

SOBRESTIMACIÓN DE LA TASA DE MORTALIDAD INFANTIL OBTENIDA CON EL MÉTODO INDIRECTO. UN MODELO PARA SU CORRECCIÓN

OVERESTIMATION OF THE INFANT MORTALITY RATE OBTAINED WITH THE INDIRECT METHOD. A MODEL FOR ITS CORRECTION

Ramiro Coa Clemente¹

Instituto de Estadística Teórica y Aplicada, Universidad Mayor de San Andrés, La Paz-Bolivia

✉ clementecoa@gmail.com

Artículo recibido: 13/08/2023

Artículo aceptado: 29/09/2023

RESUMEN

Esta investigación tiene dos objetivos, por una parte, comprobar si el método indirecto para estimar tasas de mortalidad infantil en Bolivia conduce sistemáticamente a una sobrestimación del indicador y, por otra parte, corregir dicha sobrestimación si se constata la existencia de tal problema. Asumiendo que las tasas de mortalidad infantil obtenidas con el método directo son correctas, se confirma que con el método indirecto sistemáticamente se obtienen tasas sobrestimadas. Adicionalmente, hay evidencia estadística de que mayor sobrestimación se produce en niveles bajos de mortalidad infantil y que la sobrestimación ocurre con cualquiera de las familias modelo de mortalidad de Coale y Demeny que se use para su cálculo. Para corregir el problema de sobrestimación en la tasa de mortalidad infantil se seleccionó el modelo de regresión lasso, pues, las tasas de mortalidad infantil predichas con este modelo fueron más próximas a las tasas correctas que las tasas predichas con otros modelos. Se destaca que el modelo lasso seleccionado incluye una única variable predictora, cuyo cálculo es muy simple y no requiere el uso de modelos de mortalidad: la proporción de hijos muertos en mujeres de 25-29 años de edad. Debido a que países como Bolivia aún requieren del método indirecto principalmente para estimar tasas de mortalidad infantil para pequeñas áreas, como son gran parte de los municipios, es necesario seguir explorando modelos estadísticos que permitan obtener tasas de mortalidad infantil no sobrestimadas.

Palabras clave: Método directo, Aprendizaje automático, Tablas modelo de mortalidad, Regresión Lasso.

ABSTRACT

This research has two objectives: on the one hand, to check whether the indirect method for estimating infant mortality rates in Bolivia systematically leads to an overestimation of the indicator and, on the other hand, to correct this overestimation if such a problem is found to exist. Assuming that the infant mortality rates obtained with the direct method are correct, it is confirmed that the indirect method systematically leads to overestimated rates. In addition, there is statistical evidence that greater overestimation occurs with any of the Coale and Demeny mortality model families used for its calculation. To correct the problem of overestimation in the infant mortality rate, the lasso regression model was selected, since the infant mortality rates predicted with this model were closer to the correct rates than the rates predicted with other models. It is noteworthy that the lasso model selected includes a single predictor variable, whose calculation is very simple and does not require the use of mortality models: the proportion of dead children in women aged 25-29 years. Since countries such as Bolivia still require the indirect method mainly to

¹ Director del Instituto de Estadística Teórica y Aplicada, UMSA. Ex-Director de Investigación en la Unidad de Análisis y Política Social de Bolivia. Ex-Director Nacional de la Encuesta de Demografía y Salud. M.Sc. Estadística, Pontificia Católica de Chile. Mag. Demografía, Centro Latinoamericano de Demografía. Candidato a Doctor en Demografía, Universidad Federal de Minas Gerais. ORCID: [0000-0002-2955-0204](https://orcid.org/0000-0002-2955-0204).

estimate infant mortality rates for small areas, such as most municipalities, it is necessary to continue exploring statistical models that allow us to obtain infant mortality rates that are not overestimated.

Keywords: Direct method, Machine learning, Mortality model tables, Lasso regression

1. INTRODUCCIÓN

La elevada mortalidad infantil que aún persiste en algunos países o en algunos grupos de población es resultado de muchos factores que afectan negativamente la condición de salud de muchos infantes, produciendo finalmente su muerte. Baja calidad o incluso inexistencia de los servicios públicos de salud, limitado acceso a alimentos saludables y falta de acceso tanto a agua potable como a un mejor saneamiento, son algunos de esos factores. No es menos importante el incremento en los últimos años de infanticidios, hechos atribuibles a otros factores imperantes. Es por ello que la mortalidad infantil es considerado como uno de los indicadores más importantes de la situación de salud de una población así como uno de los principales indicadores de su desarrollo socioeconómico. Es más, entre los objetivos de desarrollo sostenible para el periodo 2015-2030, acordados por los países miembros de las Naciones Unidas, se encuentran aspectos relacionados a la mortalidad infantil. En efecto, una de las metas del tercer objetivo de desarrollo sostenible, referido a salud y bienestar, tiene relación con la reducción de la mortalidad infantil (United Nations, 2020).

Por lo anterior, resalta la necesidad de tener buenas estimaciones de la tasa de mortalidad infantil tanto para las distintas divisiones administrativas (municipal, departamental y país) como para diversos grupos de población (por ejemplo, poblaciones pobres y no pobres). Hay dos métodos para estimar la denominada tasa de mortalidad infantil: (i) el método directo y (ii) el método indirecto.

Debido a que las estadísticas vitales en el país aún tienen el problema de subregistro de los hechos vitales, la información usada para estimar la tasa de mortalidad infantil con cualquiera de los dos métodos generalmente proviene de encuestas y censos. El método directo requiere de información contenida en la historia de nacimientos de cada una de las mujeres en edad reproductiva, historia en la que se registra el sexo, la fecha de nacimiento, la condición de sobrevivencia de cada hijo nacido vivo, la edad actual si el niño está vivo y la edad a la que murió en caso de que ocurrió un deceso. En el país, este tipo de información es la que se obtiene sólo en las encuestas de demografía y salud. Con relación al método indirecto, éste requiere solamente de información sobre el número de hijos nacidos vivos e hijos fallecidos para cada mujer en edad fértil, información que se recolecta tanto en encuestas, incluido las encuestas de demografía y salud, como en los censos.

Además, es importante también mencionar que si bien el método indirecto es mucho más simple de aplicar, éste, y a diferencia del método directo, se basa en un conjunto de supuestos, entre los cuales se puede mencionar que el modelo de mortalidad elegido represente la estructura de la mortalidad por edad de la población en estudio. En efecto, para estimar la tasa de mortalidad infantil mediante el método indirecto, específicamente, para transformar las proporciones de defunciones en probabilidades de muerte, generalmente se hace uso de las tablas modelo de mortalidad de Coale y Demeny, las cuales se dividen en cuatro familias denominadas Este, Oeste, Norte y Sur. Cada familia de tablas

Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección

modelo, construida principalmente con base en la antigua estructura de mortalidad de países europeos, define un comportamiento particular de la mortalidad por edad. Por ejemplo, la familia Este de tablas modelo de mortalidad se caracteriza por una elevada mortalidad en la infancia y una creciente alta mortalidad después de los 50 años de edad; la familia Norte presenta baja mortalidad en la infancia y una mortalidad mucha más baja en edades por encima de los 50 años de edad; mientras la familia Sur se caracteriza por una muy elevada mortalidad debajo de los 5 años de edad, baja mortalidad entre los 40 y 50 años, y una elevada mortalidad por encima de los 65 años. En cambio, el modelo Oeste no presenta un patrón sistemático de la mortalidad en función de la edad y se lo presenta como un promedio o como una familia residual (Coale and Demeny, 1983). En el pasado, sobretodo cuando el nivel de mortalidad infantil en Bolivia era muy alto, para las estimaciones indirectas se asumía el modelo Sur de las tablas modelo de mortalidad de Coale y Demeny. Actualmente, sin embargo, se hacen algunas estimaciones indirectas de la mortalidad infantil usando el modelo Oeste (UDAPE, 2018; INE, 2021). Cabe remarcar que el método directo no se basa en ningún supuesto ni requiere del uso de tablas modelo de mortalidad.

Cualquiera sea el modelo de mortalidad asumido, el problema central con el método indirecto parece ser la sobrestimación del verdadero nivel de mortalidad infantil. Efectivamente, las estimaciones de mortalidad infantil a partir del método indirecto son generalmente más altas que la mortalidad estimada con el método directo, el cual, como se mencionó, hace uso de la historia de nacimientos (Ministry of Finance and Economic Planning of Uganda, 1995). Además del incumplimiento de algunos de los supuestos del método

indirecto, la sobrestimación se debe en gran medida a que las estructuras de mortalidad implícitas en las tablas modelo de mortalidad no corresponden a la estructura de mortalidad del país o de la población analizada. En efecto, las estimaciones de la mortalidad infantil por métodos indirectos son especialmente sensibles a la elección del modelo de mortalidad y pueden estar severamente sesgadas cuando el modelo utilizado se desvía marcadamente del patrón de mortalidad prevalente en la población de estudio (United Nations, 1990).

Por lo expuesto, el presente trabajo tiene un doble objetivo. Primero, determinar si el método indirecto para estimar tasas de mortalidad infantil en Bolivia conduce sistemáticamente a una sobrestimación del indicador y, segundo, corregir dicha sobrestimación si se constata la existencia de tal problema. Cabe mencionar que en el análisis se consideran tres familias de mortalidad de las tablas modelo de mortalidad de Coale y Demeny, esto con el propósito de determinar si el problema de sobrestimación depende de la familia de mortalidad elegida.

2. MÉTODOS E INFORMACIÓN

En este trabajo, las tasas de mortalidad infantil son estimadas tanto con el método indirecto como con el método directo usando una misma fuente de información, la encuesta de demografía y salud realizada en 2008. Por otra parte, para corregir la sobrestimación en la tasa de mortalidad infantil se consideran algunos modelos de regresión con función de penalización, denominados también técnicas de *machine learning* o aprendizaje automático. A continuación se hace una breve descripción de estos métodos.

2.1 Métodos de Estimación de la Tasa Mortalidad Infantil

El Método Indirecto

Un primer indicador del nivel de mortalidad es la proporción de hijos muertos con relación al total hijos nacidos vivos que tienen las mujeres de una determinada edad o de un grupo de edades. Esta medición de la mortalidad, sin embargo, tiene el problema de estar referida a la edad de las madres y no a la de los niños. Brass desarrolló un procedimiento que permite transformar esas proporciones en probabilidades de muerte, específicamente en la probabilidad de que un hijo muera antes de alcanzar una edad exacta (Brass, 1974). Otros autores han desarrollado variantes de la idea original de Brass, entre ellos Trussell. Haciendo uso de las tablas modelo de mortalidad de Coale y Demeny y de los modelos de fecundidad de Coale y Trussell, Trussell estableció ecuaciones de regresión que permiten derivar tanto las probabilidades de morir en el primer año de vida como su ubicación en el tiempo (Trussell, 1975). Este último procedimiento es el que generalmente se aplica para la estimación de la tasa de mortalidad infantil y también es el usado en este trabajo, es el procedimiento al que se ha denominado método indirecto.

El Método Directo

La tasa de mortalidad infantil se la define como la probabilidad de que un nacido vivo muera antes de cumplir el primer año de vida. En términos estrictos, esas probabilidades se miden en cohortes reales. Debido a que se requieren estimaciones para periodos de tiempo, más que para cohortes reales, para estimar la tasa de mortalidad infantil por el método directo se hace uso del concepto de tabla de vida para cohortes sintéticas. Con este enfoque, y con base en la experiencia de mortalidad de las cohortes reales, se obtienen

probabilidades de muerte para pequeños segmentos de edad y un periodo de tiempo definido. A partir de esas probabilidades de muerte se obtienen las probabilidades de sobrevivencia, las que, multiplicadas acumulativamente, producen la probabilidad de que un nacido vivo sobreviva a la edad exacta de un año. Finalmente, por complemento se obtiene la tasa de mortalidad infantil.

Este método, descrito en términos generales, es el método directo adoptado por el Programa de Encuestas Demográficas y de Salud (Programa DHS). Para el cálculo de las probabilidades de muerte en los distintos segmentos, el programa DHS adoptó los siguientes segmentos de edad cumplida 0, 1-2, 3-5 y 6-11 meses. En cambio, el segmento de tiempo o periodo de tiempo para el que se calcula la tasa de mortalidad infantil generalmente comprende los 5 años previos a la encuesta.

2.2 Modelos de Regresión para Corrección de la Sobrestimación

Para corregir el problema de sobrestimación en la tasa de mortalidad infantil calculada mediante el método indirecto se puede recurrir a modelos de regresión con función de penalización, denominados también técnicas de *machine learning*. Algunos de esos modelos tienen la siguiente estructura común. Dado el modelo de regresión

$$Y_i = \beta_0 + X_i' \beta + \varepsilon_i \quad i=1,2,\dots,n,$$

donde X_i es un vector $p * 1$ de covariables para la i -ésima observación, β es un vector $p * 1$ de coeficientes y ε_i es el error aleatorio asociado a la i -ésima observación, la función objetivo penalizada a ser minimizada es

$$\varphi = \sum_{i=1}^n (Y_i - \beta_0 + X_i' \beta)^2 + \lambda \sum_{j=1}^p [(1-\alpha)\beta_j^2 + \alpha|\beta_j|]$$

Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección

donde $(Y_i - \beta_0 + X_i' \beta)^2$ es el error de predicción en la muestra, λ es un coeficiente de penalización general que controla el grado de penalización y α es otro coeficiente que, dependiendo de su valor, determina una función de penalización específica. Los modelos de regresión con funciones de penalización específicas son:

Regresión Ridge

Si $\alpha = 0$ y $\lambda > 0$, la función ϕ se reduce a la función objetivo penalizada para obtener el vector de estimadores ridge.

Regresión ridge fue usada durante mucho tiempo como un método para mantener las variables altamente colineales en un modelo de regresión lineal con fines de predicción. Esto es, regresión no selecciona variables, mantiene todas las variables en el modelo, aunque muchos coeficientes estimados son pequeños en valor.

La penalización en la regresión ridge elimina la inestabilidad de los coeficientes estimados por mínimos cuadrados ordinarios en un modelo de regresión lineal clásico, inestabilidad debida a la presencia de alta multicolinealidad entre las covariables. Cuanto más grande el valor de λ , los coeficientes serán menores y más robustos a la colinealidad. Consecuentemente, regresión ridge produce estimadores puntuales que pueden ser usados para predicción.

Regresión Lasso

Si $\alpha = 1$ y $\lambda > 0$, la función ϕ se reduce a la función penalizada lasso, función que permite obtener el vector de estimadores lasso.

Regresión lasso es usada para selección del modelo y para fines de predicción. Es usada para seleccionar las variables más correlacionadas con la variable respuesta en un conjunto de datos y, posteriormente, para

probar si esas mismas variables predicen bien la respuesta en otros conjuntos de datos.

Regresión Red elástica

Si $0 < \alpha < 1$ y $\lambda > 0$, ϕ es la función objetivo penalizada que permite obtener el vector de estimadores de red elástica.

Igual que regresión lasso, regresión red elástica puede ser usada para predicción y para selección del modelo. Este modelo de regresión tiene un término de penalización que es una combinación de la penalidad usada por regresión lasso y la penalidad usada por regresión ridge. Los coeficientes estimados con regresión red elástica son más robustos que los estimados con regresión lasso a la presencia de covariables altamente correlacionadas.

Regresión Clásica

Si $\lambda=0$, la función ϕ se reduce a la función objetivo, no penalizada, que permite obtener el vector de estimadores en un modelo de regresión clásico.

El modelo de regresión clásico, sin embargo, presenta al menos tres problemas. Primero, no realiza una selección de predictores, todos los predictores se incorporan en el modelo aunque no aporten información relevante, por lo que se reduce su capacidad predictiva; segundo, en presencia de alta multicolinealidad, los estimadores de mínimos cuadrados ordinarios son inestables, es decir, sus magnitudes pueden cambiar considerablemente con diferentes muestras, además de producir estimadores con valores absolutos demasiado grandes y; tercero, la predicciones en otras muestras podrían ser muy pobres, produciendo considerables errores de predicción. Este modelo se incluye en el análisis sólo con la finalidad de constatar su debilidad para hacer predicciones en otras muestras.

2.3 Información

En Bolivia, la única fuente de información que permite abordar los dos objetivos de este trabajo es la encuesta nacional de demografía y salud². La información contenida en dicha encuesta, específicamente en la encuesta del año 2008, denominada ENDSA 2008, permite obtener estimaciones de la tasa de mortalidad infantil tanto por el método directo como por el método indirecto, además de permitir evaluar la corrección de las tasas sobrestimadas.

Con relación a la calidad de los datos, la información referida a la historia de nacimientos obtenida en la ENDSA 2008, la que permite aplicar el método directo, es razonablemente confiable y no está seriamente afectada por la omisión de niños muertos o por errores en las declaraciones de fechas de nacimiento y muerte. El apéndice C del informe final de la ENDSA 2008 sobre calidad de la información proporciona varios resultados al respecto (Coa y Ochoa, 2009), los que evidencian la calidad aceptable de los datos. Por ello, en este documento se toman como correctas las estimaciones de la tasa de mortalidad infantil generadas con el método directo. Igualmente, la información requerida por el método indirecto sobre el total de hijos nacidos vivos e hijos fallecidos de cada mujer en edad reproductiva es bastante confiable, puesto que el cuestionario de la ENDSA 2008 fue diseñado apropiadamente para reducir al máximo la omisión de nacimientos y muertes.

Las tasas de mortalidad infantil, tanto con el método directo como con el indirecto, fueron estimadas para el país, para las áreas

urbana y rural, por departamento, por nivel de educación de la madre y por quintil de riqueza. En el caso del método indirecto, esas tasas fueron estimadas usando tres de las cuatro familias modelo de mortalidad de Coale y Demeny, las familias Este, Oeste y Sur³, además de obtener el tiempo al que corresponden esas estimaciones.

Para seleccionar el modelo de regresión que permita corregir el problema de sobrestimación en las tasas de mortalidad generadas con el método indirecto se calculó un conjunto de indicadores que se espera que estén altamente asociados con la tasa de mortalidad infantil estimada con el método directo. Estos indicadores son la proporción de hijos muertos en mujeres de 20-24 años de edad (D20-24) y en mujeres de 25-29 años de edad (D25-29); la mediana de años de educación en mujeres de 15 a 49 años de edad (MedEdu); el número medio de hijos por mujer en mujeres de 15-19 años (P1), en mujeres de 20-24 años (P2) y en mujeres de 25-29 años de edad (P3); las relaciones P1/P2 y P2/P3, las cuales reflejan la estructura de la fecundidad; la tasa de mortalidad infantil indirecta derivada de mujeres de 20-24 años (*TMIInd_20_24*) y la tasa de mortalidad infantil indirecta derivada de mujeres de 25-29 años de edad (*TMIInd_25_29*).

Es importante aclarar lo que se entiende por corrección de la sobrestimación. Admitiendo que las tasas de mortalidad infantil generadas con el método directo son correctas, una postura bastante razonable, la corrección de la sobrestimación significa seleccionar un modelo - entre varios modelos potenciales - de modo que las tasas de mortalidad infantil

² En el país, la encuesta de demografía y salud se realizó en seis oportunidades, la última se llevó a cabo en el año 2016. Empero, en la encuesta del año 2016 se detectó algunos problemas en los datos referidos a historia de nacimientos, razón por la que se usa la información de la penúltima encuesta nacional de demografía y salud realizada en el año 2008.

³ No se consideró la familia Norte porque su nivel de mortalidad por encima de los 50 años de edad es mucha más baja que en la infancia, una estructura definitivamente alejada de la observada en Bolivia.

Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección

predichas con el modelo seleccionado sean muy parecidas a las tasas generadas con el método directo, que son las tasas correctas. Como se mencionó, los algoritmos de *machine learning* permiten seleccionar el modelo más apropiado para realizar predicciones confiables tanto en la muestra de entrenamiento como en otras muestras de prueba o testeó.

Adicionalmente, si las covariables predictoras en el modelo seleccionado se pueden construir con información obtenida en un censo y que para su construcción no se requiere recurrir a tablas modelo de mortalidad, entonces el modelo seleccionado sería apropiado para obtener tasas de mortalidad infantil para pequeñas áreas, por ejemplo para municipios, que no estén afectados por el problema de sobrestimación.

2.4 Medidas de Sobrestimación

Para evaluar la existencia o no de sobrestimación sistemática en la tasa de mortalidad infantil obtenida con el método

indirecto se obtuvieron dos diferencias: (i) una diferencia absoluta, definida como

$$DifAbs = TMIInd_{25_29} - TMIDir$$

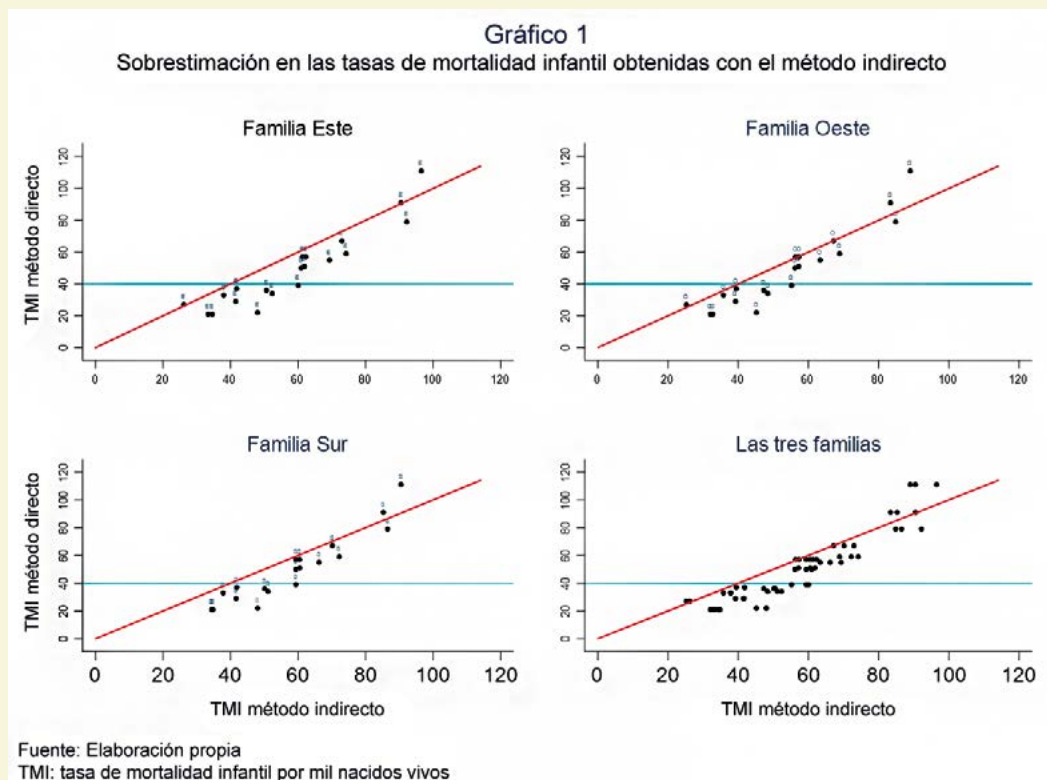
donde $TMIDir$ es la tasa de mortalidad infantil con el método directo, y (ii) una diferencia relativa, definida como

$$DifRel = [(TMIInd_{25_29} - TMIDir) / TMIDir] \times 100$$

3. RESULTADOS

3.1 Sobrestimación producida por el método indirecto

En el Gráfico 1 se exhibe la relación entre las tasas de mortalidad infantil obtenidas con el método directo y el método indirecto. El panel superior izquierdo corresponde a la familia de mortalidad de Coale y Demeny Este, el superior derecho a la familia Oeste, el inferior izquierdo a la familia Sur y en el panel inferior derecho se muestra para las tres familias (Este, Oeste y Sur) en conjunto.



Excepto en muy pocos casos, en general se observa una sobrestimación en la tasa de mortalidad infantil obtenida con el método indirecto⁴ (casi todos los puntos se encuentran a la derecha de la línea diagonal). Tal sobrestimación parece ser independiente del modelo de mortalidad elegido de las tablas modelo de mortalidad de Coale y Demeny, es decir, la sobrestimación parece producirse con cualquiera de las tres familias mortalidad. El panel inferior derecho del Gráfico 1, que incluye todos los casos, refleja el mismo patrón de sobrestimación.

Por otra parte, en el mismo Gráfico 1 se puede apreciar que la sobreestimación parece ser mayor en niveles bajos de mortalidad infantil que en niveles altos. Esto es, cuando la tasa de mortalidad infantil directa es inferior a 40 por mil nacido vivos (línea horizontal celeste), los puntos se encuentran mucho más a la derecha que cuando la tasa es superior a 40. Además, se puede observar que este segundo resultado también es independiente del modelo de mortalidad elegido.

Las magnitudes de sobrestimación se pueden observar en el Cuadro 1. Globalmente, la tasa de mortalidad infantil con el método indirecto es, en promedio, 7.8 puntos más alta que la tasa obtenida con el método directo. Esta diferencia representa una sobrestimación promedio de 25.9 por ciento. La magnitud de la sobrestimación, tanto en términos absolutos como relativos, no parece variar significativamente con la familia de mortalidad elegida, pero, se aprecian grandes diferencias en la sobreestimación cuando se considera el nivel de mortalidad infantil. En efecto, los resultados de un análisis de regresión confirman que la magnitud de la sobrestimación depende del nivel de mortalidad infantil pero no de la familia de mortalidad de Coale y Demeny elegida (Cuadro 2). Esto es, a menor nivel de mortalidad infantil, mayor es la sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto.

Cuadro 1
Promedios de sobrestimación absoluta y relativa, por nivel de mortalidad infantil y familia de mortalidad de Coale y Demeny

Nivel de mortalidad infantil	Familia			Total
	Este	Oeste	Sur	
Baja	12,7	10,2	14,1	12,3
	45,8	37,3	51,2	44,6
Alta	6,5	0,7	3,3	3,5
	12,4	3,5	8,0	7,9
Total	9,6	5,4	8,4	7,8
	29,1	20,4	28,4	25,9

Fuente: Elaboración propia

Baja: Tasa de mortalidad infantil igual o inferior a 40 por mil

Alta: Tasa de mortalidad infantil superior a 40 por mil,

⁴ Las tasas de mortalidad infantil estimadas con el método indirecto, derivadas a partir de las mujeres de 25-29 años de edad, se refieren, en promedio, a 4.4 años antes de la encuesta, es decir, en promedio corresponden al año 2003.9. En cambio, las tasas de mortalidad con el método directo se refieren al periodo 2003-2008.

Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección

Cuadro 2

**Efectos de nivel de mortalidad y familia modelo de mortalidad
sobre la sobrestimación relativa en la tasa de mortalidad infantil**

Diferencia Relativa	Coeficiente	Valor-p	Intervalo de Confianza 95%	
			LInf	LSup
TMI_directo	-0,8	0,000	-1,24	-0,41
Familia de Mortalidad				
Oeste	10,5	0,521	-22,05	43,01
Sur	-8,9	0,578	-40,72	22,94
TMI_directo#				
Familia de Mortalidad				
Oeste	-0,2	0,493	-0,80	0,39
Sur	0,0	0,987	-0,58	0,59
Constante	69,1	0,000	46,60	91,62

Fuente: Elaboración propia

TMI_directo: Tasa de mortalidad infantil con método directo

En suma, hay evidencia estadística que la tasa de mortalidad infantil obtenida con el método indirecto sobrestima sistemáticamente el nivel de la mortalidad infantil. Dicha sobrestimación, sin embargo, se produce, en similar magnitud, cualquiera sea la familia de mortalidad elegida (Este, Oeste o Sur), y es mucho mayor en niveles bajos de mortalidad infantil.

3.2 Corrección de la Sobrestimación

3.2.1 Algunas asociaciones importantes

Hay una estrecha relación, en forma de parábola, entre las tasas de mortalidad calculadas con los métodos directo e indirecto⁵ (Gráfico 2). Esta relación parabólica se observa para cada una de las tres familias de mortalidad (Este, Oeste y Sur) así como para el conjunto de datos. La misma relación, incluso más fuerte, también se observa entre las tasas de mortalidad infantil calculadas con el método directo y las proporciones de hijos muertos para mujeres de 25-29 años de edad (Gráfico 3). En efecto, con base en un modelo de regresión parabólico, la tasa de mortalidad infantil indirecta explica

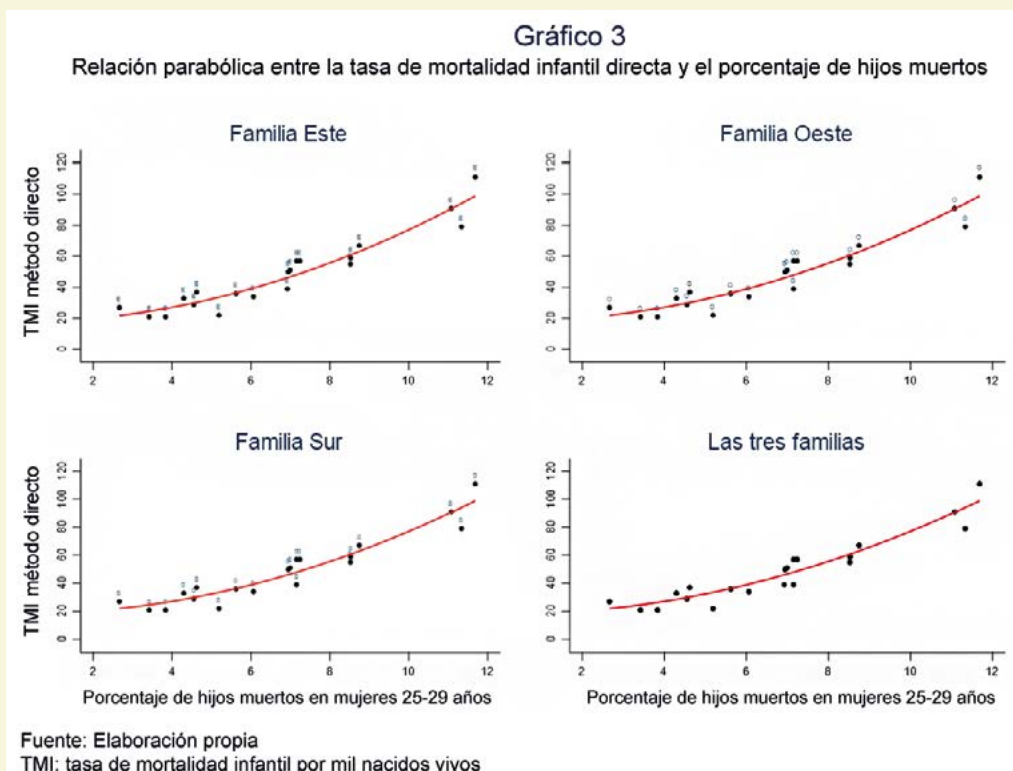
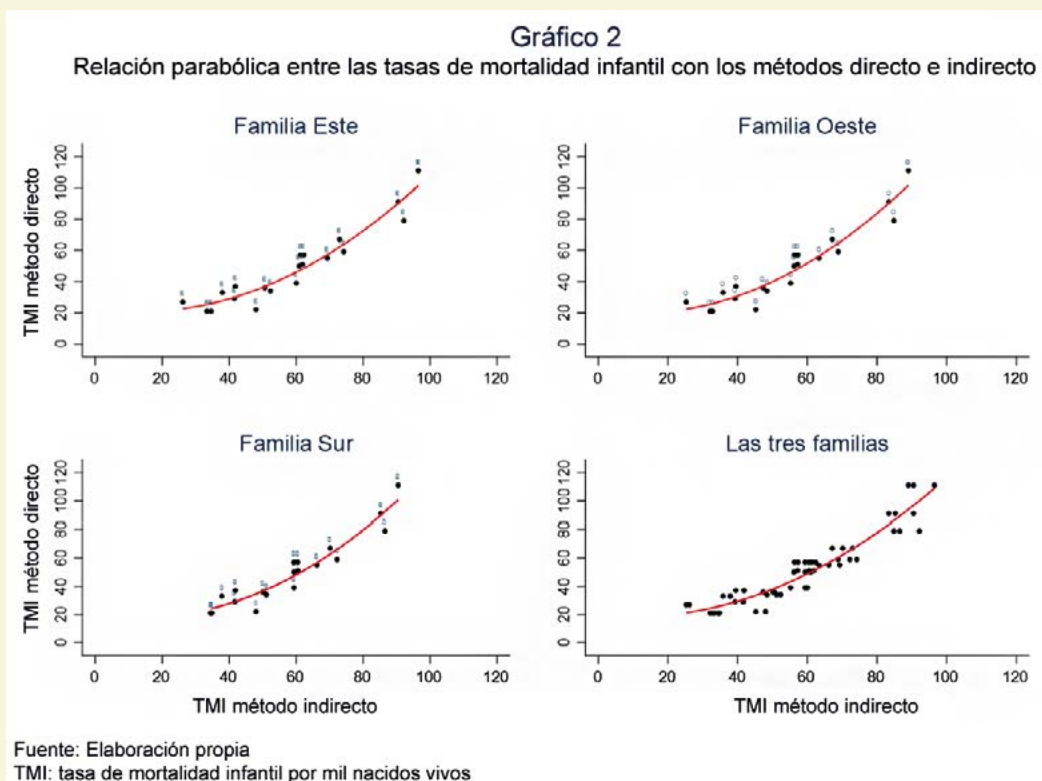
el 90.1% de la variabilidad en la tasa de mortalidad directa, mientras la proporción de hijos muertos para mujeres de 25-29 años de edad explica un 91.3% de la variabilidad en la tasa de mortalidad directa, un poco más que en el primer caso.

Además de esta leve ventaja de la proporción de hijos muertos para mujeres de 25-29 años con relación a la tasa de mortalidad indirecta, el primer indicador tiene otra ventaja más importante para fines de corrección de la sobrestimación. Resulta que para calcular la proporción de hijos muertos en mujeres de 25-29 años no se requiere adoptar ningún modelo de mortalidad, en cambio, para calcular la tasa de mortalidad infantil por el método indirecto inevitablemente se debe elegir algún modelo de mortalidad, sea de las tablas modelo de mortalidad de Coale y Demeny o de otra fuente. Otras dos ventajas de la proporción de hijos muertos en mujeres de 25-29 años son la simplicidad de su cálculo - requiere una simple división - y que se lo puede obtener con información contenida en un censo de población y vivienda.

⁵ Se recomienda hacer uso de la tasa de mortalidad infantil proveniente de mujeres de 25-29 años de edad debido a que los hijos de madres de 15-19 y 20-24 años tienen mayor riesgo de muerte.

Lo observado arriba sugiere que la proporción de hijos muertos en mujeres de 25-29 años de edad podría ser un indicador apropiado para corregir la

sobrestimación en la tasa de mortalidad infantil ocasionada con el método indirecto.



Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección

3.2.2 Corrección de la sobrestimación mediante procedimientos de *machine learning*

Para corregir el problema de sobrestimación en la tasa de mortalidad infantil causada por el método indirecto, la estrategia adoptada en esta investigación es seleccionar un modelo de regresión penalizado de modo que las tasas de mortalidad infantil predichas con ese modelo seleccionado sean muy parecidas a las tasas generadas con el método directo, tasas aceptadas como las correctas.

Un criterio importante para seleccionar el modelo más apropiado es su capacidad de generar predicciones confiables en otras muestras. Con tal propósito, la muestra total de 60 observaciones se dividió aleatoriamente en tres submuestras de igual tamaño. En terminología de *machine learning*, la primera submuestra es la de entrenamiento o *training* y las otras dos son para testear la calidad de las predicciones.

Cuadro 3
Criterios para evaluar la capacidad predictiva de los modelos penalizados

Regresión	Muestra	CME	R cuadrado
Clásica	Training	13.4	0.98
	Testing 1	211.1	0.19
	Testing 2	141.7	0.81
Ridge	Training	109.4	0.83
	Testing 1	87.8	0.66
	Testing 2	119.3	0.84
Lasso	Training	53.8	0.92
	Testing 1	44.0	0.83
	Testing 2	58.7	0.92
Red elástica	Training	53.8	0.92
	Testing 1	44.0	0.83
	Testing 2	58.7	0.92

Fuente: Elaboración propia
CME: Cuadrado medio del error

Para evaluar el ajuste del modelo seleccionado y la calidad de sus predicciones se consideraron el coeficiente de determinación y el cuadrado medio del error, respectivamente. Los resultados se muestran en el Cuadro 3.

De los cuatro modelos de regresión, los modelos lasso y red elástica tienen el mejor desempeño y en la misma magnitud. En efecto, ambos modelos producen menor error cuadrático medio cuando se usan para hacer predicciones en otras muestras (como en las dos submuestras de testeo), además de poseer los coeficientes de determinación más altos. En consecuencia, se opta por elegir el modelo de regresión lasso. Cabe observar que el modelo de regresión clásico, como se advirtió, tiene un pobre desempeño para realizar predicciones en otras muestras.

Es importante destacar que el modelo seleccionado con regresión lasso incluye una única covariable predictora, el cuadrado de la proporción de hijos muertos en mujeres de 25-29 años de edad $((D_{25-29})^2)$. Para fines de predicción, sin embargo, se adiciona al modelo la misma variable predictora elevada a la potencia uno (D_{25-29}) . Esta incorporación no afecta las predicciones.

El Cuadro 4 presenta un resumen de las tasas de mortalidad infantil predichas con el modelo seleccionado, es decir, las tasas de mortalidad infantil predichas a partir de la proporción de hijos muertos en mujeres de 25-29 años de edad. Con fines de comparación, en el mismo Cuadro 4 se incluye un resumen de las tasas de mortalidad infantil obtenidas con el método directo, consideradas en este documento y por el programa DHS como las correctas, además de un resumen de las tasas de mortalidad infantil calculadas con el método indirecto, las cuales tienen el problema de sobrestimación sistemática.

Cuadro 4
Resumen de las tasas de mortalidad infantil correctas, de tasas predichas con el modelo seleccionado y de las tasas sobrestimadas

Variable	Media	Desviación estandar	Mínimo	Máximo
TMI método directo	48,8	23,8	21,0	111,0
TMI predicha con lasso	48,8	22,9	23,5	103,1
TMI método indirecto	57,0	18,3	25,2	96,5

Fuente: Elaboración propia
 TMI: Tasa de mortalidad infantil por mil vivos

Se puede observar que, en promedio, las tasas de mortalidad infantil predichas con el modelo de regresión coinciden con las tasas correctas (48.8 por mil nacidos vivos), mientras que el promedio de las tasas de mortalidad infantil obtenidas con el método indirecto (57.0) es muy superior al promedio de las otras dos tasas (48.8). Lo mismo sucede con la desviación estándar. La dispersión de las tasas predichas y de las tasas correctas son similares (entorno de 23 puntos). Igualmente, los valores extremos de las tasas predichas están más próximos a los correspondientes valores extremos de las tasas correctas que los valores extremos de las tasas sobrestimadas.

Estos resultados sugieren que las tasas de mortalidad predichas con el modelo de regresión lasso, en promedio, se aproximan razonablemente a las tasas de mortalidad infantil correctas, a aquéllas obtenidas con el método directo. Un resultado deseado.

El Grafico 4 ilustra los anteriores resultados. La parte izquierda del gráfico muestra la relación entre las tasas obtenidas con el método directo y las estimadas con el método indirecto. Los puntos a la derecha de la línea diagonal reflejan la sobrestimación sistemática producida por el método indirecto. Esa fue la situación inicial. En cambio, la parte derecha del gráfico muestra la relación entre las tasas obtenidas con el

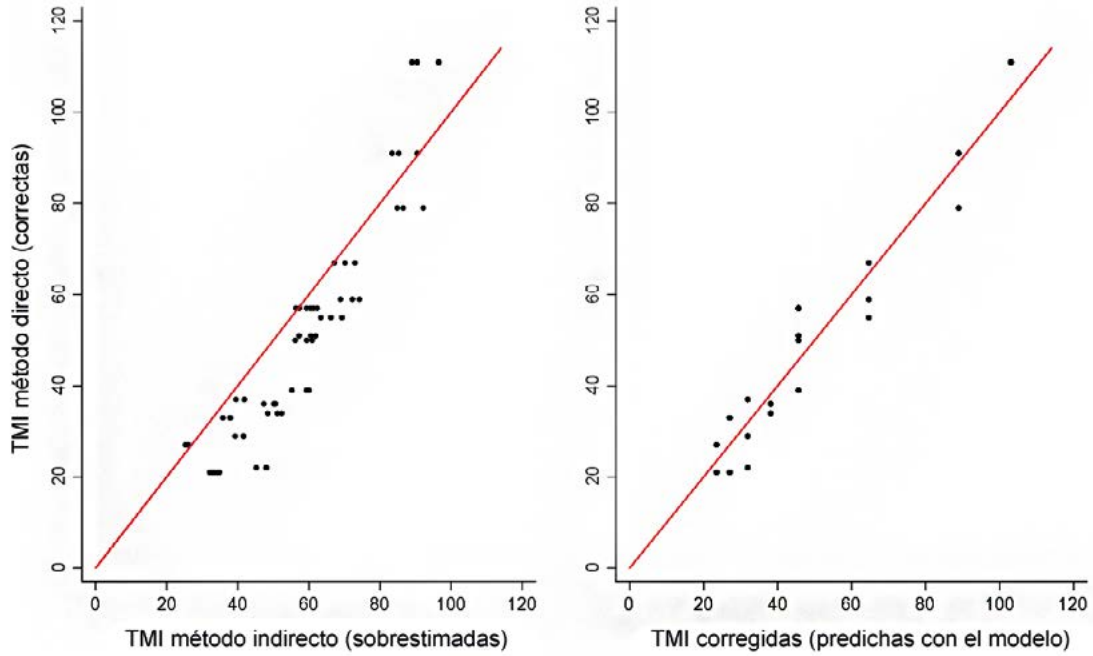
método directo y las tasas predichas con el modelo de regresión lasso. Idealmente, todos los puntos debieran estar sobre o muy próximos a la línea diagonal, lo que habría sugerido que el modelo de regresión genera muy buenas predicciones de las tasas, es decir, habría permitido eliminar totalmente el problema de sobrestimación. Las tasas de mortalidad infantil predichas, sin embargo, son razonables si se toma en cuenta que la única covariable predictora en el modelo de regresión lasso es la proporción de hijos muertos para mujeres de 25-29 años de edad, un indicador extremadamente simple de calcular y que no requiere el uso de tablas modelo de mortalidad, como las de Coale y Demeny.

El Gráfico 5 también ilustra los anteriores resultados desde otro punto de vista. Además de exhibir la relación curvilínea entre las tasas de mortalidad infantil calculadas con el método directo y la proporción de hijos muertos para mujeres de 25-29 años de edad, el gráfico también muestra la banda de confianza 99 por ciento (área sombreada) para las tasas predichas (línea continua dentro del área sombreada). Idealmente, todos los puntos debieran caer dentro de la banda de confianza, lo que habría sugerido que la banda de confianza captura todos los valores correctos de las tasas de mortalidad infantil y, consecuentemente, que el modelo de predicción es muy bueno. Empero, los valores de algunas tasas de mortalidad infantil obtenidas con el método directo se encuentran fuera de la banda de confianza, lo que sugiere que para esos casos el modelo de predicción no tiene buen desempeño. No obstante, la simplicidad del modelo predictivo es lo destacable. Podría explorarse, en futuras investigaciones, otros modelos y otras variables predictivas.

Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección

Gráfico 4

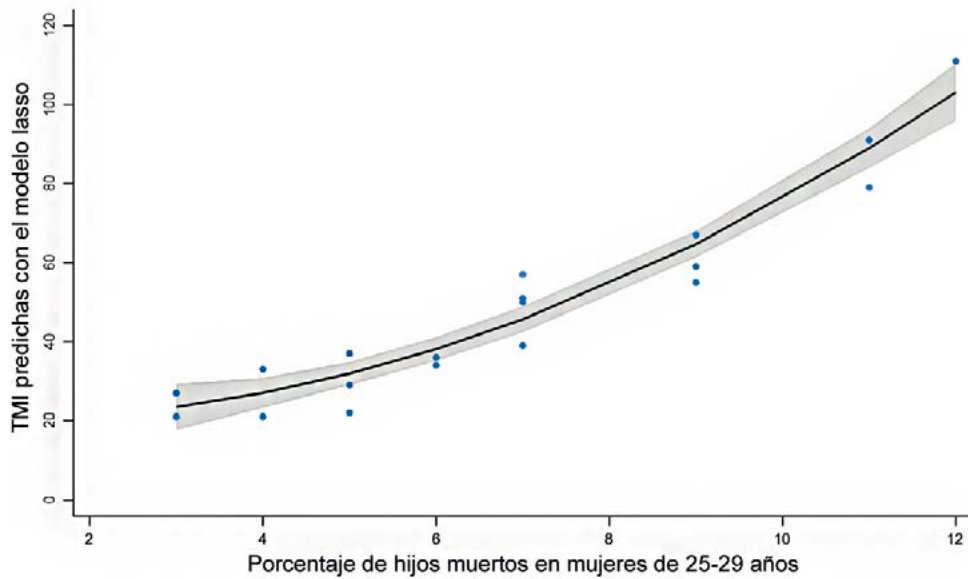
Relación de las tasas de mortalidad infantil correctas con las tasas sobrestimadas y con las tasas corregidas



Fuente: Elaboración propia
TMI: tasa de mortalidad infantil por mil nacidos vivos

Gráfico 5

Calidad predictiva del modelo de regresión lasso



Fuente: Elaboración propia
TMI: tasa de mortalidad infantil por mil nacidos vivos

4. CONCLUSIÓN

Cinco conclusiones con relación a los objetivos de esta investigación:

(i) En el caso boliviano, específicamente con datos de la ENDSA 2008, se pudo evidenciar que las tasas de mortalidad infantil obtenidas con el método indirecto están sobrestimadas sistemáticamente respecto de las tasas de mortalidad infantil obtenidas con el método directo, consideradas como las correctas. En promedio, hay un 25.9 por ciento de sobrestimación,

(ii) La sobrestimación sistemática se produce con cualquiera de las tres familias de mortalidad - Este, Oeste o Sur - de las tablas modelo de mortalidad de Coale y Demeny y en similares magnitudes.

(iii) La sobrestimación varía con el verdadero nivel de la mortalidad infantil, esto es, mayor sobrestimación a menor nivel de la mortalidad infantil. Para tasas de mortalidad infantil superiores a 40 por mil, la sobrestimación promedio es 8 por ciento; mientras para tasas de mortalidad infantil inferiores a 40 por mil, la sobrestimación promedio asciende a 45 por ciento. Este es un problema serio cuando se trata de evaluar los progresos en materia de mortalidad en la infancia, particularmente cuando se evalúan tendencias y diferencias entre municipios, puesto que muchos municipios del país han alcanzado niveles de mortalidad relativamente bajos, en cuyo caso el método indirecto conduce a tasas de mortalidad infantil demasiado sobrestimadas.

(iv) Entre los modelos de regresión analizados, el modelo de regresión lasso es el más apropiado para predecir tasas de mortalidad infantil que se asemejen

a las tasas correctas, a las obtenidas con el método directo. Es destacable que el modelo de regresión lasso seleccionado incluye sólo una covariable predictora, la proporción de hijos muertos para mujeres de 25-29 años de edad, un indicador muy simple, rápido de calcular y que no requiere de modelos de mortalidad, como los de Coale y Demeny.

(v) Se sugiere seguir explorando otros modelos y otras variables predictivas para obtener predicciones de tasas de mortalidad infantil que no estén afectadas por el problema de sobrestimación ocasionado por el método indirecto.

5. DISCUSIÓN

En general, tanto en el ámbito nacional como en el internacional, son pocos los estudios que mencionan, de manera explícita o implícita, que el método indirecto para estimar la tasa de mortalidad infantil conduce sistemáticamente a tasas sobrestimadas. Algunos de esos pocos estudios expresan, por ejemplo, que “las estimaciones de mortalidad infantil a partir del método indirecto son generalmente más altas que las estimaciones de mortalidad a partir de la historia de nacimientos...Además, [esas estimaciones] usan modelos de mortalidad que pueden o no aplicarse al país...” (Ministry of Finance and Economic Planning of Uganda, 1995); que “este método [indirecto] puede resultar en una sobrestimación de la mortalidad en la niñez” (Mokhayeri Y. et.al., 2020); que “estimaciones de la mortalidad infantil... son especialmente sensibles a la elección del modelo [de mortalidad] y pueden estar severamente sesgados cuando el modelo usado se desvía marcadamente del patrón de mortalidad prevalente en la población bajo estudio” (United Nations, 1990); o que “el nivel de las tasas de mortalidad observadas a

Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección

partir de los censos, se encuentran por encima de las tasas que se obtienen con el resto de las fuentes de información (Encuestas de Demografía y salud y Encuestas de Hogares del INE)” (UDAPE, 2018).

Son mucho más escasos, sin embargo, los estudios que evalúan la magnitud de la sobrestimación en las tasas de mortalidad infantil obtenidas con el método indirecto y que también realizan la corrección de la sobrestimación, además de una evaluación de la calidad de esa corrección. En el ámbito nacional, UDAPE estimó tasas ajustadas de mortalidad infantil y en la niñez para los municipios del país, para el año 2016, a partir de la información de los censos nacionales de población y vivienda de los años 1992, 2001 y 2012. Para el ajuste de las tasas municipales se construyó un factor de ajuste departamental calculado a partir de las diferencias observadas entre las estimaciones departamentales de mortalidad utilizando todas las fuentes de información [censos, ENDSAS y encuestas de hogares] y las estimaciones departamentales utilizando

sólo la información de los censos ... Cada factor de ajuste departamental, se aplicó, posteriormente, a las tasas de mortalidad municipales estimadas para el año 2016 (UDAPE, 2018). Este procedimiento, no suficientemente detallado, no contempla una evaluación de la calidad del ajuste realizado a las tasas de mortalidad estimadas por municipio, un aspecto sin duda muy importante para determinar si los ajustes introducidos son o no satisfactorios.

Debido a que muchos países, como en Bolivia, aún requieren del método indirecto principalmente para estimar tasas de mortalidad infantil para pequeñas áreas, como son los municipios, es necesario seguir explorando procedimientos que permitan obtener tasas de mortalidad infantil, para esas pequeñas áreas, que no estén afectadas por el problema de sobrestimación.

CONFLICTO DE INTERESES

El autor declara que no hay conflicto de intereses con respecto a la publicación de este documento.

REFERENCIA BIBLIOGRÁFICA

- Barber D. (2015). Bayesian Reasoning and Machine Learning. www.cs.ucl.ac.uk/staff/D.Barber/brml
- Behm H. y Rosero L. (1977). La mortalidad en los primeros años de vida en Países de América Latina: Ecuador 1969-1970, Centro Latinoamericano de Demografía. San José, Costa Rica.
- Brass W. (1974). Métodos para estimar la fecundidad y la mortalidad en poblaciones con datos limitados. CELADE, Serie E, No 14.
- Brass W. and Coale A. J. (1968). Methods of Analysis and Estimation, en Brass W., Coale A. J., Demeny P., Heisel D., Lorimer F., Romaniuk A. and de Wale E. The Demography of Tropical Africa, pp. 88-150. Princeton, NJ: Princeton University Press.
- Coa R. y Ochoa L. (2009). Bolivia: Encuesta Nacional de Demografía y Salud. Informe final. Ministerio de Salud y Deportes, Programa Reforma de Salud e Instituto Nacional de Estadística.
- Coale A. J. and Demeny P. (1983). Regional Model Life Tables an Stable Populations. Academic Press.

- Coale A. and Trussell J. (1974). Model fertility schedules: variations in the age structure of childbearing in human populations. *Population Index*, Vol. 40, No. 2, pp. 185-258.
- Coale A. J. and Trussell J. (1978). Estimating the Time to which Brass Estimates Apply, en Preston S. H. & Palloni A., *Fine-tuning Brass-type Mortality Estimates with Data on Ages of Children Surviving*, pp. 87-89. *Population Bulletin of the United Nations*, núm. 10-1977.
- Feeney G. (1983). Estimación de la mortalidad infantil y de la niñez en condiciones de mortalidad variable. CELADE, Serie D, No 1034.
- Hastie T., Tibshirani R. and Friedman J. (2008). *The elements of statistical learning. Data mining, inference and prediction*. Springer, 2nd Ed.
- INE. (2021). *Estimaciones y proyecciones de población de Bolivia, departamentos y municipios. Revisión 2020*. La Paz, Bolivia.
- Ministry of Finance and Economic Planning of Uganda. (1995). *Uganda: Demographic and health survey. Macro International Inc.* <https://www.dhsprogram.com/pubs/pdf/FR69/FR69.pdf>
- Mokhayeri Y., Mohammad S., Rafiei E., Asadgol Z. and Hashemi S. (2020). Indirect estimation of child mortality using 2011 census data in the Islamic Republic of Iran. *EMHJ*, vol. 26 No. 2.
- Pullum T., Assaf S. and Staveteig S. (2017). Comparisons of DHS Estimates of Fertility and Mortality with Other Estimates. *DHS Methodological Reports No. 21*.
- Shalev S. and Ben S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Somoza J. L. (1980). *Illustrative Analysis: Infant and Child Mortality in Colombia*. En WFS Scientific Report no. 10. Voorburg, Netherlands: International Statistical Institute.
- Sullivan, J. (1972). "Models for the estimation of the probability of dying between birth and exact ages of early childhood". *Population Studies*, vol.26, núm.1, pp.79-97.
- Trussell J. (1975). A re-estimation of the multiplying factors for the Brass technique for determining childhood survivorship rates. *Population Investigation Committee, Population Studies*, Vol. 29, No. 1, pp. 97-107.
- UDAPE. (2018). *La Mortalidad Infantil y en la Niñez en Bolivia: Estimaciones por Departamento y Municipio*. La Paz, Bolivia.
- UNICEF, WHO, The World Bank and UN Population Division. (2007). *Levels and Trends of Child Mortality in 2006: Estimates developed by the Inter-agency Group for Child Mortality Estimation*. New York, 2007.
- United Nations. (1990). *Step-by-step guide to the estimation of child mortality*. *Population studies*, No 107.
- United Nations (2020). *Objetivos y metas de desarrollo sostenible*. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>