

Efectos de diseño para indicadores sociales en América Latina

Función generalizada
de varianza para
estimadores directos
provenientes de
encuestas de hogares

Andrés Gutiérrez
Giovany Babativa-Márquez



NACIONES UNIDAS

CEPAL



Gracias por su interés en esta publicación de la CEPAL



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

Deseo registrarme



NACIONES UNIDAS



www.cepal.org/es/publications



www.instagram.com/publicacionesdelacepal



www.facebook.com/publicacionesdelacepal



www.issuu.com/publicacionescepal/stacks



www.cepal.org/es/publicaciones/apps

SERIE

ESTUDIOS ESTADÍSTICOS

106

Efectos de diseño para indicadores sociales en América Latina

Función generalizada de varianza para
estimadores directos provenientes de
encuestas de hogares

Andrés Gutiérrez

Giovany Babativa-Márquez



NACIONES UNIDAS

CEPAL

Este documento fue preparado por Andrés Gutiérrez, Experto Regional en Estadísticas Sociales, y Giovany Babativa-Márquez, Consultor de la Unidad de Estadísticas Sociales, División de Estadísticas de la Comisión Económica para América Latina y el Caribe (CEPAL), bajo la coordinación de Xavier Mancero, Jefe de dicha Unidad. La elaboración del documento contó con el apoyo financiero del decimotercer tramo de la Cuenta de las Naciones Unidas para el Desarrollo y la contribución del Fondo de Población de las Naciones Unidas (UNFPA).

Se agradece la colaboración de Felipe Ortiz en la preparación de la primera versión de los códigos computacionales.

Las Naciones Unidas y los países que representan no son responsables por el contenido de vínculos, enlaces o marcadores a sitios externos incluidos en esta publicación, ni por las menciones de sociedades mercantiles o nombres comerciales de productos y servicios, y no deberá entenderse que existe adhesión a sitios, su contenido, sus responsables ni a los productos o servicios que se mencionen u ofrezcan.

Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad de los autores y pueden no coincidir con las de la Organización o las de los países que representa.

Publicación de las Naciones Unidas
ISSN: 1680-8789 (versión electrónica)
ISSN: 1994-7364 (versión impresa)
LC/TS.2023/95
Distribución: L
Copyright © Naciones Unidas, 2023
Todos los derechos reservados
Impreso en Naciones Unidas, Santiago
S.23-00594

Esta publicación debe citarse como: A. Gutiérrez y G. Babativa-Márquez, "Efectos de diseño para indicadores sociales en América Latina: función generalizada de varianza para estimadores directos provenientes de encuestas de hogares", *serie Estudios Estadísticos*, N° 106 (LC/TS.2023/95), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL), 2023.

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

Índice

| | |
|--|-----------|
| Resumen | 5 |
| Introducción | 7 |
| I. Técnicas tradicionales en la estimación de error estándar de los estimadores de muestreo | 15 |
| II. Función generalizada de varianza y su relación con el efecto de diseño..... | 15 |
| A. Conceptos fundamentales | 15 |
| B. El efecto de diseño | 18 |
| III. Estimación de la varianza mediante FGV en el repositorio BADEHOG de la CEPAL..... | 21 |
| A. Tres modelos FGV | 21 |
| B. Fuentes de información | 24 |
| C. Flujo de trabajo | 26 |
| IV. Resultados..... | 29 |
| A. FGV para los parámetros de interés | 30 |
| B. Estimación del efecto diseño | 34 |
| V. Conclusiones | 35 |
| Bibliografía..... | 37 |
| Anexos | 39 |
| Anexo 1 | 40 |
| Anexo 2 | 41 |
| Serie Estudios Estadísticos: números publicados..... | 59 |

Cuadros

| | | |
|------------|--|----|
| Cuadro 1 | Cantidad de registros por país e información del diseño de muestreo..... | 20 |
| Cuadro 2 | Variables disponibles sobre necesidades básicas insatisfechas en cada país | 25 |
| Cuadro 3 | Descripción de las variables de interés..... | 29 |
| Cuadro 4 | Parámetros estimados para las FGV | 33 |
| Cuadro A1 | Medidas de calidad del ajuste para los modelos LK..... | 40 |
| Cuadro A2 | Medidas de calidad del ajuste para los modelos LE..... | 40 |
| Cuadro A3 | Cantidad de registros por país e información del diseño de muestreo..... | 40 |
| Cuadro A4 | Efectos de diseño para Argentina | 42 |
| Cuadro A5 | Efectos de diseño para Bolivia | 42 |
| Cuadro A6 | Efectos de diseño para Brasil | 43 |
| Cuadro A7 | Efectos de diseño para Chile | 44 |
| Cuadro A8 | Efectos de diseño para Colombia..... | 45 |
| Cuadro A9 | Efectos de diseño para Costa Rica | 46 |
| Cuadro A10 | Efectos de diseño para Republica Dominicana..... | 47 |
| Cuadro A11 | Efectos de diseño para Ecuador..... | 48 |
| Cuadro A12 | Efectos de diseño para Guatemala | 49 |
| Cuadro A13 | Efectos de diseño para Honduras..... | 50 |
| Cuadro A14 | Efectos de diseño para México..... | 51 |
| Cuadro A15 | Efectos de diseño para Nicaragua..... | 52 |
| Cuadro A16 | Efectos de diseño para Panamá..... | 53 |
| Cuadro A17 | Efectos de diseño para Perú..... | 54 |
| Cuadro A18 | Efectos de diseño para Paraguay | 55 |
| Cuadro A19 | Efectos de diseño para el Salvador | 56 |
| Cuadro A20 | Efectos de diseño para Uruguay | 57 |
| Cuadro A21 | Efectos de diseño para Venezuela | 57 |

Gráficos

| | | |
|-----------|--|----|
| Gráfico 1 | Estimación del error estándar usando la FGV y el método de linealización para el ingreso | 31 |
| Gráfico 2 | Estimación del error estándar usando la FGV y el método de linealización para la tasa de desempleo | 35 |

Resumen

El error de muestreo se refiere a la diferencia entre la estimación obtenida de una muestra y el parámetro que se obtendría si se midiera exhaustivamente a toda la población. La estimación del error de muestreo es importante para evaluar la precisión de los resultados y reducir la variabilidad en los resultados de la encuesta. Sin embargo, en algunos países de América Latina, las bases de datos de las encuestas se publican sin variables que proporcionen información sobre el diseño de muestreo, lo que dificulta la estimación adecuada del error.

Este documento propone un método simple para estimar el error de muestreo en ausencia de información completa del diseño de muestreo. Se introduce el concepto de efecto de diseño, un factor multiplicativo que se aplica a la estimación simple de la varianza para obtener medidas de incertidumbre. Además, se presentan estimaciones del efecto de diseño para un conjunto de indicadores, que pueden ser de utilidad para calcular tamaños de muestra adecuados para estudios observacionales utilizados para la evaluación de políticas públicas.

Este trabajo utiliza los datos del Banco de Datos de Encuestas de Hogares (BADEHOG) de la CEPAL y modelos que permiten encontrar una Función Generalizada de Varianza (FGV) para estimar los errores de muestreo en países que no cuentan con información completa del diseño de muestreo. El método basado en la FGV es consistente con el método tradicional de linealización de Taylor y permite estimar el error estándar como función de la estimación puntual y el tamaño de la muestra. Los efectos de diseño estimados para las variables de interés en 18 países de América Latina pueden ayudar a diseñar o rediseñar encuestas de hogares para medir constructos sociales específicos.

Introducción

Las encuestas de hogares son una herramienta valiosa para entender las necesidades y las condiciones de vida de la población en América Latina, puesto que permiten a los investigadores recolectar información detallada sobre los ingresos, el empleo, la educación, la salud, la vivienda y otros aspectos importantes de la vida de las personas, del comportamiento en los hogares y las características de las viviendas. La información recogida en las encuestas de hogares es utilizada por gobiernos, empresas y organizaciones no gubernamentales para diseñar políticas y programas que aborden las necesidades de la población y juegan un papel crucial en la toma de decisiones y en el desarrollo de estrategias para mejorar la calidad de vida de las personas en América Latina.

Al determinar el error de muestreo, es posible concluir acerca de si la muestra seleccionada es representativa y si los resultados obtenidos pueden generalizarse a toda la población de interés. Contar con estimaciones del error de muestreo en las encuestas de hogares de América Latina es importante porque permite evaluar la precisión de los resultados obtenidos a partir de la muestra seleccionada. Además, las estimaciones del error de muestreo permiten identificar las fuentes de variabilidad en los resultados de la encuesta y tomar medidas para reducirlas.

La tarea de estimar estos errores es fundamental para garantizar la validez de las estimaciones provenientes de las encuestas de hogares. Para lograr un proceso de estimación exitoso es necesario contar con información en la base de datos que permita representar todas las fuentes de variabilidad de las encuestas para tenerlas en cuenta en los cálculos de los errores. Dichas fuentes de variabilidad están representadas por la estratificación de la población, las múltiples etapas de selección y la ponderación desigual de hogares y personas. Por ende, las bases de datos que se utilizan en las Oficinas Nacionales de Estadística para realizar este proceso de estimación contendrán variables específicas con la identificación de los estratos, las unidades primarias y secundarias de muestreo y los factores de expansión.

Sin embargo, en algunos países de América Latina, las bases de datos de las encuestas de hogares son publicadas sin las variables que dan cuenta de la información del diseño de muestreo, generalmente debido a consideraciones relacionadas con la confidencialidad de los entrevistados. Esto dificulta que

los usuarios puedan estimar apropiadamente una medida que cuantifique la incertidumbre de las estimaciones que se obtienen con los microdatos de dichas encuestas.

El presente documento busca dar una solución a esta problemática desde una visión general que le permita a los usuarios:

- Obtener una medida del error de la estimación basado en un método que sea simple de aplicar y que requiera de poca información
- Aplicar un factor multiplicativo, conocido como efecto de diseño, a la estimación simple de la varianza para automáticamente obtener medidas de incertidumbre
- Conocer la magnitud del efecto de los diseños de muestreo en los países para diseñar (o rediseñar) encuestas de hogares que midan los constructos sociales que se abordan en este documento.

Este último apartado es particularmente valioso al realizar estudios de impacto de intervenciones de política pública en los países. Cada vez más, se utilizan estudios observacionales (Glewwe y Todd, 2022; White y Raitzer, 2017) que permitan inferir apropiadamente si una política pública tuvo el impacto adecuado o no en la población de interés. Sin embargo, por lo general, el diseño de estos estudios no está liderado por las Oficinas Nacionales de Estadística (ONE). Con esta publicación, las instituciones que planifican dichas evaluaciones en los países podrán contar con cifras precisas que les permitan calcular los tamaños de muestra adecuados para realizar este tipo de estudios.

A partir del Banco de Datos de Encuestas de Hogares (BADEHOG) de la Comisión Económica para América Latina y el Caribe (CEPAL), administrado desde la División de Estadísticas, y utilizando las bases de datos que cuentan con la información del diseño de muestreo, se ajustaron diferentes modelos que permiten encontrar una Función Generalizada de Varianza (FGV) para los estimadores directos (basados en el diseño de muestreo) de algunos parámetros trazadores en estudios de investigación social.

Esta función pudo ser aplicada a los países que no cuentan con la información del diseño de muestreo en sus bases de datos para estimar apropiadamente los errores de muestreo. Los resultados permiten llegar a la conclusión de que el método basado en la FGV es exitoso, ya que es consistente con el método tradicional de la linealización de Taylor (Gutiérrez, 2009), y permite estimar el error estándar como función de la estimación puntual y el tamaño de la muestra con el que fue obtenido. A partir de lo anterior, se estiman los efectos de diseño para las variables de interés por diferentes dominios para 18 países de América Latina.

En el marco de las políticas de datos abiertos que se han implementado en los diferentes países de la región, varias de las Oficinas Nacionales de Estadística han puesto a disposición los microdatos de las encuestas de hogares para contribuir a la investigación externa por parte de otras instituciones del gobierno, organismos internacionales, sociedad civil, academia, entre otros.

Tener acceso a los microdatos permite generar análisis para dominios que no han sido publicados por las ONE, o que no fueron planificados en el diseño de las encuestas, obteniendo así una mayor desagregación en el análisis de los datos. Ello hace necesario contar con una medida de la precisión de las estimaciones obtenidas para los indicadores en los subgrupos de interés en cada investigación, que permita discernir si las estimaciones provenientes de los análisis son confiables y precisas o si, por el contrario, tienen asociada una incertidumbre demasiado alta, en cuyo caso es mejor abstenerse de realizar algún proceso inferencial.

Una de las medidas que regularmente se utiliza para medir la precisión de los estimadores es el error estándar, el cual se define como la raíz cuadrada de la varianza del estimador directo. Su naturaleza absoluta define la amplitud de los intervalos de confianza asociados a la

estimación puntual, permitiendo reflejar el grado de incertidumbre que el proceso de muestreo genera en la inferencia.

Evidentemente, entre más cercano a cero sea el error estándar, habrá una mayor precisión y, como consecuencia, los intervalos de confianza serán más angostos. En efecto, en el hipotético caso de que la muestra pudiese abarcar a toda la población de interés, y dependiendo de la estrategia de muestreo definida por la ONE, el error estándar sería cero, pues no habría error de muestreo.

En general, la estimación del error estándar depende de los procedimientos que se hayan implementado para calcular y ajustar los factores de expansión, y no deben supeditarse solo a los valores obtenidos como ponderaciones, ya que omitir la estructura del diseño de muestreo conlleva a estimaciones sesgadas y, por consiguiente, erradas del error estándar (Lu y Gelman 2003, Gutiérrez 2009). En la literatura se encuentran diferentes métodos que permiten estimar el error estándar de un estimador directo basados a partir de un diseño de muestreo probabilístico, algunas se basan en fórmulas teóricas, linealización de Taylor y aproximaciones, otras usan procedimientos de remuestreo, o incluso se pueden encontrar funciones de estimación de varianzas, errores estándar y coeficientes de variación que utilizan modelos estadísticos (Wolter 2007).

Aunque los investigadores tienen libertad para elegir el diseño de muestreo y el estimador que usarán, una vez fijada la estrategia de muestreo (diseño de muestreo y estimador) es necesario seguir los procedimientos de estimación de varianzas apropiados para estimar las medidas de confiabilidad y precisión. El muestreo se basa en la inferencia estadística, por lo que es importante seguir las normas básicas para estimar el margen de error, que es una medida estandarizada del error de muestreo y cuantifica la incertidumbre alrededor de las estimaciones en una encuesta. La forma de calcular el error estándar dependerá de la complejidad del diseño de muestreo (por ejemplo, estratificación, selección proporcional al tamaño, múltiples etapas) y del estimador utilizado (como ajuste de pesos por ausencia de respuesta, calibración, razón de totales, medias, percentiles, coeficientes de regresión).

En general, podría afirmarse que tradicionalmente existen cuatro 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 (Kish 1965). 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 (Särndal y otros 2003). Asimismo, es posible apoyarse en los métodos computacionales modernos y aplicar los principios de los pesos replicados (Wolter 2007) para aproximar la varianza de cualquier estimador en una encuesta de hogares. Por último, también es posible utilizar un enfoque de modelización del error de muestreo, conocido como función generalizada de varianza (FGV) (Wolter 2007).

Este documento presenta la estimación del error estándar de algunos indicadores sociales de interés, utilizando una FGV en donde se involucraron los microdatos de las encuestas de hogares de 18 países en la región. Para obtener la FGV se utilizó información del repositorio BADEHOG de la CEPAL, que incluye bases de datos de encuestas de hogares, aunque no en todos los casos se dispone de información para especificar el diseño de muestreo.

Este documento se organiza de la siguiente manera: el capítulo 1 presenta una breve introducción; el capítulo 2 describe un preámbulo al problema de la estimación del error estándar; en el capítulo 3 se muestran los conceptos fundamentales de la FGV y su relación con el efecto de diseño; el capítulo 4 presenta la estimación de la varianza usando tres (3) modelos mediante un enfoque a través de una FGV con los datos del repositorio BADEHOG de la CEPAL. En el capítulo 5 y capítulo 6 respectivamente se presentan los resultados principales y las conclusiones de esta investigación. Finalmente, se incluyen dos anexos en los que se entregan las estimaciones de los efectos de diseño para los indicadores de interés en los 18 países de América Latina.

I. Técnicas tradicionales en la estimación de error estándar de los estimadores de muestreo

Calcular la varianza de los estimadores de muestreo no es siempre una tarea fácil. Para abordar este problema, se sigue un enfoque lógico basado en las probabilidades de inclusión de primer y segundo orden, tal como se establece en la teoría clásica de muestreo. Por ejemplo, considere el estimador de Horvitz-Thompson (HT), dado por la siguiente expresión:

$$\hat{t}_{y,\pi} = \sum_{k \in s} d_k y_k \quad (1)$$

En donde d_k corresponde al factor de expansión del elemento k , definido como el inverso de la probabilidad de inclusión de primer orden del elemento, $\pi_k = P(k \in s)$. En general, para cualquier diseño de muestreo sin reemplazo, la fórmula exacta para calcular la varianza del estimador de HT está dada por:

$$V(\hat{t}_{y,\pi}) = \sum_U \sum_U \Delta_{kl} \frac{y_k y_l}{\pi_k \pi_l} \quad (2)$$

En donde $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$. 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, $\pi_{kl} = P(k \in s, l \in s)$. En donde el subíndice $k, l \in s$ se refiere a la suma sobre todas las muestras que contienen a los k -ésimo y l -ésimo elementos.

Es evidente que, debido a limitaciones computacionales y a la imposibilidad de acceder a todas las observaciones de cada elemento en la población finita, calcular la varianza de los estimadores resulta inviable. En la práctica, nunca se puede obtener y calcular la varianza exacta de un estimador directo en las encuestas de hogares y siempre será necesario estimarla.

Por lo tanto, según Gutiérrez (2009), la siguiente expresión proporciona un estimador insesgado para esta varianza:

$$\hat{V}_1(\hat{t}_{y,\pi}) = \sum_S \sum_S \frac{\Delta_{kl} y_k y_l}{\pi_{kl} \pi_k \pi_l} \quad (3)$$

Asimismo, si el diseño es de tamaño de muestra fijo, un estimador insesgado está dado por

$$\hat{V}_2(\hat{t}_{y,\pi}) = -\frac{1}{2} \sum_S \sum_S \frac{\Delta_{kl}}{\pi_{kl}} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2 \quad (4)$$

Por ejemplo, si el diseño de muestreo es aleatorio simple, la fórmula de la estimación de la varianza es:

$$\hat{V}(\hat{t}_{y\pi}) = \frac{N^2}{n} \left(1 - \frac{n}{N} \right) S_{ys}^2 \quad (5)$$

En donde S_{ys}^2 es la varianza de los valores de la característica de interés en la muestra aleatoria s , dada por:

$$S_{ys}^2 = \frac{1}{n-1} \sum_{k \in S} (y_k - \bar{y}_s)^2 \quad (6)$$

Uno de los enfoques más comunes en la práctica estadística es la utilización de la técnica del último conglomerado (*ultimate cluster technique*) para estimar la varianza de los estimadores de interés en encuestas con múltiples etapas de selección, estratificadas y con factores de expansión desiguales (Valliant, y otros 2013). Esta técnica solo considera que la varianza de los estimadores es relevante únicamente en la primera etapa del muestreo, además de suponer que el muestreo se realizó con reemplazo. Las etapas posteriores de selección de la muestra se ignoran, a menos que se requiera un factor de corrección para poblaciones finitas en la primera etapa del muestreo, porque se asume que el porcentaje adicional de varianza de estas restantes etapas es insignificante.

Nótese que, según la teoría de muestreo, si el diseño es con reemplazo, existe o se presenta otro estimador insesgado además del estimador HT, conocido como el estimador de Hansen-Hurwitz (HH) (Gutiérrez, 2009). A diferencia del estimador HT, el estimador HH tiene una expresión de varianza fácil de calcular, lo que hace que la estimación de la varianza sea más manejable computacionalmente. En un diseño de muestreo en varias etapas, el estimador de Hansen-Hurwitz para el total poblacional está dado por la siguiente expresión:

$$\hat{t}_{y,p} = \frac{1}{m_I} \sum_{i=1}^{m_I} \frac{\hat{t}_{y_i}}{p_{I_i}} \quad (7)$$

En donde p_{I_i} 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:

$$\hat{V}(\hat{t}_{y,p}) = \frac{1}{m_I(m_I - 1)} \sum_{i=1}^{m_I} \left(\frac{\hat{t}_{y_i}}{p_{I_i}} - \hat{t}_{y,p} \right)^2 \quad (8)$$

Ahora, dado que la forma del estimador ha sido equiparada con un estimador tipo Hansen-Hurwitz, es posible utilizar su estimación de varianza como aproximación general a cualquier estrategia de muestreo planificada en las encuestas de hogares. Aún más, después de un poco de álgebra es posible tener la siguiente aproximación (conocida como el estimador de varianza del último conglomerado) cuya gran ventaja es que sólo hace uso de los factores de expansión finales d_k , que

suelen ser reportados por las ONE 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 unidades primarias de muestreo (UPM).

$$\widehat{V}(\hat{t}_{y,p}) = \frac{m_1}{m_1 - 1} \sum_{i=1}^{m_1} (\check{t}_{y_i} - \bar{\check{t}}_y)^2 \quad (9)$$

En donde $\check{t}_{y_i} = \sum_{k \in S_i} d_k y_k$ se define como la contribución de la i -ésima UPM a la estimación del total poblacional¹ y $\bar{\check{t}}_y = \frac{1}{m_1} \sum_{i=1}^{m_1} \check{t}_{y_i}$. Adaptando esta aproximación a los elementos transversales de las encuestas de hogares en América Latina, si el escenario de muestreo planteado fuese estratificado, con dos 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 la siguiente expresión:

$$\widehat{V}(\hat{t}_{y,p}) = \sum_h \frac{n_h}{n_h - 1} \sum_{i \in S_h} (\hat{t}_{y_i} - \bar{\hat{t}}_{y_h})^2 \quad (10)$$

En donde $\hat{t}_{y_i} = \sum_{k \in S_{hi}} d_k y_k$, $\bar{\hat{t}}_{y_h} = (1/n_h) \sum_{i \in S_h} \hat{t}_{y_i}$ y n_h es el número de UPM seleccionadas en el estrato h . El anterior procedimiento, propuesto por Hansen, Hurwitz, y Madow (1953) tiende a sobrestimar la varianza verdadera. Sin embargo, 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 las ONE.

Cuando se necesita estimar parámetros que tienen una forma no lineal, se puede utilizar el análisis matemático para aproximar las varianzas de los estimadores y hacer inferencia sobre cifras oficiales con errores estándar apropiados. Según Valliant, Dever y Kreuter (2013), esta técnica implica expresar el estimador como una 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, \dots, t_Q) , entonces su estimador de muestreo se debe expresar como $\hat{\theta} = f(t_1, t_2, \dots, t_Q)$.

Debido a la falta de fórmulas adecuadas, se han desarrollado diversas técnicas empíricas que ofrecen varianzas aproximadas que parecen ser suficientemente consistentes para propósitos prácticos (Kish 2005). Estos métodos usan una muestra de datos para crear submuestras y generar una distribución para estimar los parámetros de interés en cada submuestra. Los resultados de las submuestras se analizan para obtener una estimación del parámetro y los intervalos de confianza correspondientes. La técnica computacional general consiste en dividir la muestra en pequeños subconjuntos (réplicas), aplicar los mismos procesos de ajuste de ponderadores en cada réplica y realizar las estimaciones en cada subgrupo.

A este conjunto de técnicas se les conoce como métodos de re-muestreo (Särndal y otros, 2003). La varianza del estimador se calcula fácilmente como la varianza muestral de todas las estimaciones en cada repetición. Esta metodología no requiere que las bases de datos públicas contengan información sobre estratos o UPM, lo que garantiza el anonimato de los encuestados. Además, no es necesario conocer el diseño de muestreo utilizado en la encuesta ya que, al proporcionar los pesos replicados en las bases de datos, los investigadores pueden estimar el error de muestreo de manera automatizada y sin complejas fórmulas matemáticas. Estos métodos son eficientes y precisos para la mayoría de los parámetros de interés y algunas encuestas que utilizan estas metodologías incluyen 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 utilizado estas técnicas para estimar la varianza de algunos estimadores complejos.

¹ Note que la suma de estas contribuciones en la muestra de la primera etapa da como resultado la estimación del total \hat{t}_y .

Existen tres grandes enfoques para abordar el problema específico de re-muestreo: los pesos replicados repetidos balanceados (Judkins 1990; McCarthy 1969), el Jackknife (Krewski y Rao 1981) y el Bootstrap (Rao y Wu 1988). Estos métodos buscan seleccionar un conjunto de unidades primarias de muestreo (UPM) en cada réplica de la muestra completa, manteniendo todos los hogares y personas seleccionados dentro de esas UPM. Posteriormente, los pesos de muestreo ponderan para preservar la representatividad en la población finita, generando así un nuevo conjunto de pesos de muestreo para cada réplica. Con estos nuevos pesos, se puede calcular la estimación de interés, obteniendo múltiples estimaciones a partir de las réplicas definidas. Wolter (2007) proporciona los detalles teóricos necesarios para la estimación de la varianza utilizando los pesos replicados.

Con base en lo anteriormente expuesto, para poder generar estimaciones apropiadas (insesgadas) de los errores estándar de los estimadores directos en las encuestas de hogares, sería necesario tener la información del diseño de muestreo condensada en a) columnas de información en la base de datos conteniendo la identificación de los estratos, UPM, y factores de expansión o b) en sendas columnas de pesos replicados. Sin embargo, ninguna de las anteriores prácticas se ha estandarizado por completo en los países de América Latina, y el investigador se puede ver en serias dificultades a la hora de estimar la precisión de sus inferencias.

En caso de que los requisitos mencionados no estén presentes, existe un tercer enfoque que permitiría tener un acercamiento a la estimación apropiada del error estándar: este es la modelación de la varianza basada en la función generalizada de la varianza (FGV) y descrita con detenimiento en Lohr (2021). La FGV es fácil de usar ya que solo depende del tamaño de la muestra y del valor de la estimación puntual, lo que permite a los usuarios estimar el error estándar sin conocer previamente el diseño de la muestra.

II. Función generalizada de varianza y su relación con el efecto de diseño

Las encuestas de hogares utilizan diseños de muestreo probabilísticos, que por lo regular tienen algunos elementos transversales en América Latina, esto es consideran estratos, varias etapas de selección de conglomerados y tienen factores de expansión desiguales. Aunque, se debe resaltar que, la definición de cada una de estas características depende de cada país. Debido a que los microdatos de las encuestas de hogares publicados por las ONE reflejan el comportamiento de una muestra probabilística, la cual es una investigación parcial sobre la población de interés, los estimadores calculados a partir de estas fuentes tendrán un margen de error asociado y por lo tanto cada una de las estimaciones puntuales debería estar acompañada por su respectivo error estándar que, como se vio en la introducción, en diseños de muestreo complejos no es una tarea sencilla (Lohr 2021).

A. Conceptos fundamentales

La función de varianza generalizada ofrece una alternativa para estimar las varianzas de los estimadores de muestreo basados en la medida de probabilidad del diseño de muestreo (p). Esta alternativa implica el uso de un modelo estadístico basado en una medida de probabilidad diferente (m), lo que simplifica el proceso de generación de miles de estimaciones del error estándar a partir de las encuestas de hogares en la región. Según Wolter (2007), si se pueden estimar los parámetros del modelo a partir de encuestas pasadas o de un conjunto de datos reducido, entonces es posible calcular las estimaciones de la varianza (y, por ende, las estimaciones del error de muestreo) simplemente evaluando el modelo con los datos actuales de la encuesta.

Una situación específica de este tipo de relaciones surge cuando se necesitan estimaciones de encuestas a nivel subnacional para crear tablas de salida que contengan la estimación puntual y el error estándar estimado. En estos casos, es frecuente que la cantidad de celdas en las tablas de salida sea extensa, lo que hace que el uso de una función generalizada de varianza (FGV) sea una opción más eficiente en términos de computación. Aunque, se recomienda prestar atención cuando el tamaño de la

muestra de las subpoblaciones de interés es pequeño o cuando no hay suficiente heterogeneidad entre las observaciones de la muestra en la subpoblación. En estas situaciones, es probable que las estimaciones de la varianza de los estimadores de los totales, tamaños, proporciones o medias sean imprecisas, llegando incluso a ser nulas. En estos casos, se deben verificar estas estimaciones con expertos o reemplazarlas por una mejor aproximación.

En términos de notación, sea $V_{FGV}(\hat{\theta})$ la varianza suavizada del estimador directo $\hat{\theta}$. Partiendo del hecho de que se tiene acceso a un estimador insesgado de $V(\hat{\theta})$, denotado por $\hat{V}(\hat{\theta})$ se tiene que:

$$V_{FGV}(\hat{\theta}) = E_{mp}(\hat{V}(\hat{\theta})) \quad (11)$$

La igualdad anterior incluye los subíndices m y p , los cuales se refieren a las medidas de probabilidad del modelo y del diseño de muestreo, respectivamente. Es importante destacar que, aunque el diseño de muestreo produce estimaciones insesgadas de las varianzas, estas tienden a ser inestables cuando el tamaño de la muestra es pequeño, lo cual es común en la desagregación de estimaciones. En su artículo Rivest y Belmonte (2000) proponen el uso de modelos de suavizado para estimar las varianzas directas, los cuales se definen de la siguiente manera:

$$\log(\hat{V}(\hat{\theta})) = z'\beta + \varepsilon \quad (12)$$

En donde z es un vector de covariables explicativas, β es un vector de parámetros que deben ser estimados, ε son errores aleatorios con media cero y varianza constante, que se asumen idénticamente distribuidos condicionalmente sobre z . Del anterior modelo, la estimación suavizada de la varianza de muestreo está dada por la esperanza del estimador, dada por la siguiente expresión:

$$V_{FGV}(\hat{\theta}) = E_{mp}(\hat{V}(\hat{\theta})) = \exp(z'\beta) \cdot E_{mp}(\varepsilon) = \exp(z'\beta) \cdot \Delta \quad (13)$$

En donde, $E_{mp}(\varepsilon) = \Delta$. No hay necesidad de especificar una distribución paramétrica para los errores de este modelo. Suponiendo que existen D subpoblaciones para las cuales es posible estimar la varianza directa, entonces al utilizar el método de los momentos, se tiene el siguiente estimador insesgado para Δ :

$$\hat{\Delta} = \frac{\sum_{d=1}^D \hat{V}(\hat{\theta}_d)}{\sum_{d=1}^D \exp(z'_d \hat{\beta})} \quad (14)$$

De la misma forma, al utilizar mínimos cuadrados ordinarios, la estimación del coeficiente de parámetros de regresión está dada por la siguiente expresión:

$$\hat{\beta} = \left(\sum_{d=1}^D z_d z'_d \right)^{-1} \sum_{d=1}^D z_d \log(\hat{V}(\hat{\theta}_d)) \quad (15)$$

Por último, el estimador suavizado de la varianza muestral está definido por:

$$\hat{V}_{FGV}(\hat{\theta}_d) = \hat{E}_{mp}(\hat{V}(\hat{\theta}_d)) = \exp(z'_d \hat{\beta}) \cdot \hat{\Delta} \quad (16)$$

Rivest y Belmonte (2000) concluyeron que este estimador no sobrestima ni subestima la varianza suavizada, puesto que el promedio de las estimaciones suavizadas $\hat{V}_{FGV}(\hat{\theta}_d)$ coincide con el promedio de las varianzas directas estimadas. Por lo que se satisface que:

$$\frac{\sum_{d=1}^D \hat{V}_{FGV}(\hat{\theta}_d)}{D} = \frac{\sum_{d=1}^D \hat{V}(\hat{\theta}_d)}{D} \quad (17)$$

El Instituto de Estadísticas de Canadá (STATCAN) utiliza modelos de este tipo para informar las cifras oficiales del mercado laboral en 149 áreas censales (Beaumont y Bocci 2016). En primer lugar, se

excluyen del modelo todas las áreas que tienen menos de 10 personas en la fuerza laboral (denominador del indicador). Además, se eliminan del modelo todas las áreas en las que el estimador de varianza directa es cero, lo que significa que no se encontró ningún caso efectivo en el numerador del indicador. La estimación de la varianza directa se basa en el diseño de muestreo complejo, mientras que la estimación de la varianza suavizada se ajusta al modelo de regresión presentado en la ecuación (2). En este caso, el vector de variables explicativas se define como:

$$z'_d = \left(1, \log\left(\frac{N_d^{\text{EIB}}}{N_d^{15+}}\right), \log\left(1 - \frac{N_d^{\text{EIB}}}{N_d^{15+}}\right), \log(N_d^{15+}) \right) \quad (18)$$

Con N_d^{EIB} , el número de beneficiarios del seguro de desempleo en el área d y N_d^{15+} , el número de personas en la fuerza de trabajo. Para las áreas con un tamaño de casos efectivo mayor a 400, para evitar posibles sesgos en las áreas con tamaño de muestra grande, se decidió que la estimación suavizada por una FGV que fuese igual a la estimación directa; es decir, $\widehat{V}_{\text{FGV}}(\hat{\theta}) = \widehat{V}(\hat{\theta})$.

Fúquene-Patiño y otros (2021) emplearon un modelo de área que utiliza el enfoque de FGV para estimar la cantidad de migrantes internacionales en los municipios de Colombia en una aplicación práctica. Teniendo en cuenta que $\hat{\theta}_d$ representa la estimación directa en el dominio d , y n_d el tamaño de muestra efectivo, su modelo se formuló como una relación log-lineal que incorpora el siguiente vector de covariables auxiliares:

$$z'_d = \left(1, \hat{\theta}_d, \sqrt{\hat{\theta}_d}, n_d, \sqrt{n_d}, \sqrt{\hat{\theta}_d n_d} \right) \quad (19)$$

En Fay y Herriot (1979), se describe una de las aplicaciones más citadas de los modelos de estimación en áreas pequeñas. El United States Census Bureau realizó un censo con una muestra cocensal del 20% en cada estado con el fin de estimar el ingreso per cápita. Para obtener estimaciones a nivel desagregado, se empleó la estimación directa en combinación con una FGV como método de suavizado de las varianzas. Este enfoque se basó en los resultados de ocho estados y se generalizó para el resto del país. La modelación resultante permitió establecer el coeficiente de variación y la varianza como una función del tamaño del área.

En Latinoamérica el Ministerio de Desarrollo Social y Familia (MDSF) de Chile y la CEPAL emplearon un modelo estadístico que utilizó una FGV para estimar las variaciones de las tasas de pobreza en diferentes comunas de Chile, utilizando la información proporcionada por la encuesta CASEN 2020 (MDSF-CEPAL 2022). El modelo utilizó como variable dependiente el logaritmo natural de la estimación de la varianza directa de las tasas de pobreza, mientras que como covariables se incluyeron el intercepto, la estimación directa de la tasa de pobreza, el tamaño de muestra comunal, la interacción entre la tasa de pobreza estimada y el tamaño de muestra, la raíz cuadrada de la tasa de pobreza estimada, la raíz cuadrada del tamaño de muestra y la raíz cuadrada de la interacción entre la tasa de pobreza y el tamaño de muestra.

En la mencionada aplicación no se incluyó las comunas que tuvieron una tasa de pobreza nula y, por lo tanto, una estimación nula de la varianza del estimador directo. Sin embargo, se obtuvieron predicciones de las variaciones de estas comunas. El informe proporciona esquemas descriptivos que justifican la inclusión de las covariables y las relaciones establecidas en el modelo. Además, el factor de ajuste $\hat{\Delta}$ fue cercano a 1.2 en todas las series estudiadas.

En esta investigación nos centramos en encontrar una FGV que le permita a los investigadores y público en general, poder estimar los errores estándar y determinar el efecto de diseño para diferentes estimaciones de interés cuando se usan las encuestas de hogares. El modelo de FGV que se utilizará para suavizar la varianza del estimador tendrá como covariables al intercepto, al tamaño de la muestra,

a la estimación directa y sus posibles interacciones. Tanto la estimación directa como el tamaño de la muestra efectiva son dos insumos que el investigador puede obtener con facilidad desde las bases de datos de las encuestas de hogares. De esta manera, si los parámetros del modelo para una FGV pueden ser estimados, entonces se pueden obtener las estimaciones de la varianza rápidamente al reemplazar los valores apropiados (estimaciones puntuales y tamaños de muestra) en la forma funcional final de los modelos.

No obstante, aunque este enfoque es fácil de usar, se debe tener en cuenta que se requiere de una FGV diferente para cada variable de interés. Por lo tanto, en esta investigación se realiza el análisis para un conjunto limitado de indicadores sociales; entre ellos ingreso per cápita, pobreza monetaria, pobreza extrema, las dimensiones de los indicadores de necesidades básicas insatisfechas (NBI), y estados de ocupación en la fuerza de trabajo.

B. El efecto de diseño

Cuando se lleva a cabo la planeación o análisis de las encuestas de hogares, es crucial considerar la complejidad del diseño de muestreo. Para lograr la precisión deseada en la inferencia de los datos, y ante la imposibilidad de contar con un marco de muestreo que permita la selección directa de hogares y personas, es necesario aumentar el tamaño de la muestra. Además, como se mencionó anteriormente, es crucial tener en cuenta la información del diseño de muestreo para realizar una estimación correcta de los errores de muestreo. Una manera simple de incorporar esta información es mediante el uso del efecto del diseño, el cual fue definido por Kish (1965). Esta cantidad permite ajustar la varianza del estimador para reflejar la complejidad del diseño de muestreo. Por lo tanto, es fundamental considerar el efecto del diseño al analizar datos obtenidos de encuestas de hogares para garantizar una inferencia precisa y confiable. Esta cantidad se obtiene a partir de la siguiente expresión:

$$DEFF = \frac{V_p(\hat{\theta})}{V_{MAS}(\hat{\theta})} \quad (20)$$

En donde $V_p(\hat{\theta})$ denota la varianza de un estimador $\hat{\theta}$ bajo un diseño de muestreo p y $V_{MAS}(\hat{\theta})$ denota la varianza de este estimador $\hat{\theta}$ cuando el diseño de muestreo es aleatorio simple sin reemplazo (MAS). Esta medida permite cuantificar el efecto causado por utilizar un diseño de muestreo complejo para realizar la inferencia sobre un parámetro θ , en vez de un diseño simple. Es decir, da cuenta del efecto de aglomeración causado por la utilización de un diseño de muestreo complejo p , en la inferencia de un parámetro de interés θ (que puede ser un total, un promedio, una proporción, una razón, un percentil, etc.). UN (2009, 49) afirma que hay varias formas de interpretar el efecto del diseño:

- Como el factor por el cual la varianza del diseño de muestreo complejo es mayor que la de una muestra aleatoria simple del mismo tamaño.
- Como la medida de cuánto peor es el plan de muestreo real que la muestra aleatoria simple en términos de precisión.
- Como un reflejo de cuántos casos de muestra más tendrían que seleccionarse en el diseño de muestra planificado en comparación con una muestra aleatoria simple para lograr el mismo nivel de varianza de muestreo.

En general, la varianza del estimador $\hat{\theta}$ basado en el diseño de muestreo complejo p , se puede escribir en función del efecto del diseño. De la ecuación anterior se tiene que

$$V_p(\hat{\theta}) = V_{MAS}(\hat{\theta}) \cdot DEFF \quad (21)$$

Por ejemplo, suponiendo que la variable de interés y_k sea el ingreso per cápita al mes y que el parámetro de interés es la media poblacional, esto es $\hat{\theta} = \hat{y}_U$, entonces el estimador de la varianza de \hat{y}_U se puede escribir como sigue:

$$V(\hat{y}_U) = \frac{DEFF}{n} \left(1 - \frac{n}{N}\right) S_y^2 \quad (22)$$

En donde S_y^2 corresponde a la varianza de la característica de interés en toda la población finita, N es el tamaño de la población y n es el tamaño de la muestra.

Por otro lado, cuando la variable de interés es dicotómica, donde y_k es igual a 1 cuando el elemento k experimenta un fenómeno y y_k es igual a cero en caso contrario, el parámetro de interés puede corresponder con la proporción poblacional, esto es $\theta = P$, como en el caso del porcentaje de individuos de bajo de la línea de pobreza en un país; de esta forma, es posible escribir la varianza del estimador bajo el diseño de muestreo complejo como:

$$V(\hat{P}) = \frac{DEFF}{n} \left(1 - \frac{n}{N}\right) P(1 - P) \quad (23)$$

En donde P representa la proporción poblacional de individuos pobres en el país. Por otra parte, en un diseño de muestreo con dos etapas de selección, el efecto de diseño tradicionalmente se expresa de la siguiente manera:

$$DEFF = 1 + (\bar{m} - 1) \rho_y \quad (24)$$

En donde \bar{m} corresponde al promedio de hogares seleccionados en la segunda etapa de selección dentro de las UPM y ρ_y es el coeficiente de correlación intraclase entre la variable de interés y las UPM. En general, note que el efecto de diseño tiende a ser un valor más alto cuando:

- El coeficiente de correlación aumenta, lo cual depende de la estructura poblacional y disposición de las UPM en el marco de muestreo, por lo que esta cantidad no puede ser controlada de antemano. Por lo general, ρ_y será más grande cuando la distribución de la variable de interés puede ser explicada por las UPM del país. Esto es, si el indicador de interés fuera la pobreza y los hogares pobres están segregados geográficamente, que a su vez significa que están separados de los hogares más acaudalados, entonces ρ_y tendrá una tendencia a ser cada vez más grande.
- El promedio de hogares seleccionados por UPM ascienda. Esto sí puede ser controlado de antemano en la etapa de diseño; una vez definido este valor será considerado como un número fijo y transversal en la encuesta.

La estimación del efecto de diseño es un problema común cuando se trabaja con estimaciones desagregadas en subpoblaciones de interés. Así que la hemos considerado como un aspecto fundamental dentro de esta investigación. Como se presenta en el cuadro 1, algunos países no contienen la información de UPM ni estratos dentro de sus bases de datos públicas, aunque todas presentan la información concerniente al factor de expansión. Por lo tanto, en estos casos no será posible estimar la varianza de los estimadores bajo la estrategia muestral utilizada en la encuesta. Para lidiar con este problema, algunos autores han considerado que se puede estimar los errores de muestreo cuando la base de datos únicamente incorpora los factores de expansión y omite la restante información del diseño de muestreo.

West y McCabe (2012) usan la propiedad dada en la ecuación (3) para aproximar el efecto un diseño complejo. De otra parte, Park et al. (2003) consideran que el efecto del diseño se puede descomponer en tres partes, la primera debida al efecto de los factores de expansión, denotado como $DEFF^w$; la segunda debida al efecto de la estratificación, denotado como $DEFF^s$; y la tercera debida al

efecto de selección en múltiples etapas de muestreo, denotado como $DEFF^c$. Por lo tanto, se mantiene la siguiente relación:

$$DEFF = DEFF^w \cdot DEFF^s \cdot DEFF^c \quad (25)$$

Teniendo en cuenta la ecuación (4), el efecto de diseño debido al efecto del factor de expansión se expresa de la siguiente manera:

$$DEFF^w = \frac{V^w(\hat{\theta})}{V_{MAS}(\hat{\theta})} \quad (26)$$

En donde $V_w(\hat{\theta})$ se refiere a la varianza de un diseño de muestreo con ponderadores desiguales. Debido a que la variable de ponderación se encuentra en las bases de datos de hogares para todos los países de esta investigación, se puede obtener estimaciones de $DEFF^w$ y de $V_w(\hat{\theta})$ para cualquier estimador. Para ello se pueden usar los paquetes `survey` (Lumley 2020) y `srvyr` (Freedman Ellis y Schneider 2022) del software estadístico R. Nótese que, a partir de lo anterior, se puede obtener $V_{MAS}(\hat{\theta})$, como:

$$V_{MAS}(\hat{\theta}) = \frac{V^w(\hat{\theta})}{DEFF^w} \quad (27)$$

El efecto de diseño definido en la expresión (3) deberá ser estimado apropiadamente. En efecto, un estimador consistente de $V_p(\hat{\theta})$ será $\hat{V}_{FGV}(\hat{\theta})$, la estimación generada a partir de la función generalizada de varianza FGV. De esta forma, el estimador del efecto de diseño cumplirá la siguiente relación:

$$\widehat{DEFF} = \frac{\hat{V}_{FGV}(\hat{\theta})}{\hat{V}_{MAS}(\hat{\theta})} = \frac{\hat{V}_{FGV}(\hat{\theta})}{V_w(\hat{\theta})} \cdot DEFF^w \quad (28)$$

Cuadro 1
Cantidad de registros por país e información del diseño de muestreo

| País | Año | Encuesta | Observaciones | Población Expandida | Disponibilidad UPM y estrato |
|----------------------|------|---|---------------|---------------------|------------------------------|
| Argentina | 2020 | Encuesta Permanente de Hogares | 43 794 | 28 739 664 | No |
| Bolivia | 2020 | Encuesta Nacional de Hogares | 37 092 | 11 749 038 | Sí |
| Brasil | 2020 | Pesquisa Nacional por Amostragem de Domicílios Contínua (PNADC) | 355 436 | 211 096 453 | Sí |
| Chile | 2020 | Encuesta de Caracterización Socioeconómica Nacional | 185 437 | 19 545 799 | Sí |
| Colombia | 2020 | Gran Encuesta Integrada de Hogares | 747 822 | 49 427 266 | No |
| Costa Rica | 2020 | Encuesta Nacional de Hogares | 25 530 | 5 111 405 | Sí |
| República Dominicana | 2020 | Encuesta Nacional Continúa de Fuerza de Trabajo | 71 378 | 10 447 824 | Sí |
| Ecuador | 2020 | Encuesta Nacional de Empleo, Desempleo y Subempleo | 30 646 | 17 695 545 | No |
| Guatemala | 2014 | Encuesta Nacional de Condiciones de Vida | 54 822 | 16 002 004 | Sí |
| Honduras | 2020 | Encuesta Multipropósito de Hogares | 24 094 | 9 151 940 | No |
| México | 2020 | Encuesta Nacional de Ingresos y Gastos de los Hogares | 315 743 | 126 838 467 | Sí |
| Nicaragua | 2014 | Encuesta Nacional de Medición de Niveles de Vida de los Hogares | 29 381 | 6 240 908 | No |
| Panamá | 2019 | Encuesta Multipropósito | 42 925 | 4 232 695 | No |
| Perú | 2020 | Encuesta Nacional de Hogares - Condiciones de Vida y Pobreza | 120 346 | 33 221 604 | Sí |
| Paraguay | 2020 | Encuesta Permanente de Hogares Continua | 17 582 | 7 167 516 | Sí |
| El Salvador | 2020 | Encuesta Multipropósito de Hogares | 37 030 | 6 321 042 | No |
| Uruguay | 2020 | Encuesta Continua de Hogares | 145 166 | 3 530 883 | No |
| Venezuela | 2020 | Encuesta de Hogares por Muestreo | 133 097 | 30 067 613 | Sí |

Fuente: Elaboración propia, sobre la base del Banco de Datos de Encuestas de Hogares (BADEHOG) de la CEPAL.

III. Estimación de la varianza mediante FGV en el repositorio BADEHOG de la CEPAL

En este capítulo se realiza una breve descripción de tres modelos que serán utilizados para estimar la varianza de los estimadores de interés y se introducirán las fuentes de datos utilizadas, las cuales corresponden a encuestas de hogares para 18 países de la región que se encuentran dentro del repositorio BADEHOG de la CEPAL. Posteriormente, se hace una descripción de los códigos computacionales que se desarrollaron para implementar la metodología.

A. Tres modelos FGV

Con fines ilustrativos, considere la estimación de una proporción $P = t/N$, en donde:

$$t = \sum_{k=1}^N y_k = P \cdot N \quad (29)$$

denota el total poblacional asociado a una característica de interés y de naturaleza dicotómica y N denota el tamaño de la población finita. Considere además el estimador de HT para el total poblacional dado en la ecuación (1). La varianza del estimador \hat{t}_i puede ser escrita como sigue:

$$V(\hat{t}) \approx N^2 \frac{P(1-P)}{n} \cdot \text{DEFF} = a \cdot t^2 + b \cdot t \quad (30)$$

En donde n es el tamaño de la muestra, y convenientemente se han definido las siguientes cantidades:

$$\begin{aligned} a &= -\frac{\text{DEFF}}{n} \\ b &= \frac{N}{n} \cdot \text{DEFF} \end{aligned} \quad (31)$$

Ahora, suponga que no existe sólo una proporción de interés, sino k de ellas; una para cada dominio de representatividad en la encuesta. Por ejemplo, suponga que en la encuesta no sólo es posible estimar la pobreza monetaria a nivel nacional, sino que también es posible hacerlo a nivel de otros dominios de interés: por ejemplo, zona (urbano, rural), sexo (hombre, mujer), grupos de edad, divisiones administrativas mayores y ciudades. Cada uno de estos dominios de interés tendrá asociada una estimación de la pobreza monetaria, junto con sus correspondientes varianzas estimadas. Teniendo en cuenta lo anterior, es posible aproximar todas las varianzas de los D totales involucrados de la siguiente manera:

$$V(\hat{t}_d) \approx a_d \cdot t_d^2 + b_d \cdot t_d, \quad \text{para } d = 1, \dots, D. \quad (32)$$

Si los efectos de diseño DEFF para las D totales de interés son similares, entonces $a \cong a_d$ y $b \cong b_d$ serán consideradas constantes que pueden ser estimadas con los siguientes pasos (Wolter 2007):

- i) Estimar las varianzas de los D totales de interés $\hat{t}_1, \dots, \hat{t}_D$, usando algún método de estimación para la varianza (linealización, réplicas, etc.). Definiendo \hat{v}_d como el cuadrado del coeficiente de variación estimado (también conocido como la varianza relativa de \hat{t}_d), entonces, para $d = 1, \dots, D$, se tendrá que:

$$\hat{v}_d = \frac{\hat{V}(\hat{t}_d)}{\hat{t}_d^2} \quad (33)$$

- ii) Dado que es posible asumir que la varianza relativa es una función decreciente del valor esperado de \hat{t}_d (entre más pequeña la estimación, más grande la varianza relativa), es posible adoptar un modelo de regresión lineal de la siguiente forma:

$$v_d = \beta_0 + \frac{\beta_1}{\hat{t}_d} + \varepsilon_d \quad (34)$$

- iii) Una vez ajustado el anterior modelo, se tendrán estimaciones de los coeficientes de regresión, dadas por $\hat{\beta}_0$ y $\hat{\beta}_1$; luego, a partir del modelo ajustado estimar la varianza relativa de cualquier otro total, por ejemplo \hat{t}_{d^*} , como $\hat{v}_{d^*} = \hat{\beta}_0 + \frac{\hat{\beta}_1}{\hat{t}_{d^*}}$. Como \hat{v}_{d^*} es el valor estimado para $\frac{\hat{V}(\hat{t}_{d^*})}{\hat{t}_{d^*}^2}$, la estimación de la varianza FGV se escribe de la siguiente manera:

$$\hat{V}(\hat{t}_{d^*}) = \hat{\beta}_0 \cdot \hat{t}_{d^*}^2 + \hat{\beta}_1 \cdot \hat{t}_{d^*} \quad (35)$$

En el trabajo de Valliant (1987), los autores encontraron que las varianzas estimadas a través de una FGV son más estables que las estimaciones directas de la varianza cuando los efectos de diseño para los k totales estimados son similares; en caso contrario, las estimaciones pueden ser imprecisas. Esto fue resuelto por Krenzke (1995), en donde el autor propuso una manera de flexibilizar esta condición, ampliando la posibilidad del análisis para los casos cuando los efectos de diseño entre las estimaciones no pueden ser considerados como constantes o similares. Para el caso de estimadores de razón, definidos como $\hat{R} = \hat{t}_y/\hat{t}_z$ y asumiendo que no hay correlación entre \hat{R} y \hat{t}_z , el estimador de la varianza es:

$$\hat{V}(\hat{R}) = \frac{\hat{\beta}_1}{\hat{t}_z} \cdot \hat{R} \cdot (1 - \hat{R}) \quad (36)$$

Es posible ajustar una FGV para cualquier estimador $\hat{\theta}$; aunque, es conveniente considerar que, luego de ajustar el modelo lineal sin restricciones sobre \hat{v}_d , existe la posibilidad de obtener estimaciones negativas de la varianza. Por ende, se recomienda evitar que se ajuste un modelo directo sobre \hat{v}_d y, en vez de eso, ajustar el modelo con una función de vínculo logarítmica (modelo log-lineal); de esta manera se ajusta el modelo para $\log(\hat{v}_d)$ en vez de v_d . Haciendo esta precisión, el modelo de Krenzke (LK) se puede expresar de forma log-lineal como:

$$\log(v_d^{LK}) = z'_d \beta = \beta_0 + \frac{\beta_1}{\hat{\theta}_d} + \frac{\beta_2}{\hat{\theta}_d^{1/2}} + \varepsilon_i \quad (37)$$

En donde los ε_d son los errores del modelo considerados como una muestra aleatoria de una distribución normal con media nula y varianza σ_A^2 ; es decir $\varepsilon_d \sim N(0, \sigma_A^2)$. Además, como lo indican Edelman (1967) y Wolter (2007), se pueden considerar otros modelos tratando de buscar un mejor ajuste con los datos. Apoyados en lo anterior, se ha propuesto dos modelos adicionales, uno denominado log-lineal Común (LC) y otro denominado log-lineal Empírico (LE), que además de considerar el valor de la estimación, incorporan el tamaño de la muestra y se especifican de la siguiente manera:

$$\log(v_d^{LC}) = z'_d \beta = \beta_0 + \beta_1 n_d + \beta_2 \hat{\theta}_d + \beta_3 n_d \hat{\theta}_d + \varepsilon_d \quad (38)$$

$$\log(v_d^{LE}) = z'_d \beta = \beta_0 + \beta_1 n_d + \beta_2 \hat{\theta}_d + \beta_3 (n_d \hat{\theta}_d)^{1/2} + \varepsilon_d \quad (39)$$

Dado que $\hat{v}_d = \frac{\hat{V}(\hat{\theta}_d)}{\hat{\theta}_d^2}$, las predicciones FGV de la varianza de los estimadores, bajo cualquiera de los tres modelos estudiados, se calculan como:

$$\hat{V}(\hat{\theta}_d) = \exp(z'_d \hat{\beta}) \cdot \hat{\Delta} \cdot \hat{\theta}_d^2 \quad (40)$$

En donde $z'_d \hat{\beta} = \log(\hat{v}_d)$ corresponde a las estimaciones de la varianza relativa usando el modelo ajustado con sus correspondientes estimaciones $\hat{\beta}$ para los coeficientes de regresión del modelo. Por otro lado, $\hat{\Delta} = \frac{\sum_{d=1}^D \hat{v}_d(\hat{\theta})}{\sum_{d=1}^D \exp(z'_d \hat{\beta})}$ es el factor de corrección del sesgo en una regresión log-lineal; ignorar este factor de corrección produce una subestimación de las verdaderas varianzas cuando se aplica una la FGV (Morales et al. 2021). Nótese que las sumatorias en la expresión $\hat{\Delta}$ se refieren a la cantidad de dominios de interés que entran en el modelo FGV. Para comparar los tres modelos considerados, se van a tomar en cuenta las siguientes medidas:

- Criterio de información de Akaike (AIC). Este criterio hace referencia a la cantidad relativa de información perdida por un modelo determinado, cuanta menos información pierde, mejor el modelo. Si p es el número de parámetros estimados en el modelo y \hat{L} el valor máximo de la función de verosimilitud, entonces el AIC (Akaike 1998) viene dado por:

$$AIC = 2 p - 2 \ln(\hat{L}) \quad (41)$$

- Criterio de información Bayesiano (BIC). El BIC se calcula a partir de la función de verosimilitud del modelo y el número de parámetros de este, y su objetivo es seleccionar el modelo más probable de acuerdo con los datos disponibles. Este criterio penaliza los modelos con un gran número de parámetros, favoreciendo aquellos que se ajustan mejor a los datos y que conllevan una menor complejidad (Schwarz 1978). Si k representa el número de dominios observados, el BIC está dado por:

$$BIC = \ln(k) p - 2 \ln(\hat{L}) \quad (42)$$

- RMSE o raíz del error cuadrático medio. Es una medida de la diferencia entre los valores predichos por un modelo y los valores observados. Siendo \hat{v}_d el valor estimado para v_d , entonces el RMSE se define como:

$$RMSE = \sqrt{\frac{\sum_{d=1}^D (\hat{v}_d - v_d)^2}{D}} \quad (43)$$

- Criterio de información de desviación estadística (mejor conocido en la literatura estadística como deviance). Este es una generalización de la idea del RMSE. El DIC se calcula a partir de la desviación estadística del modelo y depende de su complejidad (mayor número de parámetros); se utiliza para comparar diferentes modelos y seleccionar el que mejor ajuste los datos.

Los criterios de información AIC, BIC y DIC se utilizan para comparar modelos estadísticos y seleccionar el que mejor se ajuste a los datos. El AIC y el BIC se basan en diferentes principios, por lo que pueden arrojar diferentes resultados en la selección del mejor modelo. El DIC es una extensión del BIC para modelos jerárquicos y puede ser útil en la selección de modelos muy complejos. En general, se prefiere el modelo con el valor más bajo de AIC, BIC, RMSE o DIC, ya que esto indica un mejor ajuste del modelo a los datos observados.

B. Fuentes de información

El flujo de trabajo fue desarrollado desde una perspectiva que permite reproducir y replicar los resultados, y les permite a los usuarios que puedan agregar variables o dominios con el fin de llegar a funciones generalizadas de la varianza que no fueron estudiadas en este documento. El usuario interesado puede consultar o descargar los códigos fuente, que han sido puestos a disposición en un repositorio de GitLab² de la División de Estadísticas de la CEPAL.

Todos los países han compartido en sus bases de datos la información del factor de expansión, pero en algunos de ellos, por cuestiones de anonimización de la información, no es posible acceder a las variables que hacen referencia al diseño de muestreo (UPM y estratos). El cuadro 1 presenta un resumen de las encuestas compartidas por cada país, el año de ejecución, la cantidad de registros en la base de datos, la población a la que expande la muestra, y finalmente una columna en la que se indica si la base cuenta con la información relacionada con el diseño de muestreo.

Son diez países los que cuentan con la información del diseño de muestreo, estos son: Bolivia, Brasil, Chile, Costa Rica, República Dominicana, Guatemala, México, Perú, Paraguay y Venezuela. En estos casos se tienen conjuntamente todas las variables que permiten identificar las unidades primarias de muestreo (UPM), los estratos del diseño de muestreo y el factor de expansión. Por lo tanto, los modelos de la función generalizada de la varianza se implementarán únicamente a partir de la información de estos países. Los restantes ocho países (Argentina, Colombia, Ecuador, Honduras, Nicaragua, Panamá, El Salvador y Uruguay) no se tendrán en cuenta en la etapa del modelamiento, pero sí se generarán predicciones de las varianzas y de los efectos de diseño asociadas a los estimadores sociales considerados en este documento.

Las bases de datos disponen de información para varios dominios de interés. En esta investigación se consideran los siguientes: sexo (hombre, mujer), nivel educativo (primaria incompleta, primaria completa, baja secundaria incompleta, alta secundaria incompleta, secundaria completa, terciaria incompleta, universitaria completa), área (rural, urbana), edad (cero a cinco años, 6 a 15 años, 16 a 30 años, 31 a 45 años, 46 a 60 años, más de 60 años), etnia (afrodescendiente, indígena, resto).

En este documento se busca analizar algunos indicadores que miden las condiciones de vida por cada uno de los dominios enunciados. Para ello se ha considerado los siguientes constructos de interés:

- Ingreso y pobreza: promedio del ingreso per cápita estandarizado (i.e. ingreso per cápita dividido entre la línea de pobreza), porcentaje de personas en situación de pobreza monetaria (extrema y no extrema).

² Véase <https://git.CEPAL.org/social-surveys/FGV>.

- Mercado del trabajo: tasa de desocupación y tasa de ocupación.
- Acceso a servicios: porcentaje de personas sin acceso a servicios de saneamiento, agua potable, energía eléctrica, combustible; porcentaje de personas con materialidad precaria en pisos, techos y paredes; porcentaje de personas con tenencia de computador y acceso a internet en el hogar.
- Necesidades básicas insatisfechas: porcentaje de personas que tienen necesidades básicas insatisfechas en el acceso a diferentes servicios (saneamiento, agua potable, energía eléctrica, combustible, internet en el hogar o tenencia de computador en el hogar) o precariedad en algún material de la vivienda (pisos, techos o paredes).

Las variables disponibles en cada país para la medición del NBI y sus variables de interés (acceso a servicios y materiales precarios) se presenta en el cuadro 2. Para los años utilizados, en las bases de datos de Brasil y Uruguay no se contó con información sobre estas variables. En las aplicaciones FGV para los países, además de realizar la estimación de la varianza para cada una de las variables definidas en el cuadro 2, también se incluyó una variable dicotómica (denotada como NBI global) que toma el valor de uno, si la persona en la encuesta tiene al menos una necesidad insatisfecha y cero, si la unidad en la encuesta tiene todas las necesidades satisfechas (de acuerdo con las definiciones nacionales).

Cuadro 2
Variables disponibles sobre necesidades básicas insatisfechas en cada país

| País | Saneamiento | Agua potable | Energía eléctrica | Combustible | Internet en el hogar | Computador en el hogar | Pisos | Techo | Paredes |
|--------------------------------------|-------------|--------------|-------------------|-------------|----------------------|------------------------|-------|-------|---------|
| Argentina | X | X | | X | | | X | X | |
| Bolivia (Estado Plurinacional de) | X | X | X | | X | X | | | |
| Chile | X | X | | | | | | | |
| Colombia | X | X | X | X | X | X | X | | X |
| Costa Rica | X | X | X | X | X | X | X | X | X |
| República Dominicana | X | X | X | X | X | X | X | X | X |
| Ecuador | X | X | X | X | | X | X | X | X |
| Guatemala | X | X | X | X | X | | X | X | X |
| Honduras | X | X | X | X | X | X | X | X | X |
| México | X | X | X | X | X | X | X | X | X |
| Nicaragua | X | X | X | X | X | | X | X | X |
| Panamá | X | X | X | X | X | X | X | X | X |
| Perú | X | X | X | X | X | X | X | X | X |
| Paraguay | X | X | X | X | X | X | X | X | X |
| El Salvador | X | X | X | X | | X | X | X | X |
| Venezuela (República Bolivariana de) | X | X | X | X | X | X | X | X | X |

Fuente: Elaboración propia, sobre la base del Banco de Datos de Encuestas de Hogares (BADEHOG) de la CEPAL.

Todo el procesamiento estadístico de las bases de datos y la generación de estimaciones directas, estimación de varianzas, modelamiento, estimaciones y predicciones fue programado en lenguaje R (R Core Team 2022), y está diseñado en un flujo de trabajo que le permite al usuario agregar nuevos dominios o incluir nuevas variables de análisis, de modo que es completamente replicable y reproducible.

C. Flujo de trabajo

El flujo de trabajo se encuentra estructurado desde tres carpetas: *input*, *src* y *output*³. Esto busca facilitar la navegación por el proyecto y llevar a cabo la ejecución de cada una de las etapas a partir de varias tareas, que buscan establecer una lógica que facilite el entendimiento del flujo y permita reproducir o replicar los resultados de una forma simple. Este flujo se basa en la estructura dada por Ball (2016).

La idea es que el proyecto sea una colección de tareas independientes que se realizan en orden hasta llegar a los resultados finales. La carpeta *input/* debe contener las bases de datos del repositorio BADEHOG que serán utilizadas en el proyecto. Estos archivos no deben ser modificados y son únicamente de lectura. Los códigos fuente de R que se utilizan para ejecutar las diferentes tareas se almacenan en la carpeta *src/*, mientras que los archivos de salida que genera cada tarea se encuentran almacenados en la carpeta *output/*.

El flujo de trabajo fue programado en su totalidad en el software estadístico R y está compuesto por 6 funciones (una función matriz y cinco tareas principales) que han sido implementadas para que el código pueda ser replicado en otros proyectos que busquen el mismo objetivo, o para estimar la función generalizada de la varianza para otras variables o dominios. Los códigos para el procesamiento presentan una estructura secuencial, por lo cual, por ejemplo, para ejecutar la tarea 3 es necesario haber ejecutado previamente la tarea 2, y a su vez la tarea 1. A continuación se describe detalladamente los insumos, procesos y salidas que realiza cada uno de los códigos.

Función de funciones: 0.aux_functions.R

Este archivo contiene funciones que son utilizadas de manera recurrente a lo largo de cada uno de los diferentes códigos computacionales. A su vez, contiene funciones para la lectura de información, codificación y unificación de categorías, ordenamiento de salidas y consolidación de las estimaciones utilizadas por otros códigos.

Primera tarea: 1.sam_estimates.R

La primera etapa del proceso consiste en realizar las estimaciones a partir de las fuentes que contienen la información del diseño de muestreo. Esto es llevado a cabo en esta etapa mediante este código que lee la información desde el repositorio BADEHOG para cada país, utilizando la función de lectura contenida en el código *0.aux_functions.R*; allí mismo se definen los nombres que identifican en la base de datos a las variables relacionadas con este estudio; es decir, se definen o se crean las variables sobre necesidades básicas insatisfechas, pobreza, pobreza extrema, ingresos, ocupación, etc. También se definen las variables que van a ser consideradas como dominios de interés. A partir de la información de cada país, se obtienen las estimaciones de los indicadores de interés para cada uno de los dominios establecidos. Los resultados son alojados en el archivo *output/1.sam_estimates.RData*, el cual se convierte en el insumo de la siguiente tarea.

Segunda tarea: 2.consol_sam_estimates.R

Como la salida de la primera tarea obedece a una estructura de tipo lista, donde cada elemento contiene las estimaciones de los indicadores de interés para cada país, entonces esta tarea se encarga de consolidar las salidas y dejarlas en un formato de tabla. El resultado de la consolidación se almacena en *output/2.consol_sam_estimates.RData*.

³ Los códigos fuente se encuentran disponibles en el repositorio público de GitLab de la División de Estadísticas de la CEPAL: <https://git.CEPAL.org/social-surveys/FGV>.

Tercera tarea: 3.GVF.R

Esta tarea se consolida la salida anterior y, para cada uno de los indicadores de interés, ajusta los modelos LC, LE y LK. Con los resultados se grafican los efectos de diseño obtenidos en la primera tarea con el fin de evaluar el comportamiento a lo largo de los diferentes dominios para los indicadores de interés. El resultado de esta tarea se aloja en *output/3.GVF.RData*.

Cuarta tarea: 4.summary_models.R

La cuarta tarea permite realizar un resumen de las medidas de ajuste de los modelos, estimando los errores estándar obtenidos a través de la FGV; además, grafica los resultados que comparan las estimaciones obtenidas a través del método de linealización de Taylor con las obtenidas por los tres modelos FGV. Por último, exporta las estimaciones de los parámetros de los modelos LC, LE y LK, así como las estimaciones de la varianza suavizada. Los resultados se alojan en *output/4.summary_models.RData*.

Quinta tarea: 5.predictions.R

La quinta y última tarea corresponde a la predicción del error estándar y del efecto de diseño de los indicadores de interés, basado en el modelo seleccionado en la tarea 4, para aquellos países que no brindaron información sobre las UPM ni estratos. Los resultados se almacenan en el archivo *output/5.predictions.RData*.

Dado que los anteriores procesamientos se realizan para cada país, hay un alto costo computacional asociado, por lo que ha optado por utilizar un enfoque de procesamiento en paralelo, para evitar el uso de ciclos computacionales y obtener resultados de una manera mucho más ágil. El tiempo de procesamiento dependerá de los recursos computacionales disponibles. En un procesador Ryzen 7 3700X, con 32 GB de RAM, el procesamiento completo dura aproximadamente 40 minutos.

IV. Resultados

Como se ha señalado anteriormente, esta investigación se enfoca principalmente en algunos indicadores sociales de interés (pobreza, pobreza extrema, necesidades básicas insatisfechas, tasa de ocupación, tasa de desempleo e ingreso). Tanto las estimaciones puntuales como la estimación de las varianzas son procesos que se realizan en las ONE. Sin embargo, algunos usuarios externos pueden estar interesados en realizar análisis más detallados que no son proporcionados por las ONE; por ejemplo, en niveles de desagregación que no son reportados en las publicaciones regulares de los resultados de las encuestas de hogares.

El cuadro 3 muestra las variables de interés, su naturaleza y los niveles que la componen. Las estimaciones directas de estos indicadores se calculan para cada uno de los diez países que sí brindan la información del diseño de muestreo en sus microdatos, y para cada uno de los dominios de interés, a saber: sexo (2 niveles), área (2 niveles), etnia (3 niveles), edad (6 niveles) y nivel educativo (7 niveles). Luego, se estiman de manera insesgada sus respectivas varianzas, las cuales son obtenidas a partir de técnicas de linealización de Taylor: asimismo se estiman los correspondientes efectos de diseño. En total, se originan $k = 10 * (2 + 2 + 3 + 6 + 7) = 200$ estimaciones para cada indicador de interés que permiten ajustar cada uno de los modelos FGV.

Cuadro 3
Descripción de las variables de interés

| Variable | Descripción | Naturaleza | Niveles |
|--------------------|-----------------------------|-------------|---------------------------|
| nbi_agua_ee | NBI de agua | Cualitativa | Con carencia/Sin carencia |
| nbi_saneamiento_ee | NBI de saneamiento | Cualitativa | Con carencia/Sin carencia |
| nbi_elect_ee | NBI de electricidad | Cualitativa | Con carencia/Sin carencia |
| nbi_interhog_ee | NBI de internet hogar | Cualitativa | Con carencia/Sin carencia |
| nbi_compuhog_ee | NBI de computador hogar | Cualitativa | Con carencia/Sin carencia |
| nbi_combus_ee | NBI de combustible | Cualitativa | Con carencia/Sin carencia |
| nbi_piso_ee | NBI en material del piso | Cualitativa | Con carencia/Sin carencia |
| nbi_techo_ee | NBI en material del techo | Cualitativa | Con carencia/Sin carencia |
| nbi_pared_ee | NBI en material de la pared | Cualitativa | Con carencia/Sin carencia |

| Variable | Descripción | Naturaleza | Niveles |
|-------------|-----------------------|--------------|--|
| pobreza_ext | Pobreza extrema | Cualitativa | En pobreza extrema/Sin pobreza extrema |
| pobreza_rec | Pobreza | Cualitativa | En condición de pobreza/Sin condición de pobreza |
| conduct3 | Ocupación | Cualitativa | Ocupado/Desocupado/Inactivo |
| tdesocup | Tasa de desempleo | Cualitativa | Desocupado/Ocupado |
| ingreso | Ingreso estandarizado | Cuantitativa | NA |

Fuente: Elaboración propia, sobre la base del Banco de Datos de Encuestas de Hogares (BADEHOG) de la CEPAL.

A. FGV para los parámetros de interés

Las estadísticas para la bondad del ajuste de los modelos (6), (7) y (8) se pueden observar en los cuadros 5, 6 y 7 disponibles en el Anexo 1. Se destaca que los modelos LC y LE presentan un comportamiento similar, aunque el modelo LC mostró un mejor rendimiento. Por otra parte, el modelo LK aunque no resultó ser el mejor en los estadísticos de bondad de ajuste, presentó un comportamiento muy similar a los resultados obtenidos por el método de linealización de Taylor.

A partir de lo anterior, se tomó el modelo LC como referente para los análisis. El cuadro 4 presenta los parámetros estimados para el modelo LC cuando se usa una FGV para cada uno de los indicadores investigados. Se puede observar que la relación entre la varianza relativa y los coeficientes de regresión del modelo presenta un comportamiento similar para todos los indicadores estudiados. Excepto en el modelo de FGV para el ingreso estandarizado, donde el coeficiente asociado al tamaño de la muestra es muy cercano a cero, en los demás indicadores, el modelo presenta un signo negativo, indicando que, a mayor tamaño de la muestra, se espera una reducción en el logaritmo de la varianza relativa. El coeficiente asociado a la estimación también presenta un signo negativo en todos los casos, mostrando de igual manera una relación inversa entre el logaritmo de la varianza relativa y el valor estimado. En cuanto al efecto de la interacción entre el tamaño de la muestra y la estimación obtenida, excepto para el ingreso per cápita, en todos los demás casos se observa una relación positiva con el logaritmo de la varianza relativa.

Para obtener el error estándar estimado de alguna desagregación particular de los indicadores analizados, bastará con usar los coeficientes del modelo LC para el indicador particular y obtener la raíz cuadrada de la ecuación (8). A continuación, se realiza un breve análisis de los resultados obtenidos para cada indicador.

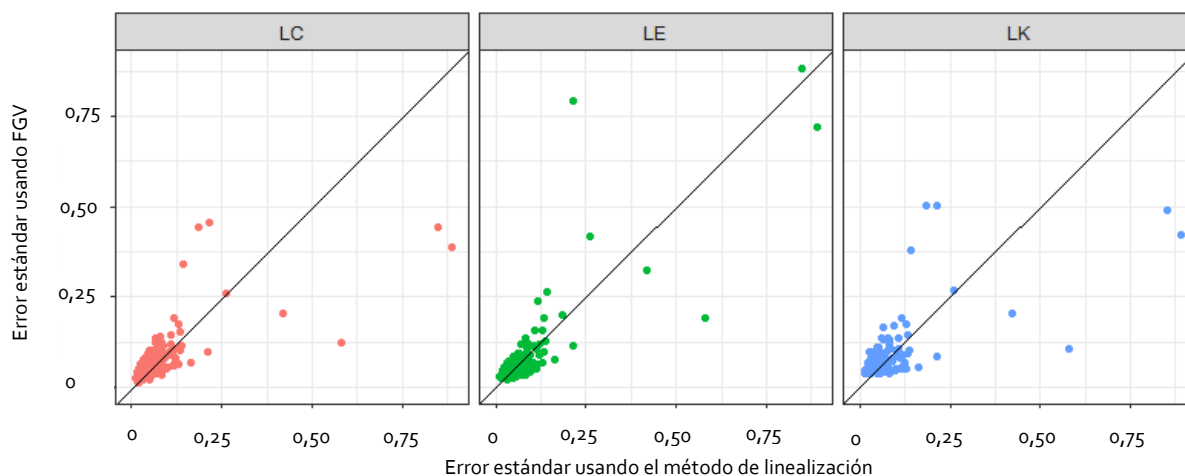
Ingreso medio per cápita estandarizado

El gráfico 1 muestra el error estándar estimado cuando se usa el método de linealización de Taylor a partir de la información del diseño muestral y cuando se usa la FGV para los diferentes modelos propuestos. Se destaca que los modelos LC y LE presentan un comportamiento similar, y para este caso, las estadísticas de bondad del ajuste mostraron que el modelo LE cuenta con un mejor rendimiento. Por otra parte, el modelo LK, aunque no resultó con el mejor rendimiento en los estadísticos de bondad de ajuste, se observa que tiende a presentar resultados cerca de la línea identidad. Así que la estimación de la varianza y en particular del error estándar a partir de una FGV resultó satisfactorio.

Los demás gráficos que comparan los resultados obtenidos por el método de linealización y por el modelo de FGV mostraron resultados similares, así que no se ilustrarán en cada caso.

Gráfico 1

Estimación del error estándar usando la FGV y el método de linealización para el ingreso



Fuente: Elaboración propia.

El cuadro 4 presenta la estimación de los parámetros del modelo LE que pueden ser utilizados para la estimación del error estándar basado en la FGV, la cual se obtiene como:

$$\widehat{SE}_{LE}(\hat{\theta}_d) = \sqrt{\exp(\hat{\beta}_0 + \hat{\beta}_1 \cdot n_d + \hat{\beta}_2 \cdot \hat{\theta}_d + \hat{\beta}_3 \cdot (n_d \cdot \hat{\theta}_d)^{1/2}) \cdot \hat{\Delta} \cdot \hat{\theta}_d^2} \tag{44}$$

En donde $\hat{\Delta}$ tiene la forma funcional dada en la ecuación (3); además, $\sigma_A^2 = 0.7564$, $\beta_0 = -7.323$, $\beta_1 = 1.1e - 04$, $\beta_2 = 0.33113$, $\beta_3 = -6.806e - 02$, n_d es el tamaño de la muestra con el cual fue calculada la estimación $\hat{\theta}_d$ para el dominio d.

Porcentaje de personas en condición de pobreza

Los indicadores de pobreza son un insumo vital para la generación de políticas públicas. Los usuarios pueden realizar estimaciones de este indicador para algunos niveles de desagregación que no son ofrecidos por las ONE, y con la metodología presentada en este documento, podrán realizar las estimaciones del error estándar a partir de los parámetros estimados.

Los estadísticos del ajuste para los modelos (6), (7) y (8), permiten observar que el modelo LC presenta el menor AIC, BIC, *Deviance* y RMSE. El cuadro 4 presenta la estimación de los parámetros del modelo LC que pueden ser utilizados para la estimación del error estándar basado en la FGV, la cual se obtiene como:

$$\widehat{SE}_{LC}(\hat{\theta}_d) = \sqrt{\exp(\hat{\beta}_0 + \hat{\beta}_1 \cdot n_d + \hat{\beta}_2 \cdot \hat{\theta}_d + \hat{\beta}_3 \cdot n_d \cdot \hat{\theta}_d) \cdot \hat{\Delta} \cdot \hat{\theta}_d^2} \tag{45}$$

En donde $\hat{\Delta}$ tiene la forma funcional dada en la ecuación (3); además, $\sigma_A^2 = 1.345$, $\beta_0 = -5.631$, $\beta_1 = -4.945$, $\beta_2 = -1.566e - 04$, $\beta_3 = 2.079e - 04$, n_d es el tamaño de la muestra con el cual fue calculada la estimación $\hat{\theta}_d$ para el dominio d.

Porcentaje de personas en condición de pobreza extrema

El indicador de pobreza extrema determina el estado más bajo de la pobreza, este indicador puede ser obtenido desde las bases de datos de las encuestas de hogares y a partir de la metodología expuesta en esta investigación será posible calcular el error estándar para cualquier nivel de desagregación usando solo la información del valor estimado y el tamaño de la muestra utilizado.

Al ajustar los modelos (6), (7) y (8), se puede observar que el modelo LC cuenta con menor AIC, BIC, *Deviance* y RMSE. Además, la estimación del error estándar por el método de linealización de Taylor y mediante la FGV permitió observar que las diferencias entre los métodos no son relevantes. A partir de la estimación de los parámetros del modelo LC para la FGV presentados en el cuadro 4, se puede observar que este indicador presenta una mayor sensibilidad que los demás indicadores cuando hay incrementos del tamaño de la muestra, puesto que su coeficiente presenta un signo negativo y por cada unidad de muestra adicional, el logaritmo de la varianza relativa presenta una mayor disminución en comparación con los demás indicadores, mientras los demás factores permanecen constantes.

Con la estimación de los parámetros del modelo LC se puede estimar el error estándar para otras desagregaciones de interés del indicador de pobreza extrema usando la raíz cuadrada de la ecuación (8). Así como se ilustró para el indicador del porcentaje de personas en condición de pobreza.

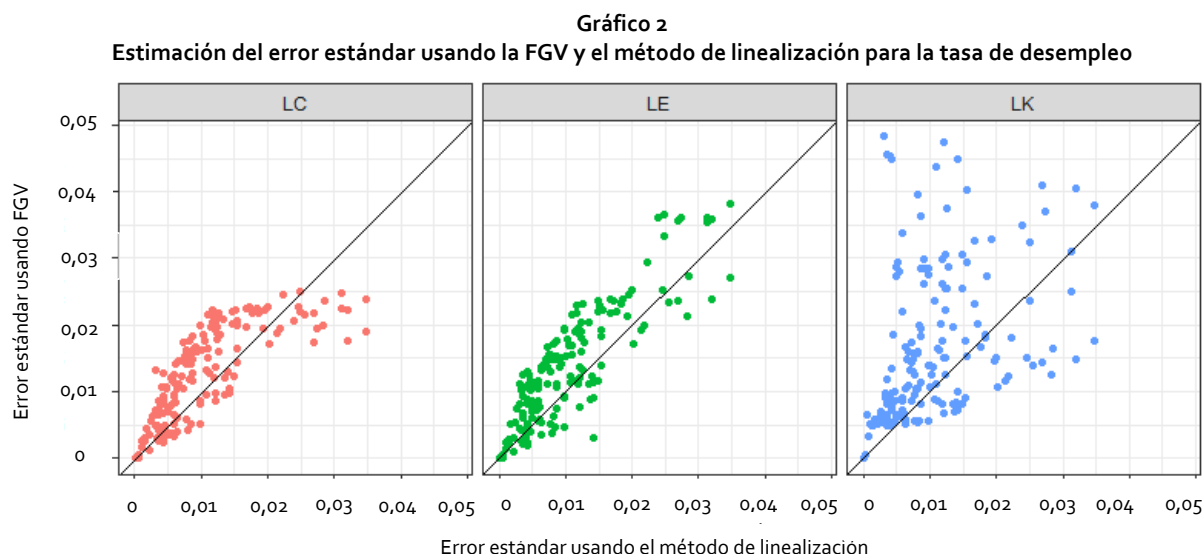
Tasa de ocupación

El ajuste de los modelos (6), (7) y (8) para el indicador de la tasa de ocupación, también permite concluir que el modelo LC presenta la mejor bondad de ajuste dado que tiene el menor AIC, BIC y *Deviance*. Los parámetros del modelo LC para la estimación del error estándar mediante la FGV se presentan en el cuadro 4, y al igual que en los casos, para obtener la estimación del error estándar se debe extraer la raíz cuadrada de la ecuación (8) al usar los parámetros estimados para este indicador.

Tasa de desempleo

Para la tasa de desempleo definida como el cociente entre la población en estado desocupado sobre el total de población en estado ocupado o desocupado, también se encontró que el modelo LC es el que tiene el menor AIC, BIC y *Deviance*. En el cuadro 4 se presenta la estimación de los parámetros del modelo LC para la FGV. Al igual que en todos los casos, para obtener la estimación del error estándar se debe extraer la raíz cuadrada de la ecuación (8) al usar los coeficientes presentados acá.

El gráfico 2 presenta la estimación del error estándar por el método de linealización de Taylor y por el método de la FGV, se observa que el modelo LC y LE son más estables. Siendo el modelo LC el que presenta el mejor desempeño en las estadísticas de bondad de ajuste. La manera de obtener el error estándar para estimaciones de la tasa de desempleo para cualquier dominio de desagregación sigue el mismo procedimiento enunciado previamente.



Fuente: Elaboración propia.

Porcentaje de personas sin acceso a servicios

El NBI se conforma una serie de indicadores que permiten identificar los aspectos en que los hogares presentan necesidades. Así que en ocasiones hay un interés particular por desagregar el indicador por sus componentes, por lo que se ha considerado la estimación de la FGV, no solo para el NBI global sino para cada uno de ellos.

El ajuste de los modelos (6), (7) y (8) se presenta en el Anexo 1, para los indicadores del porcentaje de personas sin acceso a servicios de saneamiento, agua potable, energía eléctrica, combustible; porcentaje de personas con materialidad precaria en pisos, techos y paredes; porcentaje de personas con tenencia de computador y acceso a internet en el hogar, así como el indicador de necesidades básicas insatisfechas que corresponde al porcentaje de personas que tienen al menos una necesidad insatisfecha de las enunciadas. De donde puede observar que el modelo LC cuenta con el menor AIC, BIC y *Deviance* en todos los indicadores. El cuadro 4 presenta la estimación de los parámetros del modelo LC dado en la ecuación (7) para la FGV de todos los indicadores señalados excepto para el ingreso, en donde el modelo LE de la ecuación (8) obtuvo el mejor rendimiento, y la estimación del error estándar para cualquier desagregación de estos indicadores se obtiene de la misma forma que se ilustró en los casos anteriores.

Cuadro 4
Parámetros de regresión estimados para las FGV

| Indicador | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}_A^2$ |
|-----------------|-----------------|-----------------|-----------------|-----------------|--------------------|
| Pobreza | -5,631 | -4,945 | -1,566e-04 | 2,079e-04 | 1,345 |
| Pobreza extrema | -4,278 | -5,545 | -6,567e-04 | 9,745e-04 | 1,224 |
| Ocupación | -5,786 | -4,586 | -9,845e-05 | 9,626e-05 | 1,176 |
| Desocupación | -3,764 | -5,417 | -7,759e-04 | 7,812e-04 | 0,979 |
| Ingreso | -7,323 | 1,1e-04 | 0,33113 | -6,806e-02 | 0,7564 |
| Agua potable | -3,905 | -4,773 | -1,096e-03 | 1,694e-03 | 1,182 |
| Saneamiento | -4,474 | -4,869 | -2,738e-04 | 3,710e-04 | 1,694 |
| Electricidad | -2,554 | -4,368 | -1,968e-03 | 2,346e-03 | 2,146 |
| Combustible | -4,604 | -3,973 | -4,463e-04 | 6,393e-04 | 1,423 |
| Internet | -5,679 | -4,524 | -1,537e-04 | 2,326e-04 | 1,539 |
| Computador | -6,504 | -4,333 | -6,963e-05 | 8,933e-05 | 1,094 |
| Pisos | -4,945 | -4,094 | -7,883e-05 | 7,916e-05 | 3,106 |
| Techos | -3,574 | -3,999 | -1,025e-04 | 1,010e-04 | 3,725 |
| Paredes | -4,090 | -4,060 | -9,262e-05 | 9,169e-05 | 3,301 |

Fuente: Elaboración propia.

B. Estimación del efecto diseño

Como se ha venido mencionando a lo largo del documento, para los años estudiados, las bases de datos de las encuestas de algunos países no brindan la información del diseño de la muestra en las bases de datos que son publicadas, lo que dificulta que los usuarios puedan calcular apropiadamente las medidas de variabilidad para los estimadores. Nótese que anteriormente se había definido que la estimación de la varianza FGV estaría dada por:

$$\widehat{V}(\hat{\theta}_d) = \exp(z'_d \hat{\beta}) \cdot \widehat{\Delta} \cdot \hat{\theta}_d^2 \quad (46)$$

A partir de la sección anterior, y habiendo escogido el modelo LC (que resultó tener el mejor desempeño en los indicadores estudiados en este documento) para la generación de la FGV, una forma de obtener una estimación del error estándar para cualquier indicador, en cualquier subgrupo poblacional, es utilizar los coeficientes de regresión apropiados del modelo y aplicarlos debidamente. Por ejemplo, para un país que no brinde la información del diseño en sus bases de datos, la estimación del error estándar para un parámetro estimado $\hat{\theta}_i$ (que puede ser un total, promedio o proporción) bajo el modelo de la FGV con el método LC estaría dada por la siguiente expresión:

$$\widehat{SE}_{LC}(\hat{\theta}_d) = \sqrt{\exp(\hat{\beta}_0 + \hat{\beta}_1 \cdot n_d + \hat{\beta}_2 \cdot \hat{\theta}_d + \hat{\beta}_3 \cdot n_d \cdot \hat{\theta}_d) \cdot \widehat{\Delta} \cdot \hat{\theta}_d^2} \quad (47)$$

En donde n_d es el tamaño de muestra asociado al dominio de interés. Una vez obtenido $\widehat{SE}_{LC}(\hat{\theta}_d)$, es posible reconstruir la estimación efecto diseño para $\hat{\theta}_d$ de la siguiente manera:

$$DEFF_{LC}(\hat{\theta}_d) = \frac{\widehat{SE}_{LC}^2(\hat{\theta}_d)}{\widehat{V}_{MAS}(\hat{\theta}_d)} \quad (48)$$

En donde $\widehat{V}_{MAS}(\hat{\theta}_d)$ es la varianza estimada de $\hat{\theta}_d$ asumiendo un diseño aleatorio simple sin reemplazo. Ahora, considerando lo mencionado en la ecuación (5), entonces:

$$\widehat{DEFF} = \frac{\widehat{SE}_{LC}^2(\hat{\theta}_d)}{V_w(\hat{\theta})} DEFF^w \quad (49)$$

En el anexo 2 se presentan los efectos de diseño para todos los indicadores estudiados en este documento desagregados por los dominios de interés de los 18 países estudiados.

V. Conclusiones

Este documento realiza una estimación de la función generalizada de varianza (FGV) para inferir el error de muestreo de diferentes variables y tipos de estimadores con base en información de las encuestas de hogares de la región. Se evaluaron cuatro modelos para estimar la FGV. El modelo log-lineal Común (LC) mostró el mejor desempeño, evaluado según las medidas de bondad de ajuste de AIC, BIC, *Deviance* y RMSE, para todas las variables analizadas y mostró ser consistente con el método de linealización de Taylor en todas las situaciones estudiadas.

Para cada indicador de interés se generó un modelo que permite estimar el error estándar a partir del modelo de la FGV por el método LC. Una de las principales ventajas del método basado en la FGV es su facilidad para ser utilizado, ya que solo se requiere disponer del valor estimado para el indicador de interés y el tamaño de muestra correspondiente a dicha estimación. De esta forma, el usuario cuenta con una excelente herramienta que le permitirá estimar la incertidumbre de las estimaciones de estos indicadores para desagregaciones que no son publicadas por las ONE o cuando la información de diseño de la encuesta no está disponible junto con los microdatos.

Se debe tener en cuenta que el método basado en la FGV puede no ser apropiado para indicadores distintos a los considerados en este documento, por ejemplo, informalidad en el empleo o seguridad alimentaria. No obstante, queda a disposición de los usuarios interesados un repositorio en Gitlab, con los códigos de R que fueron utilizados y que permiten, mediante un flujo de trabajo reproducible y replicable, ser aplicados a otras variables de interés o dominios no considerados en esta investigación.

Finalmente, el documento presenta los efectos de diseño para cada variable y nivel de desagregación para cada uno de los 18 países latinoamericanos cuyas bases de datos están disponibles en BADEHOG. Este es un insumo que puede ser utilizado como referencia para determinar adecuadamente el tamaño de la muestra de otras investigaciones sobre temas como los aquí analizados.

Bibliografía

- Akaike, Hirotugu (1998), "Information theory and an extension of the maximum likelihood principle". En Selected papers of hirotugu akaike, 199-213. Springer.
- Ball, Patrick (2016), "The Task Is a Quantum of Workflow". <https://hrdag.org/2016/06/14/the-task-is-a-quantum-of-workflow/>.
- Beaumont, J.-F., y C. Bocci (2016), "Small area estimation in the Labour Force Survey". Paper presented at the Advisory Committee on Statistical Methods, May 2016, Statistics Canada.
- Edelman, MW (1967), "Curve Fitting of Keyfitz Variances". Unpublished memorandum. Washington, DC: US Bureau of the Census.
- Fay, Robert E, y Roger A Herriot (1979), "Estimates of income for small places: an application of James-Stein procedures to census data". *Journal of the American Statistical Association* 74 (366a): 269-77.
- Freedman Ellis, Greg, y Ben Schneider (2022), *srvyr: 'dplyr'-Like Syntax for Summary Statistics of Survey Data*. <https://CRAN.R-project.org/package=srvyr>.
- Fúquene-Patiño, Jairo, César Cristancho, Mariana Ospina, y Domingo Morales Gonzalez (2021), "Fay-Herriot Model-Based Prediction Alternatives for Estimating Households with Emigrated Members.". *Journal of Official Statistics (JOS)* 37 (3).
- Glewwe, Paul y Todd, Petra (2022), *Impact Evaluation in International Development: Theory, Methods, and Practice*. ISBN - 978-1-4648-1497-6. The World Bank.
- Gutiérrez, Andrés (2009), *Estrategias de muestreo diseño de encuestas y estimacion de parametros*. Universidad Santo Tomas, Bogota (Colombia).
- Hansen, Morris H, William N Hurwitz, y William G Madow (1953), "Sample survey methods and theory. Vol. I. Methods and applications."
- Judkins, David R (1990), "Fay's method for variance estimation". *Journal of Official Statistics* 6 (3): 223-39.
- Kish, Leslie (1965), "Survey Sampling. New York: John Wiley & Sons". Inc. Kish Survey Sampling 1965.
- (2005), *Statistical design for research*. John Wiley & Sons.
- Krenzke, Thomas (1995), *Reevaluating generalized variance model parameters for the National Crime Victimization Survey*. US Census Bureau.
- Krewski, D, y Jon NK Rao (1981), "Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods". *The Annals of Statistics*, 1010-19.
- Lohr, Sharon L (2021), *Sampling: design and analysis*. Chapman; Hall/CRC.

- Lu, Hao, y Andrew Gelman (2003), "A method for estimating design-based sampling variances for surveys with weighting, poststratification, and raking". *Journal of Official Statistics* 19 (2): 133.
- Lumley, Thomas (2020), "survey: analysis of complex survey samples".
- McCarthy, Philip J (1969), "Pseudo-replication: Half samples". *Revue de l'Institut International de Statistique*, 239-64.
- MDSF-CEPAL (2022), "Estimaciones Comunales de Pobreza por ingresos en Chile Mediante Métodos de Estimación en Áreas Pequeñas - Informe Metodológico". Santiago de Chile, 2022.
- Morales, Domingo, María Dolores Esteban, Agustín Pérez, y Tomáš Hobza (2021), "A course on small area estimation and mixed models". *Methods, theory, and applications in R*.
- Park, Inho, Marianne Winglee, Jay Clark, Keith Rust, Andrea Sedlak, y David Morganstein (2003), "Design effects and survey planning". *Proceedings of the American Statistical Association Joint Statistical Meetings, Section on Survey Research Methods 1*: 3179-86.
- R Core Team (2022), *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rao, Jon NK, y CFJ Wu (1988), "Resampling inference with complex survey data". *Journal of the American Statistical Association* 83 (401): 231-41.
- Rivest, LOUIS-PAUL, y E Belmonte (2000), "A conditional mean squared error of small area estimators". *Survey Methodology* 26 (1): 67-78.
- Särndal, C.-E., Swensson, B., & Wretman, J (2003), *Model Assisted Survey Sampling*. Springer Science & Business Media.
- Schwarz, Gideon (1978), "Estimating the dimension of a model". *The Annals of Statistics*, 461-64.
- UN, DESA (2009), *Diseño de Muestras para Encuestas de Hogares: Directrices Prácticas. Estudios de métodos - Serie F No. 98*. United Nations Publications.
- Valliant, Richard (1987), "Generalized variance functions in stratified two-stage sampling". *Journal of the American Statistical Association* 82 (398): 499-508.
- Valliant, Richard, Jill A Dever, y Frauke Kreuter (2013), *Practical tools for designing and weighting survey samples*. Vol. 1. Springer.
- West, Brady T, y Sean Esteban McCabe (2012), "Incorporating complex sample design effects when only final survey weights are available". *The Stata Journal* 12 (4): 718-25.
- White, Howard y Raitzer, David (2017), *Impact Evaluation of Development Interventions. A Practical Guide*. Asian Development Bank. ISBN 978-92-9261-058-6.
- Wolter, Kirk M (2007), *Introduction to variance estimation*. Vol. 53. Springer.

Anexos

Anexo 1

Medidas de bondad de ajuste para los modelos

Cuadro A1
Medidas de calidad del ajuste para los modelos LK

| Indicador | <i>Deviance</i> | AIC | BIC | RMSE |
|-----------------|-----------------|-------|-------|----------|
| Pobreza | 252,5 | 579,7 | 592,5 | 0,067056 |
| Pobreza extrema | 249,1 | 577,3 | 590,0 | 0,099632 |
| Ocupación | 372,9 | 626,6 | 639,2 | 0,077850 |
| Desocupación | 335,9 | 606,2 | 618,7 | 0,163077 |
| Ingreso | 217,2 | 554,6 | 567,4 | 0,9998 |
| Agua potable | 214,6 | 509,0 | 521,3 | 0,127202 |
| Saneamiento | 394,2 | 609,1 | 621,4 | 0,170120 |
| Electricidad | 448,7 | 556,3 | 567,9 | 0,462962 |
| Combustible | 311,3 | 474,0 | 485,3 | 0,117386 |
| Internet | 286,7 | 510,8 | 522,6 | 0,068177 |
| Computador | 167,1 | 392,6 | 403,8 | 0,043558 |
| Pisos | 874,2 | 796,4 | 809,2 | 0,263304 |
| Techos | 1 325,5 | 866,7 | 879,4 | 0,862415 |
| Paredes | 1 088,3 | 835,4 | 848,2 | 0,501208 |

Fuente: Elaboración propia.

Cuadro A2
Medidas de calidad del ajuste para los modelos LE

| Indicador | <i>Deviance</i> | AIC | BIC | RMSE |
|-----------------|-----------------|-------|-------|----------|
| Pobreza | 273,1 | 595,9 | 611,9 | 0,063274 |
| Pobreza extrema | 263,6 | 589,5 | 605,5 | 0,095242 |
| Ocupación | 229,3 | 545,4 | 561,1 | 0,047730 |
| Desocupación | 180,7 | 502,8 | 518,5 | 0,072205 |
| Ingreso | 133,9 | 469,1 | 485,1 | 0,7898 |
| Agua potable | 227,4 | 520,3 | 535,7 | 0,129091 |
| Saneamiento | 284,6 | 558,6 | 574,0 | 0,108317 |
| Electricidad | 292,1 | 499,9 | 514,5 | 0,440684 |
| Combustible | 262,2 | 454,7 | 468,8 | 0,104053 |
| Internet | 257,0 | 497,2 | 512,0 | 0,062889 |
| Computador | 137,8 | 371,1 | 385,1 | 0,037133 |
| Pisos | 550,1 | 716,0 | 731,9 | 0,219530 |
| Techos | 733,4 | 763,9 | 779,8 | 0,609196 |
| Paredes | 643,8 | 744,0 | 759,9 | 0,367293 |

Fuente: Elaboración propia.

Cuadro A3
Medidas de calidad del ajuste para los modelos LC

| Indicador | <i>Deviance</i> | AIC | BIC | RMSE |
|-----------------|-----------------|-------|-------|----------|
| Pobreza | 236,8 | 570,2 | 586,1 | 0,061878 |
| Pobreza extrema | 215,5 | 553,2 | 569,2 | 0,085458 |
| Ocupación | 196,4 | 519,0 | 534,7 | 0,045979 |
| Desocupación | 162,5 | 484,7 | 500,4 | 0,071407 |
| Ingreso | 162,0 | 503,6 | 519,6 | 0,9998 |
| Agua potable | 184,4 | 486,7 | 502,1 | 0,121712 |
| Saneamiento | 266,0 | 547,7 | 563,2 | 0,098325 |
| Electricidad | 283,3 | 495,8 | 510,3 | 0,338280 |
| Combustible | 170,8 | 401,6 | 415,7 | 0,082353 |
| Internet | 212,4 | 470,1 | 484,9 | 0,057889 |
| Computador | 129,1 | 363,1 | 377,1 | 0,039939 |
| Pisos | 540,4 | 712,8 | 728,7 | 0,176431 |
| Techos | 644,5 | 741,0 | 756,9 | 0,359930 |
| Paredes | 574,3 | 723,6 | 739,6 | 0,299906 |

Fuente: Elaboración propia.

Anexo 2

Cuadros para los efectos de diseño en los países de América Latina

La estimación de los efectos de diseño se realiza para los 18 países y son calculados a partir de las bases de datos de las encuestas de hogares, disponibles en el repositorio BADEHOG de la CEPAL en el año 2022.

Las tablas permiten identificar el efecto de diseño para las variables de fuerza laboral, NBI, tipo de NBI, pobreza, pobreza extrema, tasa de ocupación y tasa de desempleo, desagregados por algunos dominios que, dependiendo de la disponibilidad de la variable en cada país, son calculados para el total nacional, por sexo, nivel educativo, área, edad y pertenencia étnica.

Cuando el dominio no está presente en la base de datos del país, entonces la celda estará vacía, mientras que en aquellos casos donde no se logró una buena estimación, se encontrará un asterisco (*) en el cuadro.

Cuadro A4
Efectos de diseño para Argentina

| Variable ^a | Total | Sexo: hombre | Sexo: mujer ^a | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área urbana | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|---------------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|----------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 2,6 | 1,6 | 2,0 | 5,2 | 5,3 | 13,3 | 10,4 | 10,6 | 9,0 | 7,8 | 2,6 | 10,2 | 4,8 | 14,4 | 9,1 | 7,8 | 7,0 |
| NBI Material del techo | 4,2 | 8,0 | 8,0 | 6,2 | 5,9 | 6,6 | 6,6 | 6,4 | 2,7 | 1,8 | | 5,7 | 3,6 | 7,1 | 6,7 | 6,3 | 5,9 |
| NBI Saneamiento | 3,9 | 8,8 | 8,8 | 9,0 | 8,1 | 9,4 | 9,7 | 9,4 | 4,8 | 2,0 | | 6,4 | 7,5 | 9,8 | 9,6 | 8,0 | 10,1 |
| Pobreza | 5,4 | 5,3 | 5,2 | 5,8 | 5,3 | 5,7 | 5,9 | 5,8 | 3,9 | 1,3 | | 2,3 | 5,6 | 6,0 | 6,0 | 5,2 | 6,1 |
| Pobreza extrema | 2,7 | 1,9 | 1,8 | 2,4 | 1,9 | 2,4 | 2,3 | 2,3 | 1,5 | * | | * | 2,1 | 2,4 | 2,5 | 2,0 | 2,4 |
| Tasa de desempleo | 1,7 | 2,8 | 3,0 | 4,2 | 3,7 | 3,6 | 1,8 | 3,7 | 3,8 | 2,6 | | 2,1 | | 2,7 | 3,9 | 4,0 | * |
| Tasa de ocupación | 8,4 | 6,1 | 6,7 | 6,0 | 4,7 | 7,0 | 2,5 | 7,0 | 7,4 | 7,1 | | 5,2 | | 7,7 | 6,8 | 7,3 | * |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A5
Efectos de diseño para Bolivia

| Variable ^a | Total | Sexo: hombre | Sexo: mujer ^a | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|----------------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|---------------|----------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 20,6 | 13,3 | 13,4 | 5,9 | 1,6 | 1,8 | 11,2 | 4,9 | 10,1 | * | 3,8 | 23,2 | 2,8 | 5,1 | 11,1 | 7,0 | 5,0 | 7,9 |
| NBI Agua | 1,5 | 8,9 | 8,8 | 7,7 | 4,9 | 6,4 | 11,4 | 7,0 | 6,0 | * | 11,8 | 5,9 | 10,1 | 7,2 | 7,1 | 7,8 | 8,8 | 7,9 |
| NBI Computador Hogar | 8,9 | 5,3 | 5,4 | 4,9 | 2,2 | 2,6 | 5,4 | 4,9 | 3,8 | * | 5,8 | 5,4 | 5,1 | 1,4 | 5,1 | 5,1 | 5,0 | 5,3 |
| NBI Electricidad | 23,6 | 20,7 | 20,6 | 17,6 | 11,0 | 16,1 | 24,2 | 15,8 | 7,0 | | * | 5,1 | 23,5 | 18,6 | 21,4 | 21,4 | 19,5 | 21,6 |
| NBI Internet Hogar | 13,6 | 8,0 | 8,0 | 6,9 | 3,2 | 3,7 | 8,3 | 6,5 | 5,2 | * | 8,6 | 7,3 | 7,7 | 5,9 | 7,2 | 7,0 | 7,0 | 7,5 |
| NBI Saneamiento | 12,4 | 14,2 | 14 | 14,9 | 8,1 | 9,5 | 14,8 | 13,2 | 9,9 | * | 16,1 | 11,5 | 16,4 | 13,5 | 15,1 | 15,1 | 14,9 | 15,7 |
| Pobreza | 8,2 | 5,4 | 5,3 | 5,9 | 2,9 | 3,5 | 5,4 | 5,4 | 4,8 | * | 5,9 | 5,1 | 5,2 | 5,7 | 6,1 | 6,1 | 5,5 | 6,1 |

| Variable ^a | Total | Sexo: hombre | Sexo: mujer ^a | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|-----------------------|-------|--------------|--------------------------|---|---|------------------------------------|--------------------------------------|--------------------------------------|---------------------------------------|---|------------|-------------|-----------|-----------|-------------|-------------|-------------|------------|
| Pobreza extrema | 5,3 | 5,2 | 5,2 | 6,6 | 4,0 | 5,1 | 5,4 | 5,9 | 4,7 | * | 6,2 | 4,9 | 5,3 | 6,4 | 6,5 | 6,7 | 6,8 | 6,1 |
| Tasa de desempleo | 11,7 | 3,5 | 2,9 | 4,5 | 1,2 | 1,0 | 3,3 | 4,4 | 3,3 | * | 2,6 | 6,7 | 2,0 | | 3,0 | 4,4 | 4,2 | 1,9 |
| Tasa de ocupación | 8,2 | 6,0 | 6,6 | 6,9 | 2,7 | 3,1 | 7,5 | 7,0 | 7,1 | 1,7 | 7,5 | 5,6 | 6,6 | | 7,1 | 6,9 | 7,3 | 4,4 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A6
Efectos de diseño para Brasil

| Variable ^a | Total | Hombre | Mujer ^a | Nivel Educativo: Alta secundaria incompleta | Nivel Educativo: Baja secundaria incompleta | Nivel Educativo: Primaria completa | Nivel Educativo: Primaria incompleta | Nivel Educativo: Secundaria completa | Nivel Educativo: Terciaria incompleta | Nivel Educativo: Universitaria completa | Área rural | Área urbana | Etnia: Afro | Etnia: Indígena | Etnia: No indígena ni afro | Edad >60 | Edad 0-5 | Edad 16-30 | Edad 31-45 | Edad 46-60 | Edad 6-15 |
|-----------------------------|-------|--------|--------------------|---|---|------------------------------------|--------------------------------------|--------------------------------------|---------------------------------------|---|------------|-------------|-------------|-----------------|----------------------------|----------|----------|------------|------------|------------|-----------|
| Ingresos | 7,0 | 12,3 | 11,1 | 11,6 | 21,3 | 7,6 | 12,2 | 6,7 | 29,1 | * | 19,8 | 7,1 | 8,9 | * | 14,1 | 22,4 | 5,5 | 20,6 | 16,9 | 19,9 | 9,8 |
| NBI Material de las Paredes | 6,9 | 1,1 | * | * | 20,3 | * | 9,7 | * | 7,7 | * | 2,4 | * | * | * | 2,1 | 29,4 | * | 16,4 | 13,3 | 17,3 | * |
| NBI Material de Pisos | 11,5 | 1,6 | 1,4 | 35,6 | 16,7 | * | 9,1 | * | 7,7 | * | 2,7 | * | 1,1 | 32,3 | 2,7 | 22,4 | * | 14,3 | 12,0 | 14,6 | 36,3 |
| NBI Material del techo | 12,1 | * | * | * | 24,8 | * | 10,8 | * | 8,2 | * | 2,3 | * | * | * | 2,0 | 37,8 | * | 19,2 | 15,2 | 20,6 | * |
| Pobreza | * | 6,2 | 7,2 | 10,3 | 8,3 | 11,4 | 6,2 | 10,7 | 10,5 | 7,4 | 3,5 | * | 26,2 | 2,9 | 8,8 | 11,7 | 13,4 | 8,2 | 8,7 | 10,7 | 7,7 |
| Pobreza extrema | * | 5,2 | 6,5 | 7,6 | 5,5 | 10,4 | 4,0 | 10,6 | 8,5 | 10,1 | 1,5 | 38,2 | 49,6 | 3,3 | 6,4 | 11,5 | 10,8 | 6,0 | 6,3 | 7,2 | 5,3 |
| Tasa de desempleo | 3,7 | 1,0 | 1,0 | 6,2 | 9,6 | 8,5 | 11,3 | 15,3 | 1,2 | 12,0 | 4,4 | 18,0 | * | 2,1 | 2,4 | 9,4 | | 1,1 | 2,4 | 7,0 | 6,0 |
| Tasa de ocupación | 4,1 | 2,4 | 2,7 | 12,3 | 10,2 | 9,8 | 12,7 | 14,1 | 3,5 | 9,2 | 4,1 | 23,4 | 1,9 | 2,0 | 3,0 | 14,0 | | 6,9 | 3,3 | 4,3 | 2,0 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A7
Efectos de diseño para Chile

| Variable ^a | Total | Sexo: hombre | Sexo: mujer ^a | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Etnia: Indígena | Etnia: No indígena ni afro | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: : 46- 60 | Edad: : 6- 15 |
|-----------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|---------------|----------------|--------------------|----------------------------------|--------------|--------------|----------------|----------------|----------------------|---------------------|
| Ingresos | 6,7 | 6,4 | 23,4 | 18,2 | 6,5 | 10,3 | 19,4 | 27,1 | 4,0 | 15,5 | 1,4 | 13,7 | 15,9 | 7,7 | 9,7 | 8,0 | 21,9 | 4,0 | 11,8 | 11,6 |
| NBI Agua | 8,6 | 9,7 | 10,4 | 7,5 | 6,9 | 8,8 | 7,1 | 8,8 | 10,2 | 7,7 | 19,1 | 9,1 | 6,8 | * | 7,2 | 9,4 | 8,8 | 9,3 | 7,8 | 10,0 |
| NBI Saneamiento | 7,1 | 2,5 | 2,4 | 2,7 | 1,0 | 1,1 | 2,7 | 2,6 | 1,6 | * | 1,5 | 4,7 | 2,7 | 2,4 | 2,6 | 1,6 | 2,8 | 2,4 | 2,7 | 2,3 |
| Pobreza | 3,5 | 7,9 | 8,2 | 9,2 | 5,2 | 5,5 | 8,6 | 9,3 | 9,4 | 4,2 | 7,6 | 48,6 | 8,5 | 30,9 | 8,8 | 9,1 | 9,2 | 9,8 | 9,6 | 9,6 |
| Pobreza extrema | 8,3 | 6,5 | 7,2 | 7,7 | 4,8 | 5,4 | 7,3 | 7,9 | 9,1 | 5,9 | 7,4 | * | 7,4 | * | 8,7 | 8,6 | 7,9 | 8,5 | 7,9 | 8,6 |
| Tasa de desempleo | 4,7 | 3,0 | 3,4 | 8,0 | 5,3 | 5,8 | 9,6 | 4,7 | 8,0 | 13,0 | 8,8 | 14,5 | 10,8 | 12,3 | 11,8 | | 4,4 | 6,5 | 6,8 | * |
| Tasa de ocupación | 2,2 | 6,4 | 7,2 | 12,6 | 4,5 | 5,7 | 10,3 | 9,3 | 11,9 | 14,8 | 11,3 | 27,8 | 12,4 | 30,9 | 13,9 | | 12,5 | 9,0 | 9,1 | * |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A8
Efectos de diseño para Colombia

| Variable ^a | Total | Sexo: hombre | Sexo: mujer ^a | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Etnia: Afro | Etnia: Indígena | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|----------------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|---------------|----------------|----------------|--------------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 12,1 | * | * | 14,1 | 13,2 | 15,2 | 6,6 | 12,6 | 22,7 | * | 12,6 | * | 10,4 | 4,4 | 14,3 | 12,2 | 7,3 | 8,2 | 11,3 | 10,1 |
| NBI Agua | 4,8 | 3,8 | * | 8,2 | 2,5 | 3,6 | 3,2 | 1,3 | 3,4 | 3,0 | * | * | * | 6,8 | 2,7 | 4,0 | * | * | 1,7 | * |
| NBI Combustible | * | 13,6 | 7,3 | 16,1 | 12,3 | 9,8 | 12,0 | 13,7 | 9,2 | 3,0 | * | 7,2 | 14,9 | 10,2 | 8,3 | 16,0 | 8,0 | 9,4 | 8,2 | 9,0 |
| NBI Computador Hogar | 24,5 | 2,6 | 1,6 | 11,9 | 6,7 | 7,4 | 2,5 | 2,7 | 7,7 | 8,4 | 12,0 | * | 10,1 | 16,5 | 4,5 | 11,3 | 2,7 | 4,0 | 4,5 | 5,6 |
| NBI Electricidad | * | 23,7 | 17,6 | 27,1 | 34,7 | 33,6 | 27,5 | 31,3 | 7,0 | 3,5 | * | 14,7 | 34,1 | 15,0 | 28,4 | 35,9 | 28,6 | 30,4 | 29,6 | 27,5 |
| NBI Internet Hogar | 3,1 | 3,9 | 1,6 | 11,8 | 3,3 | 3,7 | 1,4 | 1,1 | 7,5 | 11,7 | 19,8 | * | 5,9 | 21,5 | 2,0 | 7,8 | * | 1,6 | 1,9 | 2,4 |
| NBI Material de Pisos | * | 3,7 | 2,3 | * | * | * | 33,6 | 46,6 | * | * | * | * | * | * | 36,7 | * | 13,2 | 23,9 | 29,9 | 45,6 |
| NBI Material del techo | * | * | * | * | 12,1 | 18,9 | * | * | 5,6 | 15,2 | * | * | * | * | * | 36,3 | * | * | * | * |
| NBI Saneamiento | * | 20,8 | 14,8 | 26,8 | 26,4 | 29,5 | 19,5 | 22,7 | 18,1 | 9,6 | 37,9 | 5,9 | 10,6 | 32,4 | 23,2 | 30,5 | 18,7 | 22,7 | 24,6 | 23,7 |
| Pobreza | * | * | * | 7,8 | 1,5 | 2,9 | * | * | 4,4 | 10,8 | 7,5 | * | 9,3 | 20,1 | 1,7 | 2,3 | * | * | * | * |
| Pobreza extrema | * | * | * | 1,3 | * | * | * | * | * | 5,2 | * | * | 2,6 | 7,2 | * | * | * | * | * | * |
| Tasa de desempleo | 5,1 | * | * | 12,0 | 4,6 | 6,1 | 9,6 | * | * | * | 26,2 | * | 9,9 | 16,8 | 9,2 | | * | * | * | 4,4 |
| Tasa de ocupación | 7,2 | * | * | 20,5 | 13,3 | 11,7 | 12,3 | * | 3,6 | 2,8 | 17,9 | * | 17,4 | 21,2 | 8,5 | | * | * | * | 11,3 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro Ag
Efectos de diseño para Costa Rica

| Variable ^a | Total | Sexo: hombre | Sexo: mujer ^a | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|---------------------------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|---------------|----------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 7,4 | 9,9 | 10,2 | 2,4 | 2,8 | 7,2 | 6,5 | 4,3 | 5,4 | 7,8 | 6,3 | 13,5 | 4,9 | 1,2 | 5,8 | 4,6 | 4,5 | 3,3 |
| NBI Agua | 9,3 | 2,6 | 2,7 | 1,8 | 2,0 | 2,8 | 2,7 | 1,8 | 1,5 | * | 3,1 | 2,9 | 2,5 | 1,8 | 2,8 | 2,6 | 2,6 | 2,6 |
| NBI Combustible: Tiene carencia | 3,3 | 3,7 | 3,7 | 1,5 | 1,9 | 3,4 | 3,6 | 1,7 | 1,3 | * | 4,7 | 3,4 | 3,2 | 1,4 | 3,2 | 2,6 | 3,0 | 2,6 |
| NBI Computador Hogar | 5,4 | 3,3 | 3,3 | 2,2 | 2,4 | 3,3 | 3,3 | 2,4 | 1,3 | * | 3,2 | 3,5 | 3,1 | 1,9 | 3,1 | 3,1 | 3,0 | 2,8 |
| NBI Internet Hogar | 9,7 | 2,4 | 2,3 | 1,3 | 1,4 | 2,4 | 2,4 | 1,4 | * | * | 2,4 | 2,3 | 2,4 | 1,0 | 1,9 | 2,0 | 2,1 | 1,7 |
| NBI Material de las Paredes | 1,1 | 1,4 | 1,4 | * | * | 1,3 | 1,4 | * | * | * | 1,2 | 1,6 | * | * | 1,3 | 1,1 | 1,1 | * |
| NBI Material del techo | 1,1 | 1,8 | 1,8 | * | 1,1 | 1,3 | 1,8 | 1,0 | * | 1,2 | 1,4 | 2,4 | 1,6 | * | 1,7 | 1,5 | 1,2 | * |
| Pobreza | 9,2 | 2,7 | 2,6 | 2,3 | 2,5 | 3,0 | 2,8 | 2,1 | 1,2 | * | 2,9 | 2,5 | 2,1 | 2,4 | 3,0 | 3,0 | 2,6 | 3,0 |
| Pobreza extrema | 3,6 | 1,6 | 1,5 | 1,6 | 1,8 | 2,0 | 1,6 | 1,4 | * | * | 1,8 | 1,3 | 1,1 | 1,8 | 2,0 | 2,0 | 1,8 | 2,0 |
| Tasa de desempleo | 4,1 | 3,1 | 3,5 | 4,4 | 4,3 | 4,7 | 4,0 | 4,6 | 4,7 | 2,6 | 4,3 | 3,0 | 2,3 | | 3,6 | 4,2 | 4,7 | * |
| Tasa de ocupación | 4,2 | 3,7 | 4,3 | 3,0 | 3,0 | 4,3 | 2,9 | 4,0 | 4,1 | 3,9 | 4,1 | 4,0 | 2,8 | | 4,4 | 4,2 | 4,4 | * |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A10
Efectos de diseño para República Dominicana

| Variable ^a | Total | Sexo: hombre | Sexo: mujer ^a | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|--------------------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|---------------|----------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 30,6 | 25,0 | 24,6 | 21,8 | 5,2 | 5,1 | 29,2 | 17,5 | 12,4 | 2,0 | 27,6 | 27,3 | 9,5 | 9,8 | 18,5 | 14,2 | 14,3 | 13,9 |
| NBI Agua | 8,1 | 10,0 | 9,4 | 8,8 | 7,1 | 6,8 | 7,7 | 10,0 | 9,0 | * | 5,5 | * | 8,8 | 8,6 | 8,4 | 8,8 | 9,0 | 8,4 |
| NBI Combustible | 9,9 | 7,5 | 6,8 | 5,2 | 2,1 | 2,7 | 8,9 | 2,8 | * | | 8,4 | 6,9 | 5,1 | 4,4 | 6,4 | 5,8 | 4,5 | 5,2 |
| NBI Computador Hogar | 5,9 | 8,4 | 8,7 | 8,8 | 5,3 | 5,1 | 7,9 | 9,2 | 8,4 | * | 6,4 | 22,7 | 8,8 | 8,0 | 8,6 | 9,0 | 8,9 | 8,8 |
| NBI Electricidad | 13,8 | 18,9 | 18,3 | 16,4 | 7,3 | 8,0 | 24,8 | 5,7 | 1,4 | | 22,4 | 21,4 | 15,9 | 17,3 | 19,4 | 20,3 | 15,5 | 17,2 |
| NBI Internet Hogar | 9,2 | 14,6 | 15,8 | 12,5 | 10,9 | 10,8 | 12,2 | 14,4 | 13,7 | * | 7,4 | * | 14,4 | 13,6 | 11,9 | 13,2 | 13,9 | 12,9 |
| NBI Material de las Paredes | 13,6 | 33,2 | 33,5 | 30,0 | 11,4 | 12,4 | 33,1 | 17,5 | 5,3 | * | 30,7 | 36,6 | 14,1 | 22,6 | 33,0 | 31,3 | 20,3 | 21,9 |
| NBI Material de Pisos | 12,8 | 2,6 | 2,7 | 1,9 | * | 1,0 | 2,7 | 1,1 | * | | 2,6 | 2,7 | 1,1 | 1,9 | 2,6 | 2,4 | 1,5 | 2,1 |
| NBI Material del techo | 14,4 | 3,0 | 3,0 | 2,7 | * | 1,4 | 3,0 | 2,3 | * | | 2,4 | 3,9 | 1,7 | 2,0 | 2,8 | 2,4 | 1,9 | 2,5 |
| NBI Saneamiento | 6,6 | 13,8 | 13,9 | 15,4 | 8,5 | 8,6 | 13,9 | 12,8 | 5,7 | * | 13,2 | 23,0 | 12,3 | 14,2 | 15,2 | 15,4 | 13,9 | 15,0 |
| Pobreza | 7,6 | 6,4 | 6,4 | 7,5 | 4,3 | 3,9 | 6,3 | 7,2 | 4,6 | * | 6,1 | 11,1 | 4,0 | 7,3 | 7,4 | 7,4 | 5,4 | 7,2 |
| Pobreza extrema | 6,2 | 4,8 | 4,8 | 6,1 | 3,6 | 3,4 | 4,7 | 5,9 | 3,7 | * | 4,8 | 6,8 | 3,2 | 6,1 | 6,0 | 6,1 | 4,4 | 5,5 |
| Tasa de desempleo | 4,6 | 3,6 | 2,7 | 4,3 | 1,4 | * | 3,5 | 4,0 | 4,3 | * | 3,9 | 3,3 | * | | 2,7 | 4,2 | 3,6 | * |
| Tasa de ocupación | 7,3 | 7,6 | 8,9 | 10,1 | 5,5 | 5,1 | 9,6 | 10,4 | 11,1 | 3,3 | 8,2 | 11,7 | 9,2 | | 9,7 | 9,2 | 10,2 | * |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A11
Efectos de diseño para Ecuador

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Etnia: Afro | Etnia: Indígena | Etnia: No indígena ni afro | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|-----------------------------|-------|--------------|-------------|---|---|------------------------------------|--------------------------------------|--------------------------------------|---------------------------------------|---|------------|-------------|-------------|-----------------|----------------------------|-----------|-----------|-------------|-------------|-------------|------------|
| Ingresos | 8,6 | 6,2 | 6,9 | 4,4 | 2,4 | 3,3 | 2,8 | 4,5 | 2,5 | 3,3 | 4,6 | 8,5 | * | 2,0 | 11,4 | 2,4 | * | 7,2 | 2,8 | 3,3 | 4,7 |
| NBI Agua | 4,9 | 4,5 | 4,6 | 4,2 | 3,9 | 4,4 | 4,6 | 3,6 | * | * | 5,3 | 5,2 | 3,3 | 4,7 | 4,1 | 2,7 | 4,6 | 4,5 | 4,2 | 3,2 | 4,9 |
| NBI Combustible | 2,6 | 4,0 | 4,0 | 2,8 | 2,2 | 3,3 | 4,0 | 2,1 | * | * | 10,7 | 2,1 | * | 4,3 | 3,3 | 2,5 | 2,8 | 3,4 | 2,7 | 2,6 | 3,8 |
| NBI Computador Hogar | 4,8 | 4,6 | 4,6 | 4,1 | 3,1 | 4,4 | 4,7 | 4,3 | 1,5 | 1,0 | 4,3 | 4,9 | 2,0 | 4,1 | 6,7 | 3,9 | * | 4,5 | 4,4 | 4,1 | 4,7 |
| NBI Electricidad | 6,7 | 8,9 | 9,1 | 7,4 | 7,2 | 8,6 | 9,0 | 5,3 | * | * | * | 1,6 | * | 10,9 | 6,3 | 4,0 | 7,8 | 8,9 | 7,1 | 6,5 | 9,6 |
| NBI Material de las Paredes | 8,5 | 8,6 | 8,6 | 5,3 | 6,1 | 7,4 | 8,6 | 4,3 | * | * | 8,6 | 8,6 | 2,7 | 2,5 | 15,1 | 5,6 | 4,3 | 7,1 | 7,1 | 6,3 | 8,0 |
| NBI Material de Pisos | 12,1 | 19,7 | 19,8 | 15,2 | 11,9 | 17,4 | 20,1 | 12,8 | 3,2 | 1,6 | 22,4 | 18,6 | 7,6 | 19,3 | 18,0 | 13,4 | 13,8 | 18,1 | 15,3 | 13,7 | 19,3 |
| NBI Material del techo | 12,6 | 12,2 | 12,2 | 8,2 | 5,9 | 8,1 | 12,2 | 9,3 | 2,4 | 1,6 | 12,2 | 12,1 | 6,0 | 11,1 | 10,9 | 3,6 | 8,8 | 9,8 | 10,3 | 5,1 | 11,9 |
| NBI Saneamiento | 2,4 | 2,0 | 2,1 | 2,1 | 1,8 | 2,1 | 2,2 | 1,5 | * | * | 2,0 | 2,2 | * | 1,7 | 2,0 | 1,1 | 2,1 | 2,3 | 1,9 | 1,7 | 2,3 |
| Pobreza | 2,2 | 3,8 | 3,7 | 4,0 | 3,3 | 4,1 | 4,1 | 3,7 | 1,3 | * | 3,8 | 3,7 | 2,4 | 4,2 | 3,4 | 2,4 | 3,8 | 4,2 | 4,1 | 3,3 | 4,3 |
| Pobreza extrema | 2,9 | 3,3 | 3,3 | 4,2 | 3,6 | 4,1 | 3,7 | 3,7 | 1,2 | * | 3,1 | 3,7 | 2,8 | 3,5 | 2,9 | 2,4 | 4,4 | 4,1 | 4,1 | 3,2 | 4,1 |
| Tasa de desempleo | 1,9 | 1,8 | 1,5 | 1,5 | * | 1,4 | * | 1,8 | 2,1 | 2,1 | 2,0 | 1,5 | 1,9 | * | 1,5 | * | | 1,5 | 2,1 | 1,8 | * |
| Tasa de ocupación | 4,6 | 5,5 | 6,0 | 5,7 | 3,6 | 6,6 | 5,7 | 6,6 | 4,9 | 4,6 | 6,0 | 5,6 | 2,7 | 6,3 | 6,6 | 5,3 | * | 6,6 | 6,5 | 6,5 | 2,3 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A12
Efectos de diseño para Guatemala

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Etnia: Afro | Etnia: Indígena | Etnia: No indígena ni afro | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|-----------------------------------|-------|-----------------|----------------|---|---|---|---|---|--|--|---------------|----------------|----------------|--------------------|----------------------------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 4,3 | 3,3 | 3,1 | * | 2,0 | 2,3 | 2,9 | * | 2,8 | * | 3,4 | 3,8 | * | 6,6 | 4,2 | 1,3 | * | 2,7 | 1,4 | * | 1,9 |
| NBI Agua | 7,8 | 9,6 | 9,9 | 6,4 | 6,5 | 8,8 | * | 2,4 | 2,4 | * | * | 9,5 | * | 16,6 | 5,6 | 7,5 | 8,3 | 7,9 | 9,1 | 9,0 | 8,0 |
| NBI Combustible | 2,4 | 13,1 | 13,8 | 13,9 | 13,6 | 13,0 | * | 9,1 | 11,0 | 1,1 | 44,8 | 11,2 | * | 14,7 | 19,4 | 14,0 | 12,9 | 9,2 | 12,6 | 14,6 | 9,7 |
| NBI Electricidad | 21,8 | 3,0 | 2,7 | 40,9 | 39,5 | 20,2 | * | 22,8 | 27,6 | 4,5 | 47,3 | 7,9 | | 7,9 | * | 37,9 | 9,5 | 4,5 | 15,1 | 33,2 | 4,5 |
| NBI Internet Hogar | 11,2 | 14,4 | 16,0 | 13,0 | 11,5 | 13,5 | 35,6 | 9,2 | 11,5 | 2,6 | 16,0 | 15,9 | * | 13,3 | 22,2 | 12,4 | 13,7 | 11,7 | 13,5 | 13,7 | 12,0 |
| NBI Material de las Paredes | 10,2 | 14,4 | 14,4 | 3,2 | 4,5 | 6,7 | 18,1 | 1,3 | 1,8 | | 14,8 | 14,2 | | 14,4 | 14,4 | 4,9 | 12,2 | 13,9 | 10,1 | 6,7 | 13,5 |
| NBI Material del techo | 18,2 | 29,1 | 29,1 | 7,5 | 7,1 | 18,1 | 34,3 | 1,8 | 3,3 | | * | 10,1 | | 30,4 | 27,9 | 9,3 | 24,3 | 27,2 | 18,4 | 14,8 | 27,3 |
| NBI Saneamiento | 9,4 | 17,8 | 18,6 | 16,8 | 13,9 | 19,9 | 26,0 | 9,0 | 11,5 | 1,7 | 14,9 | 26,1 | * | 19,1 | 19,6 | 16,6 | 20,9 | 19,5 | 21,3 | 18,5 | 19,9 |
| Pobreza | 7,3 | 8,0 | 8,1 | 6,3 | 6,2 | 9,5 | 16,0 | 2,4 | 2,6 | * | 10,2 | 9,2 | | 9,0 | 7,1 | 7,3 | 9,9 | 9,3 | 9,8 | 8,7 | 8,8 |
| Pobreza extrema | 1,5 | 6,6 | 6,7 | 4,8 | 5,3 | 8,2 | 29,8 | 1,3 | 1,6 | * | 18,6 | 8,2 | | 9,6 | 5,3 | 7,2 | 8,3 | 8,0 | 8,7 | 7,8 | 6,8 |
| Tasa de desempleo | 2,1 | 1,1 | 1,3 | 1,5 | 1,1 | 1,4 | 1,5 | 1,6 | 1,6 | * | 1,5 | * | | 1,5 | * | * | * | * | 1,4 | * | * |
| Tasa de ocupación | 6,3 | 6,9 | 8,2 | 7,0 | 5,1 | 8,4 | 7,2 | 5,6 | 7,7 | 3,3 | 6,6 | 7,6 | * | 7,9 | 6,8 | 6,1 | | 7,7 | 8,5 | 8,4 | 6,0 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A13
Efectos de diseño para Honduras

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|-----------------------------|-------|-----------------|----------------|---|---|---|---|---|--|--|---------------|----------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 9,1 | 5,0 | 5,9 | 1,3 | 1,5 | 4,1 | 5,7 | * | 2,8 | * | 5,8 | 6,9 | 1,1 | 1,1 | 3,4 | 3,0 | 1,7 | 3,3 |
| NBI Agua | 5,7 | 6,1 | 6,1 | 2,9 | 3,1 | 5,9 | 7,7 | 1,6 | 2,1 | * | 10,2 | 6,0 | 4,4 | 5,6 | 6,0 | 5,8 | 5,0 | 6,0 |
| NBI Combustible | 12,8 | 15,8 | 16,2 | 9,8 | 9,3 | 12,9 | 19,7 | 5,1 | 9,5 | 2,1 | 44,8 | 15,4 | 12,6 | 12,5 | 12,9 | 13,0 | 12,7 | 12,8 |
| NBI Computador Hogar | 4,2 | 4,6 | 4,6 | 2,7 | 2,3 | 4,5 | 4,6 | 1,4 | 3,1 | * | 4,6 | 4,7 | 3,5 | 3,6 | 4,6 | 4,2 | 3,7 | 4,5 |
| NBI Electricidad | 11,2 | 23,1 | 22,8 | 8,8 | 11,1 | 25,0 | 31,6 | 5,0 | 4,2 | * | * | 18,9 | 19,0 | 24,7 | 24,4 | 24,5 | 22,3 | 25,0 |
| NBI Internet Hogar | 5,4 | 9,7 | 9,8 | 6,0 | 5,4 | 9,1 | 10,0 | 3,2 | 6,5 | 1,6 | 10,5 | 9,6 | 7,6 | 8,0 | 9,2 | 8,7 | 7,9 | 9,1 |
| NBI Material de las Paredes | 6,7 | 10,4 | 10,5 | 3,3 | 3,3 | 8,4 | 10,7 | * | 2,2 | * | 12,1 | 9,2 | 5,1 | 8,0 | 9,9 | 7,9 | 5,7 | 9,6 |
| NBI Material de Pisos | 4,2 | 12,6 | 12,6 | 3,2 | 3,9 | 11,1 | 12,8 | 1,1 | 1,8 | * | 21,3 | 8,0 | 5,8 | 9,4 | 12,2 | 9,5 | 7,5 | 11,9 |
| NBI Material del techo | 3,9 | 5,0 | 5,0 | 1,9 | 2,7 | 4,1 | 4,9 | * | 2,1 | * | 5,0 | 4,9 | * | 4,1 | 4,9 | 3,5 | * | 3,9 |
| NBI Saneamiento | 3,5 | 8,6 | 8,6 | 5,4 | 5,6 | 9,4 | 8,4 | 2,9 | 4,9 | * | 8,9 | 8,5 | 6,8 | 8,5 | 9,6 | 8,8 | 7,9 | 9,6 |
| Pobreza | 4,6 | 5,4 | 5,4 | 3,9 | 3,7 | 6,2 | 5,3 | 1,9 | 3,4 | * | 5,4 | 5,9 | 4,9 | 5,7 | 6,2 | 5,9 | 5,3 | 6,2 |
| Pobreza extrema | 5,2 | 5,2 | 5,2 | 3,8 | 4,4 | 6,5 | 5,9 | 1,7 | 3,2 | * | 9,2 | 7,0 | 6,1 | 6,6 | 6,4 | 6,7 | 6,1 | 6,2 |
| Tasa de desempleo | 1,9 | 1,5 | 1,3 | 1,5 | * | 1,9 | 1,6 | 1,6 | 1,7 | 1,2 | 1,9 | 1,1 | * | | 1,1 | 1,9 | 1,1 | * |
| Tasa de ocupación | 3,1 | 4,3 | 5,0 | 3,5 | 2,5 | 5,1 | 5,1 | 2,2 | 4,8 | 2,9 | 4,6 | 4,5 | 3,8 | * | 5,0 | 5,1 | 5,0 | 2,6 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A14
Efectos de diseño para México

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área: rural | Área: urbana | Etnia: Indígena | Etnia: No indígena ni afro | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|-----------------------------|-------|--------------|-------------|---|---|------------------------------------|--------------------------------------|--------------------------------------|---------------------------------------|---|-------------|--------------|-----------------|----------------------------|-----------|-----------|-------------|-------------|-------------|------------|
| Ingresos | 1,4 | 5,3 | 8,6 | 13,6 | 4,1 | 11,7 | 16,1 | 11,7 | 16,6 | 5,0 | 5,2 | 5,6 | 16,4 | 4,2 | 8,7 | 15,6 | 11,9 | 15,5 | 7,9 | 13,8 |
| NBI Agua | 8,6 | 8,8 | 8,6 | 6,1 | 11,1 | 7,6 | 5,6 | 11,3 | 9,1 | 3,6 | 5,8 | 13,5 | 12,7 | 7,0 | 8,4 | 9,5 | 6,0 | 6,8 | 8,1 | 6,2 |
| NBI Combustible | 4,1 | 3,5 | 4,7 | 2,2 | 13,8 | 5,4 | 3,5 | 13,4 | 14,2 | 7,2 | * | 4,8 | 26,2 | 2,2 | 6,3 | 10,0 | 2,6 | 3,8 | 5,0 | 3,3 |
| NBI Computador Hogar | 3,9 | 3,7 | 4,4 | 4,1 | 12,1 | 8,3 | 3,0 | 11,7 | 13,0 | 8,3 | * | 27,4 | 3,5 | 13,0 | 8,7 | 10,7 | 5,2 | 6,4 | 7,7 | 6,4 |
| NBI Electricidad | 4,4 | 24,0 | 23,4 | 26,9 | 20,5 | 27,5 | 23,5 | 18,9 | 6,1 | 1,9 | 18,8 | 34,5 | 22,2 | 24,9 | 24,9 | 25,7 | 26,9 | 26,8 | 29,0 | 26,8 |
| NBI Internet Hogar | 3,7 | 5,8 | 7,6 | 2,8 | 15,2 | 6,5 | 2,7 | 13,7 | 17,9 | 12,5 | * | 47,0 | 3,3 | 34,5 | 6,5 | 10,9 | 3,7 | 4,8 | 6,1 | 4,8 |
| NBI Material de las Paredes | 11,2 | 15,7 | 15,7 | 14,8 | 4,9 | 10,7 | 15,7 | 9,5 | 6,5 | 1,6 | 15,6 | 15,9 | 15,1 | 15,2 | 11,5 | 7,8 | 14,7 | 12,3 | 13,3 | 12,0 |
| NBI Material de Pisos | 12,8 | 33,2 | 33,4 | 29,5 | 16,6 | 27,6 | 33,4 | 13,8 | 4,4 | 1,6 | 32,7 | 35,2 | 32,1 | 32,1 | 21,7 | 23,2 | 32,1 | 29,2 | 25,5 | 31,5 |
| NBI Saneamiento | 4,8 | 5,5 | 6,7 | 6,0 | 30,1 | 15,0 | 3,6 | 28,5 | 34,4 | 22,7 | * | * | 4,2 | 13,1 | 19,3 | 21,6 | 6,4 | 10,3 | 14,1 | 9,0 |
| Pobreza | 4,7 | 3,1 | 4,0 | 3,8 | 14,0 | 9,0 | 2,4 | 13,8 | 15,2 | 7,4 | * | 32,2 | 2,5 | 9,9 | 12,8 | 9,6 | 4,7 | 5,8 | 9,8 | 4,0 |
| Pobreza extrema | 3,9 | 3,4 | 5,1 | 3,3 | 11,3 | 6,7 | 3,1 | 12,3 | 10,9 | 5,0 | 1,2 | 11,9 | 4,2 | 4,7 | 12,2 | 6,8 | 3,9 | 4,6 | 8,5 | 2,8 |
| Tasa de desempleo | 1,5 | 4,4 | 7,3 | 6,5 | 7,3 | 9,6 | 9,0 | 9,8 | 11,1 | 10,8 | 4,9 | 11,6 | 8,0 | 5,3 | 9,0 | | 4,5 | 9,0 | 10,2 | 4,2 |
| Tasa de ocupación | 2,4 | 1,9 | 2,6 | 3,1 | 15,9 | 11,7 | 11,3 | 10,8 | 15,3 | 19,9 | 1,0 | 8,1 | 3,0 | 3,0 | 15,2 | | 3,7 | 3,7 | 5,6 | 13,2 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A15
Efectos de diseño para Nicaragua

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Etnia: Indígena | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|-----------------------------------|-------|-----------------|----------------|---|---|---|---|---|--|--|---------------|----------------|--------------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 6,2 | 3,2 | 4,8 | * | 2,2 | 1,8 | 3,4 | 1,2 | 1,8 | 1,4 | 1,1 | 7,4 | * | * | * | 2,4 | 2,2 | 1,5 | 3,4 |
| NBI Agua | 4,6 | 8,3 | 7,9 | 4,7 | 5,3 | 7,3 | 13,6 | 5,0 | 1,9 | 1,6 | 12,2 | 3,8 | 5,6 | 5,4 | 7,5 | 7,5 | 7,7 | 6,5 | 7,6 |
| NBI Combustible | 12,7 | 15,8 | 15,3 | 8,4 | 9,4 | 11,9 | 27,2 | 8,1 | 3,2 | 1,7 | 35,1 | 5,6 | 9,2 | 9,6 | 12,4 | 13,0 | 12,6 | 11,1 | 12,8 |
| NBI Electricidad | 16,1 | 22,9 | 23,0 | 12,2 | 15,1 | 22,1 | 43,1 | 7,8 | 2,0 | 1,2 | * | 10,3 | 18,7 | 18,1 | 25,8 | 25,2 | 24,6 | 21,6 | 24,6 |
| NBI Internet Hogar | 6,1 | 12,8 | 13,2 | 9,1 | 9,6 | 10,8 | 14,4 | 9,2 | 7,0 | 6,5 | 16,8 | 13,4 | 8,0 | 9,1 | 11,3 | 11,5 | 11,5 | 10,4 | 11,8 |
| NBI Material de las Paredes | 7,5 | 17,7 | 17,7 | 7,8 | 10,1 | 10,6 | 17,7 | 7,3 | 4,2 | 2,6 | 18,1 | 17,5 | 3,6 | 2,4 | 15,3 | 17,5 | 14,9 | 5,4 | 15,5 |
| NBI Material de Pisos | 12,8 | 32,9 | 32,7 | 14,1 | 18,4 | 23,0 | 33,8 | 13,4 | 5,2 | 2,2 | 36,0 | 29,1 | 7,4 | 14,4 | 26,3 | 32,8 | 28,8 | 18,8 | 30,8 |
| NBI Material del techo | 13,2 | 30,9 | 30,9 | 13,0 | 15,3 | 15,3 | 31,5 | 13,0 | 6,8 | 4,7 | 31,4 | 30,4 | 6,1 | 4,3 | 26,0 | 30,2 | 24,7 | 8,1 | 27,3 |
| NBI Saneamiento | 12,1 | 18,0 | 17,7 | 15,5 | 16,4 | 19,5 | 21,2 | 15,2 | 8,6 | 4,8 | 24,4 | 12,5 | 16,7 | 15,7 | 20,4 | 19,0 | 20,6 | 18,0 | 20,4 |
| Pobreza | 4,9 | 5,2 | 5,1 | 3,5 | 4,4 | 5,2 | 5,2 | 3,5 | 1,5 | * | 5,9 | 4,2 | 4,1 | 3,9 | 5,6 | 5,9 | 5,8 | 4,3 | 5,9 |
| Pobreza extrema | 5,5 | 4,2 | 4,1 | 2,8 | 3,8 | 5,0 | 4,4 | 3,5 | * | * | 4,6 | 3,3 | 4,8 | 4,0 | 5,3 | 5,2 | 5,3 | 4,1 | 5,2 |
| Tasa de desempleo | 1,7 | 1,7 | 2,0 | 2,2 | 2,3 | 2,1 | 2,4 | 2,5 | 2,0 | 2,4 | 2,6 | 1,5 | 1,5 | * | | 1,6 | 2,5 | 1,8 | * |
| Tasa de ocupación | 4,1 | 5,1 | 5,5 | 4,5 | 4,6 | 5,3 | 6,2 | 4,9 | 3,6 | 4,2 | 6,6 | 4,5 | 2,8 | 3,5 | | 5,7 | 5,8 | 5,6 | 2,6 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A16
Efectos de diseño para Panamá

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Etnia: Afro | Etnia: Indígena | Etnia: No indígena ni afro | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|-----------------------------|-------|--------------|-------------|---|---|------------------------------------|--------------------------------------|--------------------------------------|---------------------------------------|---|------------|-------------|-------------|-----------------|----------------------------|-----------|-----------|-------------|-------------|-------------|------------|
| Ingresos | 5,1 | 13,4 | 12,7 | 4,2 | 4,3 | 4,4 | 4,8 | 10,5 | 5,8 | 11,0 | 6,3 | 18,7 | 8,2 | 4,4 | 17,2 | 5,0 | 1,4 | 7,5 | 7,2 | 7,1 | 5,3 |
| NBI Agua | 1,9 | 8,1 | 8,1 | 7,1 | 6,7 | 7,5 | 8,7 | 5,4 | 3,0 | 1,8 | 24,3 | 9,9 | 6,8 | 9,8 | 8,6 | 6,2 | 8,0 | 7,3 | 7,2 | 7,0 | 8,0 |
| NBI Combustible | 4,5 | 9,6 | 9,3 | 5,2 | 4,8 | 7,4 | 11,0 | 2,6 | * | * | * | 1,8 | 4,0 | 11,7 | 8,5 | 6,1 | 7,6 | 7,5 | 6,4 | 5,9 | 8,7 |
| NBI Computador Hogar | 5,8 | 6,0 | 6,0 | 5,7 | 4,1 | 5,5 | 5,9 | 5,5 | 2,6 | 2,2 | 5,2 | 6,8 | 6,0 | 5,4 | 6,4 | 5,4 | 5,0 | 5,9 | 5,5 | 5,5 | 5,9 |
| NBI Electricidad | 11,4 | 23,2 | 23,5 | 29,0 | 30,0 | 28,8 | 29,8 | 17,6 | 5,8 | 3,7 | * | 23,1 | 24,4 | 36,9 | 24,6 | 27,1 | 32,9 | 29,2 | 30,5 | 28,7 | 28,3 |
| NBI Internet Hogar | 8,9 | 8,9 | 8,9 | 7,5 | 5,9 | 8,2 | 8,5 | 6,7 | 2,8 | 2,1 | 10,8 | 9,7 | 8,2 | 8,0 | 9,1 | 8,3 | 7,3 | 8,0 | 7,3 | 7,8 | 8,5 |
| NBI Material de las Paredes | 7,4 | 17,8 | 17,7 | 7,0 | 8,1 | 11,8 | 18,6 | 3,6 | * | * | * | 2,7 | 2,7 | 42,1 | 6,0 | 7,9 | 15,3 | 14,6 | 11,8 | 8,8 | 17,7 |
| NBI Material de Pisos | 8,1 | 15,8 | 15,9 | 9,0 | 8,5 | 12,3 | 15,9 | 3,8 | 1,1 | * | 28,8 | 9,7 | 5,5 | 16,8 | 14,7 | 9,2 | 13,4 | 13,5 | 11,4 | 9,1 | 15,7 |
| NBI Material del techo | 7,8 | 42,6 | 42,5 | 17,5 | 15,9 | 26,8 | 46,9 | 4,3 | 1,1 | * | | | 2,7 | * | 20,1 | 17,1 | 38,0 | 33,8 | 27,2 | 19,9 | 42,0 |
| NBI Saneamiento | 12,3 | 11,7 | 11,7 | 12,5 | 10,3 | 11,9 | 11,5 | 10,3 | 5,4 | 4,5 | 10,3 | 12,6 | 13,4 | 10,8 | 13,3 | 9,1 | 12,3 | 13,0 | 12,3 | 11,6 | 13,1 |
| Pobreza | 3,4 | 3,9 | 3,8 | 3,9 | 3,2 | 4,1 | 3,8 | 2,6 | * | * | 3,7 | 4,7 | 3,9 | 3,8 | 4,3 | 2,7 | 4,4 | 4,3 | 4,0 | 3,4 | 4,4 |
| Pobreza extrema | 5,5 | 4,4 | 4,3 | 5,0 | 4,5 | 5,4 | 4,3 | 3,8 | 1,2 | * | 7,5 | 6,2 | 3,5 | 4,3 | 5,3 | 2,9 | 5,7 | 5,4 | 5,4 | 4,5 | 5,1 |
| Tasa de desempleo | 2,8 | 2,4 | 2,1 | 3,1 | 2,3 | 1,9 | * | 2,8 | 2,9 | 2,7 | 2,8 | 2,0 | 2,9 | 2,3 | 2,0 | * | | 1,9 | 3,0 | 2,7 | * |
| Tasa de ocupación | 3,9 | 6,5 | 7,4 | 8,0 | 4,9 | 7,3 | 5,8 | 8,5 | 7,1 | 8,5 | 6,3 | 7,6 | 8,2 | 6,6 | 6,8 | 6,6 | | 8,1 | 8,0 | 8,1 | 2,0 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A17
Efectos de diseño para Perú

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Etnia: Afro | Etnia: Indígena | Etnia: No indígena ni afro | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|----------------------------|-------|-----------------|----------------|---|---|---|---|---|--|--|---------------|----------------|----------------|--------------------|--|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 19,9 | 22,5 | 22,6 | 9,5 | 8,1 | 4,1 | 27,4 | 16,9 | 17,3 | 15,5 | 21,2 | 24,1 | 5,8 | 16,3 | 25,1 | 13,6 | 8,7 | 19,2 | 13,4 | 14,0 | 16,2 |
| NBI Agua | 8,1 | 7,5 | 7,8 | 10,5 | 9,4 | 8,1 | 4,2 | 9,5 | 13,5 | 9,2 | * | 40,0 | 10,9 | 5,4 | 14,6 | 8,2 | 8,3 | 5,9 | 6,3 | 7,2 | 3,8 |
| NBI Combustible | 7,4 | 6,8 | 8,5 | 13,8 | 13,2 | 11,9 | 7,6 | 10,2 | 17,2 | 10,9 | 22,2 | 16,2 | 18,4 | 9,2 | 9,5 | 10,1 | 12,6 | 6,7 | 7,3 | 7,8 | 4,7 |
| NBI Computador Hogar | 9,5 | 8,6 | 9,0 | 8,7 | 8,7 | 8,2 | 7,7 | 10,5 | 9,2 | 4,6 | 4,4 | 17,9 | 7,1 | 7,5 | 9,7 | 9,7 | 9,1 | 10,0 | 9,8 | 9,6 | 9,1 |
| NBI Electricidad | 16,1 | 8,6 | 10,7 | 36,6 | 34,3 | 28,6 | 11,7 | 37,3 | 35,4 | 13,3 | * | * | 42,2 | 18,1 | 17,3 | 31,4 | 29,3 | 21,4 | 23,5 | 26,4 | 12,1 |
| NBI Internet Hogar | 10,8 | 13,2 | 15,5 | 14,9 | 14,8 | 13,9 | 9,8 | 13,9 | 16,7 | 10,7 | 2,9 | * | 13,9 | 9,9 | 21,6 | 12,9 | 14,7 | 12,2 | 12,2 | 12,6 | 10,1 |
| NBI Saneamiento | 9,2 | 17,5 | 17,9 | 17,6 | 17,1 | 16,4 | 16,8 | 21,9 | 17,1 | 8,5 | 12,1 | 35,1 | 12,8 | 15,3 | 15,4 | 16,9 | 18,8 | 20,6 | 20,8 | 19,6 | 19,7 |
| Pobreza | 2,9 | 8,1 | 8,4 | 9,8 | 10,0 | 9,4 | 7,6 | 11,3 | 10,0 | 5,0 | 4,9 | 15,6 | 6,9 | 7,1 | 7,8 | 9,9 | 10,4 | 10,4 | 10,2 | 10,7 | 8,9 |
| Pobreza extrema | 4,1 | 6,7 | 7,2 | 9,4 | 9,8 | 9,0 | 5,8 | 9,4 | 9,5 | 5,3 | 3,5 | 16,3 | 6,4 | 5,7 | 6,7 | 7,5 | 9,4 | 8,6 | 8,4 | 9,7 | 6,8 |
| Tasa de desempleo | 2,8 | 3,9 | 4,8 | 5,7 | 4,6 | 2,6 | 4,4 | 6,1 | 6,3 | 6,3 | 2,6 | 18,8 | 4,6 | 6,4 | 3,8 | 4,4 | | 4,2 | 6,5 | 6,7 | 4,0 |
| Tasa de ocupación | 4,5 | 7,5 | 8,0 | 12,1 | 11,1 | 11,3 | 10,8 | 12,4 | 14,9 | 13,6 | 5,6 | 12,4 | 12,1 | 8,8 | 8,3 | 13,1 | | 12,3 | 10,5 | 10,8 | 5,4 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A18
Efectos de diseño para Paraguay

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área: rural | Área: urbana | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|----------------------------|-------|-----------------|----------------|---|---|---|---|---|--|--|----------------|-----------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 12,7 | 6,8 | 7,6 | 2,1 | 1,2 | 2,2 | 5,7 | 2,4 | 3,6 | 1,7 | 4,9 | 9,2 | 1,7 | 1,7 | 4,9 | 3,1 | 2,3 | 2,7 |
| NBI Agua | 3,6 | 2,5 | 2,6 | 2,0 | 1,8 | 2,5 | 2,5 | 1,4 | * | * | 2,6 | 2,9 | 1,9 | 2,1 | 2,7 | 2,6 | 2,1 | 2,7 |
| NBI Combustible | 9,1 | 12,2 | 11,3 | 8,0 | 7,2 | 9,3 | 11,5 | 6,8 | 3,8 | * | 18,7 | 11,9 | 7,3 | 8,2 | 10,2 | 9,3 | 8,5 | 10,1 |
| NBI Computador Hogar | 3,9 | 3,3 | 3,3 | 2,7 | 2,1 | 2,8 | 3,3 | 2,6 | 1,8 | * | 3,2 | 3,4 | 2,1 | 2,6 | 3,4 | 3,1 | 2,5 | 3,2 |
| NBI Electricidad | 1,2 | 1,4 | 1,5 | * | * | 1,3 | 1,4 | * | * | | 1,4 | 1,5 | 1,3 | 1,2 | 1,5 | 1,3 | * | 1,0 |
| NBI Internet Hogar | 8,4 | 7,5 | 7,4 | 6,0 | 5,1 | 6,4 | 7,3 | 5,6 | 4,6 | 1,3 | 7,2 | 7,8 | 5,0 | 6,2 | 7,3 | 7,0 | 5,7 | 7,2 |
| NBI Material de Pisos | 2,8 | 5,6 | 5,6 | 2,7 | 3,0 | 4,4 | 5,6 | 2,0 | * | * | 6,0 | 5,3 | 3,0 | 3,9 | 5,2 | 4,3 | 3,4 | 5,3 |
| NBI Material del techo | 3,9 | 5,8 | 5,8 | 2,4 | 3,0 | 4,8 | 5,8 | 1,6 | * | | 8,9 | 3,8 | 3,9 | 3,3 | 4,4 | 4,9 | 4,3 | 5,1 |
| NBI Saneamiento | 3,2 | 7,0 | 7,1 | 6,6 | 5,2 | 6,8 | 7,1 | 6,5 | 4,9 | 1,4 | 7,3 | 7,4 | 4,9 | 6,8 | 7,9 | 7,7 | 6,1 | 7,8 |
| Pobreza | 4,5 | 2,2 | 2,2 | 2,1 | 1,8 | 2,2 | 2,2 | 1,6 | * | * | 2,1 | 2,3 | 1,6 | 2,3 | 2,5 | 2,4 | 1,9 | 2,6 |
| Pobreza extrema | 3,6 | 1,8 | 1,9 | 2,0 | 2,0 | 2,3 | 1,8 | 1,6 | * | * | 1,6 | 2,4 | 1,6 | 2,3 | 2,4 | 2,4 | 2,0 | 2,3 |
| Tasa de desempleo | 2,2 | 1,4 | 1,1 | 1,6 | 1,1 | 1,2 | 1,1 | 1,6 | 1,6 | * | 1,6 | * | * | | 1,0 | 1,6 | 1,1 | * |
| Tasa de ocupación | 3,4 | 3,8 | 4,2 | 3,8 | 2,6 | 3,9 | 3,8 | 4,3 | 4,2 | 2,6 | 4,0 | 3,9 | 2,7 | | 4,4 | 4,4 | 4,2 | 1,1 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A19
Efectos de diseño para El Salvador

| Variable ^a | Total | Sexo: hombre | Sexo: mujer | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|--------------------------------|-------|-----------------|----------------|---|---|---|---|---|--|--|---------------|----------------|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 19,3 | 12,3 | 12,8 | 6,8 | 3,6 | 3,5 | 16,2 | 3,8 | 3,2 | 1,6 | 12,9 | 14,0 | 5,1 | 4,0 | 9,9 | 4,3 | 4,5 | 6,9 |
| NBI Agua | 3,9 | 7,0 | 7,7 | 7,1 | 5,7 | 5,8 | 7,8 | 5,7 | 2,1 | * | 6,6 | 8,2 | 6,3 | 6,1 | 6,8 | 7,1 | 6,5 | 7,0 |
| NBI Combustible | 10,8 | 8,9 | 8,5 | 5,8 | 4,5 | 4,9 | 10,8 | 2,7 | * | * | 16,8 | 7,3 | 6,9 | 5,1 | 7,6 | 6,8 | 6,0 | 6,7 |
| NBI Computador Hogar | 4,9 | 6,3 | 6,5 | 6,3 | 4,3 | 4,7 | 6,2 | 5,3 | 2,4 | 1,7 | 5,8 | 7,0 | 5,6 | 4,8 | 6,4 | 6,3 | 5,8 | 6,0 |
| NBI Electricidad | 20,2 | 19,1 | 19,8 | 31,1 | 32,0 | 34,3 | 22,5 | 32,9 | 9,9 | 3,2 | 36,2 | 21,3 | 34,0 | 31,1 | 19,9 | 30,5 | 34,7 | 27,8 |
| NBI Material de las Paredes | 2,7 | 2,7 | 2,7 | 1,9 | 1,2 | 1,4 | 2,7 | * | * | * | 3,2 | 2,3 | 1,7 | 1,7 | 2,6 | 2,4 | 1,8 | 1,6 |
| NBI Material de Pisos | 10,2 | 23,8 | 23,7 | 18,6 | 13,2 | 12,9 | 23,9 | 7,7 | 1,1 | * | 25,3 | 23,4 | 15,4 | 15,5 | 23,3 | 20,5 | 16,7 | 20,8 |
| NBI Saneamiento | 6,2 | 6,7 | 6,6 | 7,2 | 4,8 | 4,5 | 6,5 | 5,3 | 1,5 | * | 6,5 | 6,8 | 3,7 | 6,6 | 7,3 | 7,3 | 4,5 | 7,3 |
| Pobreza | 5,5 | 5,6 | 5,5 | 6,3 | 4,7 | 5,0 | 5,4 | 4,1 | 1,9 | * | 5,2 | 5,9 | 5,4 | 5,8 | 6,3 | 6,4 | 5,4 | 6,3 |
| Pobreza extrema | 5,6 | 4,7 | 4,6 | 5,7 | 4,7 | 5,0 | 4,7 | 3,9 | 1,1 | * | 4,9 | 5,5 | 5,5 | 5,6 | 5,7 | 6,0 | 5,0 | 5,7 |
| Tasa de desempleo | 3,6 | 2,0 | 2,7 | 2,9 | 1,8 | 1,8 | 2,9 | 3,3 | 2,5 | 2,2 | 2,6 | 2,0 | 1,8 | | 2,0 | 3,2 | 2,6 | * |
| Tasa de ocupación | 4,4 | 5,7 | 6,5 | 7,1 | 4,4 | 5,1 | 6,6 | 7,0 | 4,4 | 4,8 | 6,0 | 6,1 | 5,2 | | 6,7 | 6,9 | 7,1 | 1,7 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A20
Efectos de diseño para Uruguay

| Variable ^a | Total | Sexo: hombre | Sexo: mujer ^a | Nivel educativo: Alta secundaria incompleta | Nivel educativo: Baja secundaria incompleta | Nivel educativo: Primaria completa | Nivel educativo: Primaria incompleta | Nivel educativo: Secundaria completa | Nivel educativo: Terciaria incompleta | Nivel educativo: Universitaria completa | Área rural | Área urbana | Etnia: Afro | Etnia: Indígena | Etnia: No indígena ni afro | Edad :>60 | Edad : 0-5 | Edad : 16- 30 | Edad : 31- 45 | Edad : 46- 60 | Edad: 6-15 |
|-----------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|---------------|----------------|----------------|--------------------|----------------------------------|--------------|---------------|---------------------|---------------------|---------------------|---------------|
| Ingresos | * | * | * | * | 30,9 | 37,4 | 40,1 | 25,5 | * | * | 14,7 | * | 8,3 | 4,5 | * | * | 11,7 | 36,7 | 42,8 | 47,2 | 24,0 |
| Pobreza | 18,7 | 3,2 | 3,1 | 3,5 | 3,6 | 3,3 | 3,4 | * | * | * | 2,5 | 5,9 | 2,5 | * | 4,9 | 1,2 | 3,3 | 3,6 | 3,5 | 2,5 | 3,6 |
| Pobreza extrema | 8,9 | * | * | * | * | * | * | * | * | * | * | 1,1 | * | * | * | * | * | * | * | * | * |
| Tasa de desempleo | 1,8 | 5,5 | 4,4 | 6,5 | 10,5 | 10,9 | 8,3 | 10,2 | 9,9 | 4,7 | 5,5 | 42,2 | 9,3 | 2,2 | 14,8 | 4,7 | | 4,8 | 8,9 | 9,6 | * |
| Tasa de ocupación | 2,5 | 8,0 | 7,1 | 10,9 | 15,6 | 14,7 | 9,8 | 14,6 | 13,0 | 13,4 | 11,1 | * | 10,6 | 4,8 | 48,6 | 11,8 | | 14,8 | 9,6 | 8,8 | * |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.

Cuadro A21
Efectos de diseño para Venezuela

| Variable ^a | Total | Sexo: Hombre | Sexo: Mujer ^a | Nivel Educativo: Alta secundaria incompleta | Nivel Educativo: Baja secundaria incompleta | Nivel Educativo: Primaria completa | Nivel Educativo: Primaria incompleta | Nivel Educativo: Secundaria completa | Nivel Educativo: Terciaria incompleta | Nivel Educativo: Universitaria completa | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|--------------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|--------------|--------------|----------------|----------------|----------------|---------------|
| Ingresos | 29,4 | 37,8 | 37,8 | 15,5 | 14,6 | 22,7 | 34,0 | 27,2 | 24,0 | 21,6 | 22,6 | 16,9 | 35,4 | 26,0 | 27,0 | 25,5 |
| NBI Agua | 6,9 | 11,2 | 7,4 | 10,6 | 10,8 | 8,0 | 6,8 | 9,3 | 10,7 | 8,6 | 8,9 | 9,4 | 5,7 | 7,1 | 8,8 | 6,7 |
| NBI Combustible | 7,5 | 8,2 | 7,7 | 4,2 | 4,6 | 5,7 | 7,5 | 4,8 | 3,9 | 3,0 | 4,1 | 5,6 | 7,0 | 6,5 | 5,4 | 6,2 |
| NBI Computador Hogar | 8,6 | 9,0 | 8,5 | 8,8 | 8,7 | 9,3 | 8,2 | 9,6 | 8,5 | 6,3 | 8,0 | 9,3 | 8,9 | 9,8 | 9,2 | 9,8 |
| NBI Electricidad | 13,7 | 14,3 | 14,2 | 8,5 | 7,9 | 12,5 | 13,9 | 11,3 | 7,7 | 7,9 | 11,3 | 13,9 | 14,7 | 13,1 | 10,5 | 14,3 |
| NBI Internet Hogar | 12,1 | 13,4 | 11,5 | 14,9 | 14,7 | 11,0 | 6,7 | 11,8 | 15,5 | 15,3 | 12,4 | 13,8 | 7,7 | 10,3 | 11,6 | 10,1 |
| NBI Material de Pisos | 23,1 | 23,9 | 24,0 | 11,1 | 15,4 | 20,6 | 23,8 | 12,3 | 6,4 | 3,0 | 8,6 | 19,1 | 23,1 | 20,6 | 15,0 | 23,1 |
| NBI Saneamiento | 16,9 | 14,1 | 14,2 | 10,3 | 13,1 | 15,6 | 14,2 | 11,5 | 6,0 | 2,6 | 9,0 | 15,8 | 15,7 | 15,7 | 12,9 | 16,0 |

| Variable ^a | Total | Sexo: Hombre | Sexo: Mujer ^a | Nivel Educativo: Alta secundaria incompleta | Nivel Educativo: Baja secundaria incompleta | Nivel Educativo: Primaria completa | Nivel Educativo: Primaria incompleta | Nivel Educativo: Secundaria completa | Nivel Educativo: Terciaria incompleta | Nivel Educativo: Universitaria completa | Edad: >60 | Edad: 0-5 | Edad: 16-30 | Edad: 31-45 | Edad: 46-60 | Edad: 6-15 |
|-----------------------|-------|-----------------|-----------------------------|---|---|---|---|---|--|--|--------------|--------------|----------------|----------------|----------------|---------------|
| Pobreza | 2,1 | 8,3 | 8,4 | 11,0 | 11,2 | 11,0 | 7,9 | 11,2 | 10,1 | 5,6 | 7,0 | 11,1 | 9,5 | 11,0 | 10,7 | 9,4 |
| Pobreza extrema | 6,7 | 6,3 | 7,0 | 11,1 | 11,0 | 9,5 | 5,9 | 10,3 | 10,7 | 6,0 | 8,8 | 8,6 | 6,7 | 8,9 | 10,9 | 6,0 |
| Tasa de desempleo | 4,7 | 4,6 | 5,0 | 8,0 | 7,9 | 8,3 | 7,5 | 7,8 | 8,4 | 8,7 | 5,5 | | 4,7 | 7,4 | 8,5 | 3,8 |
| Tasa de ocupación | 4,1 | 7,6 | 8,4 | 13,6 | 12,9 | 13,2 | 12,4 | 12,3 | 14,2 | 14,0 | 10,9 | | 10,3 | 9,4 | 11,5 | 2,8 |

Fuente: Elaboración propia.

^a Cálculos basados en las encuestas de hogares del repositorio BADEHOG de la CEPAL.

Nota: Las celdas con * indican que el dato no está disponible.



NACIONES UNIDAS

Serie

C E P A L

Estudios Estadísticos

Números publicados

Un listado completo, así como los archivos pdf están disponibles en
www.CEPAL.org/publicaciones

106. Efectos de diseño para indicadores sociales en América Latina: función generalizada de varianza para estimadores directos provenientes de encuestas de hogares, Andrés Gutiérrez Giovany y Babativa-Márquez (LC/TS.2023/95), 2023.
105. Modelos de unidad para la generación de mapas de pobreza a nivel subnacional, Andrés Gutiérrez, Xavier Mancero, Gabriel Nieto, Felipe Molina y Diego Lemus (LC/TS.2022/191), 2022.
104. Cambio de año de referencia de los agregados regionales anuales en las cuentas nacionales, C. de Camino y otros (LC/TS.2022/158), 2022.
103. Predicciones agregadas de pobreza con información a escala micro y macro: evaluación, diagnóstico y propuestas, Walter Sosa Escudero y Magdalena Cornejo (LC/TS.2022/95), 2022.
102. La medición de la discriminación en base al autorreporte: estado de situación y desafíos, Pablo Villatoro (LC/TS.2021/87), 2021.
101. Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional, Andrés Gutiérrez, Xavier Mancero, Álvaro Fuentes, Felipe López, Felipe Molina (LC/TS.2020/52), 2020.
100. Desafíos en el diseño de medidas de pobreza multidimensional, María Emma Santos (LC/TS.2019/5), 2019.
99. Non monetary indicators to monitor SDG targets 1.2 and 1.4: Standards, availability, comparability and quality, María Emma Santos (LC/TS.2019/4), 2018.
98. Una propuesta de estimación del producto interno bruto trimestral de América Latina y el Caribe, Giannina López, Alberto Malmierca (LC.TS.2018/88), 2018.
97. Desagregación de datos en encuestas de hogares. Metodologías de estimación en áreas pequeñas, Isabel Molina (LC/TS.2018/82), 2018.

ESTUDIOS ESTADÍSTICOS

Números publicados:

- 106 Efectos de diseño para indicadores sociales en América Latina
Función generalizada de varianza para estimadores directos provenientes de encuestas de hogares
Andrés Gutiérrez y Giovany Babativa-Márquez
- 105 Modelos de unidad para la generación de mapas de pobreza a nivel subnacional
Andrés Gutiérrez, Xavier Mancero, Gabriel Nieto Felipe Molina y Diego Lemus
- 104 Cambio de año de referencia de los agregados regionales anuales en las cuentas nacionales
Claudia de Camino, Giannina López, Patricia Marchant, Ana Victoria Vega y Alfonso Zerón
- 103 Predicciones agregadas de pobreza con información a escala micro y macro
Evaluación, diagnóstico y propuestas
Walter Sosa Escudero y Magdalena Cornejo