Diseño y análisis estadístico en las encuestas de hogares de América Latina

Andrés Gutiérrez[1]

2021-12-29

[1]Experto regional en estadísticas sociales - Unidad de Estadística Social - Comisión Económica para América Latina y el Caribe (CEPAL) - andres.gutierrez@cepal.org

Índice general

Índice de cuadros

Índice de figuras

Prefacio

Figura 1: Licencia de Creative Commons

Este libro es el resultado de un compendio de las experiencias internacionales prácticas adquiridas por el autor como Experto Regional en Estadísticas Sociales de la CEPAL.

Resumen

Las encuestas de hogares son un instrumento necesario para realizar seguimiento a un conjunto amplio de indicadores requeridos para el diseño y evaluación de las políticas públicas. Las encuestas de hogares que se implementan en América Latina son de tipo y características diversas. Aunque los conceptos y procesos para su diseño y análisis guardan similitudes, este documento se enfoca principalmente en los procesos referidos a las encuestas de empleo y de propósitos múltiples, con las que los países estiman los principales indicadores relacionados con el mercado laboral, el nivel y distribución de ingresos y la condición de pobreza y las principales características sociodemográficas de la población. Se realiza un recorrido por los diferentes diseños de muestreo, las metodologías más usadas en la selección de las muestras y las estrategias de estimación de los parámetros de interés. También se revisan las técnicas utilizadas para medir el error de muestreo y los métodos disponibles para encarar desafíos como la ausencia de respuesta y la desactualización de los marcos de muestreo.

UNBIS Keywords. Encuestas por muestreo, encuestas de hogares, indicadores socioeconómicos.

Capítulo 1

Introducción

Las encuestas de hogares son un caso particular de investigación social que indaga acerca de características específicas a nivel del individuo, del hogar o de la vivienda, con el fin de obtener inferencias precisas acerca de constructos de interés. Por su naturaleza, estas investigaciones están relacionadas con variables de salud, educación, ingresos, gastos, situación laboral, acceso y uso de servicios, entre muchas otras. En algunas ocasiones, las encuestas de hogares tienen como objetivo la estimación de uno o varios indicadores que resumen un constructo económico o social. Sin embargo, existe una tendencia creciente de extender las encuestas a constructos más diversos. Es así como cada vez tienen más espacio las encuestas de propósitos múltiples como una fuente relevante de información que permite monitorear indicadores sociales.

En este tipo de encuestas, el hogar es la unidad de análisis, la cual ha sido definida por la División de Estadística de la Organización de las Naciones Unidas (ONU, 2011) como:

a. Un grupo de dos o más personas que se combinan para ocupar la totalidad o parte de una vivienda y para proporcionarse alimentos y posiblemente otros artículos esenciales para la vida. El grupo puede estar compuesto solo de personas relacionadas o de personas no relacionadas o de una combinación de ambos. El grupo también puede compartir sus ingresos.
b. Una persona que vive sola en una vivienda separada o que ocupa, como huésped, una habitación (o habitaciones) separada de una vivienda pero que no se une a ninguno de los otros ocupantes de la vivienda para formar parte de una hogar de múltiples personas.

Nótese que la anterior definición refleja la dinámica natural del cambio en las poblaciones de hogares, por lo cual se deben tener distintos enfoques para abordar el problema de la medición de indicadores sociales. En América Latina, existen una gran variedad de encuestas que abordan diferentes problemáticas sociales. Todas y cada una de ellas han sido diseñadas cuidadosamente para que respondan a las necesidades de la sociedad. Este documento plantea una recopilación de las técnicas usadas tanto en su diseño, como en su análisis.

No todas las encuestas se diseñan de la misma forma y por ende debe haber una distinción entre ellas. Por ejemplo, Kalton and Citro (1993) afirman que las encuestas de hogares pueden clasificarse en varios tipos:

- *Encuestas repetidas*, definidas como una serie de encuestas transversales aplicadas en diferentes momentos del tiempo con el mismo diseño metodológico, en donde la selección de hogares se hace de forma independiente para cada aplicación.

- *Encuestas tipo panel*, para las cuales los datos son recolectados en diferentes momentos del tiempo utilizando la misma muestra de hogares en el tiempo.
- *Encuestas rotativas*, en donde un porcentaje de hogares se mantiene en un periodo de tiempo respondiendo la encuesta y en cada aplicación algunos hogares son reemplazados por nuevos hogares de forma planificada.

El diseño de la encuesta dependerá sistemáticamente del objetivo de la medición. Por ejemplo, Kalton (2009) afirma que es prudente hacer un buena inversión en el desarrollo e implementación de un buen diseño para amortizar los costos de todo el estudio. Por lo tanto, lo que se quiere al diseñar una encuesta de hogares es que sea un instrumento confiable, que brinde estimaciones exactas y precisas, puesto que de lo contrario no se podrían monitorear las políticas públicas y los indicadores de interés de forma consistente. Por ejemplo, uno de los indicadores sociales con mayor impacto es la tasa de desocupación, que mide la razón entre la cantidad de personas que se encuentran desocupados, pero que forman parte del mercado de trabajo. Las encuestas de empleo tienen características particulares, diferentes a las de las encuestas que miden otro tipo de constructos. Duncan and Kalton (1987) mencionan que las encuestas de hogares pueden proveer estimaciones de los parámetros poblacionales en distintos puntos del tiempo, por ejemplo, la estimación de la tasa de desocupación mensual; proveer estimaciones del cambio neto de los parámetros poblacionales entre periodos de tiempo, por ejemplo, el cambio en la tasa de desocupación entre dos periodos consecutivos; o incluso medir varios componentes de cambio individual, por ejemplo cambios brutos en la situación laboral de los jefes de hogar, para lo cual se requiere que la encuesta contemple un diseño de panel o de panel rotativo.

La medición de los indicadores en el mercado de trabajo es sólo un pequeño componente en el basto universo de posibilidades de medición que brindan las encuestas de hogares. Por esta razón, este tipo de levantamientos se ha convertido en una herramienta fundamental para medir indicadores sociales en todo el mundo y que, en particular, permiten que las naciones de América Latina puedan hacer seguimiento a su desarrollo económico y social. A continuación se introducen algunas temáticas de interés para las cuales, su seguimiento depende en gran manera de la realización de encuestas de hogares.

Objetivos de Desarrollo Sostenible

Las encuestas de hogares pueden ser utilizado como herramienta para monitorear el progreso de los países en términos de metas y objetivos comunes. Es así como en 2015, la Asamblea General de la Organización de las Naciones Unidas aprobó una resolución que plantea un plan de acción en favor de las personas, el planeta y la prosperidad (ONU, 2015). Esa resolución propone el seguimiento de 17 Objetivos de Desarrollo Sostenible (ODS) y 169 metas de carácter integrado e indivisible que se conjugan en las dimensiones económica, social y ambiental. Para realizar el seguimiento a los ODS es posible utilizar diferentes fuentes de información, como censos, registros administrativos, registros estadísticos, proyecciones demográficas y también las encuestas de hogares (ONU, 2016). En particular, cada una de las metas de los ODS contiene indicadores, muchos de los cuales no pudieran ser estimados de no ser por la información disponible en las encuestas de hogares.

Por ejemplo, el primer objetivo busca *poner fin a la pobreza en todas sus formas en todo el mundo*. La primera meta de este objetivo establece la *erradicación de la pobreza extrema para todas las* *personas en el mundo*. Asimismo, la segunda meta de este objetivo motiva a los países a *reducir al menos a la mitad la proporción de hombres, mujeres y niños y niñas de todas las edades que viven en la pobreza en todas sus dimensiones con arreglo a las definiciones nacionales*. Para realizar una medición sistemática de la pobreza monetaria las encuestas de hogares son un insumo fundamental. En una primera instancia se deben definir a nivel nacional los umbrales monetarios (líneas de pobreza) sobre los cuales se clasifican a los hogares como pobres extremos, pobres relativos o no pobres. Estos umbrales vienen supeditados directamente a la realización de las encuestas de ingresos y gastos, las cuales se realizan cada cinco o diez años en los países de la región (CEPAL 2018). De forma sistemática, la medición de la pobreza monetaria se realiza con encuestas continuas que contienen módulos específicos de ingreso, los cuales indagan por todas las fuentes de ingreso, tanto de las personas como del hogar. Con base en las líneas de pobreza definidas anteriormente, se clasifica a las personas en alguna de las categorías de la pobreza.

De la misma manera, el objetivo **8** busca *promover el crecimiento económico sostenido, inclusivo y sostenible, el empleo pleno y productivo y el trabajo decente para todos*. Claramente de este objetivo se desprenden indicadores que permiten conocer la evolución de los países en la consecución de las metas. Dentro de este objetivo, se encuentra la meta **8.6** que apunta a reducir sustancialmente la proporción de jóvenes sin empleo y sin educación o entrenamiento. Esta meta se mide con el indicador **8.6.1** definido como la proporción de jóvenes (entre 15 y 24 años de edad) sin educación y sin empleo.

Son muchísimos más los ejemplos que se pueden enumerar en los cuales las encuestas de hogares juegan un rol fundamental para la medición de los indicadores y metas de los ODS definidos en la Agenda 2030; en este sentido la División de Estadísticas de las Naciones Unidas ha establecido en un análisis preliminar que un total de 77 de los indicadores de los Objetivos de Desarrollo Sostenible pueden obtenerse a partir de encuestas de hogares, cubriendo 13 de los 17 Objetivos; aunque con mayor concentración en las áreas de salud, educación, igualdad de género, pobreza, hambre, trabajo y justicia.

Mercado de Trabajo

Desde otra perspectiva, en el marco de la Decimotercera Conferencia Internacional de Estadísticos del trabajo en 1982, la Organización Internacional del Trabajo (OIT) adoptó algunas directrices concernientes con la medición y análisis de estadísticas oficiales de la fuerza de trabajo, del empleo y del desempleo con miras a mejorar la comparabilidad de las cifras y mejorar su utilidad en los países (OIT, 1982). En esta resolución se hace un énfasis especial en que las encuestas de hogares constituyen un medio apropiado de recopilación de datos sobre la población económicamente activa y que la planeación de estas investigaciones en los países debería ceñirse a las normas internacionales. Por consiguiente, la resolución afirma que las encuestas de hogares deberían:

1. Brindar datos de la población económicamente activa, definida por las personas en edad laboral que se han integrado al mercado de trabajo (trabajadores o personas en búsqueda de empleo).
2. Proveer estadísticas básicas de sus actividades durante el año, así como las relaciones entre el empleo, ingreso y otras características económicas y sociales.
3. Proveer datos sobre otros temas particulares para responder a las necesidades a largo plazo y de índole permanente.

En el año 2013, la OIT decidió revisar esta resolución y propuso algunos cambios en el marco de la decimonovena Conferencia Internacional de Estadísticos del Trabajo en donde se acogieron algunas modificaciones en términos de los objetivos de medición y el alcance de los sistemas nacionales de estadísticas del trabajo, el concepto de trabajo en todas sus formas, el empleo, la medición de las personas en situación de subutilización de la fuerza de trabajo, métodos de recopilación de datos, entre otras (OIT, 2013). Las Oficinas Nacionales de Estadística (ONE) de América Latina actualizan los instrumentos de medición de las encuestas de hogares para que puedan responder a los nuevos retos en términos de la estimación de los parámetros de interés del trabajo remunerado o no remunerado para mantener la comparabilidad de las estadísticas laborales entre los países, proporcionando nuevos y mejores indicadores para contribuir al análisis de la dinámica del mercado laboral para poder brindar la información que la sociedad necesita a medida que evoluciona este constructo social.

Ingresos y gastos

Es importante resaltar que los indicadores de bienestar (en términos de ingresos y gastos) también hacen parte del conjunto de parámetros que se pueden estimar desde las encuestas de hogares. Medir el ingreso a partir de las encuestas de hogares se constituye en un reto metodológico para los institutos nacionales de estadística en el mundo, y particularmente en América Latina. Es recomendable seguir las directrices de la Comisión Económica para Europa que revisten una actualización de los estándares internacionales, recomendaciones y buenas prácticas en la medición del ingreso en los hogares. Por ejemplo, el llamado Grupo de Canberra ha revisado exhaustivamente el tópico de la estimación del ingreso estudiando las prácticas de algunos países en términos del aseguramiento de la calidad y la publicación de este tipo de estadísticas oficiales y ha provisto la siguiente definición de ingreso en el hogar (ONU, 2011):

> *El ingreso del hogar se compone de las entradas monetarias, en especie o en servicios que por lo general son frecuentes y regulares, están destinadas al hogar o a los miembros del hogar por separado y se reciben a intervalos anuales o con mayor frecuencia. Durante el período de referencia en el que se reciben, tales entradas están potencialmente disponibles para el consumo efectivo y, habitualmente, no reducen el patrimonio neto del hogar.*

Con base en lo anterior, el uso de las encuestas de hogares para estimar el ingreso reviste retos metodológicos mayores puesto que los entrevistados deben responder con precisión cuando se les indague por este constructo que contiene los ingresos personales de cada individuo en el hogar, como sueldos y salarios, ganancias, ingresos del empleo, pensiones, etc. y también los ingresos del hogar, incluidas las rentas por alquiler y los ingresos generados por el comercio. Por lo tanto, el diseño de la encuesta debe tener en cuenta la definición de un instrumento que sea relevante para el respondiente y le permita identificar y, en algunas ocasiones, recordar la información con un cierto grado de exactitud.

Por ejemplo, si el respondiente es empleado regular, el instrumento de medición debería planearse de tal manera que el entrevistado pueda recordar la información de interés, como los rubros de seguridad social hechos por su empleador. Por otro lado, si se requiere que el respondiente brinde información acerca de un determinado periodo de tiempo, el planteamiento de la pregunta, la forma de indagar y el entrenamiento de los encuestadores pueden sesgar sistemáticamente la respuesta y por consiguiente inducir estimaciones poco confiables. Mucho se ha investigado al respecto de cómo realizar preguntas certeras en este tipo de levantamientos y el lector interesado puede consultar los trabajos de Biemer and Lyberg (2003), Presser et al. (2004), y Groves et al. (2009).

Esquema del documento

Este documento pretende revisar algunas de las metodologías más usadas por las ONE de América Latina en cuanto al diseño y análisis estadístico de las encuestas de hogares y puede servir de guía técnica a los estadísticos de la región que se encuentran involucrados en los procesos técnicos de este tipo de encuestas. De la misma forma, este documento considera conjuntamente los dos principales momentos de las encuestas: el diseño y el análisis. Nótese que estos momentos están escindidos por el levantamiento de la información en campo y parten la realización de la encuesta en dos. Los lectores que están familiarizados con la investigación social a través de las encuestas de hogares encontrarán que estas operaciones estadísticas se planean teniendo en cuenta muchos pormenores que podrían suceder en campo. Es por esto que el trabajo de las encuestas asciende cuando se logra plasmar la información recolectada en una de base de datos. En este segundo momento es cuando se debe asegurar que lo que se planificó efectivamente sea incorporado en el análisis de esta información.

Desde esta perspectiva, este documento se aborda en tres partes sustantivas que definen la planeación y el análisis de la mayoría de las encuestas de hogares en América Latina y el Caribe. La primera parte se refiere a la planeación de una encuesta y a la definición del diseño estadístico, que comprende - entre otras cosas - la generación de una medida de probabilidad discreta que soportará la inferencia basada en el principio de representatividad. En esta parte se aborda con más detalle los elementos básicos que se consideran por lo regular en los diseños de las encuestas de hogares. Un aspecto relevante es que, si bien este documento considera que las encuestas de hogares tienen muchos elementos en común, diferencia de forma cuidadosa las particularidades de cada tipo de encuesta. Por ejemplo, se trata el tema del diseño de las encuestas rotativas y se profundiza en los diferentes parámetros que se pueden considerar en este tipo de operaciones; asimismo, describe las características metodológicas que se deben considerar al momento de diseñar la encuesta y revisa los conceptos esenciales que determinarán el tipo de aplicación que se debe considerar. Asimismo, se describen los principales diseños de muestreo que se utilizan en este tipo de estudios y se expone de forma estándar los conceptos de estratificación y aglomeración de las poblaciones. Estos conceptos se complementan con varias aplicaciones prácticas para determinar el tamaño de muestra adecuado para lograr los objetivos de una investigación planeada con base en las encuestas de hogares. A pesar de que la literatura relacionada con la práctica del muestreo es relativamente abundante, existen pocos ejemplos prácticos que logren representar la problemática del tamaño de muestra y el lector podrá encontrar herramientas ilustrativas basadas en múltiples escenarios de la problemática social.

La segunda parte aborda los principios metodológicos para el correcto procesamiento de las encuestas transversales, analizadas para representar un momento específico en el tiempo. Se revisan con detenimiento los procesos ponderación en la encuesta y generación de los factores de expansión que se aplicarán a la información contenida en la base de datos para que se poder generar exitósamente inferencias a nivel nacional o regional. Si hay algo que distingue el análisis de las encuestas de cualquier otro tipo de estudio estadístico es que las propiedades importantes como insesgamiento, consistencia y eficiencia están basadas en el diseño de muestreo y no en supuestos metodológicos ligados a algún modelo estocástico. Además de analizar las principales metodologías de estimación, se presta especial atención a la estimación del error de muestreo, que no es otra cosa que una función de la varianza de las estimaciones, y se presentan las metodologías más comunes en términos de aproximaciones teóricas y computacionales al error de muestreo. Los procesos de imputación y ausencia de respuesta también son abordados, con el objetivo de recuperar tanta información como sea posible para que el investigador pueda contar con una base de datos rectangular y completa. En aquellos casos en donde la imputación no resulta ser una técnica adecuada para completar la información faltante, es necesario realizar ajustes sistemáticos en los factores de expansión para que la muestra efectiva siga siendo una muestra representativa de toda la población. El último capítulo de esta parte muestra algunos enfoques útiles en la detección de datos atípicos, y la mitigación del impacto del error no muestral en las respuestas obtenidas.

La tercera parte del documento avanza hacia los principios básicos de procesamiento en un sistema integrado de encuestas de hogares que permite la agregación y/o combinación de diferentes oleadas de las encuestas, y así permitir la inferencia en un lapso más amplio. De esta forma, se presentan detalladamente los procesos que se surten cuando se agregan encuestas a lo largo del tiempo. Para aquellas encuestas que se definen a partir de estructuras rotativas, se presenta un enfoque metodológico que permite crear bases de datos longitudinales (tipo panel) para la estimación de flujos brutos, entre otros. Acudiendo a la perspectiva de CEPAL, también se presentan los criterios de calidad que se deberían tener en cuenta para decidir si una cifra, resultante de un proceso de estimación estadística basada en encuestas de hogares, debería ser o no publicada a la sociedad.

Por último, en los apéndices del documento se presenta una discusión acerca del uso presente de las encuestas de hogares y los retos que depara el futuro en materia de la medición de indicadores sociales a través de las encuestas de hogares. Asimismo, se contempla una revisión del software que se utiliza actualmente en los ONE para llevar a cabo esta ardua tarea de diseñar y analizar las encuestas de hogares, una revisión rápida de algunas de las encuestas de la región, así como algunas directrices que se deberían considerar al momento de documentar los procesos asociados a las encuestas de hogares.

Santiago de Chile.

Diciembre de 2021.

Andrés Gutiérrez, Ph.D.

Parte I

Diseño estadístico de las encuestas de hogares

Capítulo 2

El paradigma del error total

Todos los procesos en la encuesta deben estar planificados y probados de antemano, antes de la recolección de los datos. Por ejemplo, el cuestionario debe estar muy bien diseñado para que las respuestas de las personas describan acertadamente las características de los entrevistados. De la misma forma, el subconjunto de personas que participan en la encuesta debe ser expandido con precisión y confiabilidad a la población de interés.

En una encuesta, el interés no se centra en las características particulares de un individuo sino en las características de la población a la cual ese individuo pertenece. De esta forma, la inferencia siempre se realiza teniendo en mente agregados (indicadores) poblacionales. Las siguientes son las dos fuentes principales de error cuando se realiza una encuesta:

1. **Error de muestreo**: ocurre porque no se incluyeron a todas las personas de la población y se seleccionó una muestra.
2. **Error no muestral**: se refiere a las posibles desviaciones de las respuestas provistas por un entrevistado con respecto al verdadero atributo que se desea medir.

Por ejemplo, en una encuesta de fuerza laboral mensual, puede haber confusión en el respondiente si no se hace hincapíe en el periodo de referencia; no es lo mismo indagar por la semana pasada, que por el mes pasado y el respondiente debe ser guiado para evitar equivocaciones. Además pueden existir no respondientes en algún subgrupo de interés, o incluso el marco puede estar desactualizado. Uno de los objetivos de la planeación concienzuda de la encuesta es minimizar los errores no muestrales. Es necesario minimizar las discrepancias encontradas entre la respuesta verdadera a una pregunta y la respuesta final.

Groves et al. (2009) plantea que durante todo el siglo pasado, ha surgido una serie de teorías y principios que ofrecen un marco de referencia unificado en el diseño, implementación y evaluación de encuestas. Este marco de referencia se conoce comúnmente como el paradigma del error total de muestreo y ha encaminado la investigación moderna hacia una mejor calidad de las encuestas.

I. ¿Qué es una encuesta?

Groves et al. (2009) afirma que una encuesta es un método sistemático para obtener información de (una muestra de) entes, con el fin de construir descriptores cuantitativos de los atributos de una población más grande, de la cual los entes son miembros. Por otro lado, Wikipedia afirma

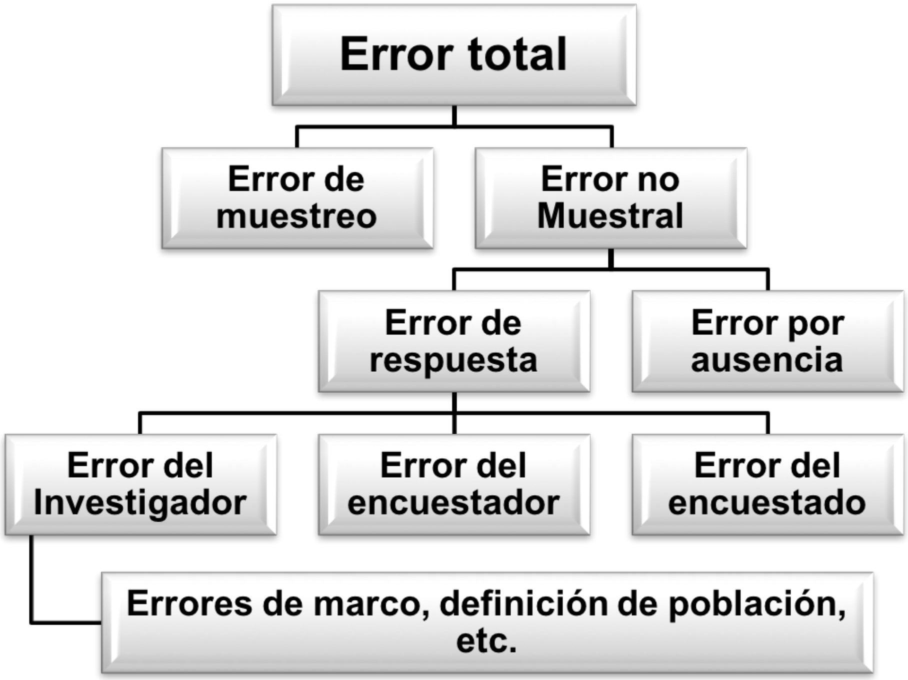

Figura 2.1: *El paradignma del error total. Fuente: adaptación de Groves et al. (2009)*

que una encuesta es un estudio observacional en el cual el investigador busca recaudar datos por medio de un cuestionario pre-diseñado, y no modificar el entorno ni controlar el proceso que está en observación (como sí lo hace en un experimento).

Además, los datos se obtienen a partir de realizar un conjunto de preguntas normalizadas dirigidas a una muestra representativa o al conjunto total de la población estadística en estudio, formada a menudo por personas, empresas o entes institucionales, con el fin de conocer estados de opinión, características o hechos específicos. Hay una diferencia sustancial entre los sondeos y las encuestas; de esta forma,

- **Sondeo**: es la traducción de *poll*, que a su vez viene del antiguo alemán referente a cabeza y se utilizaba para contar: "contar cabezas".
- **Encuesta**: es la traducción de *survey*, que a su vez viene del latín *super* (sobre) y *videre* (observar).

En general, la primera expresión aparece muchas más veces en el sector privado, en estudios de opinión y de consumo. Un sondeo no será jamás utilizado para obtener estadísticas oficiales en estudios gubernamentales o en dominios científicos. Sin embargo, los sondeos muchas veces opacan la perspectiva científica de las cifras y pueden llevar a conclusiones inexactas acerca de la realidad de una problemática. No todos los procesos de recolección de información se pueden llamar encuestas; para efectos de este documento seguiremos la definición de Groves et al. (2009), quienes afirman que una tendrá las siguientes características:

1. Los datos son recopilados mediante preguntas a personas.
2. Las respuestas son compiladas cuando: a) un encuestador pregunta y graba las respuestas del entrevistado o b) el encuestado lee y graba sus propias respuestas.
3. Los datos son recolectados de un subgrupo de personas pertenecientes a la población de

interés

II. Sesgos generados en las encuestas

Gutiérrez (2016) plantea que existen diferentes fuentes de sesgo en las encuestas y resume de la siguiente forma las dos fuentes de sesgo más importantes:

Sesgo de selección

Este tipo de sesgo ocurre cuando parte de la población objetivo no está en el marco de muestreo, o cuando el marco está incompleto y presenta deficiencias. Una muestra a conveniencia[1] es sesgada pues las unidades más fáciles de elegir o las que más probablemente respondan a la encuesta no son representativas de las unidades más difíciles de elegir. Lohr (2000) afirma que se presenta este tipo de sesgo si:

- La selección de la muestra depende de cierta característica asociada a las propiedades de interés. Por ejemplo: si la encuesta se realiza ingresando a un portal web, y precisamente las personas que no tienen cobertura de internet difieren significativamente de quienes sí tienen acceso.
- La muestra se realiza mediante elección deliberada o mediante un juicio subjetivo. Por ejemplo, si el parámetro de interés es la cantidad promedio de gastos en compras en un centro comercial y el encuestador elige a las personas que salen con muchos paquetes, entonces la información estaría sesgada puesto que no está reflejando el comportamiento promedio de las compras.
- Existen errores en la especificación de la población objetivo. Por ejemplo, en encuestas electorales, cuando la población objetivo contiene a personas que no están registradas como votantes ante la organización electoral de su país.
- Existe sustitución deliberada de unidades no disponibles en la muestra. Si, por alguna razón, no fue posible obtener la medición y consecuente observación de la característica de interés para algún individuo en la población, la sustitución de este elemento debe hacerse bajo estrictos procedimientos estadísticos y no debe ser subjetiva en ningún modo.
- Existe ausencia de respuesta. Este fenómeno puede causar distorsión de los resultados cuando los que no responden a la encuesta difieren críticamente de los que si respondieron.
- La muestra está compuesta por respondientes voluntarios. Los foros radiales, las encuestas de televisión y los estudios de portales de internet no proporcionan información confiable.

Sesgo de medición

Este tipo de sesgo ocurre cuando el instrumento con el que se realiza la medición tiene una tendencia a diferir del valor verdadero que se desea averiguar. Este sesgo debe ser considerado y minimizado en la etapa de diseño de la encuesta. Nótese que ningún análisis estadístico puede revelar que una pesa añadió a cada persona 2Kg de más en un estudio de salud. Lohr (2000) cita algunas situaciones en donde se presenta este sesgo de medición:

[1] A pesar de que las muestras por conveniencia o por juicio no pueden ser utilizadas para estimar parámetros de la población, éstas sí pueden proporcionar información valiosa en las primeras etapas de una investigación o cuando no es necesario generalizar los resultados a la población.

- Cuando el respondiente miente. Esta situación se presenta a menudo en encuestas que preguntan acerca del ingreso salarial, alcoholismo y drogadicción, nivel socioeconómico e incluso edad.
- Difícil comprensión de las preguntas. Por ejemplo: ¿No cree que este no es un buen momento para invertir? La doble negación en la pregunta es muy confusa para el respondiente.
- Las personas tienden a olvidar. Es bien sabido que las malas experiencias suelen ser olvidadas; esta situación debe acotarse si se está trabajando en una encuesta de criminalidad.
- Distintas respuestas a distintos entrevistadores. En algunas regiones es muy probable que la raza, edad o género del encuestador afecte directamente la respuesta del entrevistado.
- Leer mal las preguntas o polemizar con el respondiente. El encuestador puede influir notablemente en las respuestas. Por lo anterior, es muy importante que el proceso de entrenamiento del entrevistador sea riguroso y completo.
- La muestra está compuesta por respondientes voluntarios. Los foros radiales, las encuestas de televisión y los estudios de portales de internet no proporcionan, en general, información confiable. En este caso también se presenta sesgo de selección.

III. Evolución de las encuestas estandarizadas

Cuando el mundo occidental superó los grandes traumatismos del siglo XX (dos guerras mundiales y una recesión a larga escala), la investigación social tuvo un auge sobresaliente a través de las encuestas por correo postal. Desde entonces, existen tres preguntas, en continua dinámica, que se deben responder para planificar, ejecutar y analizar una encuesta: ¿cómo se diseñarán las preguntas? ¿cómo se seleccionará la muestra? y ¿cómo se recolectarán las respuestas?

Inicio de los cuestionarios estandarizados

La práctica de realizar las mismas preguntas en forma de cuestionario es reciente. En el principio cada encuestador preguntaba lo mismo, pero con diferentes palabras. Difícilmente, dos personas distintas eran entrevistadas con las mismas preguntas. Se encontró que la forma en cómo se preguntaba y cómo se recopilaba la información afectaba dramáticamente los resultados de las encuestas. Fue así como se decidió que los encuestadores deberías ser entrenados (pre-operativo) formalmente.

Desde la psicometría se implementó el formalismo del cuestionario. Intentando medir estados psicológicos, afectivos e intelectuales, se desarrollaron técnicas primitivas para hacer comparables las respuestas. Likert (1932) demostró que era posible realizar este tipo de comparaciones, evadiendo los largos instrumentos de medición, al formular una sola pregunta - a todos los encuestados - con una serie de respuestas en forma de escala.

Inicio de los métodos de muestreo

En un principio, los investigadores trataban de recolectar datos sobre todos los elementos de la población de interés. Esta práctica resultaba logísticamente inadecuada cuando se trataba de poblaciones con un gran tamaño. Los cálculos de los indicadores sobre toda una población resultaban muy demandantes. Groves et al. (2009) afirman que, aunque la teoría de la probabilidad tuvo sus orígenes en el siglo XVIII, no fue hasta la segunda década del siglo XX que se utilizó para realizar encuestas. La primera aplicación fue la selección sistemática de un elemento en una población enlistada. Para realizar esta selección, los registros censales se dividían en secciones y se procedía a seleccionar un elemento de la sección.

Más adelante, cuando la estadística permeó la agricultura, se definieron otros tipos de muestreo (menos demandantes) y se dio origen al muestreo de áreas. Es así como hoy en día es posible seleccionar muestras de bloques, zonas amanzanadas, secciones y sectores cartográficos, o áreas de empadronamiento censal. Se descubrió que era posible generalizar el muestreo de áreas y se creó el muestreo multietápico que permitió la selección de grandes bloques dentro de una ciudad, y áreas dentro de los bloques y el submuestreo sucesivo de unidades dentro hasta llegar a la unidad de interés. Todos estos submuestreos se realizan de forma probabilística.

La segunda guerra mundial y la gran depresión en EE.UU. fueron catalizadores de las encuestas a gran escala. En ese entonces, al igual que hoy, la tasa de desempleo era una cifra importante. Las políticas públicas empezaron a decidirse de acuerdo con las estadísticas oficiales, puesto que las grandes encuestas se realizaron mensualmente. Hoy en día existen cientos de encuestas mensuales que dan cuenta de la realidad de las sociedades en la región.

Inicio de la recolección de datos

Debido a que en un principio no existía un cuestionario estandarizado, entonces las respuestas abiertas eran la única opción de recopilar información. Esta práctica demandaba un gran esfuerzo en términos de resumir y sintetizar todo el corpus de palabras que los entrevistados usaban para responder.

En la mitad de la década del sesenta del siglo pasado, empezó una proliferación masiva de las entrevistas por correo en EE. UU. Los países con registros administrativos actualizados pueden contemplar este escenario puesto que induce altas tasas de cobertura a precios más económicos (pues se prescinde del encuestador). Las bajas tasas de respuestas (pues el encuestado debe llenar un formulario con sus respuestas y devolverlo a la oficina postal) hicieron que paulatinamente esta forma de recolección no fuese tan apetecida (Groves et al., 2009).

Un camino intermedio entre las entrevistas cara a cara y las formularios auto-administrados por correo postal son las entrevistas telefónicas. Hoy en día, la mayoría de encuestas en investigación de medios y de mercado se realiza por teléfono.

IV. El ciclo de vida de una encuesta

Atendiendo al modelo de Groves et al. (2009), se puede afirmar que en todas la encuestas se tienen dos niveles de inferencia: el individual y el grupa. El proceso de inferencia individual trata con los mismos respondientes que proveen la información primaria en el estudio; mientras que el el proceso de inferencia grupal, basado en una aproximación inductiva, va desde lo particular (la muestra) a lo general (la población).

A. Inferencia individual

Constructo

Gutiérrez (2016) menciona que los constructos son las ideas abstractas (ambiguas) sobre las cuales el investigador desea inferir y que, a su vez, dan origen a la investigación al ser la simiente de la

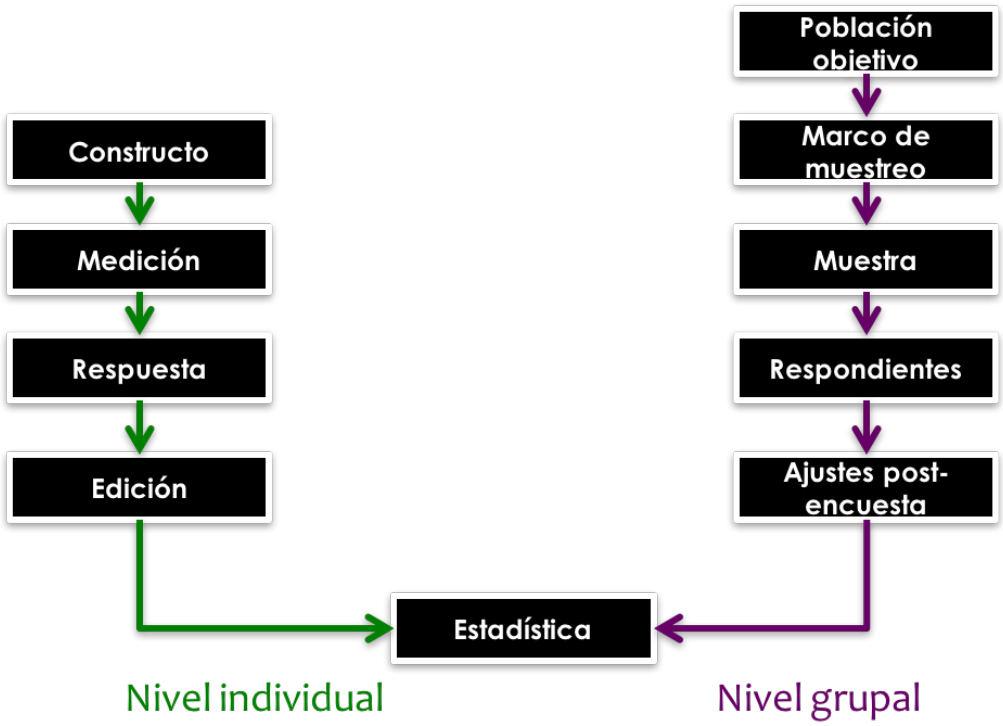

Figura 2.2: *Los niveles de inferencia en una encuestas. Fuente: adaptación de Groves et al. (2009)*

encuesta. Las palabras con que se describen los constructos son siempre simples, pero la redacción elaborada de los constructos no siempre es precisa. Por ejemplo:

- En una *encuesta de victimización* que mida la cantidad de incidentes relacionados con crímenes en un año determinado, es necesario definir muy apropiadamante qué se entiende por crimen, o cómo se define a una víctima, entre otros muchos aspectos.
- En una *encuesta de goce efectivo de derechos ciuidadanos* sobre menores de edad se puede medir la efectividad del estado al garantizar los derechos básicos a la primera infancia. Sin emabargo, es necesario definir qué es un derecho, o cómo se define primera infancia.

Mientras que algunos constructos son más abstractos que otros (optimismo en la economía, confianza inversionista, percepción del Plan Nacional de Desarrollo), algunos otros son observables más fácilmente (consumo de alcohol y otras drogas, nutrición en la primera infancia, productividad de una intervención en el sector agrícola, factores de riesgo asociados a una enfermedad).

Mediciones

La medición es una caracterización mucho más concreta que el constructo, puesto que representa una forma de obtener información de los constructos de interés. La cuestión clave para realizar una buena medición es realizar preguntas que induzcan respuestas que reflejen claramente los constructos que se desean medir. Groves et al. (2009) indican indican que estas preguntas pueden ser comunicadas en forma oral (encuestas cara a cara o telefónicas), o comunicadas en forma visual (atributos de un producto - marketing). Así mismo, también pueden existir observaciones directas del encuestador (condiciones de la vivienda), u observaciones provenientes de dispositivos electrónicos o físicos (precios de productos en supermercados, muestra de agua, muestra de sangre, etc.).

Respuesta y edición

El resultado de la medición es la respuesta y la naturaleza de las respuesta está determinada por la naturaleza de las preguntas. Después de que los entrevistados han respondido, los datos deben pasar por un proceso de edición y validación de inconsistencias.

En este proceso de edición se debe examinar la distribución completa de las respuestas y buscar *datos atípicos* para que sean revisados con detenimiento. Los datos editados constituyen el insumo para realizar todo el proceso de inferencia estadística pertinente para que las cifras resultantes sean confiables y precisas.

B. Inferencia grupal

La población objetivo

De las definiciones concernientes a agregados, esta es la más abstracta. En general, la *población objetivo* representa el conjunto de unidades que serán estudiadas. Por ejemplo, en una encuesta es posible definir la población objetivo como los adultos nacionales. Sin embargo, esta definición de población no contempla el periodo de referencia de la medición, tampoco aclara si se incluyen los adultos residentes en el exterior y, no precisa cómo se verificará la nacionalidad de un entrevistado.

Por ende, la definición de la población objetiva tiene que ser lo más precisa posible. Por ejemplo, la Gran Encuesta Integrada de Hogareas de Colombia define a su población objetivo como la Población civil no institucionalizada (PCNI), la cual contiene a todas las personas que no hacen parte de la fuerza pública y no pertenecen a instituciones de aislamiento como prisiones, hospitales, sanatorios, ancianatos, etc. La PCNI contiene a la población en edad de trabajar (PET) y a los no pertenecientes a la fuerza laboral. La edad para empezar a trabajar en el área rural es 10 años, y en la ciudad es 12 años. A su vez, la PET contiene a Inactivos y Ocupados. La clasificación de ocupado es una variable derivada que está inducida por muchos filtros.

La población enmarcada

No es posible realizar una encuesta probabilística sin un *marco de muestreo*, definido como un dispositivo que permite ubicar e identificar (ambas acciones al mismo tiempo) las unidades pertenecientes a la población de interés. Es necesario darse cuenta de que todos los marcos de muestreo presentan algún nivel de desactualización con respecto a la población de interés. Por ejemplo, un marco de muestreo de líneas telefónicas puede no contener a todos lo residentes de una ciudad.

De la misma forma, un marco de muestreo de áreas, basado en la cartografía del último ejercicio censal, puede estar desactualizado. Nótese que con un marco de áreas es posible entrevistar a la misma persona en varias ocasiones (múltiples residencias), o incluso nunca realizar la entrevista a una persona que no tienen un lugar fijo de residencia.

La población enmarcada está definida por el conjunto de miembros de la población objetivo que efectivamente tienen una probabilidad no nula de ser seleccionados en una muestra probabilística. En general para definir quién pertenece a un hogar del marco existen dos alternativas:

1. Regla *de iure*: quien habitualmente reside en el hogar es miembro de ese hogar.

- Una situación *de iure* es aquella que está reconocida por la legalidad vigente o por la autoridad competente en virtud de algún acuerdo o acto formal.
- Evita la subcobertura de individuos que no residen usualmente en su hogar, considerándolo suyo.

2. Regla *de facto*: quien pasó la noche anterior en una residencia de un hogar es miembro de ese hogar.
   - Una situación *de facto* es aquella que, existiendo en la realidad, no ha sido reconocida formalmente.
   - Evita la sobrecobertura de individuos que tienen más de una residencia.

La muestra

El tamaño de muestra define directamente la precisión y confiabilidad de las estimaciones. Este debería incrementarse a medida que lo hagan los niveles de desagregación (grupos etarios, regiones geográficas, niveles de escolaridad, etc.). Sin embargo, dependiendo de la caracterización de la estrategia de muestreo, pueden existir escenarios en donde una encuesta con un tamaño de muestra menor induzca menores errores de muestreo que una encuesta con un mayor tamaño de muestra.

No obstante, en algunas ocasiones los esfuerzos realizados para que los individuos seleccionados en la muestra respondan no son fructíferos. De esta manera, los individuos que son efectivamente entrevistados se denominan respondientes efectivos; mientras que al complemento de este conjunto se les denomina no respondientes.

Los respondientes

Pueden existir casos de no respondientes parciales (no respondientes de ítems), para los cuales debe existir un proceso de *decisión* en términos de su reemplazo. Asimismo, no todas las ausencias parciales son reemplazadas. Groves et al. (2009) afirman que algunos de los factores que inciden en el aumento de la ausencia de respuesta pueden ser causados por:

- *Contenido*: por preguntas sensibles (encuestas relacionadas con el uso de drogas, finanzas, victimización). En este caso, se puede acotar la tasa de respuesta si se ordenan las preguntas de manera adecuada.
- *Encuestadores*: aplicar métodos estándar de mejoramiento de la calidad para aumentar la precisión y tasa de respuesta de los entrevistadores involucrados en el estudio.
- *Método de recolección*: las encuestas telefónicas y por correo tienen una tasa de respuesta menor que las entrevistas personales.
- *Diseño de cuestionario*: mala planificación en el pase de las preguntas que conforman el instrumento.
- *Tiempo de la encuesta y agobio*: algunas temporadas arrojan tasas de no respuestas más altas que otras. De la misma forma, algunos cuestionarios largos son propensos a inducir una mayor ausencia de respuesta parcial por el agotamiento del respondiente. En general, las encuestas demasiado largas pueden indisponer al respondiente.

Los ajustes post-encuesta

Toda encuesta cuenta con personas que no quisieron responder y/o con un marco de muestreo que no cubre a toda la población. Por ende, es necesario reajustar los factores de expansión para evitar, sobretodo, la sub-estimación de los parámetros de interés, o implementar métodos de imputación para suplir la información faltante. De esta forma se puede utilizar una reponderación diferencial cuando es evidente que hay un patrón de ausencia de respuesta en algunos subgrupos de la población; por ejemplo: si los desempleados no responden sistemáticamente, o si las tasas de respuestas a nivel urbano son menores que las tasas de respuesta a nivel rural.

También es posible imputar (cuya raíz inglesa es *input*, traducido como introducir valores) los valores perdidos en un subconjunto de observaciones de la muestra seleccionada. En este caso es factible utilizar metodologías estocásticas complejas para imputar valores, o técnicas simples sistemáticas. Sin embargo, en cualquier caso, siempre es preferible obtener la respuesta directa del entrevistado.

V. El proceso de respuesta

No todas las encuestas se planean de tal forma que exista una interacción directa entre respondiente y entrevistador en todo tiempo. Por ejemplo:

1. *La comprensión*, en donde el respondiente interpreta la pregunta. Groves et al. (2009) afirman que en este momento se involucran todos aquellos procesos de atención a la pregunta y entendimiento de las instrucciones. La primera tarea del respondiente es interpretar la pregunta y, al hacerlo surgen procesos de análisis y asignación de un significado a los elementos sustantivos de la pregunta. Además el respondiente debe hacer una inferencia sobre el propósito de la pregunta, determinar los límites de la respuesta, así como acotar los posibles traslapes sobre las respuestas permitidas.
2. *El recaudo*, en donde el respondiente recolecta la información necesaria para brindar una respuesta. En algunas ocasiones se accede a la memoria de largo plazo que almacena todo el contenido autobiográfico y el conocimiento general. Nótese que muchas cosas pueden afectar el desempeño de la memoria de largo plazo (cuando los eventos en cuestión no se distinguen con facilidad o cuando los eventos no tienen un gran impacto personal). Aunque la memoria de largo plazo no provea la información exacta, sí provee la información relevante para que el entrevistado proporcione una respuesta adecuada. Este ciclo de recaudo de información continúa hasta que el entrevistado dé una respuesta acertada o simplemente no quiera recordar más (algunas situaciones son más difíciles de recordar) (Groves et al., 2009). Para ayudar a la memoria de largo plazo se pueden diseñar señales o pistas auto-contenidas en la pregunta. Las mejores señales son las que ofrecen un nivel de detalle más profundo.
3. *El juicio*, momento en donde se combina, se pondera y se resume la información recolectada. En esta etapa se surten procesos que complementan los recaudos que el entrevistado ha contemplado anteriormente. El juicio puede llenar los vacíos de la memoria, combinar los recaudos o ajustarlos por omisión. Por ejemplo, en una encuesta de ingresos y gastos, las personas, por lo general, no llevan la cuenta del número de veces que compraron cierto artículo o no tienen una respuesta predefinida al número de veces que han salido de compras. Por ende, el respondiente tratará de contar el número de veces que experimentó una situación, y si ese número es muy grande, seguramente se acercará a la respuesta mediante una estimación. La estrategia de estimación del respondiente (llevar la cuenta, construir una escala mediante la recordación de eventos, realizar una estimación gruesa o adivinar al azar) depende del número de sucesos, su duración, la regularidad de los mismos y el periodo de referencia de la encuesta (Groves et al., 2009).
4. *El reporte*, que es el momento en donde el respondiente formula su respuesta y la estandariza

en el formato inducido por el cuestionario. Este es el proceso de selección y comunicación de una respuesta, que incluye el encuadre de la respuesta dentro de las opciones que provee la pregunta (también implica alterar la respuesta para que se ajuste a las opciones aceptables). La forma en que se reporta la respuesta final dependerá del ajuste que se realice en los procesos de recaudo y estimación y las restricciones que la pregunta impone. En este sentido, si para una pregunta de percepción la mayoría de opciones de respuesta son negativas, la respuesta estará sesgada en esa dirección. Asimismo, los respondientes pueden dar mayor importancia a ciertas opciones de respuesta (Groves et al., 2009).

El investigador debe saber que el solo hecho de haber experimentado una situación, no implica que el respondiente haya compilado la suficiente información para reportarla como respuesta. Groves et al. (2009) afirman que se ha visto que los testigos presenciales de una situación omiten detalles importantes acerca de la situación de la cual son testigos. Además, las personas no pueden proveer la información que no tienen. Si la gente no compila la información necesaria, ninguna pregunta ni formulación logrará obtener la respuesta real. Por lo que se recomienda llevar a cabo un pre-test para validar el cuestionario. Por otro lado, aunque el respondiente conozca con exactitud la respuesta a una pregunta, no será capaz de reportarla correctamente si no hay una buena interpretación de la misma.

- *Tiempo de ocurrencia*: los eventos que sucedieron hace mucho tiempo son más difíciles de recordar.
- *Límites temporales e impacto emocional*: los eventos cercanos a límites temporales que generan impacto emocional son más fáciles de recordar. Por ejemplo, eventos catastróficos, atentados terroristas o desastres naturales.
- *Señales en las preguntas*: la asignación de múltiples señales en la redacción de la pregunta ayuda a activar el proceso de recordación.

Las preguntas cerradas con escala ordenada tienden a producir un sesgo de respuesta positivo, pues los respondientes tienden a evadir las opciones negativas de la escala (encuestas de satisfacción). Schwarz et al. (1991) demostró que las etiquetas numéricas afectan el proceso de respuesta, por lo cual recomendó que el encuestador no lea los números en las opciones de respuesta, así como acotar el número de opciones en preguntas de opinión (no muy pocas, no tantas).

Nótese que la generación de pocas opciones hace que se pierda el poder de discriminación en la respuesta, mientras que utilizar muchas opciones puede hacer que los encuestados no distingan fácilmente entre las categorías adyacentes. Además, es posible que el respondiente no quiera esperar a que el entrevistador lea exhaustivamente todas las opciones de respuesta. En este caso se presentan dos fenómeno que es necesario evadir. En primer lugar el efecto de primacía, el cual incrementa el riesgo de que el respondiente escoja una de las primeras opciones; y el efectos de recencia, en donde el respondiente siempre escogerá una de las últimas opciones.

Algunos respondientes podrán desviarse del modelo de respuesta mediante la escogencia de rutas alternas de evasión (el encuestado hará el mínimo esfuerzo para satisfacer las demandas del entrevistador). Es así como podríamos encontrar respondientes que seleccionan sistemáticamente las opciones *No sabe* o *No responde*, o que escogen siempre la misma opción para cada pregunta. Inclusive, dependiendo de la apariencia del entrevistador, el respondiente puede estar sesgado a siempre estar de acuerdo (aquiescencia). De la misma manera, es posible que el respondiente quiera presentarse a sí mismo de manera favorable, omitiendo sus atributos no deseables (Groves et al., 2009).

Capítulo 3

Elementos básicos

El fortalecimiento continuo de las investigaciones sociales es un objetivo que los institutos nacionales de estadística procuran cumplir de forma sistemática. En el caso de aquellas operaciones que conllevan la recolección de información primaria y que involucran la selección y medición de hogares y sus miembros, mantener una documentación adecuada que describa las razones por las cuales se ha optado por cierta metodología de recolección en particular es un requisito fundamental para cumplir este cometido. En este apartado se exploran diferentes métodos de recolección de la información y se discuten las diferentes particularidades en la planeación de una encuesta de hogares.

I. Universo, muestra y unidades

El término encuesta se encuentra directamente relacionado con una población finita compuesta de individuos a los cuales es necesario observar y medir. Este proceso muchas veces es realizado por medio de una entrevista. El conjunto de unidades de interés recibe el nombre de *población objetivo* o *universo* y sobre ellas se obtiene la información de interés para el estudio. Por ejemplo, *la Encuesta Nacional de Empleo y Desempleo* de Ecuador define su población objetivo como todas las personas mayores de 10 años residentes en viviendas particulares en Ecuador (INEC, 2018).

Las *unidades de análisis* corresponden a los diferentes niveles de desagregación establecidos para consolidar el diseño de la encuesta y sobre los que se presentan los resultados de interés. En México, la *Encuesta Nacional de Ingresos y Gastos de los Hogares* define como unidades de análisis el ámbito al que pertenece la vivienda: urbano alto, complemento urbano y rural. Por otro lado, la *Gran Encuesta Integrada de Hogares* de Colombia tiene cobertura nacional y sus unidades de análisis están definidas por trece grandes ciudades junto con sus áreas metropolitanas (DANE, 2017).

Como se explicará más adelante, es muy difícil contar con una lista actualizada de todos los hogares del país; por lo tanto, para recolectar la información de la población objetivo, el diseño de una encuesta de hogares en América Latina plantea la necesidad de seleccionar en varias etapas ciertas *unidades de muestreo* que sirven como medio para seleccionar finalmente a los hogares y personas que participarán de la muestra. Cuando se requiere seleccionar personas, se hace necesario seleccionar un subconjunto de zonas geográficas; para cada zona seleccionada, se procede a seleccionar a su vez un subconjunto de secciones cartográficas, que antecede a la selección de hogares. Finalmente, el cuestionario es administrado en cada hogar a un respondiente idóneo, que proporciona la información de todos los integrantes del hogar. Dependiendo de la encuesta, en algunos casos se seleccionan aleatoriamente respondientes individuales dentro del hogar; siendo estas las unidades de observación. Por ejemplo, se puede citar la experiencia de Brasil con la *Pesquisa Nacional por Amostra de Domicilios* que se realiza por medio de una muestra de viviendas en tres etapas: las unidades primarias de muestreo (UPM) son los municipios, mientras que las unidades secundarias de muestreo (USM) son los sectores censales, que conforman una malla territorial definida en el último Censo Demográfico. Por último, las unidades finales en ser seleccionadas son las viviendas (IBGE-BR, 2014).

Duncan and Kalton (1987, pág. 105) afirman que la composición de la población de interés en las encuestas de hogares cambia durante el tiempo, puesto que lo individuos nacen, mueren, migran, e incluso pasan a ser parte de organizaciones que hacen que pierdan su estatus de elegibilidad como unidades de observación en una encuesta. Nótese que la población objetivo de la mayoría de encuestas de hogares en América Latina se refiere a la población civil excluyendo a los miembros de organizaciones militares, personas recluidas en cárceles, personas que se encuentran en hospitales, etc. De igual forma, se debe tener en cuenta que los hogares pueden crearse o desintegrarse rápidamente. Por ende, los equipos técnicos de las ONE que están a cargo del diseño de las encuestas de hogares, que miden de forma transversal a la población de interés, deben tener en cuenta que, aunque los objetivos de la encuesta no cambian en el tiempo, sí lo hace la población objetivo y se deben plantear esquemas de seguimiento y actualización que den cuenta de esta realidad.

II. Periodicidad en el tiempo

Las Oficinas Nacionales de Estadística - que son los entes encargados de administrar, diseñar, analizar y difundir los resultados de las encuestas - no realizan este tipo de levantamientos de manera aislada; de hecho una característica fundamental de estas operaciones estadísticas es que se han convertido en un insumo fundamental para realizar un seguimiento periódico de muchos indicadores de interés. Por lo tanto, muchas encuestas de hogares se realizan de forma sistemática en el tiempo, aunque algunas otras no tienen una periodicidad predefinida. Es por esto que la planeación de la encuesta debe contemplar este tipo de esquemas continuos para que el levantamiento de la información primaria en campo se haga de manera más eficiente y, de la misma forma, que la estimación de los indicadores de interés se pueda realizar ajustándose a los recursos de la operación. Como se mencionó anteriormente, dado que la población es dinámica en el tiempo, la planeación y análisis de este tipo de encuestas es desafiante, puesto que si la composición de la población y las características de los elementos se considerara fija, una encuesta transversal (realizada una sola vez en un periodo de tiempo largo) sería suficiente para realizar estimaciones precisas que resuelvan los objetivos del estudio.

En algunas ocasiones, basta con realizar un medición simple en un punto específico del tiempo para completar los objetivos de la investigación. Este es el caso de las encuestas de ingresos y gastos cuya periodicidad es, en general, no menor a cinco años y las cuales son utilizadas para, entre muchos otros propósitos, actualizar la canasta básica familiar, de la cual se derivan los insumos básicos para la medición de la pobreza (CEPAL, 2018). Para otro tipo de problemáticas, como por ejemplo el seguimiento a las estadísticas derivadas del mercado de trabajo, es necesario recurrir a la medición periódica a través de encuestas de hogares, en donde los cambios naturales en las características de la población hacen que realizar una medición simple en un punto del tiempo sea insuficiente a la luz del seguimiento y monitoreo de los indicadores de interés.

Por consiguiente, al momento de realizar la planeación de una encuesta continua o periódica se debe tener en cuenta que, a pesar de que crezca la dificultad en el diseño, es posible obtener información más oportuna para la toma de decisiones y la formulación de políticas públicas. De esta manera, y teniendo en cuenta que el tiempo hace que la estructura de las poblaciones cambie, sin importar si la constituyen individuos, hogares, familias, negocios, etc., las unidades de observación deben ser consideradas como parte de la población de interés cuando nacen, inmigran o alcanzan un umbral predefinido de edad. Asimismo, las unidades ya no harán parte de la población de interés cuando mueran, emigren, o se involucren en instituciones (como el servicio militar). Por ejemplo, si las unidades de interés son los hogares, es evidente que la población no es la misma en diferentes puntos del tiempo (por ejemplo, en dos años distintos) puesto que se crean nuevas unidades cuando los jóvenes dejan a sus padres y forman nuevos hogares independientes, o cuando ocurre una separación o un divorcio; en donde un hogar se divide en dos. Además, los hogares en donde todos sus miembros han fallecido dejan de ser parte de la población objetivo. De la misma forma, dos hogares dejan de ser parte de la población objetivo cuando se unen a través de un matrimonio o algún otro tipo de unión civil. Teniendo en cuenta el papel dinámico de las poblaciones y los objetivos de investigación es posible plantear diferentes tipos de levantamientos; a continuación enumeramos algunas categorías de encuestas que las ONE realizan en la región.

Encuestas transversales

Este tipo de encuestas son diseñadas para recolectar información únicamente en un punto específico del tiempo, o sobre un periodo de referencia, y proveen toda la información pertinente acerca de la población particular restringida a un tiempo y periodo de recolección específico. Puesto que el propósito fundamental de este tipo de encuestas no se centra en las comparaciones intertemporales, no es posible estimar cambios de ningún tipo, a no ser que se realicen indagaciones retrospectivas. La siguiente tabla muestra un esquema de este tipo de operaciones estadísticas en donde se observa una muestra de una población específica en un periodo de tiempo específico (Tiempo 2). Dado que es una muestra transversal, no hay un patrón de repetición en los restantes periodos.

Cuadro 3.1: *Esquema de una encuesta transversal.*

Hogar	Tiempo 1	Tiempo 2	Tiempo 3	Tiempo 4	...	Tiempo T
1		**x**				
2		**x**				
3		**x**				
4		**x**				
...		**x**				
n		**x**				

Encuestas repetidas

Cuando existe interés en realizar un seguimiento del fenómeno en observación durante el tiempo, se utilizan encuestas repetidas que recolectan información de manera periódica. Este tipo de encuestas proveen información acerca de la dinámica de la composición de la población en el tiempo. De esta forma, en cada levantamiento se observa una muestra de la población en un tiempo determinado. Por ejemplo, la siguiente tabla muestra un acercamiento gráfico a este tipo de encuestas en donde se evidencia el carácter sistemático de estas operaciones estadísticas; además de mostrar que no es posible medir cambios individuales porque las muestras son independientes en el tiempo.

Cuadro 3.2: *Esquema de una encuesta repetida.*

Hogar	Tiempo 1	Tiempo 2	Tiempo 3	Tiempo 4	...	Tiempo T
1	**x**					
2		**x**				
3			**x**			
4				**x**		
...					**x**	
n						**x**

Encuestas panel

Las encuestas en panel están diseñadas para recolectar información periódica sobre la misma muestra en diferentes puntos del tiempo. Por definición, las unidades de muestreo son las mismas en los diferentes periodos de tiempo y, de manera general, se miden las mismas variables en cada levantamiento. Por la caracterización propia de este tipo de encuestas, sí es posible estimar los cambios individuales, así como los cambios netos sobre la población. Sin embargo, como la muestra no cambia en ningún momento del tiempo, las inferencias que se realicen estarán supeditadas a la población de la cual se seleccionó la muestra en un principio (Tiempo 1). Si la población cambia su estructura, no será posible captar este cambio puesto que las inferencias resultantes de este tipo de encuestas no son representativas de la población actual. La siguiente tabla muestra un esquema propio de las encuestas de panel en donde los individuos que fueron seleccionados la primera vez son observados a lo largo del tiempo.

Cuadro 3.3: *Esquema de una encuesta tipo panel.*

Hogar	Tiempo 1	Tiempo 2	Tiempo 3	Tiempo 4	...	Tiempo T
1	**x**	**x**	**x**	**x**	**x**	**x**
2	**x**	**x**	**x**	**x**	**x**	**x**
3	**x**	**x**	**x**	**x**	**x**	**x**
4						
...						
n						

Encuestas de panel dividido

Para hacerle frente a las dificultades propias de las encuestas de panel y poder observar tanto los cambios individuales, como los cambios en la estructura de la población, se definen las encuestas de panel dividido. Estas operaciones estadísticas son una combinación del diseño de panel puro y del diseño repetido y su objetivo es realizar inferencias precisas acerca de los cambios de una cohorte a través del tiempo y, al mismo tiempo, del cambio en estructura de la población actual. De esta forma, se realiza el seguimiento continuo, periódico y sistemático de una muestra a través del tiempo, pero en cada levantamiento se incluyen nuevos elementos seleccionados de la población actual. Como se señalará más adelante, este tipo de encuestas cubre con eficiencia la mayoría de indicadores de interés en un estudio de investigación social. La siguiente tabla muestra una caracterización de estos levantamientos que fijan una muestra de panel a lo largo del tiempo, y a la vez que se añaden nuevas observaciones.

Cuadro 3.4: *Esquema de una encuesta de panel dividido.*

Hogar	Tiempo 1	Tiempo 2	Tiempo 3	Tiempo 4	...	Tiempo T
1	**x**	**x**	**x**	**x**	**x**	**x**
2	**x**					
3		**x**				
4			**x**			
5				**x**		
...					**x**	
n						**x**

Encuestas de panel rotativo

Mantener una muestra de panel es un proceso costoso desde una perspectiva económica y logística, pero también se debe tener en cuenta el desgaste de la fuente, que tenderá a brindar menos información a medida que avanza el estudio. Además, es evidente que a medida que el tiempo transcurra la propensión a responder será más baja, puesto que el entrevistado se sentirá agotado al ser visitado una y otra vez. Por lo tanto, se definen las encuestas de panel rotativo para poder realizar inferencias parciales - restringidas a periodos de tiempo específicos - del cambio individual y a la vez captar el cambio estructural de la población. Estas encuestas incorporan nuevos elementos de la población y a la vez mantienen elementos comunes con mediciones anteriores. Obviando las dificultades que acarrea la ausencia de respuesta, las encuestas panel definen un traslape completo entre las muestras de dos puntos cualesquiera en el tiempo; sin embargo, en las encuestas rotativas existe un traslape parcial, por lo que se reduce el efecto del desgaste del panel (sobre la población inicial) y el efecto de la pérdida de muestra. Además, la inclusión de nuevos elementos en la muestra provee información pertinente del cambio en la composición estructural de la población. La siguiente tabla ejemplifica el diseño de las encuestas rotativas.

Cuadro 3.5: *Esquema de una encuesta de panel rotativo.*

Hogar	Tiempo 1	Tiempo 2	Tiempo 3	Tiempo 4	Tiempo 5	Tiempo 6
1	**x**					
2	**x**	**x**				
3	**x**	**x**	**x**			
4		**x**	**x**	**x**		
5			**x**	**x**	**x**	
6				**x**	**x**	**x**
...					**x**	**x**
n						**x**

III. Rotación de paneles

Tal como se describió anteriormente, algunas encuestas de hogares en América Latina permiten que un hogar sea visitado en más de una ocasión con el fin de tener estimaciones precisas acerca de los cambios de estado que el hogar o las personas que lo habitan puedan sufrir. Por ejemplo, un hogar que en un periodo estuvo en condición de pobreza extrema, puede estar en otro periodo en condición de pobreza relativa o inclusive puede pasar a estar fuera de la pobreza; en las encuestas de fuerza laboral, una persona puede pasar de estar empleada en un periodo a desempleada en otro periodo. Estos cambios y la dinámica propia que conllevan son de interés para los investigadores y deben ser contemplados desde una perspectiva más amplia en cuanto a su diseño. Nótese que este tipo de variaciones sobre los individuos necesariamente tiene que ser captada a través de un componente de panel, por lo que las encuestas transversales o repetidas no serían viables para realizar estas estimaciones.

En América Latina hay una gran variedad de encuestas de hogares que utilizan diseños rotativos (ver apéndice). Por ejemplo, la *Encuesta Permanente de Hogares* en Argentina renueva periódicamente el conjunto de hogares que serán entrevistados mediante un esquema[1] de rotación 2(2)2 que selecciona a las viviendas para ser entrevistadas en dos periodos consecutivos; luego los siguientes dos periodos esas viviendas salen de la selección, para finalmente volver a ser encuestadas en los siguiente dos periodos (INDEC, 2018). De esta forma, dado que la rotación es trimestral, un hogar es seguido a lo largo de 18 meses y esto permite cumplir con los objetivos de la encuesta. Este esquema induce algunas propiedades interesantes, que pueden ser ejemplificadas usando la siguiente tabla definido para los cuatro trimestres de los años 2016, 2017, 2018 en cuatro grupos de muestra: A, B, C y D.

- Entre el primer y el segundo periodo de medición hay un traslape del 50 % de hogares. En particular, nótese que entre 2016-T1 y 2016-T2, la muestra se conserva en un 50 %, puesto que $a1$ y $d1$ se repiten. Esto mismo sucede en cada trimestre del esquema rotacional.
- En el tercer periodo no habrá traslape con el primer periodo. Nótese que entre 2016-T1 y 2016-T3 no existe ningún elemento en común. De la misma manera, entre 2016-T2 y 2016-T4, no existe ningún elemento en común. Este mismo patrón se encuentra a lo largo del esquema rotacional.
- En el cuarto periodo se tendrá un 25 % de traslape con el primer periodo. Nótese, por ejemplo, que entre 2017-T1 y 2017-T4, $c3$ se repite; de la misma manera, entre 2017-T4 y 2018-T3, $d4$ se repite.
- Finalmente en el quinto periodo se volverá a tener un 50 % de traslape con respecto al primer periodo. Por ejemplo, 2016-T1 y 2017-T1 comparten el 50 % de la muestra $a1$ y $b1$; asimismo, 2017-T1 y 2018-T1 comparten el 50 % de la muestra $c3$ y $b3$.

Cuadro 3.6: *Rotación de paneles en un diseño 2(2)2.*

Año	Trimestre	A	B	C	D
2016	T1	*a1*	*b1*	*c1*	*d1*
	T2	*a1*	*b2*	*c2*	*d1*
	T3	*a2*	*b2*	*c2*	*d2*
	T4	*a2*	*b1*	*c3*	*d2*

[1] Un esquema de rotación $x(y)z$, se define como aquel en donde la vivienda entra al panel por x periodos, se excluye por los siguientes y periodos y este patrón se repite z veces en el tiempo. Nçotese que los periodos pueden ser definidos como meses, o trimestres; además un hogar es visitado un total de $x \times z$ veces.

Año	Trimestre	A	B	C	D
2017	T1	*a1*	*b1*	*c3*	*d3*
	T2	*a1*	*b2*	*c4*	*d3*
	T3	*a2*	*b2*	*c4*	*d4*
	T4	*a2*	*b3*	*c3*	*d4*
2018	T1	*a3*	*b3*	*c3*	*d3*
	T2	*a3*	*b4*	*c4*	*d3*
	T3	*a4*	*b4*	*c4*	*d4*
	T4	*a4*	*b3*	*c5*	*d4*

Otro ejemplo de una encuesta que utiliza rotación de paneles es la *Encuesta Continua de Empleo* de Bolivia que, aplicada por el Instituto Nacional de Estadística, hace uso de una metodología mixta que permite el seguimiento continuo y transversal a la tasa de desempleo y a la tasa de subocupación, así como el seguimiento a los cambios que se presentan entre los periodos de interés (trimestres y semestres), a través del análisis longitudinal de los datos en el sector urbano (pues el diseño no es rotativo en el sector rural, debido a la baja incidencia de desempleo en esta zona). En este esquema rotacional 4(0)1 una vivienda es entrevistada durante cuatro trimestres consecutivos, y luego sale del panel definitivamente. Un ejemplo de este tipo de esquemas se presenta en la siguiente tabla.

- Nótese que entre el primer y el segundo periodo de medición hay un traslape del 75 % de hogares. En particular, entre 2016-T1 y 2016-T2, la muestra se conserva en tres cuartas partes puesto que $a1$, $c1$ y $d1$ se repiten. Esto mismo sucede en cada trimestre del esquema rotacional.
- Por otro lado, entre el primer y el tercer periodo habrá un traslape del 50 %. Nótese que entre 2016-T1 y 2016-T3, la mitad de la muestra se conserva puesto que $a1$ y $d1$ se repiten. Este mismo patrón se encuentra a lo largo del esquema rotacional.
- Entre el primer y el cuarto periodo se tendrá un 25 % de traslape. Nótese, por ejemplo, que entre 2017-T1 y 2017-T4, $a2$ se repite; de la misma manera, entre 2017-T4 y 2018-T3, $d3$ se repite.
- Finalmente entre el primer y quinto periodo no se tiene ningún tipo de traslape.

Cuadro 3.7: *Rotación de paneles en un diseño 4(0)1.*

Año	Trimestre	A	B	C	D
2016	T1	*a1*	*b1*	*c1*	*d1*
	T2	*a1*	*b2*	*c1*	*d1*
	T3	*a1*	*b2*	*c2*	*d1*
	T4	*a1*	*b2*	*c2*	*d2*
2017	T1	*a2*	*b2*	*c2*	*d2*
	T2	*a2*	*b3*	*c2*	*d2*
	T3	*a2*	*b3*	*c3*	*d2*
	T4	*a2*	*b3*	*c3*	*d3*
2018	T1	*a3*	*b3*	*c3*	*d3*
	T2	*a3*	*b4*	*c3*	*d3*
	T3	*a3*	*b4*	*c4*	*d3*

Año	Trimestre	A	B	C	D
	T4	*a3*	*b4*	*c4*	*d4*

Los diseños de las encuestas de hogares deben tener en cuenta la rotación de los paneles y el número de veces que es visitado un hogar. Esta caracterización depende directamente de los indicadores a los cuales la encuesta debe responder. Por ejemplo, el diseño de rotación debe ser diferente si el interés se centra en indicadores de cambio trimestral, a si se requieren indicadores de cambio anual. Por ejemplo, el diseño 4(0)1 es conveniente si el objetivo está en comparar las estimaciones de la tasa de desocupación el mismo mes entre diferentes años, pero no lo será si se quiere conocer el cambio de estado en la situación del trabajo para las mismas personas en dos meses iguales de diferentes años. Nótese que un aspecto importante en la definición de los esquemas longitudinales radica en el tiempo en el que un hogar pertenecerá al panel. Por supuesto, hay que tener en cuenta que la tasa de ausencia de respuesta y pérdida de muestra por desgaste del respondiente crecerá en la medida en que se le pida a un hogar una participación más duradera en el tiempo.

La definición de los indicadores de interés debe primar sobre el diseño de las encuestas de hogares. Por ejemplo, si el objetivo de la encuesta se centra en la estimación del cambio del indicador en dos periodos de tiempo, entonces el cálculo de la precisión de las estimaciones debe tener en cuenta que las muestras no son independientes y por lo tanto se debe calcular la varianza de la primera ronda, la varianza de la segunda ronda y la correlación entre las dos rondas de interés. Estos tres componentes deben intervenir en el cálculo de los coeficientes de variación, así como en la determinación del tamaño de muestra en cada ronda. En efecto, como lo afirma McLaren and Steel (2001, pág. 236), para la estimación de tendencias, definidas a partir de series de tiempo macroeconómicas de los parámetros de interés en los estudios de fuerza laboral, el mejor patrón encontrado es el 1(2)m, en donde la vivienda entra en un primer mes en el panel, se excluye por los siguientes dos meses y este patrón se repite m veces consecutivas. A partir de allí, la vivienda ya no vuelve a ser incluida en el estudio. En resumen, por la naturaleza de las encuestas de hogares en la región, al momento de pensar en incluir o cambiar la estructura rotacional en el sistema de encuestas de hogares, se debería considerar en primer lugar el esquema de repartición mensual de paneles. Una mirada más profunda de este tipo de análisis longitudinales se encuentra presente en los capítulos posteriores a lo largo de este documento.

IV. Parámetros e indicadores de interés

Las encuestas son usadas para producir estimaciones de parámetros que describen la situación de una población, respondiendo a los objetivos de la investigación. En general, es posible clasificar en dos grandes grupos los indicadores o parámetros de interés en una encuesta:

1. Indicadores descriptivos, incluyendo:
   - Medias: promedio de años en educación.
   - Proporciones: porcentaje de personas que votarán por un candidato.
   - Totales: Total de personas víctimas del desplazamiento forzado.
2. Indicadores analíticos, incluyendo:
   - Correlación: relación entre la cantidad de libros leídos y los años de escolaridad.
   - Regresión: razón de incremento entre ingreso y años de experiencia

Por lo general, el conocimiento de la población a cualquier nivel está reflejado en forma de totales, o de funciones de totales. Es por esta razón que este documento se enfoca y profundiza en las características inferenciales de los totales, puesto que la generalización a otros parámetros es inmediata. De esta manera, un **total poblacional** se define como la suma de las observaciones de una variable de interés, notada como y, en la población y se calcula mediante la siguiente ecuación:

$$t_y = \sum_{k \in U} y_k$$

En donde U hace referencia al universo de estudio, mientras que y_k hace referencia a la variable de interés en el k-ésimo individuo. Por ejemplo, en una investigación social se puede realizar una encuesta para estimar el total de gasto de los hogares de un país en productos específicos de comida y bebidas no alcohólicas. En este ejemplo, la población U corresponde a los hogares, mientras que la variable y corresponde al gasto en comida y bebidas no alcohólicas, que es observada en el k-ésimo hogar, y notada como y_k.

Un caso particular de este parámetro es el **tamaño poblacional** que mide la cantidad de unidades que conforman una población y se denota como N. Por lo general, este parámetro es regularmente conocido, o al menos se tiene una aproximación de esta cantidad. En una encuesta de hogares, este parámetro podría denotar el número de hogares en el país (el cual no es conocido literalmente, aunque sí se conocen aproximaciones (o proyecciones) a esta cantidad con base en los resultados de los censos de población y vivienda) o el número de habitantes del país (el cual tampoco es conocido exactamente, aunque sí se cuente con proyecciones poblacionales). Este parámetro también toma la forma de un total poblacional:

$$N = \sum_{k \in U} 1$$

Tal vez el parámetro más relevante en la investigación social lo constituye el **promedio poblacional** que describe la cantidad que debería ser asignada a cada individuo de la población si hubiese una asignación equitativa de la variable de interés. De esta forma, el promedio se define como la suma de las observaciones de la variable en la población dividida por el tamaño poblacional N y se calcula mediante la siguiente expresión:

$$\bar{y}_U = \frac{t_y}{N}$$

Por ejemplo, en una encuesta de hogares es posible estimar el ingreso medio por hogar de la población, definido como el total de los ingresos de todos los hogares del país dividido entre el número de hogares del país. En este caso la variable de interés y es el ingreso del hogar. De la misma forma, también se podría estimar el gasto promedio de los hogares en educación; en donde la variable de interés y es el gasto de todos lo miembros del hogar en este concepto (sin importar la edad ni el nivel propedéutico) y N sería el número de hogares del país.

Un parámetro que es de particular interés es el **tamaño absoluto de un dominio poblacional** que mide la cantidad de unidades que conforman una subpoblación de interés U_d y que se denota como N_d. Por ejemplo, en las encuestas de fuerza laboral, es muy importante estimar con una alta precisión el número de personas que están desocupadas en un periodo de tiempo, y comparar su evolución a través del tiempo. En este caso, la subpoblación de interés, o dominio poblacional, estará definida por los desocupados. Nótese que este parámetro está definido como un total sobre una variable dicotómica z_{d_k} que toma el valor de 1, si el k-ésimo individuo tiene el atributo de interés y de 0, en otro caso. Este parámetro se calcula de la siguiente manera:

$$N_d = \sum_{k \in U} z_{d_k} = \sum_{k \in U_d} 1$$

De la misma forma, la incidencia relativa de los fenómenos sociales sobre los hogares o personas puede ser medida a través de la **proporción de un dominio poblacional**. Por ejemplo, la proporción de personas en condición de pobreza o de pobreza extrema son proporciones sobre toda la población, en donde la variable de interés z_{d_k} indica si el ingreso per cápita de un individuo es menor que la línea de pobreza; CEPAL (2018) presenta los pormenores metodológicos del cálculo de la pobreza en los países de América Latina y el Caribe. Este parámetro se calcula mediante la siguiente ecuación:

$$P_d = \frac{N_d}{N}$$

En algunos casos es de interés conocer el total de una variable en una subpoblación. Por ejemplo, el total del ingreso en las mujeres, o el total de gasto en el área rural. En estas situaciones el parametro se conoce como **total del dominio** y se puede calcular mediante la siguiente expresión:

$$t_{y_d} = \sum_{k \in U} y_k \ z_{d_k} = \sum_{k \in U_d} y_k$$

Así mismo, puede ser de interés calcular medidas relativas en el dominio, como por ejemplo la **media del dominio**. De esta forma, es posible calcular la media de los ingresos entre hombres y mujeres, o calcular la media de los ingresos en los ocupados, o la media del gasto en comida para la población indígena. Este parámetro puede ser calculado con la siguiente expresión:

$$\bar{y}_{U_d} = \frac{t_{y_d}}{N_d}$$

Finalmente, la **razón poblacional** se calcula como el cociente entre dos totales, el primer total t_y asociado a una variable de interés y, el segundo total t_x asociado a una variable de interés x. Por ejemplo, en la medición del mercado de trabajo, la tasa de desocupación es una razón entre el total de personas desocupadas y el total de personas activas. Nótese que para clasificar a una persona como desocupada, ocupada, activa o inactiva, es necesario realizar una indagación en la encuesta a cada uno de los miembros del hogar; por lo tanto ambas cantidades, numerador y denominador, corresponden a cantidades desconocidas de antemano. Es más, la condición de ocupación de las personas puede variar entre los periodos de observación. Este parámetro se calcula mediante la siguiente expresión:

$$R_U = \frac{t_y}{t_x}$$

En efecto, los indicadores de pobreza pueden expresarse como razones poblacionales; es el caso de la brecha de pobreza y de la incidencia de la pobreza expresada en términos de un umbral de poder adquisitivo (Foster et al., 1984). Este tipo de indicadores complejos se pueden expresar mediante la siguiente relación

$$F_\alpha = \frac{1}{N}\sum_U \left(\frac{u-y_k}{u}\right)^\alpha I_{(y_k<u)}$$

En donde y_k determina el ingreso del individuo k, u se refiere al umbral que establece la línea de pobreza y $\alpha \geq 0$. Por ejemplo, en el caso en el que $\alpha = 0$, este indicador calcula la tasa de pobreza, que es la incidencia de este fenómeno en la población; si $\alpha = 1$, este indicador calcula la brecha de la pobreza, que es la cantidad de dinero relativa que se necesitaría en promedio para que un país no tuviera personas en situación de pobreza. Por último si $\alpha = 2$, este indicador medirá la severidad de la pobreza, como una combinación entre la incidencia de la pobreza de los hogares, la brecha absoluta de ingreso de los hogares en situación de pobreza y la desigualdad de ingresos entre los hogares en situación de pobreza.

En este punto vale la pena resaltar que, en la definición de los parámetros básicos que se quieren estimar en una encuesta, el papel de los totales poblacionales es absolutamente relevante. De igual manera, existen otros parámetros que pueden ser considerados complejos - no por su forma funcional, sino por los procesos complejos que hay detrás del levantamiento de la información primaria - pero que al igual que los mencionados anteriormente resultan ser también una función de totales poblacionales. Por ejemplo, considere el **cambio neto** de los totales de la variable de interés y en dos periodos de tiempo (t_1 y t_2) dado por la siguiente expresión:

$$\Delta_y = t_{y^{(2)}} - t_{y^{(1)}}$$

En donde $t_{y^{(2)}}$ es el total de interes en el tiempo $t = 2$, y $t_{y^{(1)}}$ lo es en el tiempo $t = 1$. Este tipo de parámetros son muy comunes en las encuestas que se realizan para conocer la estructura y los cambios del mercado de trabajo. Por ejemplo, la siguiente tabla muestra la composición del mercado de trabajo en una población observada en dos periodos de interés. De esta forma, los totales marginales de la tabla corresponden a los **cambios netos** que permiten una comparación simple con el periodo anterior. Específicamente, es posible observar que hay 313 mil empleados menos, 80 mil desempleados menos y 393 mil inactivos más en el segundo periodo, en comparación al primero.

Cuadro 3.8: *Composición del mercado de trabajo en dos periodos de tiempo (cifras en miles de personas). Las columnas corresponden al segundo periodo y las filas al primero.*

Condición	Ocupado	Desocupado	Inactivo	**Total**
Ocupado	9222	128	662	**10012**
Desocupado	221	322	151	**694**
Inactivo	256	164	5941	**6361**
Total	**9699**	**614**	**6754**	**17067**

Una comparación más profunda está dada en términos de los **cambios brutos**, que corresponden a las entradas de la tabla cruzada. De esta manera, los cambios en la fuerza de trabajo de un periodo a otro, se explican porque el $92.1\,\% = (9222/10012) \times 100\,\%$ de los empleados conservó su empleo; el $31.8\,\% = (221/694) \times 100\,\%$ de los desempleados y el $4.0\,\% = (256/6361) \times 100\,\%$ de los inactivos consiguió un nuevo empleo; el $6.6\,\% = (662/10012) \times 100\,\%$ de los empleados es ahora inactivo en la fuerza laboral y el $1.3\,\% = (128/10012) \times 100\,\%$ de los empleados perdió su empleo. Así mismo, el $46.4\,\% = (322/694) \times 100\,\%$ de los desempleados conservó su clasificación; el $2.6\,\% = (256/6361) \times 100\,\%$ de los inactivos entró a la fuerza laboral como desempleado y el $21.8\,\% = (151/694) \times 100\,\%$ de los desempleados es ahora inactivo.

Algunos ejemplos de indicadores de interés y su relación con los tipos de encuestas

En esta sección se relacionan algunos de los parámetros anteriormente mencionados con los tipos más comunes de encuestas. Estos ejemplos nos presentan algunas indicaciones del tipo de encuestas que se encuentran en América Latina y examinan el raciocinio detrás de estos levantamientos. Tomando en consideración las características generales de las encuesta de hogares, Duncan and Kalton (1987) mencionan las siguientes situaciones, ejemplificadas a continuación.

- **Estimación de parámetros poblacionales en un punto del tiempo**. Por ejemplo, suponga que se quiere estimar el *ingreso per cápita promedio por área (rural - urbano) en las regiones de un país*. En este tipo de estudios, las encuestas aptas serían las transversales, las repetidas, las de panel rotativo y las de panel dividido. Nótese que las encuestas de panel puro no son aptas para estimar este parámetro puesto que la muestra no es representativa de la población en el momento actual, sino que, por el contrario, es representativa de la población en el momento en la cual se extrajo la muestra.

- **Estimación de cambios netos**. Si se quisiera estimar la *diferencia en el número de ocupados de la fuerza de trabajo entre el segundo trimestre de 2021 y el primer trimestre de 2021 en un país*, entonces las encuestas aptas serían las repetidas, las de panel rotativo y las de panel dividido. Una encuesta transversal no sería apta para lograr esta estimación, puesto que su frecuencia de realización no es trimestral. De la misma forma que en el parámetro anterior, las encuestas de panel puro no son aptas para captar este parámetro puesto que la muestra no es representativa de la población en el momento actual.

- **Estimación de cambios brutos y componentes individuales**. Para estimar el *porcentaje de personas ocupadas en el segundo trimestre de 2021 que estuvieron desocupadas en el primer trimestre de 2021 en un país* es necesario que la encuesta tenga algún patrón de selección de los mismos individuos en los dos periodos. De esta forma, las únicas encuestas aptas para estimar este tipo de cambios brutos son las de panel, panel rotativo y panel dividido. Las encuestas transversales o repetidas no podrían arrojar este tipo de estimativas puesto que su diseño no considera a los mismos individuos en la muestra en dos periodos de tiempo.

- **Estimación de la incidencia de eventos en un periodo de tiempo**. Suponga que se quiere estimar la *proporción de mujeres que fueron víctimas de un evento de violencia en los últimos seis meses en un país*. En este caso todas las encuestas resultarían aptas mediante ligeras modificaciones en el diseño. Por ejemplo, la encuesta transversal debería preguntar de forma retrospectiva; las encuestas repetidas podrían ser agregadas en los últimos seis meses, las encuestas de tipo panel rotativo y divididas deberían preguntar en cada medición de los últimos seis meses por este evento.

- **Estimación de la incidencia de eventos raros en el tiempo**. Por ejemplo, si se quisiera estimar la *proporción de personas con una enfermedad rara*, es posible que las encuestas transversales y de tipo panel no sean las más apropiadas En el primer caso, dado que el evento es raro por definición, los requerimientos de tamaño de muestra en una encuesta transversal sobrepasarían el presupuesto y los costos de una encuesta regular; en el segundo caso, además de las consideraciones anteriormente planteadas del tamaño de muestra, por la misma definición de evento raro, tampoco sería plausible que en el panel se presentaran estos eventos en los individuos a través del tiempo. Por otro lado, al agregar las encuestas repetidas, las de panel rotativas y la parte nueva del panel dividido, podría ser posible llegar al tamaño de muestra adecuado para poder captar esta incidencia de forma precisa y eficiente.

Estos últimos ejemplos muestran la importancia de contar con procedimientos adecuados de acumulación de datos y encuestas a lo largo de un periodo de interés, por ejemplo de forma anual o semestral. La acumulación de datos genera una buena base inferencial para poder estimar todo tipo de parámetros en una ventana más amplia del tiempo. Es posible acumular datos eficientemente por medio de la agregación de encuestas repetidas. De esta forma se definiría una agregación de datos vertical que añade filas, puesto que en cada levantamiento aparecen nuevos individuos, dado que el diseño de las encuestas repetidas selecciona diferentes individuos en cada punto del tiempo. Este es el caso de la *Gran Encuesta Integrada de Hogares de Colombia* que está diseñada para tener representatividad a niveles de desagregación mayores, juntando los individuos observados en los doce levantamientos continuos en un año.

Por otro lado, las encuestas de panel permiten un tipo diferente de agregación, no basado en individuos, sino en variables en el tiempo. A diferencia de las encuestas repetidas, las encuestas de panel, panel rotativo o panel dividido permiten observar a los individuos en diferentes periodos de tiempo y la agregación puede hacerse de forma horizontal, manteniendo a los individuos en las filas y añadiendo columnas cada vez que se observe una nueva medición en un periodo de tiempo diferente.

Capítulo 4

Definición del marco muestral

Todo procedimiento de muestreo probabilístico requiere de un dispositivo que permita identificar y ubicar a todos y cada uno de las unidades pertenecientes a la población objetivo, las cuales posteriormente participarán en el proceso de selección aleatoria que definirá la muestra. Este dispositivo se conoce con el nombre de **marco de muestreo**.

La mayoría de encuestas de hogares que son probabilísticas se caracterizan por usar marcos de muestreo de áreas (agregados cartográficos en todas sus formas). Aunque también es posible construir marcos de líneas telefónicas fijas y móviles. En general, sin esta herramienta no es posible realizar ningún procedimiento de muestreo probabilístico, y es por esto que la etapa de definir y alistar un buen marco de muestreo es tomada con bastante rigurosidad en las ONE.

I. El marco de muestreo

Como se verá en los capítulos posteriores, dependiendo de la naturaleza del marco de muestreo se pueden proponer diferentes tipos de diseños muestrales. Por ejemplo, cuando se dispone de un marco de elementos, se puede aplicar un diseño de muestreo de elementos; aunque, en algunas ocasiones se utilizan diseños de muestreo de conglomerados aunque se disponga de un marco de elementos. Si no se dispone de un marco de elementos (o es muy costoso construirlo) se debe recurrir a diseños de muestreo en conglomerados; es decir, que se utilizan marcos de conglomerados. Por ejemplo, al realizar una encuesta cuya unidad de observación sean las personas que viven en una ciudad, es muy difícil poder acceder a un marco de muestreo de las personas. Sin embargo, en una primera instancia, se puede tener acceso a la división cartográfica de la ciudad y así seleccionar algunas comunas, localidades, o barrios de la ciudad, para luego seleccionar a las personas en una segunda o tercera instancia. En el ejemplo anterior, las comunas, localidades, o barrios son un ejemplo claro de los conglomerados, que son agrupaciones de elementos que tienen la característica de aparecer naturalmente.

Cuando se dispone de listados de unidades, por ejemplo, el listado de empleados de una entidad, es posible aplicar un diseño de muestreo de elementos, realizar la correspondiente selección aleatoria y de acuerdo a ese mismo diseño realizar las estimaciones necesarias. Sin embargo, al realizar la planeación de una encuesta de hogares, es muy poco probable que se utilicen marcos de elementos, a no ser que el muestreo definido sea en dos fases: con una primera fase de selección de hogares y enlistamiento de personas o unidades, y una segunda fase de selección de personas o unidades. Por ejemplo, el Instituto Nacional de Estadística y Censos (INEC) de Costa Rica realiza la Encuesta Nacional de Microempresas de los Hogares con base en la muestra de la Encuesta Nacional de Hogares (primera fase), en donde se identifican las actividades económicas de los respondientes y se enlistan los trabajadores autónomos. En una segunda fase se selecciona una submuestra con base en este marco de elementos. En general, se pueden listar dos tipos de marcos de muestreo; a saber:

1. **De Lista**: listados físicos o magnéticos, ficheros o archivos de expedientes que permiten identificar y ubicar a los objetos que participarán en el sorteo aleatorio.
2. **De Área**: mapas de ciudades y regiones en formato físico o magnético, fotografías aéreas, imágenes de satélite o similares que permiten delimitar regiones o unidades geográficas en forma tal que su identificación y su ubicación sobre el terreno sea posible.

Es una virtud del marco si contiene *información auxiliar* que permita aplicar diseños muestrales y/o estimadores que conduzcan a estrategias de muestreo más eficientes con respecto a la precisión de los resultados. O también si la información auxiliar[1] está clasificada de forma sistemática y conveniente. La información auxiliar *discreta* en el marco de muestreo permite la desagregación de la población objetivo en categorías o grupos poblacionales más pequeños. Por ejemplo, nivel socioeconómico, región, departamento, etc. Por otro lado, la información auxiliar *continua*, en forma de una o varias características de interés de tipo continuo y positivas, que esté altamente relacionada con la característica de interés permitirá mejorar la eficiencia de la estrategia de muestreo. Por otra parte, un marco de muestreo es defectuoso si presenta alguno o varios de los siguientes casos:

1. **Sobre-cobertura**: se presenta si en el dispositivo aparecen objetos que no pertenecen a la población objetivo. *No son todos los que están.*
2. **Sub-cobertura**: se da cuando algunos elementos de la población objetivo no aparecen en el marco de muestreo o cuando no se ha actualizado la entrada de nuevos integrantes. *No están todos los que son.*
3. **Duplicación**: se presenta si en el dispositivo aparecen varios registros para un mismo objeto. La razón más frecuente para la presencia de este defecto es la construcción no cuidadosa del marco a partir de la unión de registros administrativos de dos o más fuentes de información.

Estos defectos ocasionan errores en el cálculo de las expresiones que se utilizarán para generar las correspondientes estimaciones, generando sesgo, pérdida de precisión y, en algunos casos, que los resultados del estudio se pongan en entredicho. No obstante, una vez que se ha definido el marco de muestreo, este empieza un periodo de decaimiento de su calidad y envejecimiento, conllevando dificultades en la realización de las encuestas de hogares que lo utilizan. Es por esta razón que, a partir de la realización de los censos de población y vivienda, las ONE actualizarán sus marcos de muestreo.

En resumen, el marco de muestreo es cualquier dispositivo o mecanismo usado para obtener acceso observacional a la población de interés, para identificar y seleccionar una muestra, de manera que respete el esquema de muestreo probabilístico y para establecer contacto con los elementos seleccionados, de manera presencial, por correo postal, por correo electrónico, o mediante procedimientos automatizados como los sistemas de captura CAPI (*Computer Assisted Personal Interviewing*) o CATI (*Computer Assisted Telephone Interviewing*).

[1] Toda información disponible a nivel poblacional o para todos y cada uno de los elementos del universo afecta directamente la estrategia empleada para obtener los objetivos de la investigación. Con respecto a la información auxiliar que pueda existir para cada elemento de la población es deseable que esté bien correlacionada con la variable de interés.

Por otro lado, recordando que la población objetivo constituye el conjunto de elementos sobre la cual se desea información y se requieren estimaciones exactas y precisas acerca de sus parámetros, entonces la población del marco es el conjunto de todos los elementos que son enlistados directamente como unidades en el marco o identificados mediante un marco más complejo, tal como un marco para selección en varias etapas. Además, los elementos son las entidades que componen la población y las unidades de muestreo son las entidades del marco muestral. Cuando no hay uno disponible, es posible construirlo. Luego, las siguientes características son deseables para un marco de muestreo:

- Que las unidades en el marco son identificados con un serial.
- Que cualquier unidad puede ser ubicada (dirección, teléfono).
- Que se pueda ordenar sistemáticamente (geografía, tamaño).
- Que contenga información adicional para cada unidad.
- Que especifique el dominio geográfico o socioeconómico al cual pertenece cada unidad.
- Que cada elemento de la población está presente sólo una vez.
- Que no contenga elementos que no estén en la población.
- Que todos los elementos de la población de interés estén en el marco muestral.

La calidad del marco puede ser medida mediante la relación que existe entre la población objetivo y la población del marco. Esto quiere decir que la población enmarcada y la población de interés no siempre van a coincidir plenamente.

En las encuestas de hogares que precisan de un marco de áreas para su realización, el proceso de selección sistemática de los hogares necesita contar con un marco de muestreo que sirva de vínculo entre los hogares y las unidades de muestreo de las primeras etapas y que permita tener acceso a la población de interés. Como lo afirma Gutiérrez (2016), el marco de muestreo más utilizado en este tipo de encuestas es de áreas geográficas que vinculan directamente a los hogares o personas con un listado de divisiones cartográficas completamente exhaustivas. Por esta razón, los diseños de muestreo de estas encuestas se apoyan en la aglomeración natural de los hogares en segmentos cartográficos, que a su vez están contenidos en agrupaciones mayores. ¿Cómo se aglomeran las personas y cómo podemos realizar un diseño de muestreo con base en esta forma de aglomeración? Pues bien, las personas se aglomeran en hogares, los cuales a su vez se aglomeran en comunidades más grandes: barrios, comunas, segmentos. Estas comunidades forman ciudades, veredas, centro poblados, etc. y la reunión de estas divisiones da como resultado el conjunto completo de unidades de interés en el país.

Por lo tanto, a pesar de que ningún país tiene a disposición una lista actualizada de todos los hogares junto con su ubicación e identificación, sí existe en todos los países listas de los segmentos cartográficos presentes en las zonas urbanas y rurales, que son actualizadas en cada censo. De esta forma, si se selecciona de forma probabilística una muestra de sectores y dentro de cada sector se selecciona de forma probabilística una muestra de hogares, entonces de forma indirecta estaremos seleccionando una muestra de hogares que puede representar la realidad de todo un país.

II. Los censos y su incidencia en los marcos de muestreo

Como se mencionó anteriormente, una característica esencial de los diseños de las encuestas de hogares es que la selección de las unidades finales de muestreo debe surtir varias etapas, de acuerdo a las agrupaciones definidas en los marcos de muestreo, que usualmente son marcos de área obtenidos de la división geográfica del país, región o municipio en áreas menores mutuamente excluyentes. Los institutos de estadística en América Latina hacen grandes esfuerzos para mantener actualizados sus marcos de muestreo. Por ejemplo, la *Encuesta Nacional de Hogares* de Costa Rica utiliza un marco muestral construido a partir de los censos nacionales de población y vivienda de 2011 y corresponde a un marco de áreas en donde sus unidades son superficies geográficas asociadas con las viviendas. Este marco en particular permite la definición de UPM con 150 viviendas en las zonas urbanas y 100 viviendas en las zonas rurales. En general, el marco está conformado por 10461 UPM (64.5 % urbanas y 35.5 % rurales).

Gambino and Silva (2009) mencionan que, en la práctica, la consecución de los marcos de lista de lo hogares en la última etapa del muestreo puede tornarse difícil puesto que dentro del conglomerado no es obvio observar de manera exhaustiva los hogares, especialmente cuando la frontera del conglomerado es una línea imaginaria. Por ejemplo, en la mayoría de casos, en el sector urbano, la distinción entre dos conglomerados está demarcada claramente por las calles que conforman la ciudad o el centro poblado; sin embargo, en la ruralidad, no solamente los caminos existentes sirven para delimitar los conglomerados, sino que también los accidentes topográficos y las señales naturales se utilizan para este fin. De la misma manera, esta delimitación se torna compleja cuando han ocurrido cambios en la infraestructura del área y aparecen nuevas construcciones.

Observe que, en general, ante el estudio de un fenómeno social, las desagregaciones geográficas más amplias constituyen un interés natural para los usuarios de las encuestas; es así como los investigadores que planean las encuestas quisieran poder desagregar la información por las regiones geográficas más grandes, que a su vez tienen cierta independencia política y administrativa. Las estadísticas nacionales que se publican a partir de las encuestas de hogares cobran mayor relevancia a nivel de regiones, estados o departamentos. Este tipo de desagregaciones se conocen con el nombre de *dominios de representación*, que a su vez son agregaciones de los *estratos de muestreo*. Los diseños de las encuestas de hogares han ido evolucionando para permitir que este tipo de subpoblaciones tenga representatividad en la encuesta. Aunado a lo anterior, si la característica de interés con la cual se planea la encuesta hace que la distribución de la población sea altamente sesgada, como en el caso de los ingresos o gastos, es recomendable crear estratos de inclusión forzosa con las unidades más importantes en la población. Esta práctica asegura que el error de muestreo sea más bajo.

En promedio, los países de la región realizan los censos cada diez años, aunque en algunos casos este periodo se extiende de forma desafortunada. En este levantamiento masivo de información se enlistan todos los hogares del país, se enumeran todos los habitantes del país y se observan algunas variables de interés que servirán a su vez para asentar las bases de comparación de las cifras en los siguientes diez años. El periodo que existe entre la realización de dos censos se denomina *periodo intercensal* y en este se realizan encuestas de hogares de diferentes constructos económicos y sociales. Los Institutos Nacionales de Estadística (INE) utilizan las particiones geográficas y cartográficas generadas en el levantamiento del censo con el fin de seleccionar, mediante diseños en varias etapas, muestras de hogares. Comúnmente, estas particiones reciben el nombre de secciones cartográficas y están formadas por un número determinado de hogares contiguos. En adelante nos referiremos a estas particiones como unidades primarias de muestreo (UPM), la cuales en el área urbana, pueden ser manzanas o agregaciones de manzanas, y en área rural pueden ser veredas o sectores censales definidos de antemano.

Algunos países hacen uso de la información censal para definir una estratificación socio económica sobre los segmentos cartográficos del marco de muestreo utilizando para tal fin la información recolectada en el censo de población más reciente. Esta práctica representa una ventaja metodológica porque, en la mayoría de encuestas, los parámetros de interés tienen un comportamiento estructural diferente en cada uno de los subgrupos poblacionales creados, tendiendo a tener una mayor precisión en la estimación de los parámetros de interés. Por ejemplo, a partir del censo, es posible crear un índice de condiciones de vivienda y/o bienestar (teniendo en cuenta las definiciones de las necesidades básicas insatisfechas o la pobreza multidimensional) para definir grupos de viviendas mutuamente excluyentes, que contengan viviendas parecidas dentro de ellos, pero que entre ellos sean muy disimiles. De esta forma, es posible estratificar los sectores cartográficos de todo un país y generar estimaciones más precisas de los indicadores sociales (como desocupacioón, pobreza, ingreso medio, etc.).

Para el caso de la *Gran Encuesta Integrada de Hogares* en Colombia, los criterios de estratificación forman dos grupos: el primero correspondiente a las 24 capitales junto con sus áreas metropolitanas y el segundo correspondiente al resto de cabeceras municipales, centros poblados y la ruralidad dispersa. Además, la encuesta también contempla criterios de estratificacion económica a nivel municipal como nivel de urbanización y estructura de la población, basada en la proporción de habitantes con necesidades básicas insatisfechas. De la misma manera, el diseño de la muestra maestra del Instituto Nacional de Estadística y Geografía de México contempla este tipo de estratificación basada en los indicadores generados con la información del Censo de Población y Vivienda 2010. Previo al proceso de estratificación sociodemográfica, fue necesario construir y seleccionar una serie de variables que lograran, en conjunto, separar el universo de UPM en agrupaciones que mejoraran las principales estimaciones de las diferentes encuestas usuarias del marco de muestreo (INEGI, 2012).

Ante la ausencia de un marco de muestreo de hogares y personas en los países de la región, el diseño de las encuestas de hogares se dice complejo puesto que involucra varias etapas de selección y estratificación. Por ende, los marcos de muestreo están conformados por unidades primarias de muestreo (UPM) que se definen como segmentos cartográficos individuales, como una agrupación de segmentos o incluso como una división de segmentos masivos. Por ejemplo, tomando en consideración el estrato urbano, en donde las UPM corresponden a manzanas (o agregaciones o particiones de manzanas), mientras que en el caso rural, las UPM corresponden a comunidades (o agregaciones o particiones de comunidades). En cualquier caso, la unidad de observación está constituida por las viviendas ocupadas particulares donde residen personas. En general, salvo en algunos países, las UPM no tienen el mismo tamaño dentro de los estratos; es decir no están constituidas por un número igual de viviendas. El caso es más evidente es la ruralidad, en donde podría ocurrir que una única UPM agrupe un conjunto de viviendas con demasiada heterogeneidad y una alta dispersión geográfica. Es así como es posible encontrar UPM con pocas viviendas o UPM con demasiadas viviendas. Esto constituye una desventaja técnica a la hora de establecer metodologías apropiadas para la recolección de la información primaria y además para la estimación de los errores de muestreo que se derivan de la encuestas de hogares y por esto algunos países están considerando la re-definición de las UPM como unidades con un número uniforme de viviendas.

Como se indicó anteriormente, es usual que tras el levantamiento de un nuevo censo se actualice el marco de muestreo con el que se seleccionarán las viviendas y hogares para todas las encuestas subsiguientes. Por la naturaleza de los censos, los INE deben recorrer la geografía de los países produciendo una nueva cartografía que derivará en la actualización de los marcos de muestreo. Por ejemplo, considere un país que cuente con un marco de muestreo que consta de diez mil UPM y, cada una de estas deberá ser clasificada por medio de una estratificación socioeconómica que estará basada en la información recolectada en el último censo de población y vivienda. Kish (1965, pág. 183) afirma que la selección de UPM con tamaño desigual acarrea algunos problemas técnicos como que el tamaño de muestra final se convierte en una variable aleatoria, que depende de la probabilidad de selección de las UPM más grandes o más pequeñas. Lo anterior aumenta la incertidumbre en el costo final del operativo, pues si en una primera instancia se seleccionan UPM con pocas viviendas, será necesario volver a realizar un proceso adicional de selección de nuevas UPM para cumplir con la cuota de viviendas.

Con base en lo anterior, se esperaría que la actualización de la cartografía y de los marcos de muestreo se realizara mínimo cada diez años. Es importante que estas actualizaciones conlleven a una definición de los marcos de muestreo que permitan tener mayor fluidez en los procesos logísticos de selección de hogares y que induzcan una mejora en la precisión de las estimaciones de los parámetros de interés. Por ejemplo, una forma muy conveniente de abordar este desafío es creando UPM que contengan, en la medida de lo posible, un mismo número de viviendas y, de esta manera, mantener una distribución uniforme en cada estrato. Siguiendo el consejo de Valliant et al. (2013, pág. 212), si el equipo de planeación de la encuesta tiene la flexibilidad de definir las UPM, como usualmente es el caso en las encuestas de hogares, entonces las UPM definitivamente deberían estar conformadas por una cantidad igual de viviendas.

III. Construcción de las UPM

La definición del marco de muestreo para las encuestas de hogares responde básicamente a un objetivo: la definición de las unidades primarias de muestreo. En la búsqueda de la optimización de esta solución, es necesario responder una pregunta fundamental: ¿cuál debe ser el tamaño apropiado para las UPM? No es lo mismo definir las UPM como agregaciones de 20 hogares, que de 1000 hogares. Esta pregunta debe ser abordada, en principio, desde una perspectiva técnica, en donde confluyan diferentes perspectivas (de muestreo, logísticas, presupuestales, cartográficas). Por ejemplo, Valliant et al. (2013, Tabla 9.1) mencionan el caso en el que, para diferentes definiciones del tamaño de las UPM, se evidencian pérdidas o ganancias significativas de eficiencia en los estimadores de las encuestas de hogares.

De esta manera, un primer acercamiento a la definición de las UPM es establecer la unión o colapso de los mismos lugares poblados, sectores o secciones cartográficas, o áreas de empadronamiento vinculados a los censos de población y vivienda, como insumo para la creación de las unidades primarias de muestreo. Como se discutió anteriormente, el objetivo del marco es tratar de proveer la mejor información de en la selección de las unidades, reduciendo la variabilidad de la estrategia de muestreo. Por lo tanto, después de revisar minuciosamente los conjuntos de datos censales y la información cartográfica del censo en los niveles básicos (en adelante, y sin pérdida de generalidad, lo llamaremos secciones censales) es necesario construir un algoritmo que permitía crear UPM desde la cartografía, basado en uniones contiguas de secciones censales, que respeten los siguientes principios:

- La conformación de las Unidades Primarias de Muestreo (UPM) excluye todas las estructuras que no contienen hogares particulares ocupados.
- Las nuevas UPM inducidas por la unión de sectores censales deben estar contenidas de manera en un sólo municipio del país; es decir no podrán definirse UPM que pertenezcan a dos o más municipios.
- De la misma forma, debe haber una diferenciación estricta en las áreas urbanas y rurales. Ninguna UPM podrá estar definida en ambas áreas.

Nótese que siempre será necesario realizar una actualización de las viviendas con hogares particulares ocupados en las UPM seleccionadas en la primera etapa de muestreo. Esta actualización dará lugar al cálculo de las probabilidades de inclusión de segunda etapa, sin la cual no se podrían calcular factores de expansión que induzcan el insesgamiento de los estimadores utilizados en las encuestas de hogares. Dado que este proceso es sistemático y debe ser realizado a lo largo del periodo intercensal, contar con UPM demasiado grandes (como lo pueden llegar a ser los sectores o segmentos censales, las áreas de empadronamiento o los lugares poblados) no es una alternativa viable presupuestariamente puesto que se incrementarían los costes asociados a la actualización y no habría uniformidad en los procesos de muestreo.

Usualmente el tamaño de las UPM en América Latina ronda el rango de 75 a 225 viviendas. Para que exista una mayor eficiencia (logística y estadística) a la hora de realizar un muestreo en dos etapas, se recomienda que las UPM conformadas tengan algún grado de explicación con respecto a las características de interés que se quieren medir en la población. Por consiguiente, es necesario revisar los tamaños de estas agregaciones y su comportamiento en términos del *coeficiente de correlación intraclase* (ICC). Como se puede notar en Cochran (1977) y Gutiérrez (2016), en la construcción de las UPM, el parámetro predominante que se debe considerar es ICC, que para la realización de encuestas con selección en múltiples etapas puede ser aproximado mediante la siguiente expresión (Valliant et al., 2013)

$$ICC = \frac{SCE}{SCE + SCD}$$

En donde SCE es suma de cuadrados relativa de los totales de la característica de interés entre las UPM y SCD es la suma de cuadrados relativa de los totales de la característica de interés dentro las UPM. El ICC es una medida de homogeneidad entre las variables que se desean medir y la conformación de las UPM. Además de afectar la variabilidad del estimador en muestreos multietápicos, esta medida determina el tamaño de muestra necesario para satisfacer los requerimientos de precisión en una encuesta de hogares. En algunos textos clásicos de muestreo, el ICC también es denotado como ρ.

La magnitud del ICC está directamente ligada al tamaño de las UPM. Por ende, en la conformación del marco de muestreo, es necesario ejecutar un algoritmo de control de tamaño de las UPM de tal forma que el ICC sea satisfactorio y coherente en los indicadores censales disponibles, como por ejemplo las dimensiones del índice de necesidades básicas insatisfechas (NBI), los indicadores del mercado de trabajo, los indicadores demográficos, entre otros.

En general, cuando el tamaño de las UPM es muy pequeño, las características de los elementos dentro de las UPM serán muy similares (sobre todo para indicadores socioeconómicos); por otro lado, si el tamaño de las UPM es demasiado grande, las características de los elementos serán más heterogéneas. Nótese que la disparidad en los tamaños de las UPM redunda en que los totales de las características de interés serán muy disimiles entre las UPM, y teniendo en cuenta la forma funcional de la varianza del estimador clásico, se generará más varianza en el componente B^2, por ende el ICC será más grande y se perderá precisión en el muestreo multietápico.

Valliant et al. (2013) afirman que la práctica estándar es combinar las secciones pequeñas o grupos de bloques cercanos geográficamente para que todas las UPM tengan al menos un número mínimo de personas. Dado que la variación en los tamaños de las UPM tiene un efecto marcado en el ICC (medida necesaria para diseñar una muestra), y que en el caso de las encuestas de hogares se puede tener una cierta flexibilidad en la formación de estos grupos, entonces las UPM deberían conformarse con un número casi igual de viviendas. En general el proceso de construcción de las UPM debería tener en cuenta las siguientes características:

1. *Límites y contenencia*, pues las UPM deben estar contenidas dentro de límites departamentales, municipales, y estar diferenciadas por su naturaleza urbana o rural.
2. *Tamaño y extensión*, pues se debe procurar que las UPM estén dentro de rangos predefinidos en términos del número de viviendas y personas, respetando los límites geográficos, y que su extensión en kilómetros cuadrados no sea superior a un umbral predefinido para el operativo de campo.

De esta forma las cargas de trabajo (en los procesos de actualización, supervisión y levantamiento de la información primaria) serán uniformes. Además las estimaciones resultantes serán óptimas en términos de eficiencia y precisión estadística, puesto que inducirán pesos de muestreo uniformes que minimizarán la varianza de las estimaciones directas. A partir de la información contenida en los censos de población y vivienda, diferentes variables se podrían utilizar para evaluar la idoneidad de las UPM con el coeficiente de correlación intraclase y el efecto diseño (DEFF). Por ejemplo, para evaluar la idoneidad de las UPM es posible analizar las siguientes variables agrupadas en los siguientes constructos:

1. Variables demográficas: grupos quinquenales de edad, sexo.
2. Necesidades básicas insatisfechas y sus dimensiones (acceso a la vivienda, acceso a servicios sanitarios, acceso a educación, situación en la ocupación y capacidad económica).
3. Variables de fuerza laboral: población en edad de trabajar, población económicamente activa, desocupados y ocupados.

En general, las medidas de correlación intraclase deben ser coherentes con las experiencias locales anteriores o con experiencias regionales que demuestren que el algoritmo de colapso y/o escisión de los sectores censales sí proporcione como resultado nuevas UPM que conserven las propiedades explicativas de los grupos desde el censo, con la ventaja de controlar su tamaño en viviendas.

Hansen et al. (1953) encontraron un efecto marcado en el tamaño de las UPM y la magnitud del ICC. Entre más pequeñas sean los conglomerados mayor será el ICC, entre más grandes sean los conglomerados menor será el ICC. Esta relación tiene una repercusión directa en la forma en que se llevarán a cabo las encuestas en el periodo intercensal. Si se crean UPM demasiado pequeñas, se precisará de un tamaño de muestra de UPM mucho mayor, y por ende un mayor coste logístico y económico. Si se crean UPM demasiado grandes, se precisará de un menor tamaño de muestra, pero con UPM inmanejables en su dimensión, que acarrearán operativos de actualización, supervisión y levantamiento demasiado costosos, junto con una pérdida grande de precisión estadística.

Para ejemplificar la relación entre el ICC y el tamaño de muestra, considere los siguientes escenarios:

1. Si el ICC es cercano a cero, las UPM serán demasiado heterogéneas por dentro y muy homogéneas entre, por tanto se necesitará de muy pocas UPM para tener una inferencia precisa. Esto quiere decir que hay mucha dispersión dentro de las UPM, pero a la vez hay muy poca variación entre ellas. En el caso que el ICC sea idéntico a cero, sólo se necesitaría de una UPM en la muestra para tener una estimación precisa, con un submuestreo exhaustivo de todas las unidades dentro de la UPM (puesto que todas las unidades dentro de la UPM serán diferentes).

2. Si el ICC es cercano a uno, las UPM serán demasiado homogéneas por dentro y muy

heterogéneas entre, por tanto se necesitará de una muestra grande de UPM para tener una inferencia precisa. Esto quiere decir que hay poca dispersión dentro de las UPM, pero a la vez hay mucha variación entre ellas. En el caso que el ICC sea idéntico a uno, para obtener una estimación precisa, se necesitaría de una muestra censal de UPM, en donde el submuestreo sea de una sola unidad (puesto que todas las unidades dentro de la UPM serán idénticas).

En resumen, la construcción de las UPM es un proceso que requiere de la más alta disposición de capacidades para que todas las operaciones estadísticas del periodo intercensal sean balanceadas en presupuesto y esfuerzo logístico. La función objetivo de este proceso es el ICC que, como se verá en los capítulos posteriores, determina el tamaño de muestra y la precisión de la inferencia.

Actualización continua del marco de muestreo

Duncan and Kalton (1987, pág. 105) afirman que la composición de la población de interés cambia durante el tiempo, puesto que lo individuos nacen, mueren, migran, e incluso pasan a ser parte de organizaciones que hacen que pierdan su estatus[2] de la unidad de observación. De igual forma, se debe tener en cuenta los nuevos hogares que pueden crearse o desintegrarse.

La realidad de los países latinoamericanos muestra una migración importante desde las áreas rurales hacia las áreas urbanas y esto repercute en una desactualización constante del marco de muestreo que fue construido varios años atrás. Este problema de actualización del marco lo enfrentan todos los países de la región y puede ser abordado a partir del ajuste constante a los pesos de muestreo de las UPM cada vez que se realiza un operativo de campo en donde haya evidencia de un cambio en el número de hogares para las UPM seleccionadas en la muestra de la primera etapa.

Como las UPM se seleccionan con un muestreo proporcional al tamaño de la UPM y las viviendas se seleccionan en campo mediante un muestreo simple (aleatorio simple o sistemático), previa actualización del empadronamiento y conteo de viviendas; entonces esta actualización podría usarse para reajustar los pesos de las UPM en los nuevos levantamientos. De esta forma se reflejaría el cambio que tiene la población (dinámica, por definición) de interés. Sin embargo, se recomienda no modificar las probabilidades de selección de las UPM para garantizar el insesgamiento de los estimadores de muestreo.

Por ejemplo, si en un país se define un esquema de muestreo que selecciona 12 viviendas dentro de cada una de las UPM seleccionadas en la primera etapa, entonces la probabilidad de selección de la i-ésima UPM U_i estaría dada por

$$\pi_{Ii} = Pr(U_i \in s_I) = n_I \frac{n_i}{N_i} = n_I \frac{12}{N_i}$$

En donde n_I hace referencia al número de UPM que se seleccionarán en la primera etapa, N_i representa el número de viviendas en la UPM y $n_i = 12$ es el número de viviendas seleccionadas dentro de la UPM. Ahora, si el número de viviendas se actualizara en la UPM, la probabilidad de inclusión cambiaría, lo cual generaría sesgo en la estimación. Por ende, las probabilidades de inclusión de las UPM deberían seguir estables entre los ciclos censales. El problema de subcobertura puede abordarse con el post-ajuste de los factores de expansión en la etapa de estimación.

[2]Nótese que la población objetivo de la mayoría de encuestas de hogares en la región se refiere a la población civil no institucionalizada, que excluye miembros de organizaciones militares, personas en cárceles, hospitales, etc.

Capítulo 5

Metodologías de estratificación

Para aumentar la eficiencia de la inferencia en las encuestas de hogares, es de particular interés que el marco de muestreo permita clasificar a las UPM de acuerdo con su nivel socio-económico con el fin de poder realizar selecciones independientes en cada categoría de la clasificación. De esta forma se garantiza la homogeneidad dentro de los grupos y se disminuye la incertidumbre de la estimación. Este proceso se conoce con el nombre de estratificación.

En la literatura especializada, es posible encontrar varias metodologías que clasifican a cada una de las UPM del marco y a la vez disminuyen la varianza de los estimadores de muestreo. Este capítulo realiza un resumen no exhaustivo de las principales técnicas utilizadas por los INE de la región, se proponen algoritmos para encontrar la mejor estratificación basada en los datos de los censos y se ilustran los procedimientos computacionales necesarios para implementar esta metodología. Si los estratos están conformados por unidades homogéneas que, a su vez, crean categorías heterogéneas entre sí, entonces se dice que el proceso de estratificación es eficiente y el error de muestreo se verá reducido significativamente.

Luego de definir las UPM en el marco de muestreo es necesario realizar una agrupación de éstas de acuerdo con sus características sociodemográficas agregadas con el fin de obtener una partición que conforme grupos homogéneos y que induzcan una mayor precisión en la ejecución de las estrategias de muestreo que se propongan dentro de la planificación de las encuestas de hogares que realizan los INE. Es importante señalar que se debe estudiar una multitud de escenarios de estratificación y para encontrar una óptima clasificación de las unidades primarias de muestreo, puesto que esta partición será utilizada en todas las encuestas de hogares que utilicen este marco de muestreo en el periodo intercensal.

En síntesis, los procesos que intervienen en la estratificación del marco de muestreo son los siguientes:

1. Ejecución de múltiples escenarios de estratificación de las UPM utilizando información agregada del censo[1].
2. Para cada método señalado anteriormente realizar particiones de 3, 4, o 5 grupos a nivel nacional y evaluar la pertinencia de realizarlo en las áreas rural y urbana de forma independiente.

[1]Aunque también es posible añadir información geoespacial, catastral o de cualquier índole si se tiene una cobertura completa a nivel de las UPM.

3. A raíz de las pruebas y los escenarios estudiados, evaluar su efectividad mediante una única medida de calidad, definida como el DEFF generalizado y escoger el mejor escenario en términos de esta medida en conjunción con la viabilidad logística con respecto al número de particiones.

Este capítulo presenta los diferentes procesos utilizados en el proceso de estratificación; establece la forma de agregación de las variables a nivel de las UPM para mantener una estructura uniforme que permita sacar un mejor provecho a la discriminación entre sus estructuras y, por ende, una mejor clasificación en los estratos; resume de forma no exhaustiva algunas de las metodologías usadas para la estratificación de marcos de muestreo (teniendo en consideración dos enfoques: univariados sobre medidas de resumen, y multivariados sobre toda la matriz de estratificación); presenta los criterios de evaluación de los métodos de estratificación; e ilustra los resultados finales de la estratificación de un marco de muestreo exponiendo las consideraciones más importantes.

I. Dimensiones estructurales en el marco de muestreo

El proceso de definición de un diseño de muestreo para las encuestas nacionales que necesita un país para responder a sus necesidades de información con miras en la elaboración de sus políticas públicas involucra varios procesos que hacen uso de los censos nacionales de población y el uso de una cartografía detallada del territorio nacional.

Como se indicó en el capítulo anterior, un aspecto fundamental para el diseño y desarrollo de encuestas de hogares involucra la definición de las UPM, definidas como unidades cartográficas que dividen el territorio nacional y permiten llevar a cabo los procesos de levantamiento de información y de trabajo de campo de la manera más idónea posible, y que además se construyen con el fin de facilitar la obtención de estimaciones precisas y confiables de los indicadores y parámetros de interés que requieren los tomadores de decisiones y expertos en políticas públicas.

Dependiendo de la planificación de las diferentes encuestas, las UPM pueden dar lugar a unidades secundarias de muestreo o permitir la selección directa de las unidades de análisis como las viviendas, los hogares y/o las personas. Independientemente de las unidades de muestreo y las jerarquías que se definan para llevar a cabo la implementación del diseño de muestreo para las encuestas, es fundamental llevar a cabo un proceso de estratificación de las UPM en grupos que sean en lo posible lo más homogéneos en cuanto a sus características socioeconómicas y de bienestar y que definan una partición del territorio nacional (Gutiérrez, 2016).

Estos grupos se denominan en la literatura estadística como estratos y su unión debe cubrir todo el territorio nacional. Como estos grupos determinan una partición, dos estratos cualesquiera deben ser mutuamente excluyentes. Los INE utilizan las particiones geográficas y cartográficas generadas en el levantamiento del censo con el fin de seleccionar muestras de hogares, mediante la ejecución de diseños de muestreo probabilísticos, estratificados y en varias etapas. En particular, para aumentar la eficiencia de la inferencia en las encuestas de hogares, es de particular interés que el marco de muestreo permita clasificar a las UPM de acuerdo con su estructura socioeconómica, con el fin de poder realizar selecciones independientes en cada categoría de la clasificación.

Al garantizar la homogeneidad dentro de los estratos se disminuye la incertidumbre de la estimación y se minimizan los errores de muestreo que se obtienen al realizar encuestas con procedeimientos de muestreo probabilístico. En el caso particular de los países latinoamericanos, este proceso se lleva a cabo haciendo uso de la información censal a nivel de personas, hogares y viviendas, en diferentes constructos o dimensiones asociadas a la calidad de vida y bienestar (demografía, características de la vivienda, tenencia de enseres y servicios públicos entre otros). Las variables que se definan sobre estos constructos son agregadas partiendo de variables binarias que toman el valor de uno, si el fenómeno en cuestión está asociado de forma positiva a mejores condiciones socioeconómicas, y cero, en cualquier otro caso. Por ejemplo:

- El acceso del hogar a una conexión de internet puede ser una variable de interés en la estratificación puesto que discrimina entre los hogares con mejores condiciones de bienestar. En este caso, la variable se define como uno (1) si el hogar dispone del servicio de internet y cero (0) si el hogar no dispone de dicho servicio.
- La materialidad de los pisos, paredes y techos también pueden ser variables importantes en la estratificación de las UPM. Mejores materiales se asocian a una mayor capacidad económica y mejores condiciones habitacionales. Estas variables se definirán como uno (1) si la vivienda no tiene materiales precarios y cero (0) en otro caso.

El proceso anterior se realiza basado en referentes internacionales de calidad de vida y en un análisis exploratorio de datos de las diferentes variables candidatas a participar en el proceso de estratificación. En primer lugar, es necesario tomar en consideración que la estratificación que se pretende realizar debe ser llevada a nivel de las UPM. Esto implica que una vez que las UPM estén categorizadas en algún estrato, todos sus componentes también estarán clasificados en la misma categoría; por consiguiente, las personas, los hogares y las viviendas de la UPM pertenecerán al estrato en el cual la UPM fue clasificada.

Con la información del censo se deben seleccionar y definir las variables que estén relacionadas directamente con los fenómenos que se estudiarán en las diferentes encuestas de hogares a lo largo del periodo intercensal. Una vez construidas las UPM, se calculan los agregados de las variables seleccionadas en las dimensiones observadas desde los censos, que por lo general son las siguientes:

- *Demografía y estructura de la población*: sexo, edad, parentesco, origen extranjero, pertenencia a grupos indígenas, número de hijos, número de dependientes, etc.
- *Educación*: analfabetismo, asistencia escolar, años de estudios, grado de escolaridad, etc.
- *Mercado de trabajo*: población en edad de trabajar, pertenencia a la fuerza de trabajo por sexo, condición de ocupación por sexo, rama de actividad, etc.
- *Características de la vivienda*: tipo de vivienda, materiales de construcción, hacinamiento, equipamiento, etc.
- *Acceso a servicios básicos*: fuente de agua, alcantarillado, acceso a salud, acceso a seguridad social, etc.
- *Necesidades básicas insatisfechas (NBI) o pobreza multidimensional*: viviendas con hacinamiento crítico, servicios inadecuados, alta dependencia económica, niños en edad escolar que no asisten a la escuela, precariedad en el aseguramiento en salud, entre otras.

La caracterización de estas dimensiones lleva a clasificar a las UPM en el marco. Por ejemplo, en la dimensión demográfica, es común que las UPM con mayor número de personas que se identifican como indígenas o pertenecientes a alguna etnia se relacionen con menores niveles de calidad de vida. De la misma manera, con los recientes fenómenos migratorios en la región, hay evidencia empírica de que las UPM que agrupan a extranjeros venidos de otros países latinoamericanos están relacionadas con menores condiciones de bienestar. Asimismo, las UPM con un mayor porcentaje de niños en la primera infancia y con madres cabeza de familia generalmente se asocian con dificultades en su calidad de vida.

De la misma manera a nivel de educación, las UPM con mayores tasas de analfabetismo (que por lo general están en las áreas rurales), y con niños que no asisten a la escuela se asocian a menores condiciones de bienestar; mientras que las UPM que tienen un mayor porcentaje de población con estudios de educación superior (que por lo general se encuentran en las áreas urbanas de las ciudades grandes) se asocian con mejores condiciones de bienestar.

En la dimensión ocupacional, las UPM rurales concentran una alta proporción de población ocupada que no necesariamente muestra mejores condiciones de vida. Por otra parte, las UPM que tienen una mayor incidencia de población desocupada y/o mayor proporción de personas dependientes (personas de 0 a 15 años o mayores de 65 años) pueden relacionarse con peores condiciones de vida.

Con respecto a las características de la vivienda, está bien documentado que las UPM con alto porcentaje de viviendas cuyos materiales de construcción de paredes, techos y pisos es precario se asocian con menores condiciones de bienestar y por lo general se presentan con mayor incidencia en las áreas rurales y en las áreas marginales de las zonas urbanas. De la misma manera las UPM que concentran viviendas con hacinamiento (por ejemplo, si el número de personas del hogar sobre el número de cuartos es mayor a tres) o con acceso inadecuado a las fuentes de agua potable, o con servicios sanitarios y de eliminación de aguas grises deficientes están asociadas a un menor bienestar socioeconómico.

II. Información a nivel de UPM

Cabe resaltar que, tomando en cuenta la información recolectada en el censo, es posible también clasificar a las personas o a los hogares en una primera instancia para después agregarlos hasta llegar a una clasificación única de la UPM; sin embargo, en la práctica este proceso puede resultar un poco más complejo y no son claras sus ventajas. Por lo anteriormente mencionado, este capítulo estará enfocado en la clasificación de las UPM a partir de una matriz de información a nivel de esta misma agregación.

Debido a que las UPM tienen, en estricto rigor, tamaños diferentes, la escala y el nivel en el que se midan los indicadores puede afectar los procesos de clasificación. Luego, si la matriz de información con la cual se realiza la estratificación se construye con base en el número de personas (con determinadas características) dentro de la UPM, al no tener en cuenta el tamaño de ésta, es muy probable que las metodologías de estratificación no logren agrupar de forma homogénea a las UPM. Por ejemplo, asuma que hay dos UPM con tamaño 100 y 300 hogares, que agrupan a 200 y 400 personas en la fuerza de trabajo, y además suponga que una de las variables de la matriz de información se define como el número de personas ocupadas. A su vez, asuma que la primera UPM pertenece a un sector acaudalado y la segunda UPM pertenece a un sector marginal. Es posible que el número de personas ocupadas en ambas UPM sea de 150 y que por esta razón queden erróneamente clasificadas en el mismo grupo. Por ende, definir la matriz de información en términos relativos (porcentaje de ocurrencia de cada variable) es una mejor alternativa para que el agrupamiento esté controlado por el tamaño de la UPM y supeditado únicamente a cambios estructurales en los constructos de medición del censo.

Por último, una vez que se ha definido el conjunto de variables que entrarán en la matriz de información, es necesario verificar que todos los indicadores de esta matriz apunten hacia el mismo horizonte del constructo censal. Es decir, que **todos** los indicadores estén expresados en términos de acceso al bienestar de cada uno de los constructos. Además, es necesario realizar un proceso de refinamiento sobre esta matriz para eliminar aquellas variables que puedan estar altamente correlacionadas con el resto de las variables o que puedan expresarse como combinación lineal de otras variables. De esta manera, se evitan los problemas de multicolinealidad y se asegura una estratificación parsimoniosa. Al final se debe contar con una matriz de información $\mathbf{X}$ compuesta por P columnas (variables de estratificación), y N_I filas (número de UPM en el marco de muestreo); en donde cada fila de la matriz de información representará la i-ésima observación de las UPM a nivel censal para cada una de las P variables.

La teoría estadística ha definido que la mejor estratificación es aquella que minimice los errores de muestreo de los estimadores, expresados en forma de varianzas o errores estándar. Además, una particularidad de los procesos de estratificación es que las varianzas de estos estimadores dependen a su vez de la variación de los microdatos observada en el censo. Sin embargo, lo que podría resultar ser una estratificación óptima para un indicador tal vez sea, al mismo tiempo, una estratificación ineficiente para otras indicadores Más aun, sabiendo que no todas las variables de interés que se observarán en las encuestas durante el periodo intercensal han sido medidas y observadas en el censo, se debe estudiar muy bien, por medio del estudio de numerosos escenarios, qué estratificación utilizar.

Hay un entendimiento tácito en todos los países de la región, repaldado en mayor o menor grado por evidencia empírica, de que la mayoría de los fenómenos sociales que se observan en las encuestas de hogares están supeditados a la distribución de la población en las UPM. Por ejemplo, si lo que se quiere medir es la informalidad en el mercado de trabajo, seguramente nos encontraremos con que este fenómeno está mucho más presente en aquellas UPM marginales, en donde también estarán presenten otros fenómenos como menos años de educación, menores tasas de acceso a la salud, menores ingresos y gastos, mayores tasas de embarazo adolescente, entre otros. De esta forma es necesario analizar las relaciones e incidencias de cada variable incluida. Por ejemplo:

- Analizar si la proporción de techos y paredes adecuadas se encuentra altamente correlacionada con la proporción de pisos adecuados.
- Tener en cuenta si la proporción de extranjeros es muy poco frecuente y sólo aparecen en algunas UPM muy específicas; en ese caso se recomendaría excluir esta variable dada su falta de discriminación.
- Analizar si la proporción de hogares con computadora y lavadora se correlacionan muy bien con la tenencia de internet y refrigerador, por lo cual estas variables no se considerarían en la matriz de estratificación.
- Evaluar si la tenencia de estufa y radio presentan indicadores muy altos a lo largo de las UPM y no incorporan capacidad de discriminación en el proceso de estratificación.

Este análisis exhausitvo de las características poblacionales indica que existe una alta correlación entre la UPM que se habita y la incidencia de fenómenos sociales y económicos. Por lo tanto, los ejercicios de estratificación que se deben estudiar tendrán una alta consistencia interna, de tal manera que al escoger la mejor estratificación se garantiza que los INE dispondrán de una clasificación óptima en el periodo intercensal para todas las encuestas de hogares que se ejecuten.

En general, hay dos grandes escenarios que deben ser revisados al momento de proponer una estratificación: univariados (sobre una medida de resumen de la matriz de información) y multivariados (sobre todas las variables de la matriz de información). Para cualquiera de estas, se recalca que el objetivo es encontrar la mejor partición que asegure que la varianza de los estimadores de muestreo sea mínima. A continuación, se presentan algunas técnicas que se pueden considerar y que además están disponibles en el software estadístico `R` mediante las librerías `stratification` (Baillargeon and Rivest, 2011) y `SamplingStrata` (Barcaroli, 2014). En ambos casos existe documentación disponible acerca de cómo utilizar las funciones de estratificación.

III. Metodologías univariadas sobre medidas de resumen

Es bien sabido que la mejor estratificación para una variable de interés es aquella que nace de su propia variación. Durante muchos años, se desarrollaron técnicas de estratificación sobre una sola variable de interés que dejaban de lado el carácter multipropósito de cualquier encuesta de hogares. Por esta razón, se sugiere partir de la matriz de información y resumir la variación y las correlaciones entre variables mediante alguna técnica multivariada de reducción de datos, como componentes principales, análisis factorial, o modelos no lineales. Como la matriz de información está en escala de porcentajes, es posible que la variabilidad recogida por la medida de resumen sea alta.

Por ejemplo, si se utiliza la técnica de componentes principales, entonces se tomaría como medida de resumen el primer componente, que resulta ser función del vector propio asociado al mayor valor propio de la matriz de covarianzas asociada a la matriz de información. Por otro lado, si se utilizara un análisis factorial confirmatorio, la medida de resumen podría ser el eje principal con la carga factorial más alta. La interpretación de estas medidas de resumen es una parte importante en la aplicación de las técnicas de estratificación. Nótese que la matriz de información está construida por cinco constructos censales (*demografía y estructura de la población*, *educación*, *mercado de trabajo*, *características de la vivienda* y *acceso a servicios básicos*) que deberían ser resumidos en una medida de bienestar de la UPM, que a su vez debe tener sentido en cuanto a la relación (o contribución) de las variables al componente o factor. En adelante, se utilizará la siguiente notación para referirse a la medida de resumen como función de todas las variables incorporadas en la matriz de información:

$$y = f(x_1, \ldots, x_P)$$

Nótese que se esperaría que esta variable de resumen, al estar definida como una medida de bienestar sobre las UPM, tuviera un comportamiento sesgado, tal como se puede observar en la figura 5.1. Por ende, si esta característica es altamente sesgada, puede ser recomendable crear un estrato de inclusión forzosa con estas unidades. Esta práctica asegura que el error de muestreo para este estrato sea nulo. A continuación se enumeran algunas técnicas de estratificación comúnmente utilizadas en la práctica estadística.

Partición en cuantiles (Q)

Este método divide la población de UPM en grupos creados a partir de la división en intervalos regulares de la distribución de la medida de resumen. Los cuantiles más usados son los cuartiles (que dividen la población en cuatro grupos), los quintiles (que dividen la población en cinco grupos) y los deciles (que dividen la población en 10 grupos); sin embargo, con los propósitos de estratificación, también es útil considerar la partición en terciles (que dividen la población en tres grupos).

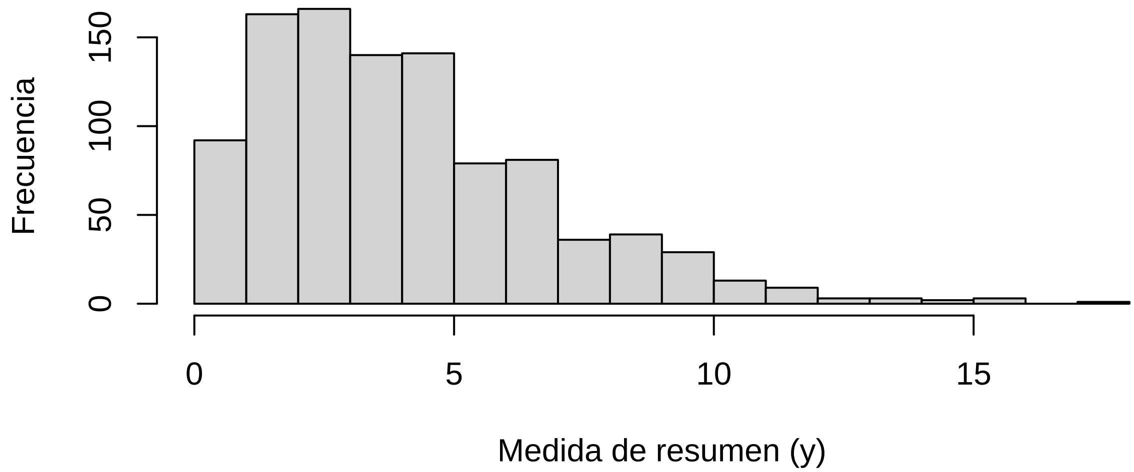

Figura 5.1: *Histograma de la medida de resumen (y) sobre las UPM*

Método de raíz de frecuencia acumulada (DH)

Dalenius and Hodges (1959) propusieron esta técnica de estratificación basada en la raíz cuadrada de las frecuencias acumuladas de la medida de resumen sobre las UPM. Esta técnica es exacta y no requiere de algún procedimiento iterativo. La idea principal de esta técnica es encontrar grupos que minimicen la siguiente función:

$$D = \sum_{h=1}^{H} W_h \sqrt{S^2_{y_h}}$$

En donde $W_h = N_h/N$ $(h = 1, \dots, H)$ es el tamaño relativo del estrato h y $S^2_{y_h}$ es la varianza de la medida de resumen en el estrato h.

Estratificación óptima (LH)

Lavallée and Hidiroglou (1988) propusieron por primera vez la construcción de una estratificación óptima para poblaciones de encuestas reales, basada en la minimización de la siguiente expresión ligada a la varianza de una estrategia de muestreo estratificada.

$$\sum_{h=1}^{H-1} \left(\frac{N_h}{N}\right)^2 \left(\frac{1}{(n-N_H)a_h} - \frac{1}{N_h}\right) S^2_{x_h}$$

En donde N_h es el número de UPM en el estrato h, n es el tamaño de muestra de las UPM, N es el número de UPM en el marco de muestreo, $S^2_{x_h}$ es la varianza de la medida de resumen en el estrato h. Finalmente a_h es la regla de asignación para el tamaño de muestra, dada por la siguiente relación:

$$a_h = \frac{\gamma_h}{\sum_h \gamma_h}$$

En donde, tomando en cuenta que $\bar{X}_h$ es la media de la medida de resumen en el estrato h, entonces, según (Baillargeon and Rivest, 2011), γ_h es proporcional al tamaño de muestra n y está definida por:

$$\gamma_h = N_h^{2q_1} \times \bar{X}_h^{2q_2} \times S_{x_h}^{2q_3}$$

Por tanto, dado que $n_h = n \times \gamma_h$, si se quisiera una estrategia de muestreo que asigne el tamaño de muestra de manera proporcional a cada uno de los estratos, entonces la regla de asignación debería estar determinada por

$$\mathbf{q} = (q_1, q_2, q_3)' = (0.5, 0, 0)'$$

La asignación de Neyman corresponderá con $\mathbf{q} = (0.5, 0, 0.5)'$; mientras que la asignación de potencia con exponente 0.7 estará dada por $\mathbf{q} = (0.35, 0.35, 0)'$. Los detalles técnicos de estos tipos de asignación pueden ser encontrados en Gutiérrez (2016).

La optimización de la función objetivo puede ser llevada a cabo de diferentes formas. En efecto, Lavallée and Hidiroglou (1988) utilizaron un algoritmo de optimización (Sethi) para encontrar los valores óptimos. Baillargeon et al. (2007) definen los pasos necesarios para implementar el procedimiento basado en el algoritmo de Sethi. Asimismo, Kozak (2004) definió un algoritmo iterativo mediante arranques aleatorios para optimizar el proceso de minimización de esta técnica de estratificación.

Estratificación geométrica (GH)

Utilizando las técnicas de estratificación mencionadas anteriormente, algunos autores se percataron de que, para poblaciones de UPM con medidas de resumen sesgadas, las varianzas relativas (coeficientes de variación) de la medida de resumen en cada estrato eran similares; es decir:

$$\frac{S_{x_1}}{\bar{X}_1} \cong \frac{S_{x_2}}{\bar{X}_2} \cong \cdots \cong \frac{S_{x_H}}{\bar{X}_H}$$

Gunning and Horgan (2004) tomaron esta evidencia en consideración y desarrollaron este método con el objetivo de que los coeficientes de variación de la medida de resumen tiendan a ser iguales dentro de los estratos y, de esta forma, encontraron que los límites que definían estos grupos estaban conformados en progresión geométrica. Siendo X la variable que contiene la información de la medida de resumen para todas la UPM del marco de muestreo, entonces los límites de los estratos estarán dados por la siguiente expresión:

$$b_h = \text{mín}(X) \left(\frac{\text{máx}\, X}{\text{mín}\, X} \right)^{h/L}; \qquad h = 1, 2, \ldots, H-1.$$

Es posible encontrar que los coeficientes de variación de los estratos conformados por estos límites son equivalentes y por ende, este método es óptimo para encontrar mejores formas de estratificar teniendo en cuenta como función objetivo la variación relativa dentro los estratos.

IV. Metodologías multivariadas sobre la matriz de información

Partiendo de la matriz de información $\mathbf{X}$ a nivel de las UPM, la cual contiene N_I filas y P columnas, es posible considerar algunos procedimientos que no necesitan de la reducción a una sola dimensión, sino que admiten tantas dimensiones como indicadores definidos en las columnas de $\mathbf{X}$. Teniendo en cuenta que en el periodo intercensal se realizarán encuestas que miden variables que están fuertemente ligadas a las observadas en el censo, entonces encontrar una estratificación que sea óptima para todo el conjunto de variables de la matriz de información asegurará una partición óptima para todas las encuestas realizadas en el periodo intercensal. Las siguientes metodologías permiten optimizar conjuntamente la eficiencia de la estratificación.

K-medias de Jarque (KmJ)

Jarque (1981) propuso utilizar una versión modificada del algoritmo de K-medias (Macqueen, 1967), cuyo objetivo es la minimización de la siguiente función de distancia:

$$\sum_{h=1}^{H}\sum_{k\in U_h}(\mathbf{x}_k-\bar{\mathbf{x}}_h)'\mathbf{\Lambda}^{-1}(\mathbf{x}_k-\bar{\mathbf{x}}_h)$$

En donde $\mathbf{x}_k$ corresponde a la medición de las P variables de la matriz de información en la k-ésima UPM, $\bar{\mathbf{x}}_h$ es el vector de medias de la matriz de información en el estrato h y $\mathbf{\Lambda}$ es una matriz diagonal de tamaño $P\times P$ cuyas entradas se definen como la varianza de las P variables de la matriz $\mathbf{X}$, es decir $\mathbf{\Lambda}[p,p]=S^2_{x_p}$, con $p=1,2,\ldots,P$. Esta modificación tiene como objetivo minimizar la relación entre la varianza de un estimador de muestreo estratificado con asignación proporcional y la de un muestreo aleatorio simple. Cuando $\mathbf{\Lambda}=\mathbf{I}$, el algoritmo resultante es idéntico al algoritmo clásico de K-medias, propuesto por Macqueen (1967).

Partición genética (BB)

Ballin and Barcaroli (2013) argumentan que la mejor estratificación es aquella partición del marco de muestreo que asegura el mínimo costo muestral que satisfaga algunas restricciones de precisión; o, que maximice la precisión de los indicadores de interés bajo las restricciones. De esta forma, el algoritmo busca minimizar la siguiente función de costos

$$c_0+\sum_{h=1}^{H}c_hn_h$$

En donde c_0 define un costo fijo y c_h es el costo promedio de observar un hogar en el estrato h. En principio, es posible definir $c_0=0$ y $c_1=c_2=\cdots=c_H=1$, lo cual da como resultado que el costo es el número de encuestas que deben realizarse en cada estrato. Este problema de optimización se complementa manteniendo las siguientes restricciones:

$$\sum_{h=1}^{H}\left(\frac{N_h^2}{n_h}\right)\left(1-\frac{n_h}{N_h}\right)S^2_{x_h,p}\leq V_{0p}\qquad p=1,2,\ldots,P.$$

En donde V_{0p} es un umbral predefinido por el usuario, que indica que la varianza de la estrategia estratificada está acotada; además, $S^2_{x_h,p}$ es la varianza poblacional de p-ésima variable de la matriz de información en el estrato h. Haciendo uso de algoritmos genéticos evolutivos, esta estratificación multivariada del marco de muestreo parte de la consideración de estratificaciones univariadas independientes (una para cada variable de la matriz de información) y de la definición del producto cartesiano resultante de todas estas particiones (estratos atómicos). Este universo de posibles estratificaciones evoluciona, uniendo grupos de forma jerárquica, sujeto a las restricciones de precisión sobre cada variable de la matriz de información, hasta converger en el número de estratos definidos de antemano H.

V. Evaluación y escogencia de la mejor estratificación

En la evaluación de los escenarios de estratificación entran las técnicas univariadas y multivariadas. Al final, el resultado de aplicar una u otra técnica es simplemente una clasificación de las UPM. Por lo tanto, cada una de las posibles estratificaciones debe ser evaluada con base en la reducción de la varianza para todos los indicadores considerados en la matriz de clasificación. La medida clásica con la que se juzgan las bondades de una estrategia de muestreo es el efecto de diseño (DEFF). Por lo tanto, la evaluación de la estratificación debe estar supeditada también a esta medida, que para la variable $p = 1, \ldots, P$, está dada por:

$$DEFF_p = \frac{Var_{ST}(\bar{x}_p)}{Var_{SI}(\bar{x}_p)} \qquad p = 1, \ldots, P.$$

En donde, $Var_{ST}(\bar{x}_p)$ y $Var_{SI}(\bar{x}_p)$ denotan la varianza del diseño estratificado y la varianza de un muestreo aleatorio simple para la media poblacional (porcentaje) de la p-ésima variable de la matriz de información. Por otro lado, Gutiérrez (2016, página 184) demuestra que, cuando la asignación es proporcional, esta relación se puede escribir de la siguiente manera:

$$DEFF_p = \frac{\sum_{h=1}^{H} W_h S^2_{x_{hp}}}{S^2_{x_p}} \cong 1 - R^2_p \qquad p = 1, \ldots, P.$$

En donde, para cada estrato $h = 1, \ldots, H$, se tiene que $S^2_{x_p}$ es la varianza de la variable x_p en la población y $S^2_{x_{hp}}$ es la varianza de la variable x_p supeditada al estrato h. Nótese que este efecto de diseño es función del coeficiente de determinación R^2_p en un modelo lineal con intercepto que relaciona la p-ésima variable de evaluación (respuesta) con los estratos (factores). Una ventaja de expresar el efecto de diseño como en la ecuación anterior es que no dependerá del tamaño de muestra. Una vez definido el criterio de evaluación de la estratificación sobre una variable x_p, es necesario definir un criterio de estratificación multivariante que contemple cada una de las P variables. Siguiendo las ideas de Jarque (1981), se propone la siguiente medida de calidad, definida como el *efecto de diseño generalizado* ($G(S)$) sobre todas las variables de la matriz de información:

$$G(S) = \sum_{p=1}^{P} DEFF_p = \sum_{p=1}^{P} \frac{1}{S^2_{x_p}} \sum_{h=1}^{H} W_h S^2_{x_{hp}}$$

Ante una estratificación pertinente, se esperaría que $Var_{ST}(\bar{x}_p) < Var_{SI}(\bar{x}_p)$, por lo tanto $0 < DEFF_p < 1$, lo que conlleva a que $0 < G(S) < P$. Luego, se debería escoger el escenario para el cual $G(S)$ fuera mínimo. Nótese que, para cada uno de los escenarios en estudio, es necesario fijar el número de estratos; en general se propende a que el número de estratos esté entre tres y cinco. Esta escogencia del número de grupos debe ser discutida al interior del INE con los equipos que determinan la rotación de las UPM en cada periodo de levantamiento de las encuestas de hogares. Escoger un número alto de estratos reducirá la varianza, pero a su vez puede tener repercusiones negativas en la logística de rotación del diseño de muestreo de las encuestas, haciendo que se agoten rápidamente las UPM dentro de los estratos geográficos y socioeconómicos. Por lo anterior, se recomienda restringir los escenarios de evaluación a la consideración de $H = 3$ y $H = 4$ estratos.

El siguiente cuadro ejemplifica la evaluación de estas técnicas para dos escenarios de estratificación (tres y cuatro estratos) en una matriz de información que contiene 8 variables. De la tabla se puede deducir varias conclusiones interesantes. Por ejemplo, para el primer indicador, la mejor estratificación es la dada por el método de raíz de frecuencia acumulada (DH) con cuatro estratos; para el segundo indicador, la mejor estratificación es la partición genética (BB) con cuatro estratos; mientras que para el último indicador, la mejor estratificación es la estratificación óptima con el algoritmo de Sethi (LH) con cuatro estratos. Como se puede notar, para cada indicador existirá un método que induzca una mayor eficiencia que para otros indicadores. Esto claramente muestra que la estratificación con respecto a un solo indicador puede ser un procedimiento inadecuado. Por lo tanto, basados en este ejemplo, el mejor método sería el de Dalenious-Hidiroglou (DH) con cuatro estratos, puesto que induce una mayor eficiencia conjunta al reducir el efecto de diseño generalizado.

Cuadro 5.1: *Efectos de diseño $DEFF_p$ y efecto de diseño generalizado $G(S)$ considerando tres ($H = 3$) y cuatro ($H = 4$) estratos para ocho variables.*

DEFF	Q (H=3)	DH (H=3)	LH (H=3)	GH (H=3)	KmJ (H=3)	BB (H=3)	Q (H=4)	DH (H=4)	LH (H=4)	GH (H=4)	KmJ (H=4)	BB (H=4)
$\bar{x}_1$	0.87	0.85	0.81	0.82	1	0.88	0.8	0.70	0.76	0.72	0.71	0.77
$\bar{x}_2$	0.89	0.82	0.95	0.97	0.94	0.88	0.79	0.74	0.75	0.77	0.75	0.71
$\bar{x}_3$	0.87	0.97	0.83	0.96	0.89	0.95	0.74	0.75	0.79	0.7	0.79	0.71
$\bar{x}_4$	0.92	0.89	0.81	0.94	0.96	1	0.77	0.73	0.73	0.7	0.71	0.74
$\bar{x}_5$	0.85	0.83	0.96	0.96	0.83	0.81	0.8	0.73	0.8	0.78	0.8	0.79
$\bar{x}_6$	0.87	0.88	0.9	0.88	0.86	0.81	0.8	0.72	0.76	0.7	0.74	0.73
$\bar{x}_7$	0.87	0.95	0.99	0.83	0.86	0.84	0.75	0.7	0.77	0.72	0.77	0.77
$\bar{x}_8$	0.93	0.82	0.91	0.99	0.93	0.88	0.77	0.74	0.72	0.78	0.76	0.75
G(S)	7.07	7.01	7.16	7.35	7.27	7.05	6.22	5.81	6.08	5.87	6.03	5.97

Para estudiar la comparabilidad y consistencia del proceso de estratificación, los algoritmos de evaluación se deberían aplicar sobre cada una de las UPM en las áreas urbanas, pero independientemente de las UPM rurales.Si la ganancia en eficiencia es mayor en este escenario, se pueden definir los estratos de forma independiente. Si, por el contrario, la comparabilidad entre estratos es imperante en el proceso de estratificación, se puede considerar únicamente el escenario conjunto en donde las UPM de la zona urbana y rural están presentes conjuntamente. En este último caso, la clasificación de las UPM de la zona urbana se regirá por las mimas condiciones que sus contrapartes urbanas.

Al margen de la técnica utilizada para encontrar la mejor clasificación de las UPM, se recalca que la viabilidad sobre el número de estratos sea discutida de forma exhaustiva por todas las áreas involucradas al interior de los INE. En forma general, es recomendable restringir los escenarios de evaluación a la consideración de H=3 o H=4 estratos. Este último componente es importante puesto que los diseños de muestreo deberían considerar un tamaño de muestra mínimo de dos UPM por estrato para poder estimar la varianza del estimador (Gutiérrez, 2016).

El efecto diseño no es el único aspecto por evaluar para la elección del procedimiento de estratificación. Es necesario verificar la estabilidad del método con respecto a los otros procedimientos de estratificación. Por ejemplo, la siguiente tabla muestra la matriz de coincidencias entre las diferentes clasificaciones de los estratos.

Cuadro 5.2: *Matriz de coincidencias, cuyas entradas están definidas como el porcentaje de UPM coincidentes en cada uno de los estratos creados por los métodos estudiados.*

Técnica	Jarque	K-means	DAL	GEO	LH-S	LH-K	Percentil
Q	1	0,64	0,92	0,84	0,89	0,89	0,82
DH	0,64	1	0,68	0,62	0,71	0,71	0,74
LH	0,92	0,68	1	0,82	0,96	0,96	0,90
GH	0,84	0,62	0,82	1	0,78	0,78	0,73
KmJ	0,89	0,71	0,96	0,78	1	1,00	0,93
BB	0,89	0,71	0,96	0,78	1,00	1	0,93

Por último, también se debe evaluar la coherencia de la distribución de las diferentes variables agregadas a nivel de UPM en los estratos. Por ejemplo, la proporción de personas mayores de 15 años alfabetizadas debería tener mayor incidencia en los estratos más altos, y este patrón también se debería observar para diferentes indicadores como la proporción de hogares con internet, la proporción de tenencia de refrigerador, la proporción de tenencia de televisión por cable, la proporción de tenencia de automóvil, la proporción de hogares con saneamiento adecuado, la proporción de hogares con pisos adecuados, la proporción de personas con educación superior, entre otras. La figura 5.2 muestra el comportamiento esperado en los estratos de muestreo para algunas variables de interés. De esta forma, el estrato uno debería presentar condiciones económicas más adversas, el estrato dos debería tener mejores condiciones, siendo el tercer estrato el que agrupa a las UPM con menores dificultades socioeconómicas. En el área rural debiesen aparecer una menor proporción de UPM en el estrato 3, dadas las condiciones menos favorables.

Si la contribución de algunas unidades al total poblacional es no significativa, y además esas unidades son de difícil acceso, es común que en algunos países de la región se opte por redefinir el universo y crear un estrato de exclusión forzosa. En este estrato no se realiza ninguna encuesta y las respectivas estimaciones no tendrán en cuenta a esta población excluida. Por último, como algunos procedimientos de clasificación se basan en la generación de números aleatorios, se recomienda documentar los códigos computacionales que se utilizaron para que los resultados puedan ser replicados, por lo que debe fijar una semilla aleatoria al comienzo del código computacional.

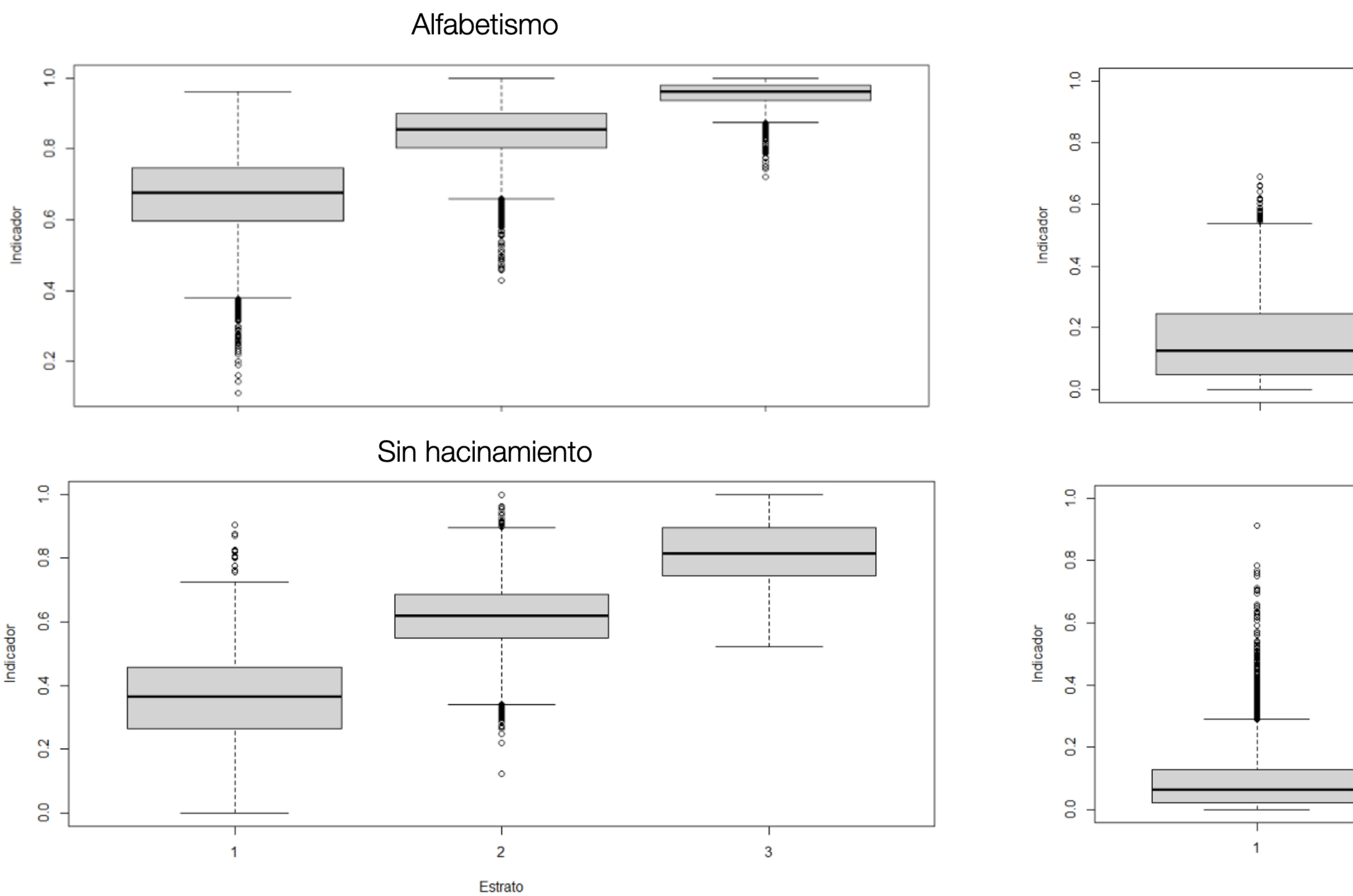

Figura 5.2: Comportamiento esperado en los estratos de muestreo para algunas variables de interés.

VI. Estratificación implícita

Los estratos explícitos definidos en la sección anterior son útiles para reducir la varianza de muestreo y asegurar la representatividad de la muestra en cada uno de los subgrupos que comparten las mismas características socioeconómicas, dentro de los mismos municipios. Además de los estratos socioeconómicos, algunas variables que se consideran en el proceso de estratificación explícita son:

- Estados o regiones de un país.
- Zona en la que está ubicado el hogar: urbana o rural. Nótese que cada país brinda su definición de ruralidad, de acuerdo con sus propias definiciones nacionales.

También es posible realizar una selección ordenada que induce una estratificación implícita, sin que necesariamente se tenga control sobre el tamaño de muestra final, y sin asumir independencia en la selección. Este tipo de estratificación es una forma de garantizar una asignación estrictamente proporcional de los hogares en todos los estratos implícitos. También puede conducir a una mayor confiabilidad de las estimaciones de la encuesta, siempre que las variables de estratificación implícita que se consideren estén correlacionadas con los indicadores de interés (por ejemplo, la tasa de desocupación, subocpupación o informalidad).

La estratificación implícita es altamente recomendada cuando la encuesta está enfocada en un tema particular (como por ejemplo el mercado de trabajo) y requiere el uso del muestreo sistemático (con probabilidades simples o desiguales) en la selección de las UPM. Según UN (2008a, pág. 46), en la mayoría de países la secuencia podría empezar con el área urbana, desagregada por departamento, a su vez desagregada por municipio; seguida del área rural, desagregada por departamento, a su vez desagregada por comuna o vereda. La selección sistemática de UPM deberá estar supeditada al ordenamiento de las UPM por el número de viviendas.

Nótese que la estratificación implícita puede constituir un método objetivo de selección de reemplazos de las UPM a las cuales no se pudo acceder en el operativo de campo; de esta forma, si una UPM fue seleccionada originalmente, pero por alguna razón operativa no puede ser empadronada, su reemplazo será la inmediatamente anterior (o posterior) en la lista estratificada implícitamente. Nótese que este procedimiento ubicará el reemplazo como la UPM ubicada en el mismo municipio, dentro del mismo departamento, en la misma zona y con un número similar de viviendas.

Aunque la estratificación implícita permite acotar el sesgo generado por la ausencia de respuesta de las UPM, Vehovar (1999, págs. 348 - 349) advierte que se debe tener precaución en cuanto a los usos de esta práctica puesto que puede conllevar sesgos importantes en las estimaciones de interés. Lo anterior se desprende del hecho de que los individuos ubicados en zonas donde sí es posible acceder puedan diferir significativamente de aquellos ubicados en las zonas de difícil acceso, las cuales difícilmente serán seleccionadas por los algoritmos de muestreo que hacen uso de la estratificación implícita.

Por esta razón es útil que, después de haber valorado los posibles sesgos, si se ha tomado la determinación de realizar las sustituciones sobre las UPM de difícil acceso, se realice un seguimiento exhaustivo en cada levantamiento que permita clasificar el esquema de recolección de información primaria y se valore su impacto en la precisión de los estimadores resultantes.

Capítulo 6

Diseño y mecanismo de selección de la muestra

Todas las encuestas de hogares en la región comparten el mismo principio inferencial: la selección de una muestra que puede representar la población de todo un país. Por supuesto, ante este objetivo tan ambicioso, es necesario contar con procedimientos robustos, probados y capaces de pasar los filtros más críticos y agudos. Tal vez en este momento de la historia, la práctica de estos procedimientos ya no genere ningún tipo de asombro, pero el lector podría animarse a contemplar todas los posibles escenarios que una sociedad enfrentaría ante la ausencia de las encuestas de hogares y sus repercusiones en materia del desarrollo social.

Es innegable la potencia y el poder que hay detrás de estas operaciones estadísticas que están sustentadas en el muestreo probabilístico que induce una inferencia que procede de lo particular a lo general, puesto que al seleccionar una muestra, esta sirve como base para obtener conclusiones acerca de la población. Al final la muestra será un vehículo adecuado para representar las características más importantes de la población en estudio, en la forma en que justamente las variables se incorporan en el formulario de la encuesta. Gutiérrez (2016) afirma que el muestreo es un procedimiento que responde a la necesidad de información estadística precisa sobre la población y los conjuntos de elementos que la conforman; asimismo comenta que el muestreo probabilístico trata con investigaciones parciales que apuntan a inferir a la población completa y en general está basado en los siguientes principios:

- *Aleatorización*: las unidades incluidas en la muestra son seleccionadas mediante un proceso probabilístico. De esta forma, además de eliminar los posibles sesgos de selección, la muestra resultante será válida para cualquier proceso de inferencia, puesto que se basa en el conjunto de todas las muestras que se pueden obtener con el esquema de muestreo definido.
- *Inclusión*: todas las unidades de la población tienen una probabilidad no nula de ser incluidas en la muestra. Lo anterior quiere decir que el procedimiento de selección le da chance de ser seleccionado a todas las unidades que componen la población. De esta manera, la muestra final puede estar compuesta por cualquier combinación plausible de hogares o individuos. Para que los anteriores principios se cumplan a cabalidad, es necesario contar con un instrumento que permita seleccionar a los hogares del país de forma exhaustiva y completa; esto quiere decir que el instrumento debería contener todos y cada uno de los hogares de la población. Dado que no existe una lista que permita identificar y ubicar a cada uno de los hogares de la población, entonces se deben contemplar otras posibilidades que permitan lograr el objetivo.

> Debido al principio natural de la aglomeración de las poblaciones humanas, es posible lograr este cometido de manera indirecta a través de la definición de los marcos de muestreo de áreas.

Las encuestas han tenido una gran trascendencia en la evolución de las mediciones de los indicadores sociales, que a su vez conllevan a que los gobiernos realicen un seguimiento y monitoreo de las cifras más importantes para la sociedad. De esta forma se podrá investigar la efectividad de las políticas públicas, para concretar las metas de mejora en las condiciones sociales y/o económicas de la ciudadanía. Tal como lo afirma Gutiérrez (2016), el muestreo es un procedimiento que responde a la necesidad de información estadística precisa sobre la población y los conjuntos de elementos que la conforman. De esta forma, una muestra bien seleccionada de unos cuantos miles de individuos puede representar con gran precisión a una población de millones de personas.

En general, se puede afirmar que un concepto apropiado por la sociedad es el que define a una muestra representativa como un modelo reducido de la población. De este concepto se desprende un argumento de validez sobre la muestra: "una buena muestra es aquella que se parece a la población, de tal forma que las categorías aparecen con las mismas proporciones que en la población". Sin embargo, en algunos casos es fundamental "sobrerepresentar" algunas categorías o incluso seleccionar unidades con probabilidades desiguales (Tillé, 2006a). La muestra no debe ser un modelo reducido de la población; debe ser una herramienta usada para obtener estimaciones válidas: exactas, confiables, precisas y consistentes.

El concepto de muestra representativa no se debe usar para referirse a que la muestra debe parecerse a la población. La teoría de muestreo se ha ocupado de estudiar estrategias óptimas que permitan asegurar la calidad de las estimaciones; en general, el concepto de representatividad debe estar asociado con la estrategia de muestreo y no sólo con la muestra seleccionada. Consecuentemente, la muestra como subconjunto de la población es una herramienta que no admite el calificativo de representativa, puesto que su objetivo no es parecerse a la población sino permitir que, mediante la correcta caracterización de una estrategia de muestreo, el proceso de inferencia logre reproducir la estructura de la población.

Lo anterior no indica que debamos abandonar del todo este adjetivo (representativo) en los proceso de muestreo. Por el contrario, el objetivo del equipo técnico experto en la selección de muestras debe estar supeditado a lograr que efectivamente este adjetivo se pueda aplicar a todo el componente de diseño y estimación. Es decir, el calificativo de representatividad es objeto de un proceso conjunto de diseño de muestreo, estimación de parámetros, acercamiento a modelos estadísticos para hacer frente a la ausencia de respuesta, entre otros. Uno de los objetivos de este capitulo será hacer precisión sobre las estructuras de selección de las muestras en las encuestas por muestreo. Al escoger un mecanismo apropiado para la selección de la muestra, será posible afirmar que la estrategia de muestreo es efectivamente representativa de la población de interés, puesto que cumple con altos estándares de rigurosidad y calidad en cada uno de los componentes del proceso.

I. Diseños de muestreo

Una vez que los marcos de muestreo se han refinado y se ha definido una estratificación apropiada para las UPM que las componen, es necesario realizar el proceso de muestreo para la selección final de los hogares. Este proceso de selección debe inducir insesgamiento, además de ser eficiente. Esto quiere decir que la inclusión de las unidades en la muestra estará supeditada a un esquema probabilístico libre de cualquier sesgo. Además de esto, se necesita que este mecanismo genere la menor dispersión posible en el proceso inferencial posterior.

El procedimiento de muestreo le asigna una probabilidad de selección conocida a cada posible muestra. Al diseñar un muestreo probabilístico, el investigador es el encargado de asignar estas probabilidades, mediante la definición del diseño de muestreo (Särndal et al., 2003). Aunque esta asignación de probabilidades se realiza de manera teórica, la pericia del equipo técnico deberá establecer cuál es la mejor forma de selección, y sobre esta escoger el mejor algoritmo de muestreo. Luego de establecer este conjunto de probabilidades, una única muestra es escogida mediante un mecanismo aleatorio que siga a cabalidad esta configuración estocástica inducida por el diseño de muestreo. Las probabilidades deben ser distintas de cero puesto que, de lo contrario, no se podría garantizar una inferencia insesgada, puesto que estaría excluyendo algunos sectores cartográficos del país. Además, estas mismas probabilidades se utilizan para crear los factores de expansión que definen todo el proceso de estimación, junto con el cálculo de los errores de muestreo, como se verá en los capítulos posteriores.

Existe una clara diferenciación entre un diseño de muestreo y un algoritmo de muestreo. El primero indica qué probabilidad de selección tendrán las posibles muestras en el soporte de muestreo, definido como el conjunto de todas las posibles muestras. Y el último se define como el proceso de selección de una única muestra que respeta las probabilidades del diseño de muestreo. En la definición de una encuesta de hogares es indispensable que se establezcan de antemano estos dos componentes. Es decir, si se ha decidido que el diseño de muestreo sea en etapas, el equipo técnico deberá documentar exhaustivamente cada etapa de muestreo, definiendo sus correspondientes unidades de muestreo y por consiguiente, los diseños de muestreos en cada etapa. Luego, es igual de importante explicar qué algoritmos de selección serán utilizados en cada etapa de muestreo. De esta forma habrá total transparencia en la selección de las unidades y esto redunda en la obtención de cifras oficiales confiables y precisas.

Existen muchas formas de seleccionar una muestra de hogares y cada una de ellas induce una medida de probabilidad sobre los elementos que conforman la población de interés. En general, asociado a cada esquema particular de muestreo se define una única función que asocia a cada hogar k con una probabilidad de inclusión en la muestra s, definida de la siguiente manera:

$$\pi_k = Pr(k \in s)$$

Si el diseño de muestreo es de tamaño fijo, estas probabilidades de inclusión de los hogares cumplirán con las siguientes propiedades

1. $\pi_k > 0$
2. $\sum_U \pi_k = n$

Observe que la primera propiedad garantiza que ningún hogar será excluido de la selección inicial. Si bien no todos lo hogares serán seleccionados para pertenecer a la muestra s, todos tendrán un chance de ser escogidos por el mecanismo de selección aleatorio. En segunda medida, el tamaño de la muestra de hogares estará inducido por la magnitud de las probabilidades de inclusión. Por esta razón, una encuesta con una tamaño de muestra grande asignará una mayor probabilidad de inclusión a todos los hogares, que una encuesta de tamaño de muestra más modesto. A continuación se presenta una lista no exhaustiva de diseños de muestreo utilizados en encuestas de hogares para la publicación de estadísticas oficiales, junto con la forma particular que toman las probabilidades de inclusión en cada esquema.

Muestreo aleatorio simple

Este diseño de muestreo supone que es posible realizar una enumeración de todas las posibles muestras de tamaño fijo y escoger una de ellas mediante una selección aleatoria que asigne la misma probabilidad a cada una. Para ejecutar este diseño de muestreo es necesario tener información suficiente y exhaustiva de la ubicación e identificación de todas las unidades de interés. Su uso es común en las etapas finales de selección de las encuestas, en donde los hogares o personas se seleccionan con las misma probabilidad. Por ejemplo, una vez se ha escogido un área de muestreo, una parte del operativo de campo deberá estar dedicada al enlistamiento de todas las viviendas en esa área seleccionada. Cuando se haya realizado este empadronamiento será posible asignarle la misma probabilidad de inclusión a cada vivienda en el área o en la UPM. Por ende, las probabilidades de inclusión en el muestreo aleatorio simple sin reemplazo son todas iguales y dadas por la siguiente expresión:

$$\pi_k = Pr(k \in s) = \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{k}} = \frac{n}{N}$$

Como se verá en los siguientes capítulos, cuando se usa el estimador de Horvitz-Thompson en este diseño de muestreo para estimar un total poblacional, y suponiendo que $S^2_{y_U}$ denota la varianza de la característica de interés en la población finita, entonces las expresiones del estimador puntual y su varianza, respectivamente, toman la siguiente forma:

$$\hat{t}_{y,\pi} = \sum_s \frac{y_k}{\pi_k}$$
$$Var(\hat{t}_{y,\pi}) = \frac{N^2}{n}\left(1 - \frac{n}{N}\right) S^2_{y_U}$$

Una variante de este tipo de esquemas de selección de muestras de hogares dentro de la UPM es el muestreo sistemático, en donde se ordena el marco con algún patrón predefinido y posteriormente se selecciona un primer hogar (como arranque aleatorio). A partir de ese primer hogar seleccionado, se incluyen los restantes hogares en la muestra mediante saltos sistemáticos equi-espaciados por el siguiente factor $a = \lfloor N/n \rfloor$, conocido como el intervalo de salto. Por ejemplo, una muestra sistemática podría ser:

$$s = \{2, 12, 22, 32, 42\}.$$

En donde el primer hogar elegido en la UPM fue el segundo y con saltos sistemáticos de diez hogares se va encuestando los restantes hogares en la lista. En este diseño la probabilidad de inclusión también es uniforme para cada hogar en la UPM y está dada por la siguiente expresión

$$\pi_k = Pr(k \in s) = \frac{1}{a} \approx \frac{n}{N}$$

Muestreo proporcional al tamaño

Este tipo de muestreo utiliza como insumo una característica de información auxiliar cuantitativa, también conocida como medida de tamaño (en inglés conocida como *measure of size*). Para la ejecución de este diseño, necesariamente el marco de muestreo deberá contener el valor correspondiente a la medida de tamaño para cada una de sus unidades. Este muestreo es utilizado con frecuencia en las etapas iniciales de selección de las muestras, particularmente en la selección de las UPM que harán parte de la muestra. De esta forma, los conglomerados o UPM con más hogares o personas (medida de tamaño) tendrán una mayor probabilidad de ser seleccionados en la muestra. Por consiguiente, las probabilidades de inclusión en la muestra para las UPM serán desiguales y proporcionales a la medida de tamaño. Observe que la cantidad de individuos en las UPM es una cifra conocida, puesto que son resultado directo de los censos de población y vivienda.

Una de las ventajas de este tipo de muestreos es que hace más eficiente la estimación de los indicadores de interés. Para que esto ocurra, la medida de tamaño debe estar linealmente relacionada con la característica de interés. Esto a menudo sucede en las problemáticas sociales indagadas en las encuestas de hogares; puesto que a mayor número de hogares, se observa una mayor incidencia de estos fenómenos. Por ejemplo, restringidos a un estrato particular, es evidente que en las UPM con mas hogares se observarán mayor número de personas pobres, o de hogares con ingresos bajos, o de personas desocupadas, etc.

Por último, la medida de tamaño no necesariamente tiene que estar definida como el conteo simple de hogares o personas dentro de las UPM, también puede definirse como una función de estos conteos; por ejemplo, la raíz cuadrada, o incluso como una función compuesta de conteos de subpoblaciones. En el caso más simple, si N_i es la medida de tamaño de la i-ésima UPM U_i, es decir el número de hogares que componen esa UPM; n_I el número de UPM que serán seleccionadas en cada estrato y N la sumatoria (o total) del número de hogares en todas las UPM del estrato (es decir, el número de hogares en el estrato) se tiene que las probabilidades de inclusión a la muestra s_I están dadas por la siguiente expresión:

$$\pi_i = Pr(U_i \in s_I) = n_I * \frac{N_i}{N}$$

Muestreo estratificado

Esta familia de diseños de muestreo permite realizar inferencias precisas en subgrupos poblacionales de interés, usualmente definidos como agregaciones geográficas grandes. Por ejemplo, si se quieren estimaciones de la incidencia de la pobreza en las regiones geográficas de un país específico, entonces es pertinente que esta división geográfica sea considerada para la definición de los estratos. Como se mencionó al inicio de este capítulo, estas divisiones territoriales se forman de manera natural, puesto que los estratos ya están definidos como regiones de interés en el seguimiento de los indicadores sociales. Por supuesto, es posible que la estrategia de muestreo cambie dependiendo de los estratos. Por ejemplo, en la planificación de las encuestas de uso de tiempo, una de las características de interés por las cuales se quiere indagar es la cantidad de horas que hombres y mujeres dedican a actividades de trabajo no remuneradas. Esta realidad cambia radicalmente entre zonas rurales y urbanas. Para este tipo de encuestas de hogares, la flexibilidad que tienen los diseños estratificados es un baluarte valioso que permite definir estrategias de muestreo más precisas.

Una consecuencia directa de la estratificación es que cada subgrupo tendrá un marco de muestreo de UPM independiente y mutuamente excluyente. Esta última caracterización induce una de las mayores ventajas del muestreo estratificado puesto que hay independencia entre los estratos. Esto significa que, al interior de cada estrato, se pueden ejecutar distintas estrategias de muestreo de forma independiente. Es común que en los países de América Latina el cruce de las áreas geográficas grandes junto con la división socioeconómica conformen los estratos (justo como se ilustró en los capítulos anteriores); asimismo una desagregación común en investigación social es la división territorial del país: urbano y rural. Evidentemente, la realidad social del entorno urbano difiere tanto del entorno rural que bien vale la pena considerar esta escisión en el diseño de muestreo de las encuestas de hogares.

Las probabilidades de inclusión definidas por este diseño de muestreo variarán en función de cada estrato h $(h = 1, \dots, H)$. Por ejemplo, si se hubiese planeado un diseño aleatorio simple en cada estrato, entonces las probabilidades de inclusión estarían dadas por la siguiente expresión

$$\pi_k = Pr(k \in s_h) = \frac{n_h}{N_h}$$

En donde s_h define la muestra seleccionada en el estrato h, N_h sería el número de hogares en ese estrato y n_h el tamaño de la muestra de hogares asociado a ese estrato.

En algunas ocasiones, se ha sugerido que el muestreo estratificado es el mejor diseño para una encuesta de hogares, lo cual es parcialmente cierto. Aunque en muchas ocasiones, la opción de estratificar es adecuada e inclusive conveniente, no es cierto estrictamente que el muestreo estratificado sea el mejor diseño de muestreo. De hecho, la varianza inducida por el diseño aleatorio estratificado puede llegar a ser más grande cuando no hay una clara homogeneidad en el comportamiento de la característica de interés dentro de los estratos.

Muestreo de conglomerados

Este diseño de muestreo surge como contraparte a la imposibilidad de generar una muestra de hogares directamente de un marco de muestreo que enliste todos y cada uno de los hogares en un país. De hecho, de forma hipotética, si fuese posible, los costos generados por una muestra aleatoria simple serían tan altos que la harían inviable desde el punto de vista presupuestario. Así, ante la ausencia de un marco de muestreo de las unidades de interés, y aprovechando el principio de aglomeración de las poblaciones humanas (que forman hogares y se aglomeran en segmentos, ciudades, regiones, etc.), la idea general detrás de este diseño es la conformación de unidades homogéneas entre sí (conglomerados), de las cuales se extraerá una muestra y para cada elemento del conglomerado se realizará un proceso exhaustivo de medición censal. De esta forma, es natural definir a las UPM como los conglomerados. Luego de seleccionar una muestra de estas UPM se realiza un censo de hogares dentro de cada una de las UPM seleccionadas. Nótese que este proceso logístico induce un esquema con ventajas económicas en términos presupuestales, puesto que limita el operativo de campo a un cierto número de UPM que se deben medir exhaustivamente.

A pesar de que esta estrategia resulte conveniente desde el punto de vista logístico y operativo, ciertamente no lo es desde el punto de vista de la eficiencia estadística; los errores de muestreo que se producen al utilizar esta metodología son bastante más elevados que en un diseño simple, puesto que al realizar el proceso de aglomeración, generalmente la variación interna de los conglomerados es muy baja y la variación entre conglomerados tiende a ser muy alta, generando mayor incertidumbre en la inferencia de la encuesta. Para superar estos inconvenientes, se podría pensar en un esquema de muestreo que aumente el tamaño de la muestra de conglomerados; sin embargo, este aumento puede llegar a ser tan grande que, en algunos estratos, se deberían seleccionar todas las UPM. Por supuesto, se trata de un esquema inviable en la práctica, pero que da paso al esquema de muestreo más común en las encuestas de hogares: la selección por etapas.

Muestreo en varias etapas

En este esquema de muestreo, la idea general es retomar los principios del muestreo de conglomerados y realizar un submuestreo de hogares dentro de los conglomerados o UPM seleccionadas inicialmente. En general, en América Latina son muy comunes los esquemas de selección en dos etapas: en la primera etapa se selecciona una muestra de UPM y en la segunda etapa se selecciona una muestra de hogares en aquellas UPM seleccionadas en la primera etapa. Aunque, también es posible encontrar en algunos países esquemas en más de dos etapas. Por ejemplo, en una primera etapa se seleccionan municipios; en una segunda etapa se seleccionan UPM dentro de los municipios seleccionados; y en la segunda etapa se selecciona una muestra de hogares en aquellas UPM seleccionadas en la segunda etapa. Si un municipio es incluido en la muestra es posible realizar un proceso de aglomeración continua sistemática, hasta llegar a la unidad de observación. Por ejemplo, en una ciudad seleccionada, es posible hacer un submuestreo de sus secciones cartográficas, luego seleccionar sectores cartográficos (contenidos en las secciones) y por último seleccionar hogares o personas.

Si el esquema de muestreo incluye la selección de municipios en la primera etapa, el diseño de muestreo apropiado en esta instancia deberá ser proporcional a una medida de tamaño, que puede ser definida como el número de habitantes de los municipios. De esta forma, con una probabilidad muy grande, a veces igual a uno, las ciudades más importantes (con más habitantes) serán siempre parte del estudio. Por otro lado, es posible que en algunas encuestas exista un submuetreo de personas dentro del hogar. En este caso, Clark and Steel (2007) aclaran que la escogencia de las personas dentro de los hogares no debería ser aleatoria simple puesto que ciertos grupos poblacionales podrían estar sub-representados o sobre-presentados. En general, el muestreo en varias etapas tiene dos características esenciales que lo hacen robusto, en términos estadísticos, y eficiente al momento de planear la logística del levantamiento de información; estas son:

- La independencia: que implica que no hay ninguna correlación en el diseño de muestreo de las unidades primarias de muestreo. Esto quiere decir que en cada UPM se puede ejecutar con independencia cualquier estrategia de muestreo que se crea apropiada para seleccionar la submuestra de hogares.
- La invarianza: que implica que sin importar qué diseño de muestreo se ejecutó en la primera etapa para seleccionar las UPM, la segunda etapa de selección podrá ejecutarse de manera independiente de la primera etapa. Es decir, el submuestreo de los hogares es independiente del muestreo de las UPM.

Un esquema de selección bastante usado en las encuestas de hogares de América Latina es el relacionado con los diseños auto-ponderados, lo cuales, en la primera etapa de muestreo se seleccionan n_I de N_I UPM con probabilidad proporcional al número de hogares N_i que la habitan; es decir:

$$\pi_i = Pr(U_i \in s_I) = n_I \frac{N_i}{N} \qquad i = 1, 2, \ldots, N_I.$$

En la segunda etapa de muestreo se seleccionan hogares dentro de las UPM que fueron incluidas en la etapa anterior. Esta selección de hogares se hace mediante un muestreo aleatorio simple, pero el tamaño de la submuestra es fijo para cada UPM. Es decir, no importa si una UPM es mucho más grande o más pequeña que las otras, el número de hogares que serán seleccionados será siempre el mismo. Por ejemplo, se podrían seleccionar $n_0 = 10$ hogares por UPM, siempre. De esta forma, en la segunda etapa, la probabilidad de que el k-ésimo hogar sea seleccionado en la submuestra s_i de la UPM U_i que fue seleccionada en la muestra de la primera etapa s_I, está dada por la siguiente expresión:

$$\pi_{k|i} = Pr(k \in s_i | U_i \in s_I) = \frac{n_0}{N_i}$$

En los esquemas auto-ponderados, a pesar de tener dos diseños de muestreo diferentes en dos etapas (proporcional al tamaño y aleatorio simple), la probabilidad de inclusión de los hogares es siempre la misma para todos los hogares, como se puede ver en la siguiente expresión:

$$\pi_k = \pi_{k|i} * \pi_i = \frac{n_0}{N_i}\frac{n_I * N_i}{N} = \frac{n_0 * n_I}{N} = \frac{n}{N}$$

Nótese que $n = n_0 * n_I$ corresponderá al número total de hogares que serán seleccionados, puesto que resulta ser la multiplicación del número de UPM que fueron seleccionadas en la primera etapa por el número de hogares que serán submuestreados en cada UPM en la segunda etapa. Este tipo de esquemas se utiliza cuando se quiere controlar el trabajo de campo y las cuotas por ciudad o municipio. Por otro lado, una particularidad de las encuestas de hogares es que, casi siempre, las personas y los hogares comparten las mismas probabilidades de inclusión. La razón de esto es que, en la mayoría de encuestas, el submuestreo de las personas es exhaustivo (censo en el hogar) y por ende, la probabilidad de inclusión en el submuestreo es forzosa.

$$\pi_k^{per} = Pr(persona \in hogar | hogar \in muestra) = 1$$

Por lo anterior, se tiene que la probabilidad de inclusión de las personas en la muestra es idéntica a la del hogar, puesto que

$$1 * \pi_{k|i} * \pi_i = 1 * \frac{n}{N} = \frac{n}{N}.$$

Muestreo en dos fases

En algunos casos en donde el marco de muestreo contiene poca o limitada información para proponer un diseño de muestreo eficiente, el investigador puede obtener información acerca de la población para construir un nuevo marco de muestreo reducido. En la primera fase, se selecciona una muestra de tamaño grande, conocida como *muestra maestra*. Para cada uno de los elementos en esa muestra se debe obtener información sobre una o más variables auxiliares, con el fin de estratificar de mejor manera, recolectar información auxiliar en la muestra, o simplemente para obtener muestras sucesivas y comparables a lo largo del ciclo de vida de la encuesta. En la segunda fase, con la ayuda de la información obtenida en la primera fase, se selecciona una submuestra mediante un diseño de muestreo conveniente, mucho más eficiente y apropiado para estimar el fenómeno en estudio.

Por ejemplo, si se requieren estimativos precisos para distintos subgrupos poblacionales, pero no existe un marco de muestreo confiable o actualizado, que permita diseñar un muestreo estratificado, entonces es necesario realizar un esquema de muestreo en dos fases. De esta forma, se selecciona una muestra aleatoria simple de tamaño moderado. Luego, se realiza un empadronamiento de los individuos en la muestra, a los cuales se les pregunta acerca de su membresía a los subgrupos poblacionales de interés. Luego, en una segunda fase, con ayuda de la información recolectada en la primera fase, se realiza un diseño estratificado.

Un ejemplo de este tipo de diseños de muestreo se da en el caso de México, en donde el INEGI ha planteado la construcción de una muestra maestra que permita seleccionar submuestras para las encuestas de hogares más importantes a la vez que se va recopilando información de los hogares pertenecientes a esta muestra maestra. En INEGI (2012), se menciona que «a partir de la construcción del Marco Maestro de Muestreo 2012, se diseñó la Muestra Maestra para lograr mantener actualizada de forma continua la información de las viviendas particulares dentro de esta muestra. El diseño de la muestra maestra consideró y respetó las UPM formadas y la estratificación con que fue construido el marco de muestreo por lo que heredó la mayoría de sus propiedades. El diseño de la Muestra Maestra está basado en la cobertura, tamaño y distribución de las encuestas continuas y periódicas del INEGI. Los tamaños de muestra en viviendas para estas encuestas junto con el promedio óptimo de viviendas a seleccionar dentro de una UPM determinaron el número de UPM a seleccionar para la Muestra Maestra 2012». De esta forma, la muestra maestra constituye un elemento esencial para el levantamiento de la Encuesta Nacional de Ocupación y Empleo, la Encuesta Nacional sobre la Confianza del Consumidor, la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública, la Encuesta Nacional de Gasto de los Hogares, entre algunas otras.

En el caso de Costa Rica, la muestra de la Encuesta Nacional de Microempresas de los Hogares sigue un diseño en dos fases. La primera fase toma como base la Encuesta Nacional de hogares, en la cual se identifican aquellos hogares cuyos integrantes desarrollan actividades económicas concernientes con emprendimientos y microempresas. A partir de este listado exhaustivo, en una segunda fase, se selecciona a todas las personas al frente de estas microempresas y se les aplica un cuestionario con el fin de obtener información sobre sus características y sus actividades económicas.

Por otro lado, en Chile se realiza el Estudio Nacional de la Discapacidad que asume un marco de muestreo reducido, en una primera fase, basado en la encuesta de hogares CASEN, en la cual se identifican los hogares que tienen miembros con alguna condición de discapacidad. En una segunda fase, se realiza una selección de hogares y mediante un cuestionario estructurado se indagan las características de las personas con esta condición.

Muestreo balanceado

El método del cubo (Tillé, 2006a) permite seleccionar muestras balanceadas, manteniendo las proporciones de la población original en la muestra en diferentes variables de equilibrio, las cuales se espera que estén correlacionadas con las variables de interés. En general, el método del cubo permite la selección de una muestra aleatoria para la cual el inverso de las probabilidades de inclusión reproduce de forma exacta el total poblacional de las variables de balanceo.

Gutiérrez (2016) afirma que este es un procedimiento general y riguroso que permite la extracción de muestras probabilísticas balanceadas y la posterior estimación de las cantidades de interés, enmarcados bajo métodos de inferencia basados en el diseño de muestreo. Dado que bajo un diseño de muestreo balanceado el estimador para los totales de un conjunto de variables auxiliares, debe ser igual al total poblacional de las mismas, entonces la varianza del estimador del total poblacional de la característica de interés se debe reducir de acuerdo con el aumento de su correlación con las variables auxiliares.

El método del cubo se compone de dos fases: la fase de vuelo y la fase de aterrizaje. En la primera, para que las restricciones sean satisfechas exactamente, se deben redondear a cero (0) o uno (1) las probabilidades de inclusión. La fase de aterrizaje consiste en el manejo adecuado del redondeo apelando a la programación lineal. Por ejemplo, aplicando el método simplex sujeto a una función de costo relacionada con la varianza del estimador.

En las encuestas de hogares es posible utilizar el algoritmo de selección del método del cubo en cada uno de los estratos conformados en el diseño de muestreo para seleccionar UPM. El método del cubo es un algoritmo de selección, que a diferencia de los algoritmos de selección tradicionales permite reproducir de forma exacta el número total de personas por grupos de edad y sexo a nivel de la UPM para este caso concreto. En Perú, la Encuesta Demográfica y de Salud Familiar utiliza este tipo de muestreo en la selección de las UPM. De esta manera, como variables de balanceo se podrían definir las siguientes:

- Una columna de unos para que exista balanceo en el número de UPM.
- El vector de probabilidades de inclusión iniciales.
- Total de personas por grupos de edad y sexo (a partir de la información de los censos de población y vivienda).

Si la encuesta se realiza de forma periódica, es necesario actualizar los marcos de muestreo y los tamaños poblacionales a través de tiempo. Llegado el caso, el investigador puede apoyarse en las proyecciones demográficas (nacimientos esperados, muertes esperadas y población proyectada) disponibles en fuentes oficiales como totales auxiliares.

II. El diseño de muestreo estándar en una encuesta de hogares

A continuación se describe de manera genérica cómo es un diseño de muestreo típico de una encuesta de hogares en la región. Por supuesto, en la práctica existen variantes que se pueden alejar un poco de esta generalización pero que, en general, mantienen la misma estructura. La mayoría de encuestas son de naturaleza multipropósito. Esto quiere decir que existen múltiples variables de interés. Por lo anterior, el investigador debe definir las variables más importantes de la investigación y sobre éstas planear el diseño de muestreo. Esta directriz implica que para obtener simultáneamente la precisión requerida en todas las estimaciones, el tamaño de muestra será un poco más exigente. Asimismo, la definición de los dominios de representatividad debe estar directamente determinada por los objetivos de la encuesta y por las unidades de muestreo.

Se debe mencionar también que el diseño de muestreo de muchas de las encuestas de hogares que se realizan actualmente en la región mantienen el mismo espíritu de los diseños que anteriormente sirvieron para levantar la información primaria. Es decir, el nivel de innovación en este campo no se da de forma intempestiva, y más bien se podría afirmar que cada vez que se rediseña una encuesta de hogares, el punto de partida será el diseño anterior de la encuesta, lo cual es oportuno si se quiere mantener la comparabilidad de las cifras entre los levantamientos periódicos. Siempre que no haya un marco de muestreo de elementos, es posible utilizar los principios del muestreo en varias etapas, mediante la selección de diferentes unidades de muestreo que contienen a los elementos de interés. Por consiguiente, el diseño de muestreo de un encuesta de hogares es generalmente probabilístico estratificado y bietápico:

- Se realiza una estratificación por zona: urbano/rural, por región y por los estratos socioeconómicos definidos en los capítulos anteriores.
- De forma independiente, dentro de cada estrato se realiza un muestreo bietápico.
  - En la primera etapa, se seleccionan áreas cartográficas, conocidas como unidades primarias de muestreo (UPM) siguiendo un diseño de muestreo proporcional al número de viviendas, hogares o personas del conglomerado.
  - En la segunda etapa, se escoge aleatoriamente un número fijo de hogares dentro de cada UPM siguiendo un diseño de muestreo aleatorio simple.

Este tipo de esquemas tienen una consecuencia importante en cuanto a la eficiencia estadística. Nótese que, en la segunda etapa de muestreo, la variación que se pueda presentar entre los hogares seleccionados en una misma UPM es muy baja con respecto a la variación que se puede presentar entre diferentes UPM. Por el principio de representatividad, las personas se aglomeran de manera natural y forman conglomerados homogéneos. Es decir, dentro de una misma UPM, los hogares tendrán características sociales bastante similares. En particular, estos hogares tendrán similares realidades en cuanto a su ingreso, gasto, desocupación, analfabetismo, educación, etc.

Adicionalmente, no es de esperarse encontrar un hogar con altos niveles de ingreso y gasto, cuyos integrantes tienen un nivel de educación muy alto, habitando una vivienda que se encuentre en un sector marginal o deprimido de la ciudad, en donde el acceso al alcantarillado es precario, y con deficiencias en los servicios de electricidad o agua potable; aunque podría suceder, no es lo que se esperaría. De la misma forma, no es de esperar que un hogar pobre, cuyo ingreso per cápita es bastante bajo y no alcanza para cubrir las necesidades básicas de sus habitantes, ocupe una vivienda ubicada en un sector acaudalado.

Por lo tanto, en este tipo de investigaciones sociales, la varianza existente entre los conglomerados es inmensa al compararla con la variación dentro de los conglomerados. Por esta razón, es de esperarse que existan diferencias significativas entre las UPM que componen la muestra, puesto que la realidad de una UPM en un sector deprimido no es la misma que la de una UPM en un sector opulento. Este es un reflejo de las desigualdades propias de América Latina, las cuales han ocupado la agenda política y legislativa de las últimas décadas. Retomaremos esta particularidad en los posteriores capítulos, cuando se aborde el tema de la eficiencia estadística y la medición del error de muestreo.

A continuación, se definirán todos los elementos involucrados en la selección de una muestra de hogares. En general, los diseños de muestreo de las encuestas de hogares estimarán el total de cada UPM t_i mediante una sub-muestra seleccionada desde el marco de muestreo compuesto por los sectores cartográficos definidos en el último censo. Suponga que la población de hogares U se divide en N_I UPM, que definen una partición de la población, llamados también conglomerados y denotadas como $U_I = \{U_1, \ldots, U_{N_I}\}$ (U_I es la población de todas las UPM en un país y N_I es el número total de UPM dentro del país). Nótese que la i-ésima UPM U_i $i = 1, \ldots, N_I$ contiene N_i hogares. Luego, el proceso de selección se surte de la siguiente manera:

- Una muestra s_I de UPM es seleccionada de U_I de acuerdo a un diseño de muestreo $p_I(s_I)$. El tamaño de la muestra de UPM se denota como n_I. Nótese que s_I representa la muestra aleatoria de UPM que fue seleccionada de acuerdo a la medida de probabilidad $p_I(s_I)$.

- Para cada UPM U_i $i = 1, \dots, n_I$ en la muestra seleccionada s_I, se realiza de forma independiente un submuestreo de hogares, de tal forma que en cada UPM existirá una muestra s_i de hogares de acuerdo a un diseño de muestreo $p_i(s_i)$. Nótese que s_i representa la muestra aleatoria de hogares que fue seleccionada en la segunda etapa de acuerdo a la medida de probabilidad $p_i(s_i)$.

Por lo tanto, en la primera etapa se ha identificado todos los sectores cartográficos de país y se ha generado el marco de muestreo de las UPM que se separan en grupos mutuamente excluyentes, según las variables de estratificación explícita previamente definidas; dentro de cada estrato se selecciona la muestra de UPM en donde la probabilidad que tiene cada UPM de pertenecer a la muestra está determinada por el número de personas o viviendas (medida de tamaño). En esta etapa es importante tener en cuenta que se seleccionará un número mayor de UPM en los estratos más grandes; evidentemente las regiones con más habitantes tendrán una muestra de UPM más grande, aunque esta relación no siempre es lineal. se recomienda que el diseño de muestreo debe ser tan simple como sea posible[1].

A pesar de que la medida de tamaño permite que las UPM con mayor cantidad de hogares tengan una mayor probabilidad de ser escogidas, esta diferencia en las probabilidades de selección se compensa en la segunda etapa de muestreo, debido a que cada hogar tendrá igual probabilidad de ser elegido en la muestra dentro del estrato. Es pertinente observar que, para la segunda etapa se requiere contar con un listado exhaustivo de todos los hogares dentro de todas las UPM seleccionadas. Este proceso de selección requerirá de un empadronamiento previo que, no solo actualice el número de hogares, sino que permita identificarlos y ubicarlos dentro de la UPM. De esta manera, y de forma aleatoria simple, se elige una muestra de hogares y su tamaño no varía entre UPM.

[1]Nótese que los esquemas de estimación se van volviendo más complejos a medida que el diseño de muestra agrega más etapas o más fases.

Capítulo 7

El efecto de diseño

Cuando se selecciona una muestra utilizando un diseño de muestreo complejo es muy improbable que exista independencia entre las observaciones. Además, como el muestreo de las encuestas de hogares es complejo, la distribución de la variable de interés no es la misma para todos los individuos. Por lo anterior, cuando se analizan datos que provienen de encuestas de hogares la inferencia correcta debe tener en cuenta estas grandes desviaciones con respecto al análisis estadístico clásico, que considera muestras aleatorias simples. Por ello, en la mayoría de ocasiones se necesita aumentar el tamaño de muestra para obtener la precisión deseada.

El efecto de diseño fue definido por Kish (1965, página 258) como *la relación entre la varianza real de una muestra y la varianza real de una muestra aleatoria simple del mismo número de elementos* y toma la siguiente expresión:

$$DEFF = \frac{Var(\hat{\theta})}{Var_{MAS}(\hat{\theta})}$$

En donde $Var(\hat{\theta})$ denota la varianza de un estimador $\hat{\theta}$ bajo un diseño de muestreo complejo $p(s)$ y $Var_{MAS}(\hat{\theta})$ denota la varianza del este estimador $\hat{\theta}$ bajo un diseño de muestreo aleatorio simple MAS. Esta cifra da cuenta del efecto de aglomeración causado por la utilización de un diseño de muestreo complejo (p), frente a un diseño de muestreo aleatorio simple MAS, en la inferencia de un parámetro de la población finita θ (que puede ser un total, un promedio, una proporción, una razón, un percentil, etc.).

I. Estimación del efecto de diseño

En la expresión de efecto de diseño se debe notar dos hechos importantes, en primer lugar, DEFF depende del diseño muestral $p(s)$, y en segundo lugar, depende del estimador del parámetro θ. De esta forma, no es correcto describir al $DEFF$ únicamente como una medida de eficiencia del diseño muestral, puesto que bajo un mismo diseño, este puede tomar diferentes valores según el parámetro que se quiera estimar.

Nótese que ninguno de los componentes del efecto de diseño se conoce y por ende deben ser estimados. En particular, un estimador aproximadamente insesgado de la varianza poblacional $S^2_{y_U}$

es la varianza muestral ponderada, la cual está dada por la siguiente expresión:

$$\hat{S}^2_{y_U} = \left(\frac{n}{n-1}\right)\frac{\sum_s w_k(y_k - \hat{\theta})^2}{\sum_s w_k - 1}$$

De esta forma, en el caso en el que θ corresponda a un promedio poblacional, una estimación de la varianza $Var_{MAS}(\hat{\theta})$ bajo muestreo aleatorio simple está dada por la siguiente expresión:

$$\widehat{Var}_{MAS}(\hat{\theta}) = \frac{1}{n}\left(1 - \frac{n}{\hat{N}}\right)\hat{S}^2_{y_U}$$

En donde $\hat{N} = \sum_s w_k$. Por lo tanto, la estimación del efecto de diseño DEFF está dada por

$$\widehat{DEFF} = \frac{\widehat{Var}(\hat{\theta})}{\widehat{Var}_{MAS}(\hat{\theta})}$$

La idea central del efecto de diseño recae en la evaluación del mismo estimador bajo diferentes escenarios de muestreo. Como el estimador que se está estudiando $\hat{\theta}$ viene ponderado por los factores de expansión de la encuesta, entonces lo más conveniente es utilizar el mismo rasero para evaluar ambas estrategias de muestreo. Es posible encontrar una discusión más profunda sobre el efecto de diseño en Gambino (2009, sección 4.), Särndal et al. (2003, página 188) y Gutiérrez et al. (2016, página 101).

II. Descomposición del efecto de diseño en las encuestas de hogares

Park (2003) propone que el efecto de diseño de cualquier encuesta se puede descomponerse en tres partes que se relacionan entre sí de forma multiplicativa. En primer lugar está el efecto debido a la ponderación desigual, $DEFF^W$; en segundo lugar se encuentra el efecto debido a la estratificación, $DEFF^S$; y por último se tiene el efecto debido al muestreo en varias etapas, $DEFF^C$. Por lo tanto:

$$DEFF = DEFF^W \times DEFF^S \times DEFF^C$$

La primera componente $DEFF^W$ del efecto de diseño general tiende a aumentar ligeramente la variación de las estrategias de muestreo. Valliant et al. (2018) afirman que esta componente puede ser estimada por medio de la siguiente expresión:

$$DEFF^W = 1 + cv^2(w_k)$$

En donde $cv(w_k)$ representa el coeficiente de variación de los pesos de muestreo w_k de las unidades en la encuesta. Si los pesos de muestreo son uniformes, entonces no habrá un incremento significativo en la varianza de la estrategia. Es por esto que los esquemas autoponderados son deseables en los diseños de muestreo de las encuestas de hogares. Por otra parte, si los pesos de muestreo tienen una variación grande, entonces habrá un incremento significativo en la varianza y, por ende, en el tamaño de muestra. Como se verá más adelante, los ajustes en el factor de expansión pueden inducir una alta variabilidad y por consiguiente se recomienda, en la medida de lo posible, crear clases o subgrupos de ajuste para mitigar y acotar la dispersión de los pesos finales de la encuesta.

Al encontrar la mejor estratificación, nos aseguramos de que la segunda componente $DEFF^S$ de esta descomposición sea menor a uno (es decir que la varianza se reduce). Lamentablemente, la reducción de la varianza no suele ser tan grande y no mitiga los efectos de aglomeración debido a las múltiples etapas de los diseños de muestreo complejos. Como lo indica Gutiérrez (2016), el efecto de diseño en el muestreo aleatorio estratificado sin reemplazo con asignación proporcional está dado por

$$DEFF^S \cong \frac{\text{Varianza dentro de los estratos}}{\text{Varianza Total}}$$

Ahora, intuitivamente tenemos que la varianza total es la suma de la varianza dentro de los estratos con la varianza entre los estratos. Por tanto se concluye que, casi siempre, esta estrategia de muestreo arrojará mejores resultados que una estrategia aleatoria simple. Por otro lado, recordando que el efecto de diseño debido a la conglomeración de la población finita en las UPM está dado por la siguiente expresión:

$$DEFF^C = 1 + (\bar{n}_{II} - 1)\rho_y$$

En donde, $\bar{n}_{II}$ es el número de hogares promedio que se seleccionan en cada UPM, y ρ_y es el coeficiente de correlación intraclase, calculado para la variable de interés sobre las UPM. Habiéndose definido el marco de muestreo en el momento del levantamiento de la información primaria censal, ya no se tendrá control sobre el valor del coeficiente de correlación intraclase (ρ_y); únicamente se tiene control sobre el número de viviendas que serán seleccionadas en promedio en las UPM ($\bar{n}_{II}$). Si el marco de muestreo quedó correctamente definido, entonces el valor de ρ_y será tan pequeño como fue posible establecerlo al proponer las UPM; de la misma manera, es recomendable que el equipo técnico dentro de los INE defina el menor número promedio posible de encuestas dentro de las UPM $\bar{n}_{II}$ para que el efecto de aglomeración sea mínimo.

En general, la disminución del $DEFF$ debido a la estratificación se matiza con el aumento del $DEFF$ debido a la desigualdad de los pesos de muestreo. Es por esto que $DEFF^C$ predomina en el efecto de diseño general y es la razón por la cual se le presta mucha atención. UN (2008b) propone que, para mitigar los efectos del muestreo multietápico, se consideren las siguientes estrategias:

1. Seleccionar tantas UPM como sea posible.
2. Definir las UPM tan pequeñas como sea posible, en términos del número de viviendas que las componen.
3. Seleccionar un número fijo de viviendas dentro de las UPM seleccionadas, en vez de un número variable.
4. Utilizar un muestreo sistemático en la UPM, en vez de seleccionar segmentos de viviendas contiguas.

Al encontrar la mejor estratificación, los funcionarios de los INE permiten que la segunda componente $DEFF^S$ de la descomposición del efecto de diseño general sea mínima para los indicadores estudiados. También es tarea de los INE asegurar que los efectos de diseño dados por el efecto de conglomeración y el uso del muestreo en varias etapas $DEFF^C$ sea mínimo. En este caso, se deberá estudiar, para cada encuesta y operación estadística que haga uso del marco de muestreo estratificado, la relación entre UPM y hogares a la luz de los indicadores de interés; en particular, es necesario decidir cuántos hogares serán seleccionados en cada UPM y cuántas UPM serán seleccionadas dentro de cada estrato

De la mima manera, y como se verá en los siguientes capítulos, el efecto debido al uso de factores de ponderación desiguales $DEFF^W$ puede ser minimizado al decidir, a la luz de la correlación entre los indicadores particulares de cada encuesta de hogares, cuáles variables de control serán utilizadas en la calibración de los estimadores. De esta forma, en esta estrategia tripartita, se asegura que el efecto de diseño general de las encuestas sea pequeño.

III. Formas comunes del efecto de diseño

Suponiendo que el parámetro de interés es la media poblacional ($\bar{y}$) de una variable de interés y (por ejemplo, el ingreso per cápita mensual), es posible escribir la varianza del estimador bajo el diseño de muestreo complejo como

$$Var(\hat{\bar{y}}) = \frac{DEFF}{n}\left(1 - \frac{n}{N}\right) S^2_{y_U}$$

En donde $S^2_{y_U}$ corresponde a la varianza poblacional de la características de interés, N es el tamaño de la población de interés U y n el tamaño de la muestra de individuos. Por otro lado, suponiendo que el parámetro de interés es la proporción poblacional (P) de una variable dicotómica y (por ejemplo, el porcentaje de individuos de bajo de la línea de pobreza en un país), es posible escribir la varianza del estimador bajo el diseño de muestreo complejo como

$$Var(\hat{P}) = \frac{DEFF}{n}\left(1 - \frac{n}{N}\right) P(1-P)$$

Cuando se trata de un diseño muestral multietápico, por ejemplo, es común seleccionar UPM en la primera etapa y posteriormente seleccionar hogares dentro de las áreas seleccionadas. En este contexto, el coeficiente de correlación intraclase está definido por

$$\rho_y = 1 - \frac{N_I}{N_I - 1}\frac{SCD}{SCT}$$

En donde, apelando a la notación clásica de los análisis de varianza, $SCT = \sum_U (y_k - \bar{y}_U)^2$ hace referencia a la suma de cuadrados total, $SCE = \sum_{U_I} N_I(\bar{y}_{U_I} - \bar{y}_U)^2$ es la suma de cuadrados entre, y $SCD = SCT - SCE$ es la suma de cuadrados entre. Cuando la característica de interés y es heterogénea entre los conglomerados, pero los conglomerados son homogéneos entre sí, entonces ρ_y es cercano a 0; mientras que si los conglomerados son heterogéneos entre sí, pero homogéneos dentro de cada uno, entonces ρ_y es cercano a 1. En este tipo de escenarios, el efecto de diseño se puede expresar como $DEFF = 1 + (\bar{n}_{II} - 1)\rho_y$. En general, nótese que el efecto de diseño será mayor cuando:

1. El coeficiente de correlación crezca, lo cual no puede ser controlado de antemano, puesto que se trata de la observación de la realidad. En general, ρ_y será más grande cuando la

distribución de la variable de interés sea explicada por las UPM en el país. Por ejemplo, si el indicador de interés es la pobreza y los hogares pobres están aglomerados, segregados y separados de los hogares más acaudalados, entonces ρ_y será más grande; además, entre más segregación haya, mayor será su valor.

2. El promedio de hogares seleccionados por UPM ascienda. Esto es controlado de antemano en la etapa de diseño y será un número fijo y transversal en la encuesta.

IV. Otras consideraciones

El efecto de diseño en subpoblaciones

La estimación del efecto de diseño es un problema común cuando se trabaja con estimaciones desagregadas en subpoblaciones de interés. Por un lado, cuando las subpoblaciones constituyen estratos (o agregaciones de estratos) planeados de antemano, para los cuales se conoce previamente su tamaño poblacional, se tiene el siguiente efecto de diseño:

$$DEFF_h = \frac{Var(\hat{\theta}_h)}{Var^h_{MAS}(\hat{\theta}_h)}$$

En donde $Var^h_{MAS}(\hat{\theta}_h)$ es la varianza del estimador restringida al estrato $h(h = 1, \ldots, H)$; en el caso en el que $\hat{\theta}_h$ corresponda al estimador del promedio poblacional en el estrato h, su valor es el siguiente:

$$Var^h_{MAS}(\hat{\theta}_h) = \frac{1}{n_h}\left(1 - \frac{n_h}{N_h}\right) S^2_{y_{U_h}}$$

Siendo n_h el tamaño de la muestra en el estrato h, N_h el tamaño poblacional del estrato h y $S^2_{y_{U_h}}$ es la varianza poblacional de la variable de interés restringida al subgrupo h. Por lo tanto, los efectos de diseño para las medias muestrales en un diseño aleatorio estratificado, serán por definición iguales a uno.

Por otro lado, cuando la subpoblación de interés no es un estrato o un post-estrato sino un subgrupo aleatorio (como por ejemplo las personas pobres, las personas ocupadas, o cualquier otro subgrupo no planeado en el diseño de la encuesta o en la etapa de calibración), en adelante notado con la letra g, cuyo tamaño de muestra no es fijo (o condicionalmente fijo por la calibración) sino aleatorio, entonces la estimación correcta del efecto de diseño es la siguiente:

$$DEFF_g = \frac{Var(\hat{\theta}_g)}{Var^U_{MAS}(\hat{\theta}_g)}$$

En donde $Var^U_{MAS}(\hat{\theta}_g)$ es la varianza del estimador de interés. En el caso en el que $\hat{\theta}_g$ corresponda al estimador del promedio poblacional en el subgrupo g, entonces su varianza estaría dada por la siguiente expresión:

$$Var^U_{MAS}(\hat{\theta}_g) = \frac{1}{n}\left(1 - \frac{n}{N}\right) S^2_{y_{gU}}$$

En donde $S^2_{y_{gU}}$ es la varianza poblacional de una nueva variable calculada en toda la población, que toma el valor de y_k, cuando la unidad k pertence al subgrupo g, y toma el valor de cero, en cualquier otro caso. Por lo tanto, en ambos efectos de diseño, la estimación de la varianza del diseño de muestreo complejo $Var(\hat{\theta}_h)$ o $Var(\hat{\theta}_g)$ es la misma, pero el denominador cambia dependiendo de si el subgrupo es un estrato o no. Es por esta razón que, al analizar una encuesta de hogares, hay coincidencia en las cifras relacionadas con la estimación puntual, errores estándar, intervalos de confianza y coeficientes de variación entre los diferentes softwares computacionales. Sin embargo, es necesario percatarse de las opciones que estos adjuntan para calcular correctamente la cifra apropiada. Nótese que, en resumen, las estimaciones de $Var^U_{MAS}(\hat{\theta}_g)$ y $Var^h_{MAS}(\hat{\theta}_h)$ serán diferentes, puesto que la primera involucra a toda la muestra, mientras que la segunda involucra únicamente a la muestra del estrato.

Lumley (2010) afirma que el efecto del diseño compara la varianza de una media o total con la varianza de un estudio del mismo tamaño utilizando un muestreo aleatorio simple sin reemplazo y que su cálculo será incorrecto si los pesos de muestreo se han re-escalado o no son recíprocos a las probabilidades de inclusión. Por ejemplo, en el caso de las subpoblaciones, la librería `survey` de `R` compara la varianza de la estimación con la varianza de una estimación basada en una muestra aleatoria simple del mismo tamaño que el de la subpoblación. Entonces, por ejemplo, en el muestreo aleatorio estratificado, el efecto de diseño calculado en un estrato será igual a uno.

El efecto de diseño general

Suponga que el diseño muestral es estratificado con H estratos; entonces por la independencia de la selección en los estratos, la varianza del estimador de un total poblacional t_y está dada por

$$Var\left(\widehat{t_{y,\pi}}\right) = \sum_{h=1}^{H} Var_h\left(\widehat{t_{y,\pi}}\right)$$

donde

$$Var_h\left(\widehat{t_{y,\pi}}\right) = DEFF_h \times Var_{MAS,h}\left(\widehat{t_{y,\pi}}\right)$$

Por otro lado,

$$Var\left(\widehat{t_{y,\pi}}\right) = DEFF \times Var_{MAS}\left(\widehat{t_{y,\pi}}\right)$$

De esta forma, se tiene que

$$DEFF = \frac{\sum_{h=1}^{H} DEFF_h Var_{MAS,h}\left(\widehat{t_{y,\pi}}\right)}{Var_{MAS}\left(\widehat{t_{y,\pi}}\right)} = \frac{\sum_{h=1}^{H} DEFF_h \frac{N_h^2}{n_h}(1-\frac{n_h}{N_h})S^2_{y,U_h}}{\frac{N^2}{n}(1-\frac{n}{N})S^2_{y,U}}$$

Es decir, el efecto de diseño puede ser visto como una combinación lineal de los efectos de diseño de los H estratos ($DEFF = \sum_{h=1}^{H} DEFF_h\ w_h$). En donde el peso w_h está dado por

$$w_h = \frac{\frac{N_h^2}{n_h}(1-\frac{n_h}{N_h})S^2_{y,U_h}}{\frac{N^2}{n}(1-\frac{n}{N})S^2_{y,U}}$$

Es claro que los pesos w_h son todos positivos, pero no necesariamente son menores a 1, y la suma de ellos tampoco es igual a 1. A continuación, examinamos la forma de w_h en el caso especial de muestras autoponderadas en todos los estratos. En este caso, se tiene que $\frac{n_h}{N_h} = \frac{n}{N}$ para todo $h = 1, \ldots, H$, y se puede ver que

$$w_h = \frac{N_h S^2_{y,U_h}}{N S^2_{y,U}} = \frac{\sum_{U_h} (y_k - \bar{y}_{U_h})^2}{\sum_U (y_k - \bar{y}_U)^2}$$

Aunque $\sum_{h=1}^H w_h \neq 1$, el peso del estrato h sí tiene una interpretación interesante, pues queda definido como la suma de cuadrados dentro del estrato, dividido por la suma de cuadrados totales de la variable de interés. Y podemos concluir que cuando los estratos están bien construidos, esto es, la variable de interés es homogénea dentro de cada estrato y los diferentes estratos son heterogéneos entre sí, los pesos w_h serán muy pequeños y el $DEFF$ general resultará mucho más pequeño que los $DEFF$ de los estratos.

Por otro lado, si los estratos no fueron construidos teniendo en cuenta la variabilidad de la característica de interés, entonces $S^2_{y,U_h} \approx S^2_{y,U}$ y $w_h = N_h/N$. De esta forma, la suma de los pesos es igual a 1, y se puede concluir que el $DEFF$ del diseño general es un promedio ponderado de los efectos de los H estratos, y el estrato que tiene mayor peso será aquel que tiene mayor representación del universo.

Finalmente, alguno de los pesos w_h puede resultar ser mayor a 1 cuando para algún estrato $\frac{n_h}{N_h} \neq \frac{n}{N}$, y cuando los estratos no están bien construidos.

El efecto de diseño en las encuestas de hogares de la región

En general, para las encuestas de hogares en la región, se planean esquemas de estratificación, aglomeración y selección de UPM con probabilidades desiguales. Heeringa et al. (2017) anotan que el efecto de estratificación reduce la varianza de las estrategias de muestreo, mientras que el efecto de selección desigual tiende a aumentarla. En general, estos dos efectos tienden a anularse entre sí. Por lo tanto, el efecto de diseño de una encuesta compleja estará únicamente en función del efecto de aglomeración, el cual puede llegar a ser grande, en comparación con los otros dos. Como ya se había comentado antes, la expresión generalizada que da cuenta del efecto de aglomeración en los diseños de muestreo complejos de las encuestas de hogares es la siguiente:

$$DEFF \approx 1 + (\bar{n}_{II} - 1)\rho_y$$

En donde se recalca que $\bar{n}_{II}$ representa el número promedio de hogares seleccionados dentro de cada UPM y ρ_y es el coeficiente de correlación intraclase, que representa el grado de homogeneidad de la variable de interés dentro de cada hogar.

Este efecto cambiará dependiendo de si la inferencia de la encuesta de hogares se quiere realizar a nivel nacional o a nivel regional. Por ejemplo, UN (2005, capítulo 7) presenta el comportamiento de esta medida a lo largo de tres encuestas de hogares en Brasil: la *Pesquisa Nacional por Amostra de Domicílios* (PNAD), la *Pesquisa Mensal de Emprego* (PME) y la *Pesquisa de Padrões de Vida* (PPV). En general, estas encuestas utilizan estratificación y selección de UPM con probabilidades desiguales; además, el tamaño promedio de las UPM es de 250 viviendas, de las cuales son seleccionadas 13 por la PNAD, 20 en la PME y 16 y 8 viviendas en la PPV en la zona rural y urbana, respectivamente.

Basado en UN (2005, capítulo 7), se nota que los efectos de diseño no solo son diferentes para cada parámetro que se desea estimar sino que varían de acuerdo a la subpoblación en la que se realice la estimación. Por ejemplo, considere al considerar el parámetro *proporción de hogares con electricidad*, se estimó que el efecto de diseño para este parámetro fue de 7.92 a nivel nacional, de 1.03 en las áreas metropolitanas, de 4.43 en las ciudades grandes y de 7.27 en las áreas rurales. Por lo anterior, y basado en la expresión que define el efecto de diseño, se observó que, fijando $\bar{n}_{II}$, el coeficiente de correlación intraclase varío dependiendo de la zona. En efecto, $\rho_y = 0.76$ a nivel nacional, $\rho_y = 0.0033$ en las zonas metropolitanas, $\rho_y = 0.38$ en las ciudades grandes y $\rho_y = 0.69$ en las áreas rurales. Lo anterior implicó que hay una mayor heterogeneidad de los hogares con electricidad entre las UPM a nivel nacional y en las áreas rurales, es decir algunos hogares tienen electricidad y otros no entre las UPM. Sin embargo, en las zonas metropolitanas la variación de esta variable entre las UPM es casi nula, es decir que todos lo hogares tienen electricidad entre las UPM de estas zonas.

Por otro lado, para la misma encuesta PNAD, los efectos de diseño para el número promedio de cuartos usados como dormitorios es de 2.14 a nivel nacional, de 2.37 en las áreas metropolitanas, de 1.72 en las ciudades grandes y de 2.09 en las áreas rurales. Considerando que $\bar{n}_{II} = 10$, el coeficiente de correlación intraclase es de $\rho_y = 0.12$ a nivel nacional, $\rho_y = 0.15$ en las zonas metropolitanas, $\rho_y = 0.08$ en las ciudades grandes y $\rho_y = 0.12$ en las áreas rurales. Lo anterior implica que hay una mayor homogeneidad del número de cuartos utilizados como dormitorio entre las UPM del país y de las zonas que lo componen.

Como se verá en los capítulos posteriores, al conocer el valor que toma el efecto de diseño para la estimación de un parámetro de interés, es posible crear escenarios de simulación que permitan establecer el tamaño de muestra en la planeación de las encuestas de hogares o en su rediseño después de la ronda de censos en una década particular.

Capítulo 8

Cálculo del tamaño de muestra

Uno de los tópicos más importantes en la literatura del diseño y análisis de encuestas de hogares es el tamaño de muestra. En general, en los libros de estadística y muestreo se establecen las características generales de los esquemas de muestreo y las propiedades estocásticas de los estimadores sin profundizar en que la muestra debe seleccionarse y que esta selección depende de cuántos hogares se necesiten en el estudio. De hecho, al hablar del tamaño de muestra en una encuesta de hogares, no solo se debe hacer referencia a los hogares, sino también a las personas.

En efecto, la determinación del tamaño de muestra debe depender del propósito de la encuesta. Por ejemplo, considere una encuesta de propósitos múltiples que se levanta cada año con el fin de indagar acerca de fenómenos demográficos, sociales, educativos, y de condiciones de vida; en este contexto, se debe tener en cuenta que el tamaño de muestra definido debe ser útil, pertinente y apropiado para todos los indicadores que se desean medir al mismo tiempo. En este capítulo, el lector podrá encontrar una guía útil para identificar la mejor ruta a la hora de abordar el cálculo del tamaño de muestra en las encuestas de hogares.

I. Confiabilidad y precisión

Antes de introducir las metodologías básicas para el cálculo del tamaño de muestra mínimo, es necesario definir los diferentes tipos de error muestral en una encuesta. En principio, se define un intervalo de confianza para el parámetro θ, inducido por su estimador insesgado $\hat{\theta}$ (que se supone con distribución normal de media θ y varianza $Var(\hat{\theta})$, como

$$IC(1-\alpha) = \left[\hat{\theta} - z_{1-\alpha/2}\sqrt{Var(\hat{\theta})}, \hat{\theta} + z_{1-\alpha/2}\sqrt{Var(\hat{\theta})}\right]$$

donde $z_{1-\alpha/2}$ se refiere al cuantil $(1-\alpha/2)$ de una variable aleatoria con distribución normal estándar. Cuando el diseño de muestreo es complejo, es necesario reemplazar el percentil de la distribución normal estándar por el presentir de una distribución $t-student$ con $N_I - H$ grados de libertad, suponiendo que hay N_I unidades primarias de muestreo y H estratos. En este orden de ideas, nótese que

$$1-\alpha = \sum_{Q_0 \supset s} p(s),$$

donde Q_0 es el conjunto de todas las posible muestras cuyo intervalo de confianza contiene al parámetro θ. Desde la expresión del intervalo de confianza, se define el *margen de error*, como aquella cantidad que se suma y se resta al estimador insesgado. En este caso, se define como

$$ME = z_{1-\alpha/2}\sqrt{Var(\hat{\theta})}$$

Desde esta expresión también es posible definir el *error estándar*, dado por

$$EE = \sqrt{Var(\hat{\theta})}$$

Las anteriores medidas sólo tienen en cuenta la precisión del estimador. Una medida que tiene en cuenta la precisión y el sesgo del estimador es el *margen de error relativo*, que se define como

$$MER = z_{1-\alpha/2}\frac{\sqrt{Var(\hat{\theta})}}{E(\hat{\theta})}$$

De la misma manera, también se define el *coeficiente de variación* o *error estándar relativo* definido por

$$CV = \frac{\sqrt{Var(\hat{\theta})}}{E(\hat{\theta})}$$

El tamaño de muestra dependerá del tipo de error que se quiera minimizar. Por ejemplo, para una población particular, el tamaño de muestra requerido para minimizar el margen de error, no será el mismo que el que se necesitará para minimizar el coeficiente de variación.

II. El efecto de diseño en la determinación del tamaño de muestra

Al momento de diseñar un estudio por muestreo con encuestas de hogares, es importante establecer el número mínimo de encuestas que se deben realizar. Esto es necesario para determinar el costo del estudio; y en el aspecto técnico, permite tener control desde la fase de diseño sobre la calidad estadística de los resultados esperados en el estudio. Como se mencionó anterioMERnte, esta calidad puede ser medida en términos del error muestral, con indicadores tales como el margen de error, el margen de error relativo o el coeficiente de variación. Todas estas medidas dependen de la varianza del estimador bajo el diseño muestral complejo; por lo tanto, contar con un valor aproximado para el efecto de diseño $DEFF$ nos permite obtener una aproximación a dicha varianza, y acercarnos al error muestral del estudio en la fase del diseño.

Uno de los primeros paradigmas con el que se debe lidiar es el de la independencia entre las observaciones. Este es un supuesto que gobierna gran parte de la teoría de análisis estadístico, pero que infortunadamente no se aplica en el contexto de las encuestas de hogares. Ante los retos que se debe enfrentar y las diversas estrategias de recolección de información, las fórmulas que se desprenden del supuesto de que los observaciones corresponden a una muestra de variables independientes e idénticamente distribuidas no son plausibles.

La estratificación, las múltiples etapas y la aglomeración de las unidades de muestreo hacen que este supuesto no se cumpla en la práctica y por tanto, utilizar las expresiones tradicionales que se encuentran en los libros introductorios de estadística guiará a tamaños de muestra insuficientes. El problema del tamaño de muestra en encuestas de hogares ha sido abordado por diferentes autores con diferentes enfoques. Quizás uno de los más aceptados es aquel que define un factor de ajuste, llamado efecto de diseño, en función de la correlación que hay entre la variable de interés con las unidades primarias de muestreo. A partir de este efecto de diseño se calcula el número de personas que deben ser encuestadas para minimizar un error de muestreo predefinido.

Cuando para la población de interés, se selecciona una muestra utilizando un diseño de muestreo de conglomerados o en varias etapas, no es imposible afirmar que existe independencia entre las observaciones. Lo anterior hace que no sea posible utilizar las fórmulas clásicas para la determinación de un tamaño de muestra, al considerar un diseño de muestreo aleatorio simple. Sin embargo, una forma sencilla de incorporar este efecto de aglomeración en las expresiones clásicas del muestreo aleatorio simple la da relación de las varianzas en el efecto de diseño:

$$DEFF(\hat{\theta}) = \frac{Var_p(\hat{\theta})}{Var_{MAS}(\hat{\theta})}$$

Esta cifra da cuenta del efecto de aglomeración causado por la utilización de un diseño de muestreo cualquiera (p), frente a un diseño de muestreo aleatorio simple (MAS) en la inferencia de un parámetro de la población finita θ (que puede ser un total, una proporción, una razón, un coeficiente de regresión, etc.). Por lo anterior, es posible escribir la varianza del estimador bajo el diseño de muestreo complejo como

$$\begin{aligned} Var_p(\hat{\theta}) &= DEFF(\hat{\theta})\ Var_{MAS}(\hat{\theta}) && (8.1)\\ &= DEFF(\hat{\theta})\ \frac{N^2}{n}\left(1-\frac{n}{N}\right)S^2_{y_U} && (8.2)\end{aligned}$$

Por lo tanto, si al implementar un muestreo aleatorio simple el tamaño de muestra n_0 es suficiente para conseguir la precisión deseada, entonces el valor del tamaño de muestra que tendrá en cuenta el efecto de aglomeración para un diseño complejo estará cercano a $n \approx n_0 \times DEFF$. UN (2008b) afirma que, dada esta relación, no es deseable aceptar estrategias de muestreo que induzcan efectos de diseño mayores a 2.5 o 3. Lo cual quiere decir que los equipos técnicos dentro de los INE deben plantear esquemas en donde el efecto de diseño para los indicadores claves de la encuesta no sea desproporcionadamente grande.

En particular, para el caso de una proporción, la calidad del estimador se puede medir en términos de la amplitud del intervalo de confianza de al menos $(1-\alpha)\times 100\,\%$; esto es, la distancia entre el estimador y el parámetro no debería superar un margen de error previamente establecido (ME). Así:

$$1-\alpha \geq \Pr\left(|\hat{P} - P| < ME\right)$$

Por ejemplo, el estimador de Horvitz-Thompson de la proporción $\hat{P}$ es insesgado para P y su distribución asintótica es gausiana con varianza dada por

$$Var\left(\hat{P}\right) = DEFF\frac{1}{n}(1-\frac{n}{N})P(1-P)$$

Al despejar el tamaño muestral n de la anterior expresión, se tiene que

$$n \geq \frac{P(1-P)}{\frac{ME^2}{DEFF\ z^2_{1-\alpha/2}}+\frac{P(1-P)}{N}}$$

De la misma manera, si el interés recae en la estimación de un promedio $\bar{y}_U$, el tamaño de muestra necesario para que la amplitud relativa del intervalo de confianza no supre un margen de error relativo previamente establecido (MER) es de

$$n \geq \frac{S^2_{y_U}DEFF}{\frac{MER^2\bar{y}^2_U}{z^2_{1-\alpha/2}}+\frac{S^2_{y_U}DEFF}{N}}$$

Por consiguiente, se evidencia que valores grandes del efecto de diseño inducirán un mayor tamaño de muestra. Claramente el incremento no es lineal, más aún, el tamaño de muestre se ve más afectado en la medida en que el $DEFF$ sea más grande.

III. Escenarios de interés

En general, en encuestas de hogares se parte de un marco de muestreo de áreas que agrupa a toda la población de un país. Estas áreas están definidas como agregaciones cartográficas o UPM y contienen a su vez a los hogares en donde se encuentran las personas que son susceptibles de ser entrevistadas. Sin embargo, debido a la agrupación natural de las personas en hogares, a veces los cálculos se hacen complejos, máxime conociendo que la población de interés es un subconjunto de los habitantes de los hogares. Por otro lado, debido a que el marco de muestreo comúnmente usado por las INE es una lista de UPM, se hace necesario más allá de calcular el tamaño de muestra de las personas, también calcular el tamaño de muestra de UPM y hogares en la muestra. Por lo tanto, en este documento se pretende sintetizar los mecanismos de asignación de muestra en tres escenarios que son comunes en la práctica estadística del diseño de encuestas de hogares:

1. Primer escenario: asignación del tamaño de muestra en problemas de inferencia que tienen que ver con la estimación de parámetros de personas. En este escenario se presenta la metodología apropiada para calcular el tamaño de muestra de UPM, hogares y finalmente personas.

2. Segundo escenario: cuando la variable de diseño y en general, las variables más importantes de la encuestas están presentes a nivel de hogar, entonces no es necesario realizar un submuestreo de personas. Partiendo de la lógica presentada en el escenario anterior, se presenta la metodología adecuada para calcular el tamaño de muestra de UPM y de hogares.

3. Tercer escenario: un caso menos común en los países de América Latina se presenta cuando el marco de muestreo empadrona las personas dentro de las UPM y además la encuesta sólo

pretende observar características asociadas a los habitantes del hogar (y por tanto no intenta observar características ni del hogar ni de la vivienda). En este caso no hay un submuestreo de hogares.

En general, al definir las expresiones de tamaño de muestra, se debe ser cuidadoso con la notación, para lo cual suponemos una población U de N elementos sobre la que se desea seleccionar una muestra s de n elementos en los cuales se quiere medir una característica de interés. En algunos casos, la población U no constituye la población de interés sino que la contiene; es decir, si se define a U_d como la población de interés, entonces $U_d \subseteq U$. En términos de notación, se tiene lo siguiente:

- N es el tamaño de la población U.
- n es el tamaño de la muestra s.
- N_I es el número de UPM en el marco de muestreo.
- n_I es el número de UPM que se selecciona en la muestra de la primera etapa s_i.
- N_{II} es el número de hogares existentes en el país.
- n_{II} es el número de hogares seleccionados en la muestra de la segunda etapa s_{II}.
- $\bar{n}$ es el número promedio de personas que se van a seleccionar en cada UPM.
- $\bar{n}_{II}$ es el número promedio de hogares que se van a seleccionar en cada UPM.
- ρ_y es el coeficiente de correlación intraclase, calculado para la variable de interés sobre las UPM.
- b es el número promedio de personas por hogar.
- r es el porcentaje promedio de personas en el hogar susceptibles de ser observadas para la característica de interés.
- $z_{1-\alpha/2}$ es el percentil $(1-\alpha/2)$ asociado a una distribución normal estándar y a la confianza que se requiera en la inferencia.

Para introducir las metodologías apropiadas, junto con las expresiones adecuadas, en cada escenario se definirán las cantidades de interés, se dará una breve introducción al problema y se realizarán los cálculos detenidamente con ejemplos de encuestas reales. Para mantener la uniformidad en los cálculos, todos los ejemplos suponen una población de tamaño $N = 50$ millones, con $N_{II} = 12$ millones de hogares, para el cual se desea obtener una muestra con una confianza del 90 %. En cada escenario se supone que el país está dividido en $N_I = 30$ mil UPM, conformadas por segmentos cartográficos (agregaciones de manzanas).

Para simplificar los cálculos y mantener la atención del lector, las expresiones que se presentarán en este capítulo corresponden al número de individuos que deberían ser seleccionados a nivel nacional, o para un solo subgrupo de interés. Por lo tanto, estos cálculos deben ser hechos tantas veces como dominios de representatividad exista en la encuesta. Por ejemplo, si el interés está en hacer inferencia en dos estratos: el rural y el urbano, entonces se debe calcular estas expresiones dos veces, una para cada área. Al final, el tamaño de muestra nacional será la sumatoria de los tamaños de muestra en cada uno de los estratos del país.

A. Tamaño de muestra para UPM, hogares y personas

Cuando la unidad de observación sean las personas, sin importar que la variable de interés esté a nivel de hogar, será necesario siempre basar nuestros cálculos en el tamaño de muestra de las personas. Por ejemplo, para tener una inferencia apropiada al estimar el ingreso medio percápita, el porcentaje de personas pobres o el porcentaje de personas con una característica particular es necesario definir a la población objetivo como todas las personas que componen un hogar para posterioMERnte medir la variable de interés que será observada para todas ellas.

Con estos elementos es posible realizar simulaciones de algunos escenarios de muestreo, que indiquen el tamaño de muestra necesario en cada una de las etapas de la selección de la muestra. Si fuese posible sistematizar los elementos más importantes a la hora de calcular el tamaño de muestra en una encuesta de hogares, sería necesario recurrir a los siguientes pasos de manera ordenada:

- **Definir la población de interés de manera explícita**. En particular, es necesario aclarar si la unidad de análisis son las personas o los hogares. De esta forma, se debe fijar los valores para r y b. Si la unidad de análisis son todas las personas del hogar, entonces el porcentaje de personas con la característica de interés será $r = 1$, de otra forma $r < 1$. Por otro lado, el número promedio de personas por hogar b dependerá del dominio de representatividad en el que se requiera el cálculo.

- **Definir el número promedio de hogares**. El número promedio de hogares que se desea encuestar en cada una de las UPM está dado por $\bar{n}_{II}$. Este proceso debería ser repetido de forma iterativa en los pasos subsiguientes para poder evaluar la calidad del diseño. De las varias opciones de $\bar{n}_{II}$ será necesario escoger solo una.

- **Calcular el número promedio de personas que serán encuestadas**. Al igual que en el paso anterior es necesario probar varios escenarios que redundarán en la escogencia de un número óptimo de personas por UPM. Los valores de $\bar{n}$ dependen directamente del paso anterior al escoger $\bar{n}_{II}$. Debido a que la selección de las personas está supeditada a la selección de los hogares, entonces $\bar{n}$ se puede descomponer manteniendo la relación con r y b, de la siguiente manera:

$$\bar{n} = \bar{n}_{II} \times r \times b$$

- **Calcular el efecto de diseño**. Es necesario definir (o calcular con encuestas o censos anteriores) la correlación intraclase de la variable de interés con el agrupamiento por UPM ρ_y. Luego de esto, se debe calcular el efecto de diseño $DEFF$ como función de ρ_y y de $\bar{n}$; esto es $DEFF \approx 1 + (\bar{n} - 1)\rho_y$. Nótese que esta cifra sólo se calcula sobre la población de interés.

- **Calcular el tamaño de muestra de personas**. A partir de las expresiones de tamaño de muestra para diseños de muestreo complejos, calcular el tamaño de muestra necesario para lograr una precisión adecuada en la inferencia. En primer lugar, si lo que se quiere estimar es un promedio $\bar{y}_U$, el tamaño de muestra necesario para alcanzar un margen de error relativo máximo de $MER \times 100\,\%$ es de

$$n \geq \frac{S^2_{y_U} DEFF}{\frac{MER^2 \bar{y}_U^2}{z^2_{1-\alpha/2}} + \frac{S^2_{y_U} DEFF}{N}}$$

Por otro lado, si lo que se quiere estimar es una proporción P, y se utiliza el margen de error, entonces la expresión apropiada para calcular el tamaño de muestra estará dada por

$$n \geq \frac{P\,(1-P)\,DEFF}{\dfrac{MER^2P^2}{z^2_{1-\alpha/2}} + \dfrac{P\,(1-P)\,DEFF}{N}}$$

- **Calcular el tamaño de muestra de hogares**. Es necesario calcular el número total de hogares que deben ser seleccionados para lograr entrevistar a todas las personas que serán observadas en el punto anterior. El número de hogares que deben ser seleccionados estará determinado por las cantidades n, b y r, de la siguiente forma

$$n_{II} = \frac{n}{r \times b}$$

- **Calcular el número de UPM**. Los hogares y las personas se observan a partir de las UPM. En este paso final es necesario calcular el número de UPM que deben ser seleccionadas en el muestreo a partir de la relación

$$n_I = \frac{n}{\bar{n}} = \frac{n_{II}}{\bar{n}_{II}}$$

Ejemplo: proporción de personas pobres

Suponga que el parámetro de interés es el porcentaje de personas pobres (cuyo ingreso está por debajo de un umbral preestablecido) y se quiere hacer inferencia con un margen de error relativo máximo del 5 %. Por estudios anteriores en este país, se ha estimado que la proporción de personas pobres está alrededor de $P = 4\,\%$. Nótese que la población objetivo está conformada por todos los habitantes del hogar puesto que $r = 100\,\%$. En este país se ha estimado que el tamaño promedio del hogar es de alrededor de $b = 3.5$ personas. Por último, según levantamientos anteriores, la correlación intraclase de la característica de interés con las unidades primarias de muestreo es $\rho_y = 0.034$.

La siguiente tabla resume los resultados del ejercicio para $\bar{n}_{II} = 10$ hogares por UPM, los cuales implican que por cada UPM se entrevistarían en promedio a $\bar{n} = 10 * 1 * 3.5 = 35$ personas. Con lo anterior se obtendría un efecto de diseño $DEFF = 2.2$, para un total de personas en la muestra de $n = 55936$ que serán observados a partir de la selección de $n_{II} = 55936/(1 * 3.5) = 15982$ hogares en $n_I = 55936/35 = 1598$ UPM.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
10	35	2.2	1598	15982	55936

Por supuesto que es posible plantear otros escenarios a medida que se evalúe el efecto que conlleva el cambio del número de hogares que se seleccionan en cada UPM. Por ejemplo, el investigador podría proponer que se seleccionarán en promedio 5 hogares por UPM, lo cual cambiaría el número de UPM que serían seleccionadas en la muestra de la primera etapa, así como también el número total de personas que serían seleccionadas en todo el operativo. Debido a que el efecto de diseño es una función del número de hogares promedio a seleccionar en las UPM, esta cifra también variará.

A continuación se muestran algunos resultados que permiten establecer estos escenarios cuando se varía el tamaño de muestra promedio de hogares por UPM. La escogencia del escenario ideal se debe dar en términos de la conveniencia logística y presupuestal en el estudio. Siguiendo las recomendaciones internacionales, se desestimarían los escenarios con efectos de diseño mayores a 3.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
5	18	1.6	2315	11575	40512
10	35	2.2	1598	15982	55936
15	52	2.8	1359	20386	71351
20	70	3.4	1239	24787	86756
25	88	3.9	1167	29186	102152
30	105	4.5	1119	33582	117538
35	122	5.1	1085	37976	132915
40	140	5.7	1059	42366	148282
45	158	6.3	1039	46754	163640

Ejemplo: ingreso promedio por persona

Suponga que se desea estimar el ingreso promedio por hogar con un margen de error relativo máximo del 2 %. La variable de interés (ingreso) es continua y se estima que la media oscila alrededor de $\bar{y}_U = 1180$ dólares con una varianza de $S^2_{y_U} = 1845.94^2$. En este caso, la población objetivo son todos los habitantes del hogar, por lo cual $r = 100\,\%$. La composición del hogar se calcula en $b = 3.79$ personas por hogar. Por último, según levantamientos anteriores, la correlación intraclase de la característica de interés es $\rho_y = 0.035$. Nótese que la correlación intraclase cambia con respecto a la característica que se desee medir.

La siguiente tabla muestra los resultados del ejercicio al seleccionar $\bar{n}_{II} = 15$ hogares por UPM, que a su vez implica que por cada UPM se encontrarían en promedio $\bar{n} = 15 * 1 * 3.79 \cong 57$ personas por UPM. Con lo anterior se obtendría un efecto de diseño $DEFF = 3$, para un total de personas en la muestra de $n = 48861$ que serán observados a partir de la selección de $n_{II} = 48861/(1*3.79) = 12892$ hogares en $n_I = 859$ UPM.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
15	57	3	859	12892	48861

A continuación se muestran algunos resultados que permiten establecer otros escenarios de muestreo cuando se varía el tamaño de muestra promedio de hogares por UPM. Recuérdese que cualquiera de estos escenarios es válido, desde el punto de vista de la eficiencia estadística, aunque no todos serán válidos si se tienen en cuenta otros aspectos como los logísticos o presupuestales. Por ejemplo, si se escogiera el penúltimo escenario, entonces para 50 hogares por UPM, se debería encuestar en promedio a 190 personas, lo cual reduciría el número de UPM a 662, pero aumentaría el tamaño de muestra general a 33098 personas, lo cual involucraría mayores costos de contratación de encuestadores, supervisores y seguramente un operativo de campo de más días de duración. Siguiendo las recomendaciones internacionales, se desestimarían los escenarios con efectos de diseño mayores a 3.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
5	19	1.6	1422	7108	26938
10	38	2.3	1000	10001	37902
15	57	3.0	859	12892	48861
20	76	3.6	789	15783	59816
25	95	4.3	747	18672	70766
30	114	4.9	719	21560	81711
50	190	7.6	662	33098	125443
100	379	14.2	619	61857	234439

Ejemplo: tasa de desocupación en adultos mayores

Suponga que la incidencia de la desocupación está alrededor de $P = 5.5\,\%$ en la población objetivo, que son las personas económicamente activas (PEA) mayores de 60 años; en este país, se ha estimado que en promedio hay $r = 4.6\,\%$ de adultos mayores por hogar que pertenecen a la PEA, cuyo tamaño promedio es de alrededor de $b = 5$ personas. Además, se quiere hacer inferencia con un margen de error relativo máximo del 15 %. Por último, según levantamientos anteriores, la correlación intraclase de la característica de interés es $\rho_y = 0.7$.

La siguiente tabla muestra los resultados del ejercicio, que implica que seleccionar $\bar{n}_{II} = 20$ hogares por UPM implicaría un promedio de $\bar{n} = 20 * 0.046 * 5 = 4.6$ adultos mayores en la PEA (personas de interés) por UPM. Con lo anterior se obtendría un efecto de diseño $DEFF = 3.5$, para un total de $n = 7272$ adultos mayores en la PEA que serán observados en la muestra a partir de la selección de $n_{II} = 7272/(0.046 * 5) \cong 31617$ hogares en $n_I = 7272/4.6 \cong 1581$ UPM.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
20	4.6	3.5	1581	31617	7272

En este caso la muestra en los 31617 hogares induce un operativo muy grande que implicaría la observación de $31617 * 5 = 158085$ personas en los hogares, de las cuales $n = 7272$ serían los casos de interés. Como se ha visto en los anteriores ejemplos, es posible plantear otros escenarios a medida que se evalúe el efecto que conlleva el cambio del número de hogares que se seleccionan en cada UPM. A continuación se muestran algunos resultados que permite establecer estos escenarios cuando se varía el tamaño de muestra promedio de hogares por UPM. Siguiendo las recomendaciones internacionales, se desestimarían los escenarios con efectos de diseño mayores a 3.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
5	1.1	1.1	1985	9926	2283
10	2.3	1.9	1716	17157	3946
15	3.5	2.7	1626	24387	5609
20	4.6	3.5	1581	31617	7272
25	5.8	4.3	1554	38848	8935
30	6.9	5.1	1536	46074	10597
50	11.5	8.3	1500	74983	17246
100	23.0	16.4	1472	147222	33861

B. Tamaño de muestra para UPM y hogares

En algunas ocasiones la variable de interés y la unidad de observación están a nivel de hogar. Por ejemplo, cuando todas las variables de interés se miden a nivel de la vivienda o del hogar. En este caso es posible modificar el algoritmo de la sección anterior para seleccionar únicamente a las viviendas u hogares en la muestra, sin necesidad de realizar un submuestreo de personas. En este caso algunas cantidades desaparecen porque no son objeto de la población de hogares, como lo son r y b; algunas otras expresiones deben ser redefinidas al contexto de los hogares, como por ejemplo, el coeficiente de correlación intraclase ρ_y, el efecto de diseño y todas las expresiones de tamaños de muestra. En todo caso, la adaptación del algoritmo se describe a continuación.

- **Definir el número promedio de hogares**. El número promedio de hogares que se desea encuestar en cada una de las UPM está dado por $\bar{n}_{II}$. Esta cifra sigue siendo el insumo más importante del algoritmo y se propone crear escenarios de muestreo a partir de su modificación y evaluación del tamaño de muestra final.

- **Calcular el efecto de diseño**. Es necesario definir (o calcular con encuestas o censos anteriores) la correlación intraclase ρ_y de la variable de interés *a nivel del hogar* con el agrupamiento por UPM definido por la división cartográfica del último censo. De igual forma, el efecto de diseño $DEFF \approx 1 + (\bar{n}_{II} - 1)\rho_y$ sigue siendo función del tamaño de muestra promedio de hogares por UPM ($\bar{n}_{II}$).

- **Tamaño de muestra de hogares**. Partiendo de las expresiones de tamaño de muestra generales para muestreos complejos y teniendo en cuenta que la población de interés son los hogares y que la variable de interés está a nivel de hogar, entonces el tamaño de muestra necesario para alcanzar un margen de error relativo máximo de $MER\%$ es de

$$n_{II} \geq \frac{S_y^2 DEFF}{\frac{MER^2 \bar{y}^2}{z_{1-\alpha/2}^2} + \frac{S_{y_U}^2 DEFF}{N_{II}}}$$

La expresión apropiada para calcular el tamaño de muestra para una proporción estará dada por

$$n_{II} \geq \frac{P\,(1-P)\,DEFF}{\frac{MER^2 P^2}{z_{1-\alpha/2}^2} + \frac{P\,(1-P)\,DEFF}{N_{II}}}$$

- **Cálculo del número de UPM**. Los hogares se observan a partir de las UPM. En este paso final es necesario calcular el número de UPM que deben ser seleccionadas en el muestreo a partir de la relación

$$n_I = \frac{n_{II}}{\bar{n}_{II}}$$

Ejemplo: gasto promedio del hogar

Suponga que se desea estimar el promedio de gasto anual en dólares en los hogares del país con un margen de error relativo máximo admisible del 3.5 %. La variable de interés (gasto) es continua y se estima que la media oscila alrededor de $\bar{y}_U = 1407$ dólares con una varianza de $S^2_{y_U} = 2228^2$. En este ejemplo se supone que el país está dividido en $N_I = 10$ mil UPM y la correlación intraclase de la característica de interés, medida a nivel del hogar, con las UPM es de $\rho_y = 0.173$.

La siguiente tabla muestra los resultados del ejercicio para $\bar{n}_{II} = 12$ hogares promedio por UPM, que serán observados a partir de la selección de $n_{II} = 16056$ hogares y $n_I = 16056/12 = 1338$ UPM, los cuales inducen un efecto de diseño $DEFF = 2.9$.

Hogares promedio por UPM ($\bar{n}_{II}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
12	2.9	1338	16056

A continuación se muestran algunos resultados que permiten establecer otros escenarios de muestreo al variar el tamaño de muestra promedio de hogares por UPM. Nótese que, por ejemplo, en el caso de seleccionar 20 hogares por UPM, se debería seleccionar una muestra de 23695 hogares en tan solo 1185 UPM. Por otro lado, si sólo se seleccionaran 2 hogares por UPM, se tendrían que visitar 3246 UPM en todo el país, aunque el número de encuestas totales descendería a 6493.

Para este tipo de operativos, en donde los cuestionarios de gasto de los hogares están acompañados de un operativo exhaustivo que le permite al investigador conocer los hábitos de consumo del hogar de forma desagregada, y en donde se visita el hogar durante un periodo de tiempo determinado, tal vez sea más conveniente estudiar la viabilidad de seleccionar más hogares por UPM y menos UPM para que el operativo de campo no exija la contratación de demasiado personal en campo. Al estar agrupados en menos UPM, se podría definir un mejor levantamiento de la información con un equipo mediano de personas; de lo contrario, se debería contratar bastante más personal que cubra las UPM dispersas a lo largo del país. Siguiendo las recomendaciones internacionales, se desestimarían los escenarios con efectos de diseño mayores a 3.

Hogares promedio por UPM ($\bar{n}_{II}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
2	1.2	3246	6493
4	1.5	2102	8407
6	1.9	1720	10320
8	2.2	1529	12233
10	2.6	1414	14145
12	2.9	1338	16056
14	3.2	1283	17967

Hogares promedio por UPM ($\bar{n}_{II}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
16	3.6	1242	19877
18	3.9	1210	21787
20	4.3	1185	23695

Ejemplo: proporción de hogares sin agua potable

Suponga que se desea obtener una muestra con un margen de error relativo máximo admisible del 10 % sobre la variable de interés (necesidades básicas insatisfechas en agua) y el parámetro de interés es el porcentaje de hogares con esta carencia. En este país, se estima que la proporción de hogares con esta condición asciende a $P = 7.5\,\%$. En este ejemplo se supone que la correlación intraclase de la característica de interés con las UPM es de $\rho_y = 0.045$.

La siguiente tabla muestra los resultados del ejercicio para $\bar{n}_{II} = 10$ hogares por UPM, que serán observados a partir de la selección de $n_{II} = 4360$ hogares en $n_I = 4360/10 = 436$ UPM, que inducen un efecto de diseño $DEFF = 1.3$.

Hogares promedio por UPM ($\bar{n}_{II}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
10	1.3	436	4360

A continuación se muestran algunos resultados que permiten establecer otros escenarios de muestreo al variar el tamaño de muestra promedio de hogares por UPM. Observe que el efecto de diseño DEFF es directamente proporcional al número de hogares entrevistados por UPM y al tamaño de muestra de hogares final; de la misma manera, es inversamente proporcional al número de UPM.

Hogares promedio por UPM ($\bar{n}_{II}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
5	1.1	758	3790
10	1.3	436	4360
15	1.5	328	4924
20	1.6	274	5490
25	1.8	242	6057
30	2.0	221	6624
35	2.2	205	7190
40	2.3	194	7757
45	2.5	185	8323

C. Tamaño de muestra para UPM y personas

En algunos casos en donde la variable de interés esté a nivel de persona y el cuestionario de la encuesta no induzca preguntas acerca del hogar, y además exista un inventario detallado de las personas que residen en cada UPM, es posible evadir la selección de los hogares e ir directamente a la selección de personas. En este caso, la lógica del cálculo del tamaño de muestra se mantiene modificando en cierta manera el algoritmo de las secciones anteriores, tal como se ilustra a continuación.

- **Definir la población de interés de manera explícita**. En este caso solo se mantiene la expresión correspondiente a r que denota el porcentaje de personas con la característica de interés en la población.
- **Definir el número promedio de personas**. El número promedio de personas (casos de la población objetivo) que se desean encuestar por cada UPM está dado por $\bar{n}$. Al igual que en las secciones anteriores, se recomienda hacer una evaluación sobre esta cantidad para determinar posibles escenarios de muestreo.
- **Calcular el efecto de diseño**. Es necesario definir el efecto de diseño $DEFF$ como una función de la correlación existente entre la variable de interés y la conformación de las UPM. De esta forma, $DEFF \approx 1 + (\bar{n} - 1)\rho_y$. Nótese que esta cifra solo podrá ser calculada sobre la población de interés.
- **Tamaño de muestra de personas**. A partir de las expresiones de tamaño de muestra para muestreos complejos, calcular el tamaño de muestra necesario para lograr una precisión adecuada en la inferencia. En este caso, las expresiones de tamaño de muestra coinciden con las del primer escenario.
- **Tamaño de muestra final**. Es necesario calcular el número total de personas que deben ser seleccionados para lograr observar a quienes hacen parte de la población objetivo. Esta cantidad está dada por n/r.
- **Cálculo del número de UPM**. Finalmente, las personas se observan a partir de las UPM. En este paso final es necesario calcular el número de UPM que deben ser seleccionadas en el muestreo a partir de la relación.
$$n_I = \frac{n}{\bar{n}}$$

Ejemplo: ingreso promedio en personas empleadas

Suponga que se desea estimar el ingreso promedio en las personas empleadas con un margen de error relativo máximo admisible del 2 %. La variable de interés (ingreso) es continua y se estima que la media oscila alrededor de $\bar{y}_U = 1458$ dólares con una varianza de $S^2_{y_U} = 2191^2$. Nótese que la población objetivo son todas las personas empleadas, cuya proporción se estima en $r = 46\,\%$. La correlación intraclase de la característica de interés es $\rho_y = 0.038$.

La siguiente tabla muestra los resultados del ejercicio al seleccionar $\bar{n} = 23$ personas de la población de interés por UPM, que a su vez implica que por cada UPM se deberían seleccionar y enlistar en promedio $23/0.46 = 50$ personas por UPM. Con lo anterior se esperarían $n = 28029$ personas empleadas en la muestra repartidas en $n_I = 28029/23 \cong 1219$ UPM; en este caso se obtendría un efecto de diseño $DEFF = 1.8$. Nótese que en este escenario, el operativo de campo abarcaría la selección y enlistamiento de $28029/0.46 \cong 60933$ personas, de las cuales se esperaría que 28029 fueran de la población de interés (personas empleadas).

Personas seleccionadas por UPM ($\bar{n}/r$)	Personas empleadas por UPM $\bar{n}$	DEFF	Muestra de UPM (n_I)	Muestra de personas empleadas (n)	Muestra de personas (n/r)
50	23	1.8	1219	28029	60933

A continuación se muestran algunos resultados que permiten establecer otros escenarios de muestreo cuando se varía el tamaño de muestra promedio de hogares por UPM. Siguiendo las recomendaciones internacionales, se desestimarían los escenarios con efectos de diseño mayores a 3.

Personas seleccionadas por UPM ($\bar{n}/r$)	Personas empleadas por UPM $\bar{n}$	DEFF	Muestra de UPM (n_I)	Muestra de personas empleadas (n)	Muestra de personas (n/r)
25	12	1.4	1857	21360	46435
50	23	1.8	1219	28029	60933
75	34	2.3	1006	34695	75424
100	46	2.7	899	41360	89913
125	58	3.1	835	48023	104398

Ejemplo: proporción de analfabetas pobres

Suponga que se quiere estimar la proporción de incidencia de la pobreza sobre la población analfabeta con un margen de error relativo máximo admisible del 15 %. Se ha estimado que alrededor del $r = 14\,\%$ de las personas del país no saben leer ni escribir. Por otro lado, tal como se vio en un ejemplo anterior, el fenómeno de la pobreza está estimado en $P = 4\,\%$. y la correlación intraclase de la característica de interés es $\rho_y = 0.0454$.

La siguiente tabla muestra los resultados del ejercicio al seleccionar un promedio de $\bar{n} = 14$ personas por UPM que no saben leer ni escribir; esto implica la selección y enlistamiento de $14/0.14 = 100$ personas por UPM. Con lo anterior se obtendría un efecto de diseño $DEFF = 1.6$, para un total de personas en la muestra de $n = 4574$ personas analfabetas, de una muestra 32671 personas enlistadas y repartidas en $n_I = 4574/14 \cong 327$ UPM.

Personas seleccionadas por UPM ($\bar{n}/r$)	Personas analfabetas por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de personas analfabetas (n)	Muestra de personas (n/r)
100	14	1.6	327	4574	32671

Es posible plantear otros escenarios a medida que se evalúe el efecto que conlleva el cambio del número de hogares que se seleccionan en cada UPM. A continuación se muestran algunos resultados que permite establecer estos escenarios cuando se varía el tamaño de muestra promedio de hogares por UPM.

Personas seleccionadas por UPM ($\bar{n}/r$)	Personas analfabetas por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de personas analfabetas (n)	Muestra de personas (n/r)
25	3.5	1.1	917	3211	22936
50	7.0	1.3	524	3665	26179
75	10.5	1.4	392	4120	29429
100	14.0	1.6	327	4574	32671
125	17.5	1.7	287	5029	35921

IV. Otros escenarios de interés

En las encuestas de hogares también surgen escenarios particulares que llevan a sugerir distintos caminos para la adopción de un determinado tamaño de muestra. En esta sección analizaremos los casos en los que los parámetros de interés son diferencias de proporciones y dobles diferencias. También se revisará el caso del planteamiento de pruebas de hipótesis y su relación con el tamaño de muestra.

A. Tamaño de muestra para la estimación de la diferencia de dos proporciones

Suponga una población U, que se encuentra particionada[1] en dos subpoblaciones U_1 de tamaño N_1 y U_2, de tamaño N_2. El interés del investigador está en conocer la diferencia de algunas proporciones entre estos grupos. Por ejemplo, suponga que se quiere conocer la diferencia entre las proporciones de niños desnutridos por sexo. Se espera que la proporción de niños desnutridos no supere la proporción de niñas desnutridas para verificar que no hay brechas de sexo. Por lo tanto, el parámetro de interés se escribe como:

$$\theta = P_1 - P_2 = \frac{N_{d1}}{N_1} - \frac{N_{d2}}{N_2}$$

En donde $N_{di} = \sum_{k \in U_i} z_{dik}$ $(i = 1, 2)$ y z_{dik} es una característica dicotómica que indica si el individuo k-ésimo de la subpoblación U_i está en estado de desnutrición. Por supuesto, bajo muestreo aleatorio simple, un estimador insesgado para θ es

$$\hat{\theta} = \hat{P}_1 - \hat{P}_2 = \frac{\hat{N}_{d1}}{N_1} - \frac{\hat{N}_{d2}}{N_2}$$

En donde, $\hat{N}_{di} = \frac{N_i}{n_i} \sum_{k \in s_i} z_{dik}$ y s_i es la muestra asociada con la población U_i. Luego, la varianza del anterior estimador es:

$$Var(\hat{\theta}) = Var\left(\hat{P}_1\right) + Var\left(\hat{P}_2\right) - 2Cov\left(\hat{P}_1, \hat{P}_2\right)$$

[1] Esta metodología también aplica en el caso en el que $U \supset (U_1 \cup U_2)$.

Por otro lado, siendo $|U_i|$ la cardinalidad del conjunto U_i, se definen las siguientes relaciones:

$$T_i = \frac{|U_1 \cap U_2|}{|U_i|} \qquad i = 1, 2.$$

De esta forma, T_1 y T_2 corresponde al porcentaje de traslape de las subpoblaciones. De la misma manera, definiendo a $R_{1,2}$ como la correlación de Pearson entre los datos observados de ambas subpoblaciones, entonces la covarianza entre este par de estimadores estaría determinada por la siguiente relación (Kish, 2004):

$$Cov(\hat{P}_1, \hat{P}_2) = \sqrt{Var(\hat{P}_1)}\sqrt{Var(\hat{P}_2)}\sqrt{T_1}\sqrt{T_2}R_{1,2}$$

En esta instancia, es útil recordar que si las poblaciones U_1 y U_2 son estratos (o agregaciones de estratos) que inducen conjuntos dijuntos y la selección de la muestra en cada uno es independiente por diseño, entonces $Cov(\hat{P}_1, \hat{P}_2) = 0$. Si, por otro lado, no existe independencia en el muestreo de ambas poblaciones, entonces $R_{1,2} \neq 0$ necesariamente. Es útil recordar que esta correlación se debe evaluar a través de las UPM. Siguiendo con el ejemplo, a pesar de que las subpoblaciones son niños y niñas, $R_{1,2} \neq 0$. Por otro lado, para encontrar el tamaño de muestra óptimo, es útil realizar los siguientes supuestos:

1. Asumir que las subpoblaciones son grandes y por ende $N_1 = N_2 = N$.
2. Por lo anterior, asumir que los tamaños de muestra pueden ser iguales, tales que $n_1 = n_2 = n$.

Nótese a su vez que, si el levantamiento de las observaciones no puede ser realizado utilizando muestreo aleatorio simple, sino que por el contrario, la muestra aleatoria fue seleccionada mediante un diseño de muestreo complejo con un efecto de diseño[2] ($DEFF$) no ignorable y mayor a uno, entonces la varianza tomaría la siguiente forma

$$Var(\hat{\theta}) = \frac{DEFF}{n}\left(1 - \frac{n}{N}\right) S^2_{\theta}$$

En donde, definiendo a $Q_i = 1 - P_i$, se tiene que:

$$S^2_{\theta} = P_1Q_1 + P_2Q_2 - 2\sqrt{T_1}\sqrt{T_2}R_{1,2}\sqrt{P_1Q_1}\sqrt{P_2Q_2}$$

De esta manera, un intervalo de confianza del 95 % para la diferencia de proporciones está dado por

$$IC(95\,\%)_{\theta} = \hat{\theta} \pm z_{1-\alpha/2}\sqrt{\frac{DEFF}{n}\left(1 - \frac{n}{N}\right) S^2_{\theta}}$$

Lo anterior quiere decir que el margen de error (ME) de la encuesta debe ser tal que:

$$ME < z_{1-\alpha/2}\sqrt{\frac{DEFF}{n}\left(1 - \frac{n}{N}\right) S^2_{\theta}}$$

Por lo tanto, despejando n, se tiene que la muestra en cada subgrupo debe mayor que:

[2] Recuerde que si el muestreo es aleatorio simple, el efecto de diseño es $DEFF = 1$.

$$n > \frac{DEFF\ S_\theta^2}{\dfrac{ME^2}{z_{1-\alpha/2}^2} + \dfrac{DEFF\ S_\theta^2}{N}}$$

Nótese que, dependiendo de los pocentajes de traslape $\sqrt{T_1}$, $\sqrt{T_2}$ y de la correlación de la característica de interés en ambas subpoblaciones $R_{1,2}$, la varianza S_θ^2 tomará diferentes formas, como se detalla a continuación:

1. Si no hay traslape, $T_1 = T_2 = 0$, y $S_\theta^2 = P_1Q_1 + P_2Q_2$.
2. Si hay traslape completo, $T_1 = T_2 = 1$ y $S_\theta^2 = P_1Q_1 + P_2Q_2 - 2R_{1,2}\sqrt{P_1Q_1}\sqrt{P_2Q_2}$.
3. Si hay traslape parcial y balanceo, $T_1 = T_2 = T$ y si además se considera que las varianzas en cada subgrupo o periodo son similares $P_1Q_1 = P_2Q_2 = PQ$, entonces $S_\theta^2 = 2PQ(1 - TR_{1,2})$.

Covarianza en comparaciones mensuales

Suponga que se quiere comparar la tasa de desempleo nacional entre dos meses consecutivos. En este escenario, asumiendo que existe independencia en el muestreo de los dos meses consecutivos, el porcentaje de traslape de muestra entre los dos meses (que por diseño es nulo) es igual a cero. Por lo tanto, $T_1 = T_2 = 0$. Luego, el término de la covarianza se anula. En resumen, la varianza del estimador en este caso sería igual a:

$$Var(\hat{P}_1 - \hat{P}_2) = Var(\hat{P}_1) + Var(\hat{P}_2)$$

Covarianza en comparaciones trimestrales o anuales

Bajo un esquema rotativo 2(2)2, suponga que se quiere comparar la tasa de desempleo nacional entre trimestres consecutivos o entre el mismo mes de dos años consecutivos. En este escenario no existe independencia en el muestreo de los dos trimestres consecutivos puestos que la estructura del panel garantiza un traslape del 50 %. En este caso $T_1 = T_2 \approx 0.5$.

Por otro lado, existe una correlación natural entre las viviendas comunes en el panel que se midieron en los periodos de interés, por lo tanto $R_{1,2} \neq 0$. Note que esta correlación se calcula sobre los individuos comunes en el panel y sobre la variable dicotómica que induce la tasa de desempleo (perteneciente a la población económicamente activa). En resumen, el término de covarianza en este caso sería igual a:

$$Cov(\hat{P}_1, \hat{P}_2) = \frac{1}{2}\sqrt{Var(\hat{P}_1)}\sqrt{Var(\hat{P}_2)}R_{1,2}$$

Covarianza en comparaciones de un mismo mes

En primer lugar, suponga que se quiere comparar la tasa de desempleo entre hombres y mujeres en un mismo mes. En este escenario no existe independencia en el muestreo de hombres y mujeres puesto que estos grupos no son estratos de muestreo. En este caso T_1 es la proporción de hombres y T_2 es la proporción de mujeres. Nótese que $T_1 \neq T_2$.

Como se comentó anterioMERnte, existe una correlación natural entre las UPM que fueron seleccionadas y que contienen tanto a hombres como a mujeres, por lo tanto $R_{12} \neq 0$. Note que esta correlación se calcula sobre todos los individuos pertenecientes a la fuerza de trabajo y sobre la variable dicotómica que induce la tasa de desempleo. En resumen, el término de covarianza en este caso sería igual a:

$$Cov(\hat{P}_1, \hat{P}_2) = \sqrt{Var(\hat{P}_1)}\sqrt{Var(\hat{P}_2)}\sqrt{T_1}\sqrt{T_2}R_{1,2}$$

Por otro lado, suponga que se quiere comparar la tasa de desempleo entre dos regiones del mismo país en un mismo mes. En este escenario existe independencia en el muestreo de las dos regiones porque la selección es independiente en cada región Esta independencia se tiene por definición del diseño de muestreo puesto que ambas regiones son agrupaciones disjuntas entre estratos de muestreo. En este caso T_1 es la proporción de personas de la primera ciudad y T_2 es la proporción de personas de la segunda ciudad. Además, tampoco no existe una correlación entre las UPM que fueron seleccionadas entre estas regiones porque la selección fue independiente, por lo tanto $R_{12} = 0$. En resumen, el término de covarianza es nulo y por ende la varianza del estimador sería igual a:

$$Var(\hat{d}) = Var(\hat{P}_1) + Var(\hat{P}_2)$$

B. Tamaño de muestra para la estimación del impacto en dos mediciones longitudinales

Para las encuestas que planean un seguimiento panel o de panel rotativo, es posible contemplar escenarios en los que se quiera estimar el efecto de una intervención, definido como la diferencia en diferencias de las proporciones de interés. De esta forma, el efecto se define como:

$$\theta = (P_{1,1} - P_{2,1}) - (P_{1,2} - P_{2,2})$$

En donde $P_{i,j}$ $(i, j = 1, 2.)$ corresponden a las proporciones del grupo i en la oleada j. Entonces el tamaño de muestra mínimo[3] necesario para lograr una estimación confiable de esta diferencia, con menos del $ME \times 100\,\%$ de margen de error, es:

$$n \geq \frac{DEFF\ S^2_\theta}{\dfrac{ME^2}{z^2_{1-\alpha/2}} + \dfrac{DEFF\ S^2_\theta}{N}}$$

En donde

$$S^2_\theta = (P_{1,1}Q_{1,1} + P_{1,2}Q_{1,2} + P_{2,1}Q_{2,1} + P_{2,2}Q_{2,2})(1 - TR)$$

En donde T corresponde a la tasa de traslape ($T = 1$ corresponde a un panel completo, $T = 0.5$ a un semi-panel con traslape del $50\,\%$ y el caso extremo $T = 0$ a una encuesta en donde no hay traslape) y R se define como la correlación entre las dos oleadas ($R = 0$ implica que no hay correlación entre

[3]Note que el tamaño de muestra de toda la encuesta es $4n$, en las dos oleadas, puesto que se debe seleccionar n elementos en cada grupo y en cada oleada.

los dos momentos, $R = -1$ implica una máxima correlación negativa entre los dos momentos y $R = 1$ implica una correlación positiva máxima entre los dos momentos).

Por ejemplo, en una encuesta de fuerza laboral inteMERdiada por alguna intervención gubernamental, puede ser de interés la evaluación del efecto de esa política de asistencia laboral entre hombres y mujeres en dos periodos de tiempo.

C. Tamaño de muestra para el contraste de hipótesis en la diferencia de proporciones

Suponga que el investigador desea realizar el contraste de una hipótesis de interés. En particular, suponga que hay dos grupos de interés en la población finita y que la hipótesis está inducida por la diferencia de las proporciones en las dos poblaciones. El investigador considera que la diferencia es significativa para el fenómeno en cuestión si es mayor que un valor D definido de antemano y conocido como el *tamaño del efecto* que el investigador desea detectar.

Nótese que la significación estadística, inducida por un valor-p, no siempre tiene la misma connotación de significación científica o económica, que puede presentarse en fenómenos raros, para los cuales no necesariamente se gozaría de significación estadística. Por lo tanto el sistema de hipótesis que se quiere contrastar es el siguiente:

$$H_o : P_1 - P_2 = 0 \qquad vs. \qquad H_a : P_1 - P_2 = D > 0$$

Nótese que, acudiendo a la distribución normal de los estimadores de las proporciones, y suponiendo independencia en el muestreo de los subgrupos, la regla de decisión en este caso induce el rechazo de la hipótesis nula cuando

$$\frac{\hat{P}_1 - \hat{P}_2}{\sqrt{Var(\hat{P}_1 - \hat{P}_2)}} > z_{1-\alpha}$$

Si las características del estudio implican que el diseño de muestreo es complejo con un $DEFF > 1$, entonces esta regla de decisión rechaza la hipótesis nula si

$$\frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{DEFF}{n}\left(1 - \frac{n}{N}\right)(P_1Q_1 + P_2Q_2)}} > z_{1-\alpha}$$

En este caso, es necesario controlar la probabilidad de cometer el error tipo-2 (aceptar una hipótesis nula, dado que ésta es falsa). A esta probabilidad se le conoce como *potencia* y, suponiendo que nuestro interés está en detectar un tamaño del efecto $P_1 - P_2 = D$, la potencia está dada por

$$
\begin{aligned}
\beta &\leq Pr\left(\frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)(P_1Q_1 + P_2Q_2)}} > z_{1-\alpha} \mid P_1 - P_2 = D\right)\\
&= Pr\left(\frac{(\hat{P}_1 - \hat{P}_2) - D}{\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)(P_1Q_1 + P_2Q_2)}} > z_{1-\alpha} - \frac{D}{\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)(P_1Q_1 + P_2Q_2)}} | P_1 - P_2 = D\right)\\
&= 1 - \Phi\left(z_{1-\alpha} - \frac{D}{\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)(P_1Q_1 + P_2Q_2)}}\right)
\end{aligned}
$$

Por lo anterior,

$$
1-\beta \geq \Phi\left(z_{1-\alpha} - \frac{D}{\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)(P_1Q_1 + P_2Q_2)}}\right)
$$

Entonces, dado que la función $\Phi()$ es creciente, se tiene que

$$
z_{1-\beta} \geq z_{1-\alpha} - \frac{D}{\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)(P_1Q_1 + P_2Q_2)}}
$$

En consecuencia, al despejar n, se tiene que la muestra en cada subgrupo debe mayor que:

$$
n \geq \frac{DEFF(P_1Q_1 + P_2Q_2)}{\frac{D^2}{(z_{1-\alpha} + z_\beta)^2} + \frac{DEFF(P_1Q_1 + P_2Q_2)}{N}}
$$

V. Algunas relaciones de interés para proporciones

Cuando se trata de estadísticas de la fuerza laboral, una variable clave para el diseño de una encuesta de hogares que mida la dinámica del mercado de trabajo es el estado de los individuos en la fuerza laboral. Para los gobiernos, es de interés proporcionar un conjunto de indicadores destinados a medir y rastrear la ocupación de los ciudadanos del país (o región). Por ejemplo, se puede obtener estimaciones de la tasa de desempleo actual (medida mensual o trimestralmente); asimismo, también son de interés la variación neta entre dos períodos y los flujos brutos entre categorías de empleo entre períodos.

Es posible mencionar tres tipos de planificación en las encuestas de hogares desde las cuales es posible abordar adecuadamente las características particulares de los estudios de fuerza laboral. La primera es a través de las encuestas repetidas,donde se realizan mediciones similares en diferentes puntos del tiempo a diferentes personas cada vez. La segunda son las encuestas de panel, donde se realizan mediciones en diferentes puntos en el tiempo a las mismas personas cada vez. La tercera son las encuestas rotativas, donde se incluyen elementos y se siguen en la muestra durante un período específico, y a medida que salen de la muestra, se agregan nuevos elementos.

Una regla general común para calcular el tamaño de la muestra afirma que como la variable de diseño es dicotómica (dependiendo del estado de empleo), la varianza de ese tipo de variables encuentra su máximo cuando la probabilidad de éxito es 0.5. Sin embargo, si las políticas públicas en un país se centran en lograr que la tasa de desempleo sea baja a través de algunas intervenciones gubernamentales que afectan (positivamente) a la fuerza laboral, y si esas estrategias son efectivas, entonces la probabilidad de éxito de la variable de diseño cambia y puede afectar el tamaño de la muestra de las encuestas de hogares.

En esta sección documento, centramos nuestra atención en el tamaño de la muestra inducido por el control del margen de error; en donde a medida que la proporción disminuye, el tamaño de la muestra aumenta sustancialmente. Sin embargo, al controlar el margen de error, debido a que la función de varianza detrás de este enfoque es simétrica alrededor de 0.5, se puede encontrar que el mismo tamaño de muestra necesario para cumplir con los requisitos de calidad para cualquier proporción (P_d) es el mismo que el requerido para satisfacer los requisitos de calidad para su complemento aditivo ($1 - P_d$).

A continuación se proporciona varios ejemplos que tipifican algunos escenarios que se pueden encontrar en la práctica. Los cálculos se pueden reproducir empleando el software estadístico `R` (R Core Team, 2020), mediante el uso de la biblioteca `samplesize4surveys` (Rojas, 2020), utilizando específicamente las funciones `ss4p` y `ss4dp`

Estimación de proporciones

1. Primer escenario: si la tasa de desempleo es baja, digamos $P = 0.05$ y el margen de error se fija en $ME = 0.0025$, entonces el intervalo de confianza esperado sería $IC = 0.05 \pm 0.0025 = (0.0475, 0.0525)$. En este caso el tamaño de muestra requerido es de alrededor de 55169.

2. Segundoescenario: si la tasa de desempleo es alta, digamos $P = 0.2$, y el margen de error de error se fija en $ME = 0.01$, entonces el intervalo de confianza sería $IC = 0.2 \pm 0.01 = (0.19, 0.21)$, y el tamaño de muestra requerido es 12144.

Nótese que ambos escenarios dan lugar al mismo margen de error relativo (MER), definido como $MER = \frac{ME}{P}$. En efecto, para el primero, tenemos $MER = (0.0025/0.05) \times 100\,\% = 5\,\%$, y para el segundo, tenemos $MER = (0.01/0.2) \times 100\,\% = 5\,\%$. Por lo tanto, incluso para el mismo margen de error relativo, el tamaño de la muestra debe ser mayor si el fenómeno que nos interesa tiene una baja incidencia en la población finita. De hecho, es posible definir una función de información para saber si el tamaño de su muestra es suficiente para cumplir con los requisitos de calidad para una proporción determinada. Esto es útil porque no se sabe exactamente qué valor tomará la proporción. Además, si la encuesta de hogares intenta estimar otras proporciones (como en una encuesta multipropósito), se encontrará rápidamente si su tamaño de muestra actual es adecuado para todo el estudio.

3. Tercer escenario: si el tamaño de la muestra se define como $n = 10000$, y la proporción es $P = 0.2$, entonces el coeficiente de variación será de 2,8 % y el margen de error será del 1.1 %. Es posible notar que todas las proporciones estimadas tendrán un margen de error inferior al 1.4 %.

4. Cuarto escenario: si el tamaño de la muestra se define como $n = 40000$, y la proporción se centra en $P = 0.05$, entonces el coeficiente de variación será del 3 % y el margen de error será de 0.2 %. Es posible notar que todas las proporciones estimadas tendrán un margen de error inferior al 0.7 %.

Teniendo en cuenta que, para una proporción P dada, el tamaño de muestra requerido para lograr un margen de error particular es el mismo que para su complemento aditivo $1 - P$, como es de esperar, si un tamaño de muestra alcanza los requisitos para una proporción establecida, también alcanzará los requisitos de calidad para cualquier proporción superior.

Sin embargo, para una proporción P, el tamaño muestral requerido para lograr un coeficiente de variación particular no es el mismo que para su complemento aditivo $1 - P$. Luego, para una proporción baja, se puede encontrar que con un tamaño de muestra dado el coeficiente de variación será mayor que para su complemento aditivo. Sobre la base de los resultados encontrados en esta sección, y bajo un MER fijo (5 % en todos los casos), encontramos que al intentar estimar proporciones (como la tasa de desempleo):

- Si la proporción es baja, anticipamos un gran tamaño de muestra.
- Si la proporción es alta, esperamos un tamaño de muestra pequeño.

Cambios netos

Ahora dirigimos nuestra atención a los cambios netos en la tasa de desempleo durante dos períodos, $\Delta = |P_1 - P_2|$. Este tipo de parámetro se puede estimar utilizando una encuesta repetida, rotativa o de panel. Sin embargo, como se evidenció en las anteriores secciones, hay una reducción en el tamaño de la muestra si se intentan estimar los cambios netos desde una encuesta rotativa o de panel. Como no estamos estimando una proporción, sino un cambio neto,tenemos que considerar qué valores son adecuados para el establecer el margen de error absoluto.

5. Quinto escenario: si no esperamos cambios significativos entre ambos períodos, y las tasas de desempleo son altas, por ejemplo $\Delta \approx |0.22 - 0.20| = 0.02$, y el margen de error se fija en $ME = 0.001$, entonces el intervalo de confianza sería $IC = 0.02 \pm 0.001 = (0.019, 0.021)$, y el tamaño de muestra requerido estaría alrededor de 96224.

6. Sextoescenario: si no esperamos cambios significativos entre períodos, y las tasas de desempleo son bajas, por ejemplo $\Delta \approx |0.05 - 0.03| = 0.02$, y el margen de error se fija en $ME = 0.001$, entonces el intervalo de confianza sería $IC = 0.02 \pm 0.001 = (0.019, 0.021)$, y el tamaño de muestra requerido debería ser de 59536.

7. Séptimo escenario: si esperamos cambios significativos entre períodos, y las tasas de desempleo difierenpor ejemplo $\Delta \approx |0.05 - 0.20| = 0.15$ y el margen de error se fija en $ME = 0.0075$, entonces el intervalo de confianza sería $IC = 0.15 \pm 0.0075 = (0.1425, 0.1575)$, y el tamaño de muestra requerido estaría alrededor de 22083.

Nótese qeu los escenarios quinto, sexto y séptimo dan como resultado el mismo MER, definido como $MER = \frac{ME}{\Delta}$. En efecto, para el séptimo escenario tenemos MER =(0.0075/0.15) %=5 %, y para el quinto y sexto, tenemos MER =(0.001/0.02) %=5 %. Por lo tanto, incluso para el mismo valor del cambio neto, el tamaño de la muestra no será el mismo y variará dependiendo de la configuración de las proporciones. Por supuesto, hay que esperar cambios más drásticos si varía la porción del traslape y la correlación entre períodos.

Además, se debe tener en cuenta que es posible encontrar diferentes configuraciones de proporciones en ambos periodos que induzcan el mismo valor en el cambio neto. Contrariamente a lo que se esperaría, si un tamaño de muestra alcanza los requisitos de calidad para un parámetro Δ, no necesariamente cumplirá con los requisitos de calidad para el mismo valor nominal del cambio neto bajo una configuración diferente en las proporciones involucradas.

Para cumplir con los requisitos de calidad, bajo el mismo MER, se necesitará un mayor tamaño de muestra si no se esperan cambios significativos en las tasas de desempleo entre los períodos involucrados. Si el cambio neto sigue siendo el mismo para ambos períodos, para cumplir con los requisitos de calidad, bajo el mismo MER, se necesitará un mayor tamaño de muestra si el fenómeno del desempleo es alto. Ahora, al intentar estimar los cambios netos de proporciones (como el cambio anual o mensual en las tasas de desempleo), encontramos que:

- Si las tasas son significativamente diferentes, esperamos un tamaño de muestra pequeño.
- Si las tasas son similares y las proporciones son bajas, requerimos un tamaño de muestra moderado.
- Si las tasas son similares, y las proporciones son grandes, esperamos un gran tamaño de muestra.

VI. Algunas consideraciones adicionales

Cuando la encuesta se ha planeado para que tenga representatividad para algún conjunto de estratos, es necesario replicar estas mismas expresiones en cada uno de los subgrupos de interés. Por otro lado, las anteriores aproximaciones al cálculo de tamaño de muestra son insuficientes ante la realidad de la ausencia de respuesta y las desactualizaciones de los marcos de muestreo. En esta sección se profundizará en estos tópicos.

A. Asignación del tamaño de muestra en los estratos de muestreo

Como se aclaró anteriormente, todas las encuestas de hogares en América Latina tienen un componente explícito de estratificación, y por ende una pregunta que surge inmediatamente es: ¿después de determinar el tamaño de muestra general, como asignarlo apropiadamente en todos los estratos de muestreo? En general, se supone que el tamaño de la muestra general es n y que hay H estratos fijos; por ende, se quiere determinar los tamaños de muestra n_h para cada estrato $(h = 1, \ldots, H)$, de tal manera que se garantice la ganancia de precisión de la estrategia de muestreo.

Existen varios tipos de asignación que pueden estudiarse para determinar la más apropiada, en términos de eficiencia; a continuación se presenta una lista no exhaustiva de ellas:

1. Asignación proporcional: en donde se selecciona una proporción de elementos en cada estrato siguiendo la estructura poblacional. Lohr (2019) afirma que este tipo de asignación se utiliza cuando es deseable que la muestra se pueda ver como una versión miniatura de la población. Gutiérrez (2016) señala que si se define la *fracción de muestreo* como $f_h = n_h/N_h$ en el estrato h, entonces al utilizar la asignación proporcional la fracción de muestreo será la misma para todos los estratos, tal que $f_h = f$. En este caso la probabilidad de inclusión de cualquier elemento en la población $\pi_k = f_h = f$ es constante y fija. De esta manera, cada unidad en la muestra representará el mismo número de elementos en la población, independientemente del estrato al que pertenezca. Bajo la asignación proporcional, el tamaño de muestra en cada

estrato está dado por

$$n_h = f \times N_h$$

2. Asignación de Neyman: en donde se selecciona una muestra de elementos en cada estrato de tal forma que se maximice la eficiencia estadística de la estrategia de muestreo. la estructura poblacional. Groves et al. (2009) mencionan que, bajo este método, se producen las menores varianzas para la media muestral comparado con otras técnicas de asignación de tamaño de muestra. Bajo la asignación de Neyman, el tamaño de muestra que minimiza la varianza de la estrategia de muestreo está dado por

$$n_h = n\frac{N_h S_{yU_h}}{\sum_{h=1}^{H} N_h S_{yU_h}}$$

donde $S_{yU_h} = \sqrt{S^2_{yU_h}}$ es la raíz de la varianza de la característica de interés en cada estrato. Gutiérrez (2016) afirma que, con respecto a la asignación de Neyman, es recomendable redondear el tamaño de muestra en cada estrato al entero más próximo.

3. Asignación de Kish: al usar la asignación proporcional en los estratos pequeños, la muestra puede resultar muy pequeña, generando problemas de eficiencia y pérdida de precisión. Por otro lado, utilizar una asignación unifoMER (selección del mismo número de elementos en cada estrato $n_h = c$) tendrá como consecuencia una variación sustancial en las fracciones de muestreo entre los estratos y, por ende, una fracción de muestreo muy grande del estrato más pequeño. Un punto inteMERdio entre la asignación proporcional y la asignación unifoMER es la asignación propuesta por Kish, la cual toma la siguiente expresión:

$$n_h = n\frac{\sqrt{\frac{1}{H^2} + I\ W_h^2}}{\sum_{h=1}^{H}\sqrt{\frac{1}{H^2} + I\ W_h^2}}$$

En donde $W_h = N_h/N$, e $I \geq 0$ es el índice de asignación de Kish, que denota la importancia relativa entre las estimaciones nacionales y las de cada estrato. A medida que este índice se hace más pequeño, menor importancia se le dará a las estimaciones nacionales. La asignación de Kish proporciona un balance entre la asignación unifoMER y la proporcional. Cuando $I = 0$, se reduce a la asignación unifoMER, mientras que si $I \to \infty$ tiende a un enfoque de asignación proporcional. Usualmente se utiliza $I = 1$ para garantizar que la precisión de las características de interés en lo nacional y en los estratos sea aproximadamente la misma.

B. Ajustes de subcobertura

Debido a las características propias de las encuestas de hogares, siempre se presentará un fenómeno que puede ser descrito como una realidad: *existirá ausencia de respuesta en las encuestas de hogares.* En estos términos, los institutos nacionales de estadística deben tomar medidas preventivas al momento de adjudicar los tamaños de muestra en cada estrato, puesto que contar con un tamaño efectivo de muestra mucho menor al planeado inicialmente puede conllevar problemas de sesgo y de precisión en las estimaciones de las cifras nacionales o regionales, con las cuales se aborda la política económica y de desarrollo de los países de la región.

En encuestas de hogares cuyo diseño es longitudinal, no solamente se debe abordar el problema de la ausencia de respuesta al momento de la aplicación de la encuesta, sino que debe ser visto de manera integral y más general debido a que un hogar que pertenezca a un panel puede decidir no participar más después de un par de visitas. Es así como la atrición se convierte en un problema que enmarca la ausencia de respuesta como un fenómeno al cual se debe prestar atención para evitar problemas de sesgo y baja confiabilidad.

Kalton (2009) advierte que el diseño de la encuesta debe tener en cuenta el ajuste de submuestras; por ejemplo, para estimar el cambio de la condición de pobreza o indigencia en los hogares es necesario realizar un ajuste al tamaño de muestra inicial para que al final de la aplicación de la encuesta el tamaño de muestra efectivo cumpla con los requerimientos de precisión de la inferencia estadística. Los INE pueden estimar, con base en su basta experiencia en la realización de encuestas, la probabilidad de que una persona (o jefe de hogar) responda al instrumento. Si esta probabilidad es denotada como $\phi = Pr(k \in s_r)$, en donde s_r denota el subconjunto de respondientes efectivos, entonces los tamaños de muestra de individuos y hogares serán ajustados al dividirlos por ϕ.

$$n_{final} = \frac{n_{inicial}}{\phi}$$

Por ejemplo, si esta probabilidad fue estimada en $\phi = 0.8$, entonces todos los tamaños de muestra calculados en los pasos anteriores deberán ser ajustados como $n_{final} = \frac{n_{inicial}}{0.8} = 1.25 \times n_{inicial}$. Por último, si la información auxiliar lo permite, este ajuste debería realizarse de manera diferenciada en cada uno de los estratos. Por ejemplo, si se conoce que este fenómeno de ausencia de respuesta tiene una mayor incidencia en lo rural que en lo urbano, entonces este ajuste debería tenerse en cuenta de forma diferenciada.

C. Sustituciones y reemplazos

Una práctica común en los operativos de campo de las encuestas de hogares en Latinoamérica es sustituir las UPM y viviendas para las cuales no se ha obtenido respuesta. Por ejemplo, se consideraría el reemplazo de las UPM cuando no se puede acceder al sitio geográfico por diferentes razones; por ejemplo, problemas de orden público o seguridad, algún cambio importante en la infraestructura de la zona, o porque no se tiene el consentimiento informado de las autoridades de la comunidad. En este caso, si no se puede acceder a la UPM, no se puede tampoco acceder a ninguno de los hogares que la integran. Los esquemas de sustituciones y reemplazos en las encuestas de hogares utiliza, por lo general, la metodología de *estratificación implícita* que permite seleccionar de manera automática a los reemplazos adecuados de acuerdo a la conformación de subgrupos poblacionales similares.

La estratificación implícita es usada cuando la encuesta está enfocada en un tema particular y, para su ejecución exitosa, requiere el uso del muestreo sistemático con probabilidades desiguales en la selección de las UPM, es decir en la definición del diseño de muestreo de la primera etapa. Según UN (2008b, pág. 46), en la mayoría de países la secuencia podría empezar con el estrato urbano, desagregado por departamento, a su vez desagregado por municipio; el estrato rural, de forma similar, es desagregado por departamento, a su vez desagregado por comuna o vereda. Observe que la selección sistemática de UPM está condicionada a la medida de tamaño utilizada en la primera etapa, es decir el número de viviendas que la componen. De esta forma, la estratificación implícita consiste en que, para cada estrato explícito (urbano, rural, regiones, etc.) se crea una lista ordenada de UPM. Esta lista estará ordenada por los estratos implícitos definidas en la planeación de la encuesta (departamento, municipio) y dentro de cada subgrupo se ordenan las UPM en orden descendente (o ascendente). De esta forma, esta metodología constituye un método objetivo de selección de reemplazos, puesto que si no se puede acceder a la UPM seleccionada originalmente, su reemplazo será la inmediatamente anterior (o posterior) en la lista estratificada implícitamente. Este procedimiento seleccionará como reemplazo a la UPM ubicada en el mismo municipio, dentro del mismo departamento, en la misma zona y con un número similar de viviendas, respetando el principio de representatividad. De otra forma, si no se considera un procedimiento similar a la estratificación implícita, los reemplazos de las UPM podrían ser seleccionados aleatoriamente en otro departamento y con un número de viviendas mucho más grande o mucho más pequeño, añadiendo sesgo a la selección inicial.

Aunque la estratificación implícita permite acotar el sesgo generado por la ausencia de respuesta de las UPM, Vehovar (1999, págs. 348 - 349) advierte que se debe tener precaución en cuanto a los usos de esta práctica puesto que también puede conllevar sesgos importantes en las estimaciones de interés. Lo anterior se desprende del hecho de que los individuos ubicados en zonas donde sí es posible acceder diferirán significativamente de aquellos individuos ubicados en las zonas de difícil acceso; es evidente que se trata de dos realidades diferentes. Por esta razón es útil que, después de haber valorado los posibles sesgos, si se ha tomado la determinación de realizar las sustituciones sobre las UPM de difícil acceso, se realice un seguimiento exhaustivo en cada levantamiento que permita clasificar el esquema de recolección de información primaria y se valore su impacto en la precisión de los estimadores resultantes.

Parte II

Procesamiento transversal de las encuestas de hogares

Capítulo 9

Estimación de parámetros

Un estimador se define como una función de la muestra aleatoria, el cual toma valores en los reales y solo depende de los elementos pertenecientes a la muestra. A su vez, un diseño de muestreo, se define por el soporte que es el conjunto Q de todas las posibles muestras. Las propiedades estadísticas de un estimador están determinadas por la medida de probabilidad discreta inducida por el diseño de muestreo. Es decir, dada la probabilidad de selección de cada muestra $s \in Q$, la esperanza, la varianza y otras propiedades de interés de los estimadores están definidas a partir de $p(s)$. En particular, la esperanza de un estimador $\hat{\theta}$ sigue la siguiente expresión:

$$E(\hat{\theta}) = \sum_{s \subset Q} \theta(s) p(s)$$

Las propiedades más comúnmente buscadas en un estimador $\hat{\theta}$ son el insesgamiento, la eficiencia y la consistencias. De esta manera, el sesgo está definido por la siguiente expresión:

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

y el error cuadrático medio, dado por

$$ECM(\hat{\theta}) = E[\hat{\theta} - \theta]^2 = Var(\hat{\theta}) + B^2(\hat{\theta}).$$

Si el sesgo de un estimador es nulo, se dice que el estimador es insesgado y cuando esto ocurre el error cuadrático medio se convierte en la varianza del estimador. Además si la varianza del estimador es pequeña en relación con otros estimadores, se dice que el estimador es eficiente. Por último, si cuando el tamaño de muestra crece, el valor del estimador se acerca al parámetro desconocido, se dice que el estimador es consistente. Särndal et al. (2003) afirman que el objetivo en un estudio por muestreo es estimar uno a más parámetros poblacionales. Gutiérrez (2016) resalta que las decisiones más importantes a la hora de abordar un problema de estimación por muestreo son

1. La escogencia de un diseño de muestreo y un algoritmo de selección que permita implementar el diseño.
2. La elección de una fórmula matemática o estimador que calcule una estimación del parámetro de interés en la muestra seleccionada.

Las anteriores no son decisiones independientes. Es decir, la escogencia de un estimador dependerá, usualmente, del diseño de muestreo utilizado. De hecho, si $\hat{\theta}$ es un estimador del parámetro θ y $p_s(\cdot)$ un diseño de muestreo definido sobre un soporte Q, entonces la estrategia de muestreo será la dupla $(p(\cdot), \hat{T})$. A continuación se presentan algunos estimadores comúnmente usados en el procesamiento de las encuestas de hogares en la región.

Aunque el marco de referencia de la teoría de muestreo es la estimación de un parámetro de interés, lo cierto es que en la práctica no solo se necesitan estimaciones que cobijen la población entera sino que también son indispensables estimaciones que involucren subgrupos poblacionales, puesto que estos inducen una partición de la población de interés. En general, es bien sabido que cuando se habla de subgrupos poblacionales se está haciendo referencia a dominios de interés, estratos o postestratos. Cuando el investigador se enfrenta a una encuesta que tiene en cuenta subgrupos poblacionales (es decir, siempre) es indispensable conocer en qué se diferencian cada uno de ellos, pues de esto depende que la investigación arroje resultados confiables mediante el planteamiento de la mejor estrategia de muestreo.

Asumiendo que $U_1, \ldots, U_g, \ldots, U_G$ denotan los subgrupos poblacionales tales que $\bigcup_{g=1}^{G} U_g = U$, y siendo N_g el tamaño absoluto del subgrupo U_g, entonces se tiene que $\sum_{g=1}^{G} N_g = N$. A partir de las anteriores definiciones, se pueden plantear algunas diferencias y similitudes entre cada uno de ellos. A continuación se resumen:

- *Dominios de interés*: este tipo de subgrupos poblacionales son aquellos para los cuales se requieren estimaciones separadas. Estos requerimientos se planean en la etapa de diseño para asegurar que el diseño de la muestra sea tal que al momento de la recolección de la información exista una cobertura apropiada en cada uno de los dominios de interés. Por lo general, lo anterior sólo se puede lograr ampliando el tamaño de muestra (n), puesto que el marco de muestreo no informa acerca de la pertenencia de los individuos a los dominios de interés. Un aspecto importante de esta clase de subgrupos poblacionales es que el número de individuos en la muestra que pertenecen a un dominio de interés es siempre aleatorio, y para algunos dominios particulares puede llegar a ser muy pequeño. Por otro lado el tamaño absoluto de cada dominio no se conoce ni antes de la etapa de diseño ni después de la etapa de estimación. Un ejemplo claro de estos subgrupos es la condición de ocupación, la condición de pobreza, la rama de actividad, entre otros.

- *Estratos*: cuando el marco de muestreo permite conocer la pertenencia de todos los individuos de la población a un subgrupo poblacional, se dice que esta clase de subgrupos se llaman estratos. Más aun, cuando se sabe que la característica de interés tiene un comportamiento distinto en cada uno de los estratos y se planea un diseño de muestreo que tenga en cuenta este aspecto mediante la selección aleatoria de unidades en cada uno de los estratos, se dice que el diseño de muestreo es estratificado. El aspecto fundamental de esta clase de subgrupos poblacionales es que el conocimiento de la pertenencia de los individuos a los estratos se incorpora en la etapa de diseño de la muestra. Nótese que a diferencia de los dominios, en los estrato se conoce el tamaño poblacional y se controla el tamaño de muestra antes de la etapa de estimación. Un ejemplo claro de estos subgrupos son las zonas urbanas o rurales, regiones y municipios.

- *Postestratos*: la propiedad que caracteriza a este tipo de subgrupos poblacionales es que aunque en la etapa de diseño el tamaño del postestrato es conocido, se desconoce el número de individuos que pertenecerán al postestrato en la muestra seleccionada. Un ejemplo claro de estos subgrupos son los grupos etarios, el sexo o la etnia que, si bien no siempre son utilizadas en la fase de diseño, sí se utilizan sus proyecciones demográficas en la fase de análisis para analizar mejorar la eficiencia de los estimadores. Al respecto Särndal et al. (2003) afirman que existen dos situaciones en las cuales se presenta esta situación, llamada comúnmente postestratificación:

- Cuando el marco de muestreo es tal que se conoce la pertenencia de todos los elementos a los subgrupos poblacionales pero el investigador decide no utilizar esta información en la etapa de diseño. Las razones para esto son diversas pero principalmente se decide obviar este tipo de información por practicidad logística. Una vez que se ha realizado la selección de la muestra, se observa la característica de interés y_k en los individuos de la muestra. El investigador decide utilizar la información auxiliar de pertenencia a los postestratos en la etapa de estimación para mejorar la eficiencia de la estrategia de muestreo, en particular del estimador propuesto.
- Mediante alguna fuente de información confiable se conocen los tamaños absolutos N_g de cada subgrupo poblacional aunque se desconoce la pertenencia de los individuos a los subgrupos, pues el marco de muestreo presenta esta deficiencia. Después de la etapa de diseño, se observa la característica de interés y se pregunta acerca de la pertenencia de los individuos seleccionados en los postestratos de tal forma que en la etapa estimación se utiliza esta información para mejorar la eficiencia de los estimadores de los parámetros de interés.

Es posible afirmar que el análisis apropiado de una encuesta no puede pasar por alto las diferencias de estos subgrupos. Es más, el diseño y rediseño de las encuestas se basan fundamentalmente en la búsqueda de estos subgrupos en la población. En todas las encuestas de hogares que se planean en América Latina se busca investigar fenómenos asociados a subgrupos poblacionales que se encuentran dispersos en la geografía de los países y, en general, se podría especificar los siguientes aspectos:

1. El tamaño de muestra de una encuesta casi siempre se basa en la incidencia de un fenómeno que clasifica a la población en algún dominio de interés.
2. Este tamaño de muestra se reparte entre los diferentes estratos geográficos para mejorar la eficiencia del levantamiento de la información y del diseño de muestreo.
3. En muchas ocasiones, las proyecciones demográficas sobre los post-estratos son utilizadas en la fase de estimación para mejorar la precisión de los estimadores.

I. El estimador de Horvitz-Thompson para totales y tamaños poblacionales

A. Estimación para totales

La mayoría de indicadores sociales a nivel nacional pueden verse como funciones de totales de una o más variables de interés. Por ejemplo, si el interés está en estimar un total $t_y = \sum_U y_k$, el estimador de Horvitz-Thompson (HT) provee una metodología que induce insesgamiento.

$$\hat{t}_{y,\pi} = \sum_s d_k y_k$$

En donde la muestra s hace referencia al subconjunto de la población que fue seleccionado siguiendo un diseño de muestreo probabilístico que induce los pesos de muestreo d_k, los cuales expanden el valor de la variable de interés y_k para el k-ésimo individuo. Nótese que d_k es el inverso multiplicativo de la probabilidad de inclusión de k-ésimo individuo en la muestra, $d_k = \pi_k^{-1}$.

Como se verá más adelante, en presencia de esquemas de estratificación y selección de conglomerados y varias etapas, esta probabilidad resulta ser el producto de las probabilidades condicionales que surgen en los subsecuentes procesos de selección probabilística. Por tanto, el peso final de muestreo resulta ser por lo general una multiplicación de factores de expansión en cada etapa o fase del esquema de muestreo. En general, este estimador toma diferentes formas a medida que el diseño de muestreo cambia. A continuación se presenta una lista no exhausitva de algunas de los diseños más importantes en la teoría de muestreo para encuestas de hogares.

Muestreo aleatorio simple

En este caso, las probabilidades de inclusión son equivalentes para cada unidad incluida en la muestra,

$$\pi_k = \frac{n}{N}$$

Por tanto el estimador toma la siguiente forma:

$$\hat{t}_{y,\pi} = \frac{N}{n} \sum_s y_k$$

Muestreo proporcional al tamaño

Este diseño de muestreo induce probabilidades de inclusión proporcionales al tamaño de una característica de información auxiliar (estrictamente positiva) disponible en el marco de muestreo; por tanto las probabilidades de inclusión obedecen la siguiente relación:

$$\pi_k = \frac{n\ x_k}{t_x} \qquad\qquad 0 < \pi_k \leq 1$$

Por tanto el estimador toma la siguiente forma:

$$\hat{t}_{y,\pi} = t_x\ \sum_s \frac{y_k}{n\ x_k}$$

Por último, no es cierto que la asignación de probabilidades desiguales en las unidades de muestreo induzca sesgo en la encuesta Por ejemplo, cuando se utiliza el estimador de expansión (Hansen-Hurwitz, para el caso de muestreos con reemplazo - Horvitz-Thompson, en muestreos sin reemplazo) el sesgo es nulo bajo estas condiciones. Sin embargo, si se utilizara un estimador que no haga uso de los factores de expansión, la inferencia sí estaría sesgada. Por ende, lo natural es que si el diseño es con probabilidades desiguales, éstas se utilicen dentro de un estimador que considere esta desigualdad; luego sería incorrecto que se utilice otro estimador diferente al estimador de expansión.

Muestreo estratificado

Si $\hat{t}_{yh,\pi}$ estima insesgadamente el total de la característica de interés t_{yh} del estrato h, entonces un estimador insesgado para el total poblacional t_y está dado por

$$\hat{t}_{y,\pi} = \sum_{h=1}^{H} \hat{t}_{yh,\pi}$$

Por ejemplo, para un diseño de muestreo aleatorio estratificado, las probabilidades de inclusión de primer orden están dadas por:

$$\pi_k = \frac{n_h}{N_h} \quad \text{si } k \in U_h$$

En este caso, siendo un s_h la muestra seleccionada en el estrato U_h, el estimador insesgado del total t_y está dado por

$$\hat{t}_{y,\pi} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in s_h} y_k$$

Muestreo de conglomerados

En el esquema general del muestreo por conglomerados, se utiliza un diseño de muestreo específico para la selección de los conglomerados en la muestra. La probabilidad de que el k-ésimo elemento, sea incluido en la muestra s es idéntica a la porbablidad de inclusión del conglomerado al que pertenece π_{Ii}; es decir

$$\pi_k = \pi_{Ii} \quad \text{si } k \in U_i$$

Si se asume que la población está dividida en N_I conglomerados y se selecciona una muestra de conglomerados s_I de tamaño n_I, entonces para un diseño de muestreo aleatorio de conglomerados, el estimador de HT del total poblacional está dado por

$$\hat{t}_{y,\pi} = \frac{N_I}{n_I} \sum_{s_I} t_{yi}$$

En donde t_{yi} hace referencia al total de la característica de interés en el conglomerado U_i. Como se mencionó en los capítulos anteriores, definir los conglomerados con tamaños muy desiguales redunda en un aumento significativo de la varianza del estimador; es por esto que, en encuestas de hogares, se intenta crear conglomerados acotados, a nivel de manzana, o vereda. Esta es una práctica muy pertinente, puesto que la varianza del estimador de expansión estará en función de la varianza de los totales de los conglomerados; si existe una alta variación en los tamaños, habrá también una alta variación en los totales y, por consiguiente, la varianza del estimador será alta. De otra forma, si se tiene conocimiento de una característica de información auxiliar a nivel de los conglomerados (medida de tamaño), es posible hacer uso de esta información del marco para reducir la varianza en el estimador.

Muestreo en dos etapas

Bajo este diseño la probabilidad de inclusión de primer orden del k-ésimo elemento está dada por

$$\pi_k = Pr(k \in s) = Pr(k \in s_i | i \in s_I)\ Pr(i \in s_I) = \pi_{k|i}\ \pi_{Ii}$$

En donde s_i corresponde a la submuestra de elementos seleccionada en el conglomerado U_i. En particular, cuando el diseño de muestreo es aleatorio simple en las dos etapas, y para cada unidad primaria de muestreo seleccionada $i \in s_I$ de tamaño N_i se selecciona una muestra s_i de elementos de tamaño n_i, entonces el estimador HT toma la siguiente forma

$$\hat{t}_{y,\pi} = \frac{N_I}{n_I} \sum_{i \in S_I} \frac{N_i}{n_i} \sum_{k \in s_i} y_k$$

Muestreo en dos fases

Este tipo de muestreo selecciona una muestra de elementos s_a en una primera fase en la cual se recolecta información de interés para crear una versión reducida y acotada del marco de muestreo. A partir de esta información, en una segunda fase, se realiza una nueva selección que define una submuestra s, en donde se observa la característica de información auxiliar. Bajo este esquema, la probabilidad de que un elemento esté en la submuestra de la segunda fase s depende de lo que haya sucedido en la muestra de la primera fase s_a; por lo tanto, la probabilidad de inclusión de cualquier elemento en la muestra final no tiene una forma cerrada y es algebraicamente intratable. Por ende, se define el estimador de Horvitz-Thompson condicionado, el cual toma la siguiente forma

$$\hat{t}_{y,\pi^*} = \sum_s \frac{y_k}{\pi_k^*} = \sum_s \frac{y_k}{\pi_{ak}\pi_{k|s_a}}$$

En la anterior expresión, π_{ak} denota la probabilidad de inclusión del elemento en la muestra de la primera fase, mientras que $\pi_{k|s_a}$ denota la probabilidad de inclusión del elemento a la submuestra de la segunda fase, condicionada a que haya sido incluido en la primera fase.

El estimador HT en una encuesta de hogares regular

Suponga un diseño regular en una encuesta de hogares; por ejemplo, asuma que se tiene un esquema estratificado de H estratos, con dos etapas de selección dentro de cada estrato (la primera etapa con selección de UPM dentro del estrato, la segunda con selección de hogares), entonces el peso de muestreo final y el estimador del total estará dado por la siguiente expresión

$$\hat{t}_{y,\pi} = \sum_s d_k y_k = \sum_h \sum_{i \in s_{Ih}} \sum_{k \in s_{hi}} w_{hik} y_{hik}$$

Por ejemplo, si dentro de cada estrato U_h $h = 1, \ldots, H$ existen N_{Ih} unidades primarias de muestreo, de las cuales se selecciona una muestra s_{Ih} de n_{Ih} unidades mediante un diseño de muestreo aleatorio simple; y además, se considera que el sub-muestreo dentro de cada unidad primaria seleccionada es también aleatorio simple, de tal manera que para cada unidad primaria de muestreo seleccionada $U_i \in s_{Ih}$ de tamaño N_i se selecciona una submuestra s_i de elementos de tamaño n_i, entonces la forma final del estimador de Horvitz-Thompson para el total poblacional quedaría expresada de la siguiente manera:

$$\hat{t}_{y,\pi} = \sum_{h=1}^{H} \hat{t}_{yh,\pi} = \sum_{h=1}^{H} \left[\frac{N_{Ih}}{n_{Ih}} \sum_{i \in S_{Ih}} \frac{N_i}{n_i} \sum_{k \in s_i} y_k \right]$$

B. Estimación para tamaños y totales en dominios

En general todas las expresiones para totales son apropiadas para tamaños poblacionales, puesto que la variable $y_k = 1 \ \forall k \in s$. De esta forma, el estimador HT para un tamaño está dado por la suma de los factores de expansión:

$$\hat{N} = \sum_{s} d_k$$

Bajo un diseño regular en una encuesta de hogares, con un esquema estratificado y dos etapas de selección, el estimador del tamaño poblacional estará dado por la siguiente expresión

$$\hat{N} = \sum_{s} d_k = \sum_{h} \sum_{i \in s_{Ih}} \sum_{k \in s_{hi}} w_{hik}$$

Al asumir un diseño de muestreo estratificado bietápico, con selección aleatoria simple en cada etapa, entonces la forma final del estimador de Horvitz-Thompson para el tamaño poblacional quedaría de la siguiente manera:

$$\hat{N}_{\pi} = \sum_{h=1}^{H} \left[\frac{N_{Ih}}{n_{Ih}} \sum_{i \in S_{Ih}} \frac{N_i}{n_i} \sum_{k \in s_i} 1 \right]$$

Como lo afirma Gutiérrez (2016), en muchas investigaciones es necesario llevar a cabo estimaciones sobre la población en general, y también sobre subgrupos de ella (denominados dominios por la subcomisión en muestreo de las Naciones Unidas). La identificación de los dominios se logra una vez la información de los elementos ha sido registrada. Los dominios tienen que cumplir las siguientes características:

1. Ningún elemento de la población puede pertenecer a dos dominios.
2. Todo elemento de la población debe pertenecer a un dominio.
3. La reunión de todos los dominios es la población del estudio.

La estimación por dominios se caracteriza por el desconocimiento previo de la pertenencia de las unidades poblacionales al dominio. Es decir, para conocer cuáles unidades de la población pertenecen al dominio, es necesario realizar el proceso de medición. En primer lugar construir una función indicadora z_{dk} de la pertenencia del elemento al dominio, la cual toma el valor 1, si el elemento k pertenece al dominio U_d ($k \in U_d$), y toma el valor 0, en otro caso. Ahora se puede utilizar los principios del estimador de Horvitz-Thompson para hallar un estimador insesgado del tamaño del dominio U_d, dado por:

$$\hat{N}_d = \sum_{s_d} d_k$$

Al multiplicar la variable de pertenencia z_{dk} por el valor de la característica de interés y_k, se crea una nueva variable y_{dk} dada por $y_{dk} = z_{dk}\ y_k$, y una vez construida es posible definir el estimador insesgado del total de la característica de interés en el dominio U_d, dado por:

$$\hat{t}_{y_d,\pi} = \sum_{s} d_k\ y_{dk} = \sum_{S_d} d_k y_k$$

II. El estimador de Hájek para medias y proporciones

Cuando se quieren estimar medias y proporciones, es muy probable que no se tenga conocimiento exacto del tamaño poblacional. Por ejemplo, para la estimación de indicadores a nivel de hogar en encuestas mensuales, es difícil tener certeza exacta del número de hogares en el país mes a mes. Por esta razón, cuando se definen estimadores de indicadores relativos, es necesario hacer un doble proceso de inferencia: a nivel de la característica de interés que se quiere investigar, y a nivel del tamaño de la población. El enfoque más comúnmente usado es el de Hájek, que define el estimador para una media de la siguiente manera:

$$\hat{\bar{y}}_s = \frac{\hat{t}_y}{\hat{N}} = \frac{\sum_s d_k\ y_k}{\sum_s d_k}$$

Para el caso de la estimación de una proporción P_d, el estimador de Hájek toma la siguiente forma:

$$\hat{P}_d = \frac{\hat{N}_d}{\hat{N}} = \frac{\sum_s d_k\ z_{dk}}{\sum_s d_k} = \frac{\sum_{s_d} d_k}{\sum_s d_k}$$

Para el caso de la estimación de una media en una subpoblación, como por ejemplo la media del gasto en el área urbana, el estimador de Hájek puede escribirse de la siguiente manera:

$$\hat{\bar{y}}_d = \frac{\hat{t}_{y_d}}{\hat{N}_d} = \frac{\sum_s d_k\ y_k\ z_{dk}}{\sum_s d_k\ z_{dk}} = \frac{\sum_{s_d} d_k\ y_k}{\sum_{s_d} d_k}$$

En general, este tipo de estimadores se pueden considerar no lineales y sus propiedades estadísticas son complejas, y puesto que tanto el numerador como el denominador son variables aleatorias, es necesaria la verificación de algunos supuestos que tienen que ver con el tamaño de la población y de la muestra (Gutiérrez, 2016). En particular, bajo un diseño de muestreo que plantee un esquema estratificado y tres etapas de selección, el estimador de la media poblacional estará dado por la siguiente expresión:

$$\hat{\bar{y}}_s = \frac{\sum_h \sum_{i \in s_{Ih}} \sum_{j \in s_{hi}} \sum_{k \in s_{hij}} w_{hijk}\ y_{hijk}}{\sum_h \sum_{i \in s_{Ih}} \sum_{j \in s_{hi}} \sum_{k \in s_{hij}} w_{hijk}}$$

Las demás expresiones para los estimadores de proporciones o medias en una subpoblación, bajo un diseño de muestreo regular, pueden ser fácilmente derivadas siguiendo los principios expuestos anteriormente.

III. Otros estimadores de muestreo

En presencia de información auxiliar es posible mejorar la eficiencia de la estimación acudiendo a diferentes formas funcionales que estiman el total; por ejemplo, con el estimador de razón:

$$\hat{t}_{y,r} = t_x \frac{\sum_s d_k y_k}{\sum_s d_k x_k}$$

En donde t_x denota el total poblacional, que se supone conocido para toda la población, de una variable auxiliar x que es preguntada en la encuesta de hogares. Por supuesto, en el análisis de este tipo de encuestas es común realizar inferencias sobre parámetros que tienen una forma no lineal. Uno de los más básicos es la razón poblacional $R_U = t_{y_1}/t_{y_2}$ cuya estimación se lleva a cabo estimando ambos componentes de la fracción

$$\hat{R} = \frac{\hat{t}_{y_1}}{\hat{t}_{y_2}} = \frac{\sum_s d_k y_{1k}}{\sum_s d_k y_{2k}}$$

Como se indicó anteriormente, la estimación de un promedio poblacional $\bar{y}_U = t_y/N$, se lleva a cabo de forma eficiente estimando el tamaño de la población y se puede ver como un caso particular de la estimación de una razón. Por otra parte, las encuestas de hogares con diseños panel o rotativos, tienen un mayor interés en la estimación del cambio de indicadores en dos periodos tiempo $\Delta = t_{y^{(t)}} - t_{y^{(t-1)}}$. Nótese que un estimador de este parámetro está dado por

$$\hat{\Delta} = \hat{t}_{y^{(t)}} - \hat{t}_{y^{(t-1)}}$$

Además, es posible mejorar la estimación del total actual $t_y^{(t)}$ al tener en cuenta la información inducida por el traslape de la encuesta en el segundo periodo, así:

$$\tilde{t}_{y^{(t)}} = \alpha \hat{t}_{y^{(t)}} + (1-\alpha)(\tilde{t}_{y^{(t-1)}} + \hat{\Delta})$$

En donde $0 < \alpha < 1$. Por otro lado, si el interés está en estimar algunas características asociadas con la pobreza, es posible utilizar estimadores más complejos. Siendo y_k el ingreso del individuo k y l el umbral de pobreza, entonces el siguiente estimador puede ser utilizado

$$\hat{F}_\alpha = \frac{1}{N} \sum_{k \in s} d_k \left(\frac{l - y_k}{l} \right)^\alpha I(y_k < l)$$

En donde $I(y_k < l)$ es una variable indicadora que toma el valor uno si $y_k < l$ o cero, en cualquier otro caso. Note que si $\alpha = 0$, se tiene una estimación de la incidencia de la pobreza y si $\alpha = 1$, se obtiene un estimación de la brecha de la pobreza (Foster et al., 1984).

La selección del estimador está altamente relacionada con el diseño de la encuesta. Por ejemplo, si se pretende estimar un indicador para un periodo de tiempo definido, el diseño de la encuesta no debería inducir un esquema de rotación que tenga traslape de hogares, puesto que la correlación del indicador induciría un aumento de su varianza y por ende pérdida de eficiencia. Sin embargo, si se desea estimar el cambio del indicador entre dos periodos de tiempos, es necesario contar con un esquema de rotación que asegure un tamaño de muestra suficiente para estimar con precisión este cambio. Cochran (1977, sección 12.13) afirma que, cuando el interés se centra tanto en la estimación del indicador en el periodo actual como en la estimación del cambio entre periodos, es recomendable tener una tasa de traslape de 2/3, 3/4 o 4/5 de una ronda a otra.

IV. Estimadores de calibración

La calibración se ha establecido como un importante instrumento metodológico en la producción de grandes masas de estadísticas (Särndal, 2007). Esta metodología integra información auxiliar en las estimaciones de la encuesta, no solo para garantizar la consistencia con las cifras oficiales reportadas por los INE, sino para hacer más eficiente el proceso de estimación. Gutiérrez (2016) da una breve descripción de este método:

1. Suponga que se tiene acceso a un vector de información auxiliar, $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{pk})$, de p variables auxiliares, el cual es conocido para los individuos seleccionados en la muestra.
2. Además, por registros administrativos u otras fuentes de confianza, se tiene el conocimiento del total del vector de información auxiliar $\mathbf{t_X} = \sum_{k \in U} \mathbf{x}_k$.
3. El propósito del estudio es estimar el total de la característica de interés incorporando la información auxiliar, dada por $\mathbf{x}_k (k \in s)$.
4. Se requiere que los pesos resultantes cumplan con la siguiente restricción
$$\sum_{k \in s} w_k \mathbf{x}_k = \mathbf{t_X}$$
la cual es conocida como la ecuación de calibración.
5. El resultado de la calibración es un nuevo conjunto de pesos w_k que son muy cercanos al inverso de la probabilidad de inclusión del k-ésimo elemento $d_k = 1/\pi_k$

En general, en América Latina, la estrategia de estimación utilizada por los INE recurre a la metodología de calibración sobre proyecciones poblacionales en los dominios de representatividad de la encuesta. Por ejemplo, departamento, zona urbana, zona rural, sexo y/o grupos de edad. Algunas ventajas de utilizar estos procedimientos es que las estimaciones tendrán un sesgo despreciable, y los errores estándares serán más pequeños al compararlos con los del estimador de Horvitz-Thompson; de esta forma se crea un sistema de ponderación que reproduce la información auxiliar disponible y que es eficiente al momento de estimar cualquier característica de interés en la encuesta. Esta coherencia entre las cifras oficiales y las que la encuesta puede producir hace que sea preferible el uso de los estimadores de calibración.

En un encuesta de hogares las restricciones de calibración pueden establecerse sobre características de hogares y características de personas al mismo tiempo. De esta forma, por ejemplo, es posible calibrar sobre las proyecciones demográficas de personas y al mismo tiempo controlar las estimaciones del número de hogares en el país de manera conjunta. Estevao and Särndal (2006) discuten una amplia variedad de casos en donde se calibra conjuntamente en distintos niveles de desagregación sobre diferentes esquema de muestreo. Por ejemplo, para la *Encuesta Continua de Empleo* de Bolivia la calibración está inducida por una post-estratificación sobre los tamaños poblacionales de los cruces resultantes entre las variable Departamento (hay 9 departamentos), Zona (rural y urbano) y PET (con dos categorías: mayor o igual a 10 años y menor de 10 años).

En resumen, utilizar este tipo de estimadores garantiza una *consistencia estética*, puesto que es deseable que las estimaciones puntuales de las encuestas coincidan con los conteos censales, proyecciones poblacionales, registros estadísticos o registros administrativos. Además, existirá un *aumento de la precisión*, porque en la búsqueda de la mejor estrategia de muestreo, el estadístico quiere obtener cifras precisas y confiables que induzcan intervalos de variación angostos y menores errores de muestreo. Por último, si existe una integración adecuada de la información auxiliar, se *disminuye el sesgo* generado por la ausencia de respuesta (debido a los individuos) o por la falta de cobertura (debido a los defectos del marco de muestreo).

A. Ganancia en eficiencia

Para mostrar cómo los estimadores de calibración inducen menores varianzas que los estimadores comunes, se planeó el siguiente experimento de simulación empírica:

1. Se generaron cuatro conjuntos de datos que guardan cierto tipo de relación específica entre la variable de interés y las variables de información auxiliar.
2. Se utilizó la metodología de calibración y se compararon, de forma empírica, para mil iteraciones, las medidas de variabilidad.

Para el primer conjunto de datos, se supuso que existe una relación lineal entre la característica de interés y una variable de información auxiliar continua. Se nota que existe homoscedasticidad en el modelo y que los residuales tienen un comportamiento coherente. A pesar de que ambos estimadores se muestran insesgados para el parámetro de interés, se nota que el estimador de calibración es menos disperso y más eficiente.

Las cuatro siguientes figuras muestran la relación lineal entra las variables (arriba-izquierda), los resudiales ajustados en un modelo de regesión simple (arriba-derecha), la dispersión (abajo-izquierda) de las estimaciones de Horvitz-Thompson (puntos grises) y de las estimaciones de calibración (puntos negros), así como la distribución empírica (abajo-derecha) del estimador de Horvitz-Thompson (línea gris) y del estimador de calibración (línea negra), en donde la línea vertical indica el valor del parámetros poblacional.

El segundo conjunto de datos asume que existe una relación lineal entre la característica de interés y una variable de información auxiliar continua. Se nota que existe heteroscedasticidad en el modelo y los residuales lo muestran. A pesar de que ambos estimadores se muestran insesgados para el parámetro de interés, el estimador de calibración es un poco más eficiente que el de Horvitz-Thompson.

El tercer conjunto de datos asume que existe una relación cuadrática entre la característica de interés y una variable de información auxiliar continua. Al utilizar un estimador de calibración lineal, los residuales muestran un comportamiento inapropiado. Sin embargo, ambos estimadores se muestran insesgados para el parámetro de interés, pero el estimador de calibración es más eficiente que el de Horvitz-Thompson.

El último conjunto de datos asume que existe una relación logística entre la característica de interés y una variable de información auxiliar dicotómica Al utilizar un estimador de calibración lineal, los residuales muestran un comportamiento inapropiado. Ambos estimadores se muestran insesgados para el parámetro de interés e igual de eficientes.

Es posible demostrar que la razón entre la varianza del estimador de calibración con la varianza del estimador de Horvitz-Thompson está supeditada al coeficiente de determinación R^2_ξ en un modelo de regresión lineal simple ξ entre la característica de interés y la información auxiliar.

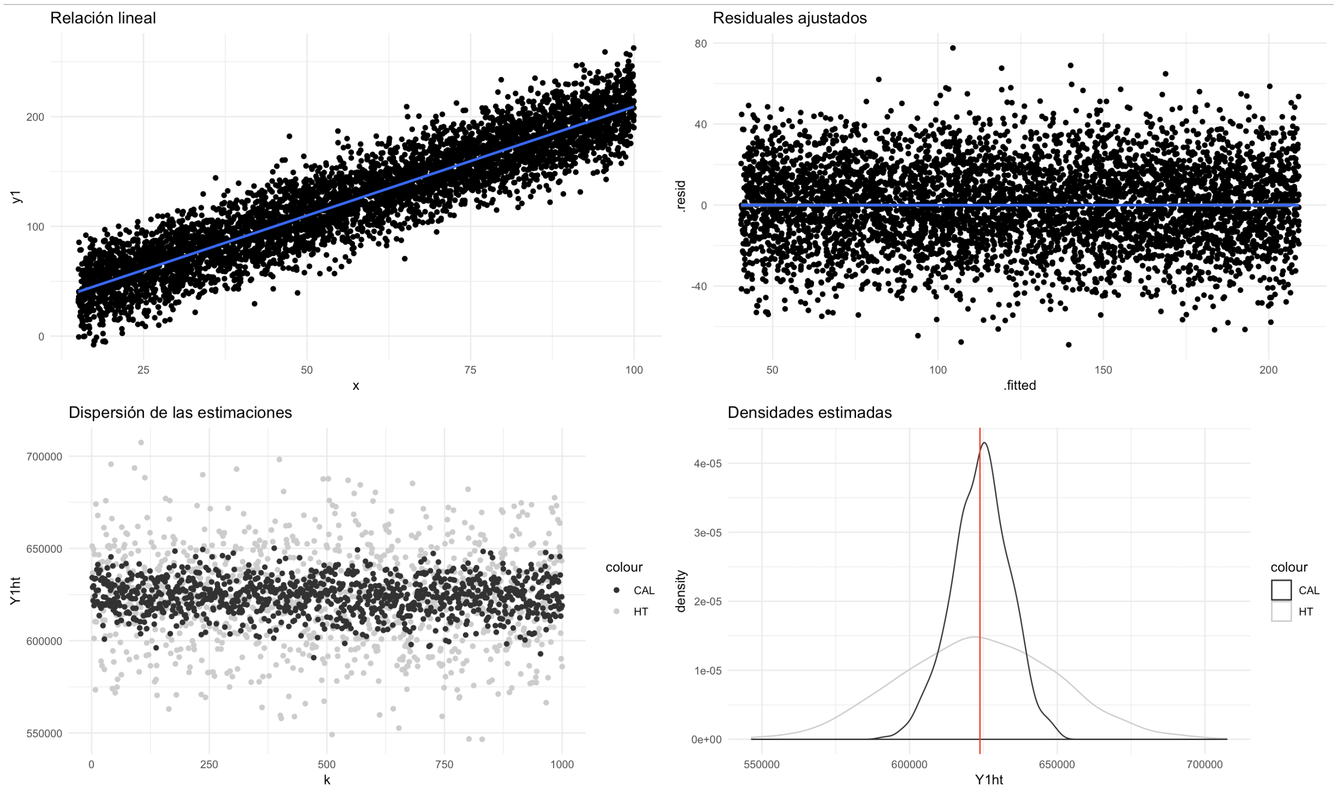

Figura 9.1: *Comportamiento del estimador de calibración en una relación de dependencia lineal*

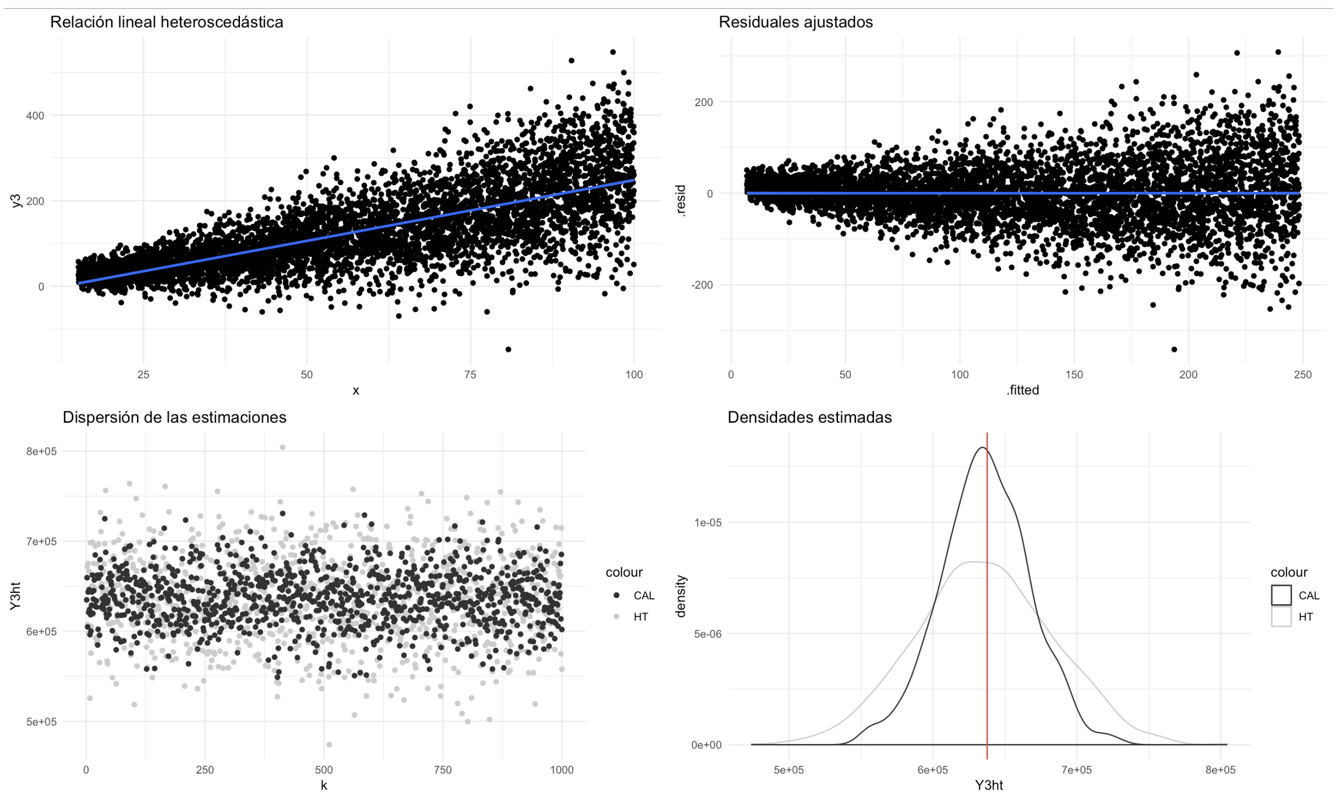

Figura 9.2: *Comportamiento del estimador de calibración en una relación de dependencia lineal con heteoscedasticidad*

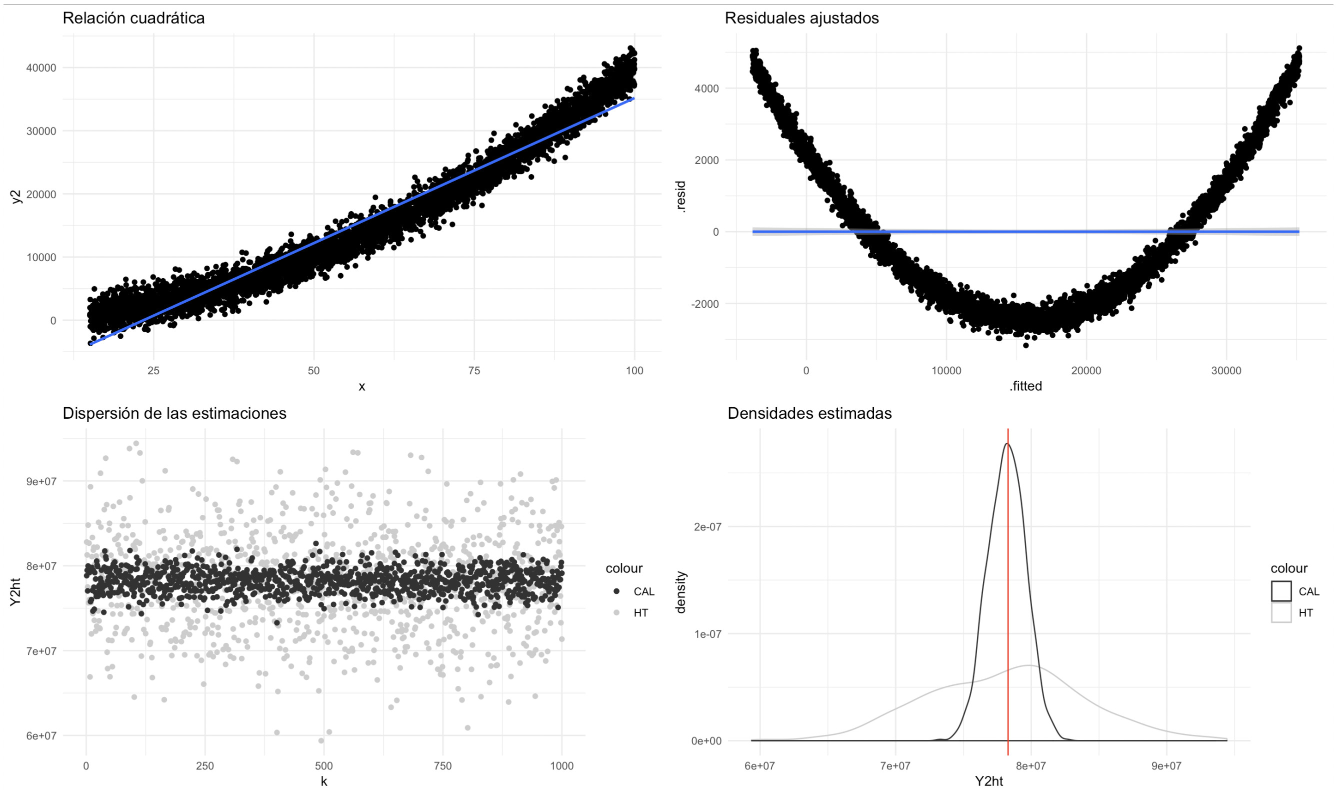

Figura 9.3: *Comportamiento del estimador de calibración en una relación de dependencia cuadrática*

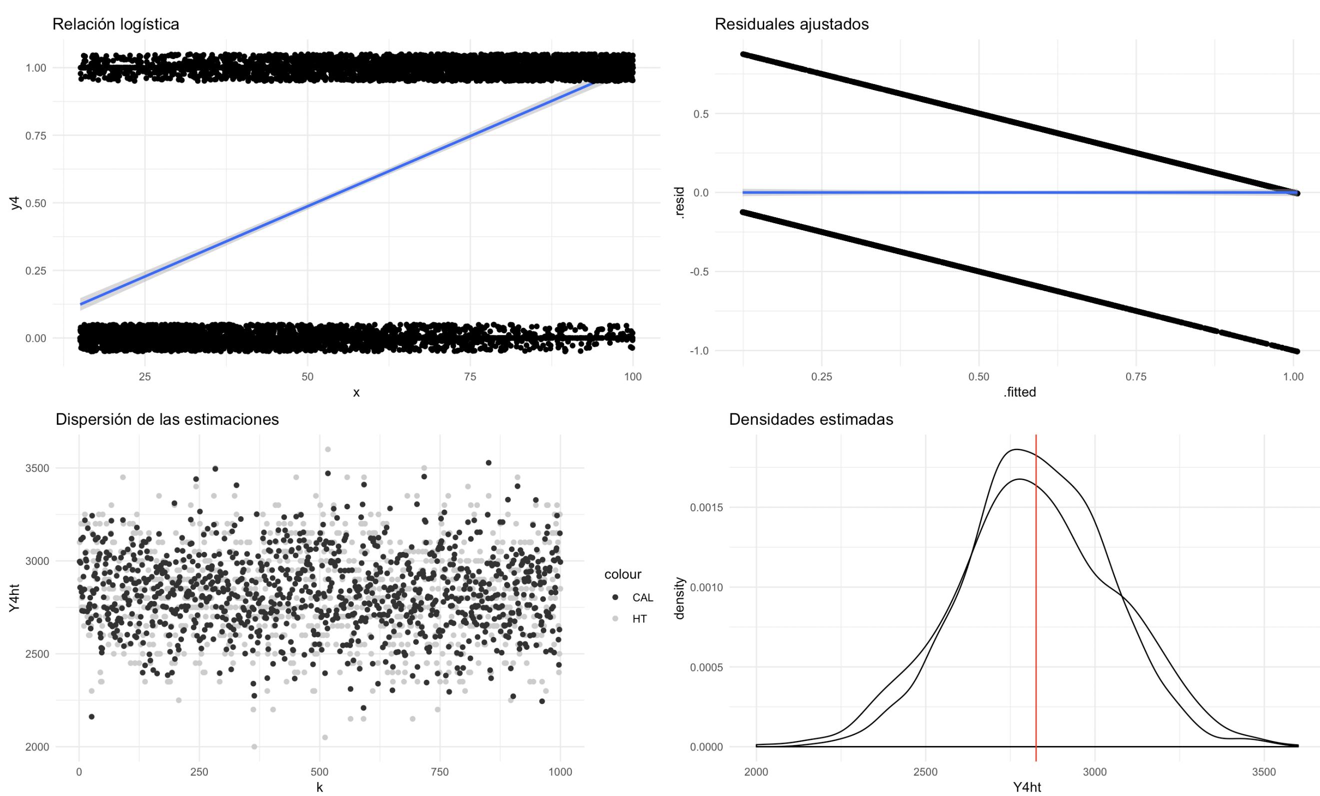

Figura 9.4: *Comportamiento del estimador de calibración en una relación de dependencia logística*

$$\frac{Var(\hat{t}_{y,cal})}{Var(\hat{t}_{y,HT})} = (1 - R^2_\xi + o(\sqrt{n})) \approx (1 - R^2_\xi)$$

Por ende, usar la metodología de calibración supone casi siempre una ganancia en la eficiencia de la estrategia de muestreo.

B. Diferentes formas del estimador de calibración

La calibración es un ajuste que se realiza a los pesos de muestreo con el propósito de que las estimaciones de algunas variables de control reproduzcan de forma perfecta los totales poblacionales de estas variables. Sin embargo, es necesario tener en cuenta las diferencias entre los métodos de calibración, que en general corresponderán con el nivel de desagregación de información auxiliar:

1. Calibración con variables continuas, que es el caso en donde la calibración se realiza con totales de variables continuas como ingreso o gasto, entre otras.
2. Post-estratificación con variables categóricas, que representa el caso en donde la calibración se realiza con los tamaños poblacionales (basados en proyecciones demográficas o registros administrativos) de subgrupos de interés.
3. *Raking* con variables categóricas, que se define como una calibración sobre los tamaños marginales de tablas de contingencia de subgrupos de interés. A diferencia del caso anterior, esta calibración no tiene en cuenta los tamaños de los cruces, sino solo los tamaños marginales; por ende, este método induce menos restricciones.

Postestratificación

La postestratificación es una de las técnicas más usadas para el ajuste de los pesos de muestreo vía calibración. Este ajuste requiere la definición de categorías poblacionales. Por ejemplo, personas en un determinado grupo de edad, en cierta región y de cierta raza. Este método se implementa dentro de cada uno de los cruces inducidos por las covariables de interés (edad, región, raza). Nótese que, es necesario tener acceso a la información auxiliar a nivel de *todos* los cruces definidos por los subgrupos. Usualmente corresponden a proyecciones demográficas. En este caso, la suma de los pesos ajustados reproducirán con exactitud los tamaños poblacionales en cada cruce.

Las categorías formadas para definir los pesos de muestreo se conocen como *post-estratos* puesto que son definidas después de que la muestra es seleccionada y los datos son recolectados. Esta es una ventaja pues estas variables no necesariamente participan en la planificación del diseño de muestreo. Por ejemplo, en una encuesta de hogares es difícil estratificar por raza, edad, sexo o ciclo educativo alcanzado. Como se conoce que estas variables pueden estar correlacionadas con la pobreza, el ingreso o la ocupación, sería una buena idea contemplarlas en la calibración. Suponiendo que se definen cuatro categorías para la raza, dos para sexo, cinco para la edad, entonces habrían 40 post-estratos y sus respectivas restricciones.

De esta manera, siendo $g = 1, \ldots, G$ el indicador del cruce poblacional (post-estrato), el estimador de postestratifiación queda definido de la siguiente manera:

$$\hat{t}_{y,pos} = \sum_{g=1}^{G} \frac{N_g}{\hat{N}_g} \hat{t}_{y_g}$$

En donde N_g corresponde al tamaño poblacional del post-estrato, $\hat{N}_g = \sum_{s_g} d_k$ y $\hat{t}_{y_g} = \sum_{s_g} d_k y_k$. Por lo anterior, el factor de expansión de estimador postestratificado queda definido como sigue:

$$w_k = d_k \frac{N_g}{\hat{N}_g} \qquad (k \in s_g)$$

Note que d_k corresponde al peso inducido por el diseño de muestreo, corregido por los ajustes de elegibilidad y por la ausencia de respuesta, los cuales serán presentados en los capítulos posteriores del documento.

Por último, se debe considerar que la cantidad de postestratos en la calibración está inducido por la cantidad de interacciones en las variables auxiliares. En algunos casos, es posible encontrar cientos de interacciones. Aunque los tamaños de los postestratos se reproduce sin error, esto puede disminuir la eficiencia de la calibración en las variables de interés. Es decir, muchas variables e interacciones hacen que las estimaciones sean inestables, sobre todo si existen cruces con celdas vacías. Es posible que el efecto de las interacciones influencie la creación de los nuevos pesos calibrados y se tengan datos atípicos en los pesos de calibración resultantes.

Raking

¿Qué sucede si los conteos poblacionales (información auxiliar) no están disponibles para todos los cruces de las variables de calibración? Es posible que los agregados poblacionales de las variables provengan de distintas fuentes y no se pueda llegar a nivel de cruce. En este caso, es factible calibrar los marginales de la tabla cruzada, sin necesidad de calibrar todas sus entradas. En este caso, el número de restricciones decrecería con respecto a la postestratifación, pues se sumaría el número de categorías, mientras que en la postestratificación se multiplican. En el escenario anterior, en donde asumimos cuatro categorías para la raza, dos para sexo, cinco para la edad, entonces habrían únicamente 11 restricciones.

Para ajustar los marginales de la tabla cruzada, es necesario realizar un procedimiento iterativo (IPFP), el cual no tiene una escritura cerrada (Gutiérrez, 2016, capítulo 10). Por ejemplo, si el *raking* es de dos marginales, se ajustan primero las filas, luego las columnas y así sucesivamente hasta alcanzar la convergencia de los pesos calibrados y el procedimiento se detiene cuando se alcanza una tolerancia prefijada. Sin pérdida de generalidad, bajo este enfoque, los pesos calibrados se escriben de la siguiente manera:

$$w_k = d_k \times exp(u_h) \times exp(v_g)$$

En donde u_h es una función de los totales marginales de las filas de la tabla cruzada y v_g es una función de los totales marginales de las columnas. El *raking* permite utilizar variables que pueden ser predictoras de las variables de interés o explicar la probabilidad de responder del hogar (o persona), además de paliar los efectos nocivos que las bajas tasas de cobertura del marco de muestreo conllevan sobre la inferencia.

C. La calibración como un cambio de paradigma

Särndal (2007) concluye que existen algunas ideas sobre las cuales vale la pena profundizar un poco más. A continuación se reproducen las ideas de Gutiérrez (2016) sobre estos criterios para enfatizar el uso práctico de los estimadores de calibración:

1. La calibración tiene un vínculo íntimo con la práctica. El uso de los métodos de ponderación de las agencias que manejan las estadísticas oficiales es una costumbre que empezó con la ponderación de unidades mediante el inverso de su probabilidad de inclusión y siguió con las ponderaciones surgidas del enfoque de post-estratificación. Las ponderaciones de calibración extienden las anteriores ideas. La calibración es reciente como término en el muestreo, pero no lo es como técnica para producir ponderaciones. Algunos autores derivaron estas ponderaciones con el argumento de que deberían diferir de la manera más mínima posible de los pesos originales. Otros autores encontraron las ponderaciones al reconocer que un estimador de regresión lineal podría ser escrito como una suma ponderada de los valores de la característica de interés. De allí surgieron términos tales como ponderación de muestreo, ponderación de regresión y ponderación de caso.

2. En algunas ocasiones el término calibración se refiere a una forma de conseguir estimativos consistentes[1]. Las ecuaciones de calibración imponen esta propiedad de consistencia sobre el vector de ponderaciones; así que, cuando éste se aplica a las variables auxiliares el resultado será consistente con los totales de estas variables. Un deseo de promover la credibilidad en las estadísticas oficiales es una razón para que las entidades busquen la consistencia. Cuando la motivación primaria para la calibración no es la concordancia con los totales de la información auxiliar sino el reducir la varianza y el sesgo debido a la ausencia de respuesta entonces el vector de ponderaciones se dice balanceado.

3. El enfoque de calibración ha ganado popularidad en las aplicaciones reales debido a que las estimaciones resultantes son fáciles de interpretar y de motivar puesto que están directamente relacionadas a los pesos inducidos por el diseño de muestreo. La calibración sobre los totales conocidos brinda al usuario una forma natural y transparente de estimación. El usuario que entiende la ponderación muestral aprecia el método de calibración puesto que modifica sutilmente los pesos originales, pero al mismo tiempo respeta los totales de la información auxiliar y mantiene el sesgo despreciable. Además, en la mayoría de aplicaciones, la calibración induce un único vector de ponderaciones aplicable a todas las variables involucradas en el estudio. Esta última razón hace que este método sea muy apetecido en las entidades oficiales que manejan encuestas muy extensas.

4. Algunos autores usan la palabra calibración en combinación con otros términos para describir varias direcciones de pensamiento. Por ejemplo, es posible encontrar términos como calibración de modelo, calibración G, calibración armonizada, calibración a un nivel más alto, calibración de regresión, calibración no lineal, calibración super-generalizada, calibración de modelos de redes neuronales, calibración basada en modelos locales polinomiales, entre otras.

5. Si la calibración representa un nuevo enfoque demarcado claramente de sus predecesores, entonces es tiempo de hacer la pregunta: ¿La calibración generaliza las teorías anteriores? ¿La calibración da mejores respuestas a las preguntas de importancia, que los enfoques de estimación anteriores? En la práctica el estadístico encuentra algunos pormenores tales como ausencia de respuestas, deficiencias del marco muestral y errores de medición. Es cierto que algunos procesos como la imputación y la reponderación para no respuestas son ampliamente difundidos y usados en la práctica. Sin embargo queda un sinsabor al utilizar estos métodos

[1]En este apartado la palabra consistente se da en el sentido de la consistencia con los totales de la información auxiliar.

pues no están enmarcados dentro de una teoría exhaustiva de inferencia en poblaciones finitas. La mayoría de artículos teóricos tratan con la estimación de parámetros bajo un mundo ideal, que no existe en la práctica, donde la ausencia de respuesta y otros errores no muestrales están ausentes.

Capítulo 10

Construcción de los factores de expansión

En todas las bases de datos de las encuestas de hogares se encuentra una columna que contiene los pesos de muestreo o factores de expansión. Con esta columna se realizan todos los análisis requeridos en la encuesta, desde estimar medias, razones, tamaños y proporciones hasta el ajuste de modelos lineales y no lineales.

La razón principal por la cual se usan los factores de expansión es para producir estimaciones que reflejen de manera precisa el comportamiento de la población objetivo. El uso correcto de los factores de expansión garantiza que la estimación sea insesgada y consistente, que el error de muestreo sea pequeño, condicionado al diseño muestral y al tamaño de la muestra; además de corregir las deficiencias de cobertura del marco de muestreo.

Los procesos de inferencia estadística establecidos en cualquier encuesta de hogares descansan sobre el principio de representatividad que afirma que es posible seleccionar una muestra y representar con bastante precisión y exactitud la realidad de la población de interés. A su vez, las propiedades estadísticas de la inferencia en encuestas de hogares descansan sobre las probabilidades de inclusión generadas por el diseño de muestreo que se implementó en la encuesta. En general el peso de muestreo d_k asociado a un individuo k en la muestra s es función de la probabilidad de inclusión del individuo, así:

$$d_k = \frac{1}{Pr(k \in s)}$$

Como se mencionó anteriormente, para conservar la estabilidad en los pesos de muestreo, es posible definir diseños de muestreo auto-ponderados, en donde las unidades finales de muestreo tengan las misma probabilidad de inclusión, sin importar el tamaño de la unidad primaria de muestreo que la contiene. Este tipo de diseños es útil porque induce mayor control sobre las estimaciones finales. Además, Valliant and Dever (2017) afirman que los pesos de muestreo se utilizan con el fin de incorporar las probabilidades de selección de las unidades en la muestra, ajustar en casos en los que no se pueda determinar si algunas unidades en la muestra son miembros de la población de interés, minimizar el sesgo causado por la ausencia de respuesta cuando algunas unidades no responden habiendo sido incluidas en la muestra, incorporar información auxiliar externa para reducir los errores muestrales de las estimaciones y compensar cuando la muestra no cubre correctamente a la población de interés.

Es de notar que la conformación de los pesos de muestreo se transforma en un reto metodológico para el investigador, puesto que debe ajustarse a la realidad de la región en donde las poblaciones de los municipios se expanden cada vez más en el sector urbano y los marcos de muestreo de las áreas geográficas se desactualizan con rapidez. Varias soluciones a este problema han sido planteadas (Gambino and Silva, 2009) y todas ellas requieren de esfuerzos económicos, logísticos y técnicos. Por ende, los equipos de los INE (a todo nivel) deben ser flexibles y adecuarse a esta realidad cambiante de la movilidad de las poblaciones, sobre todo en las áreas urbanas.

En condiciones ideales el marco de muestreo debería coincidir plenamente con la población finita. Sin embargo, en general, no es posible contar con una lista de todos los elementos de la población y, en el contexto de las encuestas a hogares, no existe una lista que enumere todos los hogares de un país de manera actualizada, por lo que la práctica estándar es construir el marco de muestreo en varias etapas, seleccionando una muestra de áreas geográficas, realizando un empadronamiento exhaustivo de todos los hogares en las áreas seleccionadas y luego seleccionando hogares. Este esquema de muestreo hace que el marco de muestreo de las encuestas a hogares presente imperfecciones.

Para hacerle frente a las imperfeccciones del marco, Valliant and Dever (2017) recomienda el uso de los códigos de disposición estandarizados por la *American Association for Public Opinion Research* (AAPOR) recomienda tratar la ausencia de respuesta de manera diferenciada y clasificar a cada unidad en la muestra en algunas de las siguientes categorías:

1. ER (*unidades elegibles que fueron respondientes efectivos*): casos elegibles para los cuales se ha recolectado una cantidad suficiente de información.
2. ENR (*unidades eligibles no respondientes*): casos elegibles para los cuales no se recolectó ningún dato o la información fue parcialmente recolectada.
3. IN (*unidades no elegibles*): casos de miembros no elegibles que no hacen parte de la población de interés.
4. UNK (*unidades con elegibilidad desconocida*): casos en donde no se puede conocer si la unidad es elegible o no.

Para construir los factores de expansión de una encuesta se recomienda seguir en este orden los siguientes procesos:

1. Creación de los pesos básicos.
2. Ajuste por elegibilidad desconocida.
3. Descarte de las unidades no elegibles.
4. Ajuste por ausencia de respuesta.
5. Calibración por proyecciones poblacionales y variables auxiliares.
6. Recorte y redondeo de los factores finales (*opcional*).

I. Creación de los pesos básicos

Este primer paso ya ha sido explicado de forma detallada en el capítulo dedicado a la selección de la muestra. Observe que, asociado a cada esquema particular de muestreo, existe una única función que vincula a cada elemento con una probabilidad de inclusión en la muestra. De esta forma:

$$\pi_k = Pr(k \in s)$$

Por lo tanto, el primer paso en la reponderación de los pesos de muestreo, es justamente la creación de los pesos básicos d_{1k} que se definen como el inverso multiplicativo de la probabilidad de inclusión

$$d_{1k} = \frac{1}{\pi_k} \qquad \forall\ k \in s$$

Estos pesos son creados incluso para aquellas unidades que serán excluidas de la muestra porque son no elegibles o porque no proveyeron ninguna información y luego serán modificados convenientemente. La siguiente figura muestra la distribución típica de los pesos originales en una encuesta de hogares. Se recomienda calcular las probabilidades de inclusión (si el muestreo es sin reemplazo) o selección (si el muestreo es con reeemplazo) a medida que avance el muestreo en sus etapas y, de esta forma, siempre confirmar la consistencia de los pesos en cada etapa y de los pesos finales.

A través de las modificaciones posteriores sobre este peso de muestreo, la distribución de los ponderadores irá sufriendo algunos cambios. Si la distribución original de los pesos básicos difiere estructuralmente con la distribución final de los ponderadores, resultante de todos los ajustes debidos a las imperfecciones del marco, entonces las propiedades estadísticas de insesgamiento, consistencia y precisión podrían desvanecerse. Lo anterior implica que el nivel de desactualización del marco de muestreo tiene implicaciones directas en la calidad de la inferencia. Por tanto, si el marco de muestreo es muy imperfecto, los ponderadores finales no inducirán una inferencia precisa.

II. Ajuste por elegibilidad desconocida

El segundo paso consiste en redistribuir el peso de las unidades cuyo estado de elegibilidad es desconocido. Por ejemplo, si la encuesta está enfocada en la población mayor de 15 años y hay personas que no proveen ninguna información acerca de su edad, entonces es necesario distribuir estos pesos. Esta situación también se puede presentar a nivel de hogar cuando no puede ser contactado porque nadie nunca atendió el llamado del encuestador (*nadie en casa*). Se acostumbra a redistribuir los pesos de las unidades con elegibilidad desconocida (UNK) entre las unidades que sí disponen de su estatus de elegibilidad (ER, ENR, IN).

Por consiguiente, si no es posible determinar la elegibilidad de algunas unidades que aparecen en el marco de muestreo, se tendrá una muestra s que contendrá el subconjunto de las unidades *elegibles respondientes* (s_{ER}), el subconjunto de las unidades *elegibles no respondientes* (s_{ENR}), el subconjunto de las unidades *no elegibles* (s_{IN}) y el subconjunto de las unidades con *elegibilidad desconocidad* (s_{UKN}). En este último caso, la elegibilidad de estas unidades se mantiene desconocida, a no ser que de manera arbitraria sean clasificadas como ENR (elegibles no respondientes), o se tenga información auxiliar en el marco de muestreo que permita imputar su estado de elegibilidad.

Se recomienda formar B $(b = 1, \ldots, B)$ categorías[1] basadas en la información del marco de muestreo. Estas categorías pueden ser estratos o cruces de subpoblaciones. Siendo s_b la muestra de unidades en la categoría b (que incluye a ER, ENR, IN y UNK), se define el factor de ajuste por elegibilidad como:

[1] Valliant and Dever (2017) recomienda formar categorías con al menos 50 casos.

$$a_b = \frac{\sum_{s_b} d_{1k}}{\sum_{s_b \cap (s_{ER} \cup s_{ENR} \cup s_{IN})} d_{1k}}$$

Para la categoría b, los pesos ajustados por elegibilidad desconocida para aquellas unidades cuya elegibilidad sí pudo ser establecida (independientemente de su estado de respuesta) estarán dados por la siguiente expresión:

$$d_{2k} = a_b * d_{1k} \qquad \forall\ k \in s_b \cap (s_{ER} \cup s_{ENR} \cup s_{IN})$$

III. Descarte de las unidades no elegibles

En esta etapa, tanto las viviendas con elegibilidad desconocida (UNK) como las viviendas que han cambiado su estado de ocupación y ahora no contienen ningún hogar particular (IN) se retirarán de la población objetivo. Por ende, la base de datos sobre la cual se prosigue el proceso ahora presentará una reducción en el número de unidades. Este tercer paso consiste en ajustar el peso de la etapa anterior de la siguiente manera:

$$d_{3k} = \begin{cases} 0, & \text{si la unidad } k \in (s_{UNK} \cup s_{IN}) \\ d_{2k}, & \text{si la unidad } k \in (s_{ER} \cup s_{ENR}). \end{cases}$$

IV. Ajuste por ausencia de respuesta

En este paso los pesos de los respondientes efectivos (ER) se ajustan para tener en cuenta a los que no respondieron (ENR). Al final del proceso, los pesos de los ER se incrementan para compensar el hecho de que algunas unidades elegibles no proveyeron información. Para el manejo efectivo de la ausencia de respuesta se consideran las siguientes variables aleatorias:

$$I_k = \begin{cases} 1, & \text{si } k \in (s_{ER} \cup s_{ENR}) \\ 0, & \text{en otro caso.} \end{cases}$$

$$D_k = \begin{cases} 1, & \text{si } k \in s_{ER} \\ 0, & \text{si } k \in s_{ENR}. \end{cases}$$

Al suponer que la distribución de las respuestas puede ser estimada, entonces la probabilidad de respuesta (*propensity score*) está dada por

$$Pr[k \in s_{ER} | k \in (s_{ER} \cup s_{ENR})] = Pr[D_k = 1 | I_k = 1] = \phi_k$$

Si el patrón de ausencia de respuesta es completamente aleatorio (en donde la no respuesta no sigue ningún patrón específico) o aleatorio (en donde el patrón de la no respuesta puede ser explicado por un conjunto de covariables $\mathbf{z}$), entonces

$$\phi_k = f(\mathbf{z}_k, \boldsymbol{\beta}) \qquad \forall\ k \in (s_{ER} \cup s_{ENR})$$

De esta forma, si fuese plausible tener acceso a las covariables $\mathbf{z}$ para los individuos elegibles en la muestra, entonces se podría estimar el patrón de ausencia de respuesta mediante la siguiente relación funcional:

$$\hat{\phi}_k = f(\mathbf{z}_k, \hat{\boldsymbol{\beta}}) \qquad \forall\ k \in (s_{ER} \cup s_{ENR})$$

Por otro lado, si el patrón de ausencia de respuesta es no aleatorio (en donde la misma estructura de la ausencia de respuesta es explicada por la variable de interés; por ejemplo cuando en una encuesta de mercado laboral son los desempleados quienes no responden), entonces

$$\phi_k = f(\mathbf{y}_k, \beta) \qquad \forall\ k \in (s_{ER} \cup s_{ENR})$$

En este caso, como no es posible tener acceso a la variables de interés para todos los individuos en la muestra de unidades elegibles (precisamente porque no todos respondieron), entonces no es posible estimar el patrón de ausencia de respuesta y por ende habrán problemas de sesgo. Por otra parte, Kim and Riddles (2012) muestran que es posible utilizar un modelo basado de estimación de las probabilidades de respuesta (*propensity score*). De esta forma, teniendo en cuenta que la probabilidad de que un individuo conteste es $\phi_k = Pr(k \in s_{ER})$, al suponer que existe acceso al vector de información auxiliar $\mathbf{z}_k$ conocido para todo $k \in (s_{ER} \cup s_{ENR})$ es posible estimarla, por ejemplo, por medio de un modelo de regresión logística; esto es,

$$\hat{\phi}_k = \frac{\exp\{\mathbf{z}_k'\hat{\boldsymbol{\beta}}\}}{1+\exp\{\mathbf{z}_k'\hat{\boldsymbol{\beta}}\}} \qquad \forall\ k \in (s_{ER} \cup s_{ENR})$$

donde $\hat{\beta}$ es el vector de coeficientes estimado de la regresión logística. Por tanto, si la ausencia de respuesta no depende de la variable de interés, es posible definir el siguiente estimador insesgado

$$\hat{t}_y = \sum_{k \in s_{ER}} d_{4k} y_k$$

En donde

$$d_{4k} = \frac{d_{3k}}{\hat{\phi}_k} \qquad \forall\ k \in s_{ER}$$

Es posible aumentar la eficiencia del estimador si se crean categorías homogéneas de individuos que tengan la misma probabilidad de responder. En este caso, los valores de las covariables pueden ser usados para crear estas categorías. Por consiguiente, siempre es necesario obtener un conjunto de covariables que esté disponible para respondientes y no respondientes a la vez.

Por ejemplo, considere un escenario simplificado en donde es posible identificar que la probabilidad de responder está relacionada únicamente con las variables edad (5 categorías) y sexo (2 categorías). En este caso, sería posible formar $Q = 10$ ($q = 1, \ldots, Q$) categorías de acuerdo al cruce de estas variables para obtener una estimación de la probabilidad de respuesta en cada clasificación y ajustar el peso de muestreo. De esta manera, siendo s_q la muestra seleccionada en la categoría q, la probabilidad de respuesta en esta categoría se estimaría como:

$$\phi_q = \frac{\sum_{s_{ER} \cap s_q} d_{3k}}{\sum_{s_q} d_{3k}}$$

El nuevo peso ajustado por la ausencia de respuesta estará dado por:

$$d_{4k} = \frac{d_{3k}}{\phi_q} = d_{3k} \frac{\sum_{s_q} d_{3k}}{\sum_{s_{ER} \cap s_q} d_{3k}}$$

En un escenario más complejo, si las probabilidades de respuesta fueron estimadas con un modelo de *propensity score* y, teniendo en cuenta que las predicciones de estas probabilidades varían entre cero y uno, es posible crear clases de individuos (respondientes y no respondientes) con probabilidades similares. En este caso, se asumiría que las unidades dentro de una misma clase tendrán la misma configuración de covariables, o al menos, una probabilidad de respuesta estimada similar $\hat{\phi}_k$. Así, dentro de cada clase, las unidades serían tratadas como si hubiesen sido aleatorizadas al tratamiento (responder) o al control (no responder).

Por lo tanto, el objetivo de este proceso es asegurar que cualquier diferencia en las covariables pueda ser ajustada. Teniendo en cuenta que, si el modelo es adecuado, la estimación $\hat{\phi}_k$ resumiría los efectos de las covariables en la respuesta del individuo, entonces una vez hayan sido creadas las clases es posible realizar el ajuste mediante alguna medida de localización en cada clase y, de esta forma, todos los individuos de una misma clase se ajustarían de la misma manera. Asumiendo que se pueden crear C clases y que s_c es la muestra de n_c unidades elegibles en la clase c ($c = 1, 2, \ldots, C$), entonces es posible utilizar alguna de las siguiente medidas (Valliant and Dever, 2017):

1. Promedio no ponderado:
$$\hat{\phi}_c = \frac{\sum_{k \in s_c} \hat{\phi}_k}{n_c}$$

2. Promedio ponderado:
$$\hat{\phi}_c = \frac{\sum_{k \in s_c} d_{3k} \hat{\phi}_k}{n_c}$$

3. Mediana no ponderada:
$$\hat{\phi}_c = mediana[\hat{\phi}_k] \qquad \forall\ k \in s_c$$

4. Tasa de repuesta no ponderada:
$$\hat{\phi}_c = \frac{\#(s_{ER} \cap s_c)}{n_c}$$

5. Tasa estimada de repuesta:
$$\hat{\phi}_c = \frac{\sum_{s_c \cap s_{ER}} d_{3k}}{\sum_{s_c} d_{3k}}$$

Nótese que, si todas las unidades dentro de una clase tienen la misma probabilidad de responder, entonces la tasa de repuesta no ponderada es la mejor opción. Además, si dentro de las clases las unidades tienen una probabilidad de responder muy disímil, entonces el promedio no ponderado (o ponderado) del PS puede usarse. De la misma manera, la tasa estimada de repuesta puede ser ineficiente si los pesos de muestreo varían demasiado, pero la probabilidad de respuesta es similar en cada clase. Por último, la mediana se considera si la distribución de la probabilidad de respuesta es sesgada.

V. Calibración de los factores de expansión

Después de conformar el sistema de ponderación de pesos de muestreo en la encuesta, es posible calibrar estos pesos con la información auxiliar disponible para cada país, a nivel nacional, por estratos de interés, e incluso por variable continuas sobre las que se tenga interés. Särndal and Lundström (2006) afirman que cuando los estudios por muestreo están afectados por la ausencia de respuesta, es deseable que la estructura inferencial que sustenta la encuesta induzca estimadores con sesgo pequeño o nulo y con errores estándares pequeños. A su vez, durante décadas los INE han optado por preferir los sistemas de pesos que sean capaces de reproducir la información auxiliar disponible[2] y que sean eficiente al momento de estimar cualquier característica de interés en un estudio multipropósito.

De ahora en adelante, y con el fin de simplificar la notación estadística, se denotará indistintamente a la muestra de elegibles respondientes como s. Como se vio en los capítulos anteriores, debido a la construcción teórica de los estimadores de calibración, los pesos calibrados responden a la siguiente restricción

$$\sum_{k \in s} w_k \mathbf{x}_k = \mathbf{t_X}$$

El ejemplo más básico se tiene cuando se desea que los pesos de muestreo reproduzcan con exactitud el tamaño de las regiones N_h de un país, y/o el tamaño del país N. Es así como, utilizar la metodología de calibración (Deville and Särndal, 1992) hace que se cumpla la siguiente ecuación de calibración sobre los nuevos pesos calibrados w_k para todos lo estratos explícitos

$$\sum_{s_h} w_k = N_h$$

Gutiérrez (2016) menciona que esta coherencia entre las cifras oficiales y las que la encuesta puede producir hace que sea preferible el uso de los estimadores de calibración. Las anteriores características son satisfechas al usar el enfoque de calibración que induce una estructura inferencial robusta en presencia de información disponible puesto que reduce tanto el error de muestreo como el error debido a la ausencia de respuesta. Una vez que se ha ejecutado el proceso de calibración, se crean nuevos pesos que, en general, pueden ser escritos como

$$w_k = g_k * d_{4k} \qquad \forall\ k \in s$$

En donde los valores g_k son dependientes de la muestra seleccionada s y de la función de optimización escogida para realizar el proceso de calibración. En general no tienen una forma cerrada, aunque dependiendo de la estructura en la información auxiliar pueden tomar valores particulares. Por lo tanto, el estimador de calibración tomará su forma clásica, dada por:

[2] Por ejemplo, el número de hogares o habitantes en el país.

$$\hat{t}_{y,cal} = \sum_{k \in s} w_k y_k$$

Los estimadores de calibración son *aproximadamente insesgados*, pero la magnitud del sesgo está dada por la siguiente expresión:

$$B(\hat{t}_{y,cal}) = E\left[\sum_{k \in s}(w_k - d_k) y_k\right]$$

Si los nuevos pesos calibrados son cercanos a los pesos originales en todas las posibles muestras, entonces el sesgo será insignificante. Ahora, si el tamaño de muestra es insuficiente no conviene utilizar este tipo de estimadores. Además, se sugiere que el coeficiente de variación del estimador de Horvitz-Thompson para las covariables (inducidas por todos los cruces y celdas considerados) sea menor del 10 % para asegurar que el sesgo de los estimadores de calibración sea despreciable.

Por otro lado, cuando se tienen múltiples variables discretas es posible que el cruce de categorías contenga muy pocas unidades para las cuales se deben ajustar los pesos originales. Esto puede inducir sesgo en algunos subgrupos. Si aún así se decide optar por mantener estas múltiples restricciones de calibración, será necesario hacer un chequeo empírico del ajuste que cada modelo pueda tener con todas las variables de la encuesta.

A. Medidas de calidad en la calibración

Silva (2004) presenta algunas consideraciones al respecto del sesgo que puede generarse al usar esta metodología en las encuestas de hogares y aborda algunos criterios para evaluar la calidad de la calibración. Estas medidas se pueden considerar como protección en contra del sesgo generado por tener demasiadas restricciones. Además, se resalta la importancia de que las variables utilizadas para la calibración sean estimadas de manera precisa por los estimadores clásicos de muestreo. Por ejemplo, si el número de personas en una región es utilizada como una variable de calibración (utilizando como total auxiliar las proyecciones demográficas), entonces el coeficiente de variación del estimador de Horvitz-Thompson sobre esta variable debería ser menor, por ejemplo, al 10 %.

La teoría afirma que entre más variables de calibración se tengan menor será la varianza asociada a las estimaciones (no así el sesgo). Sin embargo, existen problemas computacionales cuando las restricciones que se deben satisfacer son demasiadas. Una primera opción es verificar que no se tengan variables que puedan tener codependencia lineal con otras. Al descartar estas variables es posible conservar una varianza pequeña puesto que se descartan combinaciones lineales de otras variables. Se recomienda hacer un análisis de cuántas variables se deben utilizar en la calibración para optimizar el error cuadrático medio de los estimadores finales en las encuestas de hogares.

En una primera instancia, no sería adecuado utilizar demasiadas restricciones de calibración para satisfacer muchas proyecciones demográficas. Es fácil equivocarse en esta definición. Por ejemplo, si la encuesta es representativa a nivel de departamento (10 niveles), sexo (2 niveles) y edad (4 niveles), entonces podría ser contraproducente utilizar $10 \times 2 \times 4 = 80$ restricciones de calibración y se debería empezar por analizar una estrategia más parsimoniosa con $10 + 2 + 4 = 16$ restricciones de calibración. Nótese que a medida que las desagregaciones sean más profundas, el nivel de error en las proyecciones poblacionales será más grande. Ademas, entre más restricciones haya, más sesgo y varianza se introduce a la estimación. La idea general del proceso es encontrar un número de restricciones parsimonioso que permita tener estimaciones aproximadamente insesgadas con una varianza menor a la generada con los factores de expansión originales.

Por otro lado, si los pesos de calibración resultan ser menores que uno su interpretación puede tornarse difícil (aunque no reviste un problema teórico). El usuario común entiende al factor de expansión como un factor de representatividad: *es la cantidad de veces que una persona se representa a sí misma y a algunas otras más en la población.* Por ende, los pesos negativos o menores que uno no resisten esta interpretación intuitiva y natural. Además, los pesos negativos pueden conllevar a estimaciones negativas para algunos dominios en donde el tamaño de muestra es pequeño, lo cual resulta ser problemático en un contexto en donde todas las variables de estudio son no negativas.

Para garantizar que los pesos se ubiquen en un intervalo determinado, se debe minimizar una distancia que a su vez debe inducir pesos restringidos a este intervalo y que respete las ecuaciones de calibración. Es posible que no se tenga una solución exacta para todas las restricciones de calibración e incluso que el algoritmo de calibración no converja.

Con base en lo anterior, es necesario analizar los pesos g_k en perspectiva en cada dominio, estrato y postestrato de interés. Una buena idea puede ser identificar aquellos g_k que resulten potencialmente grandes o influyentes. Se recomienda postestratificar la muestra, y aplicar la calibración a aquellas unidades en los que los g_k sean estables y usar los pesos originales en el restante conjunto.

Es posible lograr que los pesos de calibración estén restringidos a un espacio predefinido por el usuario, mediante límites (L, U) sobre los g_k. De esta forma, si $w_k \geq 1$ implica $g_k \geq 1$ y por tanto $L = 1$. Se acostumbra a tomar $U > Q_3 + 1.5 * (Q_3 - Q_1)$ en donde Q_3 y Q_1 están dados en términos de la distribución de g_k y corresponden al tercer y primer cuartil, respectivamente.

Si el mecanismo que genera la ausencia de respuesta no es aleatorio (MAR) o completamente aleatorio (MCAR), es posible que los ponderadores de calibración induzcan sesgo en las estimaciones finales. En general, cuando hay ausencia de respuesta es más probable que aparezcan pesos de calibración negativos y que los pesos de calibración no convergieran a los pesos originales. Además, la varianza de los estimadores de calibración no convergerá a los resultados usuales de los estimadores de regresión.

Asumiendo que existen P variables de información auxiliar en las ecuaciones de calibración, Silva (2004) presenta algunas medidas que permiten decidir cuáles escenarios de calibración son los mejores. Asumiendo que $\hat{t}_{x_p,cal}$ es el estimador de calibración para la p-ésima variable de información auxiliar cuyo total poblacional es t_{x_p}, se define el *error relativo promedio* sobre las variables auxiliares

$$M1 = \frac{1}{P} \sum_{p=1}^{P} \frac{|\hat{t}_{x_p,cal} - t_{x_p}|}{t_{x_p}}$$

Se esperaría que $M1$ fuese nulo, si es que la distancia utilizada en la optimización de la calibración es la Ji-cuadrado. Sin embargo, como esta distancia puede arrojar valores negativos para los pesos de muestre, entonces es preferible utilizar otro tipo de distancias que pueden no converger a los totales auxiliares exactamente. En este caso, es preferible escoger aquella distancia que menores valores arroje sobre $M1$, sujeto a que los pesos tenga una interpretación adecuada.

Por otro lado, siendo $\hat{t}_{x,\pi}$ el estimador HT de la p-ésima variable de información auxiliar, se define el *coeficiente de variación relativo promedio* de la siguiente manera

$$M2 = \frac{1}{P}\sum_{p=1}^{P}\frac{\sqrt{Var(\hat{t}_{x,\pi})}}{t_x}$$

Esta medida tiene como intención que el investigador utilice variables de información auxiliar que estén bien representadas en la encuesta. De la misma manera, se define la *proporción de ponderadores extremos* menores a un límite inferior (L) predefinido , o mayores a un límite superior definido (U) predefinido. Estas medidas están respectivamente dadas por

$$M3 = \frac{1}{n}\sum_{k\in s} I(g_k < L)$$

$$M4 = \frac{1}{n}\sum_{k\in s} I(g_k > U)$$

Como los pesos de calibración están definidos como la multiplicación de los pesos ajustados d_k con un ponderador g_k, no sería deseable que hubiese ponderadores extremos, muy alejados de la unidad; puesto que sería indicio de que existe un alejamiento severo entre los pesos inducidos por el diseño y los nuevos pesos de calibración. Teniendo en cuenta que el insesgamiento está garantizado al utilizar los pesos d_k, entonces una proporción de valores g_k muy alejados sería una señal de alarma y podría indicar sesgo.

Siendo $\sigma(g)$ la desviación estandar muestral de los ponderadores g_k y $\bar{g}$ su promedio muestral, otra medida de interés es el *coeficiente de variación de los ponderadores*, la cual está supeditada a la siguiente expresión

$$M5 = \frac{\sigma(g)}{\bar{g}}$$

Una dispersión muy alta de los ponderadores sería indeseable, puesto que indicaría que hay valores influyentes que alejarían a los pesos calibrados de los pesos muestrales. Asimismo, se define la *distancia Ji-cuadrado* entre los pesos de calibración y los pesos originales, dada por:

$$M6 = \frac{1}{n}\sum_{k\in s}\frac{(w_k - d_k)^2}{d_k} = \frac{1}{n}\sum_{k\in s} d_k(g_k - 1)^2$$

Si la calibración fue exitosa, esta medida debe ser pequeña, indicando una cercanía de los pesos calibrados a los pesos originales, que señalaría que el insesgamiento se mantiene. Por otro lado, la *eficiencia* de los estimadores de calibración puede ser calculada con base en la siguiente expresión

$$M7 = \frac{1}{P}\sum_{p=1}^{P}\frac{Var(\hat{t}_{x_p,cal})}{Var(\hat{t}_{x_p,\pi})}$$

Nótese que esta medida es similar al efecto de diseño generalizado, utilizado para evaluar la eficiencia de los escenarios de estratificación. Siguiendo el mismo raciocinio, se quiere que esta medida sea pequeña y siempre menor a uno, indicando que la inferencia estadística al utilizar los pesos calibrados es mayor, que con el estimador HT. Por último, como se indicó en los capítulos anteriores, el *efecto de diseño* debido a la ponderación desigual $DEFF^w$ se define como

$$M8 = 1 + cv^2(w_k)$$

En este caso, es deseable que esta medida sea muy cercana a uno, indicando que la dispersión de los pesos finales está controlada.

B. Calibración integrada para hogares y personas

Una de las preguntas recurrentes en la calibración de encuestas de hogares es el nivel al cual se debería realizar este ajuste. En principio, es posible realizar la calibración al nivel de las personas, o al nivel de los hogares. Cada una de estas opciones trae algunas ventajas y consideraciones que se deben tener en cuenta.

- Calibrar al nivel de los hogares implica que el hogar tendrá unos nuevos pesos que cumplen con las restricciones de calibración, y esos pesos los heredará a las personas que habitan el hogar. De esta forma todas las personas pertenecientes a un mismo hogar tendrán el mismo peso de muestreo, sin importar sus diferencias en composición demográfica. Por ejemplo, hombres, mujeres, menores y mayores de 15 años tendrán el mismo peso de muestreo. Esta propiedad es atractiva puesto que emula el diseño de muestreo que se definió en la fase de planeación. Sin embargo, realizar la calibración a nivel de los hogares hace que dentro de las unidades primarias de muestreo (UPM) los hogares no tengan un peso homogéneo, lo que se distancia de las propiedades del diseño sistemático simple que se usa para la selección de los hogares dentro de las UPM.

- Por otro lado, calibrar a nivel de personas implica que los pesos de muestreo de los hogares también pueden verse alterados, y que los pesos finales de muestreo de las personas sean diferentes dentro de los hogares. De esta forma, de acuerdo a las características de las personas se tendrá un peso diferente. Por ejemplo, es posible que hombres, mujeres, menores y mayores de 15 años **no** tengan el mismo peso de muestreo. Por consiguiente cuando se calibra por personas y se utiliza un filtro sobre esa base de personas para crear una base de hogares, las características observadas de los jefes de hogares influenciarían los pesos de muestreo resultantes.

Dado que la calibración puede inducir factores de expansión diferentes para los miembros de un mismo hogar, es necesario analizar a qué nivel se realiza este procedimiento (persona, hogar). En principio, y debido al diseño de la encuesta, los pesos de muestreo originales son idénticos para todos los miembros de un mismo hogar. Sin embargo, cuando la post-estratificación trata de ajustar los totales de las restricciones de calibración, y debido a que la población no está equitativamente distribuida, entonces de igual manera se presenta un reajuste en los factores de calibración. Podría ser conveniente revisar la metodología de *raking* y su impacto en los pesos de calibración dentro de los hogares.

Por ejemplo, si la calibración se realiza a nivel de personas y se calibra sobre la población en edad de trabajar, esto traerá como consecuencia que los factores de expansión sean diferentes para los miembros de un mismo hogar, puesto que la metodología buscará ajustar los totales de las personas en edad de trabajar y las personas que no están en la fuerza de trabajo de manera independiente.

Por esta razón en la mayoría de hogares, en donde hay personas que son parte de la fuerza de trabajo y personas que no lo son, los pesos de muestreo no serán equivalentes.

En general, la mayoría de encuestas de hogares en la región tienen una naturaleza multipropósito, generando estimaciones de indicadores a nivel de persona (tasa de participación, tasa de desocupación, etc.), y al mismo tiempo, indicadores a nivel de hogar (pobreza monetaria, necesidades básicas insatisfechas y pobreza multidimensional). En este documento se enfatiza la recomendación de disponer de factores de expansión coherentes entre las diferentes unidades de análisis.

Por ejemplo, una práctica común que pone en tela de juicio las propiedades estadísticas del estimador, es generar factores de expansión a nivel de persona y endilgarle el factor de expansión del jefe de hogar al mismo hogar (Alexander, 1987). Esta es una escogencia arbitraria si es que los factores de expansión se han generado mediante una calibración que tenga en cuenta las características de las personas, como por ejemplo edad o sexo. Este acercamiento deliberado no permite sopesar las propiedades estadísticas del estimador resultante y por ende sus resultados no pueden ser interpretados confiablemente, mucho menos comparados.

Una escogencia más parsimoniosa puede ser optar por un enfoque tipo *integrated household weighting*. Nótese que, como lo expone Heldal (1992), al realizar una calibración a nivel de personas, ya no será posible agregar a las personas de un mismo hogar para obtener un único peso del hogar, pues las características de las personas del hogar serán, en general, diferentes y sus respectivos factores de expansión también lo serán. Por tanto, definiendo a $w_{k|i}$ como el factor de expansión de la persona k que pertenece al hogar i, y a $w_{II,i}$ como el factor de expansión del hogar i, es necesario que el sistema de pesos satisfaga la siguiente restricción:

$$w_{k|i} = w_{II,i} \text{ para toda persona } k \text{ en el hogar } i.$$

De esta forma, sería posible obtener pesos consistentes con las restricciones de calibración a nivel de persona y que, al mismo tiempo, permitiera la integración con los hogares al cambiar de unidad de observación. En la literatura se han descrito varios métodos para lograr esta estandarización. A continuación se profundiza en algunos de ellos.

Estevao & Sarndal

En principio, se debe notar que es posible realizar el proceso de calibración de factores de expansión sobre la base de datos de las personas o sobre la base de datos de los hogares. De estos dos escenarios calibrar sobre la base de personas parecería ser la opción más rápida puesto que, en la mayoría de los casos, las cifras que se utilizan para calibrar están al nivel de los individuos. Por ejemplo, pensando en una encuesta de fuerza laboral, es evidente que las variables más importantes de la encuesta se encuentran al nivel de las personas y que la calibración de los factores de expansión se debería realizar desde la base de datos de personas.

Ahora, suponga una situación en la cual se desea calibrar por sexo. En este caso, se debería tener acceso a las proyecciones demográficas por sexo para el periodo de referencia de la encuesta y se procedería a calibrar los factores de expansión, utilizando un enfoque de post-estratificación. En este escenario, las ecuaciones de calibración estarían dadas por la siguiente expresión:

$$\left(\sum_{k\in s} w_k x_{1k}, \sum_{k\in s} w_k x_{2i}\right) = (t_{x1}, t_{x2})$$

En donde la suma se hace sobre las personas en la muestra s; además, x_{k1} toma el valor de uno, si el individuo k es mujer, y cero en otro caso. Por supuesto, $x_{k2} = 1 - x_{k1}$, t_{x1} es la proyección demográfica del total de mujeres y t_{x2} es la proyección demográfica del total de hombres. En este caso, las covariables de la calibración son variables dicotómicas. Nótese que las ecuaciones de calibración están al nivel de la muestra s que induce una base de datos de personas. Como el muestreo ha sido en varias etapas, una posibilidad que surge al momento de calibrar los factores de expansión es utilizar la muestra de hogares s_{II} que induce una base de datos con información de los hogares y calibrar usando un enfoque de calibración general con covariables continuas. De esta forma las ecuaciones de calibración estarían dadas por la siguiente expresión:

$$\left(\sum_{i\in s_{II}} w_{II,i} z_{1i}, \sum_{i\in s_{II}} w_{II,i} z_{2i}\right) = (t_{z1}, t_{z2})$$

En donde la suma ahora se realiza al nivel de la muestra de hogares s_{II}. Nótese que $z_{1i} = \sum_{k\in s_i} x_{k1}$ se refiere al número de hombres en el hogar i; $z_{2i} = \sum_{k\in s_i} x_{k2}$ es el número de mujeres en el hogar i; y los totales de calibración $t_{z1} = t_{x1}$ y $t_{z2} = t_{x2}$ siguen siendo el número de hombres y mujeres en la población, respectivamente, por lo que coinciden plenamente.

Al respecto, nótese que calibrar con el primer escenario reproduce los totales auxiliares sobre la base de personas, mientras que calibrar sobre el segundo escenario reproduce los totales sobre la base de hogares. Sin embargo, teniendo en cuenta los principios del muestreo en varias etapas y notando que en un hogar, la probabilidad de inclusión de las personas es de uno (inclusión forzosa), entonces generar factores de expansión para las personas en el segundo escenario es muy sencillo puesto que:

$$w_{k|i} = \frac{w_{II,i}}{Pr(k \in U_i | i \in sI)} = \frac{w_{II,i}}{1} = w_{II,i}$$

Es decir que, bajo este escenario de calibración, todas las personas dentro del hogar comparten los mismos pesos de muestreo y además estos pesos son iguales al peso del hogar. Estevao and Särndal (2006, sec. 5) recrean la calibración conjunta para hogares y personas. En resumen, luego de que se ha calibrado la base de hogares, se construyen los pesos a nivel de persona recurriendo a la siguiente expresión:

$$w_k = d_{k|i} w_{II,i} \quad \forall k \in s_i$$

Como todos los individuos pertenecientes a un hogar son seleccionados para que respondan la encuesta de hogares, se tiene que $d_{k|i} = 1$, por definición. Por lo tanto, el peso del individuo (en la base de datos de la muestra de personas) será idéntico al peso calibrado del hogar; es decir $w_k = w_{II,i} \quad \forall k \in S_i$. Además, dado que el muestreo es de conglomerados en la última etapa y todos los individuos del hogar son seleccionados, entonces el peso de muestreo del hogar será el promedio de los pesos individuales.

Lemaitre & Dufour

Un segundo enfoque, condensado en Lemaitre and Dufour (1987), afirma que se deben crear nuevas variables de calibración a nivel de persona, definidas como el promedio de las variables originales en el hogar. Por ende, se definen las siguientes cantidades:

$$z_{ik} = \sum_{i \in s_{II}} x_{ik} \quad y \quad \bar{z}_{ik} = \frac{z_{ik}}{N_i}$$

En donde z_{ik} es la agregación a nivel de hogar de las covariables originales de calibración a nivel de persona y N_i es el tamaño del i-ésimo hogar. Al ejecutar el algoritmo de calibración utilizando las variables z, en vez de las variables x, se reproducen las ecuaciones de calibración a satisfacción y, dado que todos los individuos comparten las mismas covariables en la calibración, sus pesos serán idénticos para todos aquellos compartiendo un mismo hogar. Nótese que esta calibración se realiza con la base de datos a nivel de personas.

En la literatura estadística se ha estudiado este enfoque integrado. Es así como Neethling and Galpin (2006) concluyeron que, para ambos enfoques, las estimaciones resultantes redujeron el sesgo, aumentaron la precisión y proporcionaron un único conjunto de ponderaciones para los datos de las encuestas estudiadas. Además, si se opta por el segundo enfoque, en el cual el tamaño de la base de datos sería igual al número de personas entrevistadas, se tendría el suficiente margen para actualizar las restricciones de calibración con el fin de ejercer un mayor control sobre los tamaños de los subgrupos de interés.

C. Calibración sobre razones, medias y proporciones

Gutierrez et al. (2016) afirman que, además de utilizar ponderadores calibrados a tamaños o totales, también es posible imponer restricciones de calibración sobre razones, que a su vez son una generalización de medias y proporciones. Por ejemplo, considere Q subgrupos de interés (dominios, estratos o post-estratos). Si las razones para dichos subgrupos fuesen conocidas, podemos encontrar pesos w_k que satisfagan la siguiente restricción:

$$\hat{\mathbf{R}}_{cal} = (\hat{R}_{1,cal}, ..., \hat{R}_{Q,cal})' = (R_1, ..., R_Q)' = \mathbf{R}$$

donde $\hat{R}_{q,cal} = \frac{\sum_{k \in s} w_k y_{qk}}{\sum_{k \in s} w_k x_{qk}}$. De esta forma, es posible imponer la siguiente restricción en las ecuaciones de calibración

$$\hat{\mathbf{R}}_{cal} = \mathbf{R}_U.$$

Es decir, para $q = 1, ..., Q$, se define la siguiente variable de información auxiliar

$$z_{qk} = \begin{cases} y_{qk} - R_q x_{qk} & \text{si } k \in s_q \\ 0 & \text{en otro caso} \end{cases}$$

En donde

$$t_{z_q} = \sum_{k \in U} z_{qk} = \sum_{k \in U} y_{qk} - R_q x_{qk} = 0$$

Como caso particular, si las medias de los subgrupos son conocidas, la restricción queda como

$$\bar{\mathbf{y}}_{cal} = (\bar{y}_{1,cal}, ..., \bar{y}_{Q,cal})' = (\bar{y}_1, ..., \bar{y}_Q)' = \bar{\mathbf{y}}$$

Así, la restricción para las ecuaciones de calibración, $\bar{\mathbf{y}}_{cal} = \bar{\mathbf{y}}$, para cada $q = 1, ..., Q$, se define a partir de la siguiente variable de calibración:

$$z_{qk} = \begin{cases} y_{qk} - \bar{y}_q & \text{si } k \in s_q \\ 0 & \text{en otro caso} \end{cases}$$

VI. Recorte y redondeo

Recorte de pesos extremos

Un inconveniente que se genera debido a la multitud de ajustes en los factores de expansión es que, si bien el estimador resultante tendrá un sesgo cercano a cero, la distribución de los pesos puede mostrar datos extremos, sobre todo a la derecha de la distribución (valores muy grandes), que hacen que la varianza del estimador crezca y que, por ende, la precisión de la inferencia decrezca. Para hacerle frente a este problema, es posible considerar un procedimiento de *trimming* o recorte de pesos, siguiendo las recomendaciones de Valliant et al. (2018, sec. 14.4), que puede ser resumido en los siguientes pasos:

1. Recortar cualquier peso mayor a un umbral prestablecido en la distribución de pesos ajustados. Por lo general este umbral se fija alrededro de 3.5 veces la mediana de los pesos. Por tanto,

$$U = 3.5 \times mediana(\mathbf{w_k})$$

2. Cualquier peso con magnitud superior a U se trunca de la siguiente manera

$$w_k^* = \begin{cases} U, & si\ w_k \geq U \\ w_k, & en\ otro\ caso. \end{cases}$$

3. Determinar la cantidad neta perdida debido al recorte de pesos extremos, siguiendo la siguiente expresión:

$$K = \sum_{s_r} (w_k^* - w_k)$$

4. Distribuir K equitativamente entre las unidades que no fueron recortadas.
5. Iterar hasta que todos los nuevos pesos calibrados estén por debajo del umbral U.

Al final del proceso se debe asegurar que los datos extremos en los factores de expansión han sido correctamente manejados y que la distribución general de los pesos no sufrió cambios estructurales en los subgrupos poblacionales de interés.

El problema del redondeo de los factores de expansión

El principio de representatividad es el paradigma inferencial dominante en cualquier encuesta de hogares y el factor de expansión es el concepto más importante en este contexto. Por ejemplo, un hogar en una encuesta con un factor de expansión de 500 se representa a sí mismo y a otros 499 hogares más. La definición teórica del factor de expansión, inducida por el inverso multiplicativo de la probabilidad de inclusión de un hogar en la muestra, hace que la inferencia sea insesgada y confiable. Sin embargo, debido a que la probabilidad de inclusión es un número real contenido en el intervalo $(0, 1]$, entonces su inverso multiplicativo también será un número real mayor o igual que uno.

Cuando el factor de expansión no es entero, entonces su interpretación se torna compleja desde el punto de vista práctico, aunque teóricamente no tenga ninguna repercusión negativa. Sin embargo, este inconveniente puede hacer que, en la práctica, las oficinas nacionales de estadística y los usuarios de las bases de datos de encuestas de hogares tomen la decisión (bienintencionada pero errada) de redondear estas cantidades al entero más cercano. Esta práctica es perjudicial porque le añade sesgo a la inferencia y causará problemas de sobre o sub estimación en algunos dominios de estudio. Sartore et al. (2019) plantean que el redondeo de los factores de expansión puede ser problemático puesto que las estimaciones ponderadas pueden crecer o decrecer enormemente.

Los siguientes ejemplos prácticos muestran de forma directa las repercusiones perjudiciales que conlleva esta práctica y que son consecuencia directa del sesgo de redondeo:

- En encuestas de establecimientos redondear el factor de expansión en las unidades que tienen flujos de ventas grandes trae problemas de sesgo en este dominio de estudio.
- En encuestas agropecuarias, si una unidad productiva produce un cuarto de la producción nacional, el redondeo de su factor de expansión es nefasto.
- En encuestas de hogares, en donde los diseños de muestreo son generalmente auto-ponderados (en donde todas las viviendas comparten el mismo factor de expansión) dentro de los estratos, redondear el factor de expansión implica sesgar por completo todo el estrato.

Suponiendo que una muestra probabilística $s = (I_1, \ldots, I_k, \ldots, I_N)'$ fue seleccionada de una población finita U mediante un diseño de muestreo que induce probabilidades de inclusión $\pi_k = E(I_k)$ para todos los individuos $k \in U$ (en donde I_k toma el valor uno si fue seleccionado o cero en otro caso) entonces desde el punto de vista teórico los estimadores de muestreo $\hat{t}_y = \sum_s d_k\ y_k$ son insesgados cuando el factor de expansión d_k es idéntico al inverso de la probabilidad de inclusión, puesto que

$$E(\hat{t}_y) = E\left(\sum_s \frac{y_k}{\pi_k}\right) = E\left(\sum_U I_k \frac{y_k}{\pi_k}\right) = \sum_U E(I_k)\frac{y_k}{\pi_k} = \sum_U \pi_k \frac{y_k}{\pi_k} = t_y$$

De las anteriores relaciones es evidente que, cuando el factor de expansión se redondea de forma determinística, entonces $E(\hat{t}_y) \neq t_y$. Para evadir el sesgo de redondeo, es necesario emplear un método aleatorio que induzca insesgamiento en los estimadores de muestreo. En general, este problema puede ser abordado desde una perspectiva probabilística. De hecho, si en primera instancia se utiliza como redondeo la parte entera (el entero máximo que sea menor o igual) del factor de expansión, entonces bastará con añadir aleatoriamente una unidad a algunos factores de expansión para asegurar que la suma de los factores redondeados sea idéntica a la original. Con esta simple idea se le devuelve la propiedad del insesgamiento a los estimadores de muestreo.l procedimiento se describe a continuación:

1. Para $k \in s$, definir
$$\phi_k = d_k - \lfloor d_k \rfloor$$

2. Seleccionar una submuestra $s_a = (c_1, \ldots, c_k, \ldots, c_n)'$ de s con probabilidades de inclusión ϕ_k, para $k \in s$. Note que c_k tomará el valor de uno, si el elemneto k está en la submuestra y de cero, si no fue seleccionado en la submuestra.

3. Si $c_k = 0$, entonces $\tilde{d}_k = \lfloor d_k \rfloor$; en otro caso, si $c_k = 1$, entonces $\tilde{d}_k = \lfloor d_k \rfloor + 1$.

En primera instancia, nótese que la submuestra s_a no necesariamente será de tamaño fijo, puesto que $\sum_s \phi_k$ no será entera en todos los casos; por ende, es posible utilizar un algoritmo de muestreo Poisson (Gutiérrez, 2016, sección 4.1) para seleccionar esta submuestra. Sin embargo, si esta suma es entera, es posible utilizar un algoritmo de muestreo más eficiente que induzca una submuestra de tamaño fijo como por ejemplo el método de Brewer (Tillé, 2006a). Por otro lado, la esperanza de estos factores redondeados condicionados a la submuestra s_a es igual a los factores de expansión originales, tal y como se muestra a continuación

$$E(\tilde{d}_k|s_a) = \lfloor d_k \rfloor + E(c_k|s_a) = \lfloor d_k \rfloor + \phi_k = d_k$$

Por lo anterior, es importante notar que el uso de este método aleatorio de redondeo siempre induce insesgamiento en los estimadores de muestreo, puesto que

$$E\left(\sum_s \tilde{d}_k y_k\right) = E\left[E\left(\sum_s \tilde{d}_k y_k|s_a\right)\right] = E\left(\sum_s E(\tilde{d}_k|s_a) y_k\right) = E\left(\sum_s d_k y_k\right) = t_y$$

Por último, cuando los factores de expansión de la encuesta están calibrados se presenta un problema de optimización un poco más complejo, puesto que al utilizar el redondeo aleatorio, los factores de expansión perderán la propiedad de calibración. Sartore et al. (2019) y Tillé (2019) han presentado diferentes soluciones a este problema, siendo la última mucho más fácil de implementar en el software estadístico `R`. Bajo esta perspectiva, la calibración de los factores de expansión crea nuevos pesos denominados w_k que conservan la siguiente propiedad para un conjunto de totales auxiliares $\mathbf{t_x}$ disponibles para toda la población

$$\sum_s w_k \mathbf{x}_k = \mathbf{t_x}$$

El siguiente algoritmo hace uso del muestreo balanceado (Tillé, 2006b, capítulo 8), el cual representa una forma de calibración desde el diseño de muestreo y es una solución óptima para seleccionar la submuestra s_a y por ende preservar la consistencia de los pesos calibrados con los totales auxiliares.

1. Para $k \in s$, definir $\phi_k = w_k - \lfloor w_k \rfloor$ y

$$\tilde{\mathbf{x}}_k = \phi_k \ \mathbf{x}_k$$

2. Seleccionar una submuestra balanceada $s_a = (c_1, \ldots, c_k, \ldots, c_n)'$ de s con probabilidades de inclusión ϕ_k, tal que

$$\sum_{k \in s_a} \frac{\tilde{\mathbf{x}}_k}{\phi_k} \cong \sum_{k \in s} \tilde{\mathbf{x}}_k$$

3. Si $c_k = 0$, entonces $\tilde{w}_k = \lfloor w_k \rfloor$; en otro caso, si $w_k = 1$, entonces $\tilde{w}_k = \lfloor w_k \rfloor + 1$.

Es importante recalcar que la restricción en la submuestra balanceada implica que los pesos redondeados cumplan la siguiente relación

$$\sum_s c_k \ \mathbf{x}_k \cong \sum_U \mathbf{x}_k - \sum_U \lfloor w_k \rfloor \ \mathbf{x}_k$$

Lo cual conlleva inmediatamente a que los nuevos pesos, además de estar redondeados, también estén calibrados; es decir

$$\sum_s \tilde{w}_k \mathbf{x}_k \cong \mathbf{t_x}$$

Nótese que el redondeo aleatorio depende de la selección de la submuestra s_a para completar los restos de la parte entera. En esta selección intervienen diferentes algoritmos de muestreo que se pueden aplicar fácilmente utilizando la librería `sampling` (Tillé and Matei, 2016). Por ejemplo, suponga una muestra de tamaño $n = 200$ que fue seleccionada de una población de tamaño $N = 9200$ con factores de expansión desiguales que no están calibrados. Asuma que el vector de probabilidades de inclusión en la muestra toman la siguiente forma

$$\boldsymbol{\pi}_s = (\underbrace{15/500}_{50 \ veces}, \ldots, \underbrace{15/800}_{80 \ veces}, \ldots, \underbrace{15/700}_{70 \ veces})'$$

Por lo tanto, el vector de pesos de muestreo estará definido de la siguiente manera:

$$\mathbf{d}_s = (\underbrace{33.33333}_{50 \ veces}, \ldots, \underbrace{53.33333}_{80 \ veces}, \ldots, \underbrace{46.66667}_{70 \ veces})'$$

De la misma manera, el vector de excesos $\phi_k = d_k - \lfloor d_k \rfloor$ estará dado por la siguiente expresión:

$$\boldsymbol{\phi}_s = (\underbrace{0.33333}_{130 \ veces}, \ldots, \underbrace{0.66667}_{70 \ veces})'$$

Luego del cálculo de ϕ_k, se selecciona la submuestra s_a. En particular, en este caso se utiliza el algoritmo de Brewer, puesto que $\sum_s \phi_k = 90$ y es entero. Al final del proceso de redondeo aleatorio la suma de los nuevos factores coincidirá con la suma de los factores originales. Por último, si en una segunda instancia, se considera que los pesos están calibrados mediante sendas covariables de calibración, entonces es posible utilizar el método del cubo, para que la submuestra esté balanceada y los pesos redondeados sigan las restricciones de calibración bajo una tolerancia predefinida.

Capítulo 11

Estimación del error de muestreo

Después de que la muestra fue seleccionada y luego de realizar el proceso de medición, es necesario realizar la estimación de los parámetros junto con la estimación de sus errores estándar, definido como la raiz cuadrada de la varianza.

Aunque la escogencia del diseño de muestreo y el estimador sean de libre elección para los investigadores, no lo es el cálculo de las medidas de confiabilidad y precisión. Dado que la base científica sobre la cual descansa el muestreo es la inferencia estadística, se deben respetar las normas básicas para la asignación y posterior cálculo del margen de error, que constituye una medida unificada del error total de muestreo que cuantifica la incertidumbre acerca de las estimaciones en una encuesta. La forma de estimar el error estándar depende de:

- La complejidad del diseño de muestreo: estratifiación, selección proporcional al tamaño, múltiples etapas.
- La complejidad del estimador: ajuste de pesos por ausencia de respuesta, calibración, razón de totales, medias, percentiles, coeficientes de regresión.

En general, podría afirmase que existen tres alternativas para calcular el error estándar de las estimaciones en una encuesta. Basados en la estrategia de muestreo es posible encontrar las *fórmulas exactas* que describan la varianza del estimador; sin embargo cuando el estimador utilizado no es una función lineal de totales, puede ser posible utilizar un enfoque de *linealización de Taylor* para aproximar la varianza del estimador a una función lineal. Por último, es posible apoyarse en en los métodos computacionales moderno y aplicar los principios de los *pesos replicados* para aproximar la varianza de cualquier estimador en una encuesta de hogares.

I. Fórmulas exactas y linealización de Taylor

En la mayoría de casos de interés se pueden encontrar las fórmulas exactas que aplican a cada diseño de muestreo y a cada estimador usado. Para diseños simples es posible implementarlas para calcular la estimación de los errores directamente. Sin embargo, en diseños multietápicos y con estimadores simples se pueden tornar extremadamente complicadas. Más aún, en diseños multietápicos y para estimadores complejos simplemente no son una opción plausible.

La estimación de la varianza en una estrategia de muestreo es una tarea no siempre sencilla. A partir de la teoría se establece un camino lógico basado en las probabilidades de inclusión de primer y segundo orden. En general, para cualquier diseño de muestreo sin reemplazo, la fórmula exacta para calcular una varianza del estimador de Horvitz-Thompson está dada por:

$$Var(\hat{t}_{y,\pi}) = \sum_U \sum_U \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

En donde $\Delta_{kl} = \pi_{kl} - \pi_l \pi_l$. Además, la probabilidad de inclusión de segundo orden se denota análogamente como π_{kl} y define la probabilidad de que los elementos k y l pertenezcan a la muestra al mismo tiempo; esto es,

$$\pi_{kl} = Pr(k \in s,\ l \in s) = Pr(I_k\ I_l = 1) = \sum_{s \ni k,l} p(s).$$

En donde el subíndice $s \ni k, l$ se refiere a la suma sobre todas las muestras que contienen a los elementos k-ésimo y l-ésimo. Evidentemente, por razones computacionales y porque es imposible acceder a la observación de los registros sobre toda la población finita, hacer este cálculo para los estimadores de los indicadores de interés en las encuestas simplemente no es viable.

En la práctica de las encuestas de hogares nunca se podrá contar con la varianza exacta de un estimador; por consiguiente, para tener una estimación de la precisión de la estrategia de muestreo se debe estimar la varianza del estimador. Gutiérrez (2016) afrima que un estimador insesgado para esta varianza está dada por la siguiente expresión:

$$\widehat{Var}_1(\hat{t}_{y,\pi}) = \sum \sum_S \frac{\Delta_{kl}}{\pi_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

Asimismo, si el diseño es de tamaño de muestra fijo, un estimador insesgado está dado por

$$\widehat{Var}_2(\hat{t}_{y,\pi}) = -\frac{1}{2} \sum \sum_S \frac{\Delta_{kl}}{\pi_{kl}} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2$$

De esta forma, cuando el tamaño de muestra es suficientemente grande, se puede construir un intervalo de confianza de nivel $(1-\alpha)$ para el total poblacional t_y como se indica a continuación:

$$IC(1-\alpha) = \left[\hat{t}_{y,\pi} - z_{1-\alpha/2} \sqrt{Var(\hat{t}_{y,\pi})}, \hat{t}_{y,\pi} + z_{1-\alpha/2} \sqrt{Var(\hat{t}_{y,\pi})} \right]$$

donde $z_{1-\alpha/2}$ se refiere al percentil $(1-\alpha/2)$ de una variable aleatoria con distribución normal estándar. Como cada diseño de muestreo induce una forma cerrada para las probabilidades de inclusión de primer y segundo orden, las fórmulas de la estimación de la varianza se reducen ostensiblemente. Por ejemplo, si el diseño de muestreo es aleatorio simple, la fórmula de la estimación de la varianza es

$$\hat{V}ar(\hat{t}_{y\pi}) = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) S^2_{y_s}$$

En donde $S^2_{y_s}$ es la varianza de los valores de la característica de interés en la muestra aleatoria s, dada por

$$S^2_{y_s} = \frac{1}{n-1}\sum_{k\in S}(y_k - \bar{y}_s)^2$$

Por otro lado, si el diseño de muestreo es aleatorio estratificado y el parámetro de interés es una media, la fórmula del estimador de Horvitz-Thompson es $\bar{y}_\pi = \frac{1}{N}\sum_s d_k y_k = \sum_{h=1}^H W_h \bar{y}_h$; en donde $W_h = N_h/N$. Ahora, siendo S^2_{yh} la varianza muestral en el estrato h de los valores de la característica de interés y definiendo $w_h = n_h/n$, la fórmula de la estimación de la varianza es

$$\widehat{V}ar(\bar{y}_\pi) = \sum_{h=1}^H w_h^2 \frac{1-f_h}{n_h} S^2_{yh}$$

Cuando el diseño de muestreo se complejiza, también lo hace la estimación de la varianza. Por ejemplo, si el diseño de muestreo es estratificada y bietápico, de tal forma que dentro de cada estrato U_h $h = 1, \ldots, H$ existen N_{Ih} unidades primarias de muestreo, de las cuales se selecciona una muestra s_{Ih} de n_{Ih} unidades mediante un diseño de muestreo aleatorio simple; y además, se considera que el sub-muestreo dentro de cada unidad primaria seleccionada es también aleatorio simple, de tal manera que para cada unidad primaria de muestreo seleccionada $i \in s_{Ih}$ de tamaño N_i se selecciona una submuestra s_i de elementos de tamaño n_i, entonces la forma final del estimador de la varianza del estimador de Horvitz-Thompson para el total poblacional quedaría de la siguiente manera:

$$\widehat{Var}(\hat{t}_{y,\pi}) = \sum_{h=1}^H \left[\frac{N^2_{Ih}}{n_{Ih}}\left(1-\frac{n_{Ih}}{N_{Ih}}\right)S^2_{\hat{t}_{yh}S_I} + \frac{N_{Ih}}{n_{Ih}}\sum_{i\in S_{Ih}}\frac{N_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)S^2_{y_{S_i}}\right]$$

En donde $S^2_{\hat{t}_y S_I}$ y $S^2_{y_{S_i}}$ son, respectivamente, las varianzas muestrales de los totales estimados en las UPM seleccionadas y las varianzas muestrales de los hogares incluidos en la submuestra dentro de las UPM seleccionadas en la muestra de la primera etapa.

Las fórmulas computacionales requeridas para estimar la varianza de estadísticas descriptivas como la media muestral están disponibles para algunos diseños complejos que incorporan elementos como la estratificación y el muestreo por conglomerados. Sin embargo, en el caso de estadísticas analíticas más complejas, tales como coeficientes de correlación y coeficientes de regresión, no se encuentra fácilmente las fórmulas específicas en diseños muestrales que se aparten del muestreo aleatorio simple. Estas fórmulas son enormemente complicadas o, en última instancia, se resisten al análisis matemático.

II. La técnica del último conglomerado

Debido a las dificultades algebraicas y computacionales, estimar la varianza en encuestas complejas que contemplan esquemas de conglomeración, selección en varias etapas y estratificación, puede tornarse bastante tedioso, costoso y además muy demorado. En esta sección se explica por qué la técnica del último conglomerado resulta ser una buena opción a la hora de aproximar la varianza en una encuesta compleja.

Para la estimación de la varianza de los estimadores de interés en encuestas multietápicas, los programas computacionales existentes utilizan una aproximación conocida como la técnica del último conglomerado (*ultimate cluster*). Esta aproximación, que sólo tiene en cuenta la varianza de los estimadores en la primera etapa, supone que ese muestreo fue realizado con reemplazo. Los procedimientos de muestreo en etapas posteriores de la selección son ignorados a menos que el factor de corrección para poblaciones finitas no sea despreciable a nivel de la primera etapa de muestreo.

En particular, considere cualquier estimador del total poblacional dado por la siguiente combinación lineal

$$\hat{t}_{y,\pi} = \sum_{k \in s} d_k y_k = \sum_{k \in U} I_k d_k y_k$$

En donde I_k son variables indicadoras de la pertenencia del elemento k a la muestra s. Ahora, asumiendo que el factor de expansión de la encuesta d_k cumple con los supuestos básicos de un ponderador que hace insesgado a $\hat{t}_y$, es decir:

$$E_p(I_k d_k) = 1$$

Suponiendo un diseño de muestreo en varias etapas (dos o más) en donde la primera etapa supone la selección de una muestra s_I de m_I unidades primarias de muestreo (UPM) U_i $(i \in s_I)$ de tal forma que

- Si la selección se realizó con reeemplazo, la i-ésima UPM tiene probabilidad de selección p_{I_i}.
- Si la selección se realizó sin reeemplazo, la i-ésima UPM tiene probabilidad de inclusión π_{I_i}.

En las subsiguientes etapas de muestreo, se procede a seleccionar una muestra de elementos para cada una de las UPM seleccionadas en la primera etapa de muestreo. Dentro de la i-ésima UPM se selecciona una muestra s_i de elementos; en particular la probabilidad condicional de que el k-ésimo elemento pertenzca a la muestra dada que la UPM que la contiene ha sido seleccionada en la muestra de la primera etapa está dada por la siguiente expresión:

$$\pi_{k|i} = Pr(k \in s_i | i \in s_I)$$

Por ejemplo, si el muestreo es sin reemplazo en todas sus etapas, la probabilidad de inclusión del k-ésimo elemento a la muestra s está dada por

$$\begin{aligned}
\pi_k &= Pr(k \in s) \\
&= Pr(k \in s_i, i \in s_I) \\
&= Pr(k \in s_i | i \in s_I) Pr(i \in s_I) = \pi_{k|i} \times \pi_{I_i}
\end{aligned}$$

Dado que el inverso de las probabilidades de inclusión son un ponderador natural, entonces se definen las siguientes cantidades:

1. $d_{I_i} = \frac{1}{\pi_{I_i}}$, que es el factor de expansión de la i-ésima UPM.
2. $d_{k|i} = \frac{1}{\pi_{k|i}}$, que es el factor de expansión del k-ésimo elemento dentro para la i-ésima UPM.

3. $d_k = d_{I_i} \times d_{k|i}$, que es el factor de expansión final del k-ésimo elemento para toda la población U.

Desde la teoría de muestreo, es posible evidenciar que si el diseño de muestreo es con reemplazo entonces, además del estimador HT, existe otro estimador insesgado que puede considerarse, conocido como el estimador de Hansen-Hurwitz (HH) (Gutiérrez, 2016). A diferencia del estimador HT, el estimador HH tiene una expresión de varianza muy sencilla de calcular, y por consiguiente las expresiones de la estimación de la varianza del estimador HH son más manejables desde el punto de vista computacional. En efecto, bajo un diseño de muestreo en varias etapas, el estimador de Hansen-Hurwitz para el total poblacional está dada por la siguiente expresión:

$$\hat{t}_{y,p} = \frac{1}{m_I} \sum_{i=1}^{m_I} \frac{\hat{t}_{y_i}}{p_{I_i}}$$

En donde p_{Ii} corresponde a la probabilidad de selección de la unidad i, mientras que m_I es el tamaño de muestra (con reemplazo) del muestreo en la primera etapa. En este caso, la varianza estimada del estimador HH es:

$$\widehat{Var}(\hat{t}_{y,p}) = \frac{1}{m_I(m_I - 1)} \sum_{i=1}^{m_I} \left(\frac{\hat{t}_{y_i}}{p_{I_i}} - \hat{t}_{y,p} \right)^2$$

En donde las cantidades $\hat{t}_{y_i}$ representan lo totales estimados de la variable de interés en la i-ésima UPM y están dados por:

$$\hat{t}_{y_i} = \sum_{k \in s_i} \frac{y_k}{\pi_{k|i}} = \sum_{k \in s_i} d_{k|i} y_k$$

El espíritu de la técnica del último conglomerado consiste en utilizar la expresión de la estimación de la varianza del estimador HH en vez de la expresión exacta en diseños de muestreo complejos que no contemplan selecciones con reemplazo en la primera etapa. Para lograrlo, algunas cantidades deben ser equiparadas antes de poder utilizar esta aproximación. Utilizar la aproximación de la varianza requiere equiparar los términos de manera apropiada. En primer lugar, fijémonos en los estimadores $\hat{t}_{y,p}$ y $\hat{t}_{y,\pi}$. Para realizar esta comparación, se requiere que se asuma la siguiente igualdad en las probabilidades de inclusión de la primera etapa:

$$\pi_{I_i} = p_{I_i} \times m_I$$

Por lo tanto, el estimador del total poblacional quedaría definido como un estimador tipo Hanwen-Hurwitz. En efecto,

$$\hat{t}_{y,\pi} = \sum_{k \in s} d_k y_k = \sum_{i=1}^{m_I} \sum_{k \in s_i} d_k y_k = \sum_{i=1}^{m_I} \sum_{k \in s_i} \frac{1}{\pi_{I_i} \pi_{k|i}} y_k = \sum_{i=1}^{m_I} \frac{\hat{t}_{y_i}}{\pi_{I_i}} \approx \frac{1}{m_I} \sum_{i=1}^{m_I} \frac{\hat{t}_{y_i}}{p_{I_i}}$$

Ahora, dado que la forma del estimador ha sido equiparada con un estimador tipo Hanwen-Hurwitz, es posible utilizar su estimación de varianza. Aún más, después de un poco de álgebra es posible tener la siguiente aproximación, cuya gran ventaja es que sólo hace uso de los factores de expansión finales d_k, que suelen ser reportados por los INE cuando liberan los microdatos de sus encuestas, en vez de los factores de expansión de la primera etapa o los factores de expansión condicionales dentro de las UPM.

$$\begin{aligned}
\widehat{Var}(\hat{t}_{y,p}) &= \frac{1}{m_I(m_I-1)}\sum_{i=1}^{m_I}\left(\frac{\hat{t}_{y_i}}{p_{I_i}}-\hat{t}_y\right)^2\\
&= \frac{m_I}{m_I-1}\sum_{i=1}^{m_I}\frac{1}{m_I^2}\left(\frac{\sum_{k\in s_i} d_{k|i}y_k}{p_{I_i}}-\sum_{i=1}^{m_I}\sum_{k\in s_i} d_k y_k\right)^2\\
&= \frac{m_I}{m_I-1}\sum_{i=1}^{m_I}\left(\frac{\sum_{k\in s_i} d_{k|i}y_k}{m_I p_{I_i}}-\frac{1}{m_I}\sum_{i=1}^{m_I}\sum_{k\in s_i} d_k y_k\right)^2\\
&= \frac{m_I}{m_I-1}\sum_{i=1}^{m_I}\left(\frac{\sum_{k\in s_i} d_{k|i}y_k}{\pi_{I_i}}-\frac{1}{m_I}\sum_{i=1}^{m_I}\sum_{k\in s_i} d_k y_k\right)^2\\
&= \frac{m_I}{m_I-1}\sum_{i=1}^{m_I}\left(\sum_{k\in s_i} d_k y_k-\frac{1}{m_I}\sum_{i=1}^{m_I}\sum_{k\in s_i} d_k y_k\right)^2
\end{aligned}$$

Basado en lo anterior, al definir $\breve{t}_{y_i} = \sum_{k\in s_i} d_k y_k$ como la contribución[1] de la i-ésima UPM a la estimación del total poblacional y $\bar{\breve{t}}_y = \frac{1}{m_I}\sum_{i=1}^{m_I}\breve{t}_{y_i}$ como la contribución promedio en el muestreo de la primera etapa, entonces el estimador de varianza toma la siguiente forma, conocida como el estimador de varianza del *último conglomerado*.

$$\widehat{Var}(\hat{t}_{y,p}) = \frac{m_I}{m_I-1}\sum_{i=1}^{m_I}\left(\breve{t}_{y_i}-\frac{1}{m_I}\sum_{i=1}^{m_I}\breve{t}_{y_i}\right)^2 = \frac{m_I}{m_I-1}\sum_{i=1}^{m_I}\left(\breve{t}_{y_i}-\bar{\breve{t}}_y\right)^2 \tag{11.1}$$

Por ejemplo, si el escenario de muestreo planteado en la encuesta es estratificado, con tres etapas de selección dentro de cada estrato, entonces al utilizar la técnica del último conglomerado, la aproximación del estimador de la varianza estaría dada por

$$\widehat{Var}(\hat{t}_{y,p}) = \sum_h \frac{n_h}{n_h-1}\sum_{i\in s_h}\left(\hat{t}_{y_i}-\bar{\hat{t}}_{y_h}\right)^2$$

En donde $\hat{t}_{y_i} = \sum_{k\in s_{hi}} w_k y_k$, $\bar{\hat{t}}_{y_h} = (1/n_h)\sum_{i\in s_h}\hat{t}_{y_i}$ y n_h es el número de UPMs seleccionadas en el estrato h. Este procedimiento, propuesto por Hansen et al. (1953) tiende a sobrestimar la varianza verdadera, aunque resulta ser una técnica apetecida por los investigadores puesto que utiliza directamente los pesos finales de muestreo o factores de expansión que son publicados por los INE.

Utilizar la técnica del último conglomerado es una salida práctica al problema de la estimación de la varianza que, para la mayoría de encuestas que brindan estadísticas oficiales a los países, puede

[1] Note que la suma de estas contribuciones en la muestra de la primera etapa da como resultado la estimación $\hat{t}_y$.

tornarse bastante complejo. Si bien, la expresión del estimador de la varianza no constituye un estimador estrictamente insesgado, sí se considera una aproximación bastante precisa.

Por último, es importante reflexionar acerca de la definición práctica y el concepto que envuelven esta aproximación ¿Qué es un **último conglomerado**? Es la primera unidad de muestreo en un diseño complejo. Por ejemplo, considere el siguiente diseño de muestreo en cuatro etapas:

$$\underbrace{\textbf{Municipio}}_{\text{UPM}} \Rrightarrow \underbrace{\textbf{Sector}}_{\text{USM}} \Rrightarrow \underbrace{\textbf{Vivienda}}_{\text{UTM}} \Rrightarrow \underbrace{\textbf{Hogar}}_{\text{UFM}}$$

En la primera las unidades primarias de muestreo (UPM) son los municipios; dentro de cada municipio, se seleccionan unidades secundarias de muestreo (USM) que corresponden a sectores cartográficos; de esta forma, el submuestreo continua hasta seleccionar las unidades finales de muestreo (UFM) que son los hogares.

Ahora, por lo general, la primera etapa de muestreo de una encuesta está inducida por dos tipos de diseños: estratificado o con probabilidad de selección proporcional al tamaño del municipio. En cualquiera de los dos casos, se crean subgrupos de inclusión forzosa. En el muestreo estratificado serán las ciudades grandes y en el muestreo proporcional también, puesto que la medida de tamaño inducirá probabilidades de inclusión mayores a uno. Luego, para poder aplicar la aproximación en este caso, los municipios pertenecientes a este subgrupo de inclusión forzosa no serán considerados UPM, sino que inducirán un estrato de ciudades grandes. En cada ciudad de este estrato se realizará un muestreo de la siguiente manera:

$$\underbrace{\textbf{Sector}}_{\text{UPM}} \Rrightarrow \underbrace{\textbf{Vivienda}}_{\text{USM}} \Rrightarrow \underbrace{\textbf{Hogar}}_{\text{UFM}}$$

Es necesario tener en cuenta esta particularidad de algunas encuestas para poder aplicar correctamente esta técnica de aproximación de varianzas. En resumen, para aquellas ciudades que pertenecen al estrato de inclusión forzosa, las UPM serán los sectores cartográficos, y para el resto del país, las UPM serán los municipios cuya probabilidad de inclusión en la muestra de la primera etapa es menor a uno.

III. Linealización de Taylor

Cuando se trata de estimar parámetros que tienen una forma no lineal, es posible recurrir al uso de las herramientas del análisis matemático para aproximar sus varianzas con el fin de publicar las cifras oficiales con sus respectivos errores estándar. Valliant et al. (2013) mencionan que esta técnica se basa en expresar el estimador como función de estimadores lineales de totales. Por ejemplo, si el interés recae en estimar un parámetro poblacional θ que a su vez depende de Q estimadores de totales $(t_1, t_2, \ldots, t_Q)$, entonces su estimador de muestreo se debe expresar como $\hat{\theta} = f(\hat{t}_1, \ldots, \hat{t}_Q)$.

En donde $\hat{t}_j = \sum_{k \in s} w_k y_{jk}$ es un estimador del j-ésimo total. Por consiguiente, si el estimador de interés no es una función lineal de totales, entonces las propiedades estadísticas comunes como insesgamiento, eficiencia y precisión de los estimadores deben ser aproximadas. Es común usar la técnica de la linealización de Taylor para encontrar aproximaciones lineales de primer orden. Gutiérrez (2016, capitulo 8) presenta una explicación detallada de esta técnica aplicada a diferentes escenarios de estimación, en donde se consideran los siguientes pasos para construir un estimador linealizado de la varianza de una función no lineal de totales:

1. Expresar el estimador del parámetro de interés $\hat{\theta}$ como una función de estimadores de totales insesgados. Así,
$$\hat{\theta} = f(\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_Q).$$
2. Determinar todas las derivadas parciales de f con respecto a cada total estimado $\hat{t}_q$ y evaluar el resultado en las cantidades poblacionales t_q. Así
$$a_q = \left.\frac{\partial f(\hat{t}_1, \ldots, \hat{t}_Q)}{\partial \hat{t}_q}\right|_{\hat{t}_1 = t_1, \ldots, \hat{t}_Q = t_Q}$$
3. Aplicar el teorema de Taylor para funciones vectoriales para linealizar la estimación $\hat{\theta}$ con $\mathbf{a} = (t_1, t_2, \cdots, t_Q)'$. En el paso anterior, se vio que $\triangledown \hat{\theta} = (a_1, \cdots, a_Q)$. Por consiguiente se tiene que
$$\hat{\theta} = f(\hat{t}_1, \ldots, \hat{t}_Q) \cong \theta + \sum_{q=1}^{Q} a_q(\hat{t}_q - t_q)$$
4. Definir una nueva variable E_k con $k \in s$ al nivel de cada elemento observado en la muestra aleatoria, así:
$$E_k = \sum_{q=1}^{Q} a_q y_{qk}$$
5. Si los estimadores $\hat{t}_q$ son estimadores de Horvitz-Thompson, una expresión que aproxima la varianza de $\hat{\theta}$ está dada por
$$AVar(\hat{\theta}) = Var\left(\sum_{q=1}^{Q} a_q \hat{t}_{q,\pi}\right) = Var\left(\sum_S \frac{E_k}{\pi_k}\right) = \sum_U \sum_U \Delta_{kl} \frac{E_k}{\pi_k} \frac{E_l}{\pi_l}.$$

Tal como se advirtió anteriormente, Gutiérrez (2016) afirma que, para encontrar una estimación de la varianza de $\hat{\theta}$, no es posible utilizar directamente los valores E_k, porque éstos dependen de los totales poblacionales (las derivadas a_q se evalúan en los totales poblacionales que son desconocidos). Por consiguiente, los valores E_k se aproximan reemplazando los totales desconocidos por los estimadores de los mismos. Siendo e_k la aproximación de la variable linealizada dada por

$$e_k = \sum_{q=1}^{Q} \hat{a}_q y_{qk}$$

donde $\hat{a}_q$ corresponde a un estimador de a_q. La aproximación de Taylor para el estimador la varianza del estimador de Horvitz-Thompson para un total está dado por la siguiente expresión

$$\widehat{Var}(\hat{t}_{y,\pi}) = \sum \sum_S \frac{\Delta_{kl}}{\pi_{kl}} \frac{e_k}{\pi_k} \frac{e_l}{\pi_l}$$

Por ejemplo, bajo este contexto, si se quisiera estimar la tasa de desocupación (función no lineal de totales) definida como el cociente entre el total poblacional de personas que se encuentran en edad laboral pero que carecen de un empleo (t_y) sobre la cantidad de personas que pertenecen a la población económicamente activa (t_z), entonces, la estimación de la aproximación de la varianza del estimador de esta razón $\hat{\theta} = \dfrac{\hat{t}_{y,\pi}}{\hat{t}_{z,\pi}}$ estaría definida como sigue en términos de las variables linealizadas

$$e_k = \frac{1}{\hat{t}_{z,\pi}}(y_k - \hat{\theta} z_k)$$

Si, además, el muestreo de la encuesta es bietápico con selección aleatoria simple sin reemplazo en cada etapa, entonces este estimador de la varainza tomaría la siguiente forma:

$$\widehat{Var}(\hat{\theta}) = \frac{N_I^2}{n_I}\left(1 - \frac{n_I}{N_I}\right) S^2_{\hat{t}_e S_I} + \frac{N_I}{n_I}\sum_{i \in S_I} \frac{N_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right) S^2_{e_{S_i}}$$

En donde $S^2_{\hat{t}_e S_I}$ es la varianza muestral de los totales estimados t_{ei} de las UPM seleccionadas en la primera etapa del muestreo y $S^2_{e_{S_i}}$ es la varianza muestral entre los valores e_k para los elementos incluidos en la submuestra dentro de cada UPM seleccionada en la primera etapa. De la misma manera, para el caso particular de la estimación de un promedio utilizando el estimador de Hájek, las anteriores expresiones pueden adaptarse convenientemente.

Si se utiliza un estimador de calibración para el total poblacional de la característica de interés t_y, entonces siguiendo los lineamientos de Gutiérrez (2016, sección 10.6), la varianza estimada del estimador utilizando la técnica de linealización de Taylor haría uso de las siguientes variables linealizadas

$$e_k = y_k - \mathbf{x}_k' \hat{\theta}$$

En donde $\mathbf{x}_k$ son las variables relacionadas con el vector de totales auxiliares $\mathbf{t_x}$, medidas en la misma encuesta y $\hat{\theta}$ es el vector estimado de coeficientes de regresión entre los valores que toman la característica de interés y_k y el vector de información auxiliar $\mathbf{x}_k$.

En la región, tanto la *Pesquisa Nacional por Amostra de Domicilios Continua*, en Brasil, como la *Encuesta de Caracterización Socioeconómica Nacional*, en Chile, utilizan esquemas de linealización de Taylor en conjunción con el acercamiento del último conglomerado. En resumen, la linealización de Taylor supone que es posible definir una aproximación lineal de $\hat{\theta}$ así

$$\hat{\theta} - \theta \approx \sum_{j=1}^{p} \frac{\partial f(\hat{t}_1, \ldots, \hat{t}_p)}{\partial \hat{t}_j}(\hat{t}_j - t_j) = \sum_{k \in s} w_k e_k + c$$

En donde $e_k = \sum_{j=1}^{p} \frac{\partial f(\hat{t}_1, \ldots, \hat{t}_p)}{\partial \hat{t}_j} y_{jk}$ son variables linealizadas, mientras que las cantidad c representa una constante determinística que no aporta a la varianza de $\hat{\theta}$. Nótese lo conveniente de expresar esta aproximación de esta manera puesto que al final, las cantidades que intervienen en la varianza se pueden expresar como una suma ponderada de las variables e_k y por consiguiente es posible aplicar todos los principios establecidos anteriormente. De esta forma, asumiendo el escenario de muestreo planteado en las secciones anteriores, el estimador de la varianza de la aproximación lineal de $\hat{\theta}$ está dado por

$$\widehat{Var}(\hat{\theta}) = \sum_h \frac{n_h}{n_h - 1} \sum_{i \in s_h} \left(\hat{t}_{e_i} - \bar{\hat{t}}_{e_h} \right)^2$$

En donde $\hat{t}_{e_i} = \sum_{k \in s_{hi}} w_k e_k$ y $\bar{\hat{t}}_{e_h} = (1/n_h) \sum_{i \in s_h} \hat{t}_{e_i}$. Por ejemplo, si el interés estuviera en estimar una razón, entonces las nuevas variables linealizadas son $e_k = (1/\hat{t}_{y_2})(y_{1k} - \hat{\theta}\ y_{2k})$.

IV. Pesos replicados

Las complicaciones en el cálculo de los errores de muestreo pueden ser mayores dependiendo de la escogencia del estimador y del diseño de muestreo asumido para la recolección de la información primaria. En algunas ocasiones, el proceso de linealización puede resultar complicado, por lo que es posible optar por una estrategia computacional aproximada que permite pasar por alto el proceso teórico de definición de las cantidades que estiman la varianza del estimador. Este conjunto de métodos supone la idea de la selección sistemática de *submuestras* que son utilizadas para estimar el parámetro de interés, utilizando los mismos principios de estimación que con la muestra completa. Por lo anterior, se obtienen estimaciones puntuales para cada réplica, las cuales son utilizadas para estimar la varianza del estimador de interés.

En ausencia de fórmulas adecuadas, en los últimos años han aparecido una variedad de técnicas empíricas que proporcionan *varianzas aproximadas que parecen satisfactorias para fines prácticos* (Kish, 1965). Estos métodos utilizan una muestra de datos para construir submuestras y generar una distribución para las estimaciones de los parámetros de interés utilizando cada submuestra. Los resultados de la submuestra se analizan para obtener una estimación del parámetro, así como intervalos de confianza para esa estimación. El enfoque general de esta técnicas computacional se basa en:

1. Dividir la toda la muestra en pequeños subconjutnos (réplicas).
2. Repetir los mismos procesos de ajuste de ponderadores en cada réplica.
3. Hacer la estimación en cada subgrupo.
4. La varianza del estimador se calcula de manera simple como la varianza muestral de todas las estimaciones en cada réplica.

Usando esta metodología, no se requiere que las bases de datos públicas contengan la información asociada a los estratos o UPM y esto protege la anonimización de los respondientes. Además, no se requiere conocer el diseño de muestreo utilizado en la encuesta, puesto que al proveer los pesos replicados en las bases de datos, los investigadores pueden estimar el error de muestreo de forma automatizada y sin necesidad de intrincadas fórmulas matemáticas. Estos métodos ha demostrado ser eficientes y precisos para la mayoría de parámetros de interés, algunas encuestas que utilizan estas metodologías son la American Community Survey, la American Housing Survey y la Current Population Survey. En América Latina la PNADC de Brasil, la ENE de Chile y la ENEMDU de Ecuador han hecho uso de estas técnicas para la estimación de la varianza de algunos estimadores complejos.

En particular, hay tres metodologías que abordan este problema: los pesos replicados repetidas balanceadas (McCarthy, 1969; Judkins, 1990), el Jackknife (Krewski and Rao, 1981) y el Bootstrap (Rao and Wu, 1988). La idea general detrás de estos métodos es que, partiendo de la muestra completa, en cada réplica se seleccione un conjunto de UPMs manteniendo todas las unidades que hayan sido seleccionadas dentro de esas UPMs. Luego, es necesario reponderar los pesos de muestreo para que se conserve la representatividad; de esta manera, para cada réplica se obtendrá un nuevo conjunto de pesos de muestreo. Con estos pesos, se calcula la estimación de interés, obteniendo tantas estimaciones como réplicas definidas. Wolter (2007) provee todos los detalles teóricos referentes al problema de la estimación de la varianza utilizando los pesos replicados.

En lo concerniente con las técnicas de remuestreo y la utilización de los pesos replicados para el cálculo de los errores de muestreo se recalca que la técnica de *Jackknife* es útil para estimar parámetros lineales, pero no tiene un buen comportamiento cuando se trata de estimar percentiles o funciones de distribución. La técnica de *réplicas repetidas balanceadas* es útil para estimar parámetros lineales y no lineales, pero puede ser deficiente cuando se tienen dominios pequeños que pueden inducir estimaciones nulas en la configuración de los pesos. Sin embargo, como se explicará más adelante, el ajuste de Fay a la técnica anterior resulta palear todos los anteriores inconvenientes. En este caso es importante utilizar una matriz de Hadammard que induzca no más de 120 conjuntos de pesos replicados para que la publicación de la base de datos no se sobrecargue. Por último, el *bootstrap* debe ser utilizado con con detenimiento porque debe replicar el diseño de muestreo exacto y esto se logra construyendo una población a partir de los pesos de muestreo.

La técnica de Jackknife

Este método provee estimaciones eficientes para estimadores lineales y no lineales (a excepción de los percentiles). En su forma más básica, los pesos replicados se crean al retirar una UPM del análisis. Por ende, se tendrán tantos pesos replicados como UPM existan en la muestra. Además, cuando una UPM se retira en la réplica, todas las unidades dentro de esa UPM también se retiran. El desarrollo del procedimiento de Jackknife se remonta a un método utilizado por Quenouille (1956) para reducir el sesgo de las estimaciones. El refinamiento ulterior del método (Mosteller, 1968) llevó a su aplicación en una serie de situaciones de las ciencias sociales en las que las fórmulas no están fácilmente disponibles para el cálculo de errores de muestreo.

Este procedimiento ofrece mayor flexibilidad, pues el Jackknife puede implementarse en una amplia variedad de diseños muestrales; además de facilidad de uso, puesto que no requiere de software especializado. El concepto principal de esta técnica parte de una muestra de tamaño n, la cual se divide en A grupos de igual tamaño $m = n/A$, a partir de esta división, la varianza de un estimador $\hat{\theta}$ se estima a partir de la varianza observada en los A grupos.

Para cada grupo $(a = 1, 2, ..., A)$, se calcula $\hat{\theta}_{(a)}$, una estimación para el parámetro θ, calculada de la misma forma que la estimación $\hat{\theta}$ obtenida con la muestra completa, pero solo con la información restante (luego de la eliminación del grupo a). Para $a = 1, 2, ..., A$ se define

$$\hat{\theta}_a = A\hat{\theta} - (A - 1)\hat{\theta}_{(a)}$$

como un pseudovalor de θ. El estimador obtenido mediante Jackknife se presenta como una alternativa a $\hat{\theta}$ y se define como:

$$\hat{\theta}_{JK} = \frac{1}{A}\sum_{a=1}^{A}\hat{\theta}_a$$

mientras que el estimador de la varianza obtenido mediante Jackknife se obtiene como:

$$\widehat{Var}_{JK1} = \frac{1}{A(A-1)} \sum_{a=1}^{A} \left(\hat{\theta}_a - \hat{\theta}_{JK}\right)^2$$

También es posible utilizar como estimador alternativo:

$$\widehat{Var}_{JK2} = \frac{1}{A(A-1)} \sum_{a=1}^{A} \left(\hat{\theta}_a - \hat{\theta}\right)^2$$

Para diseños estratificados y multietápicos en los cuales las unidades primarias de muestreo han sido seleccionadas en el estrato h, para $h = 1, \ldots, H$, el estimador de varianza de Jackknife para la estimación de un parámetro poblacional está dado por

$$\widehat{Var}_{JK}(\hat{\theta}) = \sum_{h=1}^{H} \frac{n_{Ih} - 1}{n_{Ih}} \sum_{i=1}^{n_{Ih}} (\hat{\theta}_{h(i)} - \hat{\theta})^2$$

donde $\hat{\theta}_{h(i)}$ es la estimación de θ usando los datos de la muestra excluyendo las observaciones en la i-ésima unidad primaria de muestreo (Korn and Graubard, 1999, pg. 29 – 30). Shao and Tu (2012, Teorema 6.2) garantiza la convergencia en probabilidad de este estimador hacia la varianza teórica, de donde se puede concluir que es un estimador aproximadamente insesgado para la varianza teórica. Los pesos de la unidad k que pertenece a la UPM U_i, en el estrato U_h están dados por la siguiente expresión:

$$d_{hk}^{i} = \begin{cases} 0, \text{ si } U_i \in U_h \text{ y } k \in U_i \\ d_k, \text{ si } k \notin U_h \\ \frac{n_{Ih}}{n_{Ih}-1} d_k, \text{ si } U_i \in U_h \text{ y } k \notin U_i \end{cases}$$

En donde n_{Ih} es el número de UPM seleccionadas en el estrato U_h. Por último, para reducir el número de pesos replicados, se pueden conformar *unidades de varianza*, uniendo varias UPM dentro de un mismo estrato, y también *estratos de varianza*, colapsando estratos dentro de la muestra. En el primer caso, se podrían emparejar las UPM en cada estrato de acuerdo a la medida de tamaño. En este caso, el estimador de varianza está dado por la siguiente expresión

$$\widehat{Var}_{JK}(\hat{\theta}) = \sum_{h} \frac{n_{Ih} - n_{Ihg}}{n_{Ih}} \sum_{i \in s_{hg}} (\hat{\theta}_{h(g)} - \hat{\theta})^2$$

En donde $\hat{\theta}_{hg}$ es el estimador del parámetro retirando el g-ésimo subgrupo del estrato U_h y n_{Ihg} es el tamaño del subgrupo en la muestra denotado como s_{hg}. A continuación se ejemplifica la estructura final de una base de datos de pesos replicados con esta técnica en un conjunto reducido de tan solo ocho unidades muestrales divididas en cuatro UPM y dos estratos. Se enfatiza que habrán tantos conjuntos (columnas) de pesos replicados Jackknife como UPM existentes en la muestra de la primera etapa.

k	Estrato	UPM	$d_k^{(1)}$	$d_k^{(2)}$	$d_k^{(3)}$	$d_k^{(4)}$
1	Estrato1	UPM1	0	1,03	1,03	1,03
2	Estrato1	UPM1	0	1,03	1,03	1,03
3	Estrato1	UPM2	1,03	0	1,03	1,03
4	Estrato1	UPM2	1,03	0	1,03	1,03
5	Estrato2	UPM3	1,03	1,03	0	1,03
6	Estrato2	UPM3	1,03	1,03	0	1,03
7	Estrato2	UPM4	1,03	1,03	1,03	0
8	Estrato2	UPM4	1,03	1,03	1,03	0

El método de las réplicas repetidas balanceadas

Esta técnica conocida como BRR se desarrolló para diseños en donde dos UPM son seleccionadas por estrato. Este método funciona consistentemente para la estimación de parámetros lineales y no lineales (incluidos los percentiles) y, además, asegura máxima dispersión de las UPM a través de las regiones geográficas (estratos) (Valliant and Dever, 2017). Nótese que si el submuestreo en cada estrato es $n_{Ih} = 2$, entonces al utilizar la técnica de Jackknife deberíamos definir 2^H posibles réplicas al seleccionar aleatoriamente una UPM en cada estrato, lo cual puede ser intratable computacionalmente.

Es posible lograr la misma eficiencia reduciendo el número de pesos replicados utilizando un enfoque ortogonal con matrices de Hadamard, que son matrices cuadradas cuyas columnas deben ser ortogonales. Por ejemplo, considere la siguiente matriz:

$$\begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{pmatrix}$$

Asumiendo que el valor +1 implica que la primera UPM se mantiene como parte de la réplica y la segunda UPM es retirada de la réplica; el valor -1 implica que la segunda UPM se mantiene como parte de la réplica y la primera UPM es retirada de la réplica. Por tanto, en cada réplica se retira una UPM por estrato. Esto implica que el producto punto entre cualquier combinación de dos columnas deber ser igual a cero. Por ejemplo, tomando las columnas 2 y 4, se tiene que

$$(+1, -1, +1, -1)' \cdot (+1, -1, -1, +1) = 1 + 1 - 1 - 1 = 0$$

De esta forma, el número de réplicas ortogonales será igual al *menor múltiplo de 4 mayor o igual al número de estratos*. Las UPM que se mantienen en cada réplica se conocen como *half-samples*. Por consiguiente, el peso de los individuos en la UPM que se mantiene se multiplica por un factor de 2. Entonces, se tiene que

$$d_k = \begin{cases} 0, & \text{si } k \text{ pertence a la UPM que fue retirada} \\ 2d_k, & \text{en otro caso.} \end{cases}$$

Bajo esta metodología BRR, el estimador de la varianza toma la siguiente forma:

$$\widehat{Var}_{BRR}(\hat{\theta}) = \frac{1}{A}\sum_{a=1}^{A}(\hat{\theta}_a - \hat{\theta})^2$$

En donde $\hat{\theta}_a$ es el estimador del parámetro de interés en la réplica a. A continuación se ejemplifica la estructura final de una base de datos de pesos replicados con esta técnica en el mismo conjunto reducido, considerando que hay dos estratos.

k	Estrato	UPM	$d_k^{(1)}$	$d_k^{(2)}$
1	Estrato1	UPM1	2	0
2	Estrato1	UPM1	2	0
3	Estrato1	UPM2	0	2
4	Estrato1	UPM2	0	2
5	Estrato2	UPM3	2	0
6	Estrato2	UPM3	2	0
7	Estrato2	UPM4	0	2
8	Estrato2	UPM4	0	2

Una desventaja de este método BRR es que las unidades en dominios con muestra pequeña pueden estar ausentes en algunas combinaciones de pesos replicados por el diseño ortogonal. Lo anterior conlleva una pérdida de precisión en el cálculo del error estándar. Una solución a este problema es modificar los pesos en los pesos replicados. Para la aplicación de la Réplicas Repetidas Balanceadas es recomendable usar el método de Fay, en donde se siguen los lineamientos basados en la matriz de Hadamard, aunque las UPM no son retiradas completamente, sino que su peso se modifica de la siguiente manera:

$$d_k^a = \begin{cases} \rho * d_k, & \text{si } k \text{ pertence a la UPM que fue retirada} \\ (2-\rho)d_k, & \text{en otro caso.} \end{cases}$$

En donde $0 < \rho < 1$. Algunos estudios por simulación han mostrado una buena eficiencia para valores de ρ iguales a 0.3, 0.5 o 0.7. Bajo la metodología BRR con el ajuste de Fay, el estimador de la varianza toma la siguiente forma:

$$\widehat{Var}_{Fay}(\hat{\theta}) = \frac{1}{A(1-\rho)^2}\sum_{a=1}^{A}(\hat{\theta}_a - \hat{\theta})^2$$

En donde $\hat{\theta}_a$ es el estimador del parámetro de interés en la réplica a. A continuación se ejemplifica la estructura final de una base de datos de pesos replicados con el ajuste de Fay en el mismo conjunto reducido.

k	Estrato	UPM	$d_k^{(1)}$	$d_k^{(2)}$
1	Estrato1	UPM1	1.5	0.5
2	Estrato1	UPM1	1.5	0.5

k	Estrato	UPM	$d_k^{(1)}$	$d_k^{(2)}$
3	Estrato1	UPM2	0.5	1.5
4	Estrato1	UPM2	0.5	1.5
5	Estrato2	UPM3	1.5	0.5
6	Estrato2	UPM3	1.5	0.5
7	Estrato2	UPM4	0.5	1.5
8	Estrato2	UPM4	0.5	1.5

En general, para la aplicación de estos métodos, los pesos de muestreo se ajustan para generar los pesos replicados y, posteriormente, se repiten los ajustes por ausencia de respuesta y calibración para estos nuevos pesos. Con esta metodología se estiman los errores de muestreo y la varianza de muestreo, incluyendo el impacto de la ausencia de respuesta, el cual se espera que sea pequeño, pero relevante en el momento de calcular estimadores más precisos. Retomando las observaciones hechas anteriormente, en el caso en el que la encuesta cuente con estratos en donde se encuentre una sola UPM, el método de los pesos replicados repetidas balanceadas no es aplicable puesto que al eliminar una unidad, algunos estratos quedarán vacíos.

Método de Bootstrap

En este apartado se presenta el método de Bootstrap (Efron and Tibshirani, 1993), el cual es muy utilizado por su fácil implementación; además de ser flexible en términos del número de pesos replicados que se crean. Teniendo los pesos muestrales se procede a crear los pesos replicados con el método de remuestreo con el fin de poder calcular estimaciones de indicadores junto con las estimación de las varianzas. En el contexto de las encuestas de hogares, se trata de realizar un remuestreo a las unidades primarias de muestreo seleccionadas desde el marco de áreas.

El Bootstrap es el método basado en réplicas más versatil para el cálculo de errores estándar. Valliant and Dever (2017) mencionan que es muy eficiente en la estimación de parámetros lineales y no lineales, a diferencia del Jackknife que no es eficiente en la estimación de percentiles. Funciona también para tamaños de muestra pequeños, a diferencia del método BRR que requiere una muestra de mínimo dos UPM por estrato. Este método requiere una cantidad de pesos replicados grande, usualmente mayor a 200.

Siendo s_{BS} la submuestra Bootstrap, el peso replicado del individuo k perteneciente a la UPM i del estrato h sigue la siguiente expresión:

$$d_k^b = \begin{cases} 0, \text{ si la UPM } i \text{ no pertence a } s_{BS} \\ d_k\left[1 - \sqrt{\frac{n_{Ih}^*}{n_{Ih}-1}} + \sqrt{\frac{n_{Ih}^*}{n_{Ih}-1}}\frac{n_{Ih}}{n_{Ih}^*}n_{Ihi}^*\right], \text{ en otro caso} \end{cases}$$

En donde n_{Ih} es el úmero de UPM en la muestra original del estrato h, n_{Ih}^* es el número de UPM en la muestra Bootstrap y n_{Ihi}^* es el número de veces que la UPM i fue seleccionada en la muestra Bootstrap. En este caso se selecciona una muestra Bootstrap con $m_h^* = m_h - 1$, y los pesos toman la siguiente forma

$$d_k^a = \begin{cases} 0, & \text{si la UPM } i \text{ no pertence a } s_{BS} \\ d_k \left[1 - \sqrt{\frac{n_{Ih}^*}{n_{Ih}-1}} + \sqrt{\frac{n_{Ih}^*}{n_{Ih}-1}} \frac{n_{Ih}}{n_{Ih}^*} n_{Ihi}^*\right], & \text{en otro caso} \end{cases}$$

Bajo la metodología Bootstrap (BS), el estimador de la varianza toma la siguiente forma:

$$\widehat{Var}_{BS}(\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b - \hat{\theta})^2$$

En donde $\hat{\theta}_b$ es el estimador del parámetro de interés en la réplica b inducida por la muestra Bootstrap. En resumen, para la b-ésima réplica con los pesos resultantesse podrán calcular las estimaciones de totales, proporciones, promedios y razones y sus respectivas varianzas o desviaciones. En general, es necesario reflejar el ajuste de los pesos en los pesos replicados, por esto es necesario trabajar con la muestra originalmente seleccionada, la cual contendrá unidades no elegibles y unidades que no respondieron. Los mismos ajustes que se hicieron a la muestra original se deben realizar en cada réplica. Si hubo calibración de pesos también debe ser incluida como un proceso en cada réplica, para asegurar que el error estándar inducido por estos métodos incluirá el incremento (o decremento) de la varianza inducida por estos ajustes a los pesos. A continuación se ejemplifica la estructura final de una base de datos de pesos replicados con la metodología de bootstrap en el mismo conjunto reducido.

k	Estrato	UPM	$d_k^{(1)}$	$d_k^{(2)}$	$d_k^{(3)}$	$d_k^{(4)}$
1	Estrato1	UPM1	2	0	1	1
2	Estrato1	UPM1	2	0	1	1
3	Estrato1	UPM2	0	2	1	1
4	Estrato1	UPM2	0	2	1	1
5	Estrato2	UPM3	1	1	2	0
6	Estrato2	UPM3	1	1	2	0
7	Estrato2	UPM4	1	1	0	2
8	Estrato2	UPM4	1	1	0	2

Rao and Wu (1984) y Rao and Wu (1988) aconsejan seleccionar una muestra con reemplazo de $n_I - 1$ de las n_I UPM de la muestra, teniendo en cuenta la probabilidad de selección del diseño complejo en la primera etapa. Dado que la selección es con reemplazo, una UPM puede ser seleccionada más de una vez en esta nueva muestra. Por otro lado, también es posible realizar una selección sin reemplazo; en este caso, Preston (2009) recomiendan seleccionar una muestra con reemplazo de $n_I/2$ de las n_I UPM de la muestra, teniendo en cuenta la probabilidad de selección del diseño complejo en la primera etapa.

V. Consideraciones adicionales

En la práctica del muestreo, existen algunos paradigmas que inducen la planeación y diseño de las encuestas. En esta sección se muestran ejemplos y contra-ejemplos que permiten ilustrar algunas mitos en la estimación del error de muestreo de las encuestas de hogares. Para esto, consideremos la varianza del esimador HT, dada a continuación:

$$Var(\hat{t}_{y,\pi}) = \sum_{k \in U} \sum_{l \in U} \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

En donde $\Delta_{kl} = (\pi_{kl} - \pi_k \pi_l)$ y π_{kl} denota la probabilidad de inclusión conjunta de los elementos k y l pertenezcan a la muestra s. Bajo diseños de muestreo de tamaño fijo, existen dos estimadores insesgados para esta varianza, el primero originalmente propuesto por Horvitz and Thompson (1952) y dado por

$$\widehat{Var}_1(\hat{t}_{y,\pi}) = \sum_{k \in s} \sum_{l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

El segundo estimador propuesto por Sen (1953) y Yates and Grundy (1953), está dado por la siguiente expresión:

$$\widehat{Var}_2(\hat{t}_{y,\pi}) = -\frac{1}{2} \sum_{k \in s} \sum_{l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2$$

Estimaciones negativas de varianza

La idea de que no pueden existir estimaciones negativas de la varianza se ha instalado como un razonamiento bastante lógico e intuitivo: dado que la varianza es un parámetro positivo, entonces no puede ser estimada con cantidades negativas. Sin embargo, en la inferencia basada en el diseño de muestreo, sí es posible obtener estimativas negativas de varianza para algunas estructuras poblacionales particulares y es por esto que se requiere una experiencia mayor por parte del equipo de muestreo, que debe conocer bajo qué condiciones se podría presentar esta situación para evadirla.

Nótese que se debe diferenciar entre estimador (que es una función de variables aleatorias) y parámetro (que es un valor real desconocido). En efecto, para la varianza del estimador HT (parámetro desconocido y siempre positivo) hay estimadores (función de variables aleatorias) que pueden arrojar estimaciones negativas. Es posible que las estimaciones de la varianza arrojen resultados negativos, que no pueden ser utilizados ni interpretados. Considere el siguiente diseño de muestreo de tamaño fijo e igual a $n = 2$, el cual induce seis posibles muestras.

s	I_1	I_2	I_3	I_4	$p(s)$
s_1	1	1	0	0	0.31
s_2	1	0	1	0	0.20
s_3	1	0	0	1	0.14
s_4	0	1	1	0	0.03
s_5	0	1	0	1	0.01
s_6	0	0	1	1	0.31

En el anterior ejemplo, la probabilidad de obtener una muestra compuesta por los dos primeros elementos se fijó en 0.31; mientras que la probabilidad de obtener una muestra compuesta por el primer y el tercer elemento se fijó en 0.20 y así sucesivamente. Para esta configuración se obtienen las estimaciones puntuales para cada una de las seis posibles muestras, así como las dos posibles estimaciones de la varianza. Nótese que para ambos escenarios existen estimaciones negativas.

s	I_1	I_2	I_3	I_4	$p(s)$	$\hat{t}_{y,\pi}$	$\widehat{Var}_1(\hat{t}_{y,\pi})$	$\widehat{Var}_2(\hat{t}_{y,\pi})$
s_1	1	1	0	0	0.31	9.560440	38.099984	-0.9287681
s_2	1	0	1	0	0.20	5.883191	-4.744190	2.4710422
s_3	1	0	0	1	0.14	4.933110	-3.680428	8.6463858
s_4	0	1	1	0	0.03	7.751323	-100.252974	71.6674365
s_5	0	1	0	1	0.01	6.801242	-165.715154	323.3238494
s_6	0	0	1	1	0.31	3.123994	3.426730	-0.1793659

A pesar de los resultados negativos para las varianzas, tanto el estimador del total como los dos estimadores de su varianza siguen siendo insesgados. En efecto al multiplicar la estimación puntual por la probabilidad del diseño de muestreo se obtienen los valores poblacionales. La varianza del estimador HT para este diseño en particular es 6.744442, la cual corresponde con la esperanza de ambos estimadores de varianza. Para evitar estas estimativas negativas, Gutiérrez (2016) afirma que es necesario garantizar que la covarianza (Δ_{kl}) sea negativa para cada par de elementos en la población ($k \neq l$), lo cual no sucede con este esquema de muestreo, puesto que:

$$\Delta_{kl} = \begin{bmatrix} 0.2275 & 0.0825 & -0.1510 & -0.1590 \\ 0.0825 & 0.2275 & -0.1590 & -0.1510 \\ -0.1510 & -0.1590 & 0.2484 & 0.0616 \\ -0.1590 & -0.1510 & 0.0616 & 0.2484 \end{bmatrix}$$

Disminución de la varianza ante el aumento del tamaño de muestra

Por otro lado, Tal vez la idea de que la varianza deberá disminuir a medida que el tamaño de muestra crece se ha venido de la lógica intuitiva en donde el error de muestreo no debería existir si se realiza una medición completa de la población. Sin embargo, para algunas estrategias de muestreo es posible encontrar que existen situaciones en donde el tamaño de muestra crece, y con él la varianza del estimador. En esta sección se mostrará un ejemplo en donde sucede exactamente eso.

Para poder mostrar este hecho, vamos a utilizar un ejemplo reducido. Supongamos una población de $N = 3$ elementos $U = \{1, 2, 3\}$ y comparemos dos diseños de muestreo, el primero con un tamaño de muestra fijo de $n = 1$ y el segundo con un tamaño de muestra fijo de $n = 2$. En ambos casos la variable de interés es dicotómica que denota la presencia o ausencia del fenómeno en los individuos de la población. En el primer caso, el diseño de muestreo de tamaño de muestra $n = 1$ es el siguiente:

s	I_1	I_2	I_3	$p(s)$
s_1	1	0	0	0.5
s_2	0	1	0	0.1
s_3	0	0	1	0.4

En este esquema de muestreo, la varianza del estimador de Horvitz-Thompson es igual a $Var(\hat{t}_{y,\pi}) = 1.5$. Sin embargo, en un segundo caso, considere el siguiente diseño de muestreo de tamaño de muestra $n = 2$:

s	I_1	I_2	I_3	$p(s)$
s_1	1	1	0	0.7
s_2	1	0	1	0.2
s_3	0	1	1	0.1

Nótese que en este esquema de muestreo, la varianza del estimador de Horvitz-Thompson es igual a $Var(\hat{t}_{y,\pi}) = 2.3$. Por tanto, no es exacto afirmar que siempre que un diseño de muestreo contemple un tamaño de muestra más grande se tendrá obligatoriamente una reducción de la varianza.

Referencias

Bibliografía

Alexander, C. (1987). A class of methods for using person controls in household weighting. *Survey Methodology*, 13(1):183–198.

Baillargeon, S. and Rivest, L.-P. (2011). The construction of stratified designs in r with the package stratification. *Survey Methodology*, 37(1):53–65.

Baillargeon, S., Rivest, L.-P., and Ferland, M. (2007). Stratification en enquÊtes entreprises: Une revue et quelques avancÉes. *Assemblée annuelle de la SSC*, page 8.

Ballin, M. and Barcaroli, G. (2013). Joint determination of optimal stratification and sample allocation using genetic algorithm. *Survey Methodology*, 39(2):369–393.

Barcaroli, G. (2014). Samplingstrata: An r package for the optimization of stratified sampling. *Journal of Statistical Software*, 61(1):1–24.

Biemer, P. P. and Lyberg, L. E. (2003). *Introduction to survey quality*. Wiley series in survey methodology. Wiley-Interscience.

CEPAL (2018). *Medición de la pobreza por ingresos - Actualización metodológica y resultados*. Metodologías de la CEPAL.

Clark, R. G. and Steel, D. G. (2007). Sampling within households in household surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(1):63–82.

Cochran, W. G. (1977). *Sampling Techniques*. Wiley, third edition edition.

Dalenius, T. and Hodges, J. L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54(285):15.

DANE (2017). Gran encuesta integrada de hogares - - departamento administrativo nacional de estadística.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418):376–382.

Duncan, G. J. and Kalton, G. (1987). Issues of design and analysis of surveys across time. *International Statistical Review / Revue Internationale de Statistique*, 55(1):97.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Estevao, V. and Särndal, C.-E. (2006). Survey estimates by calibration on complex auxiliary information. *International Statistical Review / Revue Internationale de Statistique*, 74(2):127–147.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52(3):761–766.

Gambino, J. G. (2009). Design effect caveats. *The American Statistician*, 63(2):141–146.

Gambino, J. G. and Silva, P. d. N. (2009). *Chapter 16 - Sampling and Estimation in Household Surveys*, volume 29 of *Handbook of Statistics*, page 407–439. Elsevier.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., and Tourangeau, R. (2009). *Survey Methodology*. John Wiley and Sons.

Gunning, P. and Horgan, J. M. (2004). A new algorithm for the construction of stratum boundaries in skewed populations. *Survey Methodology*, 30(2):159–166.

Gutierrez, H. A., Zhang, H., and Rodriguez, N. (2016). The performance of multivariate calibration on ratios, means and proportions. *Revista Colombiana de Estadística*, 39(2):281.

Gutiérrez, A., Zhang, H., and Montaño, C. (2016). Calculo del tamaño de muestra para la estimación de una varianza en poblaciones finitas con funciones en r. *Comunicaciones en Estadística*, 9(1):109.

Gutiérrez, H. A. (2016). *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*. Ediciones de la U, segunda edición edition. Google-Books-ID: UlVmE5pkRwIC.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample survey methods and theory*, volume 1. Wiley New York.

Heeringa, S. G., West, B. T., and Berglund, P. A. (2017). *Applied survey data analysis*. Chapman and Hall CRC statistics in the social and behavioral sciences series. CRC Press.

Heldal, J. (1992). A method for calibration of weights in sample surveys. page 22.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685.

IBGE-BR (2014). Pesquisa nacional por amostra de domicilios continua - notas metodológicas.

INDEC (2018). Encuesta permanente de hogares - instituto nacional de estadística y censos.

INEC (2018). Instituto nacional de estadística y censos.

INEGI (2012). Metodología de la construcción del marco maestro de muestreo 2012 y del diseño de la muestra maestra 2012.

Jarque, C. M. (1981). A solution to the problem of optimum stratification in multivariate sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 30(2):163–169.

Judkins, D. R. (1990). Fay's method for variance estimation. *Journal of Official Statistics; Stockholm*, 6(3):223.

Kalton, G. (2009). Some issues in the design and analysis of longitudinal surveys.

Kalton, G. and Citro, C. F. (1993). Panel surveys: adding the fourth dimension. *Survey Methodology*, 19(2):205–215.

Kim, J. K. and Riddles, M. K. (2012). Some theory for propensity-score-adjustment estimators in survey sampling. *Survey Methodology*, 38(2):157–165.

Kish, L. (1965). *Survey Sampling.* John Wiley and Sons.

Kish, L. (2004). *Statistical Design for Research.* Wiley classic library edition. Wiley.

Korn, E. L. and Graubard, B. I. (1999). *Analysis of health surveys.* Wiley.

Kozak, M. (2004). Optimal stratification using random search method in agricultural surveys. *Statistic in Transition*, 6(5):797–806.

Krewski, D. and Rao, J. N. K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, 9(5):1010–1019.

Lavallée, P. and Hidiroglou, M. A. (1988). On the stratification of skewed populations. *Survey Methodology*, 14(1):33–43.

Lemaitre, G. and Dufour, J. (1987). An integrated method for weighting persons and families. *Survey Methodology*, 13(2).

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 140:55–55.

Lohr, S. (2000). *Sampling: Design and Analysis.* Thompson.

Lohr, S. L. (2019). *Sampling: Design and Analysis.* Duxbury Press.

Lumley, T. (2010). *Complex surveys: a guide to analysis using R.* Wiley series in survey methodology. Wiley.

Macqueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, page 281–297. University of California Press.

McCarthy, P. J. (1969). Pseudo-replication: Half samples. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 37(3):239–264.

McLaren, C. and Steel, D. G. (2001). Rotation patterns and trend estimation for repeated surveys using rotation group estimates. *Statistica Neerlandica*, 55(2):221–238.

Mosteller, F. (1968). *Data Analysis, Including Statistics.*

Neethling, A. and Galpin, J. S. (2006). Weighting of household survey data: A comparison of various calibration, integrated and cosmetic estimators. *South African Statistical Journal*, page 23.

OIT (1982). Resolución sobre estadísticas de la población económicamente activa, del empleo, del desempleo y del subempleo.

OIT (2013). Estadísticas del trabajo, el empleo y la subutilización de la fuerza de trabajo.

ONU (2011). *Canberra Group Handbook on Household Income Statistics.* United Nations Economic Comission for Europe, second edition edition.

ONU (2015). Transformar nuestro mundo: la agenda 2030 para el desarrollo sostenible.

ONU (2016). Global sustainable development report 2016.

Park, I. (2003). Design effects and survey planning. page 8.

Presser, S., Rothgeb, J., Couper, M., Lessler, J., Martin, E., Martin, J., and Singer, E. (2004). *Methods for Testing and Evaluating Survey Questionnaires*. John Wiley and Sons.

Preston, J. (2009). Rescaled bootstrap for stratified multistage sampling. *Survey Methodology*, 35(2):227–234.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4):353–360.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, J. N. K. and Wu, C. F. J. (1984). Bootstrap inference for sample surveys. In *Procedings of the Survey Research Methods Section*, pages 106–112. American Statistical Association.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401):231–241.

Rojas, H. A. G. (2020). *samplesize4surveys: Sample Size Calculations for Complex Surveys*. R package version 4.1.1.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33:99–119.

Sartore, L., Toppin, K., Young, L., and Spiegelman, C. (2019). Developing integer calibration weights for census of agriculture. *Journal of Agricultural, Biological and Environmental Statistics*, 24(1):26–48.

Schwarz, N., Knäuper, B., Hippler, H.-J., Noelle-Neumann, E., and Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *The Public Opinion Quarterly*, 55(4):570–582.

Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5:119–127.

Shao, J. and Tu, D. (2012). *The Jackknife and Bootstrap*. Springer Series in Statistics. Springer New York.

Silva, P. d. N. (2004). Calibration estimation: when and why, how much and how. *Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística*.

Särndal, C.-E. and Lundström, S. (2006). *Estimation in surveys with nonresponse*. Wiley series in survey methodology. Wiley, repr edition.

Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer Science and Business Media. Google-Books-ID: ufdONK3E1TcC.

Tillé, Y. (2006a). *Sampling Algorithms*. Springer Series in Statistics. Springer-Verlag.

Tillé, Y. (2006b). *Sampling Algorithms*. Springer Series in Statistics. Springer-Verlag.

Tillé, Y. (2019). A simple and efficient way of rounding calibration weights. page 3.

Tillé, Y. and Matei, A. (2016). *sampling: Survey Sampling*. R package version 2.8.

UN (2005). *Household surveys in developing and transition countries*. Studies in methods / United Nations, Department of Economic and Social Affairs, Statistics Division Series F.

UN (2008a). *Designing household survey samples: practical guidelines*. Studies in methods / United Nations, Department of Economic and Social Affairs, Statistics Division Series F. United Nations.

UN (2008b). *Designing household survey samples: practical guidelines*. Studies in methods / United Nations, Department of Economic and Social Affairs, Statistics Division Series F. United Nations.

Valliant, R. and Dever, J. A. (2017). *Survey Weights: A Step-by-step Guide to Calculation*. Stata Press, 1 edition edition.

Valliant, R., Dever, J. A., and Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer New York.

Valliant, R., Dever, J. A., and Kreuter, F. (2018). *Practical Tools for Designing and Weighting Survey Samples*. Statistics for Social and Behavioral Sciences. Springer International Publishing.

Vehovar, V. (1999). Field substitution and unit nonresponse. *Journal of Official Statistics*, 15(2):335–350.

Wolter, K. M. (2007). *Introduction to variance estimation*. Statistics for social and behavioral sciences. Springer, 2nd ed edition.

Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society B*, 15:253–261.