino que no queda registrada en el instrumento³⁴.

Por ende, en este capítulo se analizó la encuesta CASEN 2015, la cual registra la no respuesta parcial con los códigos “9” o “99” en preguntas con respuestas categóricas y

³³ De hecho, como ya se mencionó, estas preguntas correspondientes a la esfera de la informalidad laboral se empezaron a realizar en julio de 2017 en la ENE, debido a las distintas formas de ocupación identificadas a nivel nacional e internacional por la Organización Internacional del Trabajo (OIT). Para más información, visitar <http://www.ine.cl/estadisticas/laborales/informalidad-laboral>

³⁴ Dentro de su formulario, la ENE no incluye opciones de “No sabe” y “No responde”. Por protocolo, al finalizar una encuesta ésta debe tener todas las respuestas completas. Si no se cumple, se devuelve al encuestador/a para que la aplique en terreno nuevamente. En caso de no hacerlo, se elimina el hogar de la muestra. Por ende, todos los potenciales casos de no respuesta parcial quedan clasificados dentro de una respuesta válida del cuestionario. Fuente: procedimientos del Departamento de Estadísticas de Hogares del Instituto Nacional de Estadísticas, conocidos a partir de la experiencia de trabajo de la tesista en esa unidad.

numéricas, respectivamente, cuando el o la informante no conoce la respuesta o se niega a darla.

Así, se revisaron las proporciones de ambos códigos (correspondientes a “No sabe”) en las preguntas que presentaron mayores niveles de respuesta parcial en la encuesta CASEN, identificando las diferencias estadísticamente significativas entre informantes y no informantes. Al igual que en la comparación EME-ENE, se optó por calcular estas diferencias a partir de la prueba *t* para muestras independientes, identificando si éstas se deben o no a la magnitud de la muestra con la *r* para tamaño del efecto.

Si bien en el primer apartado se definió que en la encuesta CASEN la persona informante es aquella que respondió al menos una pregunta en todos los módulos, para la evaluación de la no respuesta parcial se ocupó una conceptualización diferente: dado que al buscar profundizar en la frecuencia de la no respuesta parcial según tipo de informante se debe identificar de la forma más exacta la participación del informante en las preguntas, se ocupó la participación por módulo, independiente de si participó en otros módulos del cuestionario.

Por su parte, la variable de identificación de participación en la encuesta se aplica por módulo y no por pregunta. Por ende, se trabajó bajo el supuesto de que quienes tienen código 1 (“Contesta al menos una pregunta”), respondieron todas las preguntas del módulo. Este supuesto es imposible de comprobar, pero se basa en la conjetura de que si participó fue por dos razones: o estuvo encargado de responder todo el módulo, u otra persona (el o la informante *proxy*) estaba respondiendo por él o ella y, al no saber una respuesta, se le consultó directamente. Así, existían dos posibilidades de participación:

1. La persona sobre la cual se pregunta sí tuvo participación, es decir, él o ella misma fue informante.
2. La persona sobre la cual se pregunta no tuvo participación, es decir, otra persona fue informante.

Con esta conceptualización definida, fue necesario depurar la base de datos de cara al análisis. Para esto, se creó una nueva variable por módulo a partir de las variables de identificación de informante de la CASEN mencionadas anteriormente, a modo de

diferenciar a las y los informantes *proxy* según la participación de la persona sobre la cual se pregunta. Esta variable se conformó de dos categorías:

- a. Sí (“1”), para quienes tengan un “1” (Contesta al menos una pregunta) en la variable de identificación de informante de cada módulo.
- b. No (“2”), para quienes tengan “2” (Presente, pero no contesta) o “3” (No está presente) en la variable de identificación de informante de cada módulo.

Por otro lado, las variables analizadas se recodificaron para que tuvieran dos categorías: “1” para respuestas “No sabe” y “0” para las otras respuestas válidas. Los valores perdidos se mantuvieron como tal. Así, la nueva variable permitió calcular el porcentaje de no respuesta parcial (sobre los casos válidos) y el del total de las otras categorías.

Además de contrastar la no respuesta parcial en términos globales de acuerdo a si la persona en cuestión fue informante o no, se realizó un segundo análisis problematizando el tamaño del hogar, clasificando estos según el número de integrantes, bajo el supuesto de que en los hogares con más integrantes la no respuesta parcial sería mayor. Se trabajó con cuatro categorías:

- a. Comparación entre hogares de 2 personas.
- b. Comparación entre hogares de 3 personas.
- c. Comparación entre hogares de 4 personas
- d. Comparación entre hogares de 5 personas o más.

Al analizar por tamaño del hogar se eliminaron los hogares unipersonales y aquéllos con el mismo número de informantes e integrantes por módulo, ya que no permiten una comparación al tener solo informantes directos/as. Por otro lado, permitió identificar si es que el/la informante *proxy* tenía un mejor desempeño en hogares de tamaños determinados.

Análisis

A continuación, se presentan los resultados de los análisis presentados en el marco metodológico. Todas las tablas y gráficos tienen como fuente las bases de datos indicadas en dicho apartado.

Caracterización de los y las informantes

La muestra de la ENE 2010-2017 se compone de 3.448.208 casos, de los cuales 1.085.138 son informantes. Es necesario tener en consideración que la cantidad de casos calificados como informantes no refiere a personas, pues al ser la ENE una muestra con panel, es posible que la persona informante se haya repetido en las distintas visitas. La cantidad de informantes corresponde al 31,5% de casos de la muestra total, respondiendo cada uno por 3,2 personas. Ya que el módulo de mercado laboral de la ENE es contestado por quienes tienen 15 años o más (corte de edad para la población en edad de trabajar), la encuesta se ha aplicado de manera completa para 2.725.021 personas. La cantidad de informantes equivale al 39,8% de las personas de 15 años o más. Cada uno ha contestado por 2,5 personas a las que se les aplica el cuestionario completo.

Por su parte, la CASEN 2015 cuenta con un total de 266.968 casos. De éstos se eliminaron aquellos correspondientes a hogares en que ninguno de sus integrantes había sido informante en todos los módulos, quedando un total de 256.830 casos.

De esos casos, las personas informantes corresponden a 95.556, equivalentes al 37,2% de la muestra y dando cuenta de que en promedio cada informante respondió por 2,7 personas. Las personas de 12 o más años (susceptibles de responder todos los módulos de la CASEN) son 215.529, por lo que cada informante respondió por 2,3 personas a las que se les aplicaba el cuestionario en su totalidad.

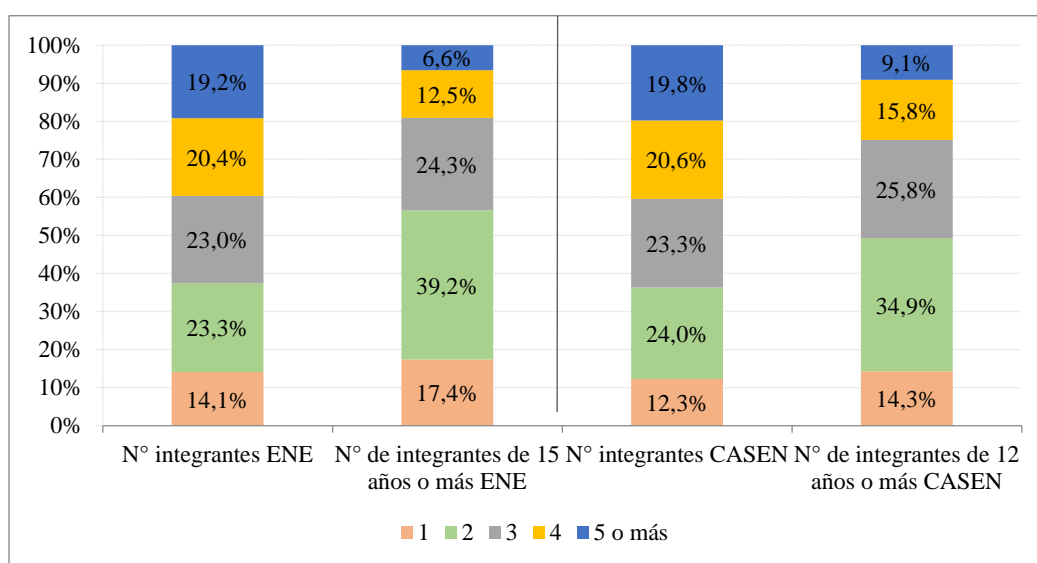
Tabla 8: Muestra e informante CASEN 2015 y ENE 2010-2017

	ENE 2010-2017	CASEN 2015
Muestra total	3.448.208	266.968
Muestra a la que se aplican todos los módulos	2.725.021	215.529
Número de informantes	1.085.138	95.556
Porcentaje de informantes respecto a la muestra a la que se aplican todos los módulos	39,8%	44,3%
Personas susceptibles de responder todos los módulos, por las que responde la persona informante	2,5	2,3

A nivel muestral, los hogares se componen en su mayoría de tres personas o menos en ambas encuestas: en la ENE son el 60,4% y en CASEN el 59,6%. Al filtrar por edad), la estructura varía: en la ENE el 80,9% de los hogares se compone de tres o menos personas susceptibles de contestar todos los módulos, mientras que en CASEN alcanza el 75,0%.

En el otro extremo, el 19,1% de los hogares de la ENE se compone de cuatro o más personas en edad de contestar todos los módulos de la encuesta. En CASEN este porcentaje es de 24,9%.

Gráfico 1: Número de integrantes en los hogares encuestados en ENE 2010-2017 y CASEN 2015



Además, existe una predominancia de informantes *proxy* sobre los directos/as, como se observa en la tabla 2:

Tabla 9: Composición de las personas informantes *proxy* según tipo de informante

Tipo de informante	ENE 2010-2017		CASEN 2015	
	n	%	n	%
Directo/a (informa sólo por sí mismo/a)	153.359	14,1	19.648	20,6
<i>Proxy</i> que informa por sí mismo/a y más personas	931.779	85,9	75.908	79,4
Total	1.085.138	100	95.556	100

Este capítulo se dividió en dos grandes apartados: el primero sobre la totalidad de las personas informantes, y el segundo sobre las particularidades de informantes directos/as y *proxy*.

Caracterización de todas las personas informantes

Una de las primeras observaciones al analizar a las personas informantes, es la predominancia de las mujeres. En CASEN el 34,3% de las personas informantes son hombres y el 65,7% mujeres, mientras que en promedio³⁵ en la ENE los/las informantes se conforman en un 33,7% de hombres y 66,3% de mujeres,

Esta tendencia ha sido consistente en la ENE desde su implementación en el 2010, constituyendo en general casi dos tercios del total.

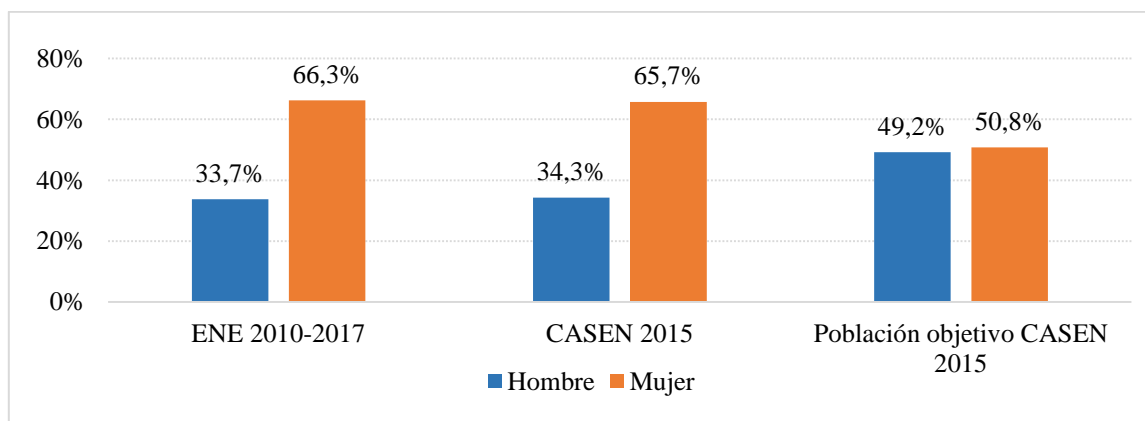
Esto contrasta con la composición total del país³⁶: de acuerdo a los datos expandidos de la CASEN 2015, en ese año los hombres eran el 49,2% y las mujeres el 50,8%³⁷, por lo que en ambas encuestas existiría una sobrerrepresentación de las mujeres como informantes.

³⁵ Los promedios de la ENE se calculan sobre el total de casos desde enero de 2010 hasta diciembre 2017, y no de manera anualizada.

³⁶ Obtenida, como se indicó en el marco metodológico, a partir de los resultados expandidos de la CASEN 2015

³⁷ Para comparar con las encuestas, los resultados de la población consideran a las personas de 12 años o más. Si bien no cumple el requisito de la ENE (15 años o más), se utiliza ya que no se observan cambios en las

Gráfico 2: Proporción según sexo de las personas informantes de la ENE 2010-2017 y Casen 2015, y en la población total según población objetivo CASEN 2015



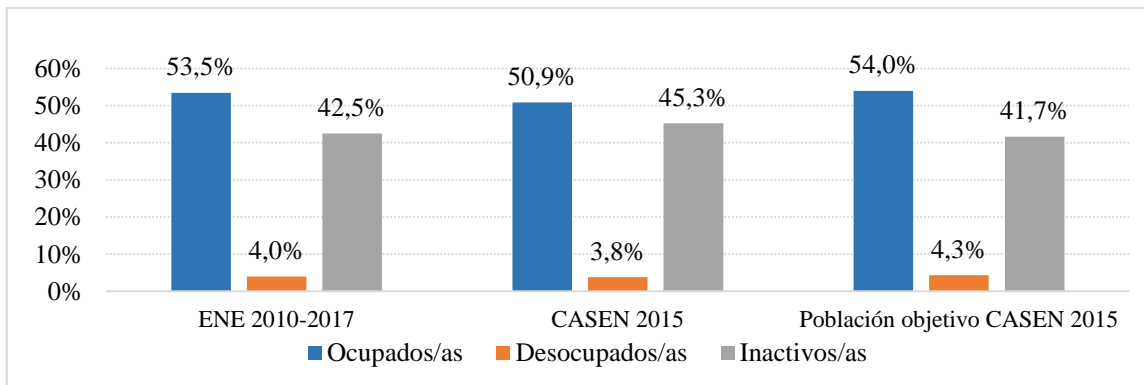
La mayor participación de mujeres como informantes en encuestas de este tipo da cuenta de que ellas o tienen mejor disposición para responder y/o se encuentran presentes en mayor medida en las viviendas. Si bien no es posible indagar en las disposiciones a responder, sí se puede realizar una caracterización de las personas informantes en cuanto a su actividad, puntualmente en su situación ocupacional, sobre todo en consideración de que las mujeres presentan mayores indicadores de inactividad que los hombres: como referencia, de acuerdo a la ENE, en el trimestre móvil abril-mayo-junio 2018, la inactividad fue de un 51,0% en mujeres en edad de trabajar (15 años o más) y de 29,0% en hombres³⁸.

Ambas encuestas muestran una distribución similar para sus informantes respecto a la población general en cuando a la condición de actividad, como se observa en el gráfico 3:

tendencias según ese filtro de edad y permite capturar la totalidad de casos de CASEN. Este dato para personas de 15 años o más sería de 46,3% de hombres y 53,7% de mujeres.

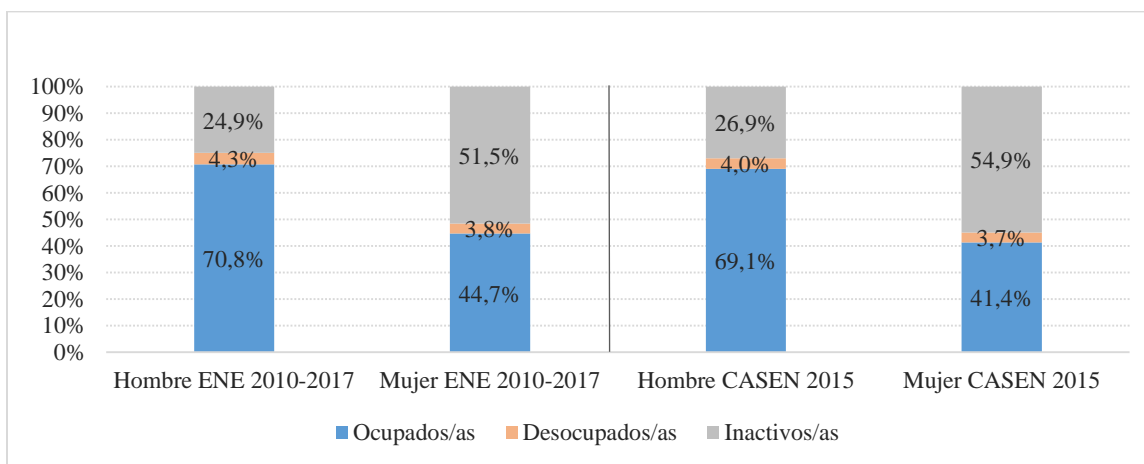
³⁸ Datos calculados a partir de base de datos disponible en <http://www.ine.cl/estadisticas/laborales/ene/base-de-datos>

Gráfico 3: Condición de actividad de informantes ENE 2010-2017 y CASEN 2015, y población objetivo CASEN 2015



Al analizar la condición de actividad entre informantes por sexo, efectivamente las mujeres son el grupo con más población inactiva, conformando el 51,5% de las mujeres informantes de la ENE y el 54,9% del mismo grupo de CASEN, cifra muy similar al parámetro poblacional indicado anteriormente. Incluso, la cifra de inactividad en informantes hombres también es cercana a la correspondiente a la población masculina.

Gráfico 4: Condición de actividad informantes ENE 2010-2017 y CASEN 2015 por sexo



Además, entre los/las informantes inactivos existen distintos motivos que explican la razón por la cual no se encuentran dentro de la fuerza de trabajo.

En la ENE, como muestra el gráfico 5, el motivo mayoritario para estar inactivo de los hombres es el que existe otro ingreso (51,0%) por jubilación, pensión o renta, seguido en una proporción notoriamente menor por los problemas de salud (16,4%). Entre mujeres la razón de inactividad se relaciona a que están a cargo de los quehaceres del hogar (47,6%), seguido de la existencia de otros ingresos (21,0%) y los problemas de salud (13,9%).

En la CASEN 2015 (gráfico 6), 7 de cada 10 hombres inactivos lo es porque tiene otra fuente de ingresos (68,2%). La frecuencia de otros motivos es mucho menor: el 12,2% declara estar inactivo por estudios, mientras que el 8,3% indica presentar problemas de salud.

En las mujeres si bien el tener otros ingresos es un motivo relevante (30,3%), la inactividad se debe principalmente a que se encargan de tareas del hogar (51,7%). En tercer lugar se ubican los problemas de salud con 6,5%.

El énfasis en los quehaceres del hogar o trabajo doméstico es relevante, ya que los hombres que están inactivos por quehaceres domésticos y que son informantes, conforman sólo el 3,9% en la ENE y el 2,7% en CASEN.

Gráfico 5: Razones de inactividad informantes ENE 2010-2017 por sexo

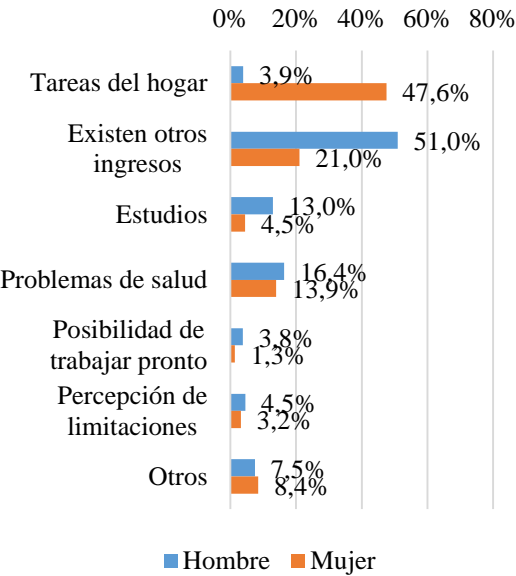
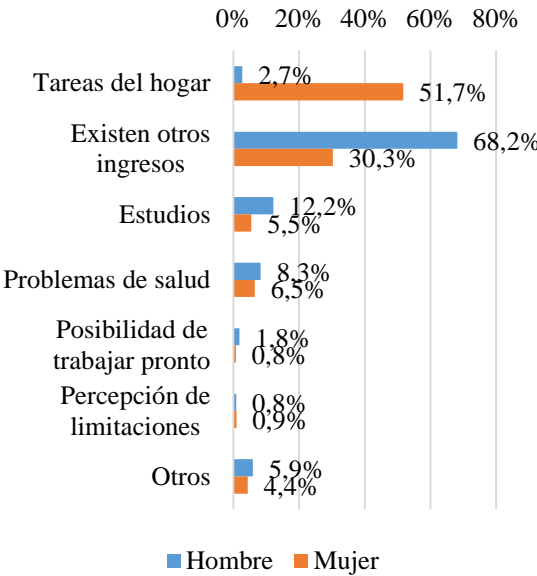


Gráfico 6: Razones de inactividad informantes CASEN 2015 por sexo



Así, la figura del trabajo doméstico adquiere una notoria presencia entre las personas informantes en ambas encuestas, siendo el segundo grupo con más frecuencia después de los/las ocupados. Además, especialmente en la CASEN 2015 existe una alta presencia también de inactivos/as que no son parte de la fuerza del trabajo porque reciben otros ingresos:

Gráfico 7: Condición de actividad y razones de inactividad informantes ENE 2010-2017

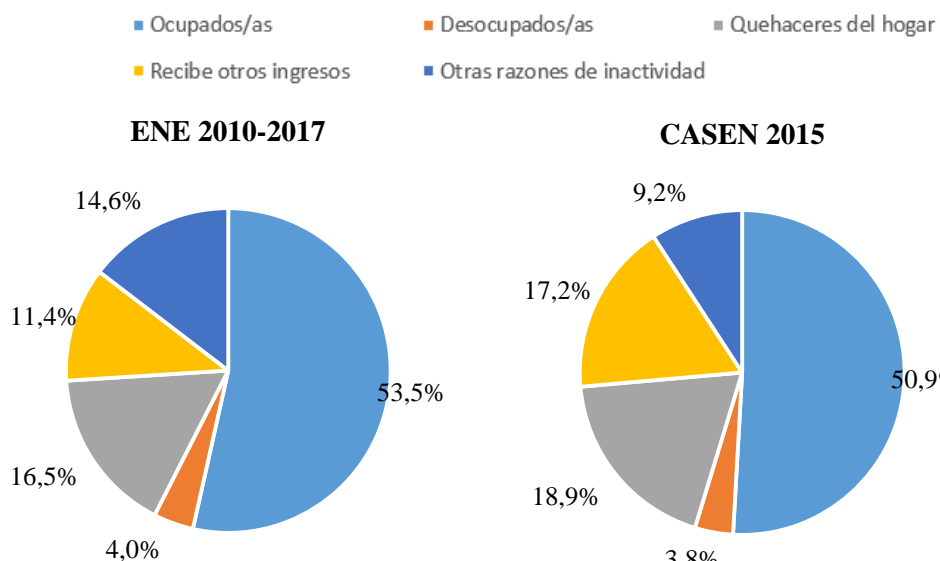
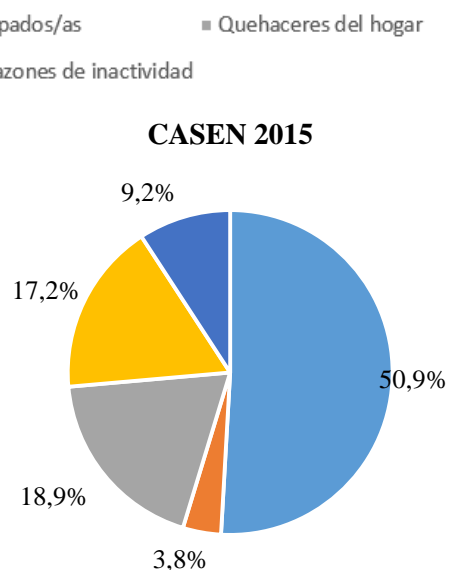


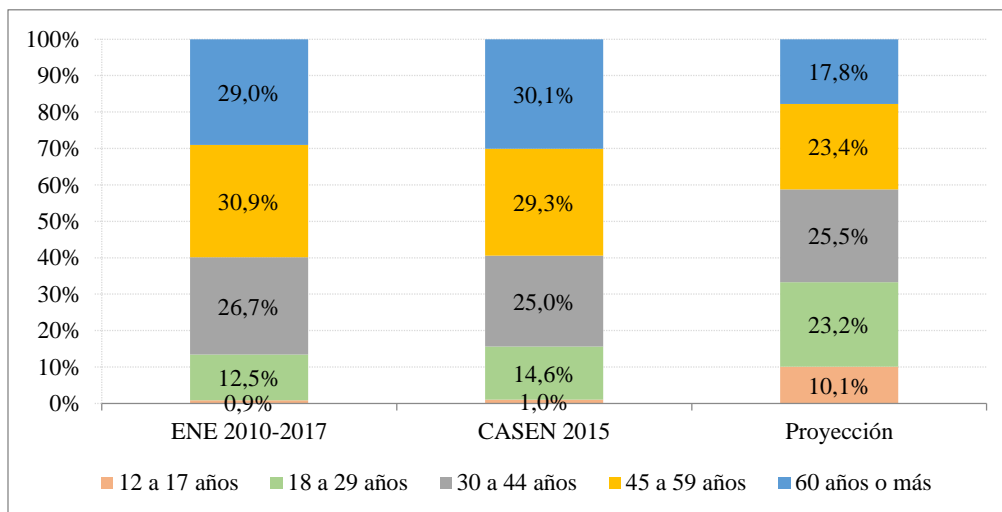
Gráfico 8: Condición de actividad y razones de inactividad informantes CASEN 2015



Al caracterizar por sexo, la mayoría de las personas informantes que se encuentran inactivas por quehaceres del hogar son mujeres, constituyendo el 16,2% en la ENE y el 18,7% en la CASEN del total de informantes, en contraste con el 9,6% y 11,2%, respectivamente, de mujeres de la población que se dedican al trabajo doméstico.

En cuanto a la edad, los promedios de tramos etarios en la ENE 2010-2017 y la CASEN 2015 son los observados en el gráfico 9:

Gráfico 9: Tramos etarios de informantes de la ENE 2010-2017 y CASEN 2015, y de las proyecciones de población INE



En ambas encuestas los tramos etarios muestran distribuciones similares, con una mayor presencia de informantes de 60 años o más (29,0% en la ENE y 30,1% en la CASEN) y de 45 a 59 años (30,9% y 29,3%, respectivamente). Esto contrasta con la población, en donde la distribución es más bien uniforme. De hecho, el 41,2% de la población tiene 45 años o más, mientras que en la ENE es el 59,9% y en CASEN el 59,4%, existiendo una diferencia de 18,7 y 18,2 puntos porcentuales, respectivamente.

Por el contrario, la presencia de informantes menores de edad es marginal, rondando el 1,0% en ambas encuestas, aun cuando en la población representan 10 veces esa proporción (10,1%).

Por otro lado, la edad promedio de las personas informantes en ambas encuestas es muy similar, siendo en la ENE de 49,5 años, y de 49,4 años en la CASEN. Esto también presenta un mayor “envejecimiento” de las encuestas, ya que en la población la edad promedio es de 42,2 años desde los 15 años y 40,8 años desde los 12.

Finalmente, se busca caracterizar a las personas informantes en términos del último nivel educacional aprobado.

En ENE y CASEN se observan distribuciones relativamente similares, las que dan cuenta de que las personas informantes mayoritariamente cuentan con educación media finalizada (35,1% y 34,8%, respectivamente). Existen matices en cuanto a los niveles inferiores: en la ENE el segundo grupo mayoritario no cuenta con un nivel educacional finalizado (25,1%), mientras que en CASEN son quienes tienen educación básica (28,0%).

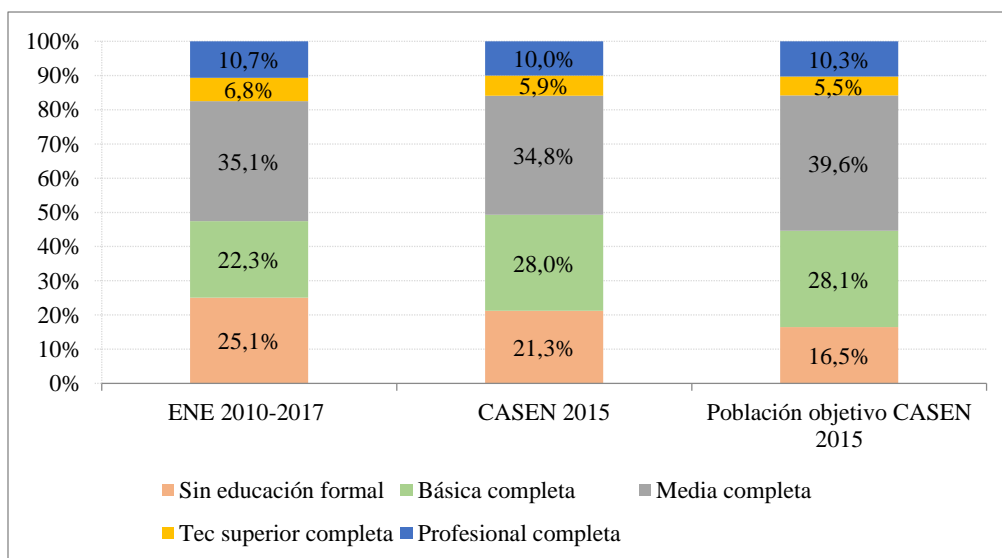
La presencia de personas informantes con educación superior finalizada es de 17,5% en la ENE (10,7% profesional y 6,8% técnica), y de 15,9% en la CASEN (10,0% y 5,9%, respectivamente), siendo en ambas encuestas el grupo con menor frecuencia entre los/las informantes.

En cuanto a la población se observan similitudes en la distribución, existiendo la mayor diferencia entre las personas que no han aprobado ningún nivel educacional: este grupo tendría menor presencia entre informantes CASEN (21,3%), teniendo 4,8 puntos porcentuales más respecto a la población objetivo de la misma encuesta (16,5%). En la ENE el 25,1% de las personas informantes no cuenta con ningún nivel educacional,

acentuando la distancia con este grupo a 8,6 puntos porcentuales. De hecho, en esta encuesta es el segundo grupo con mayor presencia, mientras que entre informantes CASEN y en la población es el tercero.

Otro grupo que marca diferencia con la población son quienes cuentan con educación media completa, distanciándose del parámetro en 4,5 y 4,8 puntos porcentuales la ENE y la CASEN, respectivamente. Finalmente, los informantes de la ENE en particular se distancian de la población respecto a las personas con educación básica en 5,8 puntos porcentuales. En ese mismo ítem la CASEN no presenta mayores diferencias.

Gráfico 10: Nivel educacional más alto aprobado para informantes ENE 2010-2017 y CASEN 2015, y población objetivo CASEN 2015



A partir de estas descripciones, se observa que las características sociodemográficas son similares para los informantes en ambas encuestas.

Sin embargo, al aplicar la prueba Phi, se concluye que las relaciones existentes entre las variables sociodemográficas con ser informante, son débiles o moderadas. Las que presentan relaciones menos robustas son la situación en el mercado laboral y el nivel educacional, mientras que el sexo y sobre todo la edad tienden a asociarse más.

Tabla 10: Valores prueba Phi en ENE 2010-2017 y CASEN 2015

Variables	Informante o no ENE		Informante o no CASEN	
	Valor p	Phi	valor p	Phi
Sexo	<0,001	0,213	<0,001	0,207
Tramo etario	<0,001	0,283	<0,001	0,356
Situación en el mercado laboral	<0,001	0,026	<0,001	0,032
Nivel educacional	<0,001	0,088	<0,001	0,199

Para conocer cuáles son las categorías de las variables de caracterización que más tienen relación con el ser o no informante, se realizó un análisis de correspondencias múltiples (ACM).

En la ENE, para el período 2015-2017³⁹, se ocuparon 999.159 casos. Éstos corresponden a las personas de 15 años o más (es decir potenciales informantes) de la ENE en el periodo indicado. De acuerdo al resumen del modelo, las dos dimensiones creadas explican el 73,3% de la inercia (proxy de varianza).

Tabla 11: Resumen del modelo de ACM para ENE 2015-2017

Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	Inercia
1	,590	1,895	,379
2	,544	1,770	,354
Total		3,665	,733
Media	,568 ^a	1,832	,366

a. El Alfa de Cronbach Promedio está basado en los autovalores promedio.

³⁹ Se ha utilizado este período debido a las limitaciones del software al ingresar más años.

En función de las coordenadas, las distancias de las categorías con ser o no informante, son:

Tabla 12: Distancias de las categorías de caracterización sociodemográfica con las categorías de “Informante” y “No informante”, ENE 2015-2017

Categorías	Distancia con Informante	Distancia con No informante
12-17	3,60	2,66
18-29	1,44	0,48
30-44	0,92	1,06
45-59	0,60	0,76
60 o más	1,18	1,72
Sin educación formal	1,19	1,62
Básica	1,47	0,43
Media	0,85	0,64
Superior	0,77	1,14
Hombre	0,95	0,21
Mujer	0,42	0,71
Ocupado/a	0,80	0,79
Desocupado/a	1,19	0,77
Otras razones	2,04	1,13
Tareas del hogar	0,13	1,22
Tiene otros ingresos	1,84	2,47

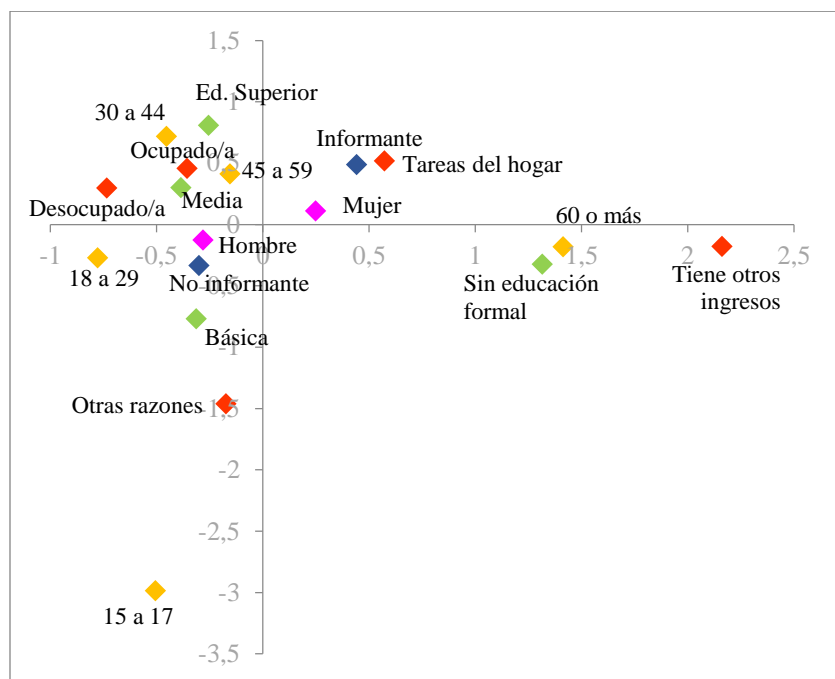
Así, las categorías que presentan una menor distancia con ser informante, son:

- Inactividad por tareas del hogar (0,13)
- Mujer (0,42)
- Edad entre 45 y 59 años (0,60)

Por el otro lado, las categorías más cercanas con no ser informante, son:

- Hombre (0,21)
- Educación básica completa (0,43)
- Edad entre 18 y 29 años (0,48)

Gráfico 11: Mapa perceptual por el Análisis de Correspondencias Múltiples ENE 2015-2017



Llama la atención las categorías que se acercan a ambas categorías: Mujer, Edad entre 45 y 59 años y Estar ocupado/a. Esto podría estar influenciado por el hecho de que los no informantes son el doble que los informantes, estando conformado por una amplia variedad de personas. Sin embargo, cabe destacar que las dos primeras categorías se encuentran más firmemente relacionadas a los informantes, mientras que estar ocupado/a se mantiene prácticamente equidistante a ambas categorías.

En el caso de la CASEN 2015, el análisis ocupó 215.529 casos, correspondientes a las personas de 12 años o más que podían ser informantes. A partir del resumen del modelo se observa la creación de dos dimensiones. Ambas tienen valores relativamente similares en la inercia, explicando el 76,7% de ésta (proxy de varianza).

Tabla 13: Resumen del modelo de ACM para CASEN 2015

Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	Inercia
1	,622	1,990	,398
2	,573	1,846	,369
Total		3,837	,767
Media	,598 ^a	1,918	,384
a. El Alfa de Cronbach Promedio está basado en los autovalores promedio.			

Por otro lado, de acuerdo a las coordenadas de cada punto, las distancias que presentan cada una de las categorías de las variables de caracterización sociodemográfica con las categorías “Informante” y “No informante”, son:

Tabla 14: Distancias de las categorías de caracterización sociodemográfica con las categorías de “Informante” y “No informante”, CASEN 2015

Categorías	Distancia con Informante	Distancia con No informantes
12-17	3,01	1,85
18-29	1,54	0,70
30-44	0,87	1,26
45-59	0,49	1,06
60 o más	1,26	1,85
Sin educación formal	1,16	1,42
Básica	1,22	0,55
Media	1,10	0,84
Superior	0,79	1,42
Hombre	1,10	0,29
Mujer	0,47	0,91
Ocupado	0,83	1,00
Desocupado	1,26	0,87
Otras razones	2,11	0,77
Tareas del hogar	0,16	1,52
Tiene otros ingresos	1,74	2,33

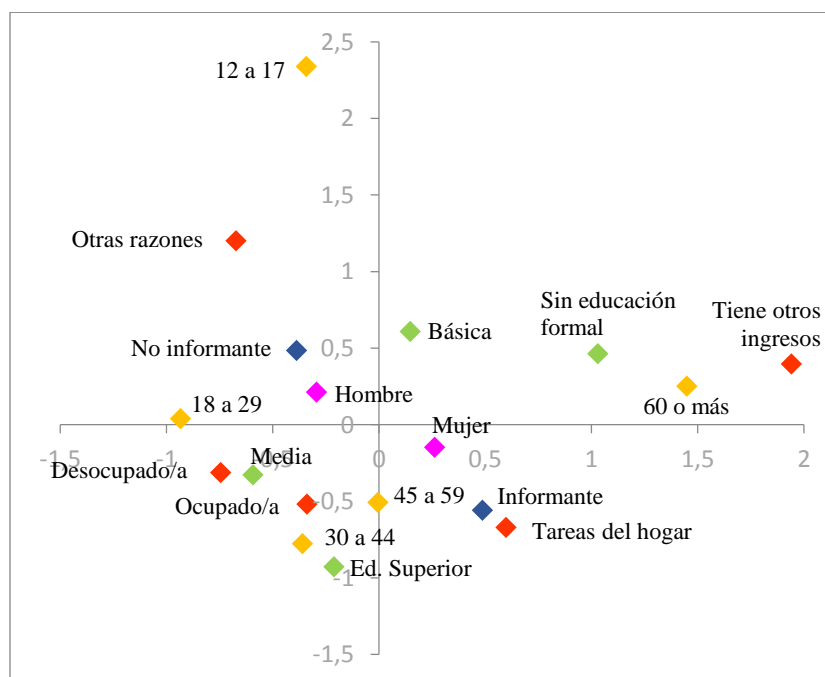
Las categorías que presentan menor distancia con ser informante son:

- Inactividad por tareas del hogar (0,16)
- Mujer (0,47)
- Edad entre 45 y 59 años (0,49)

Por el contrario, las categorías que presentan menor distancia con el punto correspondiente a los/las no informantes, son:

- Hombre (0,29)
- Educación básica completa (0,55)
- Edad entre 18 y 29 años (0,70)

Gráfico 12: Mapa perceptual por el Análisis de Correspondencias Múltiples CASEN 2015



En ambas encuestas los resultados coinciden tanto a nivel de las frecuencias como en el análisis de correspondencias, identificando las mismas categorías como las más cercanas a ser informante: dedicarse a los quehaceres del (0,13 en ENE y 0,16 en CASEN), ser mujer (0,42 y 0,47, respectivamente) y tener entre 45 y 59 años (0,60 y 0,49, respectivamente).

También en los no informantes, ambas encuestas coinciden en cuáles son las categorías más cercanas: ser hombre, con una distancia similar en ambas encuestas (0,21 en ENE y 0,29

en CASEN), seguido de contar con educación básica completa (0,43 y 0,55, respectivamente) y tener entre 18 y 29 años (0,48 y 0,70, respectivamente).

Figura 9: Categorías más relacionadas con “Informante” y “No informante”

Informante	No informante
<ul style="list-style-type: none"> •Quehaceres del hogar •Mujer •Edad entre 45 y 59 años 	<ul style="list-style-type: none"> •Hombre •Educación básica completa •Edad entre 18 y 29 años

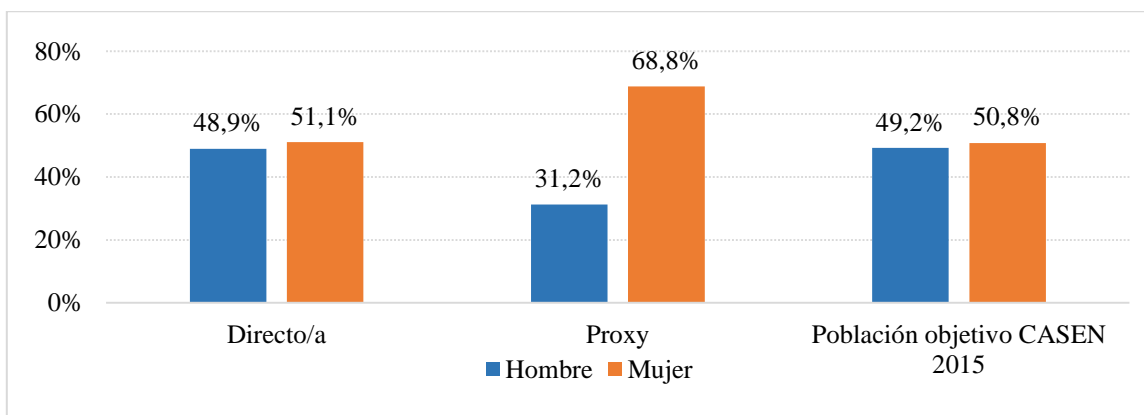
Resulta contundente que ambas encuestas, ocupando distintos marcos muestrales y períodos de aplicación, coincidan de manera tan clara en las principales categorías tanto en informantes como en no informantes. Sin duda, da cuenta de perfiles definidos de personas que responden encuestas aplicadas en viviendas particulares.

Caracterización por tipo de informante

Este apartado abordará el comportamiento de las variables sociodemográficas recién revisadas, distinguiendo entre informantes *proxy* y directos/as.

En cuanto al sexo, su distribución en informantes directos/as es prácticamente idéntica que la de la población, mientras que para los/las *proxy* existe una proporción de 2:1 de mujeres sobre hombres.

Gráfico 13: Proporción según sexo en ENE 2010-2017 y CASEN 2015 según tipo de informante, y población objetivo CASEN 2015

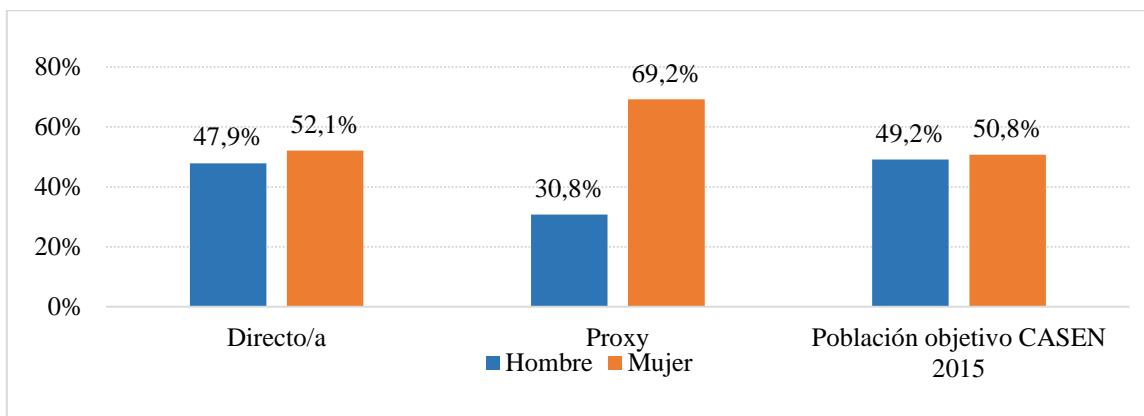


Una situación muy similar ocurre en las personas informantes de la encuesta CASEN 2015, en donde los y las informantes directos/as tienen una distribución más bien similar a la

población, mientras que la mayoría de mujeres se acentúa en informantes *proxy* (en este caso, 69,2% sobre 30,8% de hombres).

Así, en cuanto al sexo, los informantes directos/as representarían mejor el parámetro poblacional, tanto en la ENE 2010-2017 como en la CASEN 2015.

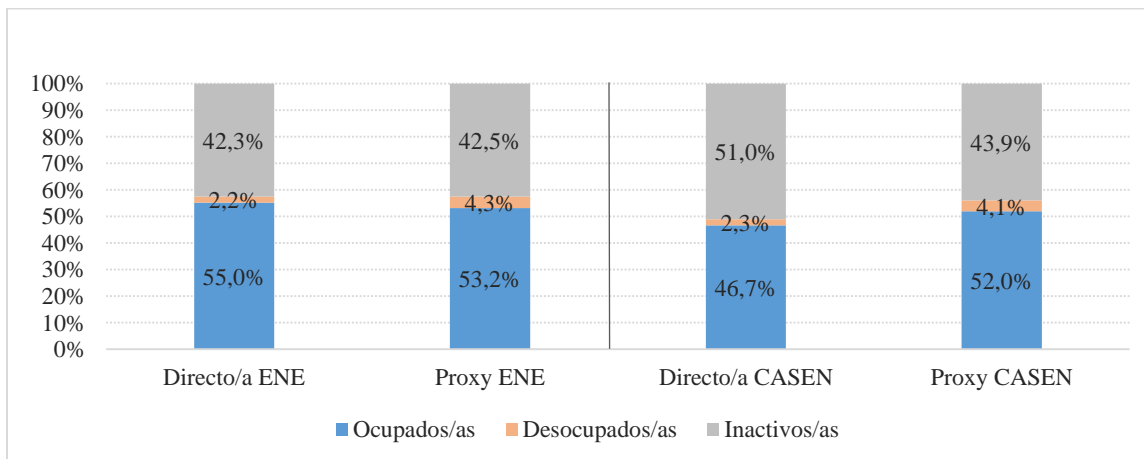
Gráfico 14: Proporciones de hombres y mujeres en las personas informantes de la CASEN 2015 según tipo de informante, y población objetivo CASEN 2015



Al analizar la condición de actividad, en la ENE existe una relativa regularidad general en ésta en ambos grupos.

En la CASEN existe una mayor presencia de inactivos/as entre informantes directos/as (51,0%) respecto a los/las proxy (43,9%). Esto indica que las personas informantes en hogares unipersonales o con la misma cantidad de informantes que de integrantes de la CASEN (definición de informante directo/a entregada en el marco metodológico), en su mayoría son personas inactivas, al contrario del resto de informantes en ambas encuestas.

Gráfico 15: Condición de actividad según tipo de informante ENE 2010-2017 y CASEN 2015



En cuanto a las razones de inactividad, mientras los/las informantes directos indican estar inactivos/as mayoritariamente por tener otros ingresos (57,6% en ENE y 67,8% en CASEN), los/las informantes *proxy* lo están debido a que realizan quehaceres del hogar (44,4% en ENE y 49,8% en CASEN).

Gráfico 16: Razones de inactividad informantes ENE 2010-2017 por tipo de informante

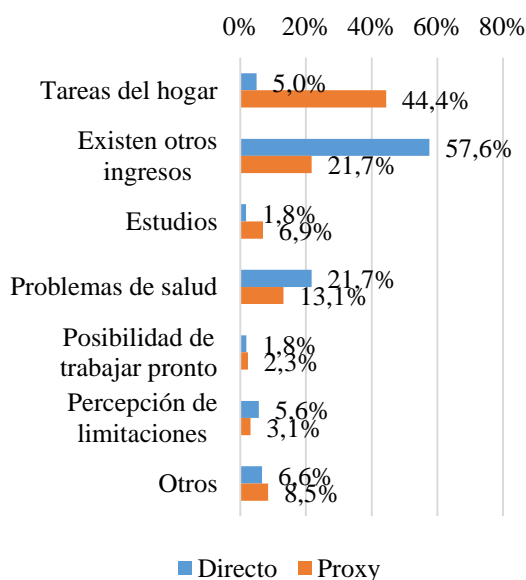
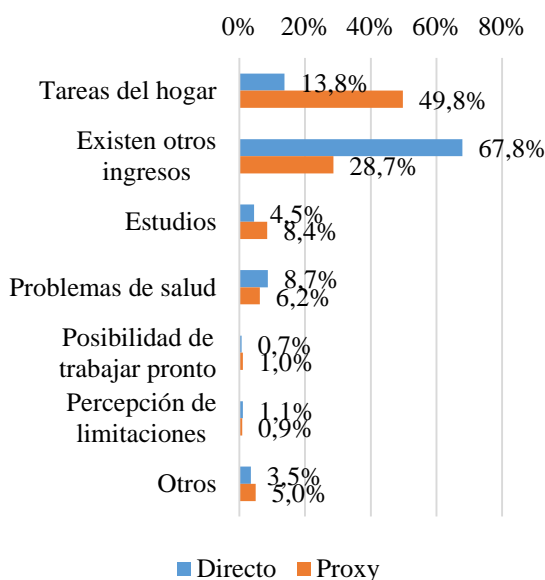
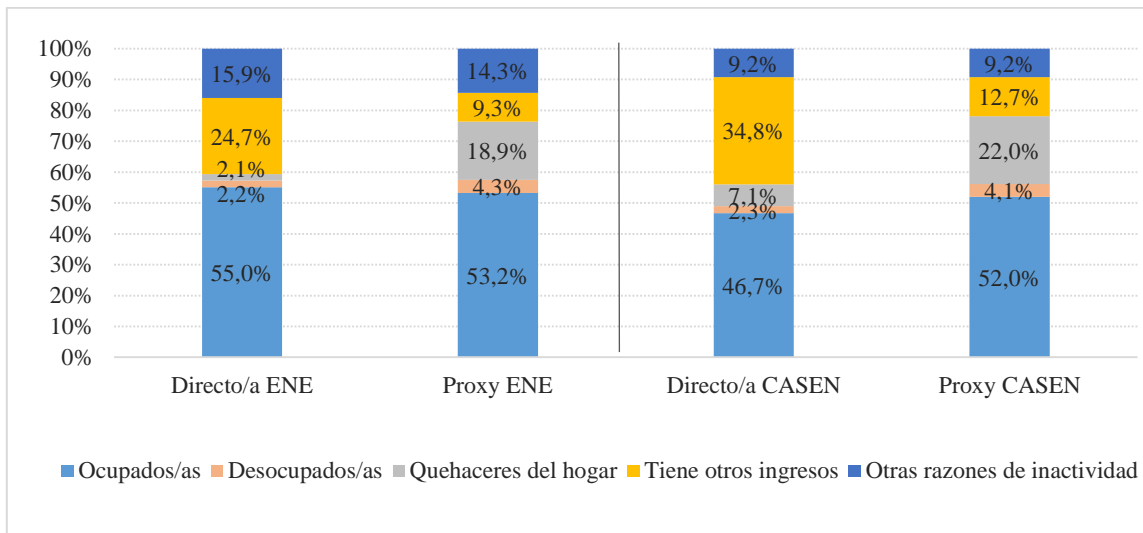


Gráfico 17: Razones de inactividad informantes CASEN 2015 por tipo de informante



Estas diferencias marcan brechas similares en ambas encuestas respecto a la condición de actividad y las razones de inactividad de sus informantes. En informantes directos/as de ENE y CASEN las personas ocupadas (55,0% y 46,7%, respectivamente) son seguidas de aquéllas que están inactivas por tener otros ingresos (24,7% y 34,8%, respectivamente). Sin embargo, es entre los/las *proxy* que el trabajo doméstico se constituye como segunda mayoría tras los/las ocupados, siendo el 18,9% de los informantes *proxy* de la ENE, y el 22,0% del mismo grupo de la CASEN.

**Gráfico 18: Condición de actividad y razones de inactividad según tipo de informante
ENE 2010-2017 y CASEN 2015**

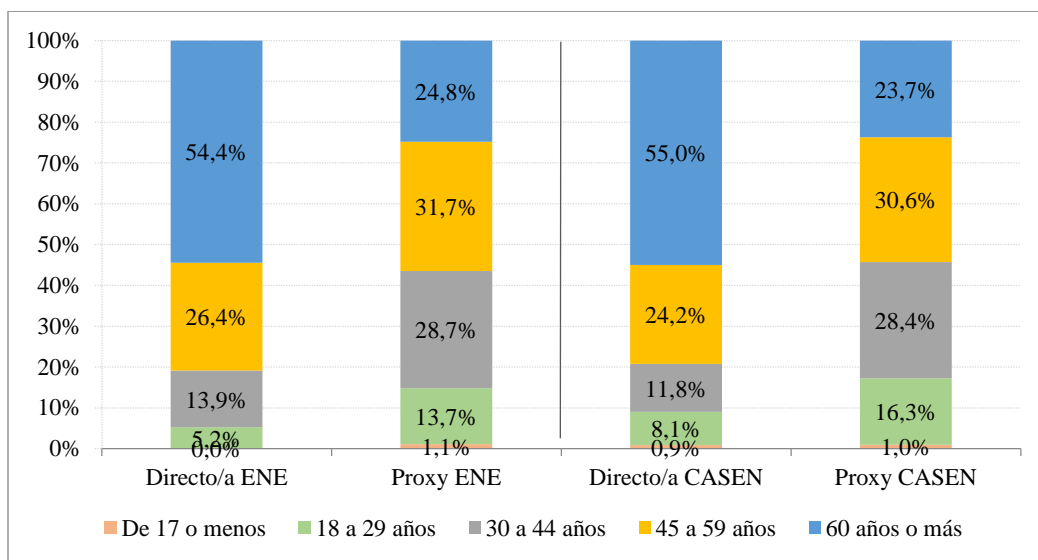


En cuanto a la edad, se hacen evidentes las diferencias en la distribución de ésta tanto en la ENE como en la CASEN: todas las categorías de edades presentan diferencias entre tipos de informantes en las dos encuestas.

La categoría “60 años o más” es la principal en informantes directos/as, abarcando al 54,4% y 55,0% de ellos/as en la ENE y CASEN, respectivamente. Entre informantes *proxy* ocupa un rol mucho menos preponderante, bajando a 24,8% y 23,7%, respectivamente. En este grupo, los tramos con mayor presencia son el de 45 a 59 años (31,7% en ENE y 30,6% en CASEN) y el de 30 a 44 años (28,7% y 28,4%, respectivamente). Además, también aumenta –respecto a los informantes directos/as–, la participación de personas de 18 a 29 años.

Por otro lado, en la ENE la media de la edad para informantes directos/as es de 59,8 años y para informantes *proxy* es de 47,8 años. EN CASEN las edades son 59,0 y 46,9 años, respectivamente. Como se mencionó en el apartado anterior, la edad promedio en la población es de 42,2 años desde los 15 años y 40,8 años desde los 12. Así, las personas informantes *proxy* tenderían a acercarse mayormente al parámetro poblacional, teniendo, en promedio, 12 años menos que los directos/as.

Gráfico 19: Tramos etarios según tipo de informante en ENE 2010-2017 y CASEN 2015



En consideración de los datos aportados desde las proyecciones de edad, se observa que los/las informantes *proxy* tienden a distribuirse de una manera más similar a la población general, en contraste con el sexo, variable que presentaba similitudes entre la población e informantes directos/as. De todas maneras, existiría entre los/las *proxy* una clara subrepresentación de las personas de 29 años o menos.

Finalmente, se analiza el nivel educacional entre tipos de informantes. La menor diferencia se presenta en informantes con educación superior finalizada, la cual fluctúa entre 14,7% y 18,0% en todos los grupos. Llama la atención que las proporciones para este nivel en general son mayores en la ENE que en la CASEN, independiente del tipo de informante.

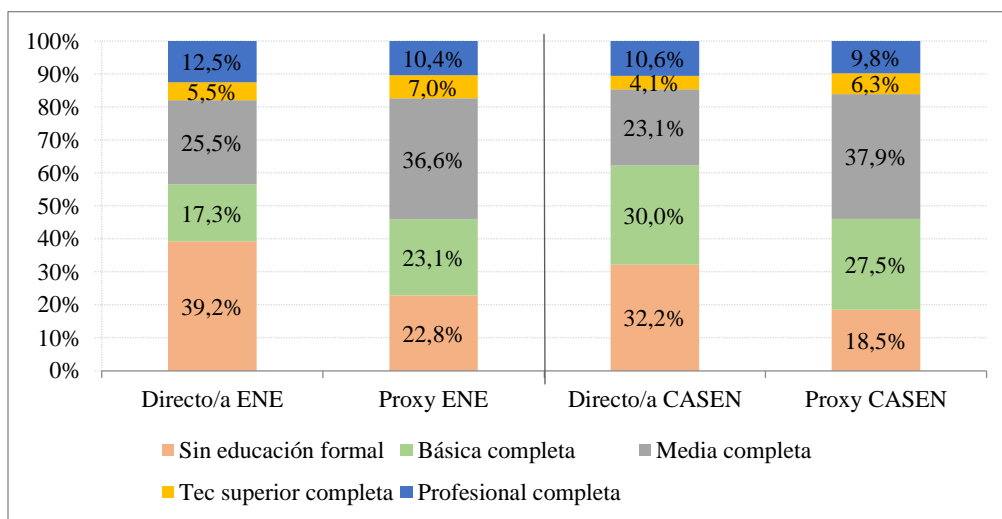
Ambos perfiles de informante directo/a se conforman principalmente por personas sin educación formal (39,2% en ENE y 32,2% en CASEN). En la ENE el segundo grupo con mayor frecuencia es la media completa (25,5%) mientras que en CASEN es la básica completa (30,0%). Este último nivel es el que plantea mayores brechas entre informantes directos/as de ENE y CASEN, existiendo 12,7 puntos porcentuales de diferencia, relegándolo al tercer lugar de predominancia entre informantes ENE.

Por otro lado, las y los informantes *proxy* cuentan principalmente con educación media (36,6% en ENE y 37,9% en CASEN), seguida de educación básica (23,1% y 27,5%, respectivamente). En tercer lugar, y en contraste con los informantes directos/as, son

quienes tienen menos informantes sin educación formal, alcanzando el 22,8% en la ENE y el 18,5% en la CASEN.

Si se compara con el parámetro poblacional, el *proxy* presentaría una adecuación en cuanto a la jerarquía de cada grupo, sobre todo el perteneciente a la CASEN.

Gráfico 20: Nivel educacional más alto aprobado según tipo de informante ENE 2010-2017 y CASEN 2015



Para ver cuáles categorías de las variables utilizadas se relacionan más con ser informante *proxy* o directo/a, se realizó nuevamente el ejercicio de análisis de correspondencias múltiples. En este caso, se excluyó la variable de ser o no ser informante, y se incluyó la de tipo de informante.

En la ENE, el modelo explica el 86,6% de la varianza a partir de las dos dimensiones creadas.

Tabla 15: Resumen del modelo de ACM para ENE 2015-2017

Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	Inercia
1	,703	2,287	,457
2	,638	2,041	,408
Total		4,328	,866
Media	,672 ^a	2,164	,433

a. El Alfa de Cronbach Promedio está basado en los autovalores promedio.

Por otro lado, la cercanía de las categorías analizadas con los tipos de informante son las siguientes:

Tabla 16: Distancias de las categorías de caracterización sociodemográfica con las categorías de “Informante” y “No informante”, ENE 2015-2017

Categorías	Distancia con Informante directo/a	Distancia con Informante <i>proxy</i>
15-17	4,53	4,43
18-29	1,74	1,02
30-44	1,45	0,60
45-59	1,10	0,21
60 o más	0,47	1,35
Sin educación formal	1,12	1,98
Básica	1,92	1,54
Media	1,33	0,44
Superior	1,35	0,61
Hombre	0,68	0,66
Mujer	0,54	0,59
Ocupado	1,22	0,34
Desocupado	1,61	0,72
Otras razones	2,40	2,24
Tareas del hogar	0,78	0,13
Tiene otros ingresos	1,11	2,01

Para los informantes directos/as, las categorías más cercanas son:

- 60 años o más (0,47)
- Mujer (0,54)
- Hombre (0,68)
- Inactividad por tareas del hogar (0,78)

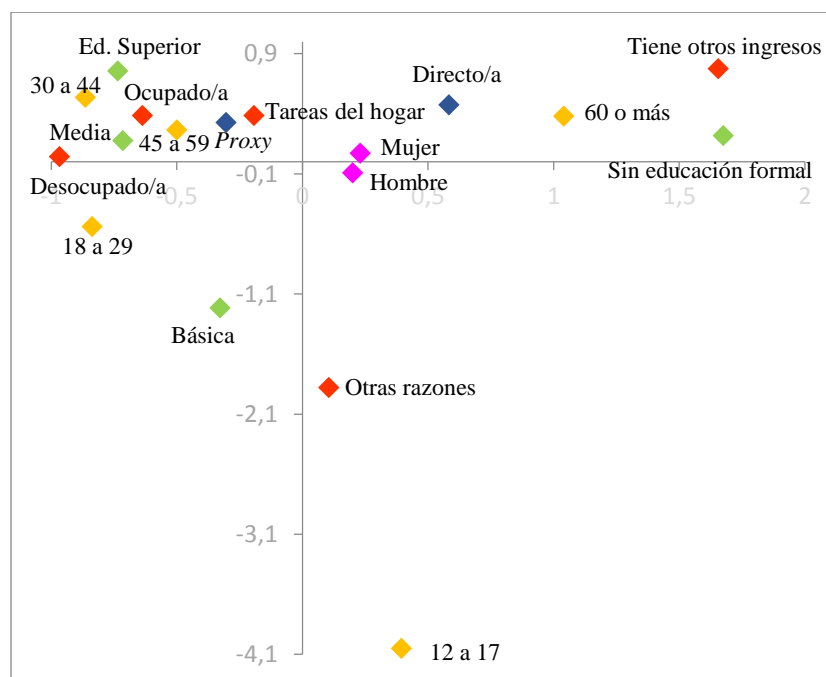
En tanto, las categorías más cercanas a las personas informantes *proxy* son:

- Inactividad por tareas del hogar (0,13)

- 45 a 59 años (0,21)
- Estar ocupado/a (0,34)
- Educación media (0,44)
- Mujer (0,59)
- 30 a 44 años (0,60)
- Educación superior (0,61)
- Hombre (0,66)
- Desocupado/a (0,72)

Un primer análisis da cuenta de cómo las personas informante *proxy* son más diversas, abarcando distintos rangos etarios, situaciones en el mercado laboral, niveles educacionales e incluso sexo, mientras que los informantes directos/as tienden a ser más homogéneos, respondiendo a un perfil muy definido: adultos/as mayores, que se encuentran fuera del mercado laboral, y que tenderían a estar más presentes en sus viviendas.

Gráfico 21: Mapa perceptual por el Análisis de Correspondencias Múltiples CASEN 2015



Por su parte, el modelo realizado en la CASEN 2015 tiene una inercia que explica el 87,3% de la varianza, a partir de la creación de dos dimensiones.

Tabla 17: Resumen del modelo de ACM para CASEN 2015

Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	Inercia
1	,713	2,329	,466
2	,636	2,036	,407
Total		4,365	,873
Media	,677 ^a	2,182	,436
a. El Alfa de Cronbach Promedio está basado en los autovalores promedio.			

La cercanía de cada categoría con ser o no informante, son las siguientes:

Tabla 18: Distancias de las categorías de caracterización sociodemográfica con las categorías de “Informante” y “No informante”, CASEN 2015

Categorías	Distancia con Informante directo/a	Distancia con Informante <i>proxy</i>
12-17	3,66	3,66
18-29	2,21	1,30
30-44	1,59	0,53
45-59	1,16	0,11
60 o más	0,43	1,51
Sin educación formal	0,97	1,81
Básica	1,47	1,25
Media	1,81	0,80
Superior	1,50	0,58
Hombre	1,17	0,87
Mujer	0,79	0,61
Ocupado	1,44	0,37
Desocupado	1,95	0,94
Otras razones	2,70	2,30
Tareas del hogar	0,90	0,21
Tiene otros ingresos	0,96	2,03

Las categorías más cercanas con los informantes directos/as, son:

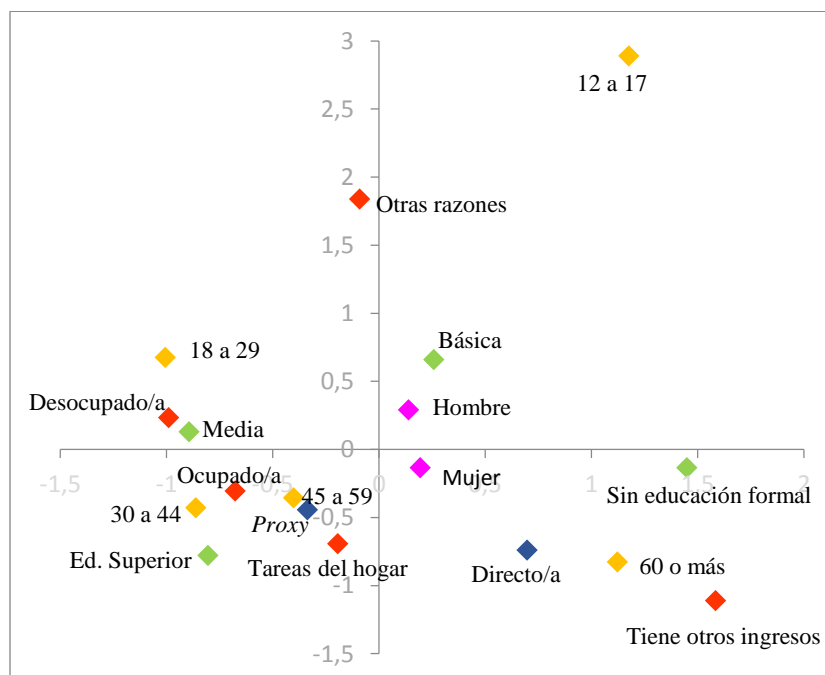
- 60 años o más (0,43)
- Mujer (0,79)
- Inactividad por quehaceres del hogar (0,90)
- Inactividad por tener otros ingresos (0,96)

Por otro lado, las categorías más cercanas a ser informante *proxy* son:

- 45 a 59 años (0,11)
- Inactividad por tareas del hogar (0,21)
- Estar ocupado/a (0,37)
- 30 a 44 años (0,53)
- Educación superior (0,58)
- Mujer (0,61)
- Educación media (0,80)

De manera similar que en la ENE, la CASEN también cuenta con un perfil mucho más definido para informantes directos/as que *proxy*.

Gráfico 22: Mapa perceptual por el Análisis de Correspondencias Múltiples CASEN 2015



Entre ambas encuestas, las características en común para ambos tipos de informantes son:

Figura 10: Categorías más relacionadas con “Informante” y “No informante”

Informante directo/a	Informante <i>proxy</i>
<ul style="list-style-type: none"> •Mujer •Mayores de 60 años •Inactivos/as 	<ul style="list-style-type: none"> •Mujer •Educación media y superior •Edad entre 30 y 59 años •Ocupados/as e inactivos/as por quehaceres del hogar.

A diferencia del análisis realizado sobre el total de informantes, en este caso existen mayores diferencias entre ambas encuestas. Si bien hay consensos (presentados en la figura anterior), sí deja ver que las particularidades asociadas a cada tipo de informante podrán depender de la encuesta y su levantamiento.

Variaciones en respuestas ENE-EME

La V EME, del año 2017, se compuso de una muestra total de 7.492 casos. Ésta se fusionó con la base correspondiente al trimestre julio-agosto-septiembre (JAS) de la ENE 2017, existiendo una coincidencia en 5.854 casos⁴⁰.

De éstos, se eliminaron quienes en la EME declararon haber cambiado de situación laboral respecto a la EME (es decir, ya no eran trabajadores/as independientes), y los que respondieron cada encuesta para un empleo distinto. Esto redujo la muestra a 5.329 casos.

Para el análisis, esta base se dividió en dos: por un lado, estaban los casos que en ambas encuestas habían sido respondidos mediante informante directo/a (grupo control), y por el otro los que en la ENE fueron respondidos por informante *proxy* (grupo experimental). El primer grupo tuvo un tamaño de 2.989 casos, y el segundo de 2.340.

Para resguardar que las comparaciones entre las respuestas dadas en los pares de preguntas fueran efectivamente entre los mismos casos, los tabulados se condicionaron a que las preguntas a comparar tuvieran respuestas en ambas encuestas. De esta manera, todas las frecuencias y tabulados comprenden solo los pares de casos que contestaron las preguntas

⁴⁰ La diferencia en ambas muestras se debe principalmente a la muestra de la ENE que tuvo su última visita en el período MAM.

en ambos instrumentos, no entrando en el análisis los casos que hayan respondido en solo una.

Esto se dirige particularmente a quienes en la ENE (cuyo levantamiento fue mayoritariamente posterior a la EME, al ocuparse el trimestre julio-agosto-septiembre y no el marzo-abril-mayo) declaran estar desocupados/as o inactivos/as, en tanto significa un cambio en la situación en el mercado laboral. Se debe tener en consideración que esto también disminuyó el tamaño de las muestras a comparar, en función de la ausencia de respuesta en cada pregunta.

Tabla 19: Muestras disponibles para los pares de preguntas a comparar

Comparaciones a realizar	Muestra disponible para grupo control	Muestra disponible para grupo experimental
Ocupación (b1 ENE/d19 EME)	2.582	2.010
Actividad económica (b14a ENE/d20 EME)	2.582	2.010
Registro en Servicio de Impuestos Internos (i4 e i5 ENE/ e40 EME)	2.489	1.877
Registros de contabilidad (i6 ENE/e38 EME)	2.489	1.877
Separación de gastos del negocio y del hogar (i7 ENE/e39 EME) ⁴¹	1.053	745

Antes de comenzar el análisis propiamente tal, se recordará la estructura de éste: existe una primera medición (momento cero, con la EME), que cuenta solo con informantes directo/as. Luego, está la segunda medición (momento uno, con la ENE), con informantes directos/as y *proxy*. Así, hubo dos grupos encuestados: el control, correspondiente a quienes en las dos encuestas contestaron de manera directa, y el experimental, que constó de quienes en la EME contestaron de manera directa y en la ENE no participaron

⁴¹ Esta pregunta es sólo para quienes responden contar con algún tipo de contabilidad en la pregunta previa.

(obteniendo las respuestas a través de un tercero, informante *proxy*). A continuación, se analiza cada dimensión:

1. Ocupación

La pregunta por ocupación, basada en la Clasificación Internacional Uniforme de Ocupaciones (CIUO) a un dígito⁴², presenta diez categorías. En el caso de la muestra revisada, se presentan nueve categorías, quedando fuera la ocupación de “Fuerza pública”.

En cuanto a las mediciones en la EME (momento cero), las distribuciones fueron las siguientes:

Tabla 20: Categorías ocupacionales en M0 para grupo control y grupo experimental

Ocupación CIUO a 1 dígito		Grupo control	Grupo experimental
Miembros del poder ejecutivo y de los cuerpos legislativos y personal directivo		6,6	5,3
Profesionales científicos e intelectuales		5,3	5,4
Técnicos y profesionales de nivel medio		7,4	7,2
Empleados de oficina		0,5	0,3
Trabajadores de los servicios y vendedores de comercios y mercados		25,8	17,2
Agricultores y trabajadores calificados agropecuarios y pesqueros		16,7	17,5
Oficiales, operarios y artesanos de artes mecánicas y de otros oficios		22,3	27,5
Operadores de instalaciones y máquinas y montadores		5,7	10,0
Trabajadores no calificados		9,8	9,7
Total	%	100	100
	n	2.582	2.010

⁴² Es la desagregación disponible de manera pública en ambas encuestas, ya que el entregar por más dígitos (y, por ende, más detalle), pone en riesgo la mantención de anonimato de las personas participantes.

En el momento 1, las distribuciones fueron las siguientes:

Tabla 21: Categorías ocupacionales en M1 para grupo control y grupo experimental

Ocupación CIUO a 1 dígito		Grupo control	Grupo experimental
Miembros del poder ejecutivo y de los cuerpos legislativos y personal directivo		2,5	3,9
Profesionales científicos e intelectuales		4,4	4,4
Técnicos y profesionales de nivel medio		7,7	8,0
Empleados de oficina		0,8	0,8
Trabajadores de los servicios y vendedores de comercios y mercados		29,8	20,0
Agricultores y trabajadores calificados agropecuarios y pesqueros		14,9	14,1
Oficiales, operarios y artesanos de artes mecánicas y de otros oficios		22,0	26,1
Operadores de instalaciones y máquinas y montadores		6,9	10,7
Trabajadores no calificados		11,0	12,0
Total	%	100	100
	n	2.582	2.010

Respecto a los cambios entre el M0 y el M1, los porcentajes de preguntas que mantuvieron sus respuestas son distintos en el grupo control y el grupo experimental, aun cuando muy cercanas. De acuerdo a los estadísticos calculados (prueba t para evaluar las diferencias y r para medir tamaño del efecto), las diferencias para cada grupo entre las respuestas mantenidas en el M0 y el M1 versus el ideal (que todas fueran iguales en las dos mediciones) son significativas al 99,5%, en tanto el valor absoluto de t está sobre el valor crítico.

Tabla 22: Promedio de preguntas sobre CIUO con misma respuesta en los dos momentos, por tipo de grupo

	Grupo control	Grupo experimental
Media	68,9%	70,1%
Significancia (valor p)	<0,001	<0,001
t	-34,101	-29,273
Grados de libertad	2.581	2.009
r (tamaño del efecto)	0,56	0,55

Contrario a lo que se podría esperar, las respuestas se mantienen en mayor medida en el grupo experimental. En éste, las respuestas comunes son el 70,1%, mientras que en el grupo control son el 68,9%. Al ser diferencias estadísticamente significativas, se profundizará en cuáles son las categorías que presentaron mayores diferencias y semejanzas entre ambos momentos.

Las respuestas comunes por categoría en ambos momentos para el grupo control son:

Tabla 23: Distribución de respuestas sobre CIUO y respuestas comunes en ambos momentos para grupo control

Ocupación CIUO a 1 dígito	M0	M1	% de respuestas comunes
Miembros del poder ejecutivo y de los cuerpos legislativos y personal directivo (1)	6,6	2,5	7,0
Profesionales científicos e intelectuales (2)	5,3	4,4	63,2
Técnicos y profesionales de nivel medio (3)	7,4	7,7	53,9
Empleados de oficina (4)	0,5	0,8	0
Trabajadores de los servicios y vendedores de comercios y mercados (5)	25,8	29,8	76,2
Agricultores y trabajadores calificados agropecuarios y pesqueros (6)	16,7	14,9	79,4
Oficiales, operarios y artesanos de artes mecánicas y de otros oficios (7)	22,3	22,0	80,3
Operadores de instalaciones y máquinas y montadores (8)	5,7	6,9	88,5

Ocupación CIUO a 1 dígito		M0	M1	% de respuestas comunes
Trabajadores no calificados (9)		9,8	11,0	54,4
Total	%	100	100	68,9
	n	2.582	2.582	1.779

Lo primero que llama la atención es la categoría 4 (Empleados de oficina), pues ninguno de los casos que indicó en el M0 pertenecer a esta categoría, lo hizo nuevamente en el M1. En el caso de miembros del poder ejecutivo y de los cuerpos legislativos y personal directivo (categoría 1), la retención de respuestas también es baja: 7,0%.

Por el contrario, las categorías con mayor retención fueron la de operadores de instalaciones y máquinas y montadores (88,5%), oficiales, operarios y artesanos de artes mecánicas y de otros oficios (80,3%), agricultores y trabajadores calificados agropecuarios y pesqueros (79,4%) y trabajadores de los servicios y vendedores de comercios y mercados (76,2%).

En el caso del grupo experimental, los resultados son los siguientes:

Tabla 24: Distribución de respuestas sobre CIUO y respuestas comunes en ambos momentos para grupo experimental

Ocupación CIUO a 1 dígito	M0	M1	% de respuestas comunes
Miembros del poder ejecutivo y de los cuerpos legislativos y personal directivo (1)	5,3	3,9	17,8
Profesionales científicos e intelectuales (2)	5,4	4,4	60,6
Técnicos y profesionales de nivel medio (3)	7,2	8,0	49,3
Empleados de oficina (4)	0,3	0,8	16,7
Trabajadores de los servicios y vendedores de comercios y mercados (5)	17,2	20,0	73,9
Agricultores y trabajadores calificados agropecuarios y pesqueros (6)	17,5	14,1	72,6
Oficiales, operarios y artesanos de artes mecánicas y	27,5	26,1	82,3

Ocupación CIUO a 1 dígito		M0	M1	% de respuestas comunes
de otros oficios (7)				
Operadores de instalaciones y máquinas y montadores (8)		10,0	10,7	87,6
Trabajadores no calificados (9)		9,7	12,0	57,2
Total	%	100	100	70,1
	n	2.010	2.010	1.409

Respecto a las que tienen el menor porcentaje de preguntas comunes, son las mismas que para el grupo control, aun cuando la proporción de las coincidencias es mayor: empleados de oficina (16,7%) y miembros del poder ejecutivo y de los cuerpos legislativos y personal directivo (17,8%). Para ambos tipos de informante, además, son las categorías con menor frecuencia.

De manera similar, las dos categorías con más retención de preguntas son operadores de instalaciones y máquinas y montadores (87,6%) y oficiales, operarios y artesanos de artes mecánicas y de otros oficios (82,3%), seguidos de los trabajadores de los servicios y vendedores de comercios y mercados (73,9%) y agricultores y trabajadores calificados agropecuarios y pesqueros (72,6%).

2. Actividad económica

La clasificación de actividad económica, realizada mediante el Clasificador de Actividades Nacionales para Encuestas Sociodemográficas (CAENES), ofrece 21 categorías a 1 dígito, 20 de las cuales estuvieron presentes en la muestra.

La distribución de las respuestas en el momento cero fue la siguiente para ambos grupos:

Tabla 25: Actividades económicas en M0 para grupo control y grupo experimental

CAENES a 1 dígito	Grupo control	Grupo experimental
Agricultura, ganadería, silvicultura y pesca	17,2	17,5
Explotación de minas y canteras	0,3	0,5

CAENES a 1 dígito		Grupo control	Grupo experimental
Industrias manufactureras		14,4	11,5
Suministro de electricidad, gas, vapor y aire acondicionado		0,0	0,0
Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación		0,2	0,4
Construcción		7,6	13,4
Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas		32,0	26,4
Transporte y almacenamiento		6,3	10,9
Actividades de alojamiento y de servicio de comidas		5,3	4,0
Información y comunicaciones		0,6	0,6
Actividades financieras y de seguros		0,1	0,3
Actividades inmobiliarias		0,8	0,4
Actividades profesionales, científicas y técnicas		3,8	3,9
Actividades de servicios administrativos y de apoyo		1,0	1,7
Administración pública y defensa; planes de seguridad social de afiliación obligatoria		0,0	0,0
Enseñanza		1,1	1,2
Actividades de atención de la salud humana y de asistencia social		1,6	1,1
Actividades artísticas, de entretenimiento y recreativas		0,8	1,1
Otras actividades de servicios		6,9	4,8
Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio		0,0	0,0
Total	%	100	100
	N	2.582	2.010

En el momento 1, las distribuciones fueron las siguientes:

Tabla 26: Actividades económicas en M1 para grupo control y grupo experimental

CAENES a 1 dígito	Grupo control	Grupo experimental
Agricultura, ganadería, silvicultura y pesca	17,5	17,8
Explotación de minas y canteras	0,2	0,5
Industrias manufactureras	13,0	10,6
Suministro de electricidad, gas, vapor y aire acondicionado	0,0	0,1
Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación	0,3	0,5
Construcción	7,3	13,4
Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	32,0	26,4
Transporte y almacenamiento	6,2	10,8
Actividades de alojamiento y de servicio de comidas	5,7	4,1
Información y comunicaciones	0,9	0,7
Actividades financieras y de seguros	0,1	0,2
Actividades inmobiliarias	0,7	0,5
Actividades profesionales, científicas y técnicas	3,9	3,5
Actividades de servicios administrativos y de apoyo	1,0	1,5
Administración pública y defensa; planes de seguridad social de afiliación obligatoria	0,2	0,4
Enseñanza	1,1	1,2
Actividades de atención de la salud humana y de asistencia social	1,9	1,1
Actividades artísticas, de entretenimiento y recreativas	0,9	1,0
Otras actividades de servicios	7,1	5,2
Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio	0,2	0,2

CAENES a 1 dígito		Grupo control	Grupo experimental
Total	%	100	100
	n	2.010	2.010

Las medias de respuestas iguales en ambas mediciones son de 85,8% para el grupo control y 83,2% para el grupo experimental. Así, a diferencia de la medición de la ocupación, en el grupo con el mismo informante se mantiene mayormente la respuesta. De todas maneras, la brecha entre ambos es pequeña (2,6 puntos porcentuales), y la diferencia entre momentos es significativa para los dos, de acuerdo a la prueba t al 99,5% (ubicándose sobre el valor crítico). En este caso, de acuerdo al r para medir tamaño del efecto, las diferencias son relevantes.

Tabla 27: Promedio de preguntas sobre CAENES con misma respuesta en los dos momentos, por tipo de grupo

	Grupo control	Grupo experimental
Media	85,8%	83,2%
Significancia (valor p)	<0,001	<0,001
t	-20,647	-20,153
Grados de libertad	2581	2009
r (tamaño del efecto)	0,38	0,41

En el grupo de control, las distribuciones y porcentaje de respuestas compartidas para cada categoría son los siguientes:

Tabla 28: Distribución de respuestas sobre CAENES y respuestas comunes en ambos momentos para grupo control

CAENES a 1 dígito	M0	M1	% de respuestas comunes
Agricultura, ganadería, silvicultura y pesca	17,2	17,5	93,2
Explotación de minas y canteras	0,3	0,2	50,0
Industrias manufactureras	14,4	13,0	70,4
Suministro de electricidad, gas, vapor y aire acondicionado	0,0	0,0	N/A

CAENES a 1 dígito	M0	M1	% de respuestas comunes	
Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación	0,2	0,3	83,3	
Construcción	7,6	7,3	84,2	
Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	32,0	32,0	91,9	
Transporte y almacenamiento	6,3	6,2	90,8	
Actividades de alojamiento y de servicio de comidas	5,3	5,7	85,3	
Información y comunicaciones	0,6	0,9	75,0	
Actividades financieras y de seguros	0,1	0,1	66,7	
Actividades inmobiliarias	0,8	0,7	85,7	
Actividades profesionales, científicas y técnicas	3,8	3,9	82,7	
Actividades de servicios administrativos y de apoyo	1,0	1,0	60,0	
Administración pública y defensa; planes de seguridad social de afiliación obligatoria	0,0	0,2	N/A	
Enseñanza	1,1	1,1	65,5	
Actividades de atención de la salud humana y de asistencia social	1,6	1,9	87,8	
Actividades artísticas, de entretenimiento y recreativas	0,8	0,9	52,4	
Otras actividades de servicios	6,9	7,1	84,4	
Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio	0,0	0,2	N/A	
Total	%	100	100	85,8
	n	2.582	2.582	2.215

* Las categorías con N/A (“no aplica”) solo presentaron casos en el M1.

Las categorías con menor retención de respuestas son la de explotación de minas y canteras (50,0%) y de actividades artísticas, de entretenimiento y recreativas (52,4%). De todas

maneras, los porcentajes de retención son considerablemente mayores que las categorías críticas del CIUO.

Por el contrario, las categorías con mayor retención de respuestas son las correspondientes a agricultura, ganadería, silvicultura y pesca (93,2%), comercio al por mayor y al por menor (91,9%) y transporte y almacenamiento (90,8%). En el grupo experimental, los números son los siguientes:

Tabla 29: Distribución de respuestas sobre CAENES y respuestas comunes en ambos momentos para grupo experimental

CAENES a 1 dígito	M0	M1	% de respuestas comunes
Agricultura, ganadería, silvicultura y pesca	17,5	17,8	92,0
Explotación de minas y canteras	0,5	0,5	72,7
Industrias manufactureras	11,5	10,6	66,2
Suministro de electricidad, gas, vapor y aire acondicionado	0,0	0,1	N/A
Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación	0,4	0,5	77,8
Construcción	13,4	13,4	83,0
Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	26,4	26,4	87,7
Transporte y almacenamiento	10,9	10,8	92,3
Actividades de alojamiento y de servicio de comidas	4,0	4,1	77,5
Información y comunicaciones	0,6	0,7	50,0
Actividades financieras y de seguros	0,3	0,2	83,3
Actividades inmobiliarias	0,4	0,5	77,8
Actividades profesionales, científicas y técnicas	3,9	3,5	73,4
Actividades de servicios administrativos y de apoyo	1,7	1,5	62,9
Administración pública y defensa; planes de seguridad social de afiliación obligatoria	0,0	0,4	N/A

CAENES a 1 dígito		M0	M1	% de respuestas comunes
Enseñanza		1,2	1,2	80,0
Actividades de atención de la salud humana y de asistencia social		1,1	1,1	81,8
Actividades artísticas, de entretenimiento y recreativas		1,1	1,0	52,2
Otras actividades de servicios		4,8	5,2	81,3
Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio		0,0	0,2	N/A
Total	%	100	100	83,2
	n	2.010	2.010	1.672

* Las categorías con N/A (“no aplica”) solo presentaron casos en el M1.

En este grupo las categorías críticas son las relacionadas a información y comunicaciones (50,0%), actividades artísticas, de entretenimiento y recreativas (52,2%). Esta última presenta diferencias en ambos grupos.

Por otro lado, las categorías con mayor retención de respuestas son las de transporte y almacenamiento (92,3%), agricultura, ganadería, silvicultura y pesca (92,0%) y de comercio (87,7%), las mismas que para el grupo control.

3. Registro en Servicio de Impuestos Internos (SII)

La comparación sobre el registro en el SII tiene una particularidad: en la ENE (M1) esta pregunta incluye no respuesta parcial, mientras que en la EME (M0) no se ofrecen como alternativas no saber o no responder. Esto insta a una primera lectura: el 4,7% de las personas informantes en la ENE indican no saber la respuesta, o bien no responden la pregunta.

En la siguiente tabla se presentan las distribuciones en el M0:

Tabla 30: Registro en SII en M0 para grupo control y grupo experimental

La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos o tiene iniciación de actividades?		Grupo control	Grupo experimental
Sí, como trabajador independiente (con boleta de honorarios)		8,4	8,2
Sí, como persona natural (el RUT de la empresa es igual al RUT del dueño)		33,2	32,1
Sí, como otro tipo de empresa		7,8	9,1
No		50,6	50,6
Total	%	100	100
	n	2.489	1.877

En el momento 1 –cuyas distribuciones se muestran en la siguiente tabla– llaman la atención las diferencias según tipo de informante en la categoría de no respuesta parcial “No sabe”. Como se podría hipotetizar, el grupo con mayor no respuesta parcial es el experimental, es decir, el que no fue respondido por el informante directo/a.

Tabla 31: Registro en SI en M1 para grupo control y grupo experimental

La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos o tiene iniciación de actividades?		Grupo control	Grupo experimental
Sí, como trabajador independiente (con boleta de honorarios)		7,5	7,2
Sí, como persona natural (el RUT de la empresa es igual al RUT del dueño)		36,9	34,6
Sí, como otro tipo de empresa		4,4	4,5
No		50,7	49,0
No sabe		0,2	4,6
No responde		0,3	0,1
Total	%	100	100
	n	2.489	1.877

Respecto a los cambios entre el momento cero y el momento uno, el porcentaje de preguntas que mantuvieron sus respuestas son distintas en el grupo control y el grupo experimental, existiendo una diferencia de 7,4 puntos porcentuales entre ambos.

Tabla 32: Promedio de preguntas con misma respuesta en los dos momentos

	Grupo control	Grupo experimental
Media	80,4%	73,0%
Significancia (valor p)	<0,001	<0,001
t	-24,633	-26,313
Grados de libertad	2.488	1.876
r (tamaño del efecto)	0,44	0,52

El grupo control mantuvo sus respuestas en el 80,4% de los casos y el experimental lo hizo en el 73,0%. A pesar de haber mayor consistencia en el primer grupo, ambos promedios presentan diferencias estadísticamente significativas al 99,5% con el parámetro ideal, en tanto los valores t son mayores al valor crítico. Además, estas diferencias son relevantes en ambos casos.

Al analizar la distribución de las respuestas y el porcentaje de consistencia en las respuestas para el grupo control (tabla 29), las categorías críticas serían “Sí, como trabajador independiente” (38,8%) y “Sí, como otro tipo de empresa” (41,5%).

Las categorías con mayor consistencia son “No” (92,9%) y “Sí, como persona natural” (81,0%). Además, el 0,5% del grupo esgrimió una no respuesta parcial aun cuando en el momento 0 había respondido una de las otras categorías.

Tabla 33: Distribución de respuestas sobre registro en SII y respuestas comunes en ambos momentos para grupo control

La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos o tiene iniciación de actividades?	M0	M1	% de respuestas comunes
Sí, como trabajador independiente (con boleta de honorarios)	8,4	7,5	38,8
Sí, como persona natural (el RUT de la empresa es	33,2	36,9	81,0

La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos o tiene iniciación de actividades?		M0	M1	% de respuestas comunes
igual al RUT del dueño)				
Sí, como otro tipo de empresa		7,8	4,4	41,5
No		50,6	50,7	92,9
No sabe			0,2	N/A
No responde			0,3	N/A
Total	%	100	100	80,4
	n	2.489	2.489	2.001

* Las categorías con N/A (“no aplica”) solo presentaron casos en el M1.

Por otro lado, en el grupo experimental, se repiten las categorías con más y menor retención de respuestas. Entre las con menos respuestas comunes están “Sí, como otro tipo de empresa” tiene el 31,6% y “Sí, como trabajador independiente” el 33,1%, mientras que “No” (86,0%) y “Sí, como persona natural” (74,6%) cuentan con mayor consistencia.

Destaca que en el grupo experimental la no respuesta parcial aumentó consistentemente, siendo 4,7%, mientras que en el grupo control fue 0,5%, lo que ya podría estar hablando de cómo el tiempo de informante incide en la no respuesta parcial⁴³. Además, cada categoría tiene comportamientos similares entre las dos mediciones: sin distinguir el tipo de informante, son las mismas categorías las que aumentan o disminuyen su frecuencia.

Tabla 34: Distribución de respuestas sobre registro en SII y respuestas comunes en ambos momentos para grupo experimental

La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos o tiene iniciación de actividades?	M0	M1	% de respuestas comunes
Sí, como trabajador independiente (con boleta de honorarios)	8,2	7,2	33,1
Sí, como persona natural (el RUT de la empresa es igual al RUT del dueño)	32,1	34,6	74,6

⁴³ Sobre esto se ahondará en el siguiente capítulo.

La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos o tiene iniciación de actividades?		M0	M1	% de respuestas comunes
Sí, como otro tipo de empresa		9,1	4,5	31,6
No		50,6	49,0	86,0
No sabe			4,6	N/A
No responde			0,1	N/A
Total	%	100	100	73,0
	n	1.877	1.877	1.370

* Las categorías con N/A (“no aplica”) solo presentaron casos en el M1.

Observando las distribuciones en ambos momentos, la categoría “Sí, como otro tipo de empresa” es la que presenta mayores diferencias: del M0 al M1 la diferencia es de 4,6 puntos porcentuales. Si bien la disminución es casi equivalente a la no respuesta parcial no existe un traspaso de respuestas entre dichas categorías, ya que quienes en el M1 respondieron “No sabe” o “No responde” en el M0 respondieron principalmente “Sí, como persona natural”. Solo el 12,5% de la no respuesta parcial en el M1 respondió “Sí, como otro tipo de empresa” en el M0.

Tabla 35: Distribución de no respuesta parcial de informantes proxy ENE en categorías de la EME

La empresa, negocio o actividad en la que trabaja, ¿está registrada en el Servicio de Impuestos Internos o tiene iniciación de actividades?		M0
		%
Sí, como trabajador independiente (con boleta de honorarios)		19,3
Sí, como persona natural (el RUT de la empresa es igual al RUT del dueño)		34,1
Sí, como otro tipo de empresa		12,5
No		34,1
Total	%	100
	n	88

4. Registros de contabilidad

Como se revisó en el marco metodológico, fue necesario recodificar la pregunta de contabilidad de la EME para que fuera comparable con la pregunta de la ENE, quedando 5 categorías: contabilidad completa, contabilidad simplificada, registros personales, ningún tipo de contabilidad, y no sabe/no responde.

Las distribuciones en el M0 (tabla 32), son las siguientes:

Tabla 36: Registros de contabilidad en M0 para grupo control y grupo experimental

La empresa, negocio o actividad por cuenta propia...		Grupo control	Grupo experimental
...acude a los servicios de un contador para llevar la contabilidad completa?		34,2	33,5
...se encuentra acogido al régimen de contabilidad simplificada?		4,9	4,6
...solo cuenta con registros personales de gastos e ingresos?		14,6	13,4
No cuenta con ningún tipo de contabilidad		46,4	48,4
No sabe / No responde		0,0	0,2
Total	%	100	100
	n	2.489	1.877

En el M1, nuevamente, se observa una disparidad en las opciones de no respuesta parcial: el grupo experimental (compuestos por informantes *proxy*) tuvo una mayor prevalencia en los “No sabe/No responde” que el grupo control.

Tabla 37: Registros de contabilidad en M1 para grupo control y grupo experimental

La empresa, negocio o actividad por cuenta propia...		Grupo control	Grupo experimental
...acude a los servicios de un contador para llevar la contabilidad completa?		34,2	33,8
...se encuentra acogido al régimen de contabilidad		6,7	6,3

La empresa, negocio o actividad por cuenta propia...		Grupo control	Grupo experimental
simplificada?			
...solo cuenta con registros personales de gastos e ingresos?		8,1	6,9
No cuenta con ningún tipo de contabilidad		50,7	49,8
No sabe / No responde		0,3	3,1
Total	%	100	100
	n	2.489	1.877

Entre ambos momentos, el grupo control mantuvo el 70,2% de las respuestas y el experimental el 67,5%. Al igual que en las variables analizadas anteriormente, ambos grupos muestran diferencias significativas entre el momento 0 y el momento 1 de acuerdo a los valores t y r.

Tabla 38: Promedio de preguntas con misma respuesta en los dos momentos

	Grupo control	Grupo experimental
Media	70,1%	67,4%
Significancia (valor p)	<0,001	<0,001
t	-32,538	-30,090
Grados de libertad	2.488	1.876
r (tamaño del efecto)	0,55	0,57

En cuanto a la consistencia de las respuestas, se identifican fácilmente dos categorías tanto en el grupo control (tabla 35) como en el experimental (tabla 36): las relacionadas a la contabilidad completa y el no contar con ningún tipo de contabilidad son las que mantienen la mayor parte de las respuestas entre mediciones (respectivamente, 78,4% y 85,4% en grupo control y 75,5% y 81,6% en grupo experimental), y las respuestas sobre contabilidad simplificada y registros personales tienen una retención menor (respectivamente, 21,3% y 19,0% en grupo control y 14,9% y 15,1% en grupo experimental).

Tabla 39: Distribución de respuestas sobre contabilidad y respuestas comunes en ambos momentos para grupo control

La empresa, negocio o actividad por cuenta propia...		M0	M1	Respuestas comunes
...acude a los servicios de un contador para llevar la contabilidad completa?		34,2	34,2	78,4
...se encuentra acogido al régimen de contabilidad simplificada?		4,9	6,7	21,3
...solo cuenta con registros personales de gastos e ingresos?		14,6	8,1	19,0
No cuenta con ningún tipo de contabilidad		46,4	50,7	85,4
No sabe / No responde		0,0	0,3	N/A
Total	%	100	100	70,1
	n	2.489	2.489	1.745

* Las categorías con N/A (“no aplica”) solo presentaron casos en el M1.

Nuevamente destaca que al incorporar informante *proxy* (grupo experimental en M1) aumenta de manera importante la proporción de la no respuesta parcial, pasando del 0,2% al 3,1%.

Tabla 40: Distribución de respuestas sobre contabilidad y respuestas comunes en ambos momentos para grupo experimental

La empresa, negocio o actividad por cuenta propia...		M0	M1	Respuestas comunes
...acude a los servicios de un contador para llevar la contabilidad completa?		33,5	33,8	75,5
...se encuentra acogido al régimen de contabilidad simplificada?		4,6	6,3	14,9
...solo cuenta con registros personales de gastos e ingresos?		13,4	6,9	15,1
No cuenta con ningún tipo de contabilidad		48,4	49,8	81,6
No sabe / No responde		0,2	3,1	0
Total	%	100	100	67,4

La empresa, negocio o actividad por cuenta propia...	M0	M1	Respuestas comunes
n	1.877	1.877	1.265

Al igual que en la variable anterior, las categorías presentan comportamientos similares en ambos grupos en cuanto al aumento y disminución entre el M0 y el M1: exceptuando la no respuesta parcial en el grupo experimental, tienden a cambiar en cantidades similares sin importar el tipo de informante en el M1.

5. Separación de gastos del negocio y del hogar

Finalmente, se analizó la pregunta destinada a conocer si la contabilidad permite separar los gastos del negocio y del hogar. Esta variable es la que tiene el menor número de casos, debido a que solo se aplica a quienes responden que llevan algún tipo de contabilidad; como se observó en el apartado anterior, el porcentaje de personas que no cuentan con ningún tipo de contabilidad van entre el 46,4% y 50,7% en las dos mediciones de ambos grupos.

En el M0 las frecuencias son relativamente similares para ambos grupos:

Tabla 41: Separación de gastos en M0 para grupo control y grupo experimental

A través de la contabilidad, ¿se pueden separar los gastos del negocio de los gastos del hogar?		Grupo control	Grupo experimental
Sí		67,2	68,1
No		32,6	31,7
No sabe / No responde		0,2	0,3
Total	%	100	100
	N	1.053	745

En el M1 se observa cómo la no respuesta parcial es mayor en el grupo experimental (tabla 42), vislumbrando nuevamente que ésta aumenta al encuestar a informantes *proxy*.

Tabla 42: Separación de gastos en M1 para grupo control y grupo experimental

A través de la contabilidad, ¿se pueden separar los gastos del negocio de los gastos del hogar?		Grupo control	Grupo experimental
Sí		76,2	77,7
No		23,1	18,3
No sabe / No responde		0,8	4,0
Total	%	100	100
	n	1.053	745

Tal como ha sido la tónica de todas las variables revisadas, ambos grupos presentan diferencias estadísticamente significativas entre mediciones de acuerdo a los valores t y r. El grupo control retiene el 62,8% de las respuestas y el experimental el 61,7%.

Tabla 43: Promedio de preguntas con misma respuesta en los dos momentos

	Grupo control	Grupo experimental
Media	62,8%	61,7%
Significancia (valor p)	<0,001	<0,001
t	-24,978	-21,470
Grados de libertad	1.052	744
r (tamaño del efecto)	0,61	0,62

En ambos grupos el porcentaje de respuestas comunes en la categoría “Sí”, siendo de 79,0% en el control y 79,9% en el experimental, llamando la atención que en éste (con informante *proxy* en la M1) sea levemente mayor. Por su parte, la categoría “No” retiene un 29,7% y 23,3% en ambos grupos, respectivamente.

Es notorio que la no respuesta parcial nunca es consistente entre mediciones: quienes indicaron “No sabe” o “No responde” en el M0, no lo vuelven a hacer en el M1. En el caso de los/las informantes *proxy*, quienes indican esas alternativas lo hacen respecto a quienes sí respondieron “Sí” o “No” anteriormente.

Tabla 44: Distribución de respuestas sobre separación de gastos y respuestas comunes en ambos momentos para grupo control

A través de la contabilidad, ¿se pueden separar los gastos del negocio de los gastos del hogar?		M0	M1	Respuestas comunes
Sí		67,2	76,2	79,0
No		32,6	23,1	29,7
No sabe / No responde		0,2	0,8	0,0
Total	%	100	100	62,8
	n	1.053	1.053	661

Una vez más, existe una diferencia respecto a la preponderancia de la no respuesta parcial, comparado con el grupo control. Así, el aumento sería proporcionalmente mayor cuando en el M1 se entrevista a un informante *proxy*.

Tabla 45: Distribución de respuestas sobre separación de gastos y respuestas comunes en ambos momentos para grupo experimental

A través de la contabilidad, ¿se pueden separar los gastos del negocio de los gastos del hogar?		M0	M1	Respuestas comunes
Sí		68,1	77,7	79,9
No		31,7	18,3	23,3
No sabe / No responde		0,3	4,0	0,0
Total	%	100	100	61,7
	n	745	745	460

En vista de los resultados presentados en este capítulo se llega a dos hallazgos: por un lado, en el caso de las preguntas analizadas de la EME y la ENE, existen diferencias estadísticamente significativas siempre, independiente del tipo de informante y del tamaño de la muestra. Esto implica que existen otros elementos, ajenos a quién responde (informante directo/a o *proxy*), que inciden en cuál es la respuesta que finalmente se da. De todas maneras, se debe tener en consideración que en todas las preguntas (exceptuando la de ocupación), las diferencias entre mediciones fueron mayores en el grupo con cambio de informante (experimental) que en el grupo control.

Al analizar todas las variaciones entre ambos momentos de mediciones, en general todas tuvieron un comportamiento común: las categorías que aumentaban o disminuían su frecuencia eran las mismas, y en proporciones similares, tanto en el grupo control como el grupo experimental. Así, no se observaron cambios drásticos en un grupo sobre otro.

Respecto a las particularidades de la EME y la ENE, algunos factores ajenos al tipo de informante que pudieron haber incidido son: los distintos equipos de levantamiento; los diferentes cuestionarios que, si bien abordan temáticas similares, tienen distintos enfoques; el proceso de codificación a partir de glosas abiertas (en el caso de las dos primeras variables); e incluso el contexto de aplicación, ya que la ENE se aplica siempre en hogares mientras que la EME se acomoda a la disponibilidad del encuestado/a, acudiendo, por ejemplo, a su lugar de trabajo. Además, podría incidir también el cambio real que pudo haber en la situación de los trabajadores/as.

Por otro lado, y de cara al tercer capítulo, es importante identificar los hallazgos en términos de la no respuesta parcial. Ésta fue mayor siempre en los/las informantes *proxy* en el M1, presentando diferencias estadísticamente significativas respecto a la persona informante directa.

No respuesta parcial CASEN

Para el análisis de la no respuesta parcial se utilizará solo la encuesta CASEN. En el apartado anterior ya se revisaron, de manera exploratoria, casi todas las preguntas que en la ENE incluyen dichas categorías de respuesta.

En función de esto, se identificaron todas las preguntas de la CASEN que incorporaran opciones de no respuesta parcial, y se seleccionaron las que presentaban una mayor frecuencia en términos absolutos. Esta opción, si bien podría maximizar la posibilidad de diferencias al presentar mayores frecuencias de NS/NR (“No sabe/No responde”), se eligió con el fin de no sufrir una falta de casos, ya que parte de este análisis contempla separar los resultados por el número de integrantes del hogar. Realizar estas aperturas, como se mencionó en el marco metodológico, implicará excluir los hogares unipersonales y aquellos con la misma cantidad de informantes que de integrantes, además de filtrar para los distintos tamaños de hogar. Todos estos factores sin duda disminuirán los valores totales de la muestra, por lo que la elección de las preguntas es solo un resguardo ante un escenario complejo como lo es terminar con una muestra pequeña para cada subgrupo de análisis.

Las variables con mayor no respuesta parcial fueron las siguientes:

Tabla 46: Variables con mayor frecuencia de no respuesta parcial en CASEN 2015

Dimensión	Pregunta	% de NS/NR	Tipo
Ocupación	o23. ¿Cuántas personas trabajan en total en ese negocio o empresa o institución en Chile? (por bandas de número de personas)	7,7	Categórica
Ocupación	o30. En los últimos 12 meses, ¿ha participado o participa actualmente en alguna capacitación laboral de a lo menos 8 horas de duración?	3,1	Categórica
Ingresos	y1. El mes pasado, ¿cuál fue su sueldo o salario líquido en su trabajo principal?	5,7	Numérica
Ingresos	y4a. En los últimos 12 meses, además de los ingresos recién declarados, ¿recibió alguno de los siguientes tipos de ingreso provenientes de su	5,6	Numérica

Dimensión	Pregunta	% de NS/NR	Tipo
	ocupación principal? Bonificaciones (monto)		
Ocupación	o28. ¿Se encuentra afiliado a algún sistema previsional (sistema de pensiones)?	1,8	Categórica
Ingresos	y4b. En los últimos 12 meses, además de los ingresos recién declarados, ¿recibió alguno de los siguientes tipos de ingreso provenientes de su ocupación principal? Gratificaciones (monto)	3,5	Numérica
Ingresos	y2h. El mes pasado, ¿a cuántos días y horas pactadas con su empleador correspondió ese sueldo o salario? Horas	3,3	Numérica
Ingresos	y2d. El mes pasado, ¿a cuántos días y horas pactadas con su empleador correspondió ese sueldo o salario? Días	2,9	Numérica
Ingresos	y3a. Además del ingreso que recién declaró, ¿recibió el mes pasado algunos de los siguientes tipos de ingresos provenientes de su ocupación principal? Horas extras (monto)	2,8	Numérica
Educación	e14. En el año escolar 2015, ¿cuánto cuesta mensualmente la carrera que estudia?	12,7	Numérica

Destaca que, en línea con lo revisado en el marco conceptual, las preguntas con mayor no respuesta parcial son principalmente numéricas, y refieren a salario, otros ingresos, días y horas de trabajo, y costo de la educación.

Para facilitar la lectura de los datos, el análisis de las variables se hará por dimensión. Además, todas las frecuencias y porcentajes revisados corresponderán exclusivamente a la no respuesta parcial, excluyendo las otras categorías de respuesta. Esto, pues algunas de las variables a revisar (sobre todo las de ingresos) contienen más de cien categorías, lo que entorpece la búsqueda del dato relevante para este análisis (la no respuesta parcial).

- **Ocupación**

Como indica la siguiente tabla, todas las preguntas presentan diferencias estadísticamente significativas en la no respuesta parcial en función de si la respuesta la entregó directamente la persona (directo/a) o si la respondió un *proxy*. Sin embargo, el *r* para medir tamaño del efecto es pequeño para las tres variables (sobre todo en o30 y o28). Así, los datos se deben analizar considerando esa salvedad.

Tabla 47: Frecuencia y porcentaje de no respuesta parcial de preguntas de Ocupación según tipo de informante

Preguntas	Directo/a		Proxy		r (tamaño del efecto)
	n	%	n	%	
o23. ¿Cuántas personas trabajan en total en ese negocio o empresa o institución en Chile?	2.478	4,5*	6.012	10,8*	0,127
o30. En los últimos 12 meses, ¿ha participado o participa actualmente en alguna capacitación laboral de a lo menos 8 horas de duración?	2.675	2,5*	3.955	3,7*	0,036
o28. ¿Se encuentra afiliado a algún sistema previsional (sistema de pensiones)?	1.259	1,2*	2.560	2,4*	0,049

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes. Los valores de todas las pruebas realizadas en este capítulo se encuentran en el Anexo.

En hogares con dos y más integrantes, la no respuesta parcial para o23 se presenta en la tabla 48:

Tabla 48: Frecuencia y porcentaje de no respuesta parcial de o23 por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
o23.	n	295	650	598	1.383	550	1.637	530	2.342
¿Cuántas personas trabajan en total en ese negocio o empresa o institución en Chile?	%	3,9*	9,4*	4,6*	10,7*	4,5*	10,7*	4,7*	11,6*
	Brecha	5,5		6,1		6,2		6,9	
	r (tamaño del efecto)	0,123		0,122		0,120		0,129	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

Entre informantes directos/as el rango de no respuesta parcial varía del 3,9% y el 4,7% entre los hogares de menor y mayor tamaño, siendo la diferencia de 0,8 puntos porcentuales. Si bien los hogares con dos personas tienen el menor porcentaje y los de cinco personas o más el mayor, el comportamiento no es lineal, ya que los hogares de cuatro personas presentan una no respuesta parcial levemente mayor (0,1 pp.) que los de tres.

Entre informantes *proxy* tiende a haber una mayor tendencia a la linealidad en tanto la no respuesta parcial aumenta de manera paulatina a medida que crece el hogar, yendo del 9,4% al 11,6%. Similar a los directos/as, existe un estancamiento entre los hogares de 3 a 4 personas, ambos teniendo 10,7%. En este caso, la diferencia de la no respuesta parcial entre los hogares de menos y más integrantes es mayor que la de informantes directos/as, alcanzando los 2,2 puntos porcentuales.

De acuerdo al tipo de hogar, la brecha entre ambos tipos de informantes crece a medida que aumentan los/las integrantes. Destaca además que en todos los hogares, independiente del número de integrantes, existen diferencias estadísticamente significativas respecto a la no

respuesta parcial para directos/as y *proxy*, siendo ésta el doble o más en el último grupo respecto al primero. Nuevamente, de acuerdo a los valores de *r* el tipo de informante explica una parte limitada de la varianza.

En el caso de o30, si bien siempre las diferencias entre directos/as y *proxy* son estadísticamente significativas, éstas se producen por el tamaño de la muestra puesto que el *r* obtenido es muy pequeño, dando cuenta de una relación muy débil.

Tabla 49: Frecuencia y porcentaje de no respuesta parcial de o30 por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
o30. En los últimos 12 meses, ¿ha participado o participa actualmente en alguna capacitación laboral de a lo menos 8 horas de duración?	n	407	527	495	881	452	1.091	526	1.456
	%	2,6*	3,8*	2,1*	3,7*	2,0*	3,8*	2,3*	3,7*
	Brecha	1,2		1,6		1,8		1,4	
	<i>r</i> (tamaño del efecto)	0,036		0,034		0,048		0,054	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

En primer lugar, se observa que la brecha en puntos porcentuales entre ambos grupos es mucho menor que en o23: va solo de 1,2 a 1,8 puntos porcentuales.

Llama la atención que para informantes directos/as el mayor porcentaje de no respuesta parcial es entre los hogares de dos integrantes (2,6%), versus incluso los de cinco personas o más (2,3%). Sin embargo, se debe tener en consideración que la variación de los NS/NR

es menor entre los distintos tipos de hogar, en tanto la diferencia de la menor y mayor tasa de no respuesta parcial es de 0,6 puntos porcentuales.

Entre informantes *proxy* el comportamiento de la no respuesta parcial es mucho más estable, siendo de 3,7% para hogares de tres y cinco o más integrantes, y 3,8% en hogares de dos o cuatro personas. En este sentido, la diferencia máxima es de 0,1 puntos porcentuales.

Finalmente, la pregunta o28 también presenta diferencias estadísticamente significativas entre ambos grupos para todos los tamaños de hogar, aunque la fuerza de esta relación es débil. De todas maneras, es relevante conocer que este tipo de preguntas presenta mayores diferencias para los grupos familiares más numerosos.

Tabla 50: Frecuencia y porcentaje de no respuesta parcial de o28 por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
o28. ¿Se encuentra afiliado a algún sistema previsional (sistema de pensiones)?	n	160	270	235	532	265	733	279	1.025
	%	1,0*	2,0*	1,0*	2,2*	1,2*	2,6*	1,2*	2,6*
	Brecha	1,0		1,2		1,4		1,4	
	r (tamaño del efecto)	0,040		0,051		0,052		0,051	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

En esta pregunta el comportamiento más estable lo presentan los informantes directos/as: el porcentaje de no respuesta parcial es 1,0% para hogares de dos y tres personas, y 1,2% para los de cuatro personas y más.

Para los *proxy* la no respuesta parcial aumenta paulatinamente, partiendo en 2,0% y terminando en 2,6% para los hogares de cuatro personas y más. Así, en ambas categorías se vería un breve aumento de los NS/NR a medida que crece el tamaño del hogar.

Esta pregunta es la que presenta las menores brechas entre tipos de informantes, yendo de 1,0 a 1,4 puntos porcentuales.

- **Educación**

Otra pregunta con una alta frecuencia fue la referente a los costos de la carrera que estudia la persona. Al igual que las variables revisadas anteriormente, existen diferencias estadísticamente significativas entre los NS/NR de las respuestas registradas por la persona directamente y las que fueron entregadas por informante *proxy*. De acuerdo a la escala presentada en el marco metodológico, una *r* de 0,22 tiene un efecto medio, en tanto el tipo de informante explica el 4,84% de la varianza de la variable (0,22 al cuadrado).

En el primer caso el porcentaje de no respuesta parcial es de 4,0%, mientras que en el segundo se cuadruplica y asciende a 17,6%.

Tabla 51: Frecuencia y porcentaje de no respuesta parcial de e14 según tipo de informante

Preguntas	Directo/a		Proxy		r (tamaño del efecto)
	n	%	n	%	
e14. En el año escolar 2015, ¿cuánto cuesta mensualmente la carrera que estudia?	218	4,0*	1.719	17,6*	0,220

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

Al analizar si el tamaño del hogar influye en la tasa de no respuesta parcial se observa que las diferencias son significativas, independiente del tamaño del hogar, teniendo el tipo de informante una influencia media en los hogares de 2 a 4 personas, y alta en los de 5 personas o más (de acuerdo a los valores de *r*). Esto significa que la brecha entre informantes directo/as y *proxy* crece a medida que el tamaño del hogar aumenta (tabla 52).

Tabla 52: Frecuencia y porcentaje de no respuesta parcial de e14 por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
e14. En el año	n	22	118	46	364	50	489	68	748
escolar 2015,	%	4,1*	16,1*	3,8*	16,2*	3,5*	16,0*	4,6*	20,0*
¿cuánto	Brecha	12,0		12,4		12,5		15,4	
cuesta	r (tamaño	0,200		0,207		0,210		0,255	
mensualmente	del								
la carrera que	efecto)								
estudia?									

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

Entre informantes directos/as el comportamiento no es lineal, es decir, la no respuesta parcial no aumenta a medida que el hogar crece. El porcentaje menor de NS/NR es en los hogares de cuatro personas (3,5%), seguidos de los de tres (3,8%), dos (4,1%) y 5 o más (4,6%). La diferencia máxima entre ellos es de 1,6 puntos porcentuales.

Entre los/las *proxy* la situación es distinta. En los hogares de hasta cuatro personas se observa un porcentaje de no respuesta parcial relativamente estable yendo del 16,0% al 16,2%, pero en los hogares de 5 integrantes o más aumenta drásticamente a 20,0%. De hecho, en ese grupo la brecha entre tipos de informantes alcanza su mayor valor: 15,4 puntos porcentuales.

- **Ingresos**

Las preguntas pertenecientes a este módulo también presentan diferencias estadísticamente significativas en la no respuesta parcial entre tipos de informantes, aunque ésta debe ser abordada con prudencia de acuerdo a los valores de r: en el caso de y2h, y2d el efecto de ser informante es bajo (valores r de 0,032 y 0,039, explicando el 0,10 y 0,15 de la varianza), mientras que en el resto es medio, explicando del 1,9% al 5,7% de la varianza de acuerdo a la participación o no de la persona en la encuesta.

Respecto a las proporciones y frecuencias, las mayores diferencias se presentan en y1, y4a, y4b e y3a: en esos casos los NS/NR se multiplican por 6, 10, 4 y 10, respectivamente, para las respuestas entregadas por un *proxy*.

Tabla 53: Frecuencia y porcentaje de no respuesta parcial de preguntas de Ingresos por tipo de informante

Preguntas	Directo/a		Proxy		r (tamaño del efecto)
	n	%	n	%	
y1. El mes pasado, ¿cuál fue su sueldo o salario líquido en su trabajo principal?	593	1,5*	4.167	9,4*	0,193
y4a. En los últimos 12 meses, además de los ingresos recién declarados, ¿recibió alguno de los siguientes tipos de ingreso provenientes de su ocupación principal? Bonificaciones (monto)	377	0,9*	4.299	9,7*	0,238
y4b. En los últimos 12 meses, además de los ingresos recién declarados, ¿recibió alguno de los siguientes tipos de ingreso provenientes de su ocupación principal? Gratificaciones (monto)	462	1,2*	2.435	5,5*	0,138
y2h. El mes pasado, ¿a cuántos días y horas pactadas con su empleador correspondió ese sueldo o salario? Horas	1.048	2,6*	1.751	4,0*	0,032
y2d. El mes pasado, ¿a cuántos días y horas pactadas con su empleador correspondió ese sueldo o salario? Días	848	2,1*	1.609	3,6*	0,039
y3a. Además del ingreso que recién declaró, ¿recibió el mes pasado algunos de los siguientes tipos de ingresos provenientes de su ocupación principal? Horas extras (monto)	205	0,5*	2.142	4,8*	0,162

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

En el caso de y1 –referente al sueldo de la actividad principal–, la no respuesta parcial en informantes *proxy* es en promedio seis veces más grande que en los directos/as, y en todos

los tipos de hogares las diferencias entre ambos grupos son estadísticamente significativas (tabla 54). Es importante considerar que el efecto de ser informante directo/a o *proxy* es medio, de acuerdo a los valores de r.

Tabla 54: Frecuencia y porcentaje de no respuesta parcial de y1 por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
y1. El mes pasado, ¿cuál fue su sueldo o salario líquido en su trabajo principal?	n	87	527	156	950	130	1.131	132	1.559
	%	1,7*	10,3*	1,6*	9,3*	1,4*	9,1*	1,6*	9,5*
	Brecha	8,6		7,7		7,7		7,9	
	r (tamaño del efecto)	0,225		0,175		0,186		0,184	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

Tanto en informantes directos/as como *proxy* se alcanza el mayor porcentaje de no respuesta parcial en los hogares de dos personas: 1,7% y 10,3%, respectivamente. En ese mismo tipo de hogar se alcanza la mayor brecha entre ambos: 8,6 puntos porcentuales.

Para los directos/as no existen variaciones importantes, incluso cuando varíe el número de integrantes del hogar. El porcentaje de no respuesta parcial varía entre 1,4% y 1,7%, siendo la diferencia de solo 0,3 puntos porcentuales.

Entre los *proxy*, la no respuesta parcial se mantiene estable excepto en los hogares de dos personas, como se revisó anteriormente; en los de tres y más personas varía entre 9,1% y 9,3%.

En la pregunta y4a, referida al monto de dinero recibido por concepto de bonificaciones, sucede algo muy similar que en y1: en todas las categorías de hogares hay diferencias estadísticamente significativas, con valores de r que dan cuenta de que el tipo de informante tiene un efecto medio, por ende se vuelve más consistente analizar según el tamaño del

hogar. Así, en los hogares de dos integrantes se observa la mayor brecha para la no respuesta parcial entre tipos de informantes. Además, este mismo tipo de hogar alcanza su valor más alto para los/las *proxy*: 10,6%, en contraste con los otros en donde se mantiene relativamente estable (entre 9,5% y 9,6%), no existiendo diferencias en torno al número de integrantes.

Tabla 55: Frecuencia y porcentaje de no respuesta parcial de y4a por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
y4a. En los últimos 12 meses, además de los ingresos recién declarados, ¿recibió alguno de los siguientes tipos de ingreso provenientes de su ocupación principal? Bonificaciones	n	49	542	81	974	88	1.202	79	1.581
	%	0,9*	10,6*	0,9*	9,5*	1,0*	9,6*	1,0*	9,6*
	Brecha	9,7		8,6		8,6		8,6	
	r (tamaño del efecto)	0,258		0,253		0,230		0,223	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

En informantes directos/as el porcentaje de la no respuesta parcial es muy similar y no muestra diferencias importantes a medida que varía el tamaño del hogar: todos los valores están entre 0,9% y 1,0%.

En y4b, dirigida a conocer los montos recibidos por gratificaciones, la no respuesta parcial presenta diferencias estadísticamente significativas de manera transversal (considerando que el efecto del tipo de informante es medio, de acuerdo al tamaño de r). En los hogares de

dos y cuatro personas, los NS/NR se quintuplican en las respuestas entregadas por un *proxy* respecto a las personas que informaron directamente.

Tabla 56: Frecuencia y porcentaje de no respuesta parcial de y4b por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
y4b. En los últimos 12 meses, además de los ingresos recién declarados, ¿recibió alguno de los siguientes tipos de ingreso provenientes de su ocupación principal? Gratificaciones	n	50	270	119	558	78	682	119	925
	%	1,0*	5,3*	1,3*	5,4*	0,9*	5,5*	1,5*	5,6*
	Brecha	4,3		4,1		4,6		4,1	
	r (tamaño o del efecto)	0,152		0,137		0,152		0,123	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

Entre informantes directos/as, si bien la mayor no respuesta parcial tiene lugar en los hogares de cinco personas o más (1,5%), no se presenta un comportamiento lineal que indique que aumentará al sumar integrantes al hogar. Además, la diferencia entre el tipo de hogar con menos y más NS/NR es de 0,6 puntos porcentuales.

Entre informantes *proxy*, la no respuesta parcial aumenta sutilmente a medida que el hogar suma un integrante: es de 5,3% en hogares de dos personas y 5,6% en los de cinco personas o más, siendo la diferencia de 0,3 puntos porcentuales.

La pregunta y2h consulta por el número de horas pactadas con el empleador. Al igual que los casos anteriores, las diferencias son estadísticamente significativas para todos los tamaños de hogares, siendo siempre la no respuesta parcial mayor para los/las *proxy*. En este caso sí se debe tener precaución con la información y observarla con prudencia, en tanto los *r* dan cuenta de que la relación es muy débil, por lo que haber encontrado diferencias significativas se debe más al tamaño de la muestra que a la diferencia entre tipos de informantes.

Dicha esa aclaración, las diferencias entre informantes directos/as y *proxy* son menores que en las otras preguntas, siendo la mayor brecha de 1,5 puntos porcentuales (hogares de cinco personas o más).

Tabla 57: Frecuencia y porcentaje de no respuesta parcial de y2h por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
y2h. El mes pasado, ¿a cuántos días y horas pactadas con su empleador correspondió ese sueldo o salario? Horas	n	124	188	265	415	256	467	223	681
	%	2,4*	3,7*	2,8*	4,0*	2,8*	3,7*	2,7*	4,2*
	Brecha	1,3		1,2		0,9		1,5	
	r (tamaño del efecto)	0,039		0,032		0,024		0,039	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

Esta pregunta tampoco presenta un comportamiento lineal respecto al aumento del tamaño del hogar, tanto en informantes directos/as como en *proxy*. En general, el comportamiento es estable: para el primer grupo la diferencia máxima de no respuesta parcial entre los distintos hogares es de 0,4 puntos porcentuales, mientras que para el segundo son 0,5.

La pregunta y2d tiene un comportamiento similar a y2h⁴⁴, en tanto no es lineal y, si bien todas las diferencias son estadísticamente significativas, las brechas por tipo de informante, en todos los tamaños de hogares, no alcanzan valores altos: la máxima diferencia es de 1,4 puntos porcentuales. Además, nuevamente, el r indicaría que esas diferencias son pequeñas.

Por otro lado, tal como en y2h los valores se mantienen relativamente estables. Entre informantes directos/as la no respuesta parcial varía entre 2,1% y 2,5%, con una diferencia de solo 0,4 puntos porcentuales para los distintos tipos de hogares. Para informantes *proxy*, la fluctuación va entre 3,4% y 3,8%, existiendo una diferencia de 0,4 puntos.

Así, si bien las diferencias entre directos/as y *proxy* estarán presentes, el tamaño del hogar no estaría ocupando un rol determinante en esto.

Tabla 58: Frecuencia y porcentaje de no respuesta parcial de y2d por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
y2d. El mes pasado, ¿a cuántos días y horas pactadas con su empleador correspondió ese sueldo o salario? Días	n	108	173	209	370	191	436	200	630
	%	2,1*	3,4*	2,2*	3,6*	2,1*	3,5*	2,5*	3,8*
	Brecha	1,3		1,4		1,4		1,3	
	r (tamaño del efecto)	0,045		0,040		0,039		0,035	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

Finalmente, la pregunta y3a –referida a los ingresos por concepto de horas extras– también presenta diferencias estadísticamente significativas entre informantes directos/as y *proxy*,

⁴⁴ Se debe tener en cuenta que y2d con y2h, en el cuestionario, es solo una pregunta que contiene dos campos.

independiente del tamaño del hogar. Los r indican que el efecto del tipo de informante es medio, explicando entre el 2,4% y el 2,7% de la varianza de la variable.

En el caso de y3a, las diferencias son acentuadas en tanto la no respuesta parcial es de ocho a nueve veces mayor en informantes *proxy* que en directos/as.

Tabla 59: Frecuencia y porcentaje de no respuesta parcial de y3a por tipo de informante, según número de integrantes del hogar

Preguntas		2 personas		3 personas		4 personas		5 personas o más	
		Dir.	Proxy	Dir.	Proxy	Dir.	Proxy	Dir.	Proxy
y3a. Además del ingreso que recién declaró, ¿recibió el mes pasado algunos de los siguientes tipos de ingresos provenientes de su ocupación principal? Horas extras (monto)	n	26	237	53	493	49	610	45	802
	%	0,5*	4,6*	0,6*	4,8*	0,5*	4,9*	0,6*	4,9*
	Brecha	4,1		4,2		4,4		4,3	
	r (tamaño del efecto)	0,164		0,165		0,164		0,157	

*Con asterisco, las diferencias estadísticamente significativas a partir de prueba t de comparación de medias para muestras independientes.

Además, el comportamiento de la no respuesta parcial es estable en ambos grupos: entre informantes directos/as varía entre 0,5% y 0,6%, mientras que en los *proxy* se mueve entre 4,6% y 4,9%. Así, la brecha entre ellos no varía en grandes términos al cambiar el tamaño del hogar.

A partir de esta revisión, existen dos grandes conclusiones: por un lado, la no respuesta parcial es mayor en informantes *proxy*, aunque esta diferencia en algunos casos es débil y

en otras es moderada. Así, las respuestas entregadas por ese tipo de informante tendrían una mayor preponderancia de “No sabe” y “No responde” que las obtenidas mediante el directo/a.

Por otro lado, a partir de un análisis descriptivo, el tamaño del hogar no es un factor que incida en el valor de la no respuesta parcial, aunque en la pregunta de educación (e14) y de número de trabajadores (o23), para los hogares de cinco o más personas, implicó un aumento en el porcentaje de los NS/NR (2,2 y 3,9 puntos porcentuales, respectivamente, respecto a los hogares de dos personas). En el resto el aumento o fue menor o no existió, pudiendo incluso disminuir la no respuesta parcial en los hogares con más integrantes.

Conclusiones

Este trabajo se realizó para responder la siguiente pregunta: ¿cómo se ven afectados los resultados de la Encuesta Nacional del Empleo y la no respuesta parcial de la Encuesta de Caracterización Socioeconómica Nacional 2015, por el uso de informantes *proxy*? Tras los análisis, es posible responderla y dar cuenta de los espacios en los que se puede seguir profundizando en pos de mejorar la calidad de las encuestas mencionadas en sus futuras aplicaciones.

En primer lugar, se descubrió que las personas que son informantes en la ENE 2010-2017 y la CASEN 2015 tienen un perfil definido: mujeres entre 45 y 59 años, inactivas debido a que se dedican al trabajo doméstico. Este perfil tiene coherencia no solo interna, en tanto las y los informantes comparten características similares entre ambas encuestas, fenómeno que tiene eco al caracterizar de manera diferenciada a *proxy* y directos/as.

Como se revisó en la teoría (Bardasi, Beegle, Dillon, & Serneels, 2011), este perfil claro puede relacionarse con el llamado sesgo de auto-selección. Este grupo de personas que está aceptando ser informante tiene características particulares, que pueden incidir en las respuestas. Por ejemplo, la ENE es una encuesta eminentemente destinada a caracterizar a la población en términos de su situación en el mercado laboral, caracterizando en profundidad el trabajo de la persona trabajadora a través de consultas de horas de trabajo, tipo de contrato, situación de subcontratación, detalles del trabajo realizado, disponibilidad para trabajar más horas, pago de previsión, etcétera. Es decir, son preguntas específicas del trabajo remunerado que requieren un grado de familiarización con el mercado laboral. Este tipo de ítems también se abordan en un módulo completo de la encuesta CASEN.

Esto se relaciona además con los procesos cognitivos de respuesta, propuestos por Tourangeau (1984): comprensión, recuperación, juicio y reporte. Si desde la primera etapa ya existen deficiencias, es probable que la respuesta no sea de calidad, en tanto ésta sería considerada correcta por el/la informante aun cuando en la realidad no lo sea, o porque decide responder una de las alternativas para no dar la impresión de no ser un/una informante idóneo/a. Esto último, de acuerdo a la definición de Kaminska y Foulsham (2013), sería una expresión del sesgo de deseabilidad social.

Por otro lado, al analizar la prevalencia de las respuestas en dos momentos (comparación EME-ENE), se observó cómo los datos cambian independiente de que la respuesta se haya obtenido mediante informante *proxy* o directo/a: las diferencias siempre fueron estadísticamente significativas para ambos grupos y de fuerza relevante (lo que se demostró a partir del cálculo de r). Así, el que las respuestas cambien respondería a otros factores externos al tipo de informante.

Sin duda, esto complejiza el pensar los procesos de levantamiento que se alejen de lo espurio, ya que existen múltiples factores que incidirían en la calidad de los datos. En el análisis realizado, como se mencionó en ese capítulo, pudieron influir elementos concretos aparte del tipo de informante (cambio de cuestionario, diferente equipo de levantamiento, período de aplicación, etcétera) que no fueron posibles de aislar, pero que sí hablan de la complejidad de diseñar cuestionario y levantar encuestas, y mantener cierta armonía. Esto, sobre todo, en el contexto de encuestas de largo alcance (como lo son la EME, ENE y CASEN).

No es menor el rol que podrían estar teniendo esos factores, sobre todo tras lo revisado en el marco conceptual. Sesgos se crean a partir de la interacción encuestador/a-informante, de las expectativas de este último respecto al rol que debe cumplir, así como de la capacidad de entender una pregunta y dar una respuesta adecuada (lo más cercana a la realidad posible) por parte de la persona que está respondiendo la encuesta.

Finalmente, se analizó la no respuesta parcial. En el ejercicio de comparación EME-ENE ya se visualizaron los primeros hallazgos: las personas tienen a responder más “no sabe” y “no responde” cuando les consultan por otro sujeto/a. Esto se confirmó al analizar los resultados de la encuesta CASEN, y se condice con lo revisado en la bibliografía en los aportes puntuales de Moore (1988), Biggs (1992), y Martin y Butcher (1982). Además, se descubrió que el número de integrantes del hogar no implica necesariamente una mayor no respuesta parcial, excepto en preguntas particulares. Si bien algunos de estos datos explican la significancia de sus resultados por el tamaño de la muestra, sin duda se constituyen como un primer acercamiento necesario y que ya visibiliza un aspecto relevante de profundizar en futuras investigaciones.

Por otro lado, la tesis permitió corroborar que las preguntas que tienen mayor tendencia a los “no sabe” y “no responde” son las numéricas. De hecho, la variable categórica con mayor no respuesta parcial (o23), es una adaptación de una pregunta netamente numérica (número de personas que trabajan en el establecimiento). Las preguntas con mayor no respuesta parcial fueron las referentes a montos de dinero (preguntas y1, y4a, y4b y e14: en total, el 40% de las preguntas con mayor no respuesta parcial en la CASEN).

El aumento de la no respuesta parcial en *proxy* no necesariamente es un defecto ya que, como plantearon Beatty y Hermann (2002), al no haber seguridad de una precisión cognitiva por parte de la persona informante, es deseable que haya presencia de no respuesta parcial (en lugar de una respuesta “incorrecta”).

Esto abre una problemática altamente relevante para la ENE, la que en la mayoría de sus preguntas no incluye no respuesta parcial. A partir de los hallazgos, de las características de sus informantes y de la problematización por la calidad del dato, es importante que una encuesta de esa relevancia incluya esas opciones de respuesta. Por los protocolos de la ENE (revisar nota al pie página 54), una vez aplicado, el formulario debe ser entregado completo sin preguntas en blanco, sino se devuelve al encuestador/a o se elimina ese hogar de la muestra. Esto, en un contexto de un trabajo de campo complejo y constante, puede inducir a que existan prácticas no deseables, como forzar una respuesta en la persona informante o incluso a que se falsifiquen. Sin duda, incluir transversalmente la no respuesta parcial aumentaría la calidad del dato y transparentaría una situación que no es anómala, sino que se espera dentro de un proceso de levantamiento de encuestas.

La implementación de los “No sabe” y/o “No responde” supone establecer criterios. Por ejemplo, qué preguntas se deben recuperar por su relevancia –aumentando el número de visitas a la vivienda– o en cuáles se puede mantener la respuesta. Crear estos criterios es trabajo que debe ser asumido sobre todo por los equipos operativos, pero a la luz de los resultados, de la bibliografía y de las otras encuestas de empleo a nivel latinoamericano (en las que sí se incluyen), sería un enorme paso para aumentar los estándares metodológicos de la ENE.

Así, es posible responder la pregunta de investigación: sí, el /la informante *proxy* influye en las respuestas, pero su efecto no es el mismo para todas las dimensiones analíticas de una encuesta.

Tiene un impacto en el aumento de la no respuesta parcial en la CASEN (principalmente, pero no exclusivo, en las variables numéricas) y, por las características de la ENE (que no incluye NS/NR), en su calidad del dato. Por otro lado, propicia la presencia de sesgos relacionados a la auto-selección y, en parte, a la deseabilidad social.

Dicho esto, ¿cuáles son los caminos para seguir generando conocimiento sobre este tema? En primer lugar, y a la luz de los hallazgos, se sugiere la realización de un ejercicio con grupos de control y experimentales. Lo realizado con la EME y la ENE fue un análisis exploratorio, y resulta interesante realizarlo en mejores condiciones, disminuyendo la presencia de factores ajenos al tipo de informante. Así, existirían dos tipos de estudio: uno similar al analizado, con mediciones en dos momentos, manteniendo la persona que encuesta, el cuestionario, y disminuyendo la brecha de tiempo entre las dos aplicaciones. El otro, podría dirigirse a visitar viviendas con hogares de más de un integrante y aplicar a cada uno/a el cuestionario por sí mismo/a y por las otras personas.

Para ambas posibilidades, tanto los equipos de la ENE como la CASEN cuentan con el marco muestral, en tanto ya manejan un listado de hogares con su número de integrantes. En el caso de la ENE, para no sobrecargar a las personas informantes –sobre todo considerando que es una encuesta panel–, se podría trabajar con las viviendas que hayan salido recientemente de la muestra.

Para llevar a cabo estos estudios, habría que realizar algunas modificaciones a los cuestionarios. En el caso de CASEN, es imperante cambiar la pregunta que identifica a la persona informante al final de cada módulo, en tanto las respuestas no se condicen con el enunciado⁴⁵. Se sugiere, en función del formato del cuestionario, que consulte por el tipo de participación de la persona sobre la cual se está preguntando, manteniendo las alternativas de respuesta o incluso dejándola dicotómica: contesta al menos una pregunta/no contesta ninguna pregunta.

⁴⁵ La pregunta es: “¿Quién (o quiénes) responde(n) el módulo?”, y las alternativas de respuesta son: “Contesta al menos una pregunta”, “Presente, pero no contesta” y “No está presente”.

En el caso de la ENE es necesario realizar cambios más profundos. Como se mencionó anteriormente, se debe incluir la no respuesta parcial. Por otro lado, otra limitante es que solo identifica a una persona como informante, no dando espacio a caracterizar la participación de otros integrantes del hogar. En esta línea, se sugiere incluir una pregunta similar a la de la CASEN (con los cambios recién señalados).

Si bien es una inversión de tiempo, dinero y trabajo, ambas encuestas tienen la oportunidad de tener un rol vanguardista a nivel latinoamericano, en tanto no existen estudios sobre el tema en el área de las ciencias sociales.

Desde la relevancia de esta tesis, es importante realizar estos ejercicios sociológicos de revisión de metodologías y de además crear conocimientos sobre ellas. Esto se constituye como fundamental del “oficio”, desde una perspectiva crítica y experta que dé espacios de mejorar los instrumentos con los que actualmente se está conociendo la realidad, sobre todo en temas importantes y sensibles como trabajo, educación y salud. Existen muchos dogmas, pero no necesariamente son aplicables a todas las encuestas, de todas las temáticas, en todos los lugares. Las especificidades tanto nacionales como regionales influirán en cómo se responden, y es necesario atenderlas.

Además, en ninguna de las dos encuestas existe (al menos de libre disposición) documentación fundada para elegir este tipo de informantes. Sin duda es un supuesto dado para los equipos técnicos y operativos, pero es importante tener el debate, yendo más allá de la conveniencia o incluso de la tradición heredada de manera irreflexiva en términos metodológicos.

Finalmente, para concluir esta tesis: este ejercicio no estuvo destinado a sugerir que es incorrecta la utilización de informantes *proxy*. Sin duda, su rol es clave en tanto minimiza el trabajo y los costos, sobre todo en encuestas a cargo del Estado, con muestras grandes y a nivel nacional, como la ENE y la CASEN.

Tras el ejercicio realizado, se observa que, si bien existe un impacto a partir del uso de informantes *proxy*, éste es uno más de los distintos factores que causan distorsión en las encuestas. Probablemente, el paso de informante *proxy* a directo/a eliminaría ese impacto,

pero crearía otros desconocidos para ambas encuestas, como lo son los asociados a informantes que responden por sí mismos/as.

En ese sentido, el uso de informantes *proxy* permite mantener, como lo planteaban Bardasi y otros (2011), un equilibrio de costo-eficiencia: economiza los tiempos de trabajo de campo y también los gastos, aun cuando tiende a aumentar la no respuesta parcial. El ocupar informante directo/a es aumentar tanto el tiempo de aplicación como los gastos, pero sin tener evidencia concluyente de los riesgos que tendría su uso en la calidad de las respuestas.

Así, es ingenuo pretender que la ENE y la CASEN se apliquen a informantes directos/as, e incluso puede cuestionarse la idoneidad de éstos por el tipo de temáticas abordadas en ambas encuestas. Pero sí, sobre todo desde la sociología, se consideró elemental dar cuenta de las limitaciones de las metodologías basadas en informantes *proxy*. El tener este conocimiento permitirá planificar mejor los trabajos de levantamiento, empezar a pensar en procedimientos de no respuesta parcial y analizarlas teniendo en cuenta cuáles son las restricciones de sus instrumentos.

Anexo

Pruebas t para igualdad de medias y r para medir tamaño del efecto

Para todas las variables revisadas, las diferencias de medias fueron significativas en tanto superaron el valor crítico (2,58). La manera de verificar esa significación fue a partir de r para medir tamaño del efecto.

1. Total de hogares

Preguntas	Prueba T para igualdad de medias			r (tamaño del efecto)
	t	Grados de libertad	Valor p	
o23	39,884	96.725,645	<0,001	0,127
o30	16,399	204.259,114	<0,001	0,036
o28	21,557	189.460,773	<0,001	0,049
e14	25,841	13.100,838	<0,001	0,220
y1	49,637	63838,820	<0,001	0,193
y4a	57,958	56183,726	<0,001	0,238
y4b	35,632	65574,328	<0,001	0,138
y2h	8,847	74760,192	<0,001	0,032
y2d	10,641	75614,918	<0,001	0,039
y3a	39,173	56943,304	<0,001	0,162

2. Hogares de 2 personas

Preguntas	Prueba T para igualdad de medias			r (tamaño del efecto)
	t	Grados de libertad	Valor p	
o23	13,395	11774,116	<0,001	0,123
o30	5,620	27070,997	<0,001	0,034
o28	6,378	25115,502	<0,001	0,040
e14	6,984	1181,196	<0,001	0,200
y1	18,742	6573,031	<0,001	0,225
y4a	20,727	6027,292	<0,001	0,258
y4b	12,610	6702,644	<0,001	0,152
y2h	3,845	9553,041	<0,001	0,039
y2d	4,328	9352,540	<0,001	0,045
y3a	13,026	6104,356	<0,001	0,164

3. Hogares de 3 personas

Preguntas	Prueba T para igualdad de medias			r (tamaño del efecto)
	t	Grados de libertad	Valor p	
o23	18,469	22713,968	<0,001	0,122
o30	10,076	44758,787	<0,001	0,048
o28	10,484	42203,854	<0,001	0,051
e14	12,336	3399,324	<0,001	0,207
y1	24,156	14151,141	<0,001	0,175
y4a	28,800	12107,120	<0,001	0,253
y4b	16,851	14895,153	<0,001	0,137
y2h	4,414	19544,049	<0,001	0,032
y2d	5,587	19232,131	<0,001	0,040
y3a	18,843	12758,497	<0,001	0,165

4. Hogares de 4 personas

Preguntas	Prueba T para igualdad de medias			r (tamaño del efecto)
	t	Grados de libertad	Valor p	
o23	19,804	26876,192	<0,001	0,120
o30	12,112	50549,242	<0,001	0,054
o28	11,608	49944,199	<0,001	0,052
e14	13,511	3944,602	<0,001	0,210
y1	25,490	18188,399	<0,001	0,186
y4a	30,118	16201,628	<0,001	0,230
y4b	20,376	17521,905	<0,001	0,152
y2h	3,427	20958,218	<0,001	0,024
y2d	5,611	21502,531	<0,001	0,039
y3a	21,085	16017,075	<0,001	0,164

5. Hogares de 5 personas o más

Preguntas	Prueba T para igualdad de medias			r (tamaño del efecto)
	t	Grados de libertad	Valor p	
o23	22,749	30533,802	<0,001	0,129
o30	9,756	55330,402	<0,001	0,041
o28	12,458	59614,548	<0,001	0,051
e14	16,164	3756,464	<0,001	0,255
y1	29,164	24147,381	<0,001	0,184
y4a	34,037	22189,155	<0,001	0,223
y4b	19,401	24538,824	<0,001	0,123
y2h	5,425	19291,400	<0,001	0,039
y2d	4,860	19105,752	<0,001	0,035
y3a	23,614	22126,313	<0,001	0,157

Bibliografía

- Bardasi, E., Beegle, K., Dillon, A., & Serneels, P. (2011). Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania. *The World Bank Economic Review*, 25(3), 418-447. doi:10.1093/wber/lhr022
- Beatty, P., & Herrmann, D. (2002). To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse. En R. Groves, D. Dillman, J. Eltinge, & R. Little, *Survey Nonresponse* (págs. 71-85). New York: Wiley.
- Belson, W. (1981). *The Design and Understanding of Survey Questions*. Londres, Inglaterra: Gower.
- Berlanga Silvente, V., Rubio Hurtado, M., & Vilà Baños, R. (enero de 2013). Cómo aplicar árboles de decisión en SPSS. *Revista d'Innovació i Reserca en Educació*, 6(1), 65-79.
- Biggs, B. (1992). *Self/Proxy Respondent Rules and Data Quality*. Obtenido de Statistics Canada: <http://www.statcan.gc.ca/pub/75f0002m/75f0002m1992008-eng.pdf>
- Bound, J., Brown, C., & Mathiowetz, N. (2000). *Measurement Error in Data Survey*. Michigan: Population Studies Center, University of Michigan.
- Bourdieu, P., Chamboredon, J.-C., & Passeron, J.-C. (2002). *El oficio de sociólogo. Presupuestos epistemológicos*. Argentina: Siglo XXI Editores Argentina S.A.
- Campanelli, P., Rothgeb, J., & Martin, E. (2005). *The Role of Respondent Comprehension and Interviewer Knowledge in CPS Labor Force Classification*. U.S. Bureau of the Census, Statistical Research Division, Washington D.C. Recuperado el 2018, de Census.gov: <https://www.census.gov/srd/papers/pdf/rsm2005-07.pdf>
- Cea D'Ancona, M. Á. (2004). *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid: Síntesis.

- Cea D'Ancona, M. Á. (2005). La senda tortuosa de la "calidad" de la encuesta. *Revista Española de Investigaciones Sociológicas*(111), 75-103. Obtenido de http://reis.cis.es/REIS/PDF/REIS_111_051168262751380.pdf
- Coe, R. (2003). Magnitud del Efecto: Una guía para investigadores y usuarios. *Revista de Psicología de la PUCP*, XXI(1), 145-177.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Dawe, F., & Knight, I. (1997). A study of proxy response on the Labour Force Survey. *Survey Methodology Bulletin*(40), 30-36.
- Del Río L., J. (sin fecha). *Introducción a SPSS Árboles de Clasificación*.
- Durkheim, E. (2001). *Las reglas del método sociológico*. México: Fondo de Cultura Económica.
- Frías Navarro, M. (2008). *Tema 5. Validez de la investigación (II): Validez de conclusión estadística*. Valencia, España: Universitat de València.
- Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Groves, R., Fowler, Jr., F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (Second ed.). New Jersey: Wiley.
- Hill, D. (1987). *Response Errors in Labor Surveys: Comparisons of Self and Proxy*. Survey of Income and Program Participation. U.S. Department of Commerce. Bureau of the Census.
- Instituto Nacional de Estadísticas. (6 de abril de 2018). *ENE*. Obtenido de Inicio: <http://www.ine.cl/docs/default-source/laborales/ene/antecedentes-metodologicos/manual-conceptual-y-metodol%C3%B3gico-dise%C3%B1o-muestral.pdf?sfvrsn=6>
- Kaminska, O., & Foulsham, T. (March de 2013). Understanding Sources of Social Desirability Bias in Different Modes: Evidence from Eye-Tracking. *ISER Working*

- Paper* *Series*(2013-04). Obtenido de <https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2013-04.pdf>
- Ledesma, R., Macbeth, G., & Cortada de Kohan, N. (2008). Tamaño del efecto: revisión teórica y aplicaciones con el sistema estadístico ViSta. *Revista Latinoamericana de Psicología*, 40(3), 425-439.
- Magidson, J. (1993). *SPSS for Windows CHAID Release 6.0*. Chicago: SPSS Inc.
- Martin, J., & Butcher, B. (1982). The Quality of Proxy Information - Some Results from a Large-Scale Study. *Journal of the Royal Statistical Society*, 31(4), 293-319.
- Maturana, H., & Varela, F. (2003). *El árbol del conocimiento. Las bases biológicas del entendimiento humano*. Santiago: Lumen / Editorial Universitaria.
- Merriam-Webster. (23 de febrero de 2018). *Proxy | Definition of Proxy by Merriam-Webster*. Obtenido de Dictionary by Merriam-Webster: America's most trusted online dictionary: <https://www.merriam-webster.com/dictionary/proxy>
- Ministerio de Desarrollo Social. (6 de abril de 2018). *Observatorio Social - Ministerio de Desarrollo Social - Gobierno de Chile*. Obtenido de Observatorio Social - Ministerio de Desarrollo Social: http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen_obj.php
- Moore, J. (1988). Self/Proxy Response Status and Survey Response Quality. *Journal of Official Statistics*, 4(2), 155-172.
- Morales Jacob, J. (2004). *Aplicación e interpretación de técnicas de reducción de datos según escalamiento óptimo (Análisis de correspondencia múltiple y análisis de componentes principales categórico)*. Tesis para optar al Título Profesional de Sociólogo. Santiago: Facultad de Ciencias Sociales, Universidad de Chile.
- Naciones Unidas. (2007). *Encuestas de hogares en los países en desarrollo y en transición*. Departamento de Asuntos Económicos y Sociales. División de Estadística. New York: Naciones Unidas. Obtenido de https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Handbooks/surveys/seriesf_96-S.pdf

- Naciones Unidas. (2009). *Clasificación Industrial Internacional Uniforme de todas las actividades económicas. Revisión 4*. Nueva York: Naciones Unidas.
- Office for National Statistics - Labour Force Survey. (2016). *User Guide, Volume 1 - LFS Background and Methodology 2016*. Obtenido de Home - Office for National Statistics:
<https://www.ons.gov.uk/file?uri=/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance/volume12016.pdf>
- Oxford Dictionaries. (19 de febrero de 2018). *proxy | Definition of proxy in English by Oxford Dictionaries*. Obtenido de Oxford Dictionaries | The World's Most Trusted Dictionary Provided: <https://en.oxforddictionaries.com/definition/proxy>
- Pardo Merino, A., & Ruiz Díaz, M. (2005). *Análisis de datos con SPSS 13 Base*. Madrid: McGraw-Hill.
- Piazzini, C. (2014). Conocimientos situados y pensamientos fronterizos: una relectura desde la universidad. *Geopolítica(s)*, 5(1), 11-33.
- Real Academia Española. (19 de febrero de 2018). *DLE: idóneo*. Obtenido de Real Academia Española: <http://dle.rae.es/?id=KvAZPI8>
- The American Association For Public Opinion Research. (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (Novena ed.). AAPOR. Obtenido de AAPOR: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf
- Tourangeau, R. (1984). Cognitive Sciences and Survey Methods. En T. Jabine, M. Straf, J. Tanur, & R. Tourangeau, *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines* (págs. 73-100). Washington: National Academy Press.
- University of Colorado Colorado Springs. (1 de octubre de 2018). *Effect Size (SE) | Effect Size Calculators*. Obtenido de University of Colorado Colorado Springs:

<https://www.uccs.edu/lbecker/effect-size#2.%20Correlation%20measures%20of%20effect>

Vidal, J. (2013). La búsqueda de la realidad o de la verdad: una aproximación a partir de la teoría sociológica. *Cinta moebio*(47), 95-114.