

Recomendaciones metodológicas
sobre la medición de la calidad
de las cifras provenientes
de encuestas de hogares



NACIONES UNIDAS

CEPAL



**Conferencia
Estadística**

de las
Américas
de la CEPAL

Gracias por su interés en esta publicación de la CEPAL



Si desea recibir información oportuna sobre nuestros productos editoriales y actividades, le invitamos a registrarse. Podrá definir sus áreas de interés y acceder a nuestros productos en otros formatos.

Deseo registrarme



NACIONES UNIDAS



www.cepal.org/es/publications



www.instagram.com/publicacionesdelacepal



www.facebook.com/publicacionesdelacepal



www.issuu.com/publicacionescepal/stacks



www.cepal.org/es/publicaciones/apps

Recomendaciones metodológicas
sobre la medición de la calidad
de las cifras provenientes
de encuestas de hogares



NACIONES UNIDAS

CEPAL



**Conferencia
Estadística**

de las
Américas
de la CEPAL

Este documento fue elaborado por el Grupo de Trabajo para la elaboración de recomendaciones metodológicas sobre la medición de la calidad de las cifras provenientes de encuestas de hogares, de la Conferencia Estadística de las Américas (CEA) de la Comisión Económica para América Latina y el Caribe (CEPAL), bienio 2022-2023. El Grupo fue coordinado por Chile (Instituto Nacional de Estadísticas) y tuvo como Secretaría Técnica a la División de Estadísticas de la CEPAL. Los países y entidades miembros del Grupo son: Argentina (Instituto Nacional de Estadística y Censos (INDEC)), Bolivia (Estado Plurinacional de) (Instituto Nacional de Estadística (INE)), Brasil (Instituto Brasileiro de Geografía y Estadística (IBGE)), Colombia (Departamento Administrativo Nacional de Estadística (DANE)), Costa Rica (Instituto Nacional de Estadística y Censos (INEC)), Cuba (Oficina Nacional de Estadística e Información (ONEI)), Ecuador (Instituto Nacional de Estadística y Censos (INEC)), El Salvador (Oficina Nacional de Estadística y Censo), Guatemala (Instituto Nacional de Estadística (INE)), México (Instituto Nacional de Estadística y Geografía (INEGI)), Nicaragua (Instituto Nacional de Información de Desarrollo (INIDE)), Paraguay (Instituto Nacional de Estadística (INE)), Perú (Instituto Nacional de Estadística e Informática (INEI)), República Dominicana (Oficina Nacional de Estadística (ONEI)) y Uruguay (Instituto Nacional de Estadística (INE)).

En la elaboración de este documento participaron las siguientes personas:

Chile (INE): Marly Olivares Lathulerie, Miguel Guerrero Herrera, Jonathan Paredes Baez, Iván Touron Romero, Denisse López Arenas, Felipe Molina Jaque

Bolivia (INE): Rosemary Villanueva Calle

Brasil (IBGE): Marcus Morais Fernandes

Colombia (DANE): María Margarita Echeverry Vásquez

México (INEGI): Eric Rodríguez Herrera, José Elías Rodríguez Muñoz

Perú (INEI): Nancy Hidalgo

Publicación de las Naciones Unidas
LC/CEA.12/11
Distribución: L
Copyright © Naciones Unidas, 2024
Todos los derechos reservados
Impreso en Naciones Unidas, Santiago
S.2301136[S]

Esta publicación debe citarse como: Grupo de Trabajo para la elaboración de recomendaciones metodológicas sobre la medición de la calidad de las cifras provenientes de encuestas de hogares de la Conferencia Estadística de las Américas, *Recomendaciones metodológicas sobre la medición de la calidad de las cifras provenientes de encuestas de hogares* (LC/CEA.12/11), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL), 2024.

La autorización para reproducir total o parcialmente esta obra debe solicitarse a la Comisión Económica para América Latina y el Caribe (CEPAL), División de Documentos y Publicaciones, publicaciones.cepal@un.org. Los Estados Miembros de las Naciones Unidas y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Solo se les solicita que mencionen la fuente e informen a la CEPAL de tal reproducción.

Índice

| | |
|---|-----------|
| Presentación..... | 5 |
| Capítulo I | |
| Antecedentes de aplicación de estándares de calidad por parte de las Oficinas Nacionales de Estadística | 7 |
| Capítulo II | |
| Marco conceptual | 13 |
| A. Encuestas por muestreo y sus características | 13 |
| 1. Marco de muestreo | 14 |
| 2. Tipos de muestreo..... | 15 |
| 3. Elementos que se deben considerar en un esquema de muestreo..... | 20 |
| B. Diseños muestrales complejos y sus características | 21 |
| C. Parámetros poblacionales y sus estimadores | 21 |
| D. Intervalo de confianza | 23 |
| E. Estimación de los errores muestrales | 24 |
| 1. La varianza y sus estimadores | 24 |
| 2. Medidas de dispersión asociadas con la varianza | 27 |
| 3. Tabulados o desagregaciones | 29 |
| Capítulo III | |
| Elementos para evaluar la calidad estadística de una estimación..... | 31 |
| A. Efecto del diseño..... | 31 |
| B. Tamaño muestral mínimo para aplicar medidas de precisión | 32 |
| 1. Tamaño muestral | 32 |
| 2. Tamaño efectivo de la muestra | 33 |
| 3. Grados de libertad..... | 33 |
| C. Tipología de los estimadores..... | 35 |
| D. Medidas de precisión utilizadas para la publicación y difusión..... | 37 |
| 1. Error estándar..... | 38 |
| 2. Coeficiente de variación | 38 |
| 3. Coeficiente de variación logarítmico | 39 |
| E. Conteo de casos no ponderado..... | 40 |
| Capítulo IV | |
| Requerimientos de calidad que se aplican a las estimaciones..... | 41 |
| Capítulo V | |
| Consideraciones finales | 45 |
| Bibliografía | 47 |

| | |
|------------------|---|
| Cuadros | |
| Cuadro II.1 | Principales parámetros, estimadores y varianza estimada en el muestreo aleatorio simple.....25 |
| Cuadro II.2 | Estimación de personas por sexo, según nivel socioeconómico..... 30 |
| Gráficos | |
| Gráfico I.1 | Frecuencia de uso de las métricas de calidad..... 11 |
| Gráfico III.1 | Percentil 0,95 de la distribución t de Student y variación porcentual según los grados de libertad..... 34 |
| Gráfico III.2 | Máximo coeficiente de variación y error estándar admitido según la estimación de <i>P</i> (primer criterio) 35 |
| Gráfico III.3 | Máximo coeficiente de variación y error estándar admitido según la estimación de <i>P</i> (segundo criterio) 36 |
| Gráfico III.4 | Comportamiento del error estándar y del coeficiente de variación según la estimación de <i>P</i> en una muestra aleatoria simple en que $n=60$ 37 |
| Gráfico III.5 | Relación entre el tamaño de la muestra y la precisión de un indicador utilizando la transformación logit..... 40 |
| Diagramas | |
| Diagrama I.1 | Dominios de estudio 10 |
| Diagrama I.2 | Otros dominios de interés..... 10 |
| Diagrama II.1 | Procesos y subprocesos del modelo MGPIE 5.1: segmento negocio 14 |
| Diagrama III.1 | Precisión de los estimadores determinada por la concentración o dispersión de los valores estimados..... 37 |
| Diagrama IV.1 | Flujograma de evaluación de las estimaciones..... 41 |
| Diagrama IV.2 | Flujograma de evaluación de los tabulados 43 |

Presentación

A nivel nacional se necesita información socioeconómica cada vez más desagregada con el fin de implementar, hacer el seguimiento y evaluar las políticas públicas. Las oficinas nacionales de estadística (ONE) satisfacen la mayor parte de esa necesidad llevando a cabo diversas operaciones estadísticas o encuestas dirigidas a los hogares con el objeto de abarcar una amplia gama de temas de índole económica, laboral, social, educativa, sanitaria y de uso del tiempo, entre muchos otros. Las ONE tienen la responsabilidad de garantizar que las estimaciones de los indicadores de interés obtenidos a través de las encuestas de hogares sean lo más robustas posible.

Por lo anterior, se hace imperativo contar con lineamientos que guíen o sirvan de referencia tanto a las ONE como a otros entes productores de estadísticas oficiales y grupos usuarios en general sobre cómo medir la calidad estadística de las estimaciones obtenidas mediante encuestas de hogares por muestreo¹.

Aunque tanto en América Latina y el Caribe como en el mundo hay diversas encuestas de hogares que tienen objetivos distintos, se observan similitudes en cuanto a la estandarización de los conceptos, la metodología, los procesos y los desafíos que se enfrentan. Por ese motivo, es posible y sumamente conveniente contar con métodos y lineamientos transversales que se puedan aplicar a toda la región.

En línea con lo anterior, en la Conferencia Estadística de las Américas de la Comisión Económica para América Latina y el Caribe (CEA-CEPAL) se definió un grupo de trabajo que en el bienio 2022-2023 tendría el objetivo general de elaborar recomendaciones destinadas a armonizar el análisis de la calidad de las cifras que surgían de las encuestas de hogares y a medir el error de muestreo aplicando métodos que fueran asequibles para todos los grupos usuarios y se pudieran aplicar a todo tipo de estimadores. De conformidad con el objetivo general, se plantearon cuatro objetivos específicos, a saber:

- i) Sistematizar el estado en que las prácticas de publicación y disseminación de cifras provenientes de las encuestas de hogares por muestreo se encontraban en ese momento.
- ii) Establecer un procedimiento estandarizado regional en que se aplicaran recomendaciones y buenas prácticas internacionales para medir la calidad y la precisión de las estimaciones resultantes del procesamiento y el análisis de las encuestas de hogares por muestreo. El procedimiento debía poder adaptarse a la realidad específica de cada país y cada encuesta.
- iii) Examinar las características particulares de los procesos de anonimización de las bases de datos y su repercusión en la forma en que los grupos usuarios de las encuestas de hogares medían el error de muestreo.
- iv) Definir un conjunto de métodos que fueran apropiados para estimar el error de muestreo con bases de datos anonimizadas de encuestas de hogares en que se utilizaran pesos replicados.

Al comienzo del presente documento se presenta un diagnóstico sobre las prácticas que las ONE de la región aplican a la hora de medir la calidad de las estimaciones, particularmente las que surgen de las encuestas de hogares por muestreo. A continuación se proporciona un marco conceptual relacionado con el ámbito de estudio y se profundiza sobre todo en las técnicas de pesos replicados destinadas a estimar los errores muestrales. Sobre la base de la revisión bibliográfica realizada, se exponen con cierto orden lógico los diversos criterios de calidad que se pueden aplicar para determinar la calidad estadística de las estimaciones, y se propone un flujo de aplicación de los criterios de calidad estudiados con el objeto de que se constituya en la referencia estándar a la hora de evaluar la calidad estadística de las estimaciones de las ONE.

¹ Las encuestas probabilísticas dirigidas a los hogares son un caso particular de las encuestas de hogares por muestreo, que constituyen una manera práctica de obtener datos actualizados sobre las condiciones y tendencias sociales, el comportamiento socioeconómico de los hogares, el acceso a los servicios básicos y el efecto de los programas sociales. Son instrumentos particularmente adecuados para crear bases de datos sobre los hogares, y abren múltiples posibilidades de establecer interrelaciones y realizar análisis utilizando los sistemas actuales de procesamiento de datos.

Capítulo I

Antecedentes de aplicación de estándares de calidad por parte de las Oficinas Nacionales de Estadística

Uno de los principios del Código Regional de Buenas Prácticas en Estadísticas para América Latina y el Caribe (CEPAL, 2011) es el compromiso con la calidad, según el cual las ONE, regidas por el sistema estadístico nacional, deben trabajar y cooperar conforme a normas, principios y estándares internacionales. A los efectos de cumplir ese precepto, en la región ha habido diversas instancias de trabajo destinadas a profundizar en los temas asociados al aseguramiento de la calidad en la producción estadística.

En 2022 se publicó la *Guía para la implementación del marco de aseguramiento de la calidad para procesos y productos estadísticos*, trabajo realizado de manera conjunta por las ONE de Colombia y México con el asesoramiento de la Secretaría Técnica de la División de Estadísticas de la CEPAL. Esa guía tiene por objeto orientar a los países de América Latina y el Caribe en la adopción y aplicación del *Manual del marco nacional de aseguramiento de la calidad para las estadísticas oficiales* de las Naciones Unidas, cuyos principios de calidad y requisitos asociados están organizados en los cuatro niveles que se indican a continuación (CEPAL, 2022):

- i) Nivel A, que se refiere a la gestión del sistema estadístico, en la que se consideran los principios fundamentales de coordinación del sistema estadístico nacional, la gestión de las relaciones y la gestión de las normas.
- ii) Nivel B, en que se contempla la gestión del entorno institucional sobre la base de los principios de independencia profesional, objetividad, transparencia, confidencialidad, compromiso con la calidad y suficiencia de recursos.
- iii) Nivel C, que alude a la gestión del proceso estadístico, que se debe regir por los principios de solidez metodológica, buena relación costo-efectividad, procedimientos apropiados y manejo de la carga del encuestado.
- iv) Nivel D, que se refiere a la gestión de los resultados estadísticos, en la que se consideran los principios fundamentales de relevancia, precisión y confiabilidad, oportunidad y puntualidad, accesibilidad y claridad, coherencia y comparabilidad, y gestión de metadatos.

El presente documento tiene por objeto profundizar en los niveles C y D en particular, a fin de contribuir a que en las estimaciones provenientes de las encuestas probabilísticas dirigidas a los hogares se apliquen los principios de solidez metodológica, procedimientos apropiados, y precisión y confiabilidad.

En línea con lo anterior, y a fin de alcanzar los objetivos planteados, el grupo de trabajo comenzó por identificar los procesos que las ONE de la región aplicaban a la hora de medir la calidad de las estimaciones, específicamente de las obtenidas mediante encuestas de hogares por muestreo. Con ese fin realizó una consulta a 11 países de la región para obtener información detallada sobre diferentes aspectos asociados a la medición de la calidad estadística de las estimaciones de interés, por ejemplo, sobre los principios y lineamientos que se aplicaban para medir la calidad de las estimaciones de los parámetros, las métricas que se empleaban, los umbrales de decisión, la clasificación de las estimaciones y los métodos de anonimización.

En lo que respecta a los lineamientos sobre la aplicación de criterios de calidad y la determinación de la usabilidad de las estimaciones de los parámetros de interés, se contemplan los definidos por las propias ONE y los basados en estándares internacionales. Entre las ONE que cuentan con lineamientos propios y documentación al respecto se encuentran las de la Argentina², Bolivia (Estado Plurinacional de) (ANDA, 2020), Chile (INE, 2020a), El Salvador, el Uruguay (INE, 2021) y el Brasil. En este último, no obstante, aún se está elaborando la versión más actualizada, y en El Salvador y el Estado Plurinacional de Bolivia no hay ningún documento oficial que haga público el estándar.

En el Catálogo Central de Datos y Microdatos del Archivo Nacional de Datos (ANDA) del Instituto Nacional de Estadística (INE) del Estado Plurinacional de Bolivia se establece la política de difusión, y en ella se hace una recomendación sobre la calidad estadística de las estimaciones que realice el usuario. En ese apartado se establecen los criterios que se deben emplear con ese fin, y se señala la importancia de tener en cuenta el coeficiente de variación y la cantidad de observaciones respecto de las cuáles se obtuvieron respuestas. Se recomienda utilizar programas estadísticos que permitan calcular los errores muestrales aplicando el diseño de muestreo que se haya utilizado en la encuesta, así como también incluir el valor del coeficiente de variación (CV) cuando se realicen estimaciones. Además, se establece un estándar para evaluar los errores muestrales de los indicadores principales y a niveles geográficos más desagregados. Ese estándar se basa en el CV y en las unidades muestrales del nivel desagregado, y en él se clasifican las estimaciones en cinco categorías: estimación óptima, estimación confiable con muy buena precisión, estimación con suficiente precisión, estimación con precisión regular y estimación no significativa (ANDA, 2020)³.

En 2020, el INE de Chile publicó dos documentos: “Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares” y “Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares”. En el primer documento se describe en detalle el análisis metodológico que el INE lleva a cabo para evaluar las medidas de calidad de las estimaciones de los parámetros de interés consideradas en el estándar, así como el marco conceptual, el uso de las medidas de calidad, y los criterios utilizados en las encuestas del instituto y en otras ONE. En ese documento el INE plantea que para determinar la precisión de los estimadores de proporción o razón no se utilice el CV sino una función cuadrática que permite evaluar el comportamiento del error estándar. Además, presenta una propuesta de lineamientos a través de una serie de flujogramas que orientan a los grupos usuarios a la hora de tomar decisiones en el momento de analizar y publicar la información (INE, 2020b). El segundo documento es un compendio del primero y en él se presentan de forma concisa los conceptos y los criterios utilizados en las estimaciones. También se incorpora el flujograma de aplicación de los criterios que se han de emplear para determinar la calidad de las estimaciones de los parámetros de interés y la de los tabulados correspondientes (INE, 2020a). En línea con esos documentos, en 2022 el INE de Chile publicó un paquete llamado “calidad” en lenguaje R (CRAN, 2023); dicho paquete permite evaluar la precisión de distintos indicadores que provienen de las encuestas de hogares considerando los estándares establecidos por la institución y por la CEPAL, lo que permite que las instituciones de la región también lo utilicen.

El Instituto Brasileño de Geografía y Estadística (IBGE), por su parte, cuenta con una guía sobre la divulgación de los errores muestrales de las encuestas por muestreo probabilístico que ese instituto realiza (IBGE, 2021a). En la guía se presentan las formas de difundir los errores de muestreo asociados a las estimaciones de los parámetros de interés obtenidas mediante encuestas por muestreo realizadas de conformidad con los fundamentos de la segunda edición del código de buenas prácticas de las estadísticas del IBGE, que se publicó en 2021 (IBGE, 2021b). También se proporcionan directrices sobre cómo hacer pública la información sobre la calidad de esas estimaciones, específicamente sobre su exactitud, con el objetivo de que los grupos usuarios puedan evaluar y utilizar las cifras producidas por el IBGE de manera adecuada. La guía contiene un marco conceptual relativo a los términos asociados a la calidad estadística de las estimaciones, y ofrece directrices sobre cómo publicar los datos sobre la precisión de estas. En la actualidad el IBGE está elaborando un nuevo documento de recomendaciones relativas a todas sus encuestas en el que se abordará la calidad estadística de las estimaciones de los parámetros de interés.

² Véanse las notas técnicas 1 a 4 del Instituto Nacional de Estadística y Censos (INDEC, s.f.).

³ El Estado Plurinacional de Bolivia no es el único país en que se emplea este sistema de archivos de metadatos basados en la Iniciativa de Documentación de Datos, sino que es bastante habitual que otras instituciones y ONE también lo utilicen. Véase ANDA (2023).

En junio de 2022, bajo la orientación de la CEPAL y de acuerdo con sus últimas recomendaciones (Gutiérrez y otros, 2020), la Dirección General de Estadística y Censos (DIGESTYC) de El Salvador publicó el documento “Recomendaciones sobre criterios de supresión para investigaciones por muestreo probabilístico,” que de forma sencilla y simplificada aporta directrices destinadas a orientar a los distintos grupos usuarios a la hora de evaluar la calidad de las estimaciones de los parámetros de interés (DIGESTYC, 2022).

Desde 2021 el INE del Uruguay cuenta con la Norma Técnica de Certificación de la Calidad de Operaciones Estadísticas INECCOE:2021-01 (INE, 2021), en que se aborda la calidad de la producción estadística con base en las dimensiones propuestas por la Oficina Estadística de la Unión Europea (Eurostat), a saber: relevancia, precisión y exactitud, oportunidad y puntualidad, accesibilidad y transparencia, comparabilidad y coherencia, y capacidad institucional. Para elaborar el documento se utilizó como marco de referencia el Modelo Genérico de Procesos Institucionales Estadísticos (MGPIE) de la Comisión Económica para Europa (CEPE). En la norma técnica se establece un conjunto de indicadores relativos a cada dimensión de la calidad, y esos indicadores se miden mediante estándares. El grado de cumplimiento de los estándares determina el nivel de desarrollo de cada indicador, es decir, cuánto se aproxima la operación estadística evaluada al nivel de calidad ideal. El nivel de desarrollo se mide en una escala que va del 1 al 4, donde 1 es el nivel más bajo de calidad esperado y 4 es el nivel máximo. En lo que atañe a la calidad de las estimaciones de los parámetros de interés, en la norma se establece que se deben calcular los errores de muestreo relacionados con las variables principales. Entre las métricas que se mencionan se encuentran el error estándar, el coeficiente de variación, los intervalos de confianza, el error cuadrático medio y los errores de diseño. Respecto de la calidad de las estimaciones, se evalúa la disponibilidad de las medidas del error de muestreo según que el dato esté disponible o no en relación con las variables principales o con el resto de las variables, y se otorga una puntuación de acuerdo con la precisión de las estimaciones de los parámetros de interés en función de los valores del CV. Si la precisión de las variables es baja, se debe proporcionar una explicación (INE, 2021).

Las ONE de la región que a la fecha no han publicado ni compartido con la comunidad documentos sobre estándares de calidad relativos a las estimaciones de los parámetros de interés declaran que se basan en lineamientos formulados por Eurostat, la Oficina de Estadísticas del Canadá, la Oficina del Censo de los Estados Unidos y la CEPAL.

Otro aspecto que se consideró en la consulta regional sobre la calidad de las estimaciones fue la desagregación de la información, es decir, los dominios de estudio garantizados en el diseño y todos los demás dominios de interés en relación con los cuales se producen estimaciones. Lo anterior se sustenta en los Objetivos de Desarrollo Sostenible (ODS), debido a que las ONE deben velar por la representatividad de toda la población considerando su vasta diversidad. En términos prácticos, las ONE deben trabajar en pro de que sus estadísticas oficiales sigan siendo robustas cuando se desagregan según las características relevantes de la población, por ejemplo, el ingreso, el sexo, la edad, la raza, la etnia, el estatus migratorio, la discapacidad y la ubicación geográfica.

Como se puede observar en el diagrama I.1, los dominios de estudio y, por tanto, aquellos en que se garantiza la representatividad de las estimaciones, generalmente obedecen a la división político-administrativa de los países considerando diferentes escalas geográficas. Lo más usual es que se hagan estimaciones a nivel nacional, por tipo de zona (urbana y rural) y por departamento (y sus diversos homólogos, por ejemplo, región, comuna, estado y otros).

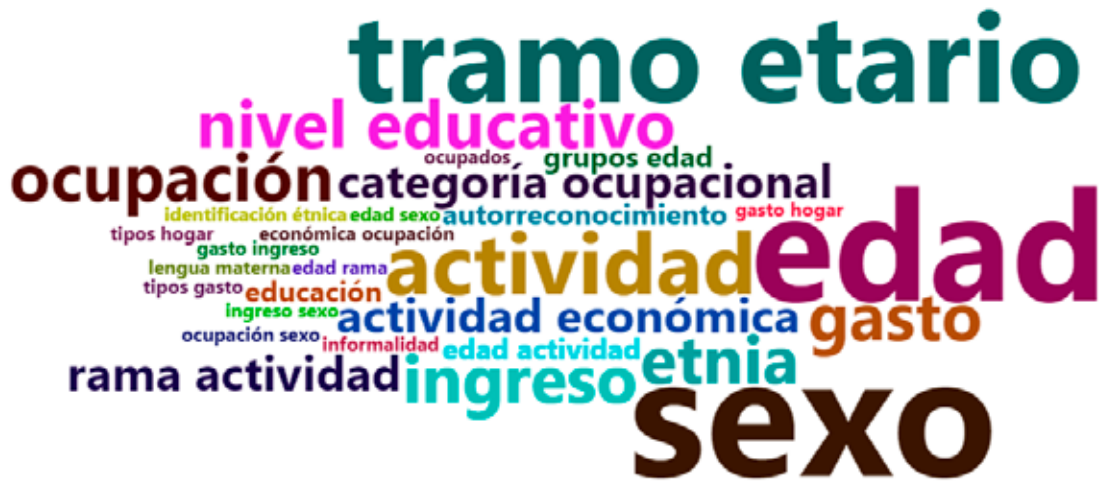
En el diagrama I.2, por otra parte, se presentan los demás dominios de interés, es decir, aquellos respecto de los cuales, en la mayoría de los casos, no se garantiza la robustez de las estimaciones. Esos dominios son los más susceptibles de carecer de calidad estadística, y entre los más comunes se encuentran el sexo, el grupo etario, el nivel de ingreso, la actividad, el gasto y otros.

Diagrama I.1
Dominios de estudio



Fuente: Comisión Económica para América Latina y el Caribe (CEPAL) e Instituto Nacional de Estadísticas de Chile (INE).
Nota: Se consideran 30 operaciones estadísticas.

Diagrama I.2
Otros dominios de interés



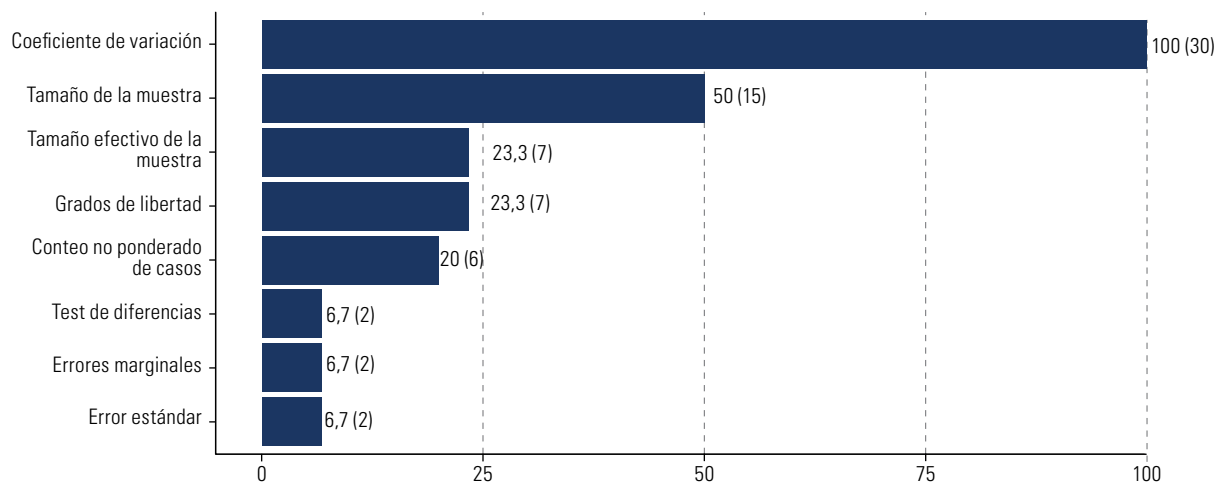
Fuente: Comisión Económica para América Latina y el Caribe (CEPAL) e Instituto Nacional de Estadísticas de Chile (INE).
Nota: Se consideran 30 operaciones estadísticas.

Según la consulta regional, y sobre la base de las 30 operaciones estadísticas que declararon las 11 instituciones tomadas en su conjunto, hay dos métricas que se emplean con frecuencia para evaluar la calidad de las estimaciones: el CV (que se utiliza en todas las operaciones estadísticas señaladas) y el tamaño muestral⁴. Otros elementos que se suelen emplear, pero en menor medida, son el tamaño efectivo de la muestra, los grados de libertad y el conteo no ponderado de casos (véase el gráfico I.1)⁵. Las diferencias observadas entre los criterios empleados por lo general vienen determinadas por el tipo de estimador (total, media, razón o percentil).

Gráfico I.1

Frecuencia de uso de las métricas de calidad

(En porcentajes)



Fuente: Comisión Económica para América Latina y el Caribe (CEPAL) e Instituto Nacional de Estadísticas de Chile (INE).

Nota: El gráfico representa las respuestas obtenidas a una pregunta de selección múltiple sobre la base de 30 operaciones estadísticas. Los números entre paréntesis representan el número de operaciones estadísticas en que se utiliza la métrica.

Respecto a los umbrales definidos para evaluar cada estimación se observa heterogeneidad en las respuestas de las ONE. En el caso del CV, el valor más usual es el 15% y, con menor frecuencia, el 20% y el 30%. En algunos casos, el CV se evalúa mediante intervalos que permiten asignar a las estimaciones distintos niveles de fiabilidad, precisión o calidad.

En cuanto a la evaluación del tamaño muestral, se identifican tres umbrales: dos relativos a las unidades elementales (20 y 60 unidades), y otro relativo a la cantidad de conglomerados que hay en la muestra lograda (dos conglomerados como mínimo)⁶. En cuanto a los grados de libertad, se indica que nueve grados es el valor mínimo necesario para garantizar la calidad de la estimación. En cuanto al conteo no ponderado de casos se declara que, si hay menos de 25 casos, la estimación no se publica y, si hay entre 25 y 49 casos, se presenta la cifra entre paréntesis a modo de alerta sobre la calidad de la estimación.

Como se ha mencionado en los párrafos previos, según la evaluación de los criterios de calidad también se establece una clasificación de las estimaciones. En algunos casos, en la clasificación se consideran cuatro categorías, por ejemplo, muy buena, buena, aceptable y referencial. En otros casos se consideran tres categorías (fiable, poco fiable y no fiable) o dos categorías (fiable o no fiable). No todas las ONE indicaron que emplearan ese tipo de clasificaciones.

⁴ El tamaño muestral se refiere al número de unidades que se tomaron en cuenta al calcular la estimación del parámetro de interés.

⁵ El tamaño efectivo de la muestra es el cociente entre el número de unidades que se tomaron en cuenta al calcular la estimación y el efecto del diseño.

⁶ Si se supone un muestreo bietápico en que las unidades secundarias son las viviendas, los conglomerados en la muestra lograda son todos aquellos en que se logró al menos una vivienda. A su vez, una vivienda lograda es aquella en que al menos se han contestado las preguntas básicas de identificación del registro de personas del hogar y otro conjunto mínimo de preguntas relacionadas con las variables principales, según los criterios propios de cada encuesta. Dos conglomerados es el valor mínimo que se necesita para calcular una varianza, sea cual fuere el dominio de estudio.

Es importante considerar que esta consulta se realizó luego de la pandemia de enfermedad por coronavirus (COVID-19), período en que las tasas de respuesta de las encuestas de hogares se vieron afectadas de manera importante. Se consultó a los países si fue necesario readaptar los criterios de calidad durante esa época. En general, la medida que adoptaron la mayoría de las ONE fue la de suprimir las estimaciones correspondientes a los niveles más desagregados debido a la reducción del tamaño muestral logrado.

Un segundo aspecto importante en cuanto a los objetivos del grupo de trabajo se refiere a que todas las personas usuarias, tanto las externas como las que pertenecen a las ONE, tengan acceso a la información que figura en las bases de microdatos que las ONE publican, para que puedan evaluar la calidad de las estimaciones de los parámetros de interés⁷. Para que eso ocurra se hace imperativo contar con lineamientos claros y precisos que permitan respetar a cabalidad los diseños muestrales con los que las encuestas fueron concebidas y que, a la vez, sean simples de ejecutar para los grupos usuarios.

En cumplimiento del principio de confidencialidad estadística, es una práctica común entre las ONE que las bases de microdatos publicadas se sometan a procesos de anonimización, lo que en ciertos casos puede conllevar la supresión de variables necesarias para estimar el error muestral. Por otro lado, hay estimadores que permiten calcular el error de muestreo mediante cálculos que son muy complejos incluso para el personal especializado. Por esos motivos, en esta consulta también se indagó sobre los métodos que se utilizaban para estimar los errores muestrales y para anonimizar los datos, a fin de dilucidar si efectivamente todos los usuarios de los microdatos tenían la posibilidad de estimar los errores de muestreo. Se constató que, en el 96,7% (29) de las operaciones estadísticas que las ONE declararon, las bases de microdatos se publicaban. Se constató asimismo que en el 90% de esas operaciones (26 de las 29) se aplicaban métodos de enmascaramiento de datos y anonimización, por ejemplo, eliminación de las variables que contenían datos personales o sensibles, como los identificadores de la vivienda. En casos menos frecuentes las ONE de la región optaban por agrupar categorías de variables que contenían información sensible, como la nacionalidad, la edad o la ocupación, otras opciones incluían limitar la desagregación geográfica, y por truncar los valores en variables como el salario.

En cuanto a los métodos de estimación de los errores muestrales, los más destacados son el método del último conglomerado, la varianza del estimador calibrado y los métodos de remuestreo, como el *bootstrap* o el método de la navaja (*jackknife*), entre otros. En ese sentido se constató que, contrario a lo que se creía, varias ONE de la región ya habían implementado métodos de remuestreo, que, según algunos expertos, representan la mejor opción a la hora de estimar los errores de muestreo, pues se pueden aplicar a cualquier tipo de estimador. Además, si los pesos replicados se ponen a disposición de las personas y de las unidades usuarias o investigadoras, resulta factible que toda persona usuaria con conocimientos básicos de estadística pueda estimar el error de muestreo de las estimaciones de los parámetros de interés, aunque se apliquen métodos de anonimización a las bases de datos. Eso ofrece la posibilidad de evaluar la calidad de la información y de hacer un uso responsable de ella.

En cuanto a los avances tecnológicos, en la región también se han dado pasos importantes respecto a la automatización de los procesos de cálculo. El INE de Chile ha confeccionado un paquete de RStudio para evaluar la calidad de las estimaciones de los parámetros de interés (INE, 2020a; CRAN, 2023)⁸. Ese paquete permite evaluar las estimaciones obtenidas a partir de encuestas por muestreo aplicando un estándar propio y también el que ha propuesto la CEPAL. Una de las virtudes del paquete es que ofrece la posibilidad de modificar los umbrales relativos a cada métrica evaluada según el estándar seleccionado. Al mismo tiempo, el Instituto Nacional de Estadística y Censos (INDEC) de la Argentina ha diseñado una aplicación llamada CEMRepBoot para aplicar métodos de remuestreo en sus encuestas. Esa aplicación permite calcular los errores de muestreo asociados a las estimaciones de los parámetros poblacionales, y ofrece recomendaciones a los grupos usuarios acerca de las estimaciones, los cálculos y el uso con fines estadísticos de los datos de las encuestas (MTESS/OISS/SRT, 2018).

⁷ En el contexto de las encuestas de hogares, la base de microdatos es una base de datos a nivel de las unidades de información en estudio, es decir, los hogares o las personas. Corresponde también a una estructura tabular-matricial de filas que representan a las personas, y de columnas que representan las diversas características o variables capturadas mediante un cuestionario y asociadas a cada sujeto.

⁸ Véase [en línea] <http://www.rstudio.com/>.

Capítulo II

Marco conceptual

Al diseñar un estudio, es fundamental tener en cuenta el objetivo de la investigación y las preguntas específicas que se desea responder, así como los recursos disponibles y las limitaciones prácticas. Elegir el enfoque de muestreo adecuado es crucial para garantizar la validez y la calidad de los resultados, y la eficiencia del proceso de investigación. En el caso de las encuestas de hogares, la información de todas las personas tiene la misma relevancia, por lo que se utilizan encuestas probabilísticas con el propósito de inferir sobre la población total⁹.

Los enfoques enunciados en los párrafos precedentes son importantes en la investigación y el análisis de los datos, y se complementan para ofrecer una visión más completa y detallada en diferentes escenarios de estudio. En la práctica, tanto el muestreo probabilístico como el no probabilístico tienen sus ventajas y desventajas, y la elección del método adecuado dependerá del contexto de la investigación, los objetivos del estudio y las limitaciones prácticas. Al combinar estos enfoques en un diseño de muestreo mixto, es posible abordar de manera efectiva una amplia variedad de preguntas de investigación y obtener un conocimiento más profundo y completo sobre el fenómeno en estudio. Las oficinas de estadística por lo general utilizan muestreos probabilísticos, ya que estos permiten lograr la representatividad poblacional y eliminar sesgos de selección.

Este capítulo tiene por objeto introducir los conceptos que engloban y dan origen a la calidad estadística de una cifra estimada mediante una encuesta por muestreo probabilístico. Se abarcará desde el diseño de la encuesta hasta el análisis de los resultados obtenidos, y el análisis de la calidad se enfocará en este último.

A. Encuestas por muestreo y sus características

Las encuestas por muestreo deben diseñarse con el fin de obtener con eficiencia información estadística de calidad¹⁰. Por ese motivo, en este apartado se proporcionan ejemplos ilustrativos de las distintas características que debe tener el diseño de la muestra y se brindan referencias bibliográficas que se pueden consultar para profundizar en las condiciones técnicas de los diseños abordados. Se señala, además, la importancia de velar por la calidad estadística de las estimaciones de los parámetros de interés desde el momento en que se formula el diseño de la encuesta.

La realización de las encuestas por muestreo puede enmarcarse en el Modelo Genérico de Procesos Institucionales Estadísticos (MGPIE) propuesto por la Comisión Económica para Europa (CEPE)¹¹. El MGPIE es un marco conceptual en que se describen las diferentes etapas y actividades involucradas en la producción de estadísticas oficiales y otros productos estadísticos. Este modelo genérico de proceso estadístico es una herramienta útil para ayudar a las distintas oficinas de estadística a comprender, diseñar, gestionar y mejorar los procesos de producción de estadísticas. El MGPIE 5.1 (segmento negocio) consta de ocho procesos, cada uno de los cuales está integrado por una serie de subprocesos (véase el diagrama II.1). Los procesos se centran

⁹ En las encuestas que se llevan a cabo en establecimientos se utilizan muestreos no probabilísticos para asegurar que en todos los levantamientos se capturen las unidades económicas que más participan en las actividades del sector.

¹⁰ Las encuestas por muestreo son aquellas en que se utilizan métodos de selección aleatoria para extraer las unidades que serán parte de la muestra a partir de una lista que contiene todas las unidades de muestreo, que generalmente se denomina marco de selección. De esta forma, la selección de cada unidad de muestreo tiene una probabilidad conocida distinta de cero de pertenecer a la muestra, y por ese motivo ese método también se denomina método de selección probabilística.

¹¹ Desde noviembre de 2013, el Comité de Modernización de Estándares, que depende del Grupo de Alto Nivel para la Modernización de la Producción y los Servicios Estadísticos, es quien se encarga del MGPIE. La versión original de este se puede consultar en la página web de la CEPE (véase [en línea] <https://unece.org/statistics/modernstats/gsbpm>).

en la planificación y gestión de los aspectos empresariales de la producción estadística, y proporcionan una guía detallada sobre la identificación de necesidades, el diseño del programa estadístico, la preparación de los recursos y el establecimiento de marcos de coordinación. El MGPIE ha sido adaptado por los institutos nacionales de estadística de diferentes países de América Latina, como Chile, Colombia, México y el Uruguay, entre otros.

Diagrama II.1

Procesos y subprocesos del modelo MGPIE 5.1: segmento negocio



Fuente: Comisión Económica para Europa (CEPE), "Generic Statistical Business Process Model (GSBPM) version 5.1", Ginebra, 2019.

Para definir el diseño de una muestra correspondiente al subproceso 2.4 del diagrama II.1 debe considerarse lo siguiente: la construcción del marco de muestreo, la elección del tipo de muestreo, la determinación del tamaño de la muestra para el muestreo probabilístico, la definición de los estimadores para el muestreo probabilístico, la identificación de las fuentes de error total de muestreo y, finalmente, la documentación del diseño de la muestra, en donde se proporcionan los detalles de cada uno de los pasos. En lo que resta de esta sección se brindará información sobre los primeros dos pasos, a saber, el marco y el tipo de muestreo.

1. Marco de muestreo

Un marco de muestreo o marco muestral es un listado de todas las unidades de la población objetivo; a partir de ese marco se pueden extraer una o varias muestras. El marco contiene información sobre las principales características de las unidades que forman parte de la población objetivo, así como los datos que permiten geolocalizarla, ubicarla o contactarla. Puede construirse utilizando diversas fuentes de información, como información censal, registros administrativos tributarios o comerciales, información cartográfica e información geoespacial, entre otras.

Al construir el marco de muestreo se debe velar por que cada una de las unidades del universo de estudio esté incluida una única vez, para asegurar que la muestra sea representativa de la población total. En el caso de las encuestas periódicas es fundamental establecer los mecanismos que permitan actualizar el marco de muestreo, y en ellos se debe considerar la creación de nuevas unidades y la eliminación de las que dejan de pertenecer al universo de estudio. Los marcos de muestreo generalmente son de dos tipos: marcos de área y marcos de lista.

Los marcos de área suelen estar formados por una jerarquía de unidades geográficas, es decir, las unidades de un nivel se pueden subdividir para formar las unidades del siguiente nivel. Todos los elementos incluidos en el marco constituyen la población de este. Las discrepancias entre la población de la encuesta y la población del marco se denominan errores de cobertura (Oficina de Estadística del Canadá, 2003, pág. 17)¹².

Los marcos de lista, por su parte, son una técnica de muestreo que se utiliza en estadística y en encuestas para identificar una población y seleccionar una muestra representativa de ella. Es una lista completa y detallada de los elementos de la población que se desea estudiar o encuestar (Kish, 1965). El marco de lista es esencial para obtener una muestra representativa, ya que proporciona una enumeración completa de los elementos de la población que se utilizarán en la selección de la muestra. Los elementos de la población pueden ser personas, empresas, hogares, objetos u otros. La lista también puede incluir información relevante sobre cada elemento, como su nombre, ubicación, tamaño y características demográficas, entre otros.

2. Tipos de muestreo

Según la información que se desee obtener y las características del universo de estudio, se debe elegir el tipo de muestreo que permita garantizar que los datos captados en la muestra sean representativos de la población total, y que la captación se realice de forma tal que asegure el uso eficiente de los recursos económicos y técnicos.

El muestreo puede ser probabilístico o no probabilístico. El primero ofrece mejores opciones para captar la información de toda la población de estudio, sobre todo en el caso de las encuestas de hogares, y es adecuado para obtener información representativa de una población completa. En los muestreos probabilísticos, cada elemento de la población tiene una probabilidad conocida y no nula de ser seleccionado en la muestra. Ese tipo de muestreo permite hacer estimaciones precisas de los parámetros poblacionales y garantizar la validez estadística de los resultados. Además, es posible medir el error de muestreo y definir intervalos de confianza en torno a las estimaciones, lo que proporciona una medida de la precisión de los resultados.

El muestreo no probabilístico, por su parte, es de gran utilidad en las encuestas de empresas, donde la participación de las unidades de observación en las diferentes variables económicas determina la importancia de que algunas sean captadas todo el tiempo y otras no: las unidades que más aportan a las variables de interés se pueden incluir de forma no probabilística, y las que aportan menos, de forma probabilística. Ese tipo de muestreo es útil cuando se requiere un enfoque más específico y dirigido hacia ciertos grupos o sectores, pues permite enfocarse en unidades de observación específicas y más relevantes en ciertos contextos. En general, quien lleva a cabo la investigación establece ciertos criterios para seleccionar algunos elementos de la población sin tener en cuenta la probabilidad de selección. A diferencia del muestreo probabilístico, este método puede dar lugar a sesgos en los resultados e impedir que las conclusiones se puedan generalizar a la población en su conjunto. Por ese motivo, para diseñar indicadores cuya principal utilidad sea definir políticas públicas se suele preferir el muestreo probabilístico en lugar del no probabilístico. Sin embargo, este último puede ser más eficiente en cuanto al tiempo y los recursos, y puede ser adecuado para estudios exploratorios o cualitativos, o para aquellos en que se busca información detallada sobre casos específicos.

A continuación se brinda información más detallada sobre cada tipo de muestreo.

¹² El error de cobertura se refiere a la diferencia entre la población objetivo y la muestra utilizada en una encuesta o estudio de investigación. Ese error puede ocurrir de dos maneras: i) sobrecobertura, que surge cuando en la muestra se incluyen personas que no forman parte de la población objetivo, y ii) subcobertura, que surge cuando algunas personas de la población objetivo no se incluyen en la muestra.

a) Muestreo probabilístico

El muestreo probabilístico es útil para realizar estimaciones de indicadores en la población objeto de estudio. Dado que cada unidad se selecciona aleatoriamente, este tipo de muestreo ofrece varias ventajas, por ejemplo, elimina el sesgo de selección (Särndal, Swensson y Wretman, 1992), y permite utilizar las probabilidades de inclusión de las unidades para hacer estimaciones confiables y calcular el error estándar asociado, lo que a su vez permite hacer inferencias acerca de la población objeto de estudio (Statistics Canada, 2003, pág. 91). A continuación se describen algunas de las modalidades de este tipo de muestreo.

i) Muestreo aleatorio simple

Se asigna un número a cada una de las personas que integran la población objeto de estudio, cuyo tamaño es N , y esos números sirven de identificador único. Se eligen tantos sujetos como sea necesario para completar el tamaño de la muestra n , con $n \leq N$, por medio de un mecanismo de selección aleatoria en que se extrae un conjunto de n números entre 1 y N . Cada número tiene la misma probabilidad de ser extraído. Este procedimiento tiene poca o nula utilidad práctica cuando la población objeto de estudio es muy grande; además, tiene un costo muy alto y supone un arduo proceso operativo. Sin embargo, constituye una referencia que se puede utilizar para comparar otras modalidades de muestreo. Existen dos tipos de selección: con reemplazo y sin reemplazo. La entropía o incertidumbre del muestreo aleatorio simple es alta, es decir, es muy difícil predecir el tipo de muestra que se obtendrá. El concepto de entropía es útil a la hora de estimar la varianza de los estimadores muestrales. Cuando la entropía de un diseño muestral es alta, es posible obtener una aproximación de las probabilidades de segundo orden en términos de las probabilidades de inclusión de primer orden, lo que es necesario para estimar la varianza (Tillé y Haziza, 2010).

ii) Muestreo sistemático

Todos los elementos de la población objeto de estudio se numeran de forma ascendente (de 1 a N), pero, en lugar de extraer n números aleatorios, solo se extrae uno, entre 1 y k , $k=N/n$, donde N es el tamaño de la población objeto de estudio. Se parte de ese número aleatorio para elegir, a intervalos constantes, todos los demás números hasta completar la muestra. Cuando la población objeto de estudio está en un orden aleatorio con respecto a la variable de interés, este método es equivalente al muestreo aleatorio simple, pero su uso facilita la extracción de la muestra. Por otro lado, cuando el orden de la población objeto de estudio tiende a presentar cambios paulatinos de la variable de interés, esta modalidad produce varianzas estimadas menores que el muestreo aleatorio simple. Esto se debe a que la muestra queda más dispersa sobre la población objeto de estudio, es decir, a que la muestra es más representativa. Este tipo de muestreo puede utilizarse, por ejemplo, en encuestas de opinión sobre la calidad del servicio de una empresa, donde los clientes se seleccionan de manera sistemática a partir de una base de datos.

iii) Muestreo estratificado

La población objeto de estudio se divide en estratos, es decir, en clases homogéneas y mutuamente excluyentes, como grupos de edad, sexo y otros. Una vez hecho esto, la muestra se distribuye de acuerdo con distintos métodos de afijación o distribución (igual en todos los estratos, proporcional al tamaño de cada estrato, distribución de Neyman y distribución óptima). Con esto se asegura que todos los estratos de interés estén representados en la muestra. En cada estrato de muestreo las muestras se seleccionan de forma independiente, por lo general utilizando el mismo método de selección, excepto en los casos en que se crean estratos especiales. Un ejemplo de estos últimos son los de inclusión forzosa, en que todos los elementos se seleccionan con probabilidad 1. Ese tipo de caso generalmente ocurre en las encuestas de empresas o establecimientos en que muy pocas empresas grandes tienen una participación considerablemente grande en las ventas o en el número de trabajadores. En las encuestas de hogares, que es el foco de este documento, los estratos en la mayoría de los casos corresponden a áreas geográficas configuradas según la división política administrativa de cada país: región, departamento, provincia y otros.

iv) Muestreo por conglomerados mono- y multietápico

Es un tipo de muestreo en que la población objeto de estudio se divide en subpoblaciones llamadas conglomerados, cuyos elementos poseen cierta característica, cualidad o atributo. Un ejemplo de esos conglomerados son las agrupaciones de viviendas ubicadas en dos o más manzanas de un área geográfica delimitada. El muestreo por conglomerados se utiliza cuando no se cuenta con un marco de muestreo completo de la población objeto de estudio o cuando es muy caro producirlo, y cuando es necesario reducir el costo de la recolección de datos. Eso ocurre principalmente cuando la población objeto de estudio está muy dispersa en un área.

Hay dos tipos de muestreo por conglomerados: de una sola etapa o monetápico, y por etapas o multietápico. El primero consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido), y luego tomar todos los elementos de cada conglomerado elegido como parte de la muestra. Cuando los conglomerados son áreas geográficas, este tipo de muestreo se suele denominar muestreo por áreas. El muestreo por conglomerados monoetápico es menos preciso que el muestreo aleatorio simple o estratificado debido a la homogeneidad inherente a las unidades de muestreo que pertenecen a los conglomerados seleccionados, ya que las unidades son físicamente cercanas y suelen tener características similares, es decir, la selección de dos o más unidades dentro del mismo conglomerado puede dar lugar a información redundante, y eso incrementa los errores muestrales de los indicadores objetivo.

El muestreo por conglomerados por etapas o multietápico se da cuando se avanza en el análisis y, dentro de cada conglomerado, se seleccionan nuevas unidades de manera aleatoria y ese proceso se repite sucesivamente hasta llegar a las unidades finales. La selección de una muestra puede ser un proceso complejo en ciertos escenarios. En algunos casos puede ser difícil localizar cada unidad de observación, lo que lleva a una selección por etapas. Además, cuando la población es demasiado grande, se vuelve ineficiente seleccionar una muestra directamente de ella, por lo que resulta conveniente formar conglomerados agrupando unidades de observación que comparten características similares. Esa agrupación permite reducir la complejidad del proceso de selección y asegurar una muestra más representativa de la población objetivo. Es importante tener en cuenta que la selección de la muestra es un aspecto crítico en cualquier estudio, ya que una muestra inadecuada puede afectar la validez y la generalización de los resultados obtenidos. Cuando la población objeto de estudio se ha dividido en conglomerados, pero hay gran variabilidad entre el tamaño de estos, se pueden formar nuevos conglomerados dentro de los que ya existen y así elevar el número de etapas de selección a dos o más. En esa situación, una alternativa consiste en combinar los conglomerados con los estratos. Los conglomerados se estratifican, se selecciona cierto número en cada estrato y, a partir de los grupos elegidos, se toman las unidades de análisis que integrarán la muestra.

El más frecuente de los muestreos por etapas es el bietápico: en la primera etapa se seleccionan de forma aleatoria los conglomerados o áreas, y en la segunda se seleccionan las unidades últimas o más elementales del conjunto poblacional, sin necesidad de seleccionar ningún otro tipo de unidad intermedia.

El hecho de aplicar el muestreo por etapas conlleva una reducción de los costos, pero también conduce a un incremento de la varianza de las estimaciones. Por otra parte, el cálculo de un estimador insesgado de la varianza puede complicarse demasiado, ya que supone estimar varios componentes de la varianza asociados con cada etapa. Por ese motivo, en algunas ocasiones se utilizan estimadores de varianza simplificados con un cierto nivel de sesgo.

v) Muestreo bifásico

El muestreo bifásico es bastante diferente del muestreo multietápico. Aunque el muestreo bifásico también implica tomar dos muestras, estas se extraen del mismo marco, y las unidades seleccionadas tienen la misma estructura en cada fase. En una muestra bifásica se recopila información de una muestra grande de unidades, y luego se recopila información más detallada para obtener una submuestra. El muestreo bifásico también se conoce como muestreo en dos fases (o muestreo doble), aunque también es posible que consista en tres o más fases. Sin embargo, al igual que ocurre con el muestreo multietápico, el diseño y el cálculo de las estimaciones de la muestra se vuelven más complicados cuando se incrementa el número de fases.

En el muestreo doble se consideran las siguientes fases:

- En la primera se selecciona una muestra grande de s_a unidades usando un diseño muestral simple o estratificado. En relación con esas unidades se recolecta información auxiliar cuya obtención es sencilla o de bajo costo.
- Con la ayuda de la información auxiliar obtenida en la primera fase, a partir de s_a se selecciona una muestra de segunda fase s , cuyo diseño muestral está condicionado a la primera muestra. Los indicadores objetivo se registran en relación con las unidades de la segunda muestra.
- La creación de un marco con información útil y confiable relativa a la muestra s_a es la base para que el muestreo doble dé resultado. Este tipo de muestreo se ha aplicado para hacer estimaciones en encuestas en que la ausencia de respuesta es considerable, o en encuestas en cuyo marco muestral no es fácil identificar la población objetivo.

Algunos ejemplos de estudios poblacionales en que resulta útil recurrir al muestreo multifásico son los relativos a poblaciones móviles (por ejemplo, personas que migran por trabajo o que viajan con frecuencia), a poblaciones ocultas o estigmatizadas (por ejemplo, debido al consumo de sustancias o drogas ilegales, o a que realizan trabajos sexuales), y a poblaciones de baja densidad (por ejemplo, comunidades rurales o aisladas).

vi) Muestreo con probabilidades proporcionales al tamaño

Este tipo de diseños de muestreo son un caso particular de los diseños con probabilidades desiguales de selección. El muestreo con probabilidades proporcionales al tamaño es un método en que se usa información auxiliar y se obtienen probabilidades de selección distintas. Si el tamaño de las unidades de la población objeto de estudio varía, y se conocen los distintos tamaños, estos se pueden usar en la selección de la muestra para mejorar la eficiencia estadística. A pesar de que el sesgo de selección no se elimina, el muestreo proporcional al tamaño puede incrementar la precisión de las estimaciones si las medidas de tamaño están correlacionadas con los indicadores objetivo, es decir, la varianza de las estimaciones ser inferior a la de un diseño muestral con probabilidades iguales (Lavrakas, 2008). El esquema en que las probabilidades de selección son proporcionales al tamaño suele utilizarse en las encuestas de hogares o viviendas, donde la cantidad de viviendas se utiliza como medida del tamaño de la población. Las variantes del muestreo con probabilidad proporcional al tamaño son las siguientes: con reemplazo, sin reemplazo, de Poisson y sistemático. Este último permite obtener muestras grandes con probabilidad proporcional al tamaño, pero no permite obtener estimaciones de varianza insesgadas, dado que la mayoría de las probabilidades de inclusión de segundo orden son cero.

vii) Muestreo con mediciones repetidas

Entre los estudios en que se emplea este tipo de muestreo destacan las encuestas longitudinales en que se hace un seguimiento de una muestra probabilística de unidades de observación durante un período, y se obtienen varias mediciones. El propósito de las encuestas longitudinales es recopilar y analizar datos sobre el crecimiento, el cambio o la tendencia de uno o varios indicadores objetivo a lo largo del tiempo. Este tipo de muestreo generalmente permite medir los cambios de una característica de la población objeto de estudio con mayor precisión que cuando se selecciona una serie de muestras independientes. Un diseño intermedio entre las muestras independientes seleccionadas de forma sucesiva y las muestras longitudinales es el muestreo con panel rotativo, en que cada vez que se lleva a cabo la encuesta se reemplaza una fracción de la muestra. El objetivo principal de las muestras de panel rotativo es obtener estimaciones a niveles agregados, y el esquema de rotación se diseña para controlar una muestra en el tiempo, garantizar que las estimaciones transversales sean insesgadas, y reducir los costos y la varianza muestral.

b) Muestreo no probabilístico

En el muestreo no probabilístico, el cálculo del tamaño y la selección de la muestra se basan en juicios y criterios subjetivos, por lo tanto, se desconoce la probabilidad de selección de las unidades de la población objeto de estudio y no es posible establecer la precisión respecto a niveles de confianza predefinidos. A pesar de ello, este tipo de muestreo, que también se denomina determinístico, representa una alternativa viable cuando aplicar

el muestreo probabilístico resulta demasiado costoso, cuando no es posible disponer de un marco de muestreo o cuando existe seguridad de que la información recabada mediante este tipo de muestreo es suficientemente útil para los fines de la investigación.

En el contexto del muestreo no probabilístico hay varias modalidades que se presentan a continuación.

i) Muestreo convencional o accidental

Consiste en recopilar datos acerca de los sujetos de estudio que resulten más accesibles. Es un esquema de muestreo rápido y de bajo costo, pero tiene deficiencias en cuanto a la representatividad. Es útil como parte del proceso de estudio exploratorio destinado a orientarse para definir la investigación, pero no es útil para caracterizar las estructuras o el comportamiento de una población objeto de estudio.

ii) Muestreo por cuotas

En este tipo de muestreo se utilizan los datos de subconjuntos o determinados estratos de la población objeto de estudio, como el sexo, la edad o la religión, entre otros, para seleccionar a los miembros que se consideran típicos según los propios fines de la investigación. El muestreo por cuotas recibe su nombre de la práctica que consiste en asignar ciertas proporciones de la muestra a determinados estratos de la población objeto de estudio. En el proceso de selección, la persona que realiza la investigación decide a quién aplicar el cuestionario de conformidad con determinados criterios generales establecidos previamente. Estos últimos resultan insuficientes para evitar que en la práctica intervengan elementos subjetivos, y por ese motivo el muestreo por cuotas no se utiliza para obtener estadísticas que requieran un buen nivel de confiabilidad. Sin embargo, en la práctica se han obtenido buenos resultados combinando el muestreo probabilístico multietápico con el muestreo por cuotas en las últimas unidades de muestreo para mejorar la representatividad de los grupos de edad y sexo.

iii) Muestreo en cadena o bola de nieve

El muestreo en cadena o de bola de nieve se basa en el supuesto de que las personas de una población constituida por pocos elementos se conocen entre sí. Para llevar a cabo un muestreo de bola de nieve es necesario identificar inicialmente a un pequeño grupo de personas que posean las características que definen a la población de interés. A continuación, se solicita a cada una de esas personas que señalen a otras que tengan las mismas características, para incluirlas en la muestra. El proceso se repite con las nuevas personas incluidas, que a su vez deben señalar a otras personas que tengan el atributo distintivo. El procedimiento continúa de forma sucesiva hasta alcanzar el tamaño de muestra deseado. El muestreo de bola de nieve permite crear una muestra bastante grande de una población específica, pero se necesitan supuestos fuertes para que los resultados obtenidos de una muestra de este tipo se puedan generalizar a la población objeto de estudio. Aunque el muestreo de bola de nieve permite identificar a miembros de una población pequeña que serían difíciles de encontrar con otros diseños, no se puede considerar que la muestra resultante se haya obtenido mediante un muestreo aleatorio simple. Las personas de la población objeto de estudio que tienen muchas conexiones tienen más probabilidades de ser incluidas en la muestra que las personas que tienen pocas conexiones; es probable que no se pueda contactar a las personas aisladas.

iv) Muestreo intencional o de juicio

La característica principal de este tipo de muestreo es que tanto el tamaño de la muestra como la selección de los elementos que la integran están sujetos al juicio del investigador, por lo que se requiere suficiente conocimiento y experiencia sobre la temática. En este tipo de muestreo, la validez de los resultados depende del conocimiento de quien investiga sobre el fenómeno en estudio y de que haya datos estadísticos que demuestren que la muestra seleccionada es útil y representativa para comprender y analizar los patrones, las tendencias o las características pertinentes en la población objetivo. El análisis de los comportamientos que se lleva a cabo mediante este tipo de muestreo no se limita a acciones humanas específicas, sino que hace alusión a patrones, tendencias o características de la población en estudio. Los comportamientos pueden variar según el tema de investigación, y pueden incluir elementos como hábitos de consumo, actitudes hacia un tema, respuestas a estímulos y prácticas culturales, entre otros.

3. Elementos que se deben considerar en un esquema de muestreo

Como se mencionó en el apartado anterior, tanto el muestreo probabilístico como el no probabilístico son técnicas de muestreo que se utilizan en la investigación. Sin embargo, existen diferencias importantes entre ambas técnicas, y las razones que llevan a usar la una o la otra dependen del objetivo de la investigación y las características de la población. El muestreo probabilístico debe utilizarse cuando se requieren estimaciones precisas y no sesgadas de la población, que son fundamentales para planificar de forma adecuada las políticas públicas, pues permiten evaluar la calidad de los resultados obtenidos y la confianza en ellos. El muestreo no probabilístico es adecuado cuando se necesita comprender un fenómeno particular en relación con el cual no es posible seleccionar muestras aleatorias.

La elección de un esquema de muestreo debe ser el resultado de la aplicación de varios principios. Tillé y Wilhelm (2017) mencionan tres: el principio de aleatorización, el de sobrerrepresentación y el de restricción. Según el primero, el diseño del muestreo debe ser lo más aleatorio posible, es decir, la selección de la muestra debe ser aleatoria y no sesgada. Según el principio de sobrerrepresentación, las unidades con mayor incertidumbre se deben sobremuestrear, pues en la muestra se debe recolectar tanta información de la población objeto de estudio como sea posible, y eso implica una tendencia a crear diseños muestrales con probabilidades de selección distintas. El principio de restricción, por su parte, consiste en seleccionar solo las muestras que tengan un conjunto de características deseables; en otras palabras, consiste en llevar a cabo un proceso de verificación (muestreo balanceado) para evitar que las muestras tengan dominios, estratos o categorías vacías.

Una vez que se ha elegido el esquema de muestreo, es necesario determinar algunos aspectos referentes a la muestra, en particular su tamaño y el procedimiento de selección.

El primer aspecto por considerar es que el error total de muestreo se refiere a todas las fuentes de sesgo y varianza de estimación que pueden afectar la precisión de los datos muestrales y que surgen en el diseño de la muestra, y en la captación, el procesamiento y el análisis de los datos de la encuesta. Según Groves y otros (2004, pág. 48), esas fuentes se pueden agrupar en dos dimensiones:

- i) Representación: se relaciona con la población objeto de estudio que se describirá en la encuesta y responde a la siguiente pregunta: ¿quiénes están incluidos en la muestra?
- ii) Medición: describe los datos que se obtendrán acerca de la unidad de observación y responde a la siguiente pregunta: ¿cuál es la temática de la encuesta?

Los costos, la carga para la persona informante, el profesionalismo, la ética y las restricciones afectan tanto la representación como la medición.

Según Groves y otros (2004, pág. 48), la representatividad de una muestra está afectada por las siguientes cuatro fuentes de error:

- i) Error de cobertura: las imperfecciones del marco de muestreo conducen a que la representatividad de la muestra sea incompleta en lo que respecta a la población objeto de estudio, lo que da lugar a que la precisión de las estimaciones muestrales sea deficiente y a que haya un sesgo de cobertura.
- ii) Error de muestreo: es el error que ocurre cuando se mide una muestra en lugar de la población completa objeto de estudio. Se compone de un error fijo (el sesgo) y de un error variable (la varianza). El sesgo muestral es la falla sistemática a la hora de observar elementos de la muestra que tienen valores distintos en el indicador objetivo. La varianza muestral se refiere a la variabilidad de las estimaciones debida a las posibles muestras obtenidas, que son afectadas por el tamaño de la muestra, y por la conglomeración, la estratificación y la ponderación.
- iii) Error debido a la falta de respuesta: la falta de respuesta total o por unidad sucede cuando no es posible encontrar a la persona informante que se había seleccionado, o cuando esta se niega a ser entrevistada o existen barreras de comunicación. El sesgo debido a la falta de respuesta surge cuando la falta de respuesta por unidad está correlacionada con uno o varios indicadores objetivo.

- iv) Error debido a ajustes posteriores: cuando la captación de los datos de una encuesta basada en muestreo probabilístico ha concluido, se llevan a cabo ajustes para considerar el sesgo de selección, los errores de cobertura y los errores debidos a la falta de respuesta. Deben tenerse en cuenta las diferencias entre los ajustes realizados debido al diseño de la muestra (por ejemplo, el ajuste de la probabilidad de selección de la persona informante según el tamaño del hogar) y los que se llevan a cabo para eliminar las diferencias entre los resultados de la estimación muestral y las estadísticas oficiales disponibles.

B. Diseños muestrales complejos y sus características

Es habitual que en el diseño muestral de una encuesta se planteen diferentes estrategias de selección y estimación, y que entre todas ellas se escoja la que satisfaga las necesidades de la investigación con eficiencia, precisión y economía. Por lo general se definen estrategias de selección diferentes de la selección directa de elementos, y se escogen estimadores que mejoran la precisión de las estimaciones, lo que conduce a que la muestra seleccionada deje de ser sencilla y se convierta en compleja.

Las muestras complejas se caracterizan por que en el proceso de selección hay varios aspectos que las diferencian del muestreo directo de elementos. Entre esos aspectos se encuentran el tipo de marco utilizado, las etapas o fases de la selección, el uso de conglomerados, los estratos muestrales, los tipos de unidades de análisis y otros.

En la mayoría de las encuestas de hogares se utilizan diseños muestrales distintos del muestreo aleatorio simple. En línea con lo que señalaron Heeringa, West y Berglund (2010), por lo general se emplea la estratificación para aumentar la eficiencia estadística y administrativa de la muestra. También se agrupan los elementos de la población en conglomerados para reducir los costos de viaje y mejorar la eficiencia de las entrevistas, lo que se traduce en diseños muestrales de varias etapas. Otra característica de los diseños muestrales complejos es que, debido a que suele aplicarse un muestreo desproporcionado de algunos elementos de la población para incrementar el tamaño muestral en subpoblaciones de especial interés, surge la necesidad de emplear la ponderación en la estimación descriptiva de las estadísticas de población.

Las encuestas de hogares por lo general tienen un marco de área en que se llega a las personas de forma indirecta definiendo varias etapas en la muestra. Además, en cada etapa se puede definir una estratificación particular, y con las submuestras de esas encuestas muchas veces se hacen otras investigaciones (fases). A veces se estiman indicadores de vivienda, otras veces de hogares y otras de personas, lo que implica una variedad importante de unidades de análisis.

Todas esas características de los planes de muestreo que suelen emplearse en las encuestas de hogares tienen efectos en la exactitud y la precisión de los estimadores, efectos que se analizan en más adelante.

Por otro lado, en este tipo de encuestas los parámetros se pueden estimar usando estimadores de regresión o de calibración. En muchos casos, las encuestas de hogares se ajustan aplicando ese tipo de métodos. Esos procesos de selección y estimación repercuten directamente en los estimadores, sobre todo en los de varianza, debido a la dificultad para definir con una fórmula matemática todos los aspectos inherentes a esos procesos.

C. Parámetros poblacionales y sus estimadores

Como señala Kish (1965), desde el punto de vista de la teoría del muestreo, un parámetro poblacional es una expresión numérica en que se sintetizan los valores de una o varias características de las N unidades elementales que conforman una población estadística completa; es una medida resumida de una cualidad de la distribución de la variable o variables en la población definida. Kish (1965) agrega que, en función de los N valores poblacionales que

toma la característica y en la k -ésima unidad de análisis, el total, la media, el tamaño de un dominio y la proporción son algunos de los parámetros poblacionales θ que se suelen estimar de forma más habitual. Al operativizar esos conceptos se obtienen las siguientes fórmulas:

$$\text{Total poblacional: } Y = f(y_1, \dots, y_N) = \sum_{k=1}^N y_k \quad (1)$$

$$\text{Media poblacional: } \mu = \bar{Y} = f(y_1, \dots, y_N) = N^{-1} \sum_{k=1}^N y_k \quad (2)$$

$$\text{Tamaño de un dominio: } N_d = f(Z_{1d}, \dots, Z_{Nd}) = \sum_{k=1}^N Z_{kd} \quad (3)$$

$$\text{Proporción poblacional: } P = f(Z_{1d}, \dots, Z_{Nd}) = \frac{N_d}{N} = N^{-1} \sum_{k=1}^N Z_{kd} \quad (4)$$

$$\text{Razón poblacional: } R = f(y_{11}, \dots, y_{1N}, y_{21}, \dots, y_{2N}) = \frac{Y_1}{Y_2} = \frac{\sum_{k=1}^N y_{1k}}{\sum_{k=1}^N y_{2k}} \quad (5)$$

En las fórmulas anteriores, Z_{kd} es una variable dicotómica que vale 1 cuando el individuo k se encuentra en el dominio d , y 0 cuando no se encuentra en él. También hay otros parámetros poblacionales de naturaleza descriptiva, a saber, la razón entre dos totales poblacionales, la mediana poblacional, los percentiles poblacionales y el tamaño poblacional, entre otros. Asimismo, hay algunos indicadores analíticos poblacionales, como el coeficiente de correlación entre variables y los coeficientes de regresión, que podrían ser de interés.

Por otra parte, Mirás (1985) plantea que, en los fenómenos de carácter estadístico, la estimación en poblaciones finitas puede plantearse, en esencia, de la siguiente forma: definida una característica y en la población $U = \{u_1, u_2, \dots, u_k, \dots, u_N\}$, característica que toma el valor numérico y_k en la unidad u_k , para estimar el valor numérico θ en función de los N valores y_k se selecciona una muestra aleatoria s y, utilizando la información obtenida a partir de la medición de las unidades de esa muestra, se atribuye a θ un valor $\hat{\theta}(s)$. Por lo tanto, un estimador es una función $\hat{\theta}$ en que a cada muestra s se le asocia el valor numérico $\hat{\theta}(s)$ que, por ende, es una estimación particular de θ dada la muestra s .

Mirás (1985) enfatiza que el diseño de un estimador depende de su distribución de probabilidad, y que esta depende del procedimiento de muestreo adoptado; por lo tanto, la formulación de estimadores, en el conjunto del diseño, no debe ser una operación independiente del procedimiento de muestreo utilizado.

Martínez (2019) argumenta que, dado que un estimador $\hat{\theta}$ es una variable aleatoria unidimensional, un buen estimador debe poseer las siguientes características:

- Ser insesgado. Se dice que un estimador $\hat{\theta}$ del parámetro poblacional θ es insesgado cuando su sesgo es nulo, es decir, cuando la esperanza del estimador $\hat{\theta}$ y el valor del parámetro θ son iguales: $B(\hat{\theta}) = E(\hat{\theta}) - \theta = 0$.
- Ser consistente. Se dice que un estimador $\hat{\theta}$ es consistente cuando conforme aumenta el tamaño de la muestra converge en probabilidad hacia el parámetro poblacional θ que se estima.

Pérez (2005) plantea que, dado que la función $\hat{\theta}$ es un estimador del parámetro poblacional θ , a los valores obtenidos en relación con cada muestra s del espacio muestral se les denomina estimaciones puntuales. En consecuencia, surge el problema de evaluar la forma de los estimadores puntuales óptimos para los parámetros poblacionales de interés, lo que conlleva, por lo general, utilizar estimadores lineales insesgados de la forma $\hat{\theta} = \sum_{k=1}^n w_k y_k$, porque ese tipo de estimadores suelen presentar las mejores propiedades; los valores w_k representan los pesos de muestreo o factores de expansión.

Concretamente, Pérez (2005) sostiene que, en el caso del muestreo sin remplazo, el estimador óptimo es el de Horvitz-Thompson, a saber, $\hat{\theta}_{HT} = \sum_{k=1}^n y_k / \pi_k$, en donde π_k es la probabilidad de que la unidad u_k esté incluida. A continuación se proporcionan las fórmulas que se obtienen al operativizar estos conceptos.

En el caso del muestreo sin reemplazo:

$$\text{Total estimado:} \quad \hat{Y}_{HT} = \sum_{k=1}^n y_k / \pi_k \quad (6)$$

$$\text{Media estimada:} \quad \hat{\bar{Y}}_{HT} = N^{-1} \sum_{k=1}^n y_k / \pi_k \quad (7)$$

$$\text{Tamaño de un dominio estimado:} \quad \hat{N}_{dHT} = \sum_{k=1}^n Z_{kd} / \pi_k \quad (8)$$

$$\text{Proporción estimada:} \quad \hat{P}_{HT} = N^{-1} \sum_{k=1}^n Z_{kd} / \pi_k \quad (9)$$

$$\text{Razón estimada:} \quad \hat{R}_{HT} = \sum_{k=1}^n y_{1k} / \pi_k \bigg/ \sum_{k=1}^n y_{2k} / \pi_k \quad (10)$$

D. Intervalo de confianza

El intervalo de confianza corresponde a un rango o recorrido de valores derivados de las estadísticas. Está relacionado con la muestra y con el diseño de muestreo, que además acota la franja en que tenderá a situarse la característica de la población que se va a estimar.

La utilización de intervalos de confianza implica elegir el nivel de confianza, es decir, el 90%, el 95% o el 99%. Dado que ese nivel varía en función de los objetivos de la encuesta y de los requisitos relativos a la precisión de las estimaciones, es importante especificar el nivel de confianza que se ha aplicado (Naciones Unidas, 2009).

Si se desean realizar inferencias sobre el parámetro de interés θ de una población definida, es común utilizar un intervalo de confianza del 95%. Para calcular ese intervalo se utiliza el valor crítico $t_{0,975, gl}$, que corresponde al percentil de la distribución t de Student con gl grados de libertad. Ese valor se utiliza para calcular el margen de error del intervalo y garantizar que este cubra el verdadero valor del parámetro con un 95% de confianza.

Es importante tener en cuenta que, cuando la muestra es grande, el valor crítico $t_{0,975, gl}$ converge hacia el mismo nivel de confianza que el percentil de la distribución normal estándar denotado por $Z_{1-\alpha/2}$, que corresponde a 1,96 cuando el nivel de confianza es del 95%. Por lo tanto, si se cuenta con una muestra lo suficientemente grande, se puede utilizar este valor crítico en lugar del valor crítico de la distribución t de Student para calcular el intervalo de confianza del 95% (INE, 2020b).

El valor de confianza sobre la población definida está dado por la siguiente ecuación:

$$(\hat{\theta} - t_{0,975, gl} * ee(\hat{\theta}), \hat{\theta} + t_{0,975, gl} * ee(\hat{\theta})) \quad (11)$$

$$LimInf = \hat{\theta} - t_{0,975, gl} * ee(\hat{\theta}) \quad (12)$$

$$LimSup = \hat{\theta} + t_{0,975, gl} * ee(\hat{\theta}) \quad (13)$$

donde:

θ es el parámetro de la población;

$\hat{\theta}$ es un estimador por muestreo correspondiente al parámetro de interés θ ;

$t_{0,975, gl}$ es el percentil 0,975 de la distribución t de Student con gl grados de libertad, y

$ee(\hat{\theta})$ es el error estándar de la estimación, donde $ee(\hat{\theta}) = \sqrt{v(\hat{\theta})}$.

La longitud de los intervalos de confianza influye en la percepción de cuán precisas son las estimaciones.

Los intervalos de confianza correspondientes a las proporciones deben estar contenidos en el intervalo (0, 1). En algunos casos, el error estándar es grande cuando se trata de estimaciones de proporción cercanas a 0 o a 1 y, por ende, los límites del intervalo son inferiores a 0 o superiores a 1. En esos casos particulares se debe considerar la posibilidad de hacer la estimación del intervalo de confianza con una variante que permita considerar esas restricciones. Una solución a este problema es considerar una transformación del estimador. De esa manera, si \hat{P} es una estimación de una proporción, se define la transformación logit de la proporción (Gutiérrez y otros, 2020):

$$\hat{L} = \log\left(\frac{\hat{P}}{1 - \hat{P}}\right) = \text{logit}(\hat{P}) \quad (14)$$

Y la aproximación de Taylor de primer orden respecto de \hat{L} es la siguiente:

$$\hat{L} \cong L(p) + \frac{\partial \hat{L}}{\partial \hat{P}} \Big|_{\hat{P}=p} (\hat{P} - p) = L(p) + \left(\frac{-1}{p(1-p)}\right) (\hat{P} - p) \quad (15)$$

Luego, la varianza de \hat{L} se puede escribir de la siguiente manera:

$$\text{Var}(\hat{L}) = A\text{Var}(\hat{L}) = \frac{\text{Var}(\hat{P})}{p^2(1-p)^2} \quad (16)$$

De esta forma, es posible definir un intervalo de 100 $(1 - \alpha)\%$ de confianza para L , del siguiente modo:

$$(\hat{L} - t_{0,975,gl} \sqrt{\text{Var}(\hat{L})}, \hat{L} + t_{0,975,gl} \sqrt{\text{Var}(\hat{L})}) = (\hat{L}_1, \hat{L}_2) \quad (17)$$

Finalmente, se obtiene la siguiente fórmula:

$$\hat{P} = \text{logit}^{-1}(\hat{L}) = \frac{\exp(\hat{L})}{1 + \exp(\hat{L})} \quad (18)$$

Por lo tanto, el intervalo de confianza correspondiente a \hat{P} está dado por la siguiente fórmula:

$$(\text{logit}(\hat{L}_1), \text{logit}(\hat{L}_2)) = \left(\frac{\exp(\hat{L}_1)}{1 + \exp(\hat{L}_1)}, \frac{\exp(\hat{L}_2)}{1 + \exp(\hat{L}_2)}\right) \subseteq (0, 1) \quad (19)$$

E. Estimación de los errores muestrales

Dado que la varianza es un indicador fundamental a la hora de determinar la precisión de las estimaciones como uno de los factores que determinan la calidad estadística, en la presente sección se desarrollará el concepto de varianza y de sus estimadores, así como también de los estimadores de otras medidas de dispersión que también se emplean para evaluar la calidad estadística.

1. La varianza y sus estimadores

Actualmente, los objetivos de las encuestas de hogares por muestreo van mucho más allá de proporcionar, desde el punto de vista descriptivo, una serie de tabulados que ofrecen estimaciones puntuales de los parámetros de interés. En efecto, cada vez es más frecuente que se incorporen objetivos analíticos destinados a ajustar

los modelos y a contrastar hipótesis que den cuenta de la fuerza y la relación de las variables en la población a diferentes niveles. Por ese motivo es necesario que al publicar los resultados se incluyan medidas apropiadas de la precisión o la exactitud de las estimaciones derivadas de los datos de la encuesta¹³.

La varianza es una de las medidas clave de la precisión en la encuesta por muestreo, y es un indicador de la variabilidad que se introduce al elegir una muestra en vez de confeccionar una lista de toda la población. Además, del cálculo de la varianza surgen otras formas de medir el error de muestreo, como el error estándar, el coeficiente de variación y el efecto del diseño, entre otras. Estas medidas se caracterizan porque guardan una relación algebraica entre sí, por lo que es posible deducir la expresión de cualquiera de ellas a partir de una que esté dada.

a) Fórmulas exactas

Cuando es posible aplicarlos, los métodos exactos de estimación de la varianza son la mejor forma de estimarla. Sin embargo, los diseños muestrales que se emplean en la mayoría de las encuestas de hogares son más complejos que el muestreo aleatorio simple o estratificado, que es el tipo de muestreo en que los métodos exactos se pueden aplicar.

Los métodos exactos dependen del diseño de la muestra, de la estimación de interés y de los procedimientos de ponderación empleados. En el cuadro II.1 se muestra la varianza correspondiente a los estimadores que se utilizan con más frecuencia en el muestreo aleatorio simple, donde n es el tamaño de la muestra extraída de una población de tamaño N .

Cuadro II.1

Principales parámetros, estimadores y varianza estimada en el muestreo aleatorio simple

| Parámetro ^a | Estimador | Estimación de la varianza |
|--|---|--|
| Media de la población (\bar{y}) | $\hat{\bar{y}} = \sum_{i=1}^n y_i/n$ | $v(\hat{\bar{y}}) = \left(1 - \frac{n}{N}\right) \cdot \frac{S_y^2}{n}$; con $S_y^2 = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1}$ |
| Total cuantitativo (τ_y) | $\hat{\tau}_y = N \cdot \hat{\bar{y}}$ | $v(\hat{\tau}_y) = N^2 \cdot v(\hat{\bar{y}})$ |
| Proporción de población correspondiente a una categoría d (P_d) | $\hat{P}_d = \sum_{i=1}^n I_i/n$ $I_i = \begin{cases} 1, & \text{si } i \in \text{clase } d \\ 0, & \text{en otro caso} \end{cases}$ | $v(\hat{P}_d) = \left(1 - \frac{n}{N}\right) \cdot \frac{\hat{P}_d \cdot (1 - \hat{P}_d)}{n-1}$ |
| Razón entre la variable y y la variable x $r_{y/x} = \frac{\tau_y}{\tau_x} = \frac{\bar{y}}{\bar{x}}$ | $\hat{r}_{y/x} = \frac{\hat{\tau}_y}{\hat{\tau}_x} = \frac{\hat{\bar{y}}}{\hat{\bar{x}}}$ | $v(\hat{r}_{y/x}) \doteq \left(1 - \frac{n}{N}\right) \cdot (S_y^2 - 2 \cdot \hat{r}_{y/x} \cdot \text{cov}(x,y) + \hat{r}_{y/x}^2 \cdot S_x^2)$ |

Fuente: W. Cochran, *Técnicas de muestreo*, Ciudad de México, Compañía Editorial Continental (CECSA), 1980.

^a Un parámetro es una función real en cuyo cálculo se consideran todos los elementos de la población. Todas las estadísticas descriptivas conocidas constituyen parámetros si se obtienen considerando todos los elementos de la población. Algunos ejemplos de esas estadísticas son la media, la proporción, el total, la razón, la varianza, el percentil y otras que se derivan de operaciones algebraicas realizadas entre ellas, como el coeficiente de variación, que es la desviación estándar dividida por la media. Cuando las funciones reales se aplican considerando solamente los elementos de la muestra seleccionada con el fin de estimar el parámetro en cuestión, se denominan estimador.

b) Método del último conglomerado

El método del último conglomerado para estimar la varianza se emplea en las estimaciones basadas en una muestra obtenida a partir de un diseño muestral complejo, generalmente polietápico, por conglomerados y estratificado (Hansen, Hurwitz y Madow, 1953, págs. 257-259).

¹³ Las medidas de precisión estadística dan cuenta de la cercanía entre el estimador proveniente de la encuesta y el valor verdadero del parámetro. También se utilizan como medida de la calidad de la estimación.

Según ese método, el último conglomerado se obtiene al considerar la unidad primaria de muestreo (UPM) como la unidad de información última en la que se agrupa la información de todas las unidades de muestreo de las etapas posteriores que se anidan en ella.

De esa forma, el diseño del muestreo se asemeja a un muestreo por conglomerados estratificados de una sola etapa, es decir, una muestra estratificada de conglomerados finales (últimos) completamente enumerados. Este enfoque por conglomerados finales produce una buena aproximación de la estimación de la varianza (Gutiérrez, 2022), y por ese motivo su uso está muy generalizado entre las ONE. Esto implica que la varianza se calcula mediante la intervarianza o la varianza entre conglomerados utilizando solo totales entre UPM, sin tener que calcular los componentes de la varianza en cada etapa de selección.

El método del último conglomerado y la estimación simplificada de la varianza en el muestreo complejo se utilizan para estimar el tamaño poblacional utilizando los mismos factores de expansión. A diferencia de otros métodos que se examinarán más adelante, este enfoque permite establecer el tamaño poblacional y aplicar la corrección por finitud para reducir la varianza, lo cual fija las poblaciones. Esto significa que, al utilizar el método del último conglomerado, se logra una mayor precisión en las estimaciones poblacionales al mantener las poblaciones fijas y ajustar la varianza mediante la corrección por finitud.

c) Aproximaciones mediante linealización de Taylor

Muchas de las estadísticas descriptivas por estimar o de las estimaciones de interés no son funciones lineales simples de los valores observados, por lo que la varianza del muestreo no puede expresarse mediante una fórmula de forma cerrada, como ocurre con la varianza de la media muestral en el caso del muestreo aleatorio simple o del muestreo estratificado.

El método de linealización se utiliza con mucha frecuencia porque puede aplicarse a casi todos los diseños muestrales y a cualquier estadística que pueda linealizarse o expresarse como una función lineal de las estadísticas comunes, como las medias o los totales, cuyos coeficientes se extraen de derivadas parciales necesarias para lograr la expansión en series de Taylor.

La linealización generalmente se utiliza para estimar razones o cocientes de dos variables que se desarrollan en una serie de Taylor infinita centrada en el valor esperado (estimado) del numerador y el valor esperado (estimado) del denominador.

A continuación, el estimador no lineal se somete a una aproximación algebraica conservando solo los primeros términos de la serie infinita de Taylor. Con eso se produce una expresión algebraica que es una función lineal de los datos muestrales, lo que significa que el estimador no lineal cociente se ha linealizado y ahora es posible obtener directamente la varianza estimada de la función linealizada (incluidos los términos de covarianza pertinentes).

Una vez linealizada, la varianza de la estimación no lineal puede aproximarse mediante los métodos exactos descritos anteriormente¹⁴.

d) Pesos replicados

Este método consiste en tomar submuestras o réplicas de la muestra total. En primer lugar, la variable de interés Y se estima utilizando la muestra total de acuerdo con el diseño probabilístico, sea cual fuere. A continuación se crean réplicas o se seleccionan submuestras de la muestra total, de manera que cada una de ellas refleje el plan de muestreo, los ajustes y los procedimientos de ponderación de la muestra completa y permita llegar a la misma estimación que la muestra completa.

Existen varias técnicas para aplicar este método, y algunos programas informáticos ya traen módulos de selección de réplicas incluidos. Entre esos módulos, los más utilizados son los grupos aleatorios, la replicación repetida equilibrada (McCarthy, 1969; Judkins, 1990), la replicación mediante el método de la navaja o *jackknife* (Krewski y Rao, 1981), el *bootstrap* y el *sub-bootstrap* (Rao y Wu, 1984).

¹⁴ Para consultar información técnica detallada sobre el proceso de linealización ilustrada con ejemplos, véanse Cochran (1977), Lohr (1999), Heeringa, West y Berglund (2010), y Valliant, Dever y Kreuter (2018).

Supongamos, por ejemplo, que se crean K réplicas a partir de una muestra dada, que a cada una de esas réplicas le corresponde una estimación $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K$ del parámetro θ , y que además la estimación basada en la muestra completa es $\hat{\theta}_0$. En ese caso, la estimación de la varianza basada en la replicación se obtiene a partir de la siguiente fórmula: $v(\hat{\theta}) = \frac{1}{c} \sum_{r=1}^K (\hat{\theta}_r - \hat{\theta}_0)^2$, donde c es una constante que depende del método de estimación.

Estos métodos de estimación de varianzas son antiguos y antes había pocos programas informáticos que los incorporaran o desarrollaran. Sin embargo, con el avance tecnológico y con el de la computación en particular, hoy en día la mayoría de los programas estadísticos tienen módulos especiales destinados al tratamiento de los datos obtenidos de muestras complejas, entendiendo por muestra compleja aquella en que se incorporan variables de ponderación, estratificación y conglomeración para estimar los principales datos estadísticos asociados a alguna variable de interés y para estimar también sus varianzas.

2. Medidas de dispersión asociadas con la varianza

Dadas la estimación del parámetro y la de la varianza como medida de dispersión por sí misma, se pueden obtener otras medidas asociadas, como el error cuadrático medio (*ECM*), el error estándar (*EE*), el coeficiente de variación (*CV*), el error absoluto (E_A), el error relativo (E_R) y el intervalo de confianza (*IC*). A su vez, dependiendo del tipo de diseño, surge otra medida de dispersión que relaciona la varianza obtenida a partir de un diseño complejo con la obtenida mediante un diseño aleatorio simple (muestreo aleatorio simple). Esa medida se denomina efecto del diseño *Deff*.

a) Error cuadrático medio

El error cuadrático medio (*ECM*) de cualquier estimador $\hat{\theta}$ tiene dos componentes, uno relacionado directamente con la variabilidad misma ($\sigma^2[\hat{\theta}]$) y otro que captura la diferencia entre el valor esperado del estimador y el verdadero valor del parámetro ($E[\hat{\theta}] - \theta$)¹⁵. Esta diferencia, denominada sesgo del estimador, es casi imposible de obtener directamente debido a que en la práctica se desconoce el valor del parámetro θ (se puede aproximar este sesgo mediante simulaciones). El *ECM* se expresa del siguiente modo:

$$ECM(\hat{\theta}) = E[\hat{\theta} - \theta]^2 = \sigma^2[\hat{\theta}] + (E[\hat{\theta}] - \theta)^2 \quad (20)$$

donde,

$$\sigma^2[\hat{\theta}] = \text{Varianza}; \text{Sesgo}(\hat{\theta}) = E[\hat{\theta}] - \theta \quad (21)$$

Se dice que el estimador es insesgado si $\text{Sesgo}(\hat{\theta}) = 0 \Leftrightarrow E[\hat{\theta}] = \theta$

En las encuestas por muestreo se utilizan estimadores insesgados o asintóticamente insesgados, de modo que el sesgo es despreciable o se anula con muestras de tamaño mayor, y por ello las medidas de precisión se enfocan en la varianza o en la raíz cuadrada de esta, que se denomina error estándar, o en una transformación de ella.

Cabe decir que el *ECM* queda en función del mismo parámetro θ , que por lo general es desconocido y se estima por medio de $\hat{\theta}$. Mediante el cálculo $E[\hat{\theta}]$, es decir, el promedio de las estimaciones sobre la base de todas las muestras posibles de cierto tamaño, se llega al valor del parámetro, siempre y cuando el estimador sea insesgado. En la práctica, si el estimador es insesgado, sobre la base de una sola muestra se llega a un valor del parámetro desconocido que es aproximadamente cercano.

¹⁵ En realidad, cuando en este documento se menciona el *ECM* se hace referencia al *ECM* empírico, que es aquel en que se supone que el parámetro poblacional por estimar es conocido y tiene un valor dado o fijado que puede corresponder a un registro administrativo o a una estimación exógena que por lo general corresponde a la obtenida con otro estimador. Se supone que esa estimación exógena es la poblacional, y entonces se calcula un sesgo empírico dado por la diferencia entre el valor obtenido con un estimador y el valor que se supone poblacional.

Cuando el estimador es insesgado, entonces el *ECM* coincide con la varianza del estimador, y también se dice que un estimador insesgado es más eficiente que otro insesgado cuando la varianza es menor. En general, el *ECM* se utiliza para comparar dos o más estimadores, sean insesgados o no, y se considera que el que tiene el menor valor es el más preciso o eficiente. Si se trata de un estimador sesgado, es deseable que sea consistente, es decir, que a medida que aumente el tamaño de la muestra, tanto el sesgo como la varianza disminuyan y converjan hacia cero (de forma asintótica).

b) Error típico o estándar

Es un indicador que brinda información sobre cuán disperso es el estimador¹⁶. Esta medida resulta más fácil de interpretar que la varianza, ya que en ella se utiliza la misma escala de medición que en la estimación:

$$EE(\hat{\theta}) = \sqrt{\sigma^2(\hat{\theta})} = \sigma(\hat{\theta}) \quad (22)$$

c) Coeficiente de variación

El coeficiente de variación (*CV*) es una medida adimensional, es decir, un número puro que no tiene unidades, lo que permite comparar el *CV* de diversas distribuciones de variables expresadas en diversas unidades. Es el cociente entre el error típico o estándar y el parámetro estimado, a saber:

$$CV(\hat{\theta}) = \sqrt{\sigma^2(\hat{\theta})} / \hat{\theta} \quad (23)$$

También se puede expresar en porcentaje:

$$100 \cdot CV(\hat{\theta}) \% = 100 \cdot \sqrt{\sigma^2(\hat{\theta})} / \hat{\theta} \% \quad (24)$$

d) Error absoluto e intervalo de confianza

El error absoluto asociado a una estimación se puede interpretar como la máxima diferencia que se espera encontrar entre el estimador del parámetro y su (verdadero) valor, con una probabilidad del $100(1-\alpha)\%$, donde $(1-\alpha)$ es el nivel de confianza utilizado o la probabilidad de que el parámetro θ esté contenido en el intervalo construido¹⁷.

La probabilidad de que la distancia entre el parámetro estimado en la muestra y el verdadero valor del parámetro equivalga como máximo al error absoluto definido $E_A(\hat{\theta})$ se expresa de la siguiente manera:

$$P(|\hat{\theta} - \theta| \leq E_A(\hat{\theta})) = 1 - \alpha \quad (25)$$

Al desarrollar la expresión (25) y despejar dentro del paréntesis en relación con el parámetro θ se llega a la siguiente expresión, que tiene forma de intervalo de confianza o estimación intervalar:

$$P[\hat{\theta} - E_A(\hat{\theta}) \leq \theta \leq \hat{\theta} + E_A(\hat{\theta})] = 1 - \alpha \Leftrightarrow \theta \in [\hat{\theta} - E_A(\hat{\theta}); \hat{\theta} + E_A(\hat{\theta})]_{1-\alpha} \quad (26)$$

Utilizando la ley de los grandes números y el teorema central del límite, si el tamaño de la muestra es lo suficientemente grande, la distribución de probabilidad del estimador converge hacia la distribución normal. No obstante, en la práctica el estimador $\hat{\theta}$ se distribuye como una *t* de Student (la varianza poblacional es desconocida

¹⁶ Un indicador o medida es una estadística descriptiva o estadígrafo, es decir, una función de los valores que toma la variable de interés en la muestra, por ejemplo, el mínimo, el máximo, el promedio, la varianza y otros. Un indicador también es una función de dos más estadígrafos, por ejemplo, el coeficiente de variación (error estándar dividido por la media) o la tasa de desocupados (total de desocupados dividido por el total de personas en la fuerza de trabajo).

¹⁷ En términos probabilísticos, en el contexto del muestreo repetido, es decir, cuando el 100% de los intervalos se construyen con muestras del mismo tamaño y del mismo nivel de confianza, la probabilidad de que el intervalo contenga el parámetro poblacional se aproxima como el porcentaje de intervalos que contendrán el parámetro.

y se estima por la misma muestra), la media o esperanza es igual al valor del parámetro θ y la varianza es igual a $\sigma^2(\hat{\theta})$. Despejando a partir de la expresión (26) se llega a la siguiente expresión correspondiente al error absoluto:

$$E_A(\hat{\theta}) = t_{1-\alpha/2}^v \cdot \sqrt{\sigma^2(\hat{\theta})} = t_{1-\alpha/2}^v \cdot EE(\hat{\theta}) \quad (27)$$

donde $t_{1-\alpha/2}^v$ es el percentil de la distribución t de Student con v grados de libertad que, en el caso de las muestras grandes (más de 30 unidades de muestreo), converge, con el mismo nivel de confianza, hacia el percentil de la distribución normal estándar denotado por $Z_{1-\alpha/2}$ que, cuando el nivel de confianza es de un 95% corresponde a 1,96.

e) Error relativo

El error relativo asociado a una estimación se puede interpretar como la máxima variación porcentual que se puede esperar entre el parámetro estimado y el verdadero parámetro poblacional con una probabilidad de $(1-\alpha)$. El error relativo se define como el cociente entre el error absoluto y la estimación:

$$E_R(\hat{\theta}) = t_{1-\alpha/2}^v \cdot \sqrt{\hat{\sigma}^2(\hat{\theta})} / \hat{\theta} = t_{1-\alpha/2}^v \cdot \frac{EE(\hat{\theta})}{\hat{\theta}} = t_{1-\alpha/2}^v \cdot CV(\hat{\theta}) = \frac{E_A(\hat{\theta})}{\hat{\theta}} \quad (28)$$

En la expresión (28) se puede observar la relación entre las distintas medidas de precisión definidas anteriormente.

f) Efecto del diseño

Kish (1965) define el efecto del diseño (*Deff*) como la relación entre la varianza de una estimación en una muestra obtenida mediante un diseño complejo $\sigma_c^2(\hat{\theta})$ y la varianza de la misma estimación en una muestra obtenida mediante un muestreo aleatorio simple $\sigma_{MAS}^2(\hat{\theta})$ (Naciones Unidas, 2009). También lo define como el factor por el que hay que multiplicar la varianza de una estimación basada en una muestra aleatoria simple del mismo tamaño para tener en cuenta las complejidades del diseño muestral derivadas de factores como la estratificación, la conglomeración o las ponderaciones. Por consiguiente, el efecto del diseño se expresa de la siguiente manera:

$$Deff(\hat{\theta}) = \frac{\sigma_c^2(\hat{\theta})}{\sigma_{MAS}^2(\hat{\theta})} \Leftrightarrow \sigma_c^2(\hat{\theta}) = \sigma_{MAS}^2(\hat{\theta}) \cdot Deff(\hat{\theta}) \quad (29)$$

El efecto del diseño depende del estimador y del propio diseño muestral, es decir, dado un diseño muestral complejo, se obtienen diferentes efectos del diseño según el parámetro de interés $\hat{\theta}$ que se ha de estimar (Chambers y Skinner, 2003; Naciones Unidas, 2009). Además, hay que considerar que la muestra no se observa de la misma manera si los diseños son distintos. Por ejemplo, en un diseño complejo bietápico por conglomerados, la muestra queda agrupada en los conglomerados seleccionados; sin embargo, en un diseño aleatorio simple, la distribución de la misma muestra es muy distinta. La única manera de que la varianza esperada sea la misma en ambos diseños es que el fenómeno se comporte de forma similar en cada conglomerado, o que cada conglomerado sea una copia fiel o representativa de la población.

Por otro lado, también se puede decir que si n se calcula usando la fórmula de muestreo aleatorio simple, entonces $n_c = n \cdot Deff$ es el tamaño de muestra que se necesita en un diseño muestral complejo para alcanzar la misma varianza que en el muestreo aleatorio simple. De allí se deriva el concepto de tamaño efectivo de la muestra, que se expresa así: $n_{ef} = \frac{n_c}{Deff}$. Por ende, si *Deff* es igual a 2 y el tamaño de la muestra en un diseño complejo es 60, el tamaño efectivo de la muestra (en un diseño aleatorio simple) es 30.

Por último, vale mencionar que el efecto del diseño no es una medida global de una encuesta, sino que depende de cada estimador en particular, lo que quiere decir que, en una misma encuesta, el efecto del diseño puede variar según el estimador y el nivel al que se esté evaluando (Chambers y Skinner, 2003; Naciones Unidas, 2009)¹⁸.

3. Tabulados o desagregaciones

Un tabulado o tabla estadística, también denominado cuadro estadístico, es una configuración matricial de filas y columnas divididas en celdas o casillas que contienen información agregada de las estadísticas descriptivas provenientes de una operación estadística o encuesta.

¹⁸ Para obtener más información sobre el efecto del diseño en las encuestas de hogares, véase Gutiérrez (2022).

Cabe mencionar que las celdas que contienen alguna estadística descriptiva, como un recuento no ponderado, una suma, un promedio, una mediana, un error estándar, un coeficiente de variación u otra medida de resumen, son por sí mismas un tabulado que, junto con otras celdas que contienen las mismas medidas de resumen, van configurando los cuadros estadísticos que se publican. En ocasiones, las estadísticas descriptivas de interés más relevantes de acuerdo con el objetivo de la publicación se ponen todas juntas, por ejemplo, el recuento ponderado y el no ponderado, la estimación del parámetro de interés y la medida de la dispersión; en general, sin embargo, se colocan separadas.

Solamente a modo de ilustración, en el cuadro II.2 se presenta el total de personas por sexo, según el nivel socioeconómico. Las estadísticas descriptivas se presentan separadas: estimaciones, errores estándar y coeficientes de variación, respectivamente.

Cuadro II.2

Estimación de personas por sexo, según nivel socioeconómico

| Variable | Estimaciones | | | Errores estándar | | | Coeficientes de variación | | |
|----------|--------------|----------|----------|------------------|-----------|-----------|---------------------------|-----------|-----------|
| | Sexo | | | Sexo | | | Sexo | | |
| | Total | Mujer | Hombre | Total | Mujer | Hombre | Total | Mujer | Hombre |
| Total | t_{00} | t_{01} | t_{02} | ee_{00} | ee_{01} | ee_{02} | cv_{00} | cv_{01} | cv_{02} |
| Bajo | t_{10} | t_{11} | t_{12} | ee_{10} | ee_{11} | ee_{12} | cv_{10} | cv_{11} | cv_{12} |
| Medio | t_{20} | t_{21} | t_{22} | ee_{20} | ee_{21} | ee_{22} | cv_{20} | cv_{21} | cv_{22} |
| Alto | t_{30} | t_{31} | t_{32} | ee_{30} | ee_{31} | ee_{32} | cv_{30} | cv_{31} | cv_{32} |

Fuente: Comisión Económica para América Latina y el Caribe (CEPAL).

Aunque cada una de las estadísticas descriptivas por separado constituye un tabulado, el tabulado de las estimaciones es el que se analiza para determinar si se puede publicar sin restricción con advertencia de fiabilidad (fiable, poco fiable o no fiable). En ese tabulado se analiza el error estándar, con recorrido entre 0 y 1 inclusive, cuando se trata de un estimador de razón (o de proporción como caso particular de este). Cuando se trata de cualquiera de los demás tipos de estimadores, se analiza el coeficiente de variación.

Las variables que dan lugar a un tabulado son variables categóricas o categorizadas (si se trata de variables numéricas), y cada categoría representa una subpoblación o partición de la población total. De este modo, si solo nos referimos al cuadro estadístico de las estimaciones, hay cuatro tabulados: el total de totales (t_{00}), los totales marginales por nivel socioeconómico (t_{10} , t_{20} y t_{30}) y por sexo (t_{01} y t_{02}), y las celdas del tabulado principal que contiene la interacción entre el nivel socioeconómico y el sexo (t_{11} , t_{12} , t_{21} , t_{22} , t_{31} , t_{32}).

En primer lugar, si la precisión del total de totales t_{00} presentara problemas, el cuadro estadístico entero no debería someterse a análisis ni publicarse.

En segundo lugar, se evalúan los totales marginales de las variables nivel socioeconómico y sexo por separado. Considerando que la variable sexo no se puede colapsar porque el análisis de género perdería sentido, si una de las categorías (mujer u hombre) no tuviera la precisión adecuada según el estándar, solo quedaría la posibilidad de analizar la categoría que sí fuera estadísticamente fiable.

En cuanto a la variable principal de estudio nivel socioeconómico, es fundamental que cada una de sus categorías sea estadísticamente fiable, ya que, si no lo fueran, al menos una de las categorías de sexo no sería fiable y eso disminuiría la cantidad de análisis de género que se podría realizar de forma comparativa. En ese caso es aconsejable, siempre que sea posible y cuando tenga sentido temático, colapsar categorías colindantes hasta lograr que la nueva categoría colapsada sea fiable.

En tercer lugar, una vez colapsadas las categorías de la variable nivel socioeconómico, se evalúan las interacciones entre esa variable y la del sexo en el tabulado principal. A partir de ese momento se aplica el estándar para determinar si las celdas del tabulado principal son fiables, poco fiables o no fiables, y luego la evaluación del tabulado prosigue para determinar si es o no es publicable. Lo que se quiere expresar es que, antes de aplicar el estándar al cuadro estadístico, es fundamental asegurarse de que la variable nivel socioeconómico se haya colapsado adecuadamente, pues eso garantiza que, al iniciar la evaluación según el flujograma del estándar, el tabulado esté preparado y no sea necesario hacer más ajustes a las categorías.

Capítulo III

Elementos para evaluar la calidad estadística de una estimación

En el principio 6 del Código Regional de Buenas Prácticas en Estadísticas para América Latina y el Caribe se establece un compromiso con la calidad y se señala que “las entidades productoras de estadísticas en el marco del sistema estadístico nacional deben trabajar y cooperar conforme a normas, principios y estándares internacionales” (CEPAL, 2011, pág. 9). Todas las ONE de la región fundamentan cada etapa del proceso de producción estadística en ese principio. Este documento se centra en los lineamientos de calidad asociados a la etapa de producción de los resultados.

Para evaluar la calidad de las estimaciones es conveniente considerar una serie de factores o criterios que ayudarán a determinar qué indicadores conviene aplicar en cada caso, ya que, dependiendo de factores como el tipo de estimador, el valor de la estimación y los niveles de estimación, la idoneidad de cada criterio puede variar.

Los criterios de calidad que se presentan a continuación están basados en los principales referentes de la región en cuanto a la formulación de lineamientos sobre la calidad estadística de las estimaciones, a saber, el documento “Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional” de la CEPAL (Gutiérrez y otros, 2020) y el “Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares” del INE (2020a) de Chile.

A. Efecto del diseño

Según Lumley (2010), en los estudios grandes los efectos del diseño (*Deff*) suelen ser mayores que 1,0, lo que implica que en los diseños complejos se necesitan muestras de tamaño más grande que en los diseños aleatorios simples. Lo anterior se observa al definir el *Deff* como un cociente entre la varianza del estimador de un muestreo complejo y la varianza proveniente de un muestreo aleatorio simple. En teoría, la varianza de un muestreo aleatorio simple debería ser menor que la de un muestreo aleatorio complejo. No obstante, en la práctica es posible que el *Deff* sea menor que 1 o sea muy superior a lo esperado dado el diseño muestral o lo observado en los diferentes niveles de estimación.

En las encuestas en que el diseño muestral es complejo, el hecho de que la estimación del *Deff* sea menor que 1 puede deberse a que los algoritmos de cálculo de la varianza se basan en aproximaciones en las que en general se aplica el método de conglomerado último, cuya aplicación puede traducirse en una subestimación de la varianza que resulte inferior a la obtenida en el marco de un diseño aleatorio simple. Ante esa situación, y bajo el supuesto de que el error de muestreo de una muestra multietápica es superior al de una muestra aleatoria simple, puede justificarse que en los cálculos se utilice un efecto de diseño acotado a 1.

Por otra parte, si quien realiza la investigación considera que los efectos estimados del diseño exceden de manera importante lo esperado, puede tomar la decisión de utilizar cotas superiores en el cálculo bajo el supuesto de que los altos valores observados se deben a factores particulares del diseño y de la muestra recolectada, dado que la inferencia se realiza sobre una muestra particular de un gran número de muestras posibles.

En particular, Gutiérrez (2022) recomienda generar una alerta sobre todas las cifras cuyo efecto del diseño sea inferior a 1.

Las situaciones mencionadas anteriormente representan una primera alerta sobre la calidad de la estimación, por lo que es responsabilidad de quien investiga indagar sobre los factores que puedan haber dado lugar a ese valor.

B. Tamaño muestral mínimo para aplicar medidas de precisión

Entre los factores que afectan la magnitud de la varianza muestral se encuentran la heterogeneidad de la variable de estudio, el tamaño de la muestra y el diseño muestral. Tal como se señala en Gutiérrez y otros (2020), el tamaño de la muestra afecta de manera indirecta la amplitud del intervalo de confianza a través del error estándar, que generalmente decrece a medida que aumenta el tamaño de la muestra.

Una vez especificadas las características que debe tener la muestra, hay que determinar el tamaño (n) que esta debe tener considerando que tiene que ser lo suficientemente representativa de la población y asegurar, dado un nivel de confianza ($1-\alpha$), que el error muestral no supere un valor máximo admisible.

Por otra parte, la Comisión Europea (2013) señala que existen dos estrategias principales para establecer requisitos de precisión: la primera, determinar los umbrales de precisión que deben cumplir los principales indicadores objetivo de la encuesta, y la segunda, establecer cuál debe ser el tamaño mínimo efectivo de la muestra.

1. Tamaño muestral

Según lo que se establece en INE (2020b), el tamaño muestral corresponde al “total de unidades de análisis consideradas para obtener las estimaciones”. En ese sentido, es importante valorar que la muestra debe tener un tamaño suficiente como para que se puedan hacer estimaciones de la población objeto de estudio y como para que, si corresponde, se pueda desglosar por sexo en lo que atañe a las variables de mayor interés.

Tanto en Gutiérrez y otros (2020) como en INE (2020a) se contempla el tamaño muestral como elemento que se debe evaluar a la hora de determinar la calidad estadística de una estimación. En Gutiérrez y otros (2020) se señala que el hecho de que el tamaño de la muestra sea adecuado garantiza que la distribución de los estimadores converja hacia la distribución teórica de donde se calculan los percentiles en el cálculo del intervalo de confianza. Por otra parte, la determinación del tamaño muestral es de vital importancia para fijar el grado de precisión de la operación estadística. No obstante, existen situaciones prácticas en que el tamaño mínimo de la muestra no se determina sobre la base de un criterio de precisión. Es común que las restricciones presupuestarias limiten el número total de entrevistas que se pueden realizar, por lo que habrán de calcularse estimaciones sin que el grado de precisión haya sido contemplado en el diseño muestral de la encuesta. En esos casos se hace aún más imperativo considerar el tamaño muestral como criterio de calidad estadística.

En Gutiérrez y otros (2020) se alude a lo sugerido por Barnett-Walker y otros (2003), quienes consideran que, para obtener una estimación, debe haber 100 unidades de análisis como mínimo. Sin embargo, en INE (2020a) se considera que 60 unidades es el tamaño mínimo necesario para que una estimación pase a ser evaluada aplicando el segundo criterio de calidad. Si hay menos de 60 unidades, la estimación se clasifica directamente como no fiable y no se recomienda utilizarla.

En lo que atañe al mecanismo destinado a controlar la falta de respuesta, la Comisión Europea (2013) señala que, cuando se trata de las estadísticas comunitarias sobre la renta y las condiciones de vida (EUSILC), en el reglamento de la Comisión Europea núm. 1982/2003 se establece que las exigencias de precisión relativas a la publicación de los datos deben expresarse en términos del número de observaciones muestrales en que se basa la estadística y del nivel de falta de respuesta parcial (además de la falta de respuesta total). En ese sentido, se dispone lo siguiente:

- No se publicará ninguna estimación que esté basada en menos de 20 observaciones muestrales, o si la falta de respuesta para la característica en cuestión supera el 50%.
- Se publicarán los datos con una marca si la estimación se basa en 20 a 49 observaciones muestrales, o si la falta de respuesta para la característica en cuestión supera el 20% y es inferior o igual al 50%.
- Los datos se publicarán de forma habitual cuando las estimaciones se basen en 50 o más observaciones muestrales y la falta de respuesta para la característica no supere el 20%.

En las recomendaciones de la DIGESTYC (2022) de El Salvador, para la Encuesta de Hogares de Propósitos Múltiples (EHPM) se considera un mínimo de 14 grados de libertad, lo que implica 15 unidades primarias de muestreo (UPM) como mínimo. Además, considerando las 12 unidades que se seleccionan en la segunda etapa de selección, se obtiene un mínimo de 180 hogares. Suponiendo una tasa de respuesta del 85%, se espera que en la muestra de la EHPM haya 153 elementos como mínimo. Para otros estudios se sugiere que en la muestra haya 100 elementos como mínimo.

2. Tamaño efectivo de la muestra

El tamaño efectivo de la muestra es el tamaño mínimo que esta debe tener para que, cuando se aplica un muestreo aleatorio simple sin reemplazo, se obtenga el mismo nivel de precisión que cuando el diseño muestral es complejo. De ello se deduce que el tamaño mínimo efectivo de la muestra en el marco del muestreo aleatorio simple debe ajustarse por los efectos del diseño $Deff$, es decir, por la variación que los componentes del diseño de muestreo, como la estratificación o la agrupación, causan en la eficiencia de este (Comisión Europea, 2013).

Como se señala en Gutiérrez y otros (2020), en las encuestas de hogares el diseño consiste en seleccionar un conjunto de hogares dentro de una misma UPM y repetir esa estrategia de selección sistemáticamente en todo el país. Por lo tanto, se puede pensar que, si la variable de interés tiene una alta correlación intraclase, entonces la realidad de las personas y de los hogares dentro de una misma UPM será muy homogénea, tanto que se podría interpretar que la información está repetida, y que las personas u hogares de una misma UPM no aportan información de manera diferenciada. Por lo tanto, debido a los efectos del diseño de muestreo complejo, la cantidad de personas que aportan a la inferencia del indicador no es el número de personas ni el número de hogares que integran la muestra, sino el tamaño efectivo de esta, n_{eff} , que permite deflactar los efectos de la aglomeración.

En línea con lo anterior, el tamaño efectivo de la muestra se calcula de la siguiente forma (Kalton, Brick y Lê, 2005):

$$n_{eff} = \frac{n}{Deff} \quad (30)$$

En cuanto a los valores de referencia que permiten evaluar el tamaño efectivo de la muestra, en Gutiérrez y otros (2020) se toman los umbrales considerados en Hornik y otros (2002) y en Klein y otros (2002), donde se establece un mínimo de 30 unidades.

3. Grados de libertad

En 1915, Sir Ronald Fisher, fundador de la estadística moderna, introdujo el concepto de los grados de libertad con referencia al conjunto de observaciones. Esos grados están dados por el número de valores que pueden asignarse de forma arbitraria antes de que el resto de las variables, luego de establecerse las que son libres, tomen un valor de forma automática con el fin de compensar e igualar un resultado conocido previamente. Según Walker (1940), en términos geométricos el número de observaciones es la dimensionalidad del espacio original, y cada relación representa un corte en ese espacio que restringe el punto de muestra a un espacio que tiene una dimensión menos. Imponer una relación a las observaciones es equivalente a estimar un parámetro a partir de ellas.

El fundamento para emplear los grados de libertad al evaluar la calidad estadística radica en que esos grados representan un elemento importante a la hora de calcular el error absoluto y determinar la amplitud de los intervalos de confianza según la distribución de probabilidad asociada y también según su relación con el tamaño de la muestra.

Como se puede observar en la expresión (31), la amplitud del intervalo de confianza del estimador del parámetro de interés ($\hat{\theta}$) depende del error de muestreo en términos del error estándar (ee) y del percentil 0,975 de la distribución t de Student con los grados de libertad (gl) correspondientes.

$$(\hat{\theta} - t_{0,975,gl} * ee(\hat{\theta}), \hat{\theta} + t_{0,975,gl} * ee(\hat{\theta})) \tag{31}$$

Tal como se señala en INE (2020b), cuando se trata de encuestas en que se utiliza un diseño muestral complejo, el cálculo de los grados de libertad para estimar la varianza no es algo trivial. Actualmente, en los paquetes estadísticos destinados al análisis de datos obtenidos de muestras complejas se emplea la regla del cálculo de grados fijos de libertad, que consiste en obtener los grados de libertad restando al número total de UPM el número total de estratos.

$$gl = \text{número de UPM} - \text{número de estratos} \tag{32}$$

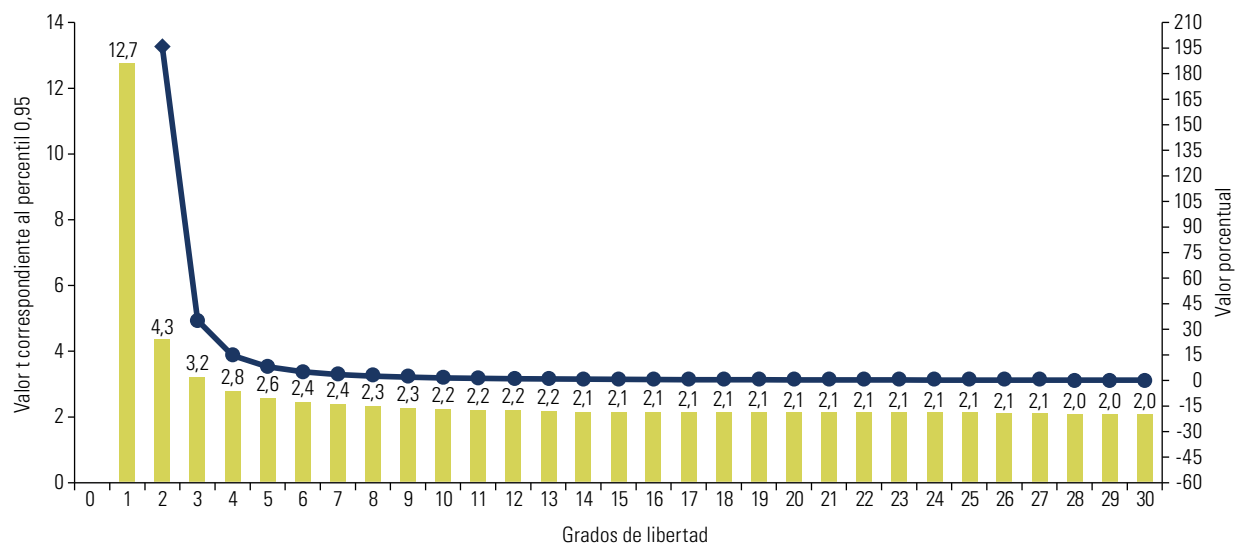
No obstante, existen pruebas de que esta metodología conduce a una sobrestimación de los grados de libertad y, en consecuencia, se obtienen intervalos de confianza más estrechos. Como alternativa, Korn y Graubard (1999) sugirieron calcular grados variables de libertad. En ese método, los grados de libertad se obtienen restando el número de UPM muestreadas que contienen a la subpoblación y el total de estratos con muestra de la subpoblación del total de conglomerados con muestra en ella. Este método resulta más conveniente, sobre todo cuando la característica de interés es poco prevalente.

En lo que respecta a la evaluación de los grados de libertad como indicador de calidad, Gutiérrez y otros (2020) hacen alusión a Parker y otros (2017), donde se considera que, si los grados de libertad inducidos por la subpoblación son menos de ocho, la cifra debería suprimirse.

Por otro lado, en INE (2020b) se analiza el comportamiento del percentil 0,95 de la distribución t de Student (percentil asociado a un nivel de confianza del 95%), ya que es el más utilizado para calcular intervalos de confianza. En el gráfico III.1 destaca el comportamiento pronunciado de la distribución al considerar pocos grados de libertad. En efecto, se puede apreciar que, cuando hay 1, 2 y 3 grados de libertad, el valor t asociado del percentil es 12,7, 4,3 y 3,2, respectivamente, lo que equivale a una variación porcentual de entre el 195,3% y el 35,2%. A medida que el número de grados de libertad aumenta, la variación porcentual se reduce y, cuando hay entre ocho y nueve grados de libertad, la variación se estabiliza.

Gráfico III.1

Percentil 0,95 de la distribución t de Student y variación porcentual según los grados de libertad



Fuente: Instituto Nacional de Estadísticas de Chile (INE), "Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares", Documento de Trabajo, Santiago, N° 13, 2020.

Es importante señalar que, si se utilizan los grados de libertad como criterio de supresión, los responsables de diseñar y planificar las encuestas con diseños muestrales bietápicos en cada país deberán tener en cuenta que, para que una cifra sea publicada en un dominio de interés, este deberá tener una cantidad suficiente de UPM (al menos 11) para que la regla no limite su disseminación (Gutiérrez y otros, 2020).

C. Tipología de los estimadores

Según lo que se señala en Comisión Europea (2013), la tipología de los estimadores desempeña un papel fundamental a la hora de definir los criterios para construir estándares que permitan evaluar la calidad estadística de las estimaciones, ya que las medidas de precisión deben adaptarse al tipo de estimador.

En diversos análisis se ha evidenciado que, cuando el estimador corresponde a una proporción o una razón, el coeficiente de variación presenta falencias como método para evaluar cuán precisa es la estimación. Según la Comisión Europea (2013), esto se debe a que el valor del porcentaje o la proporción tiene un fuerte impacto en el valor del coeficiente de variación, sobre todo cuando el porcentaje o la proporción es bajo, y a que los coeficientes de variación de los porcentajes o proporciones de cualquier característica no son simétricos. En este sentido, en Comisión Europea (2015) se menciona que, en las encuestas de hogares, los resultados a menudo se presentan como proporciones o porcentajes, y que los intervalos de confianza representan una mejor opción para presentar el error de muestreo aleatorio asociado a la estimación.

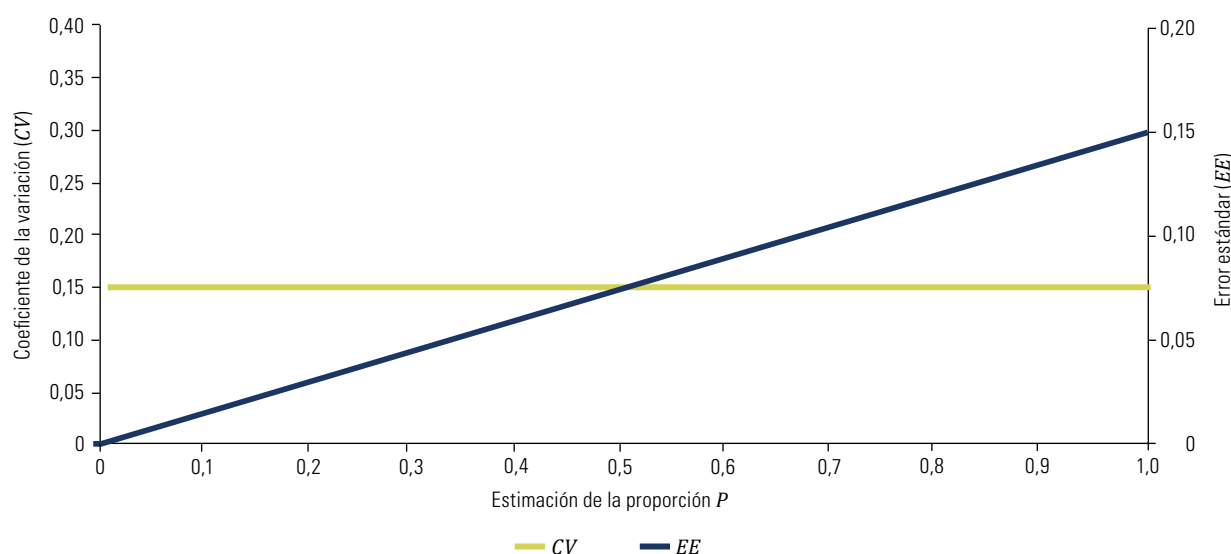
En INE (2020a), por ejemplo, la precisión de las estimaciones se evalúa considerando dos clases de estimadores: i) los que representan proporciones y razones cuyo valor se encuentra entre 0 y 1, y ii) el resto de los indicadores. En línea con ello, en INE (2020b) se analiza el comportamiento del error estándar dado el coeficiente de variación para distintos niveles de la proporción o razón, aplicando dos criterios iniciales:

- i) Según el primer criterio, la estimación de la proporción es de buena calidad si el coeficiente de variación ($CV(\hat{p})$) asociado es menor o igual que 0,15 (15%).

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. No obstante, cuando la prevalencia es baja se exigen errores estándares muy bajos, lo que no necesariamente implica que el intervalo de confianza generado no sea de utilidad para tomar decisiones (véase el gráfico III.2). Por ejemplo, las estimaciones de P que se encuentran cerca de 0,05 pueden tener un intervalo de confianza de amplitud razonable cuando el coeficiente de variación es superior a 0,15, por lo que surge la propuesta de establecer un segundo criterio que consiste en plantear cotas máximas para el error estándar según los valores que tome p (la estimación de P).

Gráfico III.2

Máximo coeficiente de variación y error estándar admitido según la estimación de P (primer criterio)



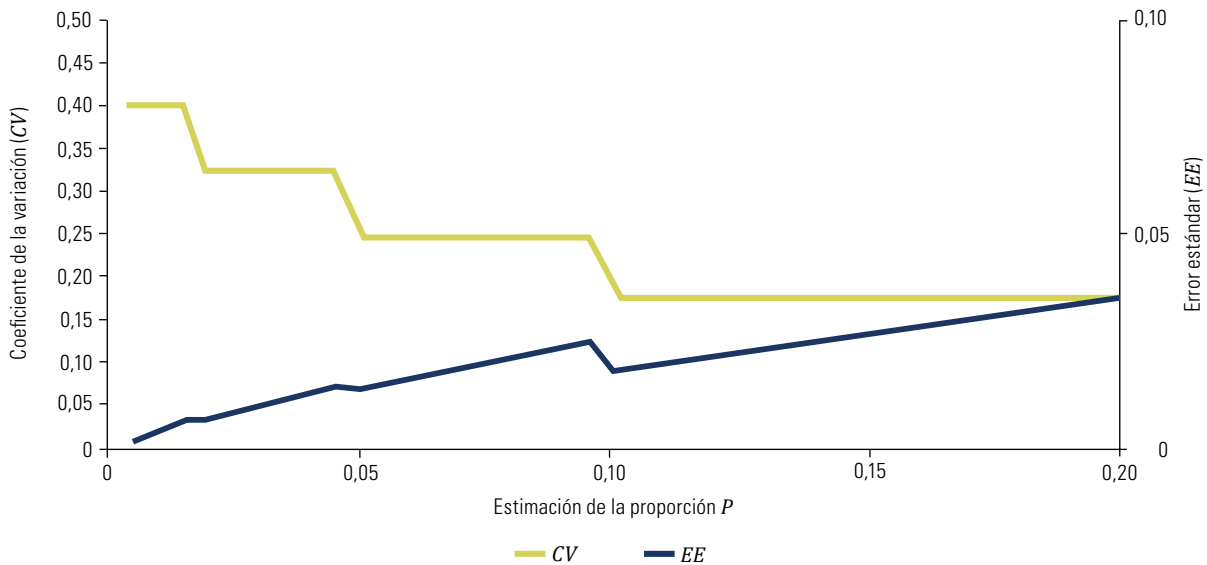
Fuente: Instituto Nacional de Estadísticas de Chile (INE), "Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares", Documento de Trabajo, Santiago, N° 13, 2020.

- ii) De acuerdo con el segundo criterio, la estimación de la proporción es de buena calidad si el coeficiente de variación ($CV(\hat{p})$) asociado coincide con el que figura en el gráfico III.3.

Al definir las cotas según los valores de p , el error estándar sigue una tendencia lineal creciente con leves decrecimientos en torno a los valores de p en que cambian las cotas del coeficiente de variación, es decir, en los valores 0,02, 0,05 y 0,10. Esto implica que, en algunos casos, a un valor de p menor se le exige un error estándar mayor, lo que rompe la tendencia natural lineal del error estándar, que aumenta a medida que los valores de p aumentan de 0 a 0,50. Eso resulta poco razonable.

Gráfico III.3

Máximo coeficiente de variación y error estándar admitido según la estimación de P (segundo criterio)



Fuente: Instituto Nacional de Estadísticas de Chile (INE), "Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares", Documento de Trabajo, Santiago, N° 13, 2020.

Frente a los hallazgos mencionados anteriormente, se descartaron los dos criterios iniciales y se determinó que, cuando la prevalencia es baja, la calidad de la estimación debe medirse considerando el error estándar y no el coeficiente de variación.

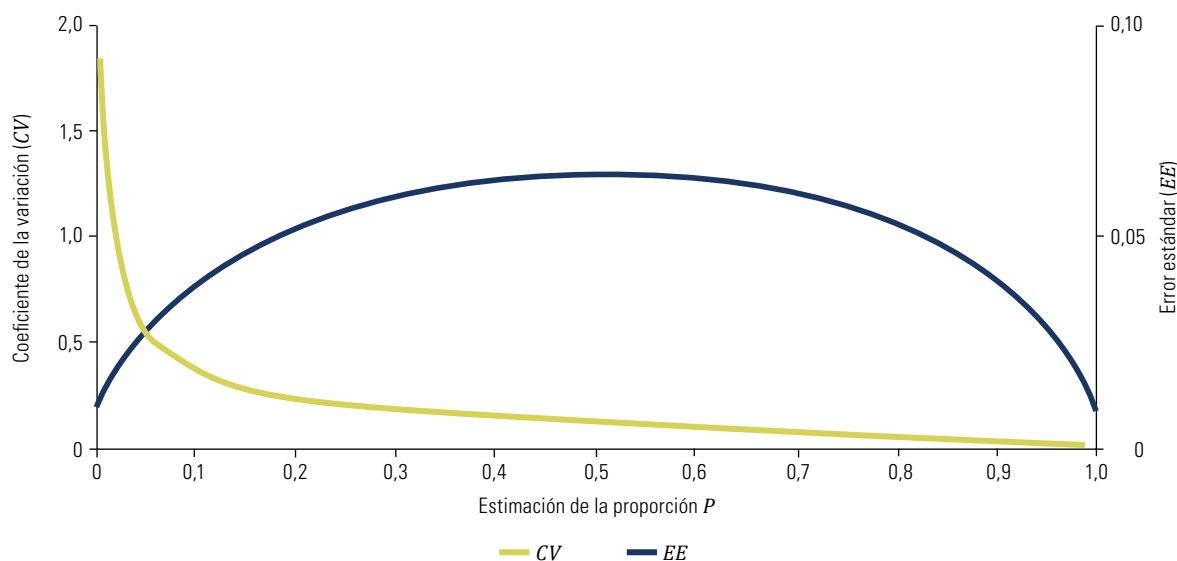
Según INE (2020b), otro aspecto por considerar al evaluar la precisión es la dicotomía de los fenómenos que se miden mediante los estimadores de proporción. Tras la estimación de la proporción de personas victimizadas, por ejemplo, subyace la estimación de las no victimizadas: ambas estimaciones tienen asociado el mismo error estándar, pero no así el CV, que será significativamente mayor para la contraparte de menor prevalencia. En el gráfico III.4 se muestra el comportamiento del error estándar y del coeficiente de variación para varios niveles de p con un tamaño muestral supuesto de 60 unidades. Se precisa que, para $p = 0,05$, el error estándar es 0,028 y equivale a un coeficiente de variación de 0,567; el complemento $1-p = 0,95$, sin embargo, tiene el mismo 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 que la de $1-p$ sí la tiene.

A partir de lo anterior, en INE (2020b) se considera que la calidad de la estimación de las proporciones y razones que tienen un valor de entre 0 y 1 en función de la dispersión debe evaluarse considerando simultáneamente el coeficiente de variación, el error absoluto y la dicotomía de los fenómenos estudiados, así como el modo en que los umbrales precisados inciden en el tamaño muestral requerido. En el próximo apartado se comentarán los indicadores de precisión empleados en INE (2020a) según el tipo de estimador.

En Gutiérrez y otros (2020) se habla de que el coeficiente de variación plantea una paradoja cuando se utiliza para evaluar estimadores de proporción, pues para un mismo fenómeno objeto de medición los coeficientes de variación de la estimación del parámetro de interés y del complemento de la proporción resultan contradictorios. En consecuencia, se presenta la alternativa del coeficiente de variación logarítmico (CV_{Log}), sobre el cual se profundiza en el próximo apartado.

Gráfico III.4

Comportamiento del error estándar y del coeficiente de variación según la estimación de P en una muestra aleatoria simple en que $n=60$



Fuente: Instituto Nacional de Estadísticas de Chile (INE), "Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares", Documento de Trabajo, Santiago, N° 13, 2020.

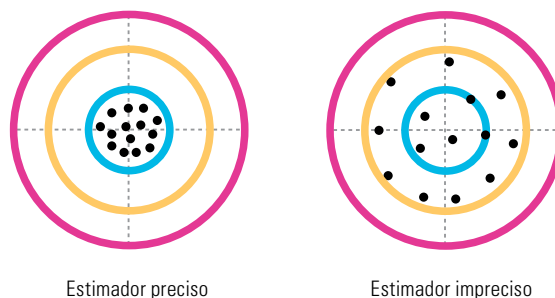
D. Medidas de precisión utilizadas para la publicación y difusión

Luego de validar los criterios de calidad asociados al efecto del diseño, el tamaño muestral y los grados de libertad, corresponde evaluar cuán precisa es la estimación teniendo en cuenta el tipo de estimador asociado.

La precisión de un estimador da cuenta de la concentración o dispersión de los valores estimados en torno al valor que se intenta estimar (error muestral), y se puede afirmar que un estimador es preciso cuando, al repetir la operación muestral reiteradas veces, la magnitud de las desviaciones es pequeña (véase el diagrama III.1).

Diagrama III.1

Precisión de los estimadores determinada por la concentración o dispersión de los valores estimados



Fuente: Comisión Económica para América Latina y el Caribe (CEPAL).

A continuación se presentan las diversas métricas que se emplean en los estándares de calidad examinados para evaluar la precisión de las estimaciones.

1. Error estándar

La precisión de un estimador se mide principalmente a través de la varianza del muestreo, definida como la suma de los desvíos al cuadrado del valor promedio de la estimación obtenido empleando todas las muestras posibles. Dado que es una medida elevada al cuadrado, para simplificar su interpretación se calcula su raíz cuadrada, que se denomina error estándar (*ee*) y constituye una de las medidas del error de muestreo¹⁹.

Debido a las falencias que el coeficiente de variación presenta a la hora de estimar la precisión de las estimaciones que consisten en proporciones o razones cuyo valor es de entre 0 y 1, en INE (2020b) se propone una nueva metodología destinada a medir la calidad de ese tipo de estimaciones en función de la precisión. Esa metodología se basa en definir umbrales del error estándar (en función de p), y en utilizar esos umbrales para clasificar las estimaciones según su fiabilidad en relación con distintos niveles de p .

Para determinar la mejor alternativa se evaluaron tres funciones, una lineal, una logarítmica y una cuadrática, y se concluyó que la función cuadrática resaltaba por ser más parsimoniosa (aspecto que es importante considerar a la hora de implementarla), y por superar las falencias del coeficiente de variación y de las otras alternativas planteadas²⁰. Además, la función cuadrática exige que el tamaño muestral sea mayor cuando la prevalencia es extrema, lo que, sin lugar a duda, crea un margen de resguardo al analizar prevalencias poco frecuentes (INE, 2020b).

La función cuadrática mencionada está definida por:

$$\text{Máximo error (ee) tolerable} = \begin{cases} \frac{\sqrt[3]{p^2}}{9}; & 0 < p \leq 0,5 \\ \frac{\sqrt[3]{(1-p)^2}}{9}; & 0,50 < p < 1 \end{cases} \quad (33)$$

2. Coeficiente de variación

El coeficiente de variación (*CV*) de una estimación es el cociente entre su error estándar y el valor promedio de la propia estimación²¹. Así, el *CV* proporciona una medida del error estándar respecto a la característica que se está midiendo (Naciones Unidas, 2009).

Este indicador es la medida más utilizada por las ONE para la publicación de cifras oficiales. Su uso es transversal porque, por su definición, es de naturaleza relativa, lo que hace que el usuario quede liberado de la unidad de medida inducida por la variable de interés (Gutiérrez y otros, 2020). Esta ventaja permite hacer comparaciones entre variables o indicadores de distinta índole.

En INE (2020a), salvo en el caso de los estimadores de razón o proporción, se emplea el *CV* para analizar la precisión de las estimaciones obtenidas, y se evalúan distintos umbrales que aluden al nivel de fiabilidad de la estimación: cuando el *CV* es menor o igual que el 15%, se dice que la estimación es fiable, cuando es mayor que el 15% pero menor o igual que el 30%, se dice que es poco fiable y, cuando supera el 30%, se dice que es no fiable.

Al examinar los criterios que emplean las distintas ONE, es común observar divergencias en cuanto a los umbrales de evaluación, y esas divergencias se observan tanto entre las encuestas como entre las oficinas de los distintos países. Por citar algunos ejemplos, según lo que se establece en Comisión Europea (2015), si el *CV* es menor o igual que el 10%, la estimación es lo suficientemente precisa y, por lo tanto, se publica sin restricciones; si el *CV* es mayor que el 10% pero menor o igual que el 30%, la estimación es menos precisa y, por lo tanto, se marca con la letra M, y, si el *CV* supera el 30%, la estimación no es suficientemente precisa para ser publicada y, por lo tanto, se sustituye por la letra N. Por otra parte, en lo que atañe a la Encuesta Permanente de Hogares (EPH) de

¹⁹ La expresión del error estándar $EE(\hat{\theta})$ se presenta en el apartado b).

²⁰ Para obtener más información, véase INE (2020b).

²¹ La expresión del coeficiente de variación $CV(\hat{\theta})$ se presenta en el apartado c).

la Argentina, el Instituto Nacional de Estadística y Censos (INDEC) de ese país considera que, si el CV es menor o igual que el 16%, la estimación es confiable, si es mayor que el 16% pero menor o igual que el 25%, es poco confiable y, cuando supera el 25%, es no fiable.

3. Coeficiente de variación logarítmico

En Gutiérrez y otros (2020) se señalan las falencias del CV en el caso particular de las proporciones. Por ejemplo, cuando la estimación del parámetro de interés es muy cercana a cero, sin importar qué tan pequeña sea su varianza, el coeficiente de variación será muy grande y no representará la calidad de la estrategia de muestreo. Sin embargo, el CV del complemento de la proporción ($1 - P$) será muy pequeño y confiable, lo que constituye una contradicción. Como alternativa, Barnett-Walker y otros (2003) plantean utilizar una transformación logarítmica sobre la proporción y aplicar lo que se denomina coeficiente de variación logarítmico ($CVLog$) para medir la precisión de las proporciones estimadas.

Si $P \leq 0,5$, se define $\hat{L} = -\log(\hat{P})$, donde la aproximación de Taylor de primer orden es:

$$\hat{L} \cong L + \frac{\partial \hat{L}}{\partial \hat{P}} \Big|_{\hat{P}=P} (\hat{P} - P) = L + \left(-\frac{1}{P}\right) (\hat{P} - P) \quad (34)$$

Luego, la varianza de \hat{L} será $Var(\hat{L}) \cong AVar(\hat{L}) = \frac{Var(\hat{P})}{P^2}$ y, por consiguiente, el error estándar de la transformación equivaldrá al coeficiente de variación de la proporción, donde:

$$EE(\hat{L}) = \sqrt{AVar(\hat{L})} = \frac{\sqrt{Var(\hat{P})}}{\hat{P}} = CV(\hat{P}) \quad (35)$$

De esta manera, podemos definir una medida de suavizamiento como el coeficiente de variación asociado a la transformación:

$$CV(\hat{L}) = \frac{EE(\hat{L})}{\hat{L}} = \frac{CV(\hat{P})}{\hat{L}} \quad (36)$$

De manera similar, para mantener la simetría y no perder generalidad, cuando $p > 0,5$ se realiza un ajuste definiendo $\hat{L} = -\log(1 - \hat{P})$. Consecuentemente, cuando las proporciones se ubican dentro del intervalo central, específicamente entre 0,2 y 0,8, observamos que los coeficientes de variación para las estimaciones, representadas como \hat{P} y $1 - \hat{P}$, muestran una similitud notable. La razón subyacente es que $1 - \hat{P}$ tiende a acercarse a valores próximos a uno en este espectro de proporciones. Este comportamiento conduce a que el coeficiente de variación de $1 - \hat{P}$, o $CV(1 - \hat{P})$, presente una correspondencia estrecha con el coeficiente de variación de \hat{P} , denotado como $CV(\hat{P})$.

En el gráfico III.5 se muestra que, al igual que ocurría con el coeficiente de variación original, el tamaño de la muestra aumentará a medida que se requiera mayor precisión en la estimación; no obstante, a diferencia del coeficiente de variación original, el tamaño de la muestra será idéntico cuando se trate de fenómenos que induzcan proporciones simétricas.

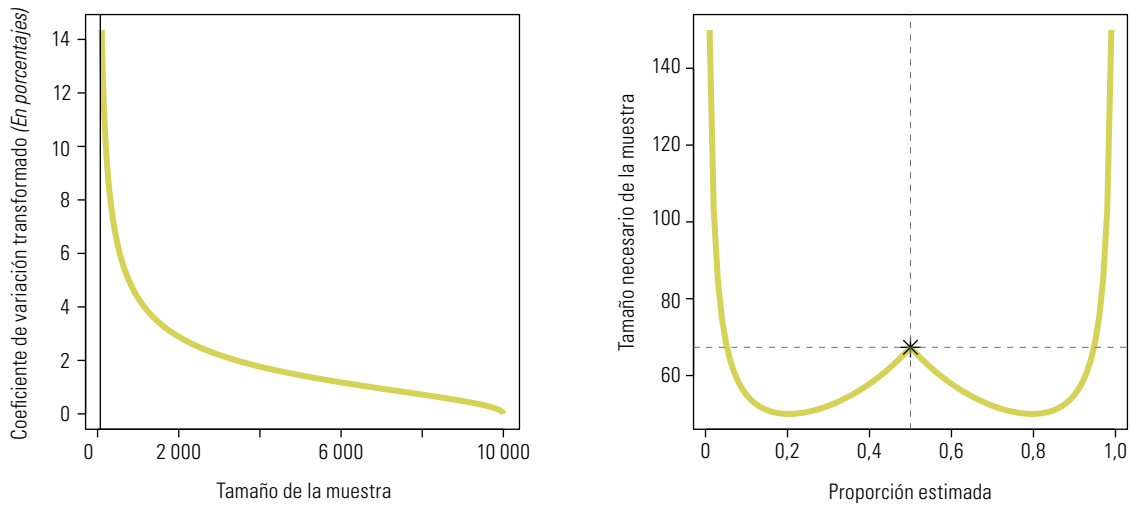
Finalmente, según el valor de \hat{P} , la expresión para evaluar el coeficiente de variación logarítmico ($CVLog$) sería la siguiente:

$$CVlog = \begin{cases} \frac{CV(p)}{-\log(p)}, & p < 0,5 \\ \frac{CV(p)}{-\log(1-p)}, & p \geq 0,5 \end{cases} \quad (37)$$

Asimismo, Gutiérrez y otros (2020) sugieren que como umbral de decisión se utilice el que señalan Barnett-Walker y otros (2003), quienes proponen que la cifra se suprima si el $CVLog$ supera el 17,5%.

Gráfico III.5

Relación entre el tamaño de la muestra y la precisión de un indicador utilizando la transformación logit



Fuente: A. Gutiérrez y otros, "Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional", *serie Estudios Estadísticos*, N° 101 (LC/TS.2020/52), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL), 2020.

E. Conteo de casos no ponderado

El conteo de casos no ponderado se refiere al número de elementos de la muestra que presentan la característica de interés. Según Gutiérrez y otros (2020), este indicador tiene un efecto directo en la determinación de la precisión del estimador de interés, y adquiere especial relevancia en el caso de los estimadores de proporción y razón. La fórmula está determinada por la siguiente expresión:

$$n = \sum_s \delta_k^y \tag{38}$$

en donde δ_k^y es una variable indicadora de cada individuo k de la muestra s , variable que toma el valor 1 si el individuo está afectado por el fenómeno que la variable de interés y representa. Nótese que esta es una cantidad aleatoria por definición, y que también es posible calcularla en la muestra de un subgrupo poblacional específico $U_{g'}$, de la siguiente manera:

$$n_y^g = \sum_s z_{gk} \delta_k^y = \sum_{s_g} \delta_k^y \tag{39}$$

Si la incidencia del fenómeno es muy baja (cuando la proporción P es cercana a cero), tanto el coeficiente de variación original como su transformación logarítmica tendrán magnitudes altas, puesto que:

$$\lim_{n_y \rightarrow 0} CV(\hat{\theta}) = \lim_{n_y \rightarrow 0} CV(\hat{L}) = \infty \tag{40}$$

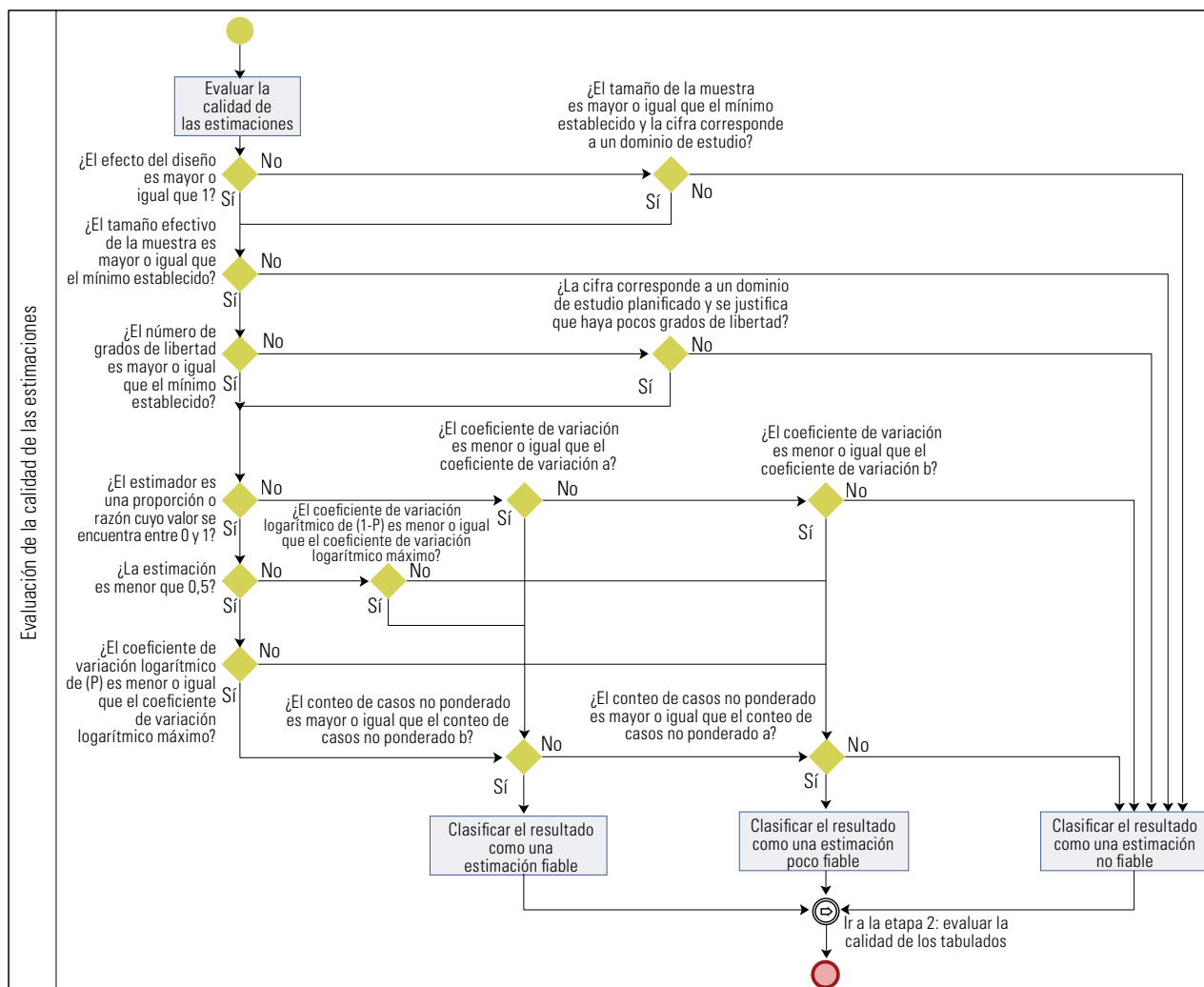
Capítulo IV

Requerimientos de calidad que se aplican a las estimaciones

En el presente apartado se pretende formular un estándar de calidad armonizado en que se propongan umbrales de decisión correspondientes a cada criterio de calidad, así como lineamientos que orienten a los grupos usuarios sobre el uso que se puede dar a las estimaciones obtenidas.

En el diagrama IV.1 se proporciona un flujo de acciones que tiene por objeto marcar el camino que se ha de seguir para evaluar la calidad de las estimaciones obtenidas mediante las encuestas de hogares por muestreo de acuerdo con lo planteado por el grupo de trabajo en la propuesta armonizada.

Diagrama IV.1
Flujograma de evaluación de las estimaciones



Fuente: Comisión Económica para América Latina y el Caribe (CEPAL).

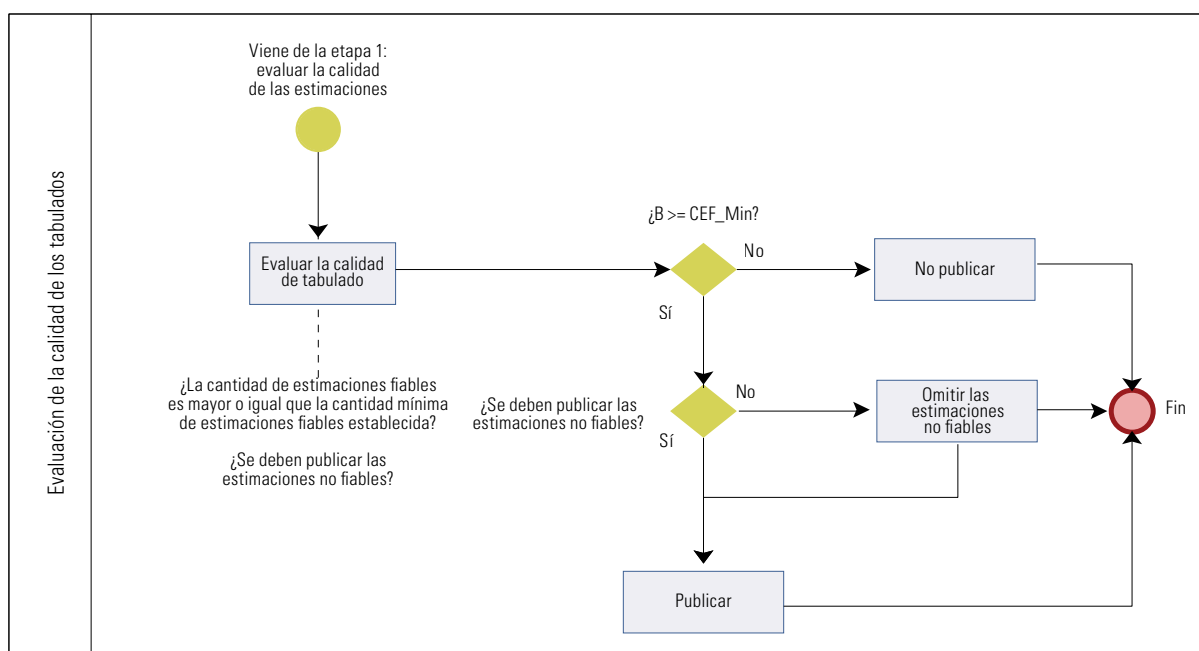
A continuación se proporciona una guía del flujo presentado en el diagrama IV.1, y en ella se indican los rangos de aceptación de cada criterio y los argumentos que los sustentan:

- Efecto del diseño (*Deff*): el *Deff* se debe calcular en relación con la variable de interés al nivel desagregado en que se encuentre la estimación. Se sugiere establecer un umbral de aceptación igual o superior a 1. Si bien es raro obtener *Deff* menores que 1, en los casos en que esto ocurra se sugiere revisar el mecanismo de estimación del *Deff* en el *software* empleado, así como estudiar la homogeneidad del fenómeno en el dominio de estudio. Si el *Deff* es inferior a 1, el tamaño muestral no es mayor que el umbral establecido en el párrafo siguiente y la cifra no corresponde a un dominio de estudio, se recomienda clasificar el resultado como una estimación no fiable. Cuando el *Deff* es superior a 1, el flujo prosigue con el tamaño efectivo de la muestra que se describe más adelante.
- Tamaño de la muestra: este elemento es crucial al determinar la ruta de publicación de un dato estadístico, ya que los avances teóricos en inferencia estadística aplicada a las encuestas se basan en este elemento. Tanto la precisión de los intervalos de confianza como la distribución de los estimadores están sujetos a que el tamaño de la subpoblación y el tamaño de la muestra correspondiente no sean reducidos. En este contexto, Barnett-Walker y otros (2003) sugieren que cualquier estimación basada en un tamaño de muestra inferior a 100 unidades debe ser eliminada o marcada como no fiable.
- Tamaño efectivo de la muestra: con base en los umbrales de aceptación del estándar del Instituto Nacional de Estadística (INE) de Chile y de la CEPAL, se sugiere que el umbral de aceptación sea de entre 60 y 100 unidades. Este umbral garantiza la convergencia de la distribución de los estimadores, y eso permite hacer inferencias sobre la población. En caso de no alcanzarse este umbral, se recomienda clasificar la estimación como no fiable; en caso contrario, se sugiere pasar a evaluar los grados de libertad.
- Grados de libertad: dado que las simulaciones muestran que a partir de los 13 grados de libertad es posible demostrar empíricamente la convergencia de los estimadores de razón, y considerando que en Parker y otros (2017) se recomienda que haya 8 grados de libertad, se sugiere que el número mínimo de grados de libertad para las estimaciones sea de entre 9 y 13. Si la estimación no tiene suficientes grados de libertad, se recomienda clasificarla como no fiable, con excepción de los casos que correspondan a dominios de estudio y sea posible justificar la razón por la que hay pocos grados de libertad. Para calcular los grados de libertad se debe emplear el cálculo de los grados variables, esto es, el número de unidades primarias de muestreo con observaciones en la subpoblación menos el número de estratos de la subpoblación que cuentan con observaciones. Una vez evaluado este elemento, el flujo prosigue con el coeficiente de variación o coeficiente de variación logarítmico.
- Coeficiente de variación: este elemento se emplea para evaluar estimaciones que no correspondan a proporciones ni razones cuyo valor se encuentre entre 0 y 1. En ese sentido, la mayoría de las oficinas nacionales de estadística (ONE) clasifican como dato no publicable las estimaciones cuyo coeficiente de variación supera el 20% (Molina, 2019). Se recomienda que el valor de aceptación para determinar que una estimación es fiable se sitúe entre 15 y 20 (lo que en el flujograma se menciona como coeficiente de variación a)). Para determinar que una estimación es poco fiable se aconseja un umbral de entre 25 y 30 y, en el caso de que el coeficiente de variación sea mayor que 30 (lo que en el flujograma se menciona como coeficiente de variación b)), se recomienda clasificar la cifra como no fiable.
- Coeficiente de variación logarítmico: este elemento se emplea para evaluar las estimaciones que son una proporción o una razón cuyo valor se encuentra entre 0 y 1. De conformidad con lo que proponen Barnett-Walker y otros (2003), cuando el tamaño efectivo de la muestra es 60, se recomienda utilizar un umbral del 17,5% para considerar que una estimación es fiable; si la cifra supera ese valor, se recomienda considerarla poco fiable o no fiable conforme a los umbrales establecidos en el siguiente punto. Este umbral de aceptación puede modificarse de acuerdo con el valor de la función logarítmica del coeficiente de variación evaluada en $p = 0,5$, según los valores que puede tomar el tamaño efectivo de la muestra.

- **Conteo de casos no ponderado:** el conteo de casos no ponderado corresponde al conteo de unidades dentro de la muestra que presentan el fenómeno de interés en estudio. Se recomienda clasificar la estimación como no fiable si el conteo es menor que 30 (lo que en el flujograma se representa como el umbral de conteo de casos no ponderado a)). En el caso de que el conteo sea mayor que esa cifra pero menor o igual que 40 o 50 (la cifra escogida dependerá de cada institución), se recomienda clasificar la estimación como poco fiable. Cuando el conteo de casos no ponderado de la estimación sea superior a 50, se recomienda clasificar la estimación como fiable (lo que en el flujograma se representa como conteo de casos no ponderado b)).

En el diagrama IV.2 se proporciona el flujo de evaluación de la calidad de los tabulados, que es la etapa siguiente a la evaluación de la calidad de las estimaciones.

Diagrama IV.2
Flujograma de evaluación de los tabulados



Fuente: Comisión Económica para América Latina y el Caribe (CEPAL).

En el flujo del diagrama IV.2 se recomienda no publicar el tabulado si la cantidad de estimaciones fiables es menor que un umbral mínimo establecido. A la hora de determinar dicho umbral se sugiere que la cantidad de estimaciones fiables sea mayor o igual que la mitad más 1 del total de las estimaciones que figuran en el tabulado (esto es, $A/2 + 1$, donde A es la cantidad de estimaciones que figuran en el tabulado). Si se decide publicar el tabulado, es necesario evaluar si se deben publicar las estimaciones no fiables o si estas se deben omitir.

El flujo correspondiente a las estimaciones y el correspondiente a los tabulados ofrecen a las ONE de la región una propuesta que les permitirá velar por la calidad de los productos estadísticos que elaboran. Considerar los umbrales y los elementos que se deben tener en cuenta respecto de la fiabilidad de las estimaciones permite definir y llevar a cabo acciones necesarias en las fases del diseño y la recolección, el procesamiento y el análisis de los datos. De conformidad con Gutiérrez y otros (2020), cuando en las estimaciones haya indicios de que hay dificultades de cobertura, se recomienda realizar un análisis de sesgo y ajustar los factores de expansión para mitigar los posibles efectos.

Capítulo V

Consideraciones finales

En el contexto de la producción estadística basada en procesos, cuyo marco de referencia es el Modelo Genérico de Procesos Institucionales Estadísticos (MGPIE) de la Comisión Económica para Europa (CEPE), las orientaciones vertidas en este documento están dirigidas a la fase de análisis de los resultados, específicamente a las etapas de preparación, finalización de resultados y aplicación de protocolos de calidad. La fase de análisis de los resultados está dedicada a la creación del contenido de los informes que se van a publicar respecto de cada operación estadística, como una encuesta, y asegura que las estadísticas y la información que se incluyan cumplan criterios de calidad y estén disponibles para que la sociedad las use, analice e interprete.

No obstante lo anterior, para garantizar la calidad de las encuestas es fundamental velar por la calidad durante todo el proceso de producción, desde la fase de diseño hasta la fase de difusión. La calidad de la encuesta depende en gran medida del diseño, y es esencial que este se estructure bien y sea coherente con los objetivos planteados. Las variables de interés, la población objetivo y el marco de muestreo deben estar claramente definidos.

Además, es crucial poner un cuidado especial en las fases de recolección y procesamiento de los datos. En esas etapas es fundamental implementar procedimientos robustos, lo que supone capacitar adecuadamente al personal de campo, aplicar mecanismos de control de calidad durante la recolección que impliquen hacer el seguimiento indicadores en el transcurso de esta, y poner en práctica procedimientos rigurosos de verificación y limpieza de los datos recogidos.

En el proceso de análisis, por su parte, la confiabilidad de las estimaciones es otro aspecto vital que no se puede pasar por alto y que depende de la calidad de los datos recolectados y procesados. Por tanto, es esencial realizar análisis detallados y rigurosos de los datos para garantizar la fiabilidad de las estimaciones, y en el presente documento se señalan elementos específicos que se deben medir y evaluar.

La revisión y mejora continua son fundamentales en este proceso, y es necesario que en cada versión de las encuestas se evalúe si los resultados cumplen los objetivos planteados y que, de ser necesario, se establezca una hoja de ruta de las mejoras que se deben considerar en el próximo ciclo de producción.

Finalmente, además de estas recomendaciones es importante recordar que la transparencia y la comunicación efectiva son esenciales para garantizar la calidad de las encuestas. La transparencia y la comunicación implican compartir los resultados de la encuesta de manera oportuna y en un formato fácil de comprender, y estar dispuestos a responder las preguntas y preocupaciones de los participantes y otras partes interesadas.

Bibliografía

- ANDA (Archivo Nacional de Datos) (2023), "Catálogo de Metadatos y Microdatos" [base de datos en línea] <https://anda.ine.gob.bo/index.php/home>.
- (2020), "Política de difusión" [en línea] <https://anda.ine.gob.bo/index.php/politicas-difusion>.
- Barnett-Walker, K. y otros (2003), *2001 National Household Survey on Drug Abuse: Statistical Inference Report*, Rockville, Administración de Salud Mental y Abuso de Sustancias (SAMHSA).
- CEPAL (Comisión Económica para América Latina y el Caribe) (2022), *Guía para la implementación del marco de aseguramiento de la calidad para procesos y productos estadísticos* (LC/CEA.11/19), Santiago.
- (2011), *Código Regional de Buenas Prácticas en Estadísticas para América Latina y el Caribe*, Santiago.
- Chambers, R. y C. Skinner (eds.) (2003), *Analysis of Survey Data*, Hoboken, Wiley.
- Cochran, W. (1980), *Técnicas de muestreo*, Ciudad de México, Compañía Editorial Continental (CECSA).
- (1977), *Sampling Techniques*, Hoboken, Wiley.
- Comisión Europea (2015), *ESS Handbook for Quality Reports*, Bruselas.
- (2013), *Handbook on Precision Requirements and Variance Estimation for ESS Households Surveys: 2013 Edition*, Bruselas.
- CRAN (Comprehensive R Archive Network) (2023), "Calidad: assesses the quality of estimates made by complex sample designs"; Viena [en línea] <https://cran.r-project.org/web/packages/calidad/index.html>.
- DIGESTYC (Dirección General de Estadística y Censos) (2022), "Recomendaciones sobre criterios de supresión para investigaciones por muestreo probabilístico", San Salvador.
- Groves, R. y otros (2004), *Survey Methodology*, Hoboken, Wiley.
- Gutiérrez, A. (2022), *Diseño y análisis estadístico en las encuestas de hogares de América Latina* (LC/PUB.2023/14-P), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).
- Gutiérrez, A. y otros (2020), "Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional"; *serie Estudios Estadísticos*, N° 101 (LC/TS.2020/52), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).
- Hansen, M., W. Hurwitz y W. Madow (1953), *Sample Survey Methods and Theory*, Nueva York, Wiley.
- Heeringa, S., B. West y P. Berglund (2010), *Applied Survey Data Analysis*, Boca Ratón, CRC Press.
- Hornik, R. y otros (2002), *Evaluation of the National Youth Anti-Drug Media Campaign: Fifth Semi-Annual Report of Findings*, Rockville, Westat.
- IBGE (Instituto Brasileño de Geografía y Estadística) (2021a), *Guia para Divulgação de Erros Amostrais nas Pesquisas por Amostragem Probabilística Realizadas pelo IBGE*, Río de Janeiro.
- (2021b), *Código de Boas Práticas das Estatísticas do IBGE*, Río de Janeiro.
- INDEC (Instituto Nacional de Estadística y Censos de la Argentina) (s.f.), "Notas técnicas" [en línea] <https://www.indec.gob.ar/indec/web/Institucional-Indec-Metodologias-2>.
- INE (Instituto Nacional de Estadísticas de Chile) (2020a), *Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares*, Santiago.
- (2020b), "Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares", *Documento de Trabajo*, N° 13, Santiago.
- INE (Instituto Nacional de Estadística del Uruguay) (2021), *Norma Técnica de Certificación de la Calidad de Operaciones Estadísticas INE-CCOE: 2021-01*, Montevideo.
- Judkins, D. (1990), "Fay's method for variance estimation", *Journal of Official Statistics*, vol. 6, N° 3, Estocolmo, Oficina de Estadística de Suecia.
- Kalton, G., J. Brick y T. Lê (2005), "Estimating components of design effects for use in sample design", *Household Sample Surveys in Developing and Transition Countries*, series F, N° 96 (ST/ESA/STAT/SER.F/96), Nueva York, Naciones Unidas.
- Kish, L. (1965), *Survey Sampling*, Nueva York, Wiley.
- Klein, R. y otros (2002), "Healthy People 2010 criteria for data suppression"; *Healthy People 2010 Statistical Notes*, N° 24, Washington, D.C., American Psychological Association (APA).
- Korn, E. y B. Graubard (1999), *Analysis of Health Surveys*, Hoboken, Wiley.
- Krewski, D. y J. Rao (1981), "Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods", *The Annals of Statistics*, vol. 9, N° 5, Durham, Instituto de Estadística Matemática.

- Lavrakas, P. (2008), "Probability proportional to size (PPS) sampling," *Encyclopedia of Survey Research Methods*, Thousand Oaks, SAGE Publications.
- Lohr, S. (1999), *Muestreo: diseño y análisis*, Ciudad de México, International Thomson Editores.
- Lumley, T. (2010), *Complex Surveys: A Guide to Analysis Using R*, Hoboken, Wiley.
- McCarthy, P. (1969), "Pseudo-replication: half samples," *Review of the International Statistical Institute*, vol. 37, N° 3, La Haya, Instituto Internacional de Estadística (IIE).
- Martínez, C. (2019), *Estadística y muestreo*, Bogotá, ECOE Ediciones.
- Mirás, J. (1985), *Elementos de muestreo para poblaciones finitas*, Madrid, Instituto Nacional de Estadística (INE).
- Molina, I. (2019), "Desagregación de datos en encuestas en hogares: metodologías de estimación en áreas pequeñas," *serie Estudios Estadísticos*, N° 97 (LC/TS.2018/82), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).
- MTESS/OISS/SRT (Ministerio de Trabajo, Empleo y Seguridad Social/Organización Iberoamericana de Seguridad Social/Superintendencia de Riesgos del Trabajo) (2018), *Encuesta Nacional a Trabajadores sobre Condiciones de Empleo, Trabajo, Salud y Seguridad (ECETSS) 2018: guía de referencia para estimaciones de errores de muestreo por medio de bootstrap*, Buenos Aires.
- Naciones Unidas (2009), "Diseño de muestras para encuestas de hogares: directrices prácticas," *Estudios de Métodos*, serie F, N° 98 (ST/ESA/STAT/SER.F/98), Nueva York.
- Oficina de Estadística del Canadá (2003), *Statistics Canada Quality Guidelines Fourth Edition – October 2003*, Ottawa.
- Parker, J. y otros (2017), "National Center for Health Statistics data presentation standards for proportions: data evaluation and methods research," *Vital and Health Statistics*, vol. 2, N° 175, Hyattsville, Centro Nacional de Estadísticas de la Salud.
- Pennell, B. y otros (2017), "A total survey error perspective on surveys in multinational, multiregional, and multicultural contexts," *Total Survey Error in Practice*, P. Biemer y otros (eds.), Hoboken, Wiley.
- Pérez, C. (2005), *Muestreo estadístico: conceptos y problemas resueltos*, Madrid, Prentice Hall.
- Rao, J. y C. Wu (1984), "Bootstrap inference for sample surveys," *Proceedings of the American Statistical Association Survey Research Methods Section*, Alexandria, American Statistical Association (ASA).
- Särndal, C., B. Swensson y J. Wretman (1992), *Model Assisted Survey Sampling*, Berlín, Springer.
- Tillé, Y. y D. Haziza (2010), "An interesting property of the entropy of some sampling designs," *Survey Methodology*, vol. 36, N° 2, Ottawa, Statistics Canada.
- Tillé, Y. y M. Wilhelm (2017), "Probability sampling designs: principles for choice of design and balancing," *Statistical Science*, vol. 32, N° 2, Durham, Instituto de Estadística Matemática.
- Valliant, R., J. Dever y F. Kreuter (2018), *Practical Tools for Designing and Weighting Survey Samples*, Berlín, Springer.
- Walker, H. (1940), "Degrees of freedom," *Journal of Educational Psychology*, vol. 31, N° 4, Washington, D.C., American Psychological Association (APA).



Comisión Económica para América Latina y el Caribe (CEPAL)
Economic Commission for Latin America and the Caribbean (ECLAC)
www.cepal.org



LC/CEA.12/11