

Documento de trabajo

Fundamentos del Estándar para la
evaluación de la calidad de las
estimaciones en encuestas de
hogares

Autores: Mesa de trabajo institucional para la
definición de lineamientos estandarizados respecto a
las medidas de precisión

N° 13, marzo de 2020



INSTITUTO NACIONAL DE ESTADÍSTICAS

Morandé 801, Santiago de Chile

Teléfono: 56 232461010

Correo: ine@ine.cl

Facebook: [@ChileINE](https://www.facebook.com/ChileINE)

Twitter: [@INE_Chile](https://twitter.com/INE_Chile)

Coordinación

Departamento de Metodologías e Innovación Estadística

Denisse López Arenas

Subdepartamento de Diseño de Marcos y Muestras

Jonathan Paredes Báez

Iván Touron Romero

Miguel Guerrero Herrera

Subdepartamento de Calidad y Estándares

Hans Kruger Rossier

Participación y colaboración

Javiera Torres Aguirre

Verónica Valdivia Medina

Agustín Arce Garcés

Alfredo Heufemann Peña

Loreto Díaz Mella

Joaquín Nilo Elgueta

Nicolás Ortiz Ruiz

Verónica Huaracán Riveros

Juan Ceccarelli Arias.

Los autores y colaboradores agradecen todas las oportunidades de mejora recibidas de parte de los equipos técnicos del INE, así como a los expertos: Carolina Casas Corderos (PhD. Jefa del programa de Magíster en Diseño y Análisis de Encuestas Sociales del Instituto de Sociología de la PUC), Isabelle Beaudry (Ph.D. Profesor Asistente, Departamento de Estadística, PUC) y Andrés Gutiérrez (Experto Regional en Estadísticas Sociales-CEPAL), cuyos comentarios y reorientaciones permitieron nutrir y enriquecer el presente trabajo.

Los Documentos de Trabajo del INE están dirigidos a investigadores, académicos, estudiantes y público en general y tienen como objetivo proporcionar un análisis exhaustivo sobre aspectos conceptuales, analíticos y metodológicos claves de los productos estadísticos que elabora la institución y, de esta forma, contribuir al intercambio de ideas entre los distintos componentes del Sistema Estadístico Nacional.

Las interpretaciones y opiniones que se expresan en los Documentos de Trabajo pertenecen en forma exclusiva a los autores y colaboradores y no reflejan necesariamente el punto de vista oficial del INE ni de la institución a la que pertenecen los colaboradores de los documentos.

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

Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares

Resumen

En este documento se presenta el trabajo metodológico realizado por el Instituto Nacional de Estadística (INE) en relación con la evaluación de las medidas de calidad de las estimaciones provenientes de encuestas a hogares por muestreo, con el propósito de proveer al Sistema Estadístico Nacional (SEN) de un conjunto de directrices que permitan orientar a la población usuaria respecto a la evaluación de la calidad, uso, análisis e interpretación de la información que produce el INE.

Bajo un marco de aseguramiento de la calidad, la importancia de contar con lineamientos para la evaluación de las estimaciones a través de criterios que consideren diferentes dimensiones (como el tamaño muestral, grados de libertad, coeficiente de variación, error estándar, entre otros) radica en que la fiabilidad estadística, entendida como el grado en que las mediciones obtenidas reflejan la realidad, permite definir cuándo una estimación cumple un requisito mínimo de calidad. En este sentido, el documento entrega un panorama respecto al marco conceptual de las medidas de calidad, su uso y criterios utilizados en las encuestas del INE, así como también en diversas Oficinas Nacionales de Estadística (ONE). De acuerdo con los antecedentes expuestos, se desarrolla una propuesta de lineamientos a través de flujogramas que orienta al conjunto de usuarios en la toma de decisiones al momento de analizar y publicar información.

Palabras clave: calidad estadística, encuestas de hogares, estimaciones, lineamientos, medidas de precisión.

Abstract

This document presents the methodological work carried out by the National Statistics Institute (INE, by its Spanish acronym) evaluating the quality measures of the estimates generated in household surveys by sampling. The purpose is to provide the National Statistical System (SEN, by its Spanish acronym) with a set of guidelines related to the evaluation of the quality, use, analysis and interpretation of the information produced by the INE.

Under a quality assurance framework, the importance of having such guidelines for the evaluation of estimates through criteria that consider different dimensions (such as sample size, degrees of freedom, coefficient of variation, error standard, among others) is that statistical reliability, understood as the degree to which measurements obtained reflect reality, allows us to define when an estimate meets a minimum quality requirement. In this regard, the document provides a conceptual overview of quality measures, their use and criteria adopted in surveys carried out by the INE and various National Statistics Offices (ONE, by its Spanish acronym). Based on the above background, a proposal for guidelines is developed through flowcharts that guides all users in decision-making when analyzing and publishing information.

Índice

1. Motivación	11
2. Marco conceptual	12
2.1. Estimadores	12
2.2. Medidas de precisión de los estimadores	13
2.2.1. Varianza	14
2.2.2. Error estándar	15
2.2.3. Coeficiente de variación	15
2.2.4. Error absoluto.....	16
2.2.5. Error relativo	16
2.2.6. Intervalo de confianza.....	17
2.3. Diseño muestral complejo	17
2.3.1. Medidas de precisión en muestras complejas	18
2.3.2. Efecto de diseño y tamaño de muestra efectivo	19
2.3.3. Grados de libertad	20
2.3.4. Fracción de muestreo y factor de corrección de poblaciones finitas	21
3. Antecedentes	22
3.1. Antecedentes internacionales. Uso de medidas de precisión en oficinas estadísticas	22
3.2. Antecedentes nacionales. Uso de medidas de precisión en el INE Chile	26
3.2.1. Descripción de las encuestas evaluadas.....	27
3.2.1.1. Encuesta Nacional de Empleo.....	27
3.2.1.2. Encuesta Suplementaria de Ingresos.....	28
3.2.1.3. Encuesta de Presupuestos Familiares	28
3.2.1.4. Encuesta Nacional sobre Uso del Tiempo	29
3.2.1.5. Encuesta Nacional de Seguridad Ciudadana	30
3.2.1.6. Encuesta de Microemprendimiento	31
3.2.1.7. Análisis de género en el INE Chile	32
3.2.2. Medidas de precisión actuales.....	33
4. Recomendaciones sobre medidas de calidad	37
4.1. Evaluación de criterios respecto a tamaño muestral y grados de libertad	38
4.1.1. Tamaño muestral	38
4.1.2. Grados de libertad.....	40
4.1.3. Fracción de muestreo.....	42
4.2. Evaluación de los criterios relacionados con la precisión	43

4.2.1.	Criterios de precisión para proporciones y razones entre 0 y 1	43
4.2.2.	Metodología para evaluar la calidad del error estándar	47
4.2.2.1.	Función lineal propuesta	47
4.2.2.2.	Función logarítmica propuesta	49
4.2.2.3.	Función cuadrática propuesta.....	51
4.2.3.	Criterios de precisión para el resto de las estimaciones	53
4.3.	Flujograma para la publicación de estadísticas.....	54
4.3.1.	Flujograma o criterios de calidad para evaluación de las estimaciones	54
4.3.2.	Flujograma o criterios de calidad para evaluación de tabulados	57
5.	Evaluación de impacto de aplicación del flujograma	57
5.1.	Pautas para aplicación del flujograma	60
5.2.	Análisis de impacto por estimaciones o celdas.....	63
5.2.1.	Análisis de criterios respecto a tamaño muestral y grados de libertad	63
5.2.2.	Análisis de criterios vinculados con medidas de precisión	64
5.3.	Análisis de impacto por tabulados	66
5.4.	Análisis de impacto por encuestas	67
5.4.1.	Evaluación de aplicación del flujograma en la ENE	68
5.4.2.	Evaluación de aplicación del flujograma en la ESI	69
5.4.3.	Evaluación de aplicación del flujograma en la EPF	71
5.4.4.	Evaluación de aplicación del flujograma en la ENUT	73
5.4.5.	Evaluación de aplicación del flujograma en la ENUSC	75
5.5.	Desagregaciones de encuestas con enfoque de género (SEEG)	77
5.5.1.	Género y Seguridad Ciudadana (SEEG-ENUSC).....	77
5.5.2.	Género e Ingresos (SEEG-ESI)	79
5.5.3.	Género y Uso del Tiempo (SEEG-ENUT).....	80
5.5.4.	Género y Microemprendimiento (SEEG-EME)	82
6.	Conclusión	84
7.	Referencias bibliográficas	86
8.	ANEXOS	89
8.1.	Anexo 1. Función cuadrática alternativa para el umbral del error estándar	89
8.2.	Anexo 2. Evaluación de las estimaciones por valor del cv según ONE	91
8.3.	Anexo 3. Indicadores evaluados por encuesta.....	92

Índice de tablas

Tabla 1. Percentil 0,975 de la distribución t-Student	20
Tabla 2. Personas microempendedoras por sexo según categoría ocupacional.....	61
Tabla 3. Porcentaje de celdas caídas para criterios relacionados con el tamaño muestral y grados de libertad respecto del total de celdas.....	64
Tabla 4. Porcentaje de celdas caídas por criterios de tamaño muestral y grados de libertad respecto del total de celdas (SEEG)	64
Tabla 5. Porcentaje de celdas caídas por criterios de medidas de dispersión respecto del total de celdas	65
Tabla 6. Porcentaje de celdas caídas por criterios de medidas de dispersión respecto del total de celdas (SEEG)	65
Tabla 7. Porcentaje de celdas caídas por evaluación del error estándar respecto del total de celdas.....	66
Tabla 8. Porcentaje de celdas caídas por evaluación del error estándar respecto del total de celdas (SEEG).....	66
Tabla 9. Tabulados caídos por encuesta según propuesta de función para el error estándar	67
Tabla 10. Tabulados caídos por encuesta de SEEG según función para error estándar	67
Tabla 11. Porcentaje de celdas caídas según criterio del flujograma, OND 2019	68
Tabla 12. Porcentaje de celdas caídas por indicador según propuesta, OND 2019	69
Tabla 13. Tabulados caídos en total y porcentaje, OND 2019.....	69
Tabla 14. Porcentaje de celdas caídas según criterio del flujograma, ESI 2018	70
Tabla 15. Porcentaje de celdas caídas por indicador, ESI 2018	70
Tabla 16. Tabulados caídos en total y porcentaje, ESI 2018	71
Tabla 17. Porcentaje de celdas caídas según criterio del flujograma, EPF 2017.....	72
Tabla 18. Porcentaje de celdas caídas por indicador, EPF 2017	72
Tabla 19. Tabulados caídos en total y porcentaje, EPF 2017.....	73
Tabla 20. Porcentaje de celdas caídas según criterio del flujograma, ENUT 2015	74
Tabla 21. Porcentaje de celdas caídas por indicador, ENUT 2015.....	74
Tabla 22. Tabulados caídos en total y porcentaje, ENUT 2015	75
Tabla 23. Porcentaje de celdas caídas según criterio del flujograma, ENUSC 2017	76
Tabla 24. Porcentaje de celdas caídas por indicador, ENUSC 2017.....	76
Tabla 25. Tabulados caídos en total y porcentaje, ENUSC 2017	77
Tabla 26. Porcentaje de celdas caídas según criterio del flujograma, SEEG-ENUSC 2017.	78

Tabla 27. Porcentaje de celdas caídas por indicador, SEEG-ENUSC 2017	78
Tabla 28. Tabulados caídos en total y porcentaje, SEEG-ENUSC 2017.....	79
Tabla 29. Porcentaje de celdas caídas según criterio del flujograma, SEEG-ESI 2018	79
Tabla 30. Porcentaje de celdas caídas por indicador, SEEG-ESI 2018	80
Tabla 31. Tabulados caídos en total y porcentaje, SEEG-ESI 2018.....	80
Tabla 32. Porcentaje de celdas caídas según criterio del flujograma, SEEG-ENUT 2018 ...	81
Tabla 33. Porcentaje de celdas caídas por indicador, SEEG-ENUT 2018.....	81
Tabla 34. Tabulados caídos en total y porcentaje, SEEG-ENUT 2018.....	82
Tabla 35. Porcentaje de celdas caídas según criterio del flujograma, SEEG-EME 2017.....	82
Tabla 36. Porcentaje de celdas caídas por indicador, SEEG-EME 2017	83
Tabla 37. Tabulados caídos en total y porcentaje, SEEG-EME 2017.....	83
Tabla 38. Indicadores evaluados en la ENE.....	92
Tabla 39. Indicadores evaluados en la ESI.....	93
Tabla 40. Indicadores evaluados en la VIII EPF.....	93
Tabla 41. Indicadores evaluados en la ENUT.....	94
Tabla 42. Indicadores evaluados en la ENUSC.....	95
Tabla 43. Indicadores evaluados en la SEEG-ENUSC.....	95
Tabla 44. Indicadores evaluados en la SEEG-ESI	96
Tabla 45. Indicadores evaluados en la SEEG-ENUT.....	99
Tabla 46. Indicadores evaluados en la SEEG-EME.....	99

Índice de gráficos

Gráfico 1. Percentil 0,95 de la distribución t-Student y variación porcentual	42
Gráfico 2. Máximo coeficiente de variación y error estándar admitido según estimación de P (primer criterio).....	44
Gráfico 3. Máximo coeficiente de variación y error estándar admitido según estimación de P (segundo escenario)	45
Gráfico 4. Comportamiento del error estándar y del coeficiente de variación según estimación de P para $n=60$, bajo muestreo aleatorio simple	46
Gráfico 5. Máximo error estándar, coeficiente de variación, tamaño muestral requerido e intervalo de confianza según función lineal (primera propuesta).....	48
Gráfico 6. Máximo error estándar, coeficiente de variación, tamaño muestral requerido e intervalo de confianza según cv logarítmico (segunda propuesta)	50

Gráfico 7. Máximo error estándar, coeficiente de variación, tamaño muestral requerido e intervalo de confianza según función cuadrática (tercera propuesta)	52
Gráfico 8. Porcentaje de tabulados a evaluar por encuesta	62
Gráfico 9. Máximo error estándar, coeficiente de variación, tamaño muestral requerido e intervalo de confianza según función cuadrática (cuarta propuesta)	90

Índice de cuadros

Cuadro 1. Estimaciones en características de la población seleccionada bajo MAS.....	13
Cuadro 2. Varianza poblacional y muestral según estimador, bajo MAS	15
Cuadro 3. Error estándar poblacional y muestral del estimador	15
Cuadro 4. Coeficiente de variación en la poblacional y muestral del estimador de θ	16
Cuadro 5. Error absoluto poblacional y muestral del estimador de θ	16
Cuadro 6. Error relativo poblacional y muestral del estimador de θ	17
Cuadro 7. Estimación de intervalo de confianza del estimador de θ	17
Cuadro 8. Medidas de precisión considerados actualmente por producto	33
Cuadro 9. Clasificación del coeficiente de variación y error relativo.....	34
Cuadro 10. Criterios actualmente considerados por producto	35
Cuadro 11. Clasificación de la estimación según cv y tamaño muestral	36
Cuadro 12. Niveles de cv aceptables.....	44
Cuadro 13. Umbrales de aceptación de los criterios de evaluación de las distintas propuestas.....	59
Cuadro 14. Categorización de las estimaciones por valor del cv según ONE.....	91

Índice de Ilustraciones

Ilustración 1. Flujograma para la evaluación de la calidad de las estimaciones	56
Ilustración 2. Flujograma para la evaluación de los tabulados.....	57

1. Motivación

El INE, en su misión de proveer al Sistema Estadístico Nacional de normas metodológicas para la producción estadística, ha elaborado el presente documento titulado “Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares”, con el propósito de poner a disposición del público usuario orientaciones para el análisis e interpretación de estadísticas oficiales.

A pesar de que las encuestas de hogares pueden generar un alto volumen de información, es posible afirmar que no siempre tienen un potencial estadístico aceptable o tolerable, si se evalúan las medidas de calidad. A raíz de esta situación, el documento de trabajo tiene como principal objetivo recomendar los tipos de medidas que se deben utilizar, cuándo emplearlas y cuáles son los rangos que se admiten para que una estimación se encuentre apta para uso, análisis y publicación.

En este sentido, el documento aborda en primer lugar el marco conceptual en que se circunscriben las medidas de calidad, entregando una definición de cada una de ellas; en segundo lugar, se presentan los antecedentes respecto al uso de medidas de calidad en encuestas de hogares por muestreo a nivel nacional e internacional. Luego, se presentan flujogramas de aplicación de los criterios a seguir sobre las medidas de calidad; y, por último, se obtiene y examina el impacto que tiene la aplicación de los lineamientos sobre una muestra intencionada de estimaciones provenientes de distintas encuestas.

Los lineamientos fueron elaborados bajo la metodología de “mesa de trabajo”, que se extendió desde diciembre de 2018 a febrero de 2020, y estuvo a cargo de unidades del INE pertenecientes al Departamento de Metodología e Innovación Estadística y a la Subdirección Técnica. Las propuestas metodológicas elaboradas por la mesa de trabajo fueron expuestas a un comité externo al INE, compuesto por académicos y expertos, quienes realizaron una serie de recomendaciones que permitieron nutrir y enriquecer el presente trabajo.

2. Marco conceptual

Para situar la discusión respecto a las medidas de calidad, es necesario referirse al rol que cumplen las encuestas de hogares como fuente de información. Dentro del sistema estadístico se reconocen tres principales fuentes de información de estadísticas sociales y demográficas que utilizan los países para la planificación y control de los programas de desarrollo socioeconómico: (i) encuestas por muestreo, (ii) censos de población, que por lo general se realizan en intervalos prolongados de aproximadamente 10 años, caracterizados por ser operaciones sumidas en grandes costos y (iii) registros administrativos, cuya fiabilidad depende del grado de exhaustividad y coherencia de sus definiciones o conceptos.

De las tres fuentes de información, las encuestas de hogares por muestreo destacan como fuente clave y fundamental, ya que figuran entre los métodos más flexibles de recopilación de datos, permitiendo atender de forma continua las necesidades de información. Otra de las ventajas que se identifican en las encuestas por muestreo, es que los costos asociados suelen ser inferiores al de un censo, debido a que se basan, por lo general, en muestras probabilísticas, lo que garantiza bajo el *principio de representatividad*, que los resultados puedan emplearse como base para extraer conclusiones certeras sobre la población total.

Bajo este principio, se espera que las muestras sean representativas de la población de donde se extrajo, lo que se logra cuando la distribución de las características de los elementos contenidos en la muestra es idéntica o similar a la distribución de la población. En este contexto, a través de la muestra se pueden conseguir estimaciones de diferentes parámetros de interés, cuyas propiedades dependerán del estimador y del diseño muestral adoptado.

A continuación, se definen algunos conceptos fundamentales en el desarrollo de este documento.

2.1. Estimadores

Los estimadores corresponden a estadísticos o estadígrafos (es decir, una función de los valores que toma una variable a través de la muestra) usados para estimar un parámetro desconocido de la población. Aun cuando las encuestas se centran en muchos propósitos, los estimadores comúnmente empleados para estimar características de la población son:

- Total, por ejemplo, total de personas en la fuerza de trabajo.
- Media, por ejemplo, ingreso promedio por hogar.

- Proporción, proporción de unidades que estén clasificadas dentro de alguna clase definida, por ejemplo, proporción de personas en situación de pobreza.
- Razón, definido como el cociente de dos totales o medias, por ejemplo, total de desocupados entre la fuerza de trabajo (tasa de desocupación).
- Percentiles, indican la posición de un valor respecto a los demás, por ejemplo, el ingreso mediano de las personas ocupadas.

Las expresiones para obtener las estimaciones dependen del tipo de muestreo¹ y, como primer acercamiento al lector se muestran en el *Cuadro 1* estas expresiones bajo muestreo aleatorio simple (MAS) sin reemplazo.

Cuadro 1. Estimaciones en características de la población seleccionada bajo MAS

Parámetro	Estimador
Población total (T_y)	$t = \hat{T}_y = N \cdot \hat{Y}$
Media de la población (\bar{Y})	$\bar{y} = \hat{\bar{Y}} = \frac{\sum_{i=1}^n y_i}{n} = \frac{\hat{T}_y}{n}$
Proporción de población para una categoría (P)	$p = \hat{P} = \frac{n_d}{n}$
Razón de la población (R)	$r = \hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{\hat{Y}}{\hat{X}}$

Fuente: Elaboración propia. INE 2020.

2.2. Medidas de precisión de los estimadores

Actualmente, los objetivos de las encuestas de hogares por muestreo van mucho más allá de proveer, desde el punto de vista descriptivo, una serie de tabulados con estimaciones puntuales de los parámetros de interés. Cada vez es más frecuente la incorporación de objetivos analíticos, donde se desea ajustar modelos y realizar contrastes de hipótesis que den cuenta de la fuerza y relación de las variables en la población a diferentes niveles. Por

¹ El tipo de muestreo se refiere al procedimiento mediante el cual se obtiene una muestra de la población; estos a su vez se dividen en probabilísticos y no probabilísticos. En los muestreos probabilísticos todas las unidades tienen probabilidad conocida de ser seleccionadas y la elección se hace aleatoriamente. Los tipos de muestreo probabilísticos más utilizados son: aleatorio simple, sistemático, estratificado y por conglomerados. Las formas para seleccionar unidades más frecuentemente empleadas son: muestreo aleatorio simple sin reemplazamiento, muestreo aleatorio simple con reemplazamiento, selección con probabilidad proporcional al tamaño con reemplazamiento y muestreo simple sistemático. Para una revisión más detallada consultar Cochran (1998).

ello se hace necesario que las publicaciones de los resultados incluyan medidas apropiadas de precisión o exactitud de las estimaciones derivadas de los datos de la encuesta.

Una de las medidas de precisión clave en las encuestas por muestreo es la varianza muestral, que es un indicador de la variabilidad que se introduce al elegir una muestra en vez de toda la población. Existen otras formas de medir el error de muestreo como: el error estándar, el coeficiente de variación, el efecto del diseño, entre otras. Estas medidas se caracterizan porque guardan entre sí relación algebraica, por lo que es posible deducir la expresión de cualquiera de ellas a partir de una dada.

A continuación, se presentan, las expresiones asociadas a las medidas de precisión más comúnmente utilizadas (bajo el MAS, sin reemplazo) para lo cual se adoptarán las siguientes notaciones:

- θ : Parámetro poblacional, definido como el valor verdadero de la medida de alguna característica de la población, y que, por lo general, suele ser estimada ya que no se conoce.
- $\hat{\theta}$: Estimador, definido como un estadístico o estadígrafo que estima a través de la muestra al parámetro poblacional.

2.2.1. Varianza

La varianza de un estimador puede definirse como el promedio de desviaciones cuadráticas del valor promedio de la estimación. No se usa muy a menudo en *estadísticas descriptivas* porque proporciona unidades de medida de la variable elevadas al cuadrado. Las expresiones para estimar la varianza poblacional y muestral de los principales estimadores se presentan en el *Cuadro 2*.

Cuadro 2. Varianza poblacional y muestral según estimador, bajo MAS

Parámetro	Varianza poblacional del estimador	Estimación de varianza
Media de la población (\bar{Y})	$V(\hat{Y}) = \left(\frac{N-n}{N}\right) \cdot \frac{\sigma^2}{n}, \sigma^2 = \sum_{i=1}^N \frac{(Y_i - \bar{Y})^2}{N}$	$v(\hat{Y}) = \left(\frac{N-n}{N}\right) \cdot \frac{s^2}{n}, s^2 = \sum_{i=1}^n \frac{(Y_i - \hat{Y})^2}{n-1}$
Total (T)	$V(\hat{T}) = N^2 \cdot V(\hat{Y})$	$v(\hat{T}) = N^2 \cdot v(\hat{Y})$
Proporción de población para una categoría (P)	$V(\hat{P}) = \left(\frac{N-n}{N-1}\right) \cdot \frac{P \cdot (1-P)}{n}$	$v(\hat{P}) = \left(\frac{N-n}{N}\right) \cdot \frac{\hat{P} \cdot (1-\hat{P})}{n-1}$

Fuente: Técnicas de Muestreo. Cochran (1998).

2.2.2. Error estándar

El error estándar corresponde a la raíz cuadrada de la varianza del estimador. Esta medida resulta más fácil de interpretar ya que utiliza la misma escala de medición de la estimación. Las expresiones para obtener el *ee* se presentan en *Cuadro 3*.

Cuadro 3. Error estándar poblacional y muestral del estimador

Error estándar poblacional del estimador	Estimación del error estándar
$EE(\hat{\theta}) = \sqrt{V(\hat{\theta})}$	$ee(\hat{\theta}) = \sqrt{v(\hat{\theta}_s)}$

Fuente: Elaboración propia. INE, 2020.

2.2.3. Coeficiente de variación

El coeficiente de variación se obtiene mediante el cociente entre el error estándar y el valor promedio de la propia estimación. Suele expresarse en porcentaje y resulta ser útil para comparar la precisión de estimaciones entre encuestas con tamaño muestral diferente, o en los casos en donde la escala de medición de las estimaciones a comparar es diferente, ya que no tiene unidad de medida (*Cuadro 4*).

Cuadro 4. Coeficiente de variación en la poblacional y muestral del estimador de θ

Coeficiente de variación poblacional del estimador	Estimación del coeficiente de variación
$CV(\hat{\theta})(\%) = \frac{EE(\hat{\theta})}{\theta} \cdot 100\%$	$cv(\hat{\theta})(\%) = \frac{ee(\hat{\theta})}{\hat{\theta}_s} \cdot 100\%$

Fuente: Elaboración propia. INE, 2020.

2.2.4. Error absoluto

El error absoluto corresponde a la máxima diferencia que se puede esperar entre el parámetro estimado y el verdadero parámetro poblacional con una probabilidad de $(1 - \alpha)$, siendo $(1 - \alpha)$ denotado como el nivel de confianza utilizado. Se expresa en la misma unidad de medida del estimador y se suele usar con un nivel de confianza de 90% o 95% (Cuadro 5) donde el valor $t_{1-\alpha/2}^v$ corresponde al percentil de la distribución *t-Student* con v grados de libertad (ver 2.3.3), que para muestras grandes converge con el mismo nivel de confianza al percentil de la distribución *Normal Estándar* denotado por $Z_{1-\alpha/2}$, y que para un 95% de confianza corresponde a 1,96. Esto asumiendo que el estimador $\hat{\theta}$ se distribuye *t-Student* (dado que la varianza poblacional es desconocida y estimada por la misma muestra) con media o esperanza igual al valor del parámetro θ y varianza igual a $V(\hat{\theta})$.

Cuadro 5. Error absoluto poblacional y muestral del estimador de θ

Error absoluto poblacional	Estimación del error absoluto
$EA(\hat{\theta}) = Z_{1-\alpha/2} \cdot EE(\hat{\theta})$	$ea(\hat{\theta}) = t_{1-\alpha/2}^v \cdot ee(\hat{\theta})$

Fuente: Elaboración propia. INE, 2020.

2.2.5. Error relativo

Corresponde a la máxima desviación porcentual que se puede esperar entre el parámetro estimado y el verdadero parámetro poblacional con probabilidad de $(1 - \alpha)$. Esta medida permite homogenizar las unidades de medida de varios indicadores, al convertirse en un porcentaje relativo al estimador (Cuadro 6).

Cuadro 6. Error relativo poblacional y muestral del estimador de θ

Error relativo poblacional del estimador	Estimación del error relativo
$ER(\hat{\theta}) = Z_{1-\alpha/2} \cdot CV(\hat{\theta})$	$er(\hat{\theta}) = t_{1-\alpha/2}^v \cdot cv(\hat{\theta})$

Fuente: Elaboración propia. INE, 2020.

2.2.6. Intervalo de confianza

El intervalo de confianza corresponde a un rango o recorrido de valores, derivado de las estadísticas de la muestra, que posiblemente contenga al verdadero parámetro poblacional de la variable de interés con una probabilidad de $(1 - \alpha)$, denominada nivel de confianza y que habitualmente se fija en 90% o 95% (Cuadro 7).

Cuadro 7. Estimación de intervalo de confianza del estimador de θ

Intervalo de confianza para θ	Estimación del Intervalo de confianza para θ
$[\hat{\theta} - EA(\hat{\theta}) ; \hat{\theta} + EA(\theta)]_{1-\alpha}$	$[\hat{\theta}_s - ea(\hat{\theta}_s) ; \hat{\theta}_s + ea(\hat{\theta}_s)]_{1-\alpha}$

Fuente: Elaboración propia. INE, 2020.

Un intervalo de confianza con nivel de confianza del 95% significa que, en promedio, de cada 100 muestras obtenidas bajo el mismo diseño, el intervalo contiene al verdadero valor del parámetro en 95 de ellas.

2.3. Diseño muestral complejo

Según la ONU (2009), prácticamente todos los diseños muestrales para las encuestas de hogares, tanto en los países en desarrollo como en los países desarrollados, son complejos, debido a sus características: varias etapas de selección, estratificación, conglomeración y probabilidades de selección desigual, entre otras².

Por su parte, Wolter (2007) utiliza la terminología “Diseño Muestral Complejo Moderno” haciendo referencia a cinco dimensiones de complejidad relacionadas con las encuestas por muestreo, a saber: (i) grado de complejidad del diseño muestral, (ii) grado de complejidad de los estimadores del estudio, (iii) múltiples características o variables de interés, (iv) uso

² Otras de las características asociadas a diseño muestrales complejos son múltiples marcos muestrales y muestreo bifásico. Se considera que un diseño muestral es complejo cuando contempla al menos dos de estas características mencionadas.

de estadísticos descriptivos e inferenciales de las encuestas, y (v) la escala o tamaño del estudio. De acuerdo con lo anterior, se puede señalar que la mayoría de las encuestas utilizadas para la generación de estadísticas oficiales contemplan diseños muestrales complejos, pudiendo ser analizadas desde las cinco dimensiones señaladas.

2.3.1. Medidas de precisión en muestras complejas

Hasta ahora se han presentado las medidas de precisión bajo el método MAS, sin embargo, tal como se mencionó anteriormente, los diseños muestrales de las encuestas de hogares corresponden a diseños complejos, por lo que se deben considerar criterios adicionales para el uso e interpretación de las medidas de precisión con el fin de evitar llegar a conclusiones o análisis incorrectos³.

Para resolver el tema del cálculo de errores muestrales tomando en cuenta el diseño muestral complejo, existen diversos paquetes estadísticos (SPSS, R, SAS, STATA, SUDAAN, entre otros) que permiten incorporar el esquema de muestreo al momento de realizar los análisis de datos en muestras complejas. De acuerdo con Siller & Tompkins (2005) varios de los paquetes estadísticos producen resultados iguales, por lo que el foco de priorización está en elegir el método para calcular las medidas de precisión.

En este sentido, según la ONU (2009), lo ideal sería utilizar métodos exactos para la estimación de las medidas de precisión; sin embargo, dada la complejidad de los diseños muestrales, no es posible obtener expresiones exactas para tal fin, por lo que, surgen otros métodos que permiten aproximaciones para estas medidas. Los métodos alternativos más usados y que se encuentran disponibles en la mayoría de los paquetes estadísticos se pueden dividir en tres grupos: conglomerado último; técnicas de replicado; y aproximaciones por linealización. A continuación, se realiza un breve resumen de los diferentes métodos⁴:

- El método de *conglomerado último* considera para el cálculo de la varianza únicamente la muestra contenida en las Unidades Primarias de Muestreo (UPM) y toma en cuenta los totales estimados de la variable en cada UPM, con independencia del muestreo hecho en posteriores etapas del diseño. Esto, simplifica considerablemente la fórmula de estimación de la varianza en muestras complejas, lo que constituye una de las principales razones de su extendido uso en encuestas.

³ Para más referencia sobre las consecuencias de no utilizar el módulo de muestras complejas para estimar los errores muestrales en diseño complejos, consultar Cassell (2008).

⁴ Para una revisión más detallada sobre los métodos de estimación de medidas de precisión bajo muestras complejas, consultar Martínez (2007), Wolter (2007) y Heeringa-West-Berglund (2010).

- Las *técnicas de replicación* consisten en tomar repetidas submuestras o réplicas de la muestra original y siguiendo el esquema de muestreo complejo, se estima la variable de interés a través de estas réplicas, para finalmente calcular el error estándar sobre la base de las estimaciones obtenidas. Las técnicas más utilizadas son: grupos aleatorios, réplicas equilibradas (BRR), Jackknife (JK1, JK2, y JK_n) y Bootstrap.
- El método de *aproximaciones por linealización* consiste en calcular la varianza de los estimadores a partir de la desagregación de los primeros términos de una serie de Taylor. Esto conlleva a expresar la estimación en función de una expansión en serie de Taylor para posteriormente aproximar la varianza mediante la parte lineal de dicha expansión aplicando los métodos ya descritos.

2.3.2. Efecto de diseño y tamaño de muestra efectivo

El efecto de diseño, denotada como *Deff*, se define como el cociente entre la varianza de un estimador en un diseño determinado y la varianza del estimador basada en MAS. Se considera también como el factor por el cual hay que multiplicar la varianza de una estimación basada en una muestra aleatoria simple, para tener en cuenta la complejidad del diseño muestral.

En cuanto a la interpretación, se puede decir que cuando el *Deff* es mayor a 1, significa que hay una mayor variabilidad en el muestreo complejo respecto al muestreo aleatorio simple, por tanto, se requiere mayor tamaño de muestra para obtener niveles de precisión similares al obtenido bajo MAS.

Por otro lado, también se puede decir que si n es calculado usando la fórmula de muestreo aleatorio simple, entonces $n_c = n \cdot Deff$ es el tamaño de muestra necesario en un diseño muestral complejo para alcanzar la misma varianza que en el muestreo aleatorio simple, de donde se deriva el concepto de *tamaño de muestra efectivo*, que se obtiene como $n_{ef} = n_c / Deff$. Por ejemplo, un tamaño de muestra bajo un diseño complejo de 60, considerando un *Deff* de 2, equivaldría a un tamaño de muestra efectivo (o bajo muestreo aleatorio simple) de 30.

2.3.3. Grados de libertad

Las distribuciones de probabilidad juegan un papel importante en la construcción de los intervalos de confianza y en los contrastes de hipótesis, así como en el cálculo del error absoluto y el error relativo (secciones 2.2.4 y 2.2.5). Entre los elementos que participan para poder emplear las distribuciones de probabilidad, las que permiten hacer inferencias respecto a parámetros poblacionales con determinada certeza, se encuentra los grados de libertad.

Los grados de libertad (*gl*) se definen como el número de observaciones independientes que son libres de variar (número de individuos en los datos) menos el número de parámetros estimados (número de relaciones impuestas a los datos)⁵. Además, según Heeringa, West & Berglund (2010), estos también dan cuenta de la precisión con la que se están estimando los parámetros a través de la muestra, por lo que, a mayores grados de libertad, las estimaciones serán más precisas, y viceversa.

A manera ilustrativa, en la *Tabla 1* se muestra la variación del percentil 0,975 de la distribución *t-Student* al incrementar los grados de libertad, valores que afectan, entre otras cosas, los resultados del intervalo de confianza, ya que, al ser este percentil el que multiplica al error estándar para obtener la amplitud del intervalo (*sección 2.2.6*) se refleja que, al tener menos grados de libertad, la amplitud del intervalo aumenta, generando estimaciones menos precisas; por el contrario, al tener mayores grados de libertad, la amplitud del intervalo disminuye, originando estimaciones más precisas.

Tabla 1. Percentil 0,975 de la distribución *t-Student*

Grados de libertad (gl)	Percentil 0,975 de la distribución <i>t-Student</i>
1	12,71
2	4,30
3	3,18
4	2,77
20	2,08
30	2,04
40	2,02
60	2,00
100	1,98
∞	$Z_{0,975}=1,96$

Fuente: Elaboración propia. INE, 2020.

⁵ Para más detalle sobre los grados de libertad consultar Pagano (2009).

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

A efectos de análisis de subpoblaciones, algunos *softwares* utilizan para el cálculo de los grados de libertad el enfoque previamente mencionado, aunque se ha demostrado en diferentes estudios que esta práctica no es la más adecuada ya que sobreestima los grados de libertad y como consecuencia, se originan intervalos de confianza más estrechos y con ello se incrementa la tasa de rechazo de la hipótesis nula. Ante esto, Korn & Graubard (1999) recomiendan un enfoque, denotado *cálculo de grados de libertad variable*, en donde se obtienen los grados de libertad mediante el total de conglomerados con muestra en la subpoblación menos el total de estratos con muestra en la subpoblación. Esto ofrece mejor resguardo en los análisis sobre todo en el caso de características poco prevalentes.

2.3.4. Fracción de muestreo y factor de corrección de poblaciones finitas

En el muestreo aleatorio simple, la relación que existe entre el tamaño de la muestra y el tamaño de la población, denotada por $f = \frac{n}{N}$, corresponde a la fracción de muestreo. El inverso $\left(\frac{N}{n}\right)$ correspondería al factor de expansión.

Por otro lado, el factor $\left(1 - \frac{n}{N}\right)$ o proporción de la población no incluida en la muestra, se conoce como factor de *corrección de población finita (cpf)* que representa el ajuste efectuado en el error estándar de la estimación para tener en cuenta el hecho de que la muestra se selecciona sin sustitución a partir de una población finita.

El término *cpf* tiene un límite superior de 1 y refleja la reducción de la varianza resultante del muestreo sin reemplazo. Si la fracción del muestreo f es pequeña, el término *cpf* tiende a 1 y su efecto sobre la varianza es mínimo, por lo que podría despreciarse. Por otra parte, si f es grande, despreciar el término *cpf* cuando la selección muestral se hace sin reemplazo, dará lugar a una sobreestimación de la varianza.

Por otro lado, las unidades primarias de muestreo (UPM) se seleccionan mediante muestreo sin reemplazo con probabilidad proporcional al tamaño (PPT), lo que hace más complejo el concepto de “fracción del muestreo”. No obstante, el número de UPM es a menudo grande y

la fracción del muestreo de las UPM en cada estrato es muy pequeña, dando un valor próximo a 1 para todos los términos *cpf* en la primera etapa.

3. Antecedentes

En esta sección se presenta la revisión realizada sobre los lineamientos para el uso de medidas de precisión que actualmente se acogen tanto a nivel internacional como a nivel nacional dentro del INE. Contar con este tipo de información, especialmente considerando la experiencia internacional en esta materia de las distintas Oficinas Nacionales de Estadísticas (ONE) es sumamente valioso al momento de realizar ajustes metodológicos, identificar las buenas prácticas y conocer las dificultades de este proceso.

Para ello, se exhibe en primera instancia la información en torno a esta temática consultada dentro de distintas ONE, y luego, para mostrar los precedentes en Chile, se presentan sucintamente las principales características de las seis encuestas de hogares con las que se trabajó para evaluar el impacto que tiene la aplicación de los lineamientos para el uso de medidas de precisión. Posteriormente se muestran los criterios vigentes hasta el año 2019 dentro de la institución para las distintas encuestas, los cuales son en general, heterogéneos y responden en forma mayoritaria a criterios particulares de cada encuesta de acuerdo con la evidencia empírica y los objetivos de cada estudio.

3.1. Antecedentes internacionales. Uso de medidas de precisión en oficinas estadísticas

Durante enero y febrero de 2019, se contactó a diversas ONE con el fin de evaluar cuáles y de qué forma usaban las medidas de precisión para sus encuestas de hogares. La consulta fue enviada a una muestra intencionada de 19 países de los cuales 13 contestaron el cuestionario.

Para ello se preparó un breve cuestionario, que indagaba sobre las siguientes dimensiones:

- Medidas de precisión utilizadas para la publicación y difusión.
- Medidas de precisión de acuerdo con tipología de estimadores.
- Uso de lineamientos o recomendaciones para la publicación y difusión de medidas de precisión.
- Requerimientos de precisión para las estimaciones.

- Acciones de la ONE cuando no se cumplen los requerimientos para las estimaciones.
- Casos especiales para el uso del coeficiente de variación.
- Aplicación de tamaño muestral mínimo para aplicar medidas de precisión.
- Estimación en áreas pequeñas.

A continuación, se presenta un breve resumen de cada una de las dimensiones indagadas:

- Medidas de precisión utilizadas para la publicación y difusión:** las medidas utilizadas por las oficinas estadísticas coinciden entre ellas, existiendo diferencia en las variantes empleadas, dentro de las cuales se encuentran el coeficiente de variación, error de muestreo, intervalos de confianza, tamaño de la muestra, error estándar, efecto de diseño y error relativo. Sin embargo, en algunas ocasiones se menciona que se hacen distinciones entre encuestas del mismo país.
- Medidas de precisión de acuerdo con tipología de estimadores:** de acuerdo con lo señalado por las oficinas estadísticas, en general se utilizan las medidas de precisión que ya habían señalado para los estimadores, independiente del tipo de indicador (total, media, proporción y razón).
- Uso de lineamientos o recomendaciones para la publicación y difusión de medidas de precisión:** esta dimensión busca conocer si las oficinas estadísticas utilizan algún lineamiento, recomendación o buenas prácticas relacionadas con las medidas de precisión. Según lo señalado, se siguen diversos tipos de recomendaciones, por ejemplo, se expone que siguen las recomendaciones que emanan de la Naciones Unidas, de otras oficinas estadísticas como INEGI⁶, STATCAN⁷, INE España⁸, el Código de buenas prácticas de Estadísticas Europeas EUROSTAT (2017) y el estándar relacionado con el reporte de calidad. Para el caso específico del INEGI, se utiliza un lineamiento respecto al diseño de la muestra en proyectos de encuesta INEGI (2011). En síntesis, se puede señalar que, si bien las oficinas declararon no contar con un lineamiento estándar, usan criterios comunes al provenir de recomendaciones en torno a las medidas de precisión acordados para los distintos tipos de estimadores.

⁶ Instituto Nacional de Estadística y Geografía de México.

⁷ Statistics Canadá.

⁸ Instituto Nacional de Estadística de España.

d. **Requerimientos de precisión para las estimaciones:** las oficinas estadísticas evidenciaron variedad de lineamientos, teniendo como criterio principal al coeficiente de variación, donde se observaron los siguientes umbrales para este indicador (más detalle en *Anexo 8.2, Cuadro 14*):

- Menor o igual a 15%, se considera aceptable o buena. Mayor a 15%, usar con precaución.
- Menor o igual a 20%, estimación confiable. Mayor a 20%, estimación no confiable.
- Menor a 16%, publicable. Entre 16 y 33,33, usar con precaución. Mayor a 33,33%, impublishable.
- Confiabilidad alta, entre 0 y 15%. Confiabilidad moderada, entre 15 y 30%. Confiabilidad baja, mayor a 30%.

Igualmente, se observaron desagregaciones que hacen referencia a la calidad de la estimación en diversos umbrales: cinco tramos (muy buena, buena, aceptable, utilizar con precaución, solo de referencia), y seis tramos (excelente, muy buena, buena, aceptable, precaución, no publicable).

Otro de los lineamientos utilizados por las encuestas es establecer una cantidad mínima de viviendas u hogares por dominio de estimación o inferencia estadística, que también varía según ONE. Por ejemplo, Paraguay establece una muestra mínima efectiva de 30 unidades, Argentina más de 50 unidades elementales y España entre 20 y 50 unidades muestrales.

e. **Acciones de la ONE cuando no se cumplen los requerimientos para las estimaciones.** Dentro de las acciones típicas que toman las oficinas estadísticas se encuentran: no publicación del estimador, incorporar advertencias a los usuarios respecto a la precisión de la estimación y agrupar categorías. En el caso particular del INEGI se señala que cuando un estimador no cumple con la precisión alta o media, la información se publica con advertencia. También existen oficinas estadísticas que, al tener menos de 30 casos en ciertas subcategorías, deciden no publicar el indicador y en caso de publicación se incorporan advertencias a los usuarios. Por su parte el INE España difiere en su tratamiento de acuerdo con las decisiones que se toman para cada encuesta; en general siguen las recomendaciones de Eurostat que establece que:

- No se debe publicar una estimación basada en menos de 20 elementos muestrales.
 - Las estimaciones obtenidas con una muestra entre 20 y 50 elementos deben llevar un asterisco para indicar que tienen una fiabilidad baja.
 - Se publican todos los indicadores con el coeficiente de variación; si exceden el umbral definido, se explica al usuario que ese resultado debe ser tomado como un valor descriptivo⁹.
- f. **Tratamiento especial para el uso del coeficiente de variación cuando la prevalencia de un atributo es baja.** En términos generales, los países declararon que no tienen un uso diferenciado o específico para este tipo de casos, en consecuencia, se da el mismo tratamiento que al resto de las estimaciones. A pesar de esto, en Argentina se advierte que la estimación de la prevalencia debe ser superior al 5% para aceptar el coeficiente de variación como medida de precisión, de lo contrario, se destaca al usuario que la calidad de la estimación podría estar comprometida y se presenta la desviación estándar como medida de precisión.
- g. **Aplicación de tamaño muestral mínimo para aplicar medidas de precisión.** De los países que declararon considerar este criterio se establecen ciertos tamaños mínimos para aplicar el criterio del tamaño muestral, por ejemplo, diferenciando para el área “rural” y “urbana”; también para las divisiones administrativas. Por ejemplo, Argentina establece un mínimo de 50 viviendas/hogares para obtener estimaciones a nivel de dominios de estudio o áreas geográficas pertenecientes a la división política administrativa o estratos de muestreo. También establece que las Unidades Primarias de Muestreo (UPM) o conglomerados de viviendas deben tener al menos cinco viviendas/hogares con respuesta (en caso contrario se colapsan para el cálculo de la estimación de la varianza) y todos los estratos a nivel de UPM deben contemplar al menos cinco unidades (en caso contrario se colapsan para el cálculo de la estimación de la varianza). Existen ONE que utilizan este criterio mínimo de casos para publicar pero

⁹ Descriptivo en el contexto de estadísticas descriptivas, es decir, se dan a conocer los distintos resultados extrapolados a la población, pero sin connotación de planteamiento de hipótesis del verdadero valor en la población, ni se analiza la significancia estadística en las diferencias observadas respecto a otros períodos. Dado que los valores extrapolados carecen de precisión suficiente de acuerdo con los umbrales establecidos a priori, no se consideran fiables por tener mucha variabilidad muestral (varían dependiendo exclusivamente de la muestra extraída, que por lo general se da en casos en que el tamaño es pequeño para recoger la verdadera variabilidad del fenómeno en la población).

que está sujeto a otras temáticas, como por ejemplo que difiere según encuesta (Canadá, Francia) o según la subpoblación que se está utilizando, al ponderar el tamaño muestral obtenido con la prevalencia que esta tiene en la realidad (Australia, Perú, Francia).

- h. Estimación en áreas pequeñas.** En general no se usa la estimación en áreas pequeñas o recién se está incorporando esta técnica a los procedimientos de las oficinas estadísticas. De acuerdo con esto, esta técnica se utiliza tomando como insumos principales los censos de población y los registros administrativos.

A partir de esta consulta a las distintas ONE se pudieron encontrar diversos criterios sobre las medidas de precisión de las estimaciones, donde se puede deducir que, en general, no existen lineamientos estándares en esta materia. Por un lado, las ONE manifestaban que existían diferencias en los criterios al momento de evaluar la calidad de los datos –que como se advierte y evidencia a continuación, también ocurre en el INE Chile- y por otro, entre los países consultados, tampoco se puede indicar con claridad un criterio uniforme, lo que conlleva a un desafío pendiente para las estadísticas oficiales.

3.2. Antecedentes nacionales. Uso de medidas de precisión en el INE Chile

En el INE se levantan y publican varias encuestas. Algunas son propias, como la Encuesta Nacional de Empleo, la Encuesta Suplementaria de Ingresos, la Encuesta de Presupuestos Familiares y la Encuesta Nacional sobre Uso del Tiempo. Otras, aunque también son diseñadas y levantadas por el INE, son producto de convenios con otras instituciones del Estado, como la Encuesta Nacional de Seguridad Ciudadana (encargada por el Ministerio del Interior) y la Encuesta de Microemprendimiento (encargada por el Ministerio de Economía).

Para efectos de establecer los lineamientos respecto a la calidad de las estimaciones y la medición del impacto en las cifras a publicar, dadas las coyunturas de las encuestas, se consideraron las mencionadas anteriormente por ser procesos continuos para los cuales se ha alcanzado cierta sistematización y experiencia profesional, compartiendo los mismos problemas de calidad en la publicación de cifras.

3.2.1. Descripción de las encuestas evaluadas

3.2.1.1. Encuesta Nacional de Empleo

La Encuesta Nacional de Empleo (ENE) es una encuesta continua a hogares que se aplica a las personas de 15 años y más que son residentes habituales en viviendas particulares ocupadas, con el objetivo de caracterizar su relación con el mercado laboral, clasificando a las personas en ocupados, desocupados y fuera de la fuerza de trabajo. El indicador principal es la tasa de desocupación. La encuesta, para el momento del diagnóstico, posee áreas de estimación publicables a nivel nacional, regional, provincial y de algunas ciudades del país.

La ENE es una encuesta de diseño muestral bietápico (selección de conglomerados de viviendas en primera etapa y de viviendas en segunda etapa) de carácter trimestral, es decir, la muestra completa se divide en tres submuestras, obteniéndose mensualmente estimaciones en base al trimestre completo anterior del mes de publicación. Por ejemplo, las estimaciones publicadas en abril de un año dado corresponden al trimestre enero-febrero-marzo (EFM) del mismo año. Esta muestra se va refrescando a razón de 1/6 mensual, de modo que para las estimaciones del trimestre EFM, 1/6 de la submuestra del mes de marzo es rotada por nuevas viviendas de refresco, conservándose para la expansión, los valores muestrales obtenidos de los dos meses anteriores (enero y febrero) y los 5/6 de la muestra del mes de marzo. Solo los valores de 1/6 de la muestra rotada son los que aportan valores muestrales nuevos al trimestre de expansión.

A lo largo de su historia, la ENE ha incorporado mejoras tanto en su diseño muestral como en los tópicos que aborda, conforme a las mejores prácticas adoptadas por países miembros de la Organización para la Cooperación y el Desarrollo Económico (OCDE) y en línea con las recomendaciones de la Organización Internacional del Trabajo (OIT).

En relación con las mejoras en el diseño muestral, durante 2019 se estableció un diseño que experimentó cambios sustanciales en los distintos ámbitos que lo definen¹⁰. La implementación de dicho diseño se realizará mediante un método de incorporación progresiva, entre enero y junio de 2020, cuyo objetivo es aminorar el efecto que pudiera producir el nuevo diseño en las estimaciones de parámetros de interés.

¹⁰ Entre los cambios se encuentran: actualización del marco muestral, cambio de la configuración de la unidad de muestreo de primera etapa, nueva estratificación muestral, ampliación de la cobertura geográfica, aumento del tamaño muestral, mejor forma de calibración del factor de expansión, entre otros. Más antecedentes en INE Chile (2020).

3.2.1.2. Encuesta Suplementaria de Ingresos

La Encuesta Suplementaria de Ingresos (ESI) es un módulo complementario que se aplica en la Encuesta Nacional de Empleo, en el trimestre octubre-noviembre-diciembre (OND) de cada año. Su objetivo es caracterizar los ingresos laborales de las personas que son clasificadas como ocupadas en la ENE y los ingresos de otras fuentes de los hogares, tanto a nivel nacional como regional, según precios de cada año.

La encuesta comprende un módulo de preguntas vinculada al formulario habitual de la encuesta de empleo, por lo que constituye una fuente de información complementaria de la caracterización de la población ocupada, respecto de los ingresos laborales que le son inherentes. Es la única encuesta que levanta anualmente información de ingresos laborales de las personas en Chile y que tiene representatividad para todas las regiones.

La encuesta mide dos fuentes de ingresos que se circunscriben en el concepto de ingreso corriente (Grupo Intersecretarial de Trabajo sobre Cuentas Nacionales, GITCN (2009)) el proveniente de la ocupación y aquel proveniente de otras fuentes, percibidos tanto por las personas como por los hogares, durante el periodo de referencia.

Los ingresos provenientes de fuentes laborales o de la ocupación son los ingresos netos, es decir, excluyen descuentos legales e impuestos, y distinguen aquellos cuya fuente de origen es el trabajo dependiente de aquellos provenientes del trabajo independiente. Por otra parte, los ingresos de otras fuentes capturan ingresos provenientes de rentas de la propiedad, activos financieros y transferencias. Con ambas fuentes de ingresos es posible estimar la composición y distribución de los ingresos de la ocupación y de los hogares.

3.2.1.3. Encuesta de Presupuestos Familiares

La Encuesta de Presupuestos Familiares (EPF) es una encuesta aplicada a hogares, que tiene como objetivo principal identificar la estructura y característica del gasto en consumo final de los hogares urbanos, en las capitales regionales de Chile y algunas de sus zonas conurbadas, con un período de referencia de un año. El objetivo secundario de la encuesta es identificar la estructura del ingreso total disponible de los hogares urbanos, con la misma cobertura geográfica y temporal antes descrita.

La periodicidad de la encuesta era de 10 años, pero con el ingreso de Chile a la OCDE en 2010 se estableció el compromiso país de actualizar el Índice de Precios al Consumidor (IPC)

con una periodicidad máxima de cinco años. Por ello la EPF, como principal fuente de información para el IPC, se comenzó a realizar con una periodicidad quinquenal.

El diseño muestral de la encuesta corresponde a un muestreo bietápico, estratificado por comuna, clasificación socioeconómica y clasificación de las manzanas según grupo de tamaño (considerando el total de viviendas que las componen).

El ámbito de representatividad de las primeras cinco versiones de la encuesta fue la zona urbana del Gran Santiago. El área de estimación se extendió a partir de la VI EPF, pasando a ser una encuesta de alcance nacional, y como consecuencia el ámbito geográfico se amplió al conjunto de capitales regionales del país, más algunas de sus zonas conurbadas.

El período de captura de información de la VIII EPF fue de 12 meses consecutivos, por lo que la muestra completa seleccionada se distribuyó en 12 submuestras mensuales, pero la expansión del gasto, ingresos y demás indicadores se realiza con la muestra completa.

La información obtenida correspondió al período comprendido entre julio de 2016 y junio de 2017. En cuanto a la representatividad de la muestra, es importante mencionar que:

- El diseño muestral de la VIII EPF permitió realizar estimaciones consistentes y estadísticamente significativas con frecuencia temporal anual y, en términos geográficos, para las áreas de estimación (i) Gran Santiago, (ii) Resto de capitales regionales y (iii) Total de capitales regionales.
- Si bien el trabajo de campo de la encuesta se realiza a un nivel más desagregado en términos geográficos (manzana, comuna, región) y temporales (mes y quincena), los datos recabados por el estudio no permiten realizar estimaciones a esos niveles, ya que no cuentan con niveles de precisión adecuados para una estimación confiable.

3.2.1.4. Encuesta Nacional sobre Uso del Tiempo

La Encuesta Nacional sobre Uso del Tiempo (ENUT) es de carácter nacional urbana. La primera versión se levantó entre septiembre y diciembre de 2015. Su objetivo general es obtener información sobre el uso del tiempo de la población de 12 años y más, respecto a actividades de trabajo realizadas en el mercado, el trabajo no remunerado y las actividades personales. Sus objetivos específicos son:

- Visibilizar la carga global de trabajo de la población de 12 años y más a través de la medición del tiempo que se dedica al trabajo en la ocupación y sus actividades conexas y al trabajo no remunerado.

- Evidenciar las desigualdades de género en el tiempo dedicado a trabajo doméstico y de cuidados no remunerado.
- Caracterizar la desigualdad en el uso del tiempo de acuerdo con variables de clasificación como edad, sexo, vínculo con el mercado laboral y nivel socioeconómico.
- Obtener información sobre el uso del tiempo de las personas de 12 años y más en actividades personales.

La población objetivo la constituyen todas las personas de 12 años y más, que tengan condiciones físicas y mentales aptas para responder la encuesta, que residan habitualmente en viviendas particulares ocupadas (elegibles) y que se encuentren en estas durante el período de levantamiento. Esta población involucra al área urbana de todas las regiones del país, y específicamente a las principales comunas (118 en total) que acumulan el 85% de la población total nacional (según proyecciones de población) ordenadas de mayor a menor en cuanto a cantidad de personas.

Las unidades de observación corresponden a los integrantes del hogar de 12 años o más, presentes en la vivienda durante el período de levantamiento de la encuesta, con capacidades físicas y mentales aptas para responder, mientras que las unidades de análisis corresponden a personas, hogares y las actividades que estos declaran al momento de la entrevista.

El diseño muestral de la ENUT es probabilístico y bietápico, en que las unidades primarias de muestreo o de primera etapa corresponden a manzanas, mientras que las unidades secundarias de muestreo o de segunda etapa, corresponden a las viviendas.

3.2.1.5. Encuesta Nacional de Seguridad Ciudadana

La Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) se enmarca en un convenio de colaboración anual entre la Subsecretaría de Prevención del Delito (SPD) del Ministerio del Interior y el INE. Esta encuesta surge a partir de la necesidad de obtener datos desagregados por regiones, de la población residente en zonas urbanas del país, que den cuenta de su percepción de inseguridad y reacción frente al delito, y la victimización de personas y hogares. Dado lo extenso de la serie (14 versiones desde 2003), la ENUSC constituye una de las experiencias de mayor riqueza a nivel nacional e internacional en la medición tanto de la victimización como de la percepción de inseguridad y reacción frente al delito.

La ENUSC busca obtener información sobre la percepción de inseguridad, la reacción frente al delito y la victimización de personas y hogares, a partir de una muestra representativa de zonas urbanas a nivel nacional y regional. Para la versión XIV de la encuesta se definieron como objetivos específicos:

- Caracterizar la percepción de inseguridad y conocer la reacción frente al delito de las personas de 15 años o más, residentes en zonas urbanas del país y de las regiones.
- Obtener datos estadísticos acerca de la victimización de los hogares y personas, residentes en zonas urbanas del país, que hayan sido objeto de intento(s) de delito(s) o de delito(s) consumado(s).
- Caracterizar los delitos e intentos de delitos que hayan afectado a los hogares y las personas de las zonas urbanas del país y de las regiones.

La población objetivo la constituyen las personas de 15 años o más y los hogares de las viviendas particulares ocupadas del área urbana de 101 comunas del país. Las unidades de análisis son los hogares y las personas de 15 años y más que los constituyen, representados por un informante seleccionado por el método de Kish¹¹, siendo este último, la unidad de información. Tal informante entrega información sobre su percepción de inseguridad y reacción frente al delito y datos sobre victimización que lo han afectado a él o a cualquier otro integrante del hogar durante los últimos 12 meses.

El diseño muestral de la ENUSC es probabilístico, trietápico, estratificado geográficamente y por tamaño poblacional en términos de vivienda. Las unidades de primera etapa son conglomerados denominados manzanas, mientras que las viviendas y las personas son las unidades de segunda y tercera etapa, respectivamente.

3.2.1.6. Encuesta de Microemprendimiento

La Encuesta de Microemprendimiento (EME) está dirigida a hogares en donde reside un dueño de un microemprendimiento, tiene carácter bienal y es realizada desde el año 2013 por el INE en conjunto con el Ministerio de Economía, Fomento y Turismo, siendo el instrumento oficial para caracterizar la heterogénea realidad de los microemprendimientos en Chile.

¹¹ Método de muestreo utilizado para seleccionar un individuo al azar dentro de un hogar. Utiliza una tabla predeterminada para seleccionar un individuo, basándose en el número total de individuos que viven en el hogar Kish (1963).

La EME es una encuesta con un diseño muestral bifásico, pues la primera fase corresponde a la Encuesta Nacional de Empleo (ENE), desde donde se selecciona (en segunda fase) una muestra representativa de viviendas en la que resida algún microempresario, los que corresponden a trabajadores por cuenta propia o empleadores dueños de una microempresa (unidad económica de hasta 10 trabajadores incluyendo al dueño).

La última versión de la encuesta cuenta con 10 módulos que contienen 89 preguntas que permiten obtener información relevante de los microempresarios y sus negocios. En detalle se obtiene información relevante de la formalidad de los negocios, su vínculo con el sistema financiero, la cantidad y el tipo de empleo generado, entre otros.

Es importante señalar que, uno de los principales objetivos de la encuesta es estimar y caracterizar las unidades económicas pertenecientes al sector informal, las que no pueden ser estudiadas mediante encuestas tradicionales a establecimientos, ya que, estas se basan en el listado de empresas registradas ante el Servicio de Impuestos Internos (SII), las cuales por definición corresponden al sector formal.

3.2.1.7. Análisis de género en el INE Chile

La Sección de Estudios y Estadísticas de Género (SEEG) tiene por objetivo incorporar el enfoque de género en la producción estadística para su uso en la elaboración de políticas públicas, y satisfacer las necesidades de información de agentes públicos y privados, así como de la población en general. Ello implica el análisis de diversos productos estadísticos a nivel institucional con el fin de visibilizar las brechas de género que existen en el país. Esto ha creado la necesidad de desagregar aún más los tabulados de las encuestas previamente presentados, creando nuevas aperturas y analizando información específica por subpoblaciones (hombres y mujeres) lo que representa un mayor desafío en cuanto a calidad de las estimaciones publicadas.

El INE ha avanzado en la incorporación del enfoque de género en la producción estadística, en especial en los procesos de “análisis” y “difusión”. Un ejemplo de ello es la elaboración de cuadros estadísticos a partir de las distintas encuestas del INE, incluyendo tanto encuestas sociales como económicas, datos de censos y registros administrativos. Estos cuadros cumplen la función de entregar información relevante respecto a temáticas de género tanto a nivel nacional como regional. En este contexto, a nivel de encuestas de hogares, los análisis de género que se evalúan de la SEEG para los lineamientos de medidas de precisión son los que pertenecen a las encuestas ENUSC, ESI, ENUT y EME.

En general, la SEEG al realizar análisis con mayor desagregación que los productores de cada encuesta utiliza los mismos criterios de evaluación de la calidad de las estimaciones establecidos por cada producto. Los criterios actualmente utilizados se desarrollan en el siguiente apartado.

3.2.2. Medidas de precisión actuales

Considerando los diversos objetivos, periodicidad y tamaños de la muestra de la encuesta, actualmente cada producto implementa sus propios lineamientos en cuanto a medidas de precisión. Si bien cada encuesta tiene sus propios criterios, ellos también cuentan con similitudes, ya sea por los criterios que utilizan para evaluar sus estimaciones (*Cuadro 8*) o por los límites establecidos a partir de ellos para determinar si se publican oficialmente los datos.

Cuadro 8. Medidas de precisión considerados actualmente por producto

Producto	Medidas de precisión que considera
ENE	Error relativo y coeficiente de variación
ESI	Error relativo, coeficiente de variación y tamaño muestral
EPF	Coeficiente de variación y tamaño muestral
ENUT	Coeficiente de variación
ENUSC	Coeficiente de variación
EME	Coeficiente de variación

Fuente: Elaboración propia. INE, 2020.

Actualmente en las encuestas de hogares del INE se utilizan tres indicadores para evaluar las medidas de precisión en las encuestas: error relativo, coeficiente de variación y tamaño muestral. Como se logra apreciar, todas las encuestas tienen en común el uso del coeficiente de variación como parámetro para determinar la precisión del dato. Asimismo, la ENE y la ESI lo complementan también observando el error relativo¹² que se asocia a ellos para clasificar la confiabilidad de sus estimaciones en términos estadísticos (*Cuadro 9*).

¹² Ver secciones 2.2.4 y 2.2.5. En términos prácticos, los límites de los rangos para el error relativo equivalen al doble de los rangos dados para el coeficiente de variación al considerar el 95% de confianza de la estimación por intervalos, esto es: $er(\hat{\theta}) = Z_{1-\frac{\alpha}{2}} \cdot ea(\hat{\theta})$, con nivel de confianza $(1 - \alpha) = 0,95$ y $Z_{1-\frac{\alpha}{2}} = Z_{0,975} \approx 1,96 \approx 2$.

Cuadro 9. Clasificación del coeficiente de variación y error relativo

Clasificación	Rango del coeficiente de variación (cv) (%)	Rango del error relativo (er) (al 95% de confianza) (%)
Muy bueno	(0 ; 5)	(0 ; 10)
Bueno	[5 ; 15)	[10 ; 30)
Regular	[15 ; 20)	[30 ; 40)
No recomendable	[20 ; ∞)	[40 ; ∞)

Fuente: Manual metodológico y conceptual de la ENE. INE, 2018.

Las encuestas presentan en común que, a partir del valor obtenido en el coeficiente de variación de la estimación, recomiendan lo siguiente:

- I. **Muy bueno.** Hasta 5% se considera que la estimación es óptima en términos de precisión.
- II. **Bueno.** Entre 5% y 15% se considera una buena estimación en términos de precisión.
- III. **Regular.** Entre 15% y 20% se podría considerar una estimación poco confiable al tener alta variabilidad. Se debe tener precaución al momento de realizar análisis y realizar conclusiones a partir de ellas.
- IV. **No recomendable.** Sobre el 20% la estimación se presenta como poco precisa, es decir, deficiente en términos de precisión y por ende no recomendable, independiente del número de observaciones que se tenga disponible.

Asimismo, la diferencia entre las encuestas, en cuanto al coeficiente de variación, radica en los límites establecidos para publicar (menor o igual que 15% o menor o igual que 20%) y cuándo ellos pueden ser utilizados exclusivamente con un uso descriptivo.

Adicionalmente, algunas de las encuestas también consideran la evaluación del número de observaciones de cada estimación. En tanto, al presentarse pocos casos¹³, se genera teóricamente, una alta dispersión en la estimación de las medidas de error, lo que dificulta su interpretación como indicador de calidad, y que se traduce en cuestionamientos en la confiabilidad del dato. Por lo mismo, tanto la ESI como la EPF definen un número mínimo de observaciones para la publicación de sus estimaciones (*Cuadro 10*).

¹³ Aunque a veces se hable de casos, unidades, elementos, unidades elementales, unidades muestrales u observaciones muestrales, todos estos términos se refieren indistintamente a las unidades de análisis, que corresponden por lo general, a la unidad de la última etapa de muestreo, es decir, la vivienda u hogar o a la persona que contesta la encuesta (esta última se denomina unidad de información u observación).

Cuadro 10. Criterios actualmente considerados por producto

Producto	Medidas de precisión que considera		
	Error relativo	Coefficiente de variación	Tamaño muestral
ENE	< 40%	< 20%	-
ESI	< 40%	< 20%	> 5 observaciones
EPF	-	≤ 15%	≥ 6 hogares
ENUT	-	< 20%	-
ENUSC	-	< 20%	-
EME	-	< 20%	-

Fuente: Elaboración propia. INE, 2020.

En el caso de la ENE, si bien se reconoce la importancia de evaluar el tamaño muestral, al no existir una recomendación oficial al respecto, debido a que los parámetros poblacionales de las variables de interés difieren entre sí, dificulta la tarea de definir un único umbral que sea transversal a todas las categorías o desagregaciones de la encuesta. Por lo anterior, dicha encuesta deja a criterio de los analistas la responsabilidad de definir el límite mínimo de observaciones que aseguren representatividad y precisión del estimador. A su vez, se recomienda que, al realizar alguna estimación, se considere que las observaciones queden distribuidas en distintos conglomerados y estratos, y que, además, exista varianza al interior de ellos, dicho de otro modo, que exista un número suficiente de observaciones para el cálculo de la varianza.

Tomando en consideración las dos medidas mayormente utilizadas por las encuestas aquí evaluadas, coeficiente de variación y número de observaciones muestrales, las estimaciones pueden ser clasificadas en cuanto a su calidad como se presenta en *Cuadro 11*.

De acuerdo con la clasificación dada, se define conceptualmente:

- I. **Estimador no confiable.** No se recomienda utilizar este estimador para predecir características poblacionales (inferencia estadística). Una alternativa es agrupar categorías. Al agrupar o colapsar las categorías se debe analizar nuevamente su precisión, ya que se obtendrán otros valores y, por ende, se debe determinar su idoneidad para un uso inferencial o solamente descriptivo.
- II. **Uso descriptivo (nota metodológica).** En los casos que la estimación realizada por el investigador no pueda ser agrupada, es necesario incluir una nota metodológica que advierta las restricciones y el error asociado a dicha estimación y que solo debe ser usado

con fines descriptivos, esto es, que en general se pueda usar para sacar conclusiones de tendencias, pero no de estimaciones de totales.

III. **Estimador confiable.** Se puede utilizar la estimación según los objetivos propuestos.

Adicional a lo presentado anteriormente, fruto del trabajo desarrollado en la SEEG, quienes siguen los criterios de cada productor, es importante relevar que el trabajar con desagregaciones por sexo, ha dado cuenta de ciertas limitaciones para la captura de información que se encuentra presente en subpoblaciones muy reducidas. Por ejemplo, en la ENE, a nivel regional, al desagregar por rama de actividad económica y sexo, se observa una feminización o masculinización de ciertas ramas, lo cual genera problemas de representatividad de las estimaciones que tienen una menor prevalencia de observaciones, donde incluso un aumento del tamaño muestral no asegura que la subpoblación con menos observaciones obtenga la representatividad que le permita cumplir con los criterios de precisión definidos por la encuesta.

Cuadro 11. Clasificación de la estimación según cv y tamaño muestral

CV	Encuesta											
	ENE Número de observaciones		ESI Número de observaciones		EPF Número de observaciones		ENUT	ENUSC	EME			
	Bajo	Alto	≤ 5	> 5	≤ 5	> 5						
(0% ; 5%)	Estimador confiable	Estimador confiable	Estimador no confiable	Estimador confiable	Estimador no confiable	Estimador confiable	Estimador confiable	Estimador confiable	Estimador confiable			
[5% ; 15%)	Descriptivo (nota metodológica)									Descriptivo (nota metodológica)/ Estimador no confiable		
[15% ; 20%)	Estimador no confiable	Descriptivo (nota metodológica)				Estimador no confiable					Estimador no confiable	Estimador no confiable
[20% ; ∞)		Estimador no confiable								Descriptivo (nota metodológica)		

Fuente: Elaboración propia. INE, 2020.

Asimismo, trabajar con mayores aperturas en la mayor parte de los productos, ha obligado a crear nuevos lineamientos dentro de los mismos criterios antes presentados. Por ejemplo, para la publicación de tabulados relativos a género e ingresos, debido a la baja prevalencia de observaciones al momento de generar desagregaciones a nivel regional por sexo con la ESI (ver *Cuadro 10*) se incorporó un criterio adicional de publicación para los tabulados a nivel regional, que exigía, además de los requisitos de precisión que exige la encuesta, cumplir con los siguientes requerimientos:

- I. Más del 50% de las celdas deben cumplir con los criterios de precisión definidos en la ESI.
- II. Más del 50% de las regiones deben cumplir con el punto anterior, para el tabulado analizado.

A raíz de lo anterior, es más fácil notar que, si bien actualmente existen criterios comunes entre todos los productos, siguen existiendo diferencias en los criterios y límites utilizados para evaluar las estimaciones, lo que conlleva a la necesidad de incorporar un lineamiento único. La importancia que tiene contar con un estándar de evaluación para la producción estadística radica en que las instituciones de la administración del Estado, así como los usuarios en general, necesitan que la información generada cumpla con los estándares de calidad para la focalización correcta de los recursos destinados a las políticas públicas, por lo cual es fundamental definir indicaciones para la correcta utilización de la información estadística producida.

4. Recomendaciones sobre medidas de calidad

Tradicionalmente la calidad de una estimación ha sido medida a través de su precisión¹⁴, sin embargo, en la práctica es posible encontrar estimaciones que si bien alcanzan buena precisión, han sido calculadas con tamaños de muestra relativamente bajos, por lo que no se puede garantizar la representatividad de esta.

A continuación, se dan a conocer una serie de criterios relacionados con la calidad de las estimaciones, que serán evaluados de manera independiente, para luego ser conjugados con la pretensión de generar lineamientos para garantizar la calidad de las estimaciones. Estos

¹⁴ La medida de precisión predilecta para medir la calidad de una estimación empleadas por la ONE es el cv, como se constata en Molina (2018) y Gutiérrez (2018) a pesar de tener algunas propiedades no deseables ampliamente desarrolladas en Heath & Borowski (2003) y Vásquez & Núñez (2011).

criterios se pueden subdividir en dos grandes bloques, a saber: (i) criterios relacionados con el tamaño muestral y grados de libertad y (ii) criterios relacionados con la precisión.

4.1. Evaluación de criterios respecto a tamaño muestral y grados de libertad

4.1.1. Tamaño muestral

Sin pérdida de generalidad, al señalar el término “tamaño muestral” se hará referencia al total de unidades de análisis consideradas para obtener las estimaciones. Con esto se busca comprender las diferentes unidades de análisis que se pueden encontrar en las encuestas de hogares por muestreo, a saber: viviendas, hogares y/o personas¹⁵.

Tanto la aleatoriedad como el tamaño muestral juegan un papel fundamental para lograr hacer inferencias acertadas a través de la muestra, ya que a medida que se aumentan los tamaños muestrales, utilizando selecciones aleatorias, esperamos poder representar de mejor manera las características presentes en la población, lo que trae consigo estimaciones de mejor calidad.

En cuanto a esto, el Teorema Central del Límite (TCL) expresa que, si la distribución del estimador es desconocida y no es *Normal*, entonces se necesita un tamaño muestral suficientemente grande para que su distribución converja a la *Normal*, lo que resulta de mucha utilidad a la hora de querer hacer inferencias de la población. Ante esto, surge la pregunta sobre cuál sería un tamaño muestral suficientemente grande. Al respecto, Lincoln (1993) comenta que, si bien se prefieren tamaños muestrales mayores o iguales a 100, en la mayoría de las aplicaciones se considera suficiente un tamaño muestral de 30 o más para permitir el empleo de la distribución *Normal*.

Por su parte, Barnett-Walker (2003) sugieren que el tamaño de muestra mínimo en cuanto a unidades de análisis sea 100. Por otra parte, Hornik et al. (2002) propone que cualquier estimación basada en una muestra efectiva de 30 o menos debe ser suprimida. Respecto a proporciones cerca de 0,01 establecen un tamaño de muestra efectivo mínimo de 140.

¹⁵ El recuento del tamaño muestral se realiza dependiendo del tipo de estimador. Para proporciones y razones entre 0 y 1, se realiza el conteo sobre las unidades de análisis del denominador; mientras que para el resto de las estimaciones se hace sobre el total de unidades de análisis que nutren la estimación.

Las recomendaciones generales de Eurostat (2013) en este sentido son las siguientes: (i) no se debe publicar una estimación basada en menos de 20 elementos muestrales y (ii) las estimaciones obtenidas con una muestra entre 20 y 50 elementos deben llevar un asterisco (*) para indicar que tienen una fiabilidad baja.

Como puede observarse, casi todas las referencias sobre el tamaño muestral adecuado se hacen en el contexto de muestreo aleatorio simple de unidades de muestreo y no se especifica si son o no las unidades de análisis.

Por otra parte, la mayoría de las encuestas de hogares, emplean técnicas de muestreo complejo para extrapolar las características de la muestra a la población, y las unidades de muestreo primarias son conglomerados¹⁶ de unidades secundarias o viviendas, en las cuales puede existir más de un hogar como unidad de análisis, o bien pudiera incorporarse otra etapa de selección, que involucre la selección de un miembro del hogar como unidad de análisis.

Pero, en muestreo complejo, la conglomeración, al igual que la ponderación desigual de las unidades de análisis, genera un efecto de diseño *Deff* que aumenta la varianza de las estimaciones en comparación con una muestra aleatoria simple de unidades de análisis. Por su parte, la estratificación por sí sola, si está correlacionada positivamente con la estimación, reduce la varianza; en caso contrario, la aumenta. En muestreo complejo, el efecto de diseño *Deff* recoge el efecto combinado de estos tres elementos.

En este contexto, debe ser considerado el efecto de diseño *Deff* para determinar el tamaño muestral efectivo (n_e) que se describe como el tamaño de muestra bajo muestreo aleatorio simple (MAS) equivalente al tamaño muestral (n) obtenido bajo un diseño muestral complejo. La forma de cálculo del tamaño muestral efectivo viene dada por:

$$n_e = n/Deff$$

¹⁶ La técnica compleja de muestreo utilizada en la mayoría de los *softwares* estadísticos es la del conglomerado último, es decir, toda la información obtenida de las unidades de análisis se aglomera en los conglomerados para el cálculo de las estimaciones y varianzas. El tamaño de muestra de conglomerados o unidades de primera etapa (existe selección aleatoria de conglomerados en primera etapa) no se considera en la determinación del tamaño muestral, sino que en los grados de libertad (ver *sección 2.3.3* y *4.1.2* siguiente) pero para el cálculo de la varianza, se requieren como mínimo dos en cada estrato de diseño.

Recomendación

Como resultado de la revisión se recomienda establecer un tamaño muestral mínimo de 60 unidades para generar una estimación fiable. Esto equivale, considerando un efecto del diseño promedio de 1,5 (2,0) a una muestra efectiva de 40 (30). No obstante, en los casos que el investigador consiga efectos de diseño mayores a 2,0 deben tomarse las precauciones pertinentes orientadas a incrementar el tamaño de muestra efectivo.

A lo anterior se añade que, aunque se ha optado por establecer un tamaño muestral mínimo de unidades por debajo de lo encontrado en la literatura (lo que en esta primera instancia facilita la implementación en términos del impacto que pudiera ocasionar este valor) se deja abierta la posibilidad de evaluar la pertinencia de aumentar el mínimo de unidades requeridas en etapas posteriores a la acogida de esta recomendación.

4.1.2. Grados de libertad

La mayoría de las técnicas utilizadas para análisis de muestras complejas realizan las estimaciones mediante el método del conglomerado último, es decir, todo se totaliza a nivel de conglomerado para calcular la varianza de los indicadores o estimadores mediante la intervianza o la varianza entre los conglomerados. Por esta razón, también toma relevancia el recuento de los conglomerados que intervienen en el cálculo de las medidas de precisión.

Por ejemplo, Vaillant, Denver & Kreuter (2013) concuerdan que 60 conglomerados o unidades primarias son suficientes para estimaciones de intervalos, y, por ende, para contrastes de hipótesis.

La cifra de 60 conglomerados asegura un comportamiento aproximadamente *Normal* o *studentizado* de la variable de interés a través de ellos, que además sirve para la construcción de los intervalos de confianza. También asegura la representatividad mínima, al ser factible la distribución de los conglomerados en los diferentes estratos o dominios de estudio.

Desde este punto de vista, 60 conglomerados prevén que el número de viviendas, hogares y/o personas sea mucho mayor, que son las posibles unidades de análisis de las cuales se indagan sus características.

En capítulos anteriores se ha expuesto la relación existente entre el número de conglomerados en la muestra y los grados de libertad (ver *sección 2.3.3*) los cuales juegan un papel sustancial a la hora de hacer inferencia dentro de subpoblaciones de interés, por

tanto, se busca controlar el tamaño muestral de primera etapa en función de los grados de libertad, quedando por definir el enfoque a adoptar para el cálculo de estos, así como el umbral de aceptación para que las estimaciones de interés sean consideradas admisibles.

En cuanto al enfoque de cálculo de los grados de libertad, se precisan dos formas de cálculo, a saber: *cálculo de grados de libertad fijos* y *cálculo de grados de libertad variables*. Dadas las consecuencias observadas¹⁷ al evaluar ambas opciones, se opta por la segunda¹⁸, pues permite obtener intervalos de confianza más amplios y por ende un mejor resguardo en cuanto a la precisión asociada a las estimaciones de subpoblaciones con pocos grados de libertad.

Respecto al umbral de aceptación de los grados de libertad, es menester conocer la naturaleza del percentil 0,95 de la distribución *t-Student* (percentil asociado a un nivel de confianza del 95%) pues es el más utilizado para el cálculo de intervalos de confianza.

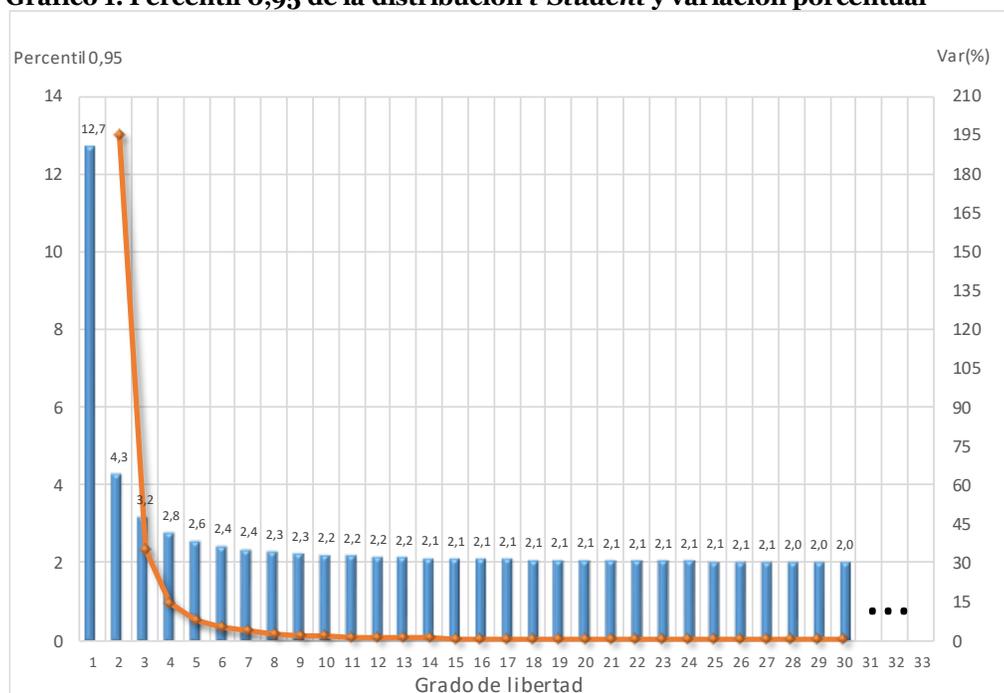
En el *Gráfico 1*, se visualiza el comportamiento de dicho percentil junto con su variación porcentual al aumentar los grados de libertad (se muestran los primeros 30) donde destaca un pronunciado comportamiento de la distribución al considerar pocos grados de libertad. Por ejemplo, cuando presenta uno, dos y tres grados de libertad, los valores asociados del percentil son 12,7; 4,3 y 3,2, respectivamente, lo que equivale a variaciones de 195,3% y 35,2%; para los siguientes grados de libertad se observa una disminución en la variación de los percentiles, estabilizándose a partir de los grados de libertad ocho y nueve.

Por otro lado, como referencia de utilización de umbral para los grados de libertad, se precisa en Parker et al. (2017) la siguiente recomendación: si los grados de libertad para alguna proporción de interés son menos de ocho, entonces la proporción debe ser marcada para revisión, cuyo resultado deriva en publicación o eliminación de la cifra.

¹⁷ En la *sección 2.3.3* se mencionan las consecuencias de adoptar esa primera opción. Para un análisis más detallado consultar Korn & Graubard (1999).

¹⁸ Esta opción contempla el cálculo de los grados de libertad mediante el total de conglomerados con muestra en la subpoblación menos el total de estratos con muestra en la subpoblación. Al igual que en el caso del tamaño muestral, dichos conteos se realizan dependiendo del tipo de estimador. Para proporciones y razones entre 0 y 1, se realiza el conteo sobre las unidades de análisis del denominador; mientras que para el resto de las estimaciones se forman sobre el total de unidades de análisis que nutren la estimación.

Gráfico 1. Percentil 0,95 de la distribución *t-Student* y variación porcentual



Fuente: Elaboración propia. INE, 2020.

Recomendación

En suma, dados estos antecedentes, el umbral de aceptación para este criterio es cuando los grados de libertad asociados a una estimación de interés son 9 o más.

4.1.3. Fracción de muestreo

La fracción de muestreo fue otro de los criterios evaluados en la estructuración de recomendaciones sobre la calidad de las estimaciones. Sin embargo, como ya fue mencionado en la *sección 2.3.4*, en las encuestas de hogares esta fracción de muestreo suele ser baja por tratarse de poblaciones extensas.

De igual manera, la fracción de muestreo va relacionada en términos generales con las características del diseño muestral del estudio, lo que hace que este concepto no aplique para subpoblaciones que no contemplen dominios de diseño¹⁹ o uniones de estos.

¹⁹Kish (1980) identifica tres tipos de subpoblaciones: (i) dominios de diseño, donde la muestra es planificada, diseñada y seleccionada de manera separada; por ejemplo: regiones, urbano-rural, entre otras (por lo general van asociadas a división político territoriales), (ii) clases cruzadas (*cross-classes*), subpoblaciones que atraviesan el diseño muestral a través de los estratos y UPM; por ejemplo: edad, sexo, ocupación, etc. y (iii) subpoblaciones mixtas, que se obtiene al combinar las dos distinciones anteriores.

Las estimaciones publicadas que consideran estos niveles se asocian a fracciones de muestreo bajas, en consecuencia, se opta por prescindir de este criterio, ya que su evaluación involucra cálculos adicionales que no aportarán información relevante. Sin embargo, se insta a que el investigador preste especial atención en la evaluación del tamaño muestral cuando se evidencien altas fracciones de muestreo.

4.2. Evaluación de los criterios relacionados con la precisión

Una vez que se ha validado la suficiencia del tamaño de muestra que da origen a la estimación, debe evaluarse la precisión asociada, cuyo mecanismo depende del tipo de estimador, pudiendo ser un total, una media, una mediana, una razón, una proporción, entre otros.

A efectos de abordar este tema, se consideran dos clases de estimadores en la evaluación de la precisión, (i) la primera clase, compuesta por los estimadores de proporciones y razones entre 0 y 1 y (ii) la segunda clase, formada por el resto de los indicadores.

4.2.1. Criterios de precisión para proporciones y razones entre 0 y 1

Para determinar el método con que se evaluará la precisión de las estimaciones se analizaron los dos criterios empleados actualmente. En este caso se estudió la pertinencia de seguir utilizando el coeficiente de variación como única medida de calidad para los estimadores de proporciones y razones entre 0 y 1.

En ese contexto se examinó el comportamiento del error estándar dado el coeficiente de variación para distintos niveles de la proporción p , considerando los dos siguientes criterios para clasificar a las estimaciones según el coeficiente de variación:

- **Primer criterio.** La estimación de la proporción posee buena calidad si el coeficiente de variación (cv) asociado es menor o igual a 0,15.
- **Segundo criterio.** La estimación de la proporción posee buena calidad si el coeficiente de variación (cv) asociado se encuentra acorde a el siguiente cuadro²⁰.

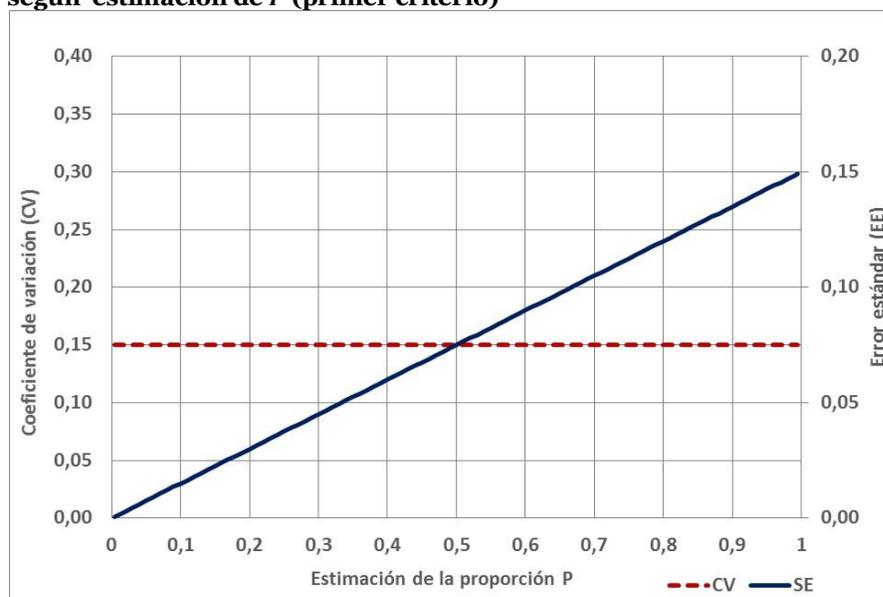
²⁰ Estos umbrales del coeficiente de variación fueron establecidos en la Encuesta Nacional de Caracterización Socioeconómica de 2015, al determinar tamaños muestrales para estimar de la tasa de pobreza por ingresos a nivel comunal, que, para valores cercanos a cero, admitían mayores límites del error relativo, equivalentes al doble de los límites admitidos del coeficiente de variación, considerando un nivel de confianza del 95%, más detalles en MDS (2015).

Cuadro 12. Niveles de cv aceptables

Proporción p	Coefficiente de variación CV
[0,00 ; 0,02)	$\leq 0,400$
[0,02 ; 0,05)	$\leq 0,325$
[0,05 ; 0,10)	$\leq 0,250$
[0,10 ; 1,00]	$\leq 0,180$

Fuente: Elaboración propia. INE, 2020.

Respecto al **primer criterio**, como se puede apreciar en el *Gráfico 2*, al establecer una cota máxima para el coeficiente de variación, y al ser el error estándar (ee) una función lineal de él ($ee(\hat{p}) = CV(\hat{p}) \cdot \hat{p}$), a medida que aumenta la proporción se admiten errores estándares más altos.

Gráfico 2. Máximo coeficiente de variación y error estándar admitido según estimación de P (primer criterio)

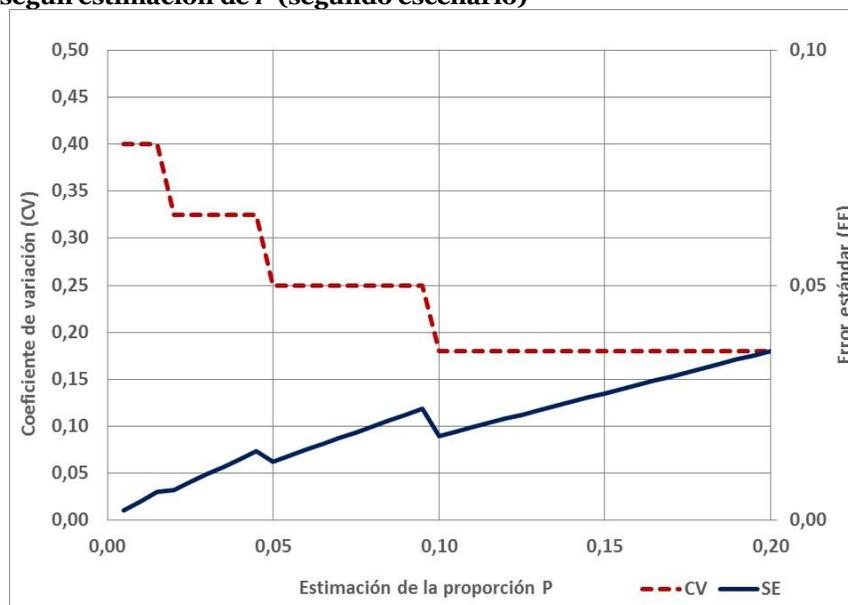
Fuente: Elaboración propia. INE, 2020.

Una limitación de este escenario radica en que para prevalencias bajas se exigen errores estándares muy bajos, lo que no necesariamente implica que el intervalo de confianza generado no sea de utilidad para la toma de decisiones. Por ejemplo, estimaciones de P que se encuentren cerca de 0,05 pueden tener intervalos de confianza de amplitud razonable cuando el coeficiente de variación es superior a 0,15.

Lo anterior sugiere que pudieran establecerse cotas máximas para el cv según los valores que tome p (la estimación de P) lo que origina el **segundo criterio** para la evaluación (ver *Gráfico 3*). Al definir dichas cotas según valores de p , el error estándar sigue una tendencia

lineal creciente con leves decrecimientos en torno a los valores de p en los que cambian las cotas del cv, es decir, en los valores 0,02; 0,05 y 0,10. Esto implica que para algunos valores de p menores se exigen errores estándares mayores, rompiendo la tendencia natural lineal creciente del error estándar a medida que los valores de p aumentan de 0 a 0,50; lo que resulta ser poco razonable.

Gráfico 3. Máximo coeficiente de variación y error estándar admitido según estimación de P (segundo escenario)



Fuente: Elaboración propia. INE, 2020.

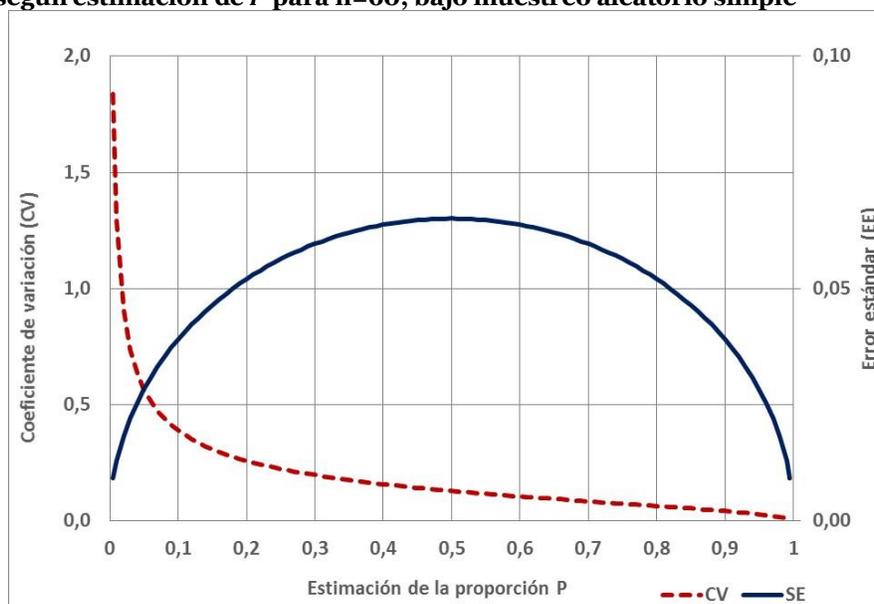
Los análisis de los dos criterios anteriores sugieren que la calidad de la estimación de prevalencias bajas debe ser medida considerando el error estándar y no el coeficiente de variación. Esto va en línea con lo que ponen de relieve Vaillant, Denver & Kreuter (2013) en un ejercicio de cálculo de tamaño muestral para una prevalencia muy baja ($p=0,01$) donde al fijar un cv esperado de 0,05 ($ee=0,0005$) obtienen un tamaño de muestra de 39.600; y que al replantear este mismo ejercicio, pero fijando el error absoluto en 0,005 ($ee=0,0026$, y $cv=0,26$) se obtiene un tamaño de muestra de 1.522, valor que dista notablemente del primer cálculo. Es decir, si para un investigador le resulta tolerable un error absoluto de 0,005 (la mitad del valor de la estimación en 0,01) podría optar por muestrear 1.522 casos en vez de los 39.600.

Otro aspecto para considerar en este análisis es la dicotomía de los fenómenos que se miden a través de los estimadores de proporción. Por ejemplo, detrás de la estimación de la proporción de personas victimizadas subyace la estimación de las no victimizadas, ambas

estimaciones tienen asociado el mismo error estándar, pero no así el coeficiente de variación, que será significativamente mayor para la contraparte de menor prevalencia.

El *Gráfico 4* muestra el comportamiento del error estándar y el coeficiente de variación para varios niveles de p con un tamaño muestral supuesto en 60 unidades, en donde se precisa que, para $p = 0,05$ el error estándar es 0,028 equivalente a un coeficiente de variación de 0,567, sin embargo, el complemento $1-p=0,95$ tiene el mismo nivel de error estándar pero un cv de 0,030. En este sentido, sería ilógico concluir que la estimación de p no tiene una calidad aceptable pero la de $1-p$ sí la tiene.

Gráfico 4. Comportamiento del error estándar y del coeficiente de variación según estimación de P para $n=60$, bajo muestreo aleatorio simple



Fuente: Elaboración propia. INE, 2020.

En consecuencia, la calidad de la estimación de proporciones y razones entre 0 y 1 en función de la dispersión, será evaluada considerando simultáneamente el coeficiente de variación, el error absoluto y la dicotomía de los fenómenos estudiados, así como la incidencia de los umbrales precisados sobre los tamaños muestrales requeridos.

Para esto se propone una metodología que permite evaluar la calidad de la medida del error estándar de acuerdo con funciones matemáticas que resguarden aspectos deseados en las medidas, mencionadas en el párrafo anterior.

4.2.2. Metodología para evaluar la calidad del error estándar

Esta metodología se basa en definir umbrales (en función de p) del error estándar, que permitan clasificar las estimaciones en fiables o no fiables, para distintos niveles de p . Para ello se proponen tres funciones, una lineal, otra logarítmica y una cuadrática.

La metodología que se propone está en línea con las recomendaciones de Eurostat (2013) que exhorta a utilizar el error estándar y otras medidas expresadas en términos absolutos para medir la precisión de proporciones y razones. En contraposición, exponen que los coeficientes de variación generalmente no se recomiendan para estimar la precisión de proporciones, lo que se debe a que el valor de la proporción tiene un fuerte impacto en el coeficiente de variación, especialmente cuando es baja, y además porque en estos casos los coeficientes de variación no son simétricos respecto a $p=0,50$, es decir, por ejemplo, el coeficiente de variación para $p=0,10$ no es el mismo que el de su complemento o simétrico $1-p=0,90$, sin embargo, el error estándar es el mismo.

4.2.2.1. Función lineal propuesta

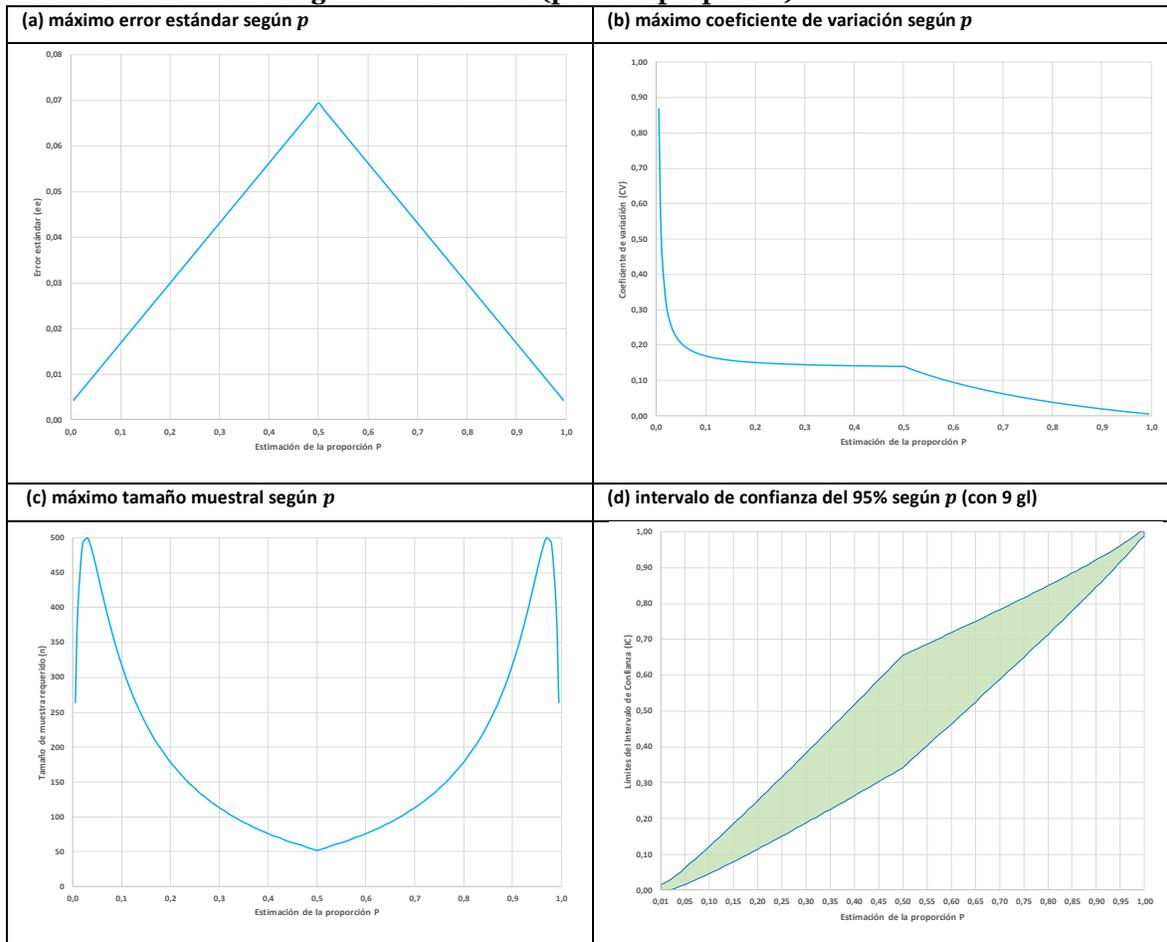
Esta propuesta emplea una función lineal que delimita los valores aceptables para el error estándar en función de los valores de p , por lo que contempla límites de aceptación para el ee que crecen gradualmente a medida que crece la estimación de p . A su vez, a medida que decrece p también decrece el límite de aceptación para el ee , lo que evita los saltos injustificables presentados en el *segundo criterio* evaluado previamente (ver *Cuadro 12* y *Gráfico 3*). La función queda definida como sigue:

$$\text{Máximo } ee \text{ tolerable} = \begin{cases} \frac{5}{38} \cdot p + \frac{7}{1900} & ; \quad 0 < p \leq 0,50 \\ \frac{5}{38} \cdot (1 - p) + \frac{7}{1900} & ; \quad 0,50 < p < 1 \end{cases}$$

Esta función lineal fue determinada mediante la fijación de dos puntos, a saber: (i) el primer punto considera un valor de $p = 0,01$, para el cual se estableció un error estándar de 0,005 (lo que origina un cv de 0,50) y (ii) el segundo punto considera un valor de $p = 0,20$ con error estándar 0,03 (lo que ocasiona un cv de 0,15). Ambos puntos representan límites para la dispersión considerados aceptables para las estimaciones realizadas en las encuestas del INE. Además, esta función también contempla simetría en todo el intervalo $[0 ; 1]$.

El *Gráfico 5* muestra el comportamiento de los umbrales de tolerancia tanto del error estándar como del coeficiente de variación para esta primera función propuesta, así como para el tamaño muestral e intervalo de confianza. Ahí destaca que, aunque se logra fijar umbrales simétricos para el ee , lo que deriva en un comportamiento aceptable para el cv (*Gráfico 5: (a)* y *(b)*) también ocasiona un comportamiento para el tamaño muestral poco aceptable²¹ (*Gráfico 5: (c)*) ya que para prevalencias extremas existen puntos a partir de los cuales ($p = 0,03$ y $p = 0,97$ donde se requiere en ambos casos $n = 500$) se empieza a exigir menos tamaño muestral, hallazgo que va en detrimento de esta propuesta.

Gráfico 5. Máximo error estándar, coeficiente de variación, tamaño muestral requerido e intervalo de confianza según función lineal (primera propuesta)



Fuente: Elaboración propia. INE, 2020.

En cuanto al intervalo de confianza (*Gráfico 5: (d)*) este tiene un comportamiento aceptable debido a que el máximo se alcanza en $p = 0,50$ como todo estimador tipo proporción o razón

²¹ Se espera que a medida que la prevalencia de una característica sea más baja se requiera mayor tamaño muestral para obtener determinado nivel de precisión.

entre 0 y 1, por lo que la amplitud de este intervalo resulta ser simétrica respecto a este máximo (con error estándar máximo tolerable de 0,07).

4.2.2.2. Función logarítmica propuesta

Otra función que presenta propiedades interesantes y que logra mitigar la falencia encontrada en el tamaño muestral de la primera propuesta, y que además ha sido empleada en Barnett-Walker (2003) y Hornik, et al. (2002) es la conocida como cv logarítmico, que para este caso²², queda definida de la siguiente forma.

$$CV[-\log(p)] = \begin{cases} \frac{ee(p)}{p \cdot \log\left(\frac{1}{p}\right)} > 0,1862 & ; 0 < p \leq 0,50 \\ \frac{ee(p)}{(1-p) \cdot \log\left(\frac{1}{1-p}\right)} > 0,1862 & ; 0,50 < p < 1 \end{cases}$$

Despejando el error estándar de la expresión anterior, se determina el error estándar máximo tolerable, dado en la siguiente expresión:

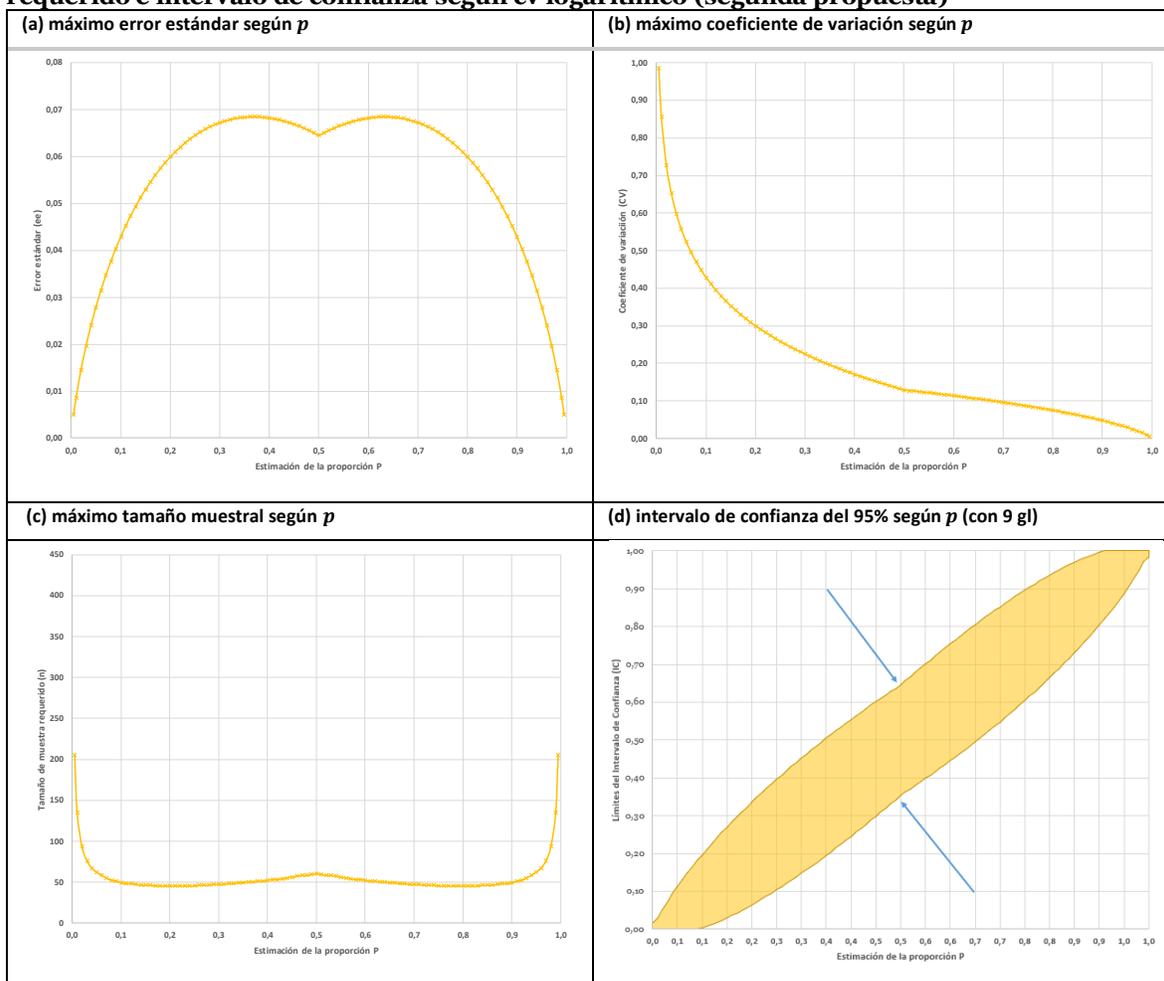
$$\text{Máximo } ee \text{ tolerable} = \begin{cases} 0,1862 \cdot p \cdot \log\left(\frac{1}{p}\right) & ; 0 < p \leq 0,50 \\ 0,1862 \cdot (1-p) \cdot \log\left(\frac{1}{1-p}\right) & ; 0,50 < p < 1 \end{cases}$$

Esta función logra superar el desacierto de la primera propuesta al exigir mayores tamaños muestrales a fenómenos con prevalencias extremas (*Gráfico 6: (c)*). Sin embargo, aunque logra calcar muy bien el comportamiento esperado del cv (*Gráfico 6: (b)*) no ocurre lo mismo con el comportamiento del *ee* entorno a las prevalencias cercanas a 0,50 (*Gráfico 6: (a)*) ya que el umbral resulta ser más exigente respecto de lo que pasa propiamente en 0,50, lo que resulta ser poco comprensible al referirnos a estimaciones tipo proporción o razón entre 0 y 1 y que ciertamente tiene también sus implicancias sobre el tamaño muestral y sobre los intervalos de confianza para este entorno.

²² El valor 0,1862 en la última función del máximo *ee* tolerable, constituye el valor que toma esta función evaluada en $p = 0,50$, en el contexto de un muestreo aleatorio simple y considerando un tamaño muestral de $n = 60$ unidades. Similar ejercicio se encuentra en Hornik, et al. (2002) y Walker (2003) utilizando $n = 41$ y $n = 68$, respectivamente.

Como los límites del intervalo de confianza (*Gráfico 6: (d)*) dependen del error absoluto, y estos, a su vez se relacionan en forma directa con el error estándar, la mayor amplitud del intervalo de confianza no se da en $p = 0,50$ como se esperaría, sino que se presenta para los valores aproximados de $p = 0,37$ y $0,63$, alcanzando un error estándar máximo de $0,069$ (en $p = 0,50$ el error estándar que se alcanza es de $0,065$).

Gráfico 6. Máximo error estándar, coeficiente de variación, tamaño muestral requerido e intervalo de confianza según cv logarítmico (segunda propuesta)



Fuente: Elaboración propia. INE, 2020.

Puestas en evidencia las falencias de las dos primeras propuestas y con el propósito de superar estas limitantes, el equipo técnico del INE acogió la tarea de identificar funciones que logren un mejor acercamiento al comportamiento esperado de los indicadores evaluados, de donde emerge la siguiente propuesta²³.

²³ También se precisó otra función cuadrática que superaba las falencias observadas en el cv logarítmico, que también fue empleada para el *benchmark* realizado, cuyo detalle se encuentra en el *Anexo 8.1*.

4.2.2.3. Función cuadrática propuesta

Una función señalada como plausible por ajustarse al comportamiento esperado para los indicadores evaluados es la siguiente función cuadrática.

$$\text{Máximo } ee \text{ tolerable} = \begin{cases} \frac{\sqrt[3]{p^2}}{9} & ; \quad 0 < p \leq 0,50 \\ \frac{\sqrt[3]{(1-p)^2}}{9} & ; \quad 0,50 < p < 1 \end{cases}$$

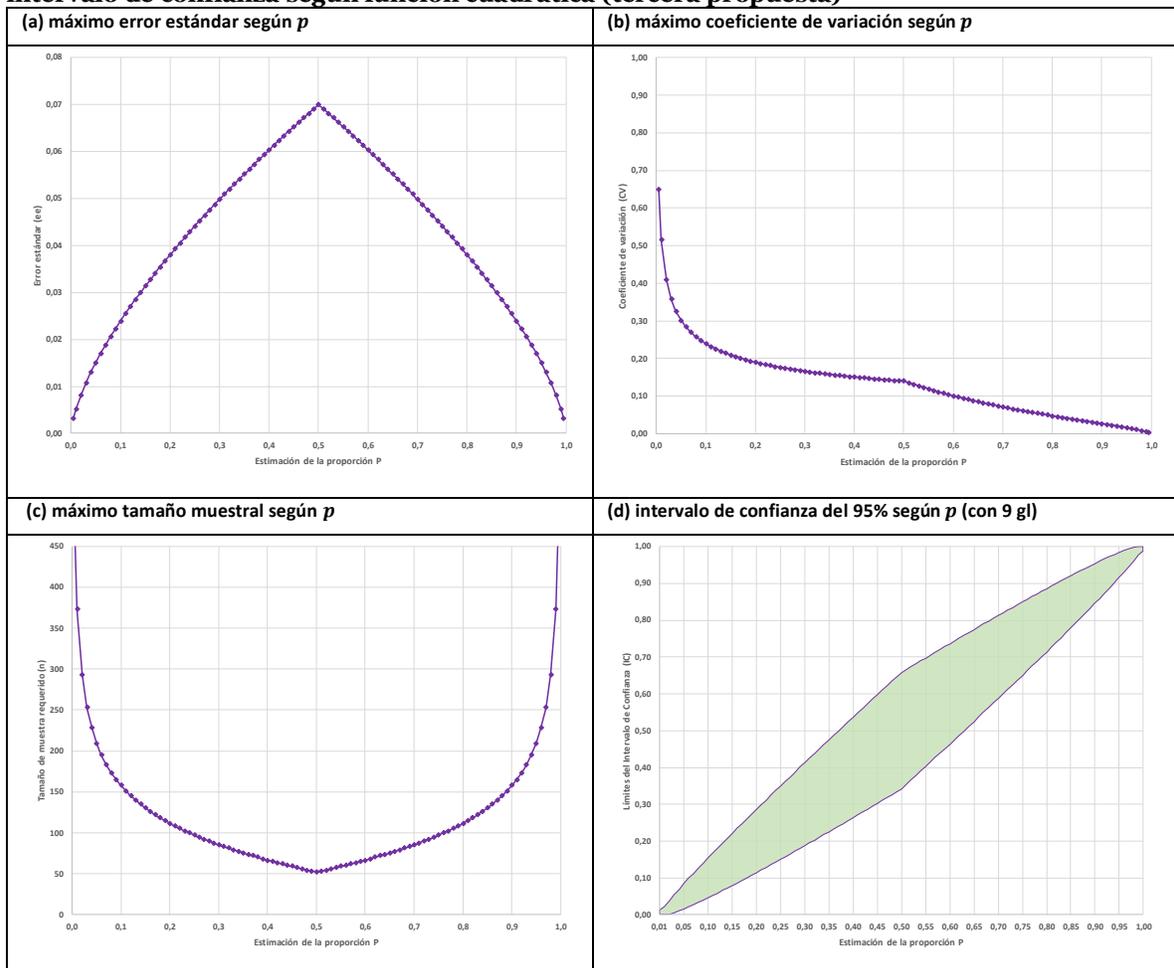
Como se observa en el *Gráfico 7*, esta función emula el comportamiento esperado tanto para el error estándar como para el coeficiente de variación (*Gráfico 7*: (a) y (b)) logrando gradualmente exigir mayores *ee* a medida que la prevalencia crece, manteniendo la simetría de esta medida respecto a $p = 0,50$. Esta función cuadrática genera errores estándares simétricos y, por tanto, intervalos de confianza simétricos (*Gráfico 7*: (d)) alcanzando el máximo error estándar (0,065) en $p = 0,50$ como se espera para este tipo de estimador.

A diferencia de la función logarítmica, para prevalencias extremas poco frecuentes, por ejemplo, para $p=0,005$, el coeficiente de variación que se alcanza es cercano a 0,7, mientras que, para este mismo nivel de la prevalencia, con la propuesta anterior se alcanza un *cv* cercano a 1,0.

Adicionalmente, esta función resalta por ser más parsimoniosa, aspecto importante a considerar para su implementación, además de exigir mayores tamaños muestrales para prevalencias extremas alrededor de 500 (*Gráfico 7*: (c)) lo que, sin lugar a duda, genera un margen de resguardo al analizar prevalencias poco frecuentes.

Dado el comportamiento excepcional de esta función se le confiere un estatus superior de preferencias respecto a las propuestas anteriores, ya que logra superar las falencias identificadas en las mismas.

Gráfico 7. Máximo error estándar, coeficiente de variación, tamaño muestral requerido e intervalo de confianza según función cuadrática (tercera propuesta)



Fuente: Elaboración propia. INE, 2019.

Recomendación

De las tres funciones para la evaluación del error estándar, se recomienda la función cuadrática anterior para delimitar el umbral o máximo error estándar tolerable de tal modo que una estimación de tipo proporción o razón entre 0 y 1 pueda ser catalogada como *fiable*.

En caso de que el error estándar de la estimación supere el umbral de la función cuadrática, la estimación debe ser catalogada como *poco fiable* y amerita una nota de advertencia en la publicación. Se recomienda en estos casos, siempre que sea posible de acuerdo con los objetivos del estudio, agrupar categorías siempre y cuando tenga sentido la nueva categoría colapsada, para de este modo, lograr disminuir el error estándar por debajo del umbral establecido.

4.2.3. Criterios de precisión para el resto de las estimaciones

Se producen varias divergencias entre encuestas, incluso a nivel de ONE de los distintos países, para el umbral máximo permitido del coeficiente de variación como medida de precisión o dispersión de las estimaciones que se publican.

Al ser el coeficiente de variación el cociente entre cantidades expresadas en las mismas unidades (el cociente entre el error estándar del estimador y el mismo estimador) este indicador está libre de unidades, lo que permite hacer comparaciones entre variables o indicadores de distinta índole, como comparación de variables expresadas en diferentes unidades y/o diferentes escalas.

Nótese que el coeficiente de variación se indetermina para valores del estimador cercanos a cero, tal como ocurre con el coeficiente de variación para proporciones.

Para variables o indicadores con recorrido en intervalos reales distintos al rango $[0; 1]$, es decir, sobre variables que no sean proporciones ni razones entre 0 y 1, la medida de precisión apropiada para la fijación de criterios a utilizar es el coeficiente de variación.

Existen otro tipo de variables o indicadores que tienen un rango y/o recorrido acotado y que toman valores en el intervalo real $[a; b]$, con $a < b$ y $b > 1$. Por ejemplo, el rango de variación de la variable tiempo promedio dedicado al cuidado de niños en la ENUT está acotado teóricamente a 24 horas. Este tipo de variables serán tratadas igualmente que las variables que teóricamente no tienen límites superiores exactos conocidos, como el nivel de ingresos o gastos.

Recomendación

La recomendación general para estas variables o indicadores es que el coeficiente de variación de la estimación no sea superior a 0,15. Cuando se supere este umbral, se recomienda evaluar la posibilidad de agrupar categorías con la intención de aumentar la precisión de la estimación, previendo que la nueva categoría tenga sentido interpretativo para el análisis e interés del estudio.

De igual modo, si para el resto de las estimaciones no consideradas en el párrafo previo, no se logra una agrupación idónea de las categorías de manera que no superen el umbral definido, se recomienda alertar en la publicación. En este punto se identifican dos tipos de alerta dependiendo del tramo donde se ubique el cv: en caso que el cv se encuentre entre 0,15 y 0,30, se consigna que la estimación es poco fiable, por lo que debe ser empleada con

precaución, ya que podría llevar a conclusiones poco acertadas de las características de la población, mientras que, si es mayor a 0,30, se declara la estimación como *no fiable*, y no se recomienda su uso.

4.3. Flujograma para la publicación de estadísticas

En el desarrollo de este documento se pone de relieve que la precisión de las estimaciones no solo depende del coeficiente de variación como se ha abordado mayormente por las ONE, sino que, también es necesario tomar en cuenta las interacciones de otros criterios como el tamaño muestral, los grados de libertad, el tipo de estimador, entre otros.

En atención a esto, se propone, en primera instancia, evaluar la calidad de las estimaciones, para luego pasar a evaluar la calidad de los tabulados.

4.3.1. Flujograma o criterios de calidad para evaluación de las estimaciones

Se entrega una propuesta del flujograma (*Ilustración 1*) que presenta además de los criterios a evaluar, un orden lógico en que deben evaluarse, lo que permite identificar cada estimación como fiable, poco fiable o no fiable. Esta categorización será tomada en cuenta en una etapa posterior para evaluar si el tabulado que integran se adecúa para su publicación oficial.

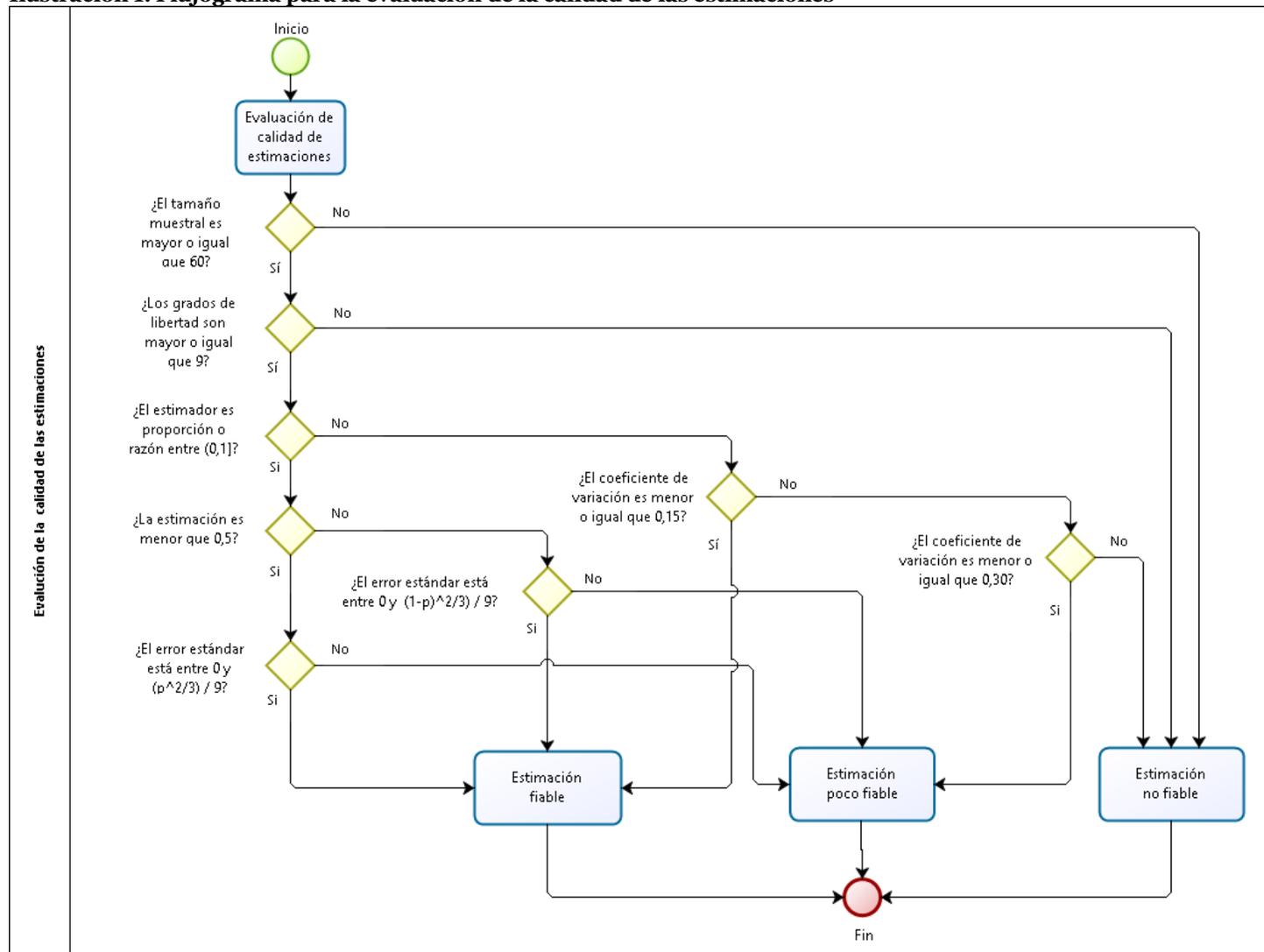
En principio se debe asegurar que el tamaño de muestra y grados de libertad sean suficientes para realizar la estimación, esto es, número de unidades de análisis y número de conglomerados. En cuanto al tamaño muestral o total de unidades de análisis, se precisa que el mínimo sea 60, mientras que el número de conglomerados se podrá validar a través de los grados de libertad, que finalmente son determinantes en la amplitud del intervalo de confianza, para lo que se considera un mínimo de 9 grados de libertad.

En relación con lo anterior, se recomienda que las estimaciones generadas con pocas observaciones muestrales, en cuanto a tamaño muestral o grados de libertad, sean catalogadas como *no fiables*. En caso contrario, si el tamaño muestral y los grados de libertad son suficientes, debe evaluarse la precisión de la estimación, para lo que se considera el tipo de estimación y su coeficiente de variación o error estándar.

Como se desprende de los análisis anteriores, el abordaje de la precisión depende del tipo de estimación. Por tanto, la precisión de las estimaciones de proporciones o razones entre 0 y 1, son medidas a través del error estándar, cuyo máximo tolerable se determinará por medio de la función cuadrática (*sección 4.2.2.3*) recomendada, la cual depende de la estimación puntual. Por otra parte, la precisión de las estimaciones diferentes de proporciones o razones entre 0 y 1, son medidas a través del coeficiente de variación, cuyo máximo tolerable se precisa en 0,15.

Finalmente, si la estimación es evaluada con el error estándar y no cumple con la precisión deseada, se recomienda publicarla con una advertencia que indique que la estimación es *poco fiable* y, en caso contrario, publicar sin observaciones adicionales. Para el resto de las estimaciones que se evalúan con el coeficiente de variación y no cumplen con la precisión deseada, si el cv es mayor que 15% y menor o igual que 30%, se recomienda publicarla como poco fiable, mientras que si el cv sobrepasa el 30% debe identificarse como no fiable y se recomienda no publicarla.

Ilustración 1. Flujograma para la evaluación de la calidad de las estimaciones



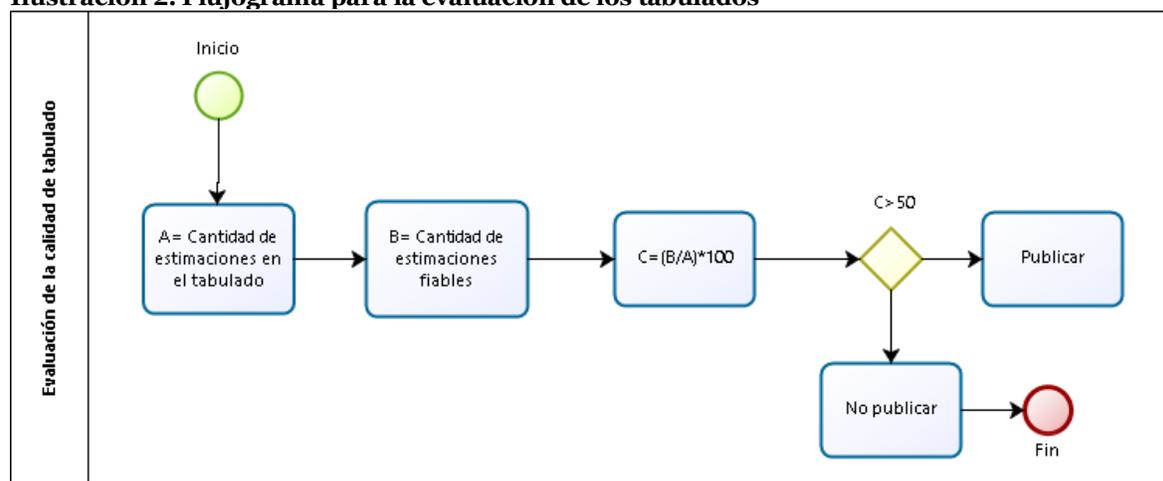
Fuente: Elaboración propia. INE, 2020.

4.3.2. Flujograma o criterios de calidad para evaluación de tabulados

Una vez que se han evaluado todas las estimaciones que componen un tabulado, corresponde evaluar la calidad este en su conjunto con el fin de determinar si es publicable o no publicable. Se considera que un tabulado es publicable si más de 50% de las estimaciones presentes en él fueron clasificadas como estimaciones fiables. En cualquier otro caso, el tabulado no debe publicarse.

La *Ilustración 2* muestra el flujograma o secuencia de criterios a aplicar para la evaluación de los tabulados.

Ilustración 2. Flujograma para la evaluación de los tabulados



Fuente: Elaboración propia. INE, 2020.

5. Evaluación de impacto de aplicación del flujograma

La evaluación del impacto de la aplicación del flujograma en las estimaciones en celdas, se realiza mediante el conteo de las que se “caen” o no cumplen los criterios de calidad establecidos, ya sea de tamaño muestral, grados de libertad y coeficiente de variación o error estándar, de acuerdo con la secuencia presentada en el flujograma (*Ilustración 1*). Posteriormente, la evaluación del impacto en tabulados se realiza mediante el conteo de tabulados que se caen por no cumplir con el criterio que más del 50% de sus estimaciones sean fiables (*Ilustración 2*).

Presentados los flujogramas para evaluar la calidad de las estimaciones y de los tabulados, se prosigue a evaluar el impacto de su implementación en las siguientes encuestas:

1. Encuesta Nacional de Empleo (ENE)
2. Encuesta Suplementaria de Ingresos (ESI)
3. Encuesta de Presupuestos Familiares (EPF)
4. Encuestas Nacional sobre Uso del Tiempo (ENUT)
5. Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC)

Asimismo, se realiza una evaluación de la información proveída por SEEG sobre las encuestas:

1. Encuesta Nacional Urbana de Seguridad Ciudadana (SEEG-ENUSC)
2. Encuesta Suplementaria de Ingresos (SEEG-ESI)
3. Encuestas Nacional sobre Uso del Tiempo (SEEG-ENUT)
4. Encuesta Microemprendimiento (SEEG-EME)

Adicionalmente, para cada una de las encuestas mencionadas, se evaluarán los impactos en las tres funciones²⁴ propuestas que emulan el comportamiento del error estándar (detalladas en la *sección 4.2.2*). Las funciones para evaluar el error estándar son parte de las secuencias propuestas para medir el impacto tanto en celdas como tabulados a publicar (*Cuadro 13*).

El conteo se hace en forma secuencial, en el sentido que primero se evalúa cuántas celdas se caen por el tamaño muestral; de las celdas que no se caen por tamaño muestral, se cuentan cuántas se caen por grados de libertad y, finalmente, de las celdas que no se caen por tamaño muestral ni por grados de libertad se cuentan cuántas celdas se caen por coeficiente de variación o error estándar. De todas maneras, la evaluación de la medida de precisión de una estimación es única, es decir, no puede realizarse evaluando el coeficiente de variación o error estándar simultáneamente, ya que depende del tipo de estimador (ver *sección 4.2*) y el rango de valores posibles a tomar, se considera solo una medida de precisión, respectivamente.

La diferencia entre las secuencias propuestas de evaluación de las estimaciones radica fundamentalmente en cómo se evalúa el error estándar cuando el tipo de estimación es proporción o razón entre 0 y 1, en que se utilizan las funciones lineal, logarítmica y cuadrática, para la primera, segunda y tercera secuencia de evaluación, en cada caso.

²⁴ De las tres funciones propuestas para la evaluación del error estándar, se recomienda la función cuadrática (*sección 4.2.2.3*) pero se conservan las otras dos (lineal y logarítmica) para fines comparativos.

Cuadro 13. Umbrales de aceptación de los criterios de evaluación de las distintas propuestas

Secuencia de Evaluación	Tamaño Muestral	Grados de Libertad	Coefficiente de Variación	Función utilizada para evaluación del Error Estándar
Primera secuencia propuesta	≥ 60	≥ 9	$\leq 15\%$	Función $f(p)$ lineal: $ee \leq \begin{cases} \frac{5}{38} \cdot p + \frac{7}{1900} & ; 0 < p \leq 0,50 \\ \frac{5}{38} \cdot (1-p) + \frac{7}{1900} & ; 0,50 < p < 1 \end{cases}$
Segunda secuencia propuesta	≥ 60	≥ 9	$\leq 15\%$	Función $f(p)$ logarítmica: $ee \leq \begin{cases} 0,1862 \cdot p \cdot \log\left(\frac{1}{p}\right) & ; 0 < p \leq 0,50 \\ 0,1862 \cdot (1-p) \cdot \log\left(\frac{1}{1-p}\right) & ; 0,50 < p < 1 \end{cases}$
Tercera Secuencia propuesta	≥ 60	≥ 9	$\leq 15\%$	Función $f(p)$ cuadrática: $ee \leq \begin{cases} \frac{\sqrt[3]{p^2}}{9} & ; 0 < p \leq 0,50 \\ \frac{\sqrt[3]{(1-p)^2}}{9} & ; 0,50 < p < 1 \end{cases}$

Fuente: Elaboración propia. INE, 2020.

Como se evalúan todas las celdas de varios tabulados en cada encuesta, y estas estimaciones corresponden a distintos tipos (proporciones, razones entre 0 y 1 o resto de estimaciones) existen celdas que se caen por el coeficiente de variación y otras por el error estándar, por lo que en una misma encuesta van a existir celdas caídas por el coeficiente de variación y otras distintas caídas por el error estándar.

1. **Primera secuencia propuesta.** Esta evaluación considera los umbrales de aceptación presentados en la *Ilustración 1* de *sección 4.3*, es decir:
 - a. Tamaño muestral ≥ 60
 - b. Grados de libertad ≥ 9
 - c. Coeficiente de variación $\leq 15\%$ y/o error estándar $\leq f(p)$ lineal.

2. **Segunda secuencia propuesta.** Se mantienen los criterios de la primera propuesta, a excepción del error estándar, el cual se evalúa utilizando una función logarítmica (error estándar $\leq f(p)$ logarítmica).
3. **Tercera secuencia propuesta.** También se mantienen los criterios descritos en la primera propuesta, salvo que para evaluar el error estándar se emplea la función cuadrática (error estándar $\leq f(p)$ cuadrática).

5.1. Pautas para aplicación del flujograma

Se definen los siguientes conceptos y pautas para medir el impacto de la aplicación de las distintas propuestas sobre las celdas, y después sobre los tabulados.

Cuadro estadístico y tabulado. Los cuadros estadísticos constituyen arreglos ordenados de los datos procesados para facilitar la lectura e interpretación de estos. Por otro lado, se entenderá por tabulado a la expresión gráfica que sintetiza un valor o estimación producto del cruce entre dos o más variables. Para la implementación del flujograma para la evaluación, se define como capacidad máxima de un tabulado, al cruce entre tres variables, una de carácter geográfico y a dos variables temáticas²⁵.

Por consiguiente, un cuadro estadístico puede contener uno o más tabulados (cruces de variables) los cuales a efectos del flujograma deben ser evaluados de forma independiente.

La *Tabla 2* muestra un ejemplo de un cuadro estadístico en donde se ilustra la definición de tabulado adoptada para la evaluación de impacto. Este cuadro contiene en total ocho tabulados, compuestos por diferentes cruces entre categoría ocupacional, sexo y región (región A cualquiera) en que se consideran tres estimaciones: la primera corresponde al “Número de personas microemprendedoras”, la segunda a la “Distribución” (porcentaje por filas) y la tercera a la “Concentración” (porcentaje por columnas).

²⁵ Una variable geográfica es aquella que entrega información desagregada con base en la división político administrativa del país, de acuerdo con los niveles de representatividad de las áreas de estimación de cada estudio en particular. Una variable temática, es aquella que busca describir o caracterizar el fenómeno o evento de interés de una encuesta en particular.

Tabla 2. Personas microempendedoras por sexo según categoría ocupacional

Categoría ocupacional	Número de personas microempendedoras			Distribución (%)			Concentración (%)		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres
Total	53.317	30.278	23.040	100	56,79	43,21	100	100	100
Empleador	10.729	6.736	3.994	100	62,78	37,23	20,12	22,25	17,34
Cuenta propia	42.588	23.542	19.046	100	55,28	44,72	79,88	77,75	82,66

Fuente: Elaboración propia. EME, 2017.

Al enfocarnos en el extracto que corresponde a “Número de personas microempendedoras” se precisan cuatro tabulados, que corresponden a los cruces: (i) región-categoría ocupacional, compuesto por la condición de empleador (10.729) y cuenta propia (42.588); (ii) región-sexo, con hombres (30.278) y mujeres (23.040); (iii) región-sexo-categoría ocupacional (6.736; 3.994; 25.542; 19.046) y (iv) total general de la región (53.317).

Por otro lado, al observar lo referente a la “Distribución” se identifican dos tabulados que son reconocidos por (i) el cruce región-sexo-categoría ocupacional (62,78; 37,23; 55,28; 44,72) y (ii) el cruce región-sexo con el total de hombres (56,79) y mujeres (43,21).

Asimismo, al examinar la “Concentración” se presentan dos tabulados que quedan determinados por: (i) el cruce región-categoría ocupacional, con los totales para empleador (20,12) y cuenta propia (79,88); y (ii) el cruce región-sexo-categoría ocupacional (22,25; 17,34; 77,75; 82,66).

De esta forma, la *Tabla 2* tiene un total de ocho tabulados distintos que generan 21 celdas que serán evaluadas, nueve correspondiente a “Número de personas Microempendedoras”, seis celdas para “Distribución” y seis también para “Concentración”.

Tabulados a evaluar. Dada las restricciones de tiempo y recursos, para cada encuesta, se prioriza el análisis a los indicadores y tabulados esenciales. La selección de los tabulados quedó a cargo de los equipos técnicos de los productos participantes en la mesa de trabajo, junto con sus respectivas jefaturas. El detalle de los indicadores seleccionados según encuesta se encuentra en el *Anexo 8.3*.

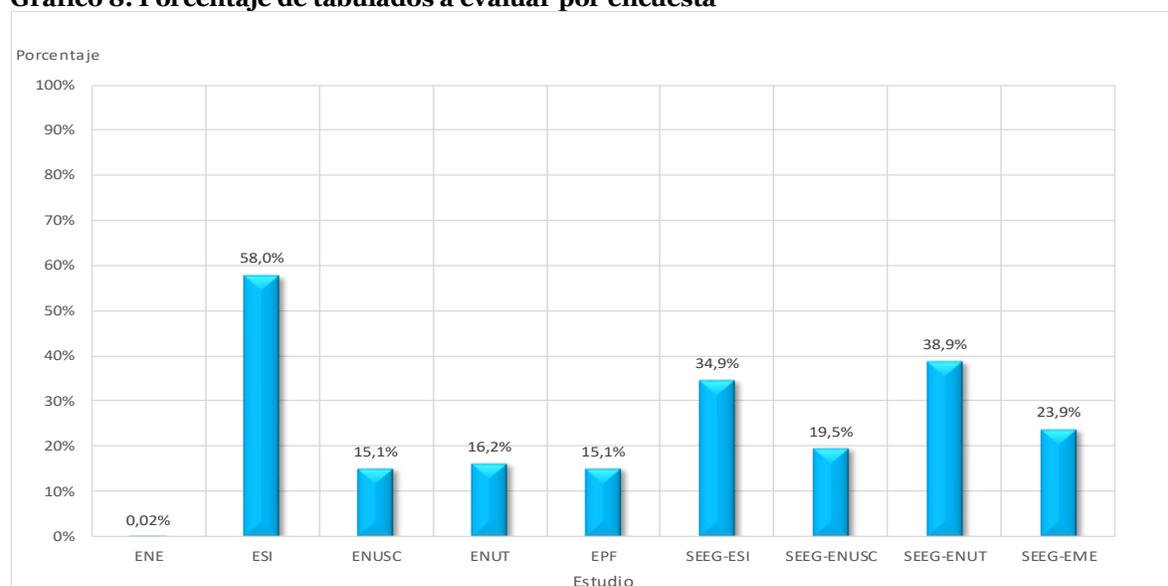
Para la evaluación de las distintas propuestas se utilizan los mismos tabulados y celdas contenidas en ellos. Como se observa en el *Gráfico 8*, en la mayoría de las encuestas se evaluó más del 15% de los tabulados producidos, destacando la ESI, donde se evaluó el 58% de los

tabulados²⁶. En cuanto a las desagregaciones que provee la SEEG, para cada encuesta se evaluó más del 19% de los tabulados que se producen, destacando la ENUT, donde se evaluó el 38,9% y ESI, con 34,9%.

Forma secuencial. Los análisis se realizaron de forma secuencial, es decir, las celdas de los tabulados que no cumplen con los criterios (o se “caen”) en alguna de las secuencias del flujograma, no son consideradas posteriormente.

Factibilidad de publicar un tabulado. Una vez que se han evaluado todas las estimaciones que componen un tabulado, corresponde evaluar la calidad del tabulado en su conjunto con el fin de determinar si es *publicable* o *no publicable*. Se considera que un tabulado es *publicable* si más de 50% de las estimaciones presentes en él fueron clasificadas como *estimaciones fiables* (ver sección 4.3.2). En cualquier otro caso, el tabulado no debe publicarse.

Gráfico 8. Porcentaje de tabulados a evaluar por encuesta



Fuente: Elaboración propia. INE, 2020.

Los resultados de la evaluación serán presentados en dos etapas: evaluación de estimaciones en celdas y evaluación de tabulados. En la primera se evalúan los criterios relacionados con las dimensiones de la muestra lograda (tamaño muestral y grados de libertad) y los criterios vinculados con medidas de precisión (coeficiente de variación o error estándar). En la segunda etapa se evalúa la cantidad de tabulados que se caen por las tres secuencias

²⁶ Para el caso de la ENE, se consideraron en la evaluación de impacto los tabulados que corresponden a la *nota estadística*, considerados fundamentales para el análisis e interpretación de resultados. El resto de tabulados se encuentran en reconfiguración dado el rediseño de la ENE.

propuestas de evaluación de celdas. Adicionalmente se realiza una comparación de los resultados a la luz de las diferentes funciones empleadas para evaluar el error estándar que son precisadas en el *Cuadro 13* anterior.

5.2. Análisis de impacto por estimaciones o celdas

La presente sección muestra los resultados de la evaluación para cada uno de los criterios del flujograma en las encuestas evaluadas. En la *sección 5.2.1* se comparan los resultados de los criterios relacionados con la dimensión de la muestra lograda, es decir, tamaño muestral y grados de libertad, mientras que en la *sección 5.2.2*, se comparan los resultados por los criterios del flujograma que corresponden a medidas de dispersión, es decir, coeficiente de variación y error estándar (para las tres funciones propuestas).

5.2.1. Análisis de criterios respecto a tamaño muestral y grados de libertad

En términos de tamaño muestral se establece un mínimo de 60 unidades, siendo esta la primera secuencia del flujograma. El siguiente criterio de evaluación corresponde a los grados de libertad, donde se establece que estos sean nueve o más.

En todas las encuestas, excepto en la EPF y la ENUSC, el mayor porcentaje de celdas caídas se concentra en estos dos resultados de evaluación. Si bien en las *secciones 5.4 y 5.5* se analizan los resultados de la evaluación por encuesta, en términos generales se puede establecer que hay ciertas características de ellas que permiten explicar las diferencias en el efecto de estos criterios. Por ejemplo, tanto en la ENE como en la ESI las aperturas son a nivel regional o de ciudades, mientras que la EPF solo presenta resultados a nivel nacional y desagregados por macrorregión. En tanto, la ENUT y la ENUSC tienen resultados a nivel nacional y regional.

Respecto al criterio de evaluación del tamaño muestral, la ESI y la ENUT presentan los mayores porcentajes de celdas caídas (38,9% y 31,8%, respectivamente). En el caso de los grados de libertad, la ENUT (así como su análisis desde la SEEG) es la encuesta con más celdas caídas (20,8% y 31,1%, respectivamente) seguida de la SEEG-EME (2,0%). Ninguna de las otras encuestas evaluadas presenta celdas caídas en este criterio del flujograma.

Tabla 3. Porcentaje de celdas caídas para criterios relacionados con el tamaño muestral y grados de libertad respecto del total de celdas

	ENE	ESI	EPF	ENUT	ENUSC
Total celdas evaluadas	243	4.505	63.363	8.640	1.488
% celdas caídas por Tamaño Muestral	7,3	38,9	0,1	31,8	5,4
% celdas caídas por Grados de Libertad	0,0	0,0	0,0	20,8	0,0

Fuente: Elaboración propia. INE, 2020.

Al analizar los resultados de la evaluación a los productos de la SEEG (*Tabla 4*), vemos que tienen una tendencia similar a los análisis de la *Tabla 3*, donde el criterio de tamaño muestral es el que resulta más estricto, especialmente en el caso de la ESI (25,6%) y EME (12,4%).

Tabla 4. Porcentaje de celdas caídas por criterios de tamaño muestral y grados de libertad respecto del total de celdas (SEEG)

	SEEG ENUSC	SEEG ESI	SEEG ENUT	SEEG EME
Total celdas evaluadas	1.145	4.095	216	1.657
% celdas caídas por Tamaño Muestral	2,4	25,6	2,1	12,4
% celdas caídas por Grados de Libertad	0,0	0,0	31,1	2,0

Fuente: Elaboración propia. INE, 2020.

5.2.2. Análisis de criterios vinculados con medidas de precisión

Como se menciona en la *sección 4.2*, la precisión de las estimaciones se evalúa supeditada al tipo de estimador. Las estimaciones del tipo proporción y razón entre 0 y 1 se evalúan a través del error estándar, para el cual se utilizan tres funciones, denominadas respectivamente función lineal, logarítmica y función cuadrática (*Cuadro 13*). El resto de las estimaciones son evaluadas a través del coeficiente de variación, estableciendo el valor 0,15 como punto de corte para que las estimaciones sean consideradas *fiabes*.

Esto significa que no todos los indicadores son evaluados por todos los criterios, pues depende del tipo de estimador que se incluyó en el flujograma. En ese sentido, en la ENUSC no se aplica el criterio de coeficiente de variación debido a que todos sus indicadores corresponden a proporciones que son evaluados con el error estándar. No obstante, al realizar la evaluación a los indicadores de la ENE, ESI, EPF y ENUT se consideran indicadores de medias (o totales) que sí se evalúan mediante el cv, al igual que en los estudios que realiza la SEEG en las encuestas ENUSC, ESI, ENUT y EME.

Al analizar las celdas a través del coeficiente de variación, se tiene (*Tabla 5*) que la EPF es el producto con mayor proporción de celdas caídas por este criterio, con 29,6%, mientras que la ESI es el que presenta mayor proporción de celdas caídas para el análisis de SEEG con 4,5% (*Tabla 6*).

Tabla 5. Porcentaje de celdas caídas por criterios de medidas de dispersión respecto del total de celdas

	ENE	ESI	EPF	ENUT	ENUSC
Total celdas evaluadas	243	4.505	63.363	8.640	1.488
% celdas caídas por Coeficiente de Variación	1,9	4,2	29,6	1,0	-

Fuente: Elaboración propia. INE, 2020.

Tabla 6. Porcentaje de celdas caídas por criterios de medidas de dispersión respecto del total de celdas (SEEG)

	SEEG ENUSC	SEEG ESI	SEEG ENUT	SEEG EME
Total celdas evaluadas	1.145	4.095	216	1.657
% celdas caídas por Coeficiente de Variación	1,8	4,5	0,0	3,0

Fuente: Elaboración propia. INE, 2020.

Respecto al análisis de impacto en relación con el error estándar, que se muestra en *Tabla 7* y *Tabla 8*, se observa una tendencia clara sobre cuál de las propuestas produce más o menos caídas de celdas, en donde la función lineal (propuesta 1) es la más estricta, seguida de la función cuadrática (propuesta 3) y luego la función logarítmica (propuesta 2). Por ejemplo, para la ENUSC, la propuesta 1 produce mayor caída de celdas (43,4%) seguida de la propuesta 3 (9,9%) y finalmente la propuesta 2 (5,2%).

También se observa que, para algunas encuestas se producen diferencias notables entre las distintas propuestas. Esto se evidencia especialmente en la ENUSC y la ENUT, mientras que, en casos como la ENE, ESI y EPF, la variación entre las tres propuestas es casi inexistente.

Tabla 7. Porcentaje de celdas caídas por evaluación del error estándar respecto del total de celdas

	ENE	ESI	EPF	ENUT	ENUSC
Total celdas evaluadas	243	4.505	63.363	8.640	1.488
% celdas caídas por propuesta 1	3,8	1,2	0,2	10,5	43,4
% celdas caídas por propuesta 2	0,0	0,1	0,1	1,6	5,2
% celdas caídas por propuesta 3	0,0	0,6	0,2	8,0	9,9

Fuente: Elaboración propia. INE, 2020.

Entre las encuestas evaluadas por la SEEG (Tabla 8) se observa que la EME es la que presenta mayor impacto en celdas caídas con la propuesta 1 (11,9%) y propuesta 2 (5,2%) mientras que con la propuesta 3 no hay impacto en celdas caídas. En segundo orden de impacto, la ENUSC presenta un 3,2% de celdas caídas con la propuesta 1, bajando levemente a 2,7% con la propuesta 2.

Tabla 8. Porcentaje de celdas caídas por evaluación del error estándar respecto del total de celdas (SEEG)

	SEEG ENUSC	SEEG ESI	SEEG ENUT	SEEG EME
Total celdas evaluadas	1.145	4.095	216	1.657
% celdas caídas por propuesta 1	3,2	3,0	0,5	11,9
% celdas caídas por propuesta 2	2,7	0,1	0,0	5,2
% celdas caídas por propuesta 3	0,0	0,0	0,0	0,0

Fuente: Elaboración propia. INE, 2020.

5.3. Análisis de impacto por tabulados

De acuerdo con lo expuesto en las secciones anteriores, la cantidad de celdas caídas por los distintos criterios varía en cada encuesta analizada. Ahora bien, la cantidad de celdas que se caen por encuesta puede traducirse o no en la caída de un tabulado²⁷. Como se menciona al inicio de este capítulo, para efectos de esta evaluación, se considerará que un tabulado amerita ser publicado cuando más del 50% de sus celdas son catalogadas *fiables*, es decir, que se encuentre en los umbrales de aceptación al evaluar cada uno de los criterios del flujograma.

²⁷ La caída de un tabulado se produce cuando más del 50% de sus celdas que lo componen se caen por cualquiera de los criterios de calidad establecidos, es decir, por tamaño muestral, grados de libertad o por las medidas de dispersión, específicamente cuando las estimaciones resultan catalogadas como poco fiables o no fiables.

Dado que en la *sección 5.2.2* se realiza una evaluación de tres funciones para el error estándar, la cantidad de tabulados caídos varía de acuerdo con cada una de ellas. La ENUT es la que tiene mayor porcentaje de tabulados con más del 50% de celdas caídas para las tres propuestas. No obstante, existen diferencias entre las propuestas, pero manteniendo la tendencia, en que la propuesta 1 es la más restrictiva (35,1%) seguida de la propuesta 3 con 32,7% y finalmente la menos restrictiva, la propuesta 2, con 25,5% (*Tabla 9*).

La ENUSC es la que presenta una mayor variación entre propuestas: en la 1, un 16,1% de los tabulados puede ser publicado y, en las demás propuestas, un 3,6%. La ENE (3,3%) la ESI (6,8%) y la EPF (3,9%) son los productos con un menor porcentaje de tabulados no publicables, y en ninguna de ellas hay variaciones entre las distintas propuestas.

Tabla 9. Tabulados caídos por encuesta según propuesta de función para el error estándar

	ENE	ESI	EPF	ENUT	ENUSC
Total de tabulados evaluados	30	116	102	208	56
% tabulados caídos por propuesta 1	3,3	6,8	3,9	35,1	16,1
% tabulados caídos por propuesta 2	3,3	6,8	3,9	25,5	3,6
% tabulados caídos por propuesta 3	3,3	6,8	3,9	32,7	3,6

Fuente: Elaboración propia. INE, 2020.

En relación con las encuestas de la SEEG, la que tiene un mayor porcentaje de tabulados caídos es la ENUT (3,6%) en la propuesta 1 (*Tabla 10*) mientras que las propuestas 2 y 3 no presentan caídas para ninguno de las encuestas evaluadas.

Tabla 10. Tabulados caídos por encuesta de SEEG según función para error estándar

	SEEG ENUSC	SEEG ESI	SEEG ENUT	SEEG EME
Total tabulados evaluados	207	462	84	302
% tabulados caídos por propuesta 1	0,0	0,9	3,6	0,6
% tabulados caídos por propuesta 2	0,0	0,0	0,0	0,0
% tabulados caídos por propuesta 3	0,0	0,0	0,0	0,0

Fuente: Elaboración propia. INE, 2020.

5.4. Análisis de impacto por encuestas

A continuación, se especifica la evaluación de impacto al aplicar el flujograma para evaluar la calidad de las estimaciones para cada una de las encuestas:

5.4.1. Evaluación de aplicación del flujograma en la ENE

En la ENE la evaluación fue realizada sobre el 0,02% de los tabulados producidos²⁸, para el periodo OND 2019²⁹, que abarca seis indicadores y 30 tabulados que derivan en 243 celdas. En general, se contempla la desagregación geográfica nacional, regional, provincial y por ciudad (*Anexo 8.3, Tabla 38*).

Al observar las celdas caídas por criterio de evaluación, se precisa que la mayor parte de los indicadores se ve afectado por el tamaño muestral (7,3%) seguido por la propuesta 1 del error estándar (3,8%) que resulta ser la única que provoca caídas según propuestas (*Tabla 11*).

Tabla 11. Porcentaje de celdas caídas según criterio del flujograma, OND 2019

N° indicador	Tamaño muestral	Grados de libertad	Coeficiente de variación	Error estándar		
				1° Propuesta	2° Propuesta	3° Propuesta
Total	7,3	0,0	1,9	3,8	0,0	0,0
1	0,0	0,0	0,0	0,0	0,0	0,0
2	0,0	0,0	0,0	30,0	0,0	0,0
3	0,0	0,0	0,0	20,0	0,0	0,0
4	15,1	0,0	4,2	0,0	0,0	0,0
5	0,0	0,0	0,0	0,0	0,0	0,0
6	0,0	0,0	0,0	0,0	0,0	0,0

Fuente: Elaboración propia sobre la base de datos provenientes de la ENE, período OND 2019.

Al observar el porcentaje total de celdas caídas por indicador de forma agregada³⁰ (*Tabla 12*) la primera propuesta muestra 13,0% de celdas caídas, mientras que las propuestas 2 y 3 muestran 9,2%. Más en detalle, destacan el indicador 2 y 3 con mayor porcentaje de celdas caídas en la propuesta 1 (30% y 20%, respectivamente) y el indicador 4, con igual porcentaje de celdas caídas para todas las propuestas, con 19,3%.

²⁸ Para el caso de la ENE, se consideraron en la evaluación de impacto los tabulados que corresponden a la *nota estadística*, considerados fundamentales para el análisis e interpretación de resultados. El resto de tabulados se encuentran en reconfiguración dado el rediseño de la ENE.

²⁹ Trimestre octubre-noviembre-diciembre (en siglas OND) 2019.

³⁰ De forma agregada significa que se van sumando las celdas caídas por los distintos criterios de evaluación. Por ejemplo, el porcentaje total de celdas que se caen con la propuesta 1 (de tabla 12, con 13%) corresponde a la suma las caídas por tamaño muestral (7,3%) grados de libertad (0,0%) coeficiente de variación (1,9) y error estándar (considerando la función lineal) con 3,8%.

Tabla 12. Porcentaje de celdas caídas por indicador según propuesta, OND 2019

Nº Indicador	% Celdas caídas propuesta 1	% Celdas caídas propuesta 2	% Celdas caídas propuesta 3
Total	13,0	9,2	9,2
1	0,0	0,0	0,0
2	30,0	0,0	0,0
3	20,0	0,0	0,0
4	19,3	19,3	19,3
5	0,0	0,0	0,0
6	0,0	0,0	0,0

Fuente: Elaboración propia sobre la base de datos provenientes de la ENE, período OND 2019.

Finalmente, considerando los 30 tabulados analizados (*Tabla 13*) para todas las propuestas, tan solo uno de los tabulados no podría ser publicado (correspondiente al 3,3%).

Tabla 13. Tabulados caídos en total y porcentaje, OND 2019

	% Tabulados caídos propuesta 1	% Tabulados caídos propuesta 2	% Tabulados caídos propuesta 3
Total tabulados analizados	30	30	30
Total tabulados con más de 50% de celdas caídas	1	1	1
Total tabulados con 50% o menos de celdas caídas	29	29	29
Porcentaje tabulados con más de 50% celdas caídas	3,3%	3,3%	3,3%

Fuente: Elaboración propia sobre la base de datos provenientes de la ENE, período OND 2019.

5.4.2. Evaluación de aplicación del flujograma en la ESI

Para la ESI 2018, se evaluaron 116 tabulados (58% del total de tabulados producidos) asociados a 4.505 celdas. Salvo en el indicador 1 (ingreso medio nominal de la población ocupada) donde además existe interacción por provincia y ciudad, todos los demás indicadores contemplan los cruces nacional y regional (*Anexo 8.3, Tabla 39*).

En la *Tabla 14* se presenta el porcentaje de celdas caídas por los distintos criterios de evaluación del flujograma, donde se puede observar que la mayor cantidad se origina por tamaño muestral, con 38,9%, seguida por el coeficiente de variación, con 4,2%. Al analizar por indicador, se observa que el mayor porcentaje de celdas caídas ocurre en el indicador 1, donde un 33,9% se cae por el criterio de tamaño muestral y 4,3%.

En relación con las propuestas de evaluación del error estándar, en términos generales, la primera propuesta es la que presenta mayor caída de celdas (1,2%) seguida de la propuesta 3 (0,6%) y la propuesta 2 (0,1%).

Tabla 14. Porcentaje de celdas caídas según criterio del flujograma, ESI 2018

N° Indicador	Tamaño muestral	Grados de libertad	Coeficiente de Variación	Error estándar		
				1° propuesta	2° propuesta	3° propuesta
Total	38,9	0,0	4,2	1,2	0,1	0,6
1	33,9	0,0	4,3	-	-	-
2	12,9	0,0	0,0	-	-	-
3	29,5	0,0	1,6	-	-	-
4	15,5	0,0	3,5	-	-	-
5	35,5	0,0	-	12,9	1,1	5,9
6	0,0	0,0	-	0,0	0,0	0,0
7	0,0	0,0	-	0,0	0,0	0,0
8	0,0	0,0	7,2	-	-	-

Fuente: Elaboración propia sobre la base de datos provenientes de la ESI 2018.

Al evaluar la cantidad de celdas caídas en términos agregados por indicador (*Tabla 15*) se observa que con la propuesta 1 se ocasiona mayor impacto, con 44,3% de celdas caídas, seguida por la propuesta 3 con 43,7% (38,9% + 0,0% + 4,2% + 0,6%, de *Tabla 14*).

Tabla 15. Porcentaje de celdas caídas por indicador, ESI 2018

N° indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
Total	44,3	43,2	43,7
1	38,2	38,2	38,2
2	12,9	12,9	12,9
3	31,1	31,1	31,1
4	19,0	19,0	19,0
5	48,4	36,6	41,4
6	0,0	0,0	0,0
7	0,0	0,0	0,0
8	7,2	7,2	7,2

Fuente: Elaboración propia sobre la base de datos provenientes de la ESI 2018.

También se observa que para cada indicador se producen los mismos porcentajes de celdas caídas independientemente de la propuesta considerada, a excepción del indicador 5 que sigue la tendencia general en que la propuesta 1 es más restrictiva, produciendo el 48,4% de

estimaciones caídas, seguida de la propuesta 3 con 41,4% y la menos restrictiva, la función logarítmica con 36,6%.

Finalmente, el porcentaje de tabulados que tienen más de 50% de celdas caídas (*Tabla 16*) no varía en las distintas propuestas y se mantiene en 5,2%.

Tabla 16. Tabulados caídos en total y porcentaje, ESI 2018

	% tabulados caídos propuesta 1	% tabulados caídos propuesta 2	% tabulados caídos propuesta 3
Total tabulados analizados	116	116	116
Total tabulados con más de 50% de celdas caídas	6	6	6
Total tabulados con 50% o menos de celdas caídas	110	110	110
Porcentaje tabulados con más de 50% celdas caídas	5,2%	5,2%	5,2%

Fuente: Elaboración propia sobre la base de datos provenientes de la ESI 2018.

5.4.3. Evaluación de aplicación del flujograma en la EPF

Se evaluaron 102 tabulados que corresponden aproximadamente al 15,1% del total de tabulados que se producen, referidos a 63.363 celdas. Dichos datos fueron seleccionados considerando las estimaciones más relevantes respecto a los objetivos de la encuesta y que pudiesen presentar mayores problemas de precisión. Dado el nivel de desagregación en que se presentan actualmente los datos de gasto promedio mensual de los hogares, se decide centrar los análisis en estos datos, junto con las estimaciones de participación (proporción) de dicho gasto (*Anexo 8.3, Tabla 40*).

En la *Tabla 17* se presentan las celdas caídas por cada uno de los criterios del flujograma, donde puede observarse que la mayor cantidad de celdas caídas se da por no cumplir el criterio del coeficiente de variación con 29,6%, siendo en el indicador 6 que se produce la mayor caída para este criterio, con 61,8%.

Tabla 17. Porcentaje de celdas caídas según criterio del flujograma, EPF 2017

Indicador	Tamaño muestral	Grados de libertad	Coeficiente de Variación	Error estándar		
				1° propuesta	2° propuesta	3° propuesta
Total	0,1	0,0	29,6	0,2	0,1	0,2
1	15,9	0,0	0,0	-	-	-
2	0,0	0,0	0,0	-	-	-
3	0,0	0,0	-	0,0	0,0	0,0
4	0,0	0,0	11,7	-	-	-
5	0,0	0,0	-	0,5	0,0	0,0
6	0,0	0,0	61,8	-	-	-
7	0,0	0,0	-	0,3	0,2	0,4

Fuente: Elaboración propia sobre la base de datos provenientes de la EPF 2017.

En términos agregados (Tabla 18) no se observan variaciones relevantes de celdas caídas con las tres propuestas de evaluación del error estándar. Se aprecia que para los indicadores 2 y 3 no se producen celdas caídas, mientras que el indicador 6 produce la mayor caída, con igual porcentaje igual a 61,8%.

Tabla 18. Porcentaje de celdas caídas por indicador, EPF 2017

N° indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
Total	29,9	29,8	29,9
1	15,9	15,9	15,9
2	0,0	0,0	0,0
3	0,0	0,0	0,0
4	11,7	11,7	11,7
5	0,5	0,0	0,0
6	61,8	61,8	61,8
7	0,3	0,2	0,4

Fuente: Elaboración propia sobre la base de datos provenientes de la EPF 2017.

Respecto al porcentaje de tabulados que tienen más de 50% de celdas caídas (Tabla 19) volvemos a observar que este no varía con las distintas propuestas, manteniéndose en 3,9%.

Tabla 19. Tabulados caídos en total y porcentaje, EPF 2017

	% tabulados caídos propuesta 1	% tabulados caídos propuesta 2	% tabulados caídos propuesta 3
Total tabulados analizados	102	102	102
Total tabulados con más de 50% de celdas caídas	4	4	4
Total tabulados con 50% o menos de celdas caídas	98	98	98
Porcentaje tabulados con más de 50% celdas caídas	3,9%	3,9%	3,9%

Fuente: Elaboración propia sobre la base de datos provenientes de la EPF 2017.

5.4.4. Evaluación de aplicación del flujograma en la ENUT

Del total de tabulados que se producen, se evaluaron 208 correspondientes a 14 grupos de actividades que se presentan con ambos indicadores principales de la encuesta: tasa de participación (104 tabulados) y tiempo promedio dedicado a la actividad (104 tabulados) donde, además, se apertura por zona geográfica (a nivel nacional y regional). Por tanto, el total de celdas analizadas corresponde a 8.640 (*Anexo 8.3, Tabla 41*).

En la *Tabla 20* se presentan las celdas caídas por criterio de evaluación. Para la ENUT se observa que las celdas caídas se concentran en los criterios del flujograma referidos al tamaño muestral y grados de libertad de la encuesta, donde la mayor cantidad se da por no cumplir el criterio de tamaño muestral (31,8%) seguido de los grados de libertad (20,8%).

Al analizar por indicador, se aprecia que el mayor porcentaje de celdas caídas ocurre en el indicador 12, donde el 73,8% de las celdas se caen por criterio de tamaño muestral, mientras que los indicadores 3, 5, 7, 11 y 13 producen el mayor porcentaje de celdas caídas por el criterio de grados de libertad, con 26,6%.

Con relación a las propuestas de evaluación del error estándar, la segunda propuesta es la que expone una menor cantidad de celdas caídas, presentando grandes diferencias respecto de las otras dos propuestas. Tanto la primera como la tercera propuesta presentan un mayor nivel de exigencia, siendo la primera, en general, la propuesta más estricta. La excepción a lo anterior resulta cuando se trata de estimaciones cercanas a 0 o 1, fenómeno que sucede en el grupo de actividades 1.

Tabla 20. Porcentaje de celdas caídas según criterio del flujograma, ENUT 2015

N° Indicador	Tamaño muestral	Grados de libertad	Coeficiente de Variación	Error estándar		
				1° Propuesta	2° Propuesta	3° Propuesta
Total	31,8	20,8	1,0	10,5	1,6	8,0
1	37,7	15,2	-	6,9	4,6	12,1
2	38,1	14,8	0,0	-	-	-
3	21,7	26,6	-	31,3	3,7	22,5
4	22,8	25,9	0,9	-	-	-
5	21,7	26,6	-	31	3	21,1
6	23,4	25,6	1,6	-	-	-
7	21,7	26,6	-	7,3	2,5	5,7
8	44,5	14,4	4,5	-	-	-
9	37,7	15,2	-	26,7	3,5	17,5
10	42,9	13,8	0	-	-	-
11	21,7	26,6	-	19,8	0,9	12,1
12	73,8	1,6	6,0	-	-	-
13	21,7	26,6	-	22,2	5,1	19,8
14	23,7	24,7	0,0	-	-	-

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUT 2015.

En cuanto a los totales (*Tabla 21*) y en relación con lo anterior, se observa que la segunda propuesta de evaluación es la que presenta menor cantidad de celdas caídas (55,2%).

Tabla 21. Porcentaje de celdas caídas por indicador, ENUT 2015

N° indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
Total	64,1	55,2	61,6
1	59,8	57,5	65,0
2	52,9	52,9	52,9
3	79,6	52,0	70,8
4	49,6	49,6	49,6
5	79,3	51,3	69,4
6	50,6	50,6	50,6
7	55,6	50,8	54,0
8	63,4	63,4	63,4
9	79,6	56,4	70,4
10	56,7	56,7	56,7
11	68,1	49,2	60,4

N° indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
12	81,4	81,4	81,4
13	70,5	53,4	68,1
14	48,4	48,4	48,4

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUT 2015.

Esta segunda propuesta es también la que presenta menor cantidad tabulados que tienen más de 50% de celdas caídas, con 25,5% (Tabla 22).

Tabla 22. Tabulados caídos en total y porcentaje, ENUT 2015

	% tabulados caídos propuesta 1	% tabulados caídos propuesta 2	% tabulados caídos propuesta 3
Total tabulados analizados	208	208	208
Total tabulados con más de 50% de celdas caídas	73	53	68
Total tabulados con 50% o menos de celdas caídas	135	155	140
Porcentaje tabulados con más de 50% celdas caídas	35,1%	25,5%	32,7%

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUT 2015.

5.4.5. Evaluación de aplicación del flujograma en la ENUSC

Para la ENUSC se evaluaron 42 tabulados correspondientes a 11 indicadores principales de la encuesta, lo que se traduce en 1.488 celdas (Anexo 8.3, Tabla 42).

En la Tabla 23 se presentan las celdas caídas por criterio de evaluación. En la ENUSC se observa que los criterios relacionados al tamaño muestral y grados de libertad no tienen impacto relevante. Como se muestra, 5,4% de las celdas se caen por criterio de tamaño muestral, variando significativamente entre los indicadores, donde algunos presentan 0% (indicadores 3, 4, 8, 9 y 10) y otros sobre 30% (indicadores 5 y 7). Ninguno de los indicadores evaluados presenta celdas caídas por grados de libertad.

Con relación a las propuestas de evaluación del error estándar, se observa que la primera propuesta es la que presenta mayor porcentaje de celdas caídas (43,4%) seguida de la tercera (9,9%) y finalmente la segunda, con 5,2%.

Tabla 23. Porcentaje de celdas caídas según criterio del flujograma, ENUSC 2017

N° indicador	Tamaño muestral	Grados de libertad	Coeficiente de variación	Error estándar		
				1° propuesta	2° propuesta	3° propuesta
Total	5,4	0,0	-	43,4	5,2	9,9
1	2,1	0,0	-	37,9	5,4	22,9
2	2,1	0,0	-	55,8	3,3	25,4
3	0,0	0,0	-	28,1	1,0	0,0
4	0,0	0,0	-	68,8	1,0	0,0
5	32,3	0,0	-	31,3	20,8	0,0
6	2,1	0,0	-	38,5	11,5	4,2
7	32,3	0,0	-	29,2	7,3	0,0
8	0,0	0,0	-	64,6	3,1	0,0
9	0,0	0,0	-	63,5	0,0	0,0
10	0,0	0,0	-	53,1	1,0	0,0
11	6,3	0,0	-	61,5	13,5	29,2

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUSC 2017.

En la *Tabla 24*, en el agregado por indicador, se observa que la propuesta 1 de evaluación es la que presenta mayor cantidad de celdas caídas (48,8%) mientras que la propuesta 2, es la que reporta menor impacto, con 10,6%.

Tabla 24. Porcentaje de celdas caídas por indicador, ENUSC 2017

N° indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
Total	48,8	10,6	15,3
1	40,0	7,5	25,0
2	57,9	5,4	27,5
3	28,1	1,0	0,0
4	68,8	1,0	0,0
5	63,6	53,1	32,3
6	40,6	13,6	6,3
7	61,5	39,6	32,3
8	64,6	3,1	0,0
9	63,5	0,0	0,0
10	53,1	1,0	0,0
11	67,8	19,8	35,5

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUSC 2017.

Asimismo, la segunda propuesta junto a la tercera son las que presentan una menor cantidad tabulados (3,6%) que tienen más de 50% de celdas caídas (*Tabla 25*) mientras que la propuesta 1 genera 16,1% de tabulados caídos.

Tabla 25. Tabulados caídos en total y porcentaje, ENUSC 2017

	% tabulados caídos propuesta 1	% tabulados caídos propuesta 2	% tabulados caídos propuesta 3
Total tabulados analizados	56	56	56
Total tabulados con más de 50% de celdas caídas	9	2	2
Total tabulados con 50% o menos de celdas caídas	47	54	54
Porcentaje tabulados con más de 50% celdas caídas	16,1%	3,6%	3,6%

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUSC 2017.

5.5. Desagregaciones de encuestas con enfoque de género (SEEG)

A continuación, se presenta la evaluación de impactos asociados a las desagregaciones con enfoque de género realizadas por SEEG. Para el enfoque de género aplicado a las cuatro encuestas evaluadas se consideran ocho indicadores (*Anexo 8.3*).

5.5.1. Género y Seguridad Ciudadana (SEEG-ENUSC)

El presente análisis se realiza considerando un 19,5% del total de tabulados producidos por la SEEG para esta encuesta.

Como se observa en la *Tabla 26* y con relación al criterio de evaluación de tamaño muestral y grados de libertad, 2,4% de las celdas se caen por tamaño muestral y 0% por grados de libertad, siendo el indicador 8 el que muestra mayor porcentaje de celdas caídas bajo este criterio con 10,5%.

Para el análisis de impacto respecto a las medidas de precisión es necesario recordar que los indicadores considerados en la ENUSC corresponden a proporciones, por tanto, deben ser evaluados por medio del error estándar. Sin embargo, algunos indicadores, empleados para el enfoque de género, están relacionados con totales (total hombres, total mujeres) y medias (tiempo medio que se dedica a ciertas actividades) los que son evaluados con el coeficiente de variación, como es el caso del indicador 6, que presenta el mayor porcentaje de celdas caídas bajo este criterio con 6,7%.

Respecto a la evaluación del error estándar, la primera propuesta muestra 3,2% de sus celdas caídas, mientras la segunda tiene un 2,7% y la tercera un 0%.

Tabla 26. Porcentaje de celdas caídas según criterio del flujograma, SEEG-ENUSC 2017

N° indicador	Tamaño muestral	Grados de libertad	Coeficiente de variación	Error estándar		
				1° propuesta	2° propuesta	3° propuesta
Total	2,4	0,0	1,8	3,2	2,7	0,0
1	0,0	0,0	-	0,0	0,0	0,0
2	0,0	0,0	-	0,7	0,0	0,0
3	0,0	0,0	-	0,0	0,0	0,0
4	0,4	0,0	-	6,3	5,6	0,0
5	0,0	0,0	0,0	-	-	-
6	1,0	0,0	6,7	-	-	-
7	2,5	0,0	2,5	-	-	-
8	10,5	0,0	6,2	-	-	-

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUSC 2017.

Como indica la *Tabla 27*, al comparar las distintas propuestas en el agregado por indicador, los resultados muestran que hay una disminución continua en la cantidad de celdas caídas entre las propuestas 1 a 3 (7,4%, 6,9% y 4,2%) respectivamente.

Tabla 27. Porcentaje de celdas caídas por indicador, SEEG-ENUSC 2017

N° indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
Total	7,4	6,9	4,2
1	0,0	0,0	0,0
2	0,7	0,0	0,0
3	0,0	0,0	0,0
4	6,7	6,0	0,4
5	0,0	0,0	0,0
6	7,7	7,7	7,7
7	5,0	5,0	5,0
8	16,7	16,7	16,7

Fuente: Elaboración propia sobre la base de datos provenientes la ENUSC 2017.

Finalmente, para todas las propuestas, no existen tabulados con más de 50% de celdas caídas (*Tabla 28*).

Tabla 28. Tabulados caídos en total y porcentaje, SEEG-ENUSC 2017

	% tabulados caídos propuesta 1	% tabulados caídos propuesta 2	% tabulados caídos propuesta 3
Total tabulados Analizados	207	207	207
Total tabulados con más de 50% de celdas caídas	0	0	0
Total tabulados con 50% o menos de celdas caídas	207	207	207
Porcentaje tabulados con más de 50% celdas caídas	0%	0%	0%

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUSC 2017.

5.5.2. Género e Ingresos (SEEG-ESI)

En el caso de Género e Ingresos, el análisis considera un total de 462 tabulados y 4.095 celdas.

Como indica la *Tabla 29* en el criterio del tamaño muestral es donde se cae la mayor cantidad de celdas con un total de 25,6%, seguido por coeficiente de variación con 4,5%. Respecto de los indicadores, el número 8 es el que pierde mayor proporción de celdas por criterio de tamaño muestral (37,9%) seguido por el indicador 7 (26,9%). En relación con el criterio del error estándar, la propuesta 1 presenta el mayor porcentaje de celdas caídas (3,0%) siendo el indicador 4, bajo esta propuesta, el que presenta mayor cantidad de celdas caídas (21,2%).

Tabla 29. Porcentaje de celdas caídas según criterio del flujograma, SEEG-ESI 2018

N° indicador	Tamaño muestral	Grados de libertad	Coeficiente de variación	Error estándar		
				1° propuesta	2° propuesta	3° propuesta
Total	25,6	0,0	4,5	3,0	0,1	0,0
1	0,0	0,0	-	0,0	0,0	0,0
2	0,0	0,0	-	4,7	0,4	0,0
3	0,0	0,0	-	0,0	0,0	0,0
4	0,0	0,0	-	21,2	0,4	0,0
5	16,2	0,0	5,4	-	-	-
6	24,7	0,0	3,4	-	-	-
7	26,9	0,0	6,3	-	-	-
8	37,9	0,0	6,7	-	-	-

Fuente: Elaboración propia sobre la base de datos provenientes de la ESI 2018.

Al comparar las distintas propuestas en términos agregados por indicador (*Tabla 30*) los resultados muestran que en promedio, la propuesta 1 es la que produce mayor porcentaje de

celdas caídas con 33,1%, siendo los indicadores 2 y 4 los que producen la diferencia en cuanto a celdas caídas respecto a las otras dos propuestas.

Tabla 30. Porcentaje de celdas caídas por indicador, SEEG-ESI 2018

Nº indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
Total	33,1	30,2	30,1
1	0,0	0,0	0,0
2	4,7	0,4	0,0
3	0,0	0,0	0,0
4	21,2	0,4	0,0
5	21,6	21,6	21,6
6	28,1	28,1	28,1
7	33,2	33,2	33,2
8	44,6	44,6	44,6

Fuente: Elaboración propia sobre la base de datos provenientes de la ESI 2018.

Finalmente, en lo que se refiere a análisis por tabulados (*Tabla 31*) los resultados muestran que para la primera propuesta de medición se caen 4 tabulados, mientras que no se cae ningún tabulado para la segunda y tercera propuesta.

Tabla 31. Tabulados caídos en total y porcentaje, SEEG-ESI 2018

	% tabulados caídos propuesta 1	% tabulados caídos propuesta 2	% tabulados caídos propuesta 3
Total tabulados Analizados	462	462	462
Total tabulados con más de 50% de celdas caídas	4	0	0
Total tabulados con 50% o menos de celdas caídas	458	462	462
Porcentaje tabulados con más de 50% celdas caídas	0,9%	0,0%	0,0%

Fuente: Elaboración propia sobre la base de datos provenientes de la ESI 2018.

5.5.3. Género y Uso del Tiempo (SEEG-ENUT)

En lo que refiere a Género y Uso del Tiempo, se analizaron un total de 216 celdas y 84 tabulados, considerando un total de cuatro indicadores.

Como lo indica la *Tabla 32*, el criterio de evaluación por grados de libertad es donde se produce mayor cantidad de celdas caídas con 31,1%, provocada principalmente por el indicador 4 con un 58,8%. Al comparar las propuestas para el tratamiento del error

estándar, el análisis muestra que la primera propuesta genera 0,5% de celdas caídas, mientras que la segunda y tercera propuesta no producen celdas caídas.

Tabla 32. Porcentaje de celdas caídas según criterio del flujograma, SEEG-ENUT 2018

N° indicador	Tamaño muestral	Grados de libertad	Coeficiente de variación	Error estándar		
				1° propuesta	2° propuesta	3° propuesta
Total	2,1	31,1	0,0	0,5	0,0	0,0
1	0,0	0,0	-	0,0	0,0	0,0
2	2,2	21,1	-	1,1	0,0	0,0
3	0,0	0,0	0,0	-	-	-
4	2,9	58,8	0,0	-	-	-

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUT 2015.

En cuanto al total de celdas caídas por indicador, según muestra la *Tabla 33*, las mediciones tienen escasa variación entre ellas en términos de porcentaje de celdas caídas, donde la primera propuesta tiene 33,7%, mientras que la segunda y tercera llegan a 33,2%. El mayor porcentaje de celdas caídas se produce en el indicador 4, el cual genera un 61,7% de celdas caídas para todas las propuestas.

Tabla 33. Porcentaje de celdas caídas por indicador, SEEG-ENUT 2018

N° indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
Total	33,7	33,2	33,2
1	0,0	0,0	0,0
2	24,4	23,3	23,3
3	0,0	0,0	0,0
4	61,7	61,7	61,7

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUT 2015.

Finalmente, en lo que se refiere a análisis por tabulados (*Tabla 34*) los resultados muestran que para la primera propuesta se caen tres tabulados, mientras que las propuestas 2 y 3 no generan tabulados caídos.

Tabla 34. Tabulados caídos en total y porcentaje, SEEG-ENUT 2018

	% tabulados caídos propuesta 1	% tabulados caídos propuesta 2	% tabulados caídos propuesta 3
Total tabulados Analizados	84	84	84
Total tabulados con más de 50% de celdas caídas	3	0	0
Total tabulados con 50% o menos de celdas caídas	81	84	84
Porcentaje tabulados con más de 50% celdas caídas	3,6%	0,0%	0,0%

Fuente: Elaboración propia sobre la base de datos provenientes de la ENUT 2015.

5.5.4. Género y Microemprendimiento (SEEG-EME)

Para Género y Microemprendimiento, se analizaron un total de 302 tabulados y 1.657 celdas.

De acuerdo con los resultados presentados en la *Tabla 35*, el tamaño muestral es el criterio donde se caen la mayor cantidad de celdas con 12,4%, seguido por la primera propuesta de tratamiento del error estándar con 11,9%.

Al evaluar por indicadores, el indicador 7 es el que muestra la mayor proporción de celdas caídas con 27,1% para el criterio del tamaño muestral, seguido por el indicador 8 con 26,4%.

Tabla 35. Porcentaje de celdas caídas según criterio del flujograma, SEEG-EME 2017

N° indicador	Tamaño muestral	Grados de libertad	Coeficiente de variación	Error estándar		
				1° propuesta	2° propuesta	3° propuesta
Total	12,4	2,0	3,0	11,9	5,2	0,0
1	0,0	0,0	-	0,0	0,0	0,0
2	0,0	4,0	-	15,3	15,3	0,0
3	6,6	0,0	-	15,8	0	0,0
4	6,7	2,7	-	25,2	10,5	0,0
5	16,9	0,0	3,4	-	-	-
6	10,6	3,3	3,9	-	-	-
7	27,1	0,8	7,6	-	-	-
8	26,4	1,1	8,9	-	-	-

Fuente: Elaboración propia sobre la base de datos provenientes de la EME 2017.

Respecto a la cantidad agregada de celdas caídas por indicador (*Tabla 36*) los resultados muestran una disminución en la cantidad de celdas caídas al pasar de la primera a la tercera propuesta, con 29,3%, 22,6% y 17,4%, respectivamente, siendo el indicador 8 el que tiene el mayor porcentaje de celdas caídas con 36,4% para todas las propuestas.

Tabla 36. Porcentaje de celdas caídas por indicador, SEEG-EME 2017

Nº indicador	% celdas caídas propuesta 1	% celdas caídas propuesta 2	% celdas caídas propuesta 3
Total	29,3	22,6	17,4
1	0,0	0,0	0,0
2	19,3	19,3	4,0
3	22,4	6,6	6,6
4	34,6	19,9	9,4
5	20,3	20,3	20,3
6	17,8	17,8	17,8
7	35,5	35,5	35,5
8	36,4	36,4	36,4

Fuente: Elaboración propia sobre la base de datos provenientes de la EME 2017.

En lo que se refiere a la evaluación del impacto en tabulados (*Tabla 37*) bajo la primera propuesta solo dos tabulados tienen más del 50% de sus celdas caídas (0,7%) cifra que desaparece para el caso de la segunda y tercera propuesta.

Tabla 37. Tabulados caídos en total y porcentaje, SEEG-EME 2017

	% tabulados caídos propuesta 1	% tabulados caídos propuesta 2	% tabulados caídos propuesta 3
Total tabulados Analizados	302	302	302
Total tabulados con más de 50% de celdas caídas	2	0	0
Total tabulados con 50% o menos de celdas caídas	300	302	302
Porcentaje tabulados con más de 50% celdas caídas	0,7%	0,0%	0,0%

Fuente: Elaboración propia sobre la base de datos provenientes de la EME 2017.

6. Conclusión

Las estimaciones obtenidas a partir de las encuestas de hogares, en general, son utilizadas para la toma de decisiones de políticas públicas que afectan a la ciudadanía, y que podrían redundar en mejoras en la calidad de vida. Por ello resulta imprescindible evaluar la calidad de las estimaciones que se producen, así como velar también por la estandarización de los lineamientos de la calidad de modo que sean comparables y aplicables a todas las estadísticas oficiales diseminadas por los miembros del SEN.

Con la pretensión de abordar este tema, se realiza una consulta internacional a diferentes ONE y se profundiza en la literatura en busca de experiencia atinente a esta necesidad. Se precisa que, actualmente entre las ONE, no existe un lineamiento estándar al respecto, y que inclusive a lo interno, para cada encuesta se acogen diferentes criterios para medir la calidad de las estimaciones.

No obstante, se identifican criterios que, inicialmente, son evaluados de forma independiente, logrando fijar umbrales de aceptación para cada uno de ellos. Luego, a modo de propuesta, se ordenan de forma secuencial, y bajo un orden lógico, se aplican siguiendo un flujograma, de tal manera de evaluar la calidad de las estimaciones bajo diferentes dimensiones; finalmente, se evalúa el tabulado que integra las estimaciones en su conjunto.

Respecto a los criterios, se especifican como relevantes el tamaño muestral, los grados de libertad y las medidas de precisión. Para el tamaño muestral se acuerda como umbral idóneo que las estimaciones contengan 60 unidades de análisis o más, lo que subyace un tamaño efectivo de 40 (asumiendo un *deff* de 1,5), valor a partir del cual, se obtienen propiedades favorables al realizar inferencias de los parámetros poblacionales. En cuanto a los grados de libertad, se estudia la distribución de los percentiles más utilizados para realizar las inferencias (nivel de confianza de 90% o 95%) donde se determina que el umbral de nueve o más, genera resguardos para garantizar precisión aceptable.

Respecto a las medidas de precisión, como hallazgo en la literatura, se releva la importancia de un trato especial a las estimaciones del tipo proporciones y razones entre 0 y 1, ya que, la medida habitualmente empleada (*cv*) para medir la calidad de estas estimaciones, presenta propiedades no favorables. Por esta razón se opta, por un lado y, siguiendo las recomendaciones internacionales, estudiar el uso del error estándar para estimadores del tipo proporción y razón entre 0 y 1, y por el otro, para el resto de los estimadores, asumir el coeficiente de variación.

Para las estimaciones a evaluar bajo el error estándar, se analizan el comportamiento de tres funciones (lineal, logarítmica y cuadrática) fijando el error estándar sin perder de vista lo que pasa con el tamaño muestral requerido, el coeficiente de variación y el intervalo de confianza; así como también la dicotomía³¹ de los fenómenos. Después de varios análisis, se elige la función cuadrática $\sqrt[3]{p^2}/9$ para fijar el umbral de aceptación, ya que cuenta con varias bondades relevantes: (i) se adapta al comportamiento esperado, (ii) es parsimoniosa, (iii) exhibe mayores tamaños muestrales requeridos para prevalencias bajas, lo que deriva en mayor resguardo en la calidad de las estimaciones.

Por otro lado, para el resto de las estimaciones, se propone que sigan siendo evaluadas por el *cv*, fijando como umbral de aceptación 15% o menos.

Ya identificados los criterios a emplear y sus respectivos umbrales de aceptación, a efecto de garantizar la calidad de las estimaciones a publicar, se implementan dos flujogramas. En el primero, se conjugan los criterios señalados de manera de lograr categorizar las estimaciones en fiables, poco fiables y no fiables; en el segundo se estudia la pertinencia de publicar el tabulado que integra las estimaciones ya evaluadas. Para esto último, se considera que un tabulado es publicable si más de 50% de las estimaciones presentes en él son clasificadas como estimaciones fiables. En cualquier otro caso, el tabulado no debe publicarse.

Con todo, este documento se ciñe a proveer directriz circunscrita al campo de las encuestas de hogares por muestreo³², específicamente en medir la calidad de las estimaciones desde un punto de vista transversal³³, por lo que quedan exentas estimaciones correspondientes a series de tiempo, o transformaciones que impliquen estimaciones de estudios levantados en diferentes períodos de corte transversal³⁴. Asimismo, se distingue la necesidad de replicar esta experiencia de trabajo para las encuestas por muestreo a empresas, con el fin de cubrir más ámbitos de la producción estadística del INE.

³¹ Por ejemplo, detrás de la estimación de la proporción de personas victimizadas subyace la estimación de las no victimizadas. Ambas estimaciones tienen asociadas el mismo error estándar, mas no el coeficiente de variación, que será significativamente mayor para la contraparte de menor prevalencia.

³² Cabe mencionar que hasta ahora la publicación de las estimaciones estaba supeditada mayormente al criterio del *cv*, por lo que, esta nueva mirada, al considerar más criterios, eleva el nivel de exigencia en la búsqueda de estimaciones robustas.

³³ Las categorías residuales (por ejemplo: no saben, no responde, otros, etcétera) quedan libres de la aplicación del flujograma, ya que, mayormente son condiciones marginales no sustanciales a efectos de los análisis.

³⁴ Durante el desarrollo de las actividades inherentes a este documento se identificaron estimaciones cuya calidad también ameritaba ser evaluada, como variaciones y brechas. No obstante, se sitúa esta evaluación a ser abordada a través de contrastes de hipótesis, tema que escapa del alcance propuesto.

7. Referencias bibliográficas

- Barnett-Walker, K. C. (2003). National Household Survey on Drug Abuse. In K. C. Barnett-Walker, *Statistical Inference Report*. Rockville, Maryland 20857, EEUU. Retrieved from samhdasupport@
- C Pardo, G. C. (2001). Métodos estadísticos multivariados en investigación social. *Simposio de estadística. Universidad Nacional de Colombia*.
- Cassell, D. L. (2008). *Wait Wait, Don't Tell Me... You're Using the Wrong Proc!* Obtenido de https://www.lexjansen.com/pnwsug/2008/DavidCassell_WaitWait.pdf
- Cochran, W. (1998). *Técnicas de Muestreo*. México D.F.: Compañía Editorial Continental, S.A. de C.V.
- Eurostat. (2013). *Handbook on precision requirements and variance estimation for ESS households surveys*. Publications Office of the European Union.
- EUROSTAT. (2017). *Código de buenas prácticas de las Estadística europeas*. Obtenido de <https://ec.europa.eu/eurostat/documents/4031688/9394048/KS-02-18-142-ES-N.pdf/e792b761-6f09-42a9-a1e0-3a3356a0de1c>
- Gallardo San Salvador, J. Á. (2017). Ampliación de Análisis de Datos Multivariantes. En J. Á. Gallardo San Salvador, *Ampliación de Análisis de Datos Multivariantes*. Obtenido de <https://www.ugr.es/~gallardo/#adm>
- Grupo Intersecretarial de Trabajo sobre Cuentas Nacionales (GITCN) (2009). Obtenido de <https://unstats.un.org/unsd/nationalaccount/docs/SNA2008Spanish.pdf>
- Gutiérrez, A. (2018). Limitaciones de las Encuestas de Hogares en la Medición de Indicadores Sociales. *Taller regional sobre desagregación de estadísticas sociales mediante métodos de estimación en áreas pequeñas*. Santiago de Chile: CEPAL.
- Heath, J. P., & Borowski, P. (2003). *Quantifying Proportional Variability*. Canada: Vancouver: Department of Mathematics, University of British Columbia.
- Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied Survey Data Analysis*. Boca ratón, FL: Taylor and Francis Group, LLC.

- Hornik, R., Maklan, D., Cadell, D., Prado, A., Barmada, C., & Jaconsohn, L. (2002). *Evaluation of the National Youth Anti-Drug Media Campaign*. Pennsylvania: National Institute on Drug Abuse, National Institute of Health.
- INE Chile. (2018a). *Informe de Calidad. VIII Encuesta de Presupuestos Familiares*. Santiago: Subdirección Técnica. Departamento de Presupuestos Familiares.
- INE Chile. (2018b). *Metodología - VIII Encuesta de Presupuestos Familiares*. Instituto Nacional de Estadísticas de Chile, Departamento de Presupuestos Familiares. Obtenido de <http://www.ine.cl/docs/default-source/ingresos-y-gastos/epf/viii-epf/documentacion/metodolog%C3%ADa-viii-epf.pdf?sfvrsn=4>
- INE Chile. (febrero de 2020). *Documento Metodológico, Encuesta Nacional de Empleo*. (D. d. Estadística, Ed.) Obtenido de https://www.ine.cl/docs/default-source/ocupacion-y-desocupacion/metodologia/espanol/metodolog%C3%ADa-encuesta-nacional-de-empleo-ene-2020.pdf?sfvrsn=793380e9_6
- INEGI. (2011). *Diseño de la muestra en proyectos de encuesta*. México: Instituto Nacional de Estadísticas.
- Instituto Nacional de Estadísticas, Chile. (Enero de 2018). *Manual conceptual y metodológico diseño muestral Encuesta Nacional de Empleo*. Obtenido de sitio web INE: <http://www.ine.cl/docs/default-source/laborales/ene/antecedentes-metodologicos/manual-conceptual-y-metodol%C3%B3gico-dise%C3%B1o-muestral.pdf?sfvrsn=6>
- Kish, L. (1980). *Design and Estimation for Domain*. USA: Michigan: Journal of the Royal Statistical Society. Series D (The Statistician), Vol. 29, No. 4.
- Kish, L. (1963). *Changing strata and selection probabilities*. Proceedings of the Social Statistics Section, American Statistical Association, 124-131.
- Korn, E., & Graubard, B. (1999). *Analysis of Health Surveys*. USA: John Wiley & Sons, INC.
- Lincoln, C. (1993). *Estadísticas para las ciencias administrativas*. Santa Fe de Bogotá: McGraw-Hill Interamericana S.A, tercera edición.
- Martínez, J. C. (2007). Una aproximación metodológica al uso de datos de encuestas en hogares. *Realidad, Datos y Espacio. Revista Internacional de Estadística y Geografía*, 68-75.

- MDS. (2015). *Metodología de Diseño Muestral Encuesta de Caracterización Socioeconómica Nacional, Casen 2015, Serie Documentos Metodológicos Casen N°33*. Ministerio de Desarrollo Social, Observatorio Social, Santiago. Obtenido de http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Metodologia_de_Disenio_Muestral_Casen_2015.pdf
- Molina, I. (2018). *Serie Estudios Estadísticos: Desagregación de datos en encuestas de hogares, Metodologías de estimación en áreas pequeñas*. Santiago: Naciones Unidas: CEPAL.
- ONU. (2009). *Diseño de muestras para encuestas de hogares. Directrices prácticas*. Recuperado el 15 de Noviembre de 2017, de https://unstats.un.org/unsd/publication/seriesf/Seriesf_98s.pdf
- Pagano, R. (2009). *Understanding Statistics in the Behavioral Sciences*. USA: Cengage Learning.
- Parker , J., Talih , M., Malec , D., Madans, J., Rothwell, C., Paulose, R., & Blumberg, S. (2017). *National Center for Health Statistics Data Presentation Standards for Proportions*. Hyattsville, Maryland : National Center for Health Statistics, Vital Health Stat, Series 2-175.
- Siller, A. B., & Tompkins, L. (2005). *The big four: Analyzing complex sample survey data using SAS, SPSS, STATA, and SUDAAN*. Hyattsville: SUGI 31. Paper 172-31. Obtenido de <https://www.lexjansen.com/nesug/nesug05/pos/pos3.pdf>
- Valliant, R., Dever, J. A., & Kreuter, F. (2012). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- Vásquez, E. R., & Núñez, A. C. (2011). *Inconsistencia del coeficiente de variación para expresar la variabilidad de un experimento en un modelo de análisis de varianza*. Cuba: Ministerio de Educación Superior.
- Wolter, K. (2007). *Statistics for Social and Behavioral Scienses, Introduction to Variance Estimation*. New York, USA: Springer.

8. ANEXOS

8.1. Anexo 1. Función cuadrática alternativa para el umbral del error estándar

Esta propuesta parte de la conocida forma geométrica de la varianza de una proporción, es decir, la gráfica del producto $p \cdot q$, cuya forma es cóncava y que luego se ajusta para pasar por puntos ya consensuados en las propuestas previas para los indicadores evaluados. La función queda determinada por la siguiente función cuadrática:

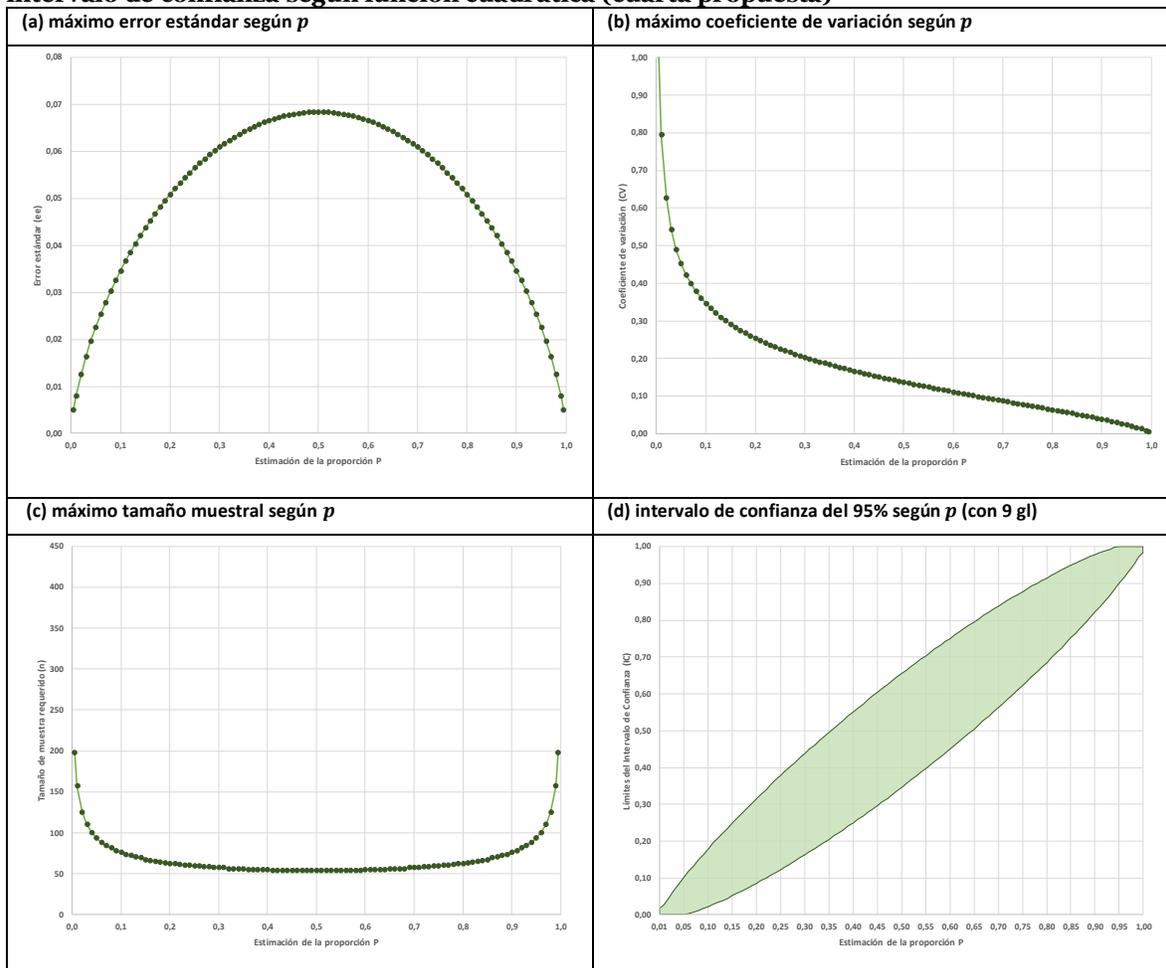
$$\text{Máximo } ee \text{ tolerable} = \frac{\sqrt[3]{(p \cdot q)^2}}{5,8} ; 0 < p < 1$$

Como se observa en el *Gráfico 9*, esta función también emula, al igual que la función de la tercera propuesta, el comportamiento deseado tanto para el *ee* como para el *cv* (*Gráfico 9*: (a) y (b)) logrando gradualmente ir exigiendo mayores *ee* a medida que la prevalencia crece, manteniendo la simetría de esta medida. Nótese además que, para prevalencias bajas (por ejemplo $p = 0,005$) el *cv* logra alcanzar valores próximos a 1, comportamiento que se condice con lo encontrado en la segunda propuesta (*cv* logarítmico) donde se consiguen tamaños muestrales máximos tolerables también similares para las prevalencias extremas, de alrededor de 200.

Al igual que la función cuadrática propuesta (*sección 4.2.2.1*) esta presenta las mismas bondades estadísticas aceptables para el error estándar (simetría y máximo en $p = 0,50$) como para el intervalo de confianza (*Gráfico 9*: (d)) en donde la amplitud máxima se alcanza en $p = 0,50$.

Esta propuesta es menos exigente que la tercera propuesta, debido a que, se consiguen tamaños muestrales máximos tolerables para las prevalencias extremas, de alrededor de 200 unidades de análisis, lo que lo hace menos robusto a la hora de analizar prevalencias poco frecuentes.

Gráfico 9. Máximo error estándar, coeficiente de variación, tamaño muestral requerido e intervalo de confianza según función cuadrática (cuarta propuesta)



Fuente: Elaboración propia. INE, 2020.

8.2. Anexo 2. Evaluación de las estimaciones por valor del cv según ONE

Cuadro 14. Categorización de las estimaciones por valor del cv según ONE

País	Categorización de las estimaciones según valores del cv
Argentina	<ul style="list-style-type: none"> - $cv \leq 16,6\%$ → publicar - $16,6\% < cv \leq 33,3\%$ → publicable con advertencia "usar con precaución" - $cv > 33,3\%$ → resultados referenciales
Australia	- "La Oficina Australiana de Estadística (ABS en sus siglas en inglés) sigue las directrices y recomendaciones internacionales establecidas, sin embargo, a menudo son muy generales y deben interpretarse en el contexto de nuestra propia situación"
Canadá	<ul style="list-style-type: none"> - $cv \leq 2\%$ → excelente - $2\% < cv \leq 4\%$ → muy bueno - $4\% < cv \leq 8\%$ → bueno - $8\% < cv \leq 16\%$ → aceptable - $16\% < cv \leq 33,33\%$ → usar con precaución - $cv > 33,33\%$, → demasiado poco confiable para ser publicado
Costa Rica	<ul style="list-style-type: none"> - $cv \leq 5\%$ → muy precisas - $5\% < cv \leq 10\%$ → estimaciones precisas - $10\% < cv \leq 20\%$ → estimación aceptable - $cv > 20\%$ → estimación poco confiables por lo que se debe usar con precaución
Ecuador	<ul style="list-style-type: none"> - $cv \leq 20\%$ → publicable - $cv > 20\%$ → publicable con advertencia "usar con precaución"
España	- No hay una norma única, depende de la encuesta
Francia	- "Ninguno: actualmente se configura más o menos por práctica. Las regulaciones de la UE están en preparación"
Italia	<ul style="list-style-type: none"> - "Como estado miembro de la Unión Europea, el Instituto Nacional Italiano de Estadística adopta los estándares europeos para la presentación de informes de calidad. El material está disponible en: https://ec.europa.eu/eurostat/web/quality/quality-reporting" - "Las encuestas estadísticas de Istat también siguen pautas generales de calidad en todo el proceso estadístico. Hay dos manuales disponibles en la siguiente página: https://www.istat.it/en/methods-and-tools/tools-for-data-quality/guidelines"

País	Categorización de las estimaciones según valores del cv
México	<ul style="list-style-type: none"> - $cv < 5\%$ → precisión alta - $10\% \leq cv < 15\%$ → precisión moderada - $cv \geq 15\%$ → precisión baja
Paraguay	<ul style="list-style-type: none"> - $cv \leq 20\%$ → publicable - $cv \geq 20\%$ → publicable con advertencia "las estimaciones no son confiables"
Perú	<ul style="list-style-type: none"> - $cv \leq 5\%$ → muy buena precisión - $5\% < cv \leq 10\%$ → buena precisión - $10\% < cv \leq 15\%$ → precisión aceptable - $cv > 15\%$ → resultados referenciales
República Dominicana	Según los lineamientos de las buenas prácticas de otros institutos u oficinas de estadística de la región y el mundo, tales como: INEGI de México, INE de España, entre otros
Venezuela	<ul style="list-style-type: none"> - $cv < 5\%$ → muy Buena - $5\% \leq cv < 10\%$ → buena - $10\% \leq cv < 15\%$ → aceptable - $15\% \leq cv < 20\%$ → utilizarla con precaución - $cv \geq 20\%$ → no es buena y sólo se tomará como cifra solo de referencia

Fuente: Elaboración propia sobre la base de consulta a las ONE sobre uso de medidas de precisión en estimaciones, 2019.

8.3. Anexo 3. Indicadores evaluados por encuesta

Tabla 38. Indicadores evaluados en la ENE

Nº	Período	Nombre del conjunto de tabulados	Zona geográfica	Total de tabulados evaluados	Total de celdas evaluadas
Total				30	243
1	OND 2019	Tasa de desocupación	Nacional, regional, provincial y conurbaciones	6	52
2		Nivel población de 15 años y más (PET)	Nacional y regional	14	102
3		Inactividad según potencialidad	Nacional	2	4
4		Inactividad según razones de inactividad	Nacional	2	12
5		Nivel rama de actividad económica	Nacional	2	22
6		Tasa Indicadores de subutilización de la fuerza de trabajo	Nacional y regional	4	51

Fuente: Elaboración propia. INE, 2020.

Tabla 39. Indicadores evaluados en la ESI

Nº	Período	Nombre del conjunto de tabulados	Zona Geográfica	Total de tabulados evaluados	Total de celdas evaluadas
Total				116	4.505
1	2017	Ingreso medio nominal de la población ocupada	Nacional, Regional, provincia y ciudad	31	2.763
2		Ingreso nominal por hora de las personas asalariadas	Nacional	20	171
3		Ingreso medio nominal de las personas asalariadas del sector no agropecuario	Nacional y regional	24	867
4		Ingreso medio real de las personas asalariadas con contrato definido	Nacional	28	165
5		Proporción de personas ocupadas que viven por debajo de la mediana de los ingresos	Nacional y regional	8	408
6		Participación de los ingresos de la ocupación en los ingresos del hogar, según decil	Nacional	2	11
7		Participación de cada decil en el ingreso total de los hogares	Nacional	1	10
8		Composición del ingreso por decil de hogares según fuente de ingresos	Nacional	2	110

Fuente: Elaboración propia. INE, 2020.

Tabla 40. Indicadores evaluados en la VIII EPF

Nº	Período	Nombre del conjunto de tabulados	Zona geográfica	Total tabulados evaluados	Total de celdas evaluadas
Total				102	63.363
1	2016-2017	Gasto promedio mensual del hogar por características del sustentador principal y su hogar	Total capitales regionales, Gran Santiago y Resto de capitales regionales	30	195
2		Gasto promedio mensual del hogar por quintiles de ingreso per cápita		6	18
3		Participación en el gasto promedio mensual del hogar por quintiles de ingreso per cápita		3	15
4		Gasto promedio mensual del hogar por quintiles de ingreso per cápita, división y grupo CCIF		24	1.530
5		Participación en el gasto promedio mensual del hogar por quintiles de ingreso per cápita, división y grupo CCIF		18	1524
6		Gasto promedio mensual del hogar por quintiles de ingreso per cápita y producto CCIF		12	30.042
7		Participación en el gasto promedio mensual del hogar por quintiles de ingreso per cápita y producto CCIF		9	30.039

Fuente: Elaboración propia. INE, 2020.

Tabla 41. Indicadores evaluados en la ENUT

Nº	Período	Nombre del conjunto de tabulados	Zona geográfica	Total de tabulados evaluados	Total de celdas evaluadas
Total				208	8.640
1	2015	Carga global de trabajo en la población ocupada de 15 años y más: tasa de participación en día tipo	Nacional y regional	12	480
2		Carga global de trabajo en la población ocupada de 15 años y más: tiempo promedio en día tipo		12	480
3		Trabajo no remunerado fuera de la frontera de producción del SCN: tasa de participación en día tipo		16	672
4		Trabajo no remunerado fuera de la frontera de producción del SCN: tiempo promedio en día tipo		16	672
5		Trabajo doméstico no remunerado para el propio hogar: tasa de participación en día tipo		16	672
6		Trabajo doméstico no remunerado para el propio hogar: tiempo promedio en día tipo		16	672
7		Trabajo de cuidados no remunerados a integrantes del hogar: tasa de participación en día tipo		16	672
8		Trabajo de cuidados no remunerados a integrantes del hogar: tiempo promedio en día tipo		16	672
9		Trabajo en la ocupación y traslados asociados: tasa de participación en día tipo		12	480
10		Trabajo en la ocupación y traslados asociados: tiempo promedio en día tipo		12	480
11		Trabajo no remunerado para otros hogares, comunidad y voluntario: tasa de participación en día tipo		16	672
12		Trabajo no remunerado para otros hogares, comunidad y voluntario: tiempo promedio en día tipo		16	672
13		Ocio y vida social: tasa de participación en día tipo		16	672
14		Ocio y vida social: tiempo promedio en día tipo		16	672

Fuente: Elaboración propia. INE, 2020.

Tabla 42. Indicadores evaluados en la ENUSC

N°	Período	Nombre del conjunto de tabulados	Zona geográfica	Total de tabulados evaluados	Total de celdas evaluadas
Total				56	1.488
1	2017	Percepción de aumento de la delincuencia en el País	Nacional y regional	8	240
2		Victimización personal agregada por delitos de mayor connotación social		8	240
3		Victimización a hogares agregada por delitos de mayor connotación social		4	96
4		Victimización a hogares agregada por delitos de connotación económica		4	96
5		Denuncia agregada de hogares por delitos de mayor connotación social		4	96
6		Robo de vehículos en hogares		4	96
7		Revictimización a hogares por delitos de mayor connotación social		4	96
8		Victimización de hogares por delitos cibernéticos		4	96
9		Victimización de hogares por delitos de vandalismo		4	96
10		Intentos de delitos		4	96
11		Percepción de exposición frente al delito		8	240

Fuente: Elaboración propia. INE, 2020.

Tabla 43. Indicadores evaluados en la SEEG-ENUSC

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
Total			207	1.145
1	Percepción del nivel de la delincuencia en los últimos 12 meses en el barrio, por sexo	Nacional, ambos sexos, concentración.	4	19
	Percepción del nivel de seguridad frente al delito cuando ya está oscuro caminando solo por su barrio, por sexo			
	Percepción de exposición frente al delito en los próximos doce meses, por sexo			
	Víctimas del delito robo por sorpresa, por sexo			
2	Percepción del nivel de la delincuencia en los últimos 12 meses en el barrio, por sexo	Regional, ambos sexos, concentración	30	135
	Percepción de exposición frente al delito en los próximos doce meses, por sexo			
3	Percepción del nivel de la delincuencia en los últimos 12 meses en el barrio, por sexo	Nacional, por sexo concentración y distribución.	8	76
	Percepción del nivel de seguridad frente al delito cuando ya está oscuro caminando solo por su barrio, por sexo			

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
	Percepción de exposición frente al delito en los próximos doce meses, por sexo			
	Víctimas del delito robo por sorpresa, por sexo			
4	Percepción del nivel de la delincuencia en los últimos 12 meses en el barrio, por sexo	Regional, por sexo, concentración y distribución.	60	540
	Percepción de exposición frente al delito en los próximos doce meses, por sexo			
5	Percepción del nivel de la delincuencia en los últimos 12 meses en el barrio, por sexo	Nacional, ambos sexos y total.	10	20
	Percepción de seguridad en terminal de buses, por sexo			
	Percepción del nivel de seguridad frente al delito cuando ya está oscuro caminando solo por su barrio, por sexo			
	Percepción de exposición frente al delito en los próximos doce meses, por sexo			
6	Percepción del nivel de la delincuencia en los últimos 12 meses en el barrio, por sexo	Regional, ambos sexos y total.	60	105
	Percepción de exposición frente al delito en los próximos doce meses, por sexo			
7	Percepción del nivel de la delincuencia en los últimos 12 meses en el barrio, por sexo	Nacional, por sexo	5	40
	Percepción de seguridad en terminal de buses, por sexo			
	Percepción del nivel de seguridad frente al delito cuando ya está oscuro caminando solo por su barrio, por sexo			
	Percepción de exposición frente al delito en los próximos doce meses, por sexo			
	Víctimas del delito robo por sorpresa, por sexo			
8	Percepción del nivel de la delincuencia en los últimos 12 meses en el barrio, por sexo	Regional, por sexo	30	210
	Percepción de exposición frente al delito en los próximos doce meses, por sexo			

Fuente: Elaboración propia. INE, 2020.

Tabla 44. Indicadores evaluados en la SEEG-ESI

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
Total			462	4.095
1	Proporción y concentración de las personas ocupadas, según grupo ocupacional.	Nacional, ambos sexos	4	31
	Proporción y concentración de las personas ocupadas dependientes, según horas habituales de trabajo.			
	Proporción y concentración de los empleadores, según tramo de ingreso mínimo neto.			

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
	Proporción y concentración de los cuenta propia, según horas habituales de trabajo.			
2	Proporción y concentración de las personas ocupadas, según grupo ocupacional.	Regional, ambos sexos	30	255
	Proporción y concentración de las personas ocupadas dependientes, según horas habituales de trabajo.			
3	Proporción y concentración de las personas ocupadas, según grupo ocupacional.	Nacional, por sexo	4	62
	Proporción y concentración de las personas ocupadas dependientes, según horas habituales de trabajo.			
	Proporción y concentración de los empleadores, según tramo de ingreso mínimo neto.			
	Proporción y concentración de los cuenta propia, según horas habituales de trabajo.			
4	Proporción y concentración de las personas ocupadas, según grupo ocupacional.	Regional, por sexo	30	510
	Proporción y concentración de las personas ocupadas dependientes, según horas habituales de trabajo.			
5	Número de personas ocupadas por sexo, según grupo ocupacional	Nacional, ambos sexos y total	29	172
	Ingreso por hora e ingreso medio y mediano mensual de las personas ocupadas dependientes por sexo, según nivel educacional.			
	Número de personas ocupadas dependientes por sexo, según nivel educacional.			
	Ingreso por hora e ingreso medio y mediano mensual de las personas ocupadas dependientes por sexo, según horas habituales de trabajo.			
	Número de personas ocupadas dependientes por sexo, según horas habituales de trabajo.			
	Ingreso medio y mediano mensual de los empleadores por sexo, según nivel educacional.			
	Número de empleadores por sexo, según nivel educacional.			
	Número de empleadores por sexo, según tramo de ingreso mínimo neto.			
	Número de cuenta propia por sexo, según rama de actividad (ciiu4.cl 2012)			
	Número de cuenta propia por sexo, según rama de actividad (CIU4.cl 2012)			
	Ingreso medio y mediano mensual de los cuenta propia por sexo, según horas habituales de trabajo.			
Número de cuenta propia por sexo, según horas habituales de trabajo.				
6	Número de personas ocupadas por sexo, según región.	Regional, ambos sexos y total	196	869
	Número de personas ocupadas por sexo, según grupo ocupacional			

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
	Ingreso por hora e ingreso medio y mediano mensual de las personas ocupadas dependientes por sexo, según nivel educacional.			
	Número de personas ocupadas dependientes por sexo, según nivel educacional.			
	Ingreso por hora e ingreso medio y mediano mensual de las personas ocupadas dependientes por sexo, según horas habituales de trabajo.			
	Número de personas ocupadas dependientes por sexo, según horas habituales de trabajo.			
7	Número de personas ocupadas por sexo, según grupo ocupacional	Nacional, por sexo y total	19	334
	Ingreso por hora e ingreso medio y mediano mensual de las personas ocupadas dependientes por sexo, según nivel educacional.			
	Número de personas ocupadas dependientes por sexo, según nivel educacional.			
	Ingreso por hora e ingreso medio y mediano mensual de las personas ocupadas dependientes por sexo, según horas habituales de trabajo.			
	Número de personas ocupadas dependientes por sexo, según horas habituales de trabajo.			
	Ingreso medio y mediano mensual de los empleadores por sexo, según nivel educacional.			
	Número de empleadores por sexo, según nivel educacional.			
	Número de empleadores por sexo, según tramo de ingreso mínimo neto.			
	Número de cuenta propia por sexo, según rama de actividad (ciiu4.cl 2012)			
	Número de cuenta propia por sexo, según rama de actividad (CIU4.cl 2012)			
	Ingreso medio y mediano mensual de los cuenta propia por sexo, según horas habituales de trabajo.			
Número de cuenta propia por sexo, según horas habituales de trabajo.				
8	Número de personas ocupadas por sexo, según región.	Regional, por sexo	150	1.862
	Número de personas ocupadas por sexo, según grupo ocupacional			
	Ingreso por hora e ingreso medio y mediano mensual de las personas ocupadas dependientes por sexo, según nivel educacional.			
	Número de personas ocupadas dependientes por sexo, según nivel educacional.			
	Ingreso por hora e ingreso medio y mediano mensual de las personas ocupadas dependientes por sexo, según horas habituales de trabajo.			
	Número de personas ocupadas dependientes por sexo, según horas habituales de trabajo.			

Fuente: Elaboración propia. INE, 2020.

Tabla 45. Indicadores evaluados en la SEEG-ENUT

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
Total			84	216
1	Tiempo promedio destinado a la carga global de trabajo y concentración porcentual por sexo	Nacional, por sexo.	14	32
	Tiempo promedio destinado a la carga global de trabajo en personas ocupadas por sexo según jornada de trabajo en la ocupación			
	Tasa de participación y tiempo promedio destinado a trabajos no remunerados por sexo			
	Distribución del trabajo doméstico no remunerado para el propio hogar por sexo según tipo de actividad			
	Tasa de participación y tiempo promedio destinado a trabajos no remunerados en día de semana de jóvenes laboralmente inactivos que no estudian			
2	Tiempo promedio destinado a la carga global de trabajo y concentración porcentual por sexo	Regional, por sexo	15	90
	Tiempo promedio destinado a la carga global de trabajo y concentración porcentual por sexo	Nacional, por sexo	10	26
	Tiempo promedio destinado a la carga global de trabajo en personas ocupadas por sexo según jornada de trabajo en la ocupación			
	Tasa de participación y tiempo promedio destinado a trabajos no remunerados por sexo			
	Tasa de participación y tiempo promedio destinado a trabajos no remunerados en día de semana de jóvenes laboralmente inactivos que no estudian			
	Tiempo promedio destinado a la carga global de trabajo y concentración porcentual por sexo	Regional, por sexo	45	68
	Tasa de participación y tiempo promedio destinado a trabajos no remunerados por sexo			

Fuente: Elaboración propia. INE, 2020.

Tabla 46. Indicadores evaluados en la SEEG-EME

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
Total			302	1657
1	Personas microempendedoras por sexo según categoría ocupacional	Nacional, ambos sexos. Concentración.	5	38
	Personas microempendedoras por sexo según trabajo asalariado previo			
	Personas microempendedoras por sexo según actividad económica del negocio			
	Personas microempendedoras por sexo según solicitud de préstamo bancario			

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
	Categoría ocupacional prevalente del personal contratado por sexo de la persona microempresaria			
2	Personas microempresarias por sexo según categoría ocupacional	Regional ambos sexos. Concentración	45	150
	Personas microempresarias por sexo según trabajo asalariado previo			
	Personas microempresarias por sexo según solicitud de préstamo bancario			
3	Personas microempresarias por sexo según categoría ocupacional	Nacional, por sexo. Concentración y distribución,	10	152
	Personas microempresarias por sexo según trabajo asalariado previo			
	Personas microempresarias por sexo según solicitud de préstamo bancario			
	Categoría ocupacional prevalente del personal contratado por sexo de la persona microempresaria			
4	Personas microempresarias por sexo según categoría ocupacional	Regional, por sexo. Concentración y distribución.	90	600
	Personas microempresarias por sexo según trabajo asalariado previo			
	Personas microempresarias por sexo según solicitud de préstamo bancario			
5	Personas microempresarias por sexo según categoría ocupacional	Nacional, ambos sexos y total	16	59
	Personas microempresarias por sexo según proveedor principal			
	Personas microempresarias por sexo según razón por la que dejó de trabajar como asalariada			
	Personas microempresarias por sexo según trabajo asalariado previo			
	Personas microempresarias por sexo según actividad económica del negocio			
	Personas microempresarias por sexo según ganancia del emprendimiento			
	Personas microempresarias por sexo según solicitud de préstamo bancario			
Categoría ocupacional prevalente del personal contratado por sexo de la persona microempresaria				
6	Personas microempresarias por sexo según categoría ocupacional	Regional, ambos sexos y total.	68	180
	Personas microempresarias por sexo según proveedor principal			
	Personas microempresarias por sexo según trabajo asalariado previo			
	Personas microempresarias por sexo según solicitud de préstamo bancario			
7	Personas microempresarias por sexo según categoría ocupacional	Nacional, por sexo.	8	118
	Personas microempresarias por sexo según proveedor principal			
	Personas microempresarias por sexo según razón por la que dejó de trabajar como asalariada			
	Personas microempresarias por sexo según trabajo asalariado previo			

N°	Nombre del conjunto de tabulados	Nivel geográfico y desagregación	Total de tabulados evaluados	Total de celdas evaluadas
	Personas microempendedoras por sexo según actividad económica del negocio			
	Personas microempendedoras por sexo según ganancia del emprendimiento			
	Personas microempendedoras por sexo según solicitud de préstamo bancario			
	Categoría ocupacional prevalente del personal contratado por sexo de la persona microempendedora			
8	Personas microempendedoras por sexo según categoría ocupacional	Regional, por sexo	60	360
	Personas microempendedoras por sexo según proveedor principal			
	Personas microempendedoras por sexo según trabajo asalariado previo			
	Personas microempendedoras por sexo según solicitud de préstamo bancario			

Fuente: Elaboración propia. INE, 2020.