

# **BOLIVIA: DISCREPANCIA ENTRE LAS TASAS GLOBALES DE FECUNDIDAD ESTIMADAS CON LOS MÉTODOS DIRECTO E INDIRECTO. UN MODELO PARA LA CORRECCIÓN**

## **BOLIVIA: DISCREPANCY BETWEEN TOTAL FERTILITY RATES ESTIMATED WITH DIRECT AND INDIRECT METHODS. A MODEL FOR CORRECTION**

Ramiro Coa Clemente<sup>1</sup>

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

✉ [clementeco@gmail.com](mailto:clementeco@gmail.com)

Patricia Loza Cruz<sup>2</sup>

Instituto Nacional de Estadística, La Paz-Bolivia

✉ [lcruzpatricia@gmail.com](mailto:lcruzpatricia@gmail.com)

Artículo recibido: 10/02/2025

Artículo aceptado: 25/03/2025

### **RESUMEN**

*Esta investigación tiene dos objetivos: por una parte, verificar si las tasas globales de fecundidad estimadas con el método indirecto discrepan sistemáticamente de aquellas generadas con el método directo, considerado este último como el método correcto y asumido por el programa de encuestas demográficas y de salud y, por otra parte, corregirlas si se evidencia una discrepancia sistemática. Con base en la información proporcionada por las encuestas de demografía y salud, se constata importantes diferencias entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto. Para el ajuste o corrección de la tasa estimada con el método indirecto se eligió el modelo de regresión red elástica de entre cuatro modelos ensayados. Las tasas globales de fecundidad predichas con red elástica están mucho más próximas a las tasas correctas que aquellas derivadas con el método indirecto. Estos resultados avalan el procedimiento - empleado en este artículo - para ajustar o corregir las estimaciones indirectas del nivel de fecundidad. En consecuencia, las tasas globales de fecundidad generadas con información de un censo requieren ser corregidas, como también las generadas con información de las encuestas de hogares.*

**Palabras clave:** Tasa global de fecundidad, Método directo, Método indirecto, Aprendizaje automático.

### **ABSTRACT**

*This research has two objectives: on the one hand, to verify whether the total fertility rates estimated with the indirect method differ systematically from those generated with the direct method, the latter being considered the correct method and assumed by the demographic and health survey program, and, on the other hand, to correct them if a systematic discrepancy is evident. Based on the information provided by the*

<sup>1</sup> 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).

<sup>2</sup> Lic. Estadística. Especialidad en Evaluación de Impacto de Programas Sociales. M.Sc. Ciencia de Datos. Consultora de instituciones públicas y privadas en análisis estadístico. ORCID: [0009-0009-6633-6264](https://orcid.org/0009-0009-6633-6264)

*demographic and health surveys, there are important differences between the total fertility rates estimated with the direct and indirect methods. For the adjustment or correction of the rate estimated with the indirect method, the elastic network regression model was chosen from among four models tested. The overall fertility rates predicted with the elastic net are much closer to the correct rates than those derived with the indirect method. These results support the procedure - used in this article - to adjust or correct the indirect estimates of the fertility level. Consequently, total fertility rates generated with information from a census need to be corrected, as do those generated with information from household surveys.*

**Keywords:** Total fertility rate, Direct method, Indirect method, Machine learning

---

## 1. INTRODUCCIÓN

La fecundidad es una de las componentes demográficas más importantes en la determinación del tamaño, composición y distribución territorial de una población. Su evolución temporal define generalmente el ritmo de crecimiento de una población. En el país, similar a lo observado en otros países de la región y del mundo, el nivel de fecundidad ha descendido continuamente en las últimas décadas, afectando el tamaño y la distribución etaria de la población. Muy vinculado a ese continuo descenso de la fecundidad está el proceso de envejecimiento de la población boliviana, iniciado hace varios años.

Una de las medidas sintéticas más usadas del nivel de fecundidad es la tasa global de fecundidad, interpretada como el promedio de hijos nacidos vivos que tendrían las mujeres durante toda su vida reproductiva, si las tasas de fecundidad por edad obtenidas en la encuesta o censo se mantuviesen invariables en el tiempo y las mujeres sobrevivieran todo su periodo reproductivo (Coa y Ochoa, 2009). Esta es una medida útil para comparar los niveles de fecundidad de varias poblaciones ya que no está afectada por diferencias que pudieran existir en las estructuras etarias de sus poblaciones femeninas en edad fértil. Además, a partir de la tasa global de fecundidad se puede derivar, a grandes rasgos, una estimación de la tasa

neta de reproducción, una medida útil para evaluar el reemplazo de una generación por la siguiente<sup>3</sup>.

En el país, sólo las encuestas demográficas y de salud, las que se han realizado mínimamente cada cinco años, proporcionan información detallada para estimar la tasa global de fecundidad con un procedimiento que no requiere supuestos y ajustes, procedimiento denominado método directo de estimación. La desventaja de estas encuestas es que las estimaciones de los niveles de fecundidad con el método directo sólo pueden calcularse, con cierta confiabilidad, para unidades o categorías definidas por el diseño de la muestra, y no para unidades administrativas pequeñas o incluso medianas, como son los municipios.

Para el caso de pequeñas unidades administrativas o, en general, para poblaciones pequeñas o medianas, la tasa global de fecundidad puede estimarse sólo con información generada a partir de un censo. Esa información censal, sin embargo, no permite aplicar el método directo de estimación, considerado el método “correcto” en este trabajo; por el contrario, se debe recurrir al denominado método indirecto, método que se basa en supuestos y que requiere de algunos ajustes en el proceso de aplicación.

<sup>3</sup> Una tasa neta de reproducción igual a 1, equivalente a aproximadamente una tasa global de fecundidad de 2,1, significa que cada generación asegura completamente su reemplazo. Por el contrario, si la tasa es menor a 1, correspondiente a una tasa global de fecundidad menor a 2,1, significa que cada generación no alcanza a reemplazarse.

Entonces, una pregunta que surge es la siguiente: ¿Cuán diferentes o similares son las estimaciones de la tasa global de fecundidad obtenidas con los métodos directo e indirecto? Si ambas estimaciones son sistemática y significativamente diferentes, la siguiente pregunta es: ¿cómo corregir la estimación de la tasa global de fecundidad producida con el método indirecto? El objetivo de esta investigación es responder estas dos preguntas. El último censo realizado en Bolivia, en marzo del año 2024, contiene preguntas diseñadas para obtener estimaciones del nivel de fecundidad mediante el método indirecto. Empero, si se confirma que las estimaciones generadas con ambos métodos son sistemáticamente diferentes, será necesario corregir la estimación derivada con el método indirecto a fin de disponer de estimaciones que permitan una buena comparación tanto entre los niveles de fecundidad de pequeñas poblaciones como entre los niveles de fecundidad de pequeñas y grandes poblaciones.

Luego de esta sección introductoria, en la segunda sección se presentan, de manera concisa, los métodos directo e indirecto para estimar la tasa global de fecundidad, así como los modelos de regresión basados en técnicas de *machine learning* usados para predecir la tasa global de fecundidad corregida; en la tercera sección se evalúan las diferencias entre las tasas globales de fecundidad generadas con ambos métodos, el directo y el indirecto, también se examinan los potenciales predictores de la tasa global de fecundidad correcta y se efectúan las predicciones con el modelo de regresión seleccionado; en la cuarta sección se presentan las principales conclusiones de esta investigación; y el artículo concluye con la quinta sección en la que se presenta una breve discusión de los aspectos centrales de esta investigación.

## **2. MÉTODOS E INFORMACIÓN**

En este trabajo, la tasa global de fecundidad se la estima usando ambos métodos - el indirecto y el directo - empleando una misma fuente de información: la encuesta de demografía y salud de los años 2008 y 2016. Por otra parte, si se confirma que las tasas estimadas con el procedimiento indirecto discrepan de las tasas correctas, aquellas generadas con el método directo, entonces las estimaciones indirectas serán corregidas o ajustadas. Para su corrección se consideran algunos modelos de regresión con función de penalización, denominados también técnicas de *machine learning* o aprendizaje automático. A continuación se hace una descripción sucinta de estos métodos.

### **2.1 Métodos de Estimación de la Tasa Global de Fecundidad**

En general, son dos los métodos más usados para estimar la tasa global de fecundidad o nivel de fecundidad: los denominados directo e indirecto. Ambos hacen uso de la información sobre hijos nacidos vivos. Mientras el método indirecto fue usado extensamente antes del directo, principalmente en países con deficiencias en sus sistemas de registros de hechos vitales, el uso del método directo fue impulsado en principio por el programa de la encuesta mundial de fecundidad y posteriormente por el programa de encuestas demográficas y de salud.

#### **El Método Indirecto**

Propuesto por Brass (1964), este método hace uso de la información recopilada en encuestas y censos acerca del número de hijos nacidos vivos durante toda la vida y sobre el número de hijos nacidos vivos durante el año previo a la fecha de la encuesta o censo, además de la cantidad

de mujeres en edad fértil. Cabe notar que este método no requiere de una completa historia de nacimientos detallada con fechas de nacimientos, sexo y otra información. Las preguntas formuladas para obtener información requerida por este método son simples, razón por la que se incluyen en los censos de población y en encuestas de hogares. En las encuestas de demografía y salud, además de recolectar información sobre la historia completa de nacimientos, también se obtiene información para aplicar este procedimiento.

En términos generales, el método consiste en comparar la fecundidad de un período, normalmente el año previo a la fecha de la entrevista, con la paridez media acumulada de toda la vida. El número promedio de hijos nacidos vivos o paridez media, obtenido al dividir el número de hijos declarados por el número de mujeres, es una medida del nivel de fecundidad experimentada por una cohorte de mujeres. Una medida similar a la paridez media puede obtenerse a partir de las tasas observadas de fecundidad específicas por edad, correspondientes al año previo a la fecha de la entrevista, denominadas también tasas de fecundidad preliminares. Al acumular estas tasas observadas se obtiene la fecundidad reciente acumulada, acumulada hasta el límite superior de cada grupo de edades. Luego de realizar una corrección a fin de que la fecundidad reciente acumulada se refiera a la edad central de cada grupo de edades, ésta es comparada con la fecundidad acumulada de toda la vida.

La comparación de la paridez media acumulada con la fecundidad reciente acumulada permite ajustar el nivel de fecundidad reciente, que se supone representa la verdadera estructura de la fecundidad. Para ajustar el nivel de fecundidad reciente se recomienda usar las razones entre ambas medidas acumuladas correspondientes

a mujeres en edades reproductivas relativamente jóvenes puesto que las declaraciones de las mujeres mayores acerca de sus hijos nacidos vivos están sujetas a errores de omisión.

En este trabajo se usan dos tipos de ajuste de las tasas de fecundidad recientes. El primero, hace uso de la razón entre ambas medidas acumuladas concernientes a mujeres de entre 20 a 24 años de edad, ajuste con el que se obtiene una Primera Estimación Indirecta de la tasa global de fecundidad; mientras el segundo ajuste se refiere al uso del promedio ponderado de las razones correspondientes a mujeres de entre 20 a 24 y a mujeres de entre 25 a 29 años de edad, ajuste con el que se genera la Segunda Estimación Indirecta de la tasa global de fecundidad. Estas dos estimaciones indirectas del nivel de fecundidad son consideradas en posteriores análisis.

### **El Método Directo**

Para la estimación de la tasa global de fecundidad por el método directo se hace uso de información acerca de la historia de nacimientos de cada mujer en edad fértil. En el país, sólo en las encuestas de demografía y salud se obtiene información detallada y completa sobre la historia de nacimientos. El sistema de registro administrativo de hechos vitales aún enfrenta problemas de cobertura y de omisión, entre otros.

Con este método, la tasa global de fecundidad combina las tasas específicas de fecundidad por grupos de edades para un periodo determinado, generalmente el periodo de 1 a 36 meses antes de la encuesta. Estas tasas específicas son las razones entre el número de nacimientos ocurridos en el periodo de tiempo a mujeres del grupo de edades en el momento del nacimiento y el número de años-mujer de exposición en el mismo

## Bolivia: Discrepancia entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto. Un modelo para la corrección

periodo de tiempo de mujeres del mismo grupo de edades.

Este método, descrito en términos generales, al que se ha denominado método directo en este trabajo, es adoptado por el Programa de Encuestas Demográficas y de Salud. Además de no requerir de supuestos ni ajustes, la calidad de los datos empleados por el método es razonablemente confiable puesto que la información sobre historia de nacimientos en las encuestas de demografía y salud, particularmente en la del año 2008 (Coa y Ochoa, 2009), no está seriamente afectada por la omisión de nacimientos o por errores en las declaraciones de fechas de nacimientos. Por lo anterior, la tasa global de fecundidad estimada con el método directo es considerada correcta.

### 2.2 Modelos de Regresión para Corrección de la Discrepancia

Esta sección se basa en la síntesis acerca de modelos de regresión penalizados, expuesto en el artículo de Coa (2023).

Para corregir la discrepancia en la tasa global de fecundidad calculada mediante el método indirecto con relación a la obtenida con el método directo, en este trabajo se recurre a modelos de regresión con función de penalización, denominados también técnicas de *machine learning*. Estos modelos tienen una estructura común, descrita a continuación.

Dado el modelo de regresión

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

donde  $Y_i$  es la respuesta para la  $i$ -ésima observación;  $\beta_0$  es un coeficiente denominado intercepto;  $X_i$  es un vector de covariables correspondiente a la  $i$ -ésima observación, de dimensión  $p \times 1$ ;  $\beta$  es un vector de coeficientes asociados a las covariables, de dimensión

$p \times 1$ ; y  $\varepsilon_i$  es el error aleatorio asociado a la  $i$ -ésima observación; la función objetivo penalizada a ser minimizada es

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

donde  $(Y_i - \beta_0 - X_i^t \beta)^2$  es el cuadrado del error de predicción en la muestra,  $\lambda$  es un coeficiente de penalización general que controla el grado de penalización y  $\alpha$  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  $\varphi$  se reduce a la función 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. Este tipo de regresión no selecciona variables, por el contrario, 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  $\lambda$ , los coeficientes serán más pequeños y más robustos a la colinealidad. Consecuentemente, la regresión ridge produce estimadores puntuales que pueden ser usados para predicción.

#### Regresión lasso

Si  $\alpha = 1$  y  $\lambda > 0$ , la función  $\varphi$  se reduce a la

función penalizada lasso, función que permite obtener el vector de estimadores lasso.

A diferencia de regresión ridge, en la que se conservan en el modelo todas las variables predictoras, es decir, ninguno de los coeficientes del modelo se hace cero; en regresión lasso la penalización tiene otro efecto, elimina las variables que no agregan poder predictivo al modelo, esto es, algunos coeficientes llegan a ser iguales a cero. Por ello, tanto la regresión lasso como la regresión ridge reducen la complejidad del modelo, aunque por medios diferentes. La regresión lasso reduce el número de covariables que afectan la variable respuesta, mientras la regresión ridge reduce el peso que tiene cada covariable sobre la respuesta.

Otro efecto de la penalización, válido para regresión ridge y lasso, es evitar un posible sobreajuste del modelo. Esto es, la penalización de los coeficientes más altos mejora la generalización del modelo estimado, contribuye a que el modelo funcione bien tanto con los datos de entrenamiento como con los datos de prueba, una característica deseada para fines de esta investigación.

Por último, regresión lasso, similar a regresión ridge, también permite enmendar el problema de multicolinealidad. Si dos o más variables predictoras son altamente colineales, sus correspondientes coeficientes de mínimos cuadrados ordinarios pueden resultar en valores espuriamente altos. En este caso, las predicciones de la variable respuesta son sensibles incluso a pequeños cambios en los datos. Esto es, el modelo fue sobreajustado en los datos de entrenamiento y podría no ser apropiado para obtener predicciones precisas con otros datos. En esta situación, ambas penalizaciones, lasso y

ridge, corrigen los coeficientes altos. Lasso corrige la multicolinealidad reduciendo esos coeficientes a cero, mientras ridge sólo reduce sus valores, no los anula.

### **Regresión red elástica**

Si  $0 < \alpha < 1$  y  $\lambda > 0$ ,  $\varphi$  es la función 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  $\varphi$  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

---

<sup>4</sup> En el país, la encuesta de demografía y salud se realizó en siete oportunidades, la última fue en el año 2023, aunque no se publicó aún la base de datos de la última encuesta. Por ello, en el presente estudio se usan los datos de las encuestas de demografía y salud de 2008 y 2016.

## **Bolivia: Discrepancia entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto. Un modelo para la corrección**

---

muy pobres, produciendo considerables errores de predicción.

### **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 salud<sup>4</sup>, ya que esta fuente permite aplicar ambos métodos de estimación de la fecundidad, el directo y el indirecto, y, por tanto, permite evaluar las discrepancias entre las tasas globales de fecundidad estimadas con ambos procedimientos.

Cabe notar que la información sobre historia de nacimientos, la que es requerida por el método directo, y particularmente aquella obtenida en la encuesta de demografía y salud del año 2008, es razonablemente confiable y no está seriamente afectada por la omisión de nacimientos vivos o por errores en las declaraciones de fechas de nacimientos. El apéndice C del informe final de la encuesta de 2008 sobre calidad de la información proporciona varios resultados al respecto (Coa y Ochoa, 2009), los que evidencian la calidad aceptable de los datos. Igualmente, la información requerida por el método indirecto sobre el número de hijos nacidos vivos y el número de nacimientos en el año previo a la encuesta es bastante confiable, puesto que el cuestionario de las encuestas de demografía y salud fue diseñado apropiadamente para reducir al máximo la omisión de nacimientos.

Las tasas globales de fecundidad fueron estimadas tanto con el método directo como con el indirecto usando la información de las encuestas de demografía y salud de los años 2008 y 2016, para diferentes categorías de análisis: país, áreas de residencia urbana y rural, departamento de residencia, nivel de educación de la madre, quintil de riqueza del hogar y pertenencia étnica de la madre. Se tuvo un total de 43 estimaciones, 24

correspondientes al año 2008 y 19 para el 2016.

Además de las tasas globales de fecundidad calculadas con el método directo para cada categoría de análisis, se obtuvieron dos estimaciones de la misma tasa con el método indirecto, descritas en la sección 2.1. Adicionalmente, a partir de las tasas de fecundidad recientes específicas por edad, se obtuvo, sin realizar ningún ajuste, una tercera estimación de la tasa global de fecundidad, denominada tasa global de fecundidad preliminar carente de ajuste. Si bien esta última medida no es, en rigor, un estimador indirecto apropiado de la tasa global de fecundidad, tiene dos virtudes: por una parte, parece estar muy asociada con la tasa global de fecundidad obtenida con el método directo, considerada como la estimación correcta y, por otra parte, su cálculo es muy simple, se la obtiene con una simple división. Por estas dos características, la tasa global de fecundidad preliminar - carente de ajuste - es un potencial predictor de la tasa global de fecundidad correcta.

Es importante notar que estas tres variables (los dos estimadores indirectos y el estimador preliminar sin ningún ajuste), consideradas potenciales predictoras de la correcta tasa global de fecundidad, además de ser simples en su cálculo, pueden ser generadas con datos de un censo o de una encuesta que no recolecta información sobre historia de nacimientos, como por ejemplo la encuesta de hogares realizada periódicamente por el Instituto Nacional de Estadística. En consecuencia, el modelo de regresión que sea seleccionado permitirá predecir o estimar la tasa global de fecundidad correcta para unidades administrativas pequeñas, por ejemplo para los 343 municipios del país.

Es más, como las encuestas de demografía y salud - en las que se obtienen datos de

historia de nacimientos - se realizan cada cierto número de años y las encuestas de hogares - en las que no se recolectan datos de historia de nacimientos - se realizan cada año, entonces también se podría predecir o estimar la tasa global de fecundidad correcta para aquellos años intermedios en los que no se haya realizado la encuesta de demografía y salud.

## 2.4 Medidas de Discrepancia

Para evaluar la existencia o no de una considerable discrepancia entre la estimación de la tasa global de fecundidad obtenida con los métodos directo e indirecto se obtienen dos diferencias. Una diferencia absoluta, definida como

$$Dif.Absoluta = ETGF_i - ETGF_{directo}$$

donde  $ETGF_i$  significa estimación de la tasa global de fecundidad con el  $i$ -ésimo procedimiento indirecto,  $i = 1$  representa la primera estimación con el método indirecto,

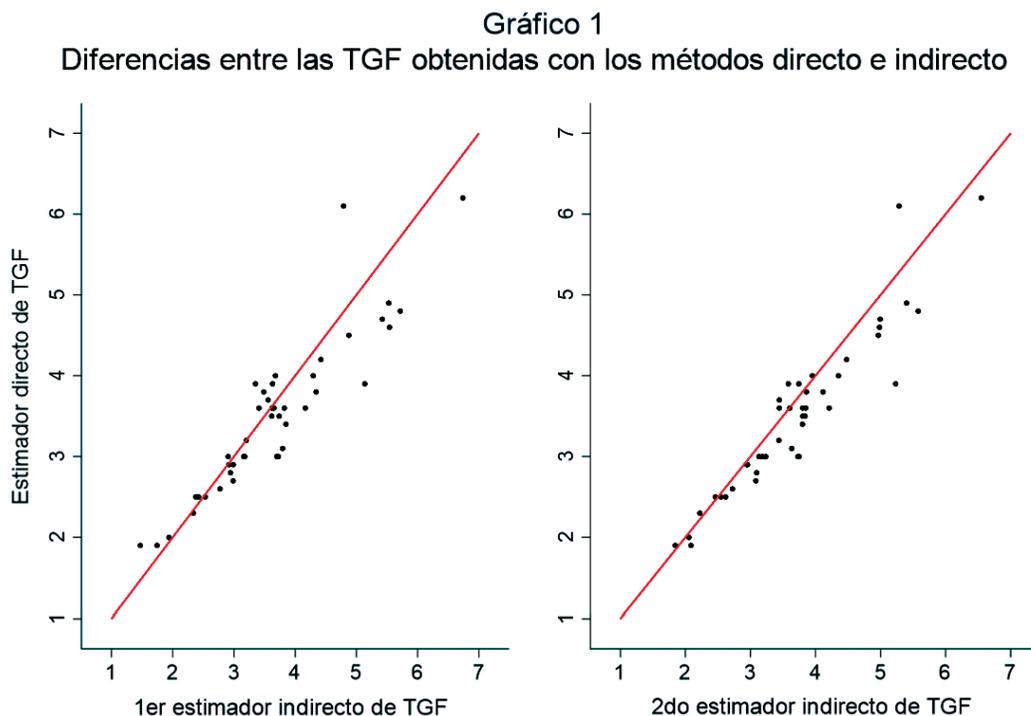
$i = 2$  se refiere a la segunda estimación con el método indirecto e  $i = 3$  a la estimación con el modelo de regresión penalizado; y  $ETGF_{directo}$  denota la estimación de la tasa global de fecundidad con el método directo; y una diferencia relativa definida como

$$Dif.Relativa = \frac{|ETGF_i - ETGF_{directo}|}{ETGF_{directo}} * 100.$$

## 3. RESULTADOS

### 3.1 Diferencias Producidas por el Método indirecto

Una primera idea acerca de las diferencias o discrepancias entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto se exhibe en el Gráfico 1. Recordemos que se tienen dos estimadores indirectos del nivel de fecundidad. En el panel izquierdo se presentan las discrepancias de la primera estimación indirecta respecto de la estimación correcta, aquella generada



Fuente: Elaboración propia  
TGF: Tasa Global de Fecundidad

## Bolivia: Discrepancia entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto. Un modelo para la corrección

Cuadro 1

Magnitud de las diferencias entre las TGF obtenidas con los métodos directo e indirecto

Diferencia	Estimación indirecta de TGF	Estadísticas resumen de magnitudes de las diferencias			
		Promedio	Desviación estándar	Mínima	Máxima
Absoluta	Primera	0,17	0,46	-1,31	1,24
	Segunda	0,22	0,35	-0,82	1,33
Relativa	Primera	9,62	7,80	0,06	31,69
	Segunda	8,51	6,95	0,11	34,13

Fuente: Elaboración propia

TGF= Tasa Global de Fecundidad

con el método directo; mientras en el panel derecho se observan las discrepancias de la segunda estimación indirecta con relación a la estimación directa. En ambos paneles se observan considerables desviaciones de las estimaciones indirectas respecto de las directas, pues casi todos los puntos están alejados de la línea diagonal, principalmente cuando la estimación indirecta es superior a 3.

La magnitud de las diferencias entre las estimaciones indirectas y directas puede observarse en el Cuadro 1. Con relación al primer estimador indirecto, sus estimaciones del nivel de fecundidad presentan importantes desviaciones, en términos absolutos, respecto de las producidas con el método directo. En un extremo, la estimación indirecta del nivel de fecundidad es 1.31 hijos menos que la estimación correcta; mientras que en el otro extremo, con la estimación indirecta se obtiene 1,24 hijos más que con la estimación correcta. Similares resultados se registran con el segundo estimador indirecto. Las considerables diferencias también se pueden apreciar en términos relativos. En promedio, se registra una desviación de 9.62% de las estimaciones indirectas con relación a las correctas, esto en el caso del

primer estimador indirecto; en tanto para el segundo estimador indirecto la desviación promedio es levemente inferior, 8.51%. También se registra que las diferencias relativas máximas, para ambas estimaciones indirectas, son altas, ascienden a más de 30%. En conjunto, estos resultados constatan importantes desviaciones de las tasas globales de fecundidad estimadas con el método indirecto respecto de las tasas correctas, aquellas estimadas con el método directo.

¿Con cuál de las dos estimaciones indirectas de la tasa global de fecundidad se producen diferencias o discrepancias más pequeñas? A primera vista, la similitud en la dispersión de los puntos en ambos paneles del Gráfico 1 sugiere que ambos indicadores indirectos de la tasa global de fecundidad producen similares discrepancias con relación a la estimación directa. Pruebas estadísticas formales confirman esta similitud. Efectivamente, pruebas t-student para la igualdad de promedios tanto de las diferencias absolutas como de las diferencias relativas no son rechazadas. Un resultado similar se obtiene con un análisis de varianza. Por tanto, las discrepancias o desviaciones de las estimaciones indirectas respecto de las

correctas son estadísticamente similares, aunque los resultados también sugieren que la segunda estimación indirecta de la fecundidad tiende a producir discrepancias levemente inferiores.

### 3.2 Un Predictor Apropriado para la Corrección de la Discrepancia

Ambos estimadores indirectos del nivel de fecundidad son potenciales predictores del correcto nivel de fecundidad. Efectivamente, estos dos indicadores están muy correlacionados con la estimación correcta de la tasa global de fecundidad, presentando una correlación de 0.91 para el primer estimador indirecto y 0.94 para el segundo (véase Cuadro 1A del anexo).

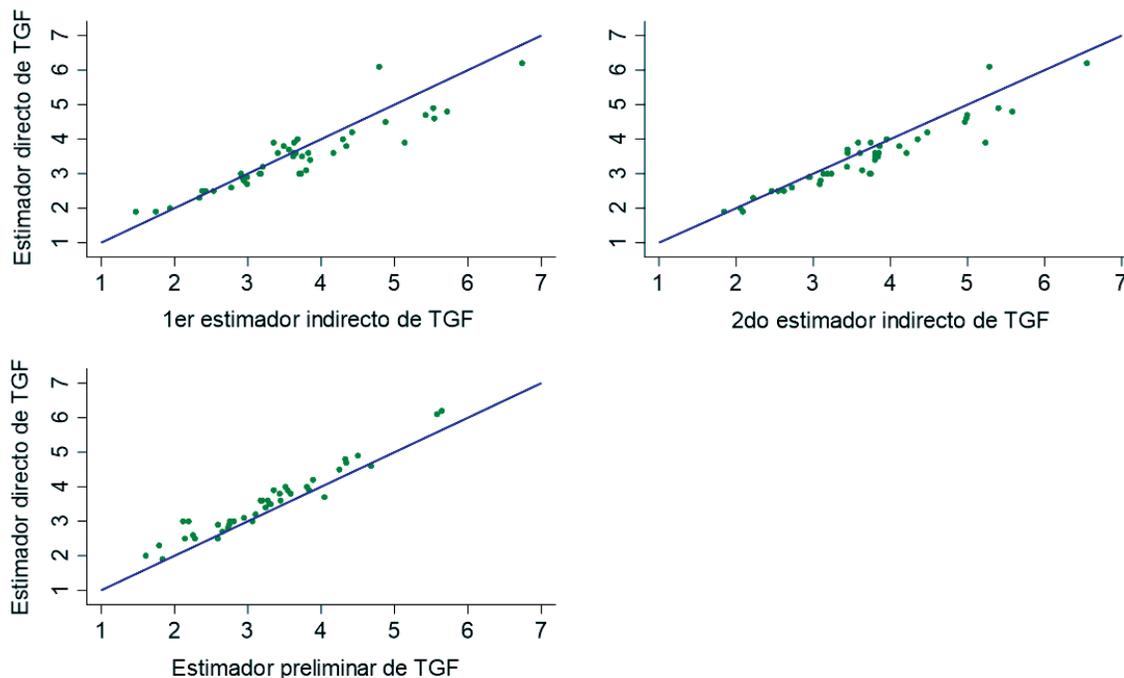
Se tiene, sin embargo, un tercer predictor potencial, el cual se encuentra más fuertemente asociado con el estimador correcto que los

dos estimadores indirectos. Ciertamente, la correlación entre los estimadores correcto y preliminar de la fecundidad es más alta (0.97). El tercer panel del Gráfico 2 devela este hecho, que el estimador correcto del nivel de fecundidad está más asociado con el estimador preliminar de la fecundidad que con los otros dos estimadores indirectos.

Además de esa asociación más estrecha, en el mismo tercer panel del Gráfico 2 se puede apreciar que casi todas las estimaciones preliminares de la tasa global de fecundidad son menores o iguales a las estimaciones correctas, sea cual sea el nivel de fecundidad (casi todos los puntos se encuentran por encima de la línea diagonal). Esto es, las estimaciones preliminares de la tasa global de fecundidad subestiman sistemáticamente la tasa correcta, independientemente del nivel de fecundidad. Con base en estas tres

Gráfico 2

Asociación del estimador directo con los estimadores indirectos y preliminar de la TGF



Fuente: Elaboración propia  
TGF: Tasa Global de Fecundidad

## Bolivia: Discrepancia entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto. Un modelo para la corrección

características del estimador preliminar - alta asociación, subestimación sistemática e independencia del nivel de fecundidad - el estimador preliminar de la tasa global de fecundidad podría ser un buen predictor de la tasa correcta. En términos gráficos, la predicción del nivel de fecundidad correcto podría consistir simplemente en desplazar levemente hacia arriba la línea diagonal en el tercer panel del Gráfico 2. La medida en que se la debe desplazar estará definida por los coeficientes del modelo de regresión penalizado que se seleccione.

### 3.3 Predicción de la Correcta Tasa Global de Fecundidad

Se evidenció en anteriores secciones que la tasa global de fecundidad obtenida con cualquiera de los dos estimadores indirectos examinados en este documento discrepa o es diferente de la estimación correcta, aquella que es producida por el método directo y que fue adoptada por el programa de encuestas de demografía y salud. En consecuencia, las estimaciones del nivel de fecundidad

generadas con información de un censo o de las encuestas de hogares, información que hace uso el método indirecto, no serán correctas. Este es un problema que se presenta cuando se estima el nivel de fecundidad para pequeñas áreas, por ejemplo para municipios.

Para subsanar este problema, la estrategia adoptada en esta investigación es seleccionar un modelo de regresión penalizado de modo que las tasas globales de fecundidad predichas con ese modelo seleccionado sean muy parecidas a las tasas generadas con el método directo, tasas aceptadas como las correctas. Con este propósito, en cada uno de los cuatro modelos ensayados se incluyeron como variables predictoras el estimador preliminar de la tasa global de fecundidad y los otros dos estimadores indirectos de la misma tasa. Los algoritmos de *machine learning* permitirán seleccionar el modelo más apropiado para realizar predicciones confiables.

Un criterio importante para seleccionar el modelo más apropiado es su capacidad de

Cuadro 2  
Criterios para evaluar la capacidad predictiva de los modelos penalizados

Regresión	Muestra	CME	R <sup>2</sup>
Clásica	Training	0,008	0,99
	Testing 1	0,115	0,89
	Testing 2	0,075	0,91
Ridge	Training	0,014	0,98
	Testing 1	0,031	0,97
	Testing 2	0,095	0,88
Lasso	Training	0,009	0,99
	Testing 1	0,073	0,93
	Testing 2	0,055	0,93
Red elástica	Training	0,010	0,99
	Testing 1	0,040	0,96
	Testing 2	0,057	0,93

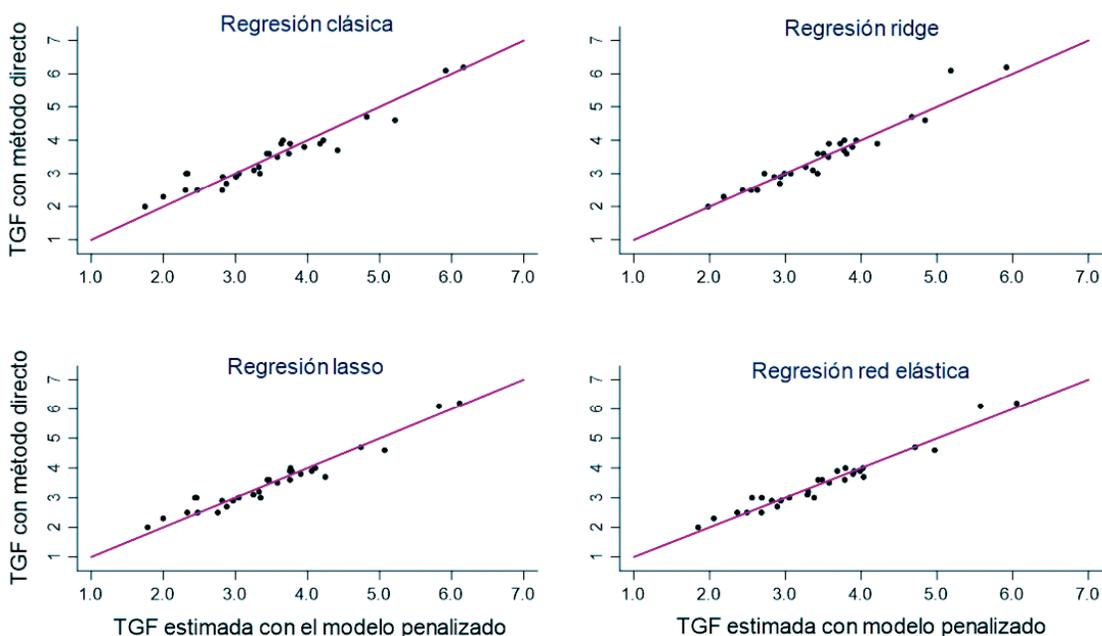
Fuente: Elaboración propia

CME: Cuadrado Medio del Error

R<sup>2</sup>: Coeficiente de Determinación

Gráfico 3

TGF estimadas con método directo comparadas con TGF estimadas con distintos modelos



Fuente: Elaboración propia  
TGF: Tasa Global de Fecundidad

generar predicciones confiables en otras muestras. Para diagnosticar este criterio, la muestra total de 43 observaciones se dividió aleatoriamente en tres submuestras de similar tamaño. En terminología de *machine learning*, la primera submuestra es la de entrenamiento o *training*, mientras las otras dos son usadas para testear la calidad de las predicciones. Con el propósito de evaluar el ajuste del modelo y la calidad de las predicciones se consideran el coeficiente de determinación y el cuadrado medio del error, respectivamente. Los valores de estas dos medidas para los cuatro modelos de regresión ensayados se muestran en el Cuadro 2.

Con base en estas dos medidas, el modelo de regresión red elástica presenta un mejor desempeño. En efecto, este modelo reporta el menor error cuadrático medio de las predicciones y un mejor ajuste en las muestras de prueba (Cuadro 2). Por tanto, para estimar

la tasa global de fecundidad correcta se elige el modelo de regresión red elástica. Cabe notar, sin embargo, que el desempeño de los otros tres modelos no es malo, incluso el de regresión clásica. Este hecho se debe principalmente a que el estimador preliminar de la tasa global de fecundidad está fuertemente correlacionado con la tasa correcta (tercer panel del Gráfico 2).

El mejor desempeño del modelo red elástica también puede apreciarse en términos gráficos. En el Gráfico 3 se observa que las predicciones de la tasa global de fecundidad generadas con regresión red elástica se encuentran más próximas a las estimaciones correctas del nivel de fecundidad - los puntos están más próximo a la línea diagonal - que las predicciones obtenidas con las otras tres regresiones, si bien regresión lasso también tiene un buen desempeño.

## Bolivia: Discrepancia entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto. Un modelo para la corrección

Cuadro 3

Magnitud de las diferencias del estimador directo de la TGF con los estimadores indirectos y con los estimadores de regresión red elástica (se consideran sólo estimaciones de la TGF para los grupos de *testing* 1 y 2)

Diferencia	Diferencia del estimador directo con:	Estadísticas resumen de magnitudes de las diferencias			
		Promedio	Desviación estándar	Mínima	Máxima
Absoluta	1er estimador indirecto	0,17	0,50	-1,31	1,24
	2do estimador indirecto	0,20	0,39	-0,82	1,33
	regresión red elástica	-0,01	0,22	-0,53	0,38
Relativa	1er estimador indirecto	9,78	8,74	0,06	31,69
	2do estimador indirecto	8,90	8,02	0,11	34,13
	regresión red elástica	5,20	3,88	0,19	14,69

Fuente: Elaboración propia

TGF= Tasa Global de Fecundidad

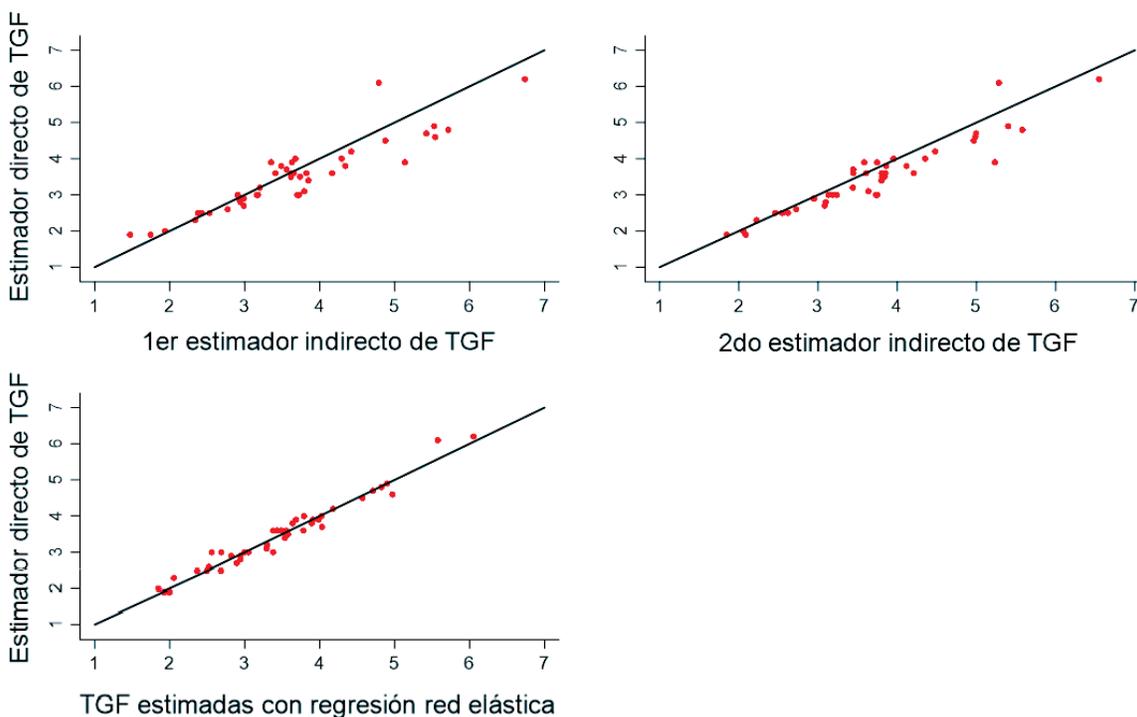
Resultados de la medida en la que el modelo de red elástica permite mejorar o corregir la estimación de la tasa global de fecundidad se exhiben en el Cuadro 3. Para generar esos resultados se han considerado sólo las dos submuestras de prueba o testeo, puesto que interesa principalmente la capacidad predictiva del modelo en muestras distintas a la de entrenamiento. Se presentan las diferencias absolutas y relativas entre las predicciones de la tasa global de fecundidad basadas en regresión red elástica y las estimaciones correctas de la misma tasa. Con fines de comparación también se presentan las diferencias entre las estimaciones indirectas y las correctas.

Entre los resultados expuestos en el Cuadro 3 se tiene que las desviaciones absolutas de las tasas globales de fecundidad estimadas con regresión red elástica respecto de las estimaciones correctas son menos variables o menos dispersas (0.22) que aquellas estimaciones generadas con el método indirecto (0.50 para el primer estimador indirecto y 0.39 para el segundo). En efecto,

con red elástica, en un extremo se estima una tasa de 0.53 hijos menos que con el método correcto y, en el otro extremo, la tasa estimada es 0.38 hijos más que la tasa correcta. Estas desviaciones o valores extremos son mucho más grandes para las dos estimaciones indirectas. Estos resultados constatan que con regresión red elástica se logró una mejora en la estimación de la tasa global de fecundidad. En términos de diferencias relativas también se observan importantes mejoras. La diferencia relativa promedio de alrededor de 9% para las estimaciones indirectas se redujo a 5.2% para las estimaciones generadas con red elástica. Igualmente, se registra una considerable reducción en las diferencias extremas. La diferencia relativa máxima de al menos 31.69% para el caso de las estimaciones indirectas se redujo a 14.69% para el caso de la estimación con red elástica. Estos resultados también constatan que las estimaciones de la tasa global de fecundidad obtenidas con el modelo de regresión red elástica están más próximas de las tasas correctas que aquéllas derivadas con el método indirecto.

Gráfico 4

Comparación del estimador directo con estimadores indirectos y con estimadores de red elástica



Fuente: Elaboración propia  
 TGF: Tasa Global de Fecundidad

Esta mejora puede también apreciarse en términos gráficos. En los dos primeros paneles del Gráfico 4 se reproducen las diferencias o discrepancias entre las estimaciones indirectas de la tasa global de fecundidad y las correctas estimaciones de la misma tasa. Se podría decir que esas diferencias representan la situación inicial. En cambio, el tercer panel exhibe las diferencias entre las estimaciones de las tasas obtenidas con red elástica y las correctas estimaciones. Claramente las estimaciones con red elástica están mucho más próximas a las estimaciones correctas (todos los puntos están muy próximos de la línea diagonal).

#### 4. CONCLUSIONES

Se constató importantes diferencias entre las tasas globales de fecundidad estimadas con los métodos indirecto y directo, este último

considerado como el método correcto, el cual fue asumido por el programa de encuestas demográficas y de salud. Esas diferencias se agudizan cuando el nivel de fecundidad es más alto, superior a 3 hijos, llegando a registrarse una discrepancia de hasta 34 por ciento. En consecuencia, la tasa global de fecundidad estimada con el método indirecto requiere de un ajuste o corrección a fin de lograr una estimación igual o al menos muy próxima a la estimación correcta de la tasa.

De los cuatro modelos de regresión ensayados para obtener estimaciones de la tasa global de fecundidad de modo que estén lo más próximo posible a las estimaciones correctas, se eligió el modelo de red elástica, un modelo que, con el uso de algoritmos de *machine learning*, reduce más que los otros tres modelos los errores de predicción de la tasa

## **Bolivia: Discrepancia entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto. Un modelo para la corrección**

---

global de fecundidad en muestras diferentes a la de entrenamiento.

Las estimaciones de la tasa global de fecundidad derivadas con regresión red elástica están mucho más próximas de las tasas correctas que aquellas derivadas con el método indirecto. Ciertamente, la diferencia promedio de alrededor de 9 por ciento entre las estimaciones indirectas y la estimación correcta de la tasa global de fecundidad se redujo a una diferencia promedio de 5.2 por ciento entre las generadas con red elástica y las estimaciones correctas. Igualmente, se registró una considerable reducción en las diferencias extremas con relación a la estimación correcta; de al menos 31.69% para el caso de las estimaciones indirectas se redujo a 14.69% para el caso de la estimación con red elástica.

### **5. DISCUSIÓN**

El método directo para estimar la tasa global de fecundidad, considerado como el correcto, sólo es aplicable con información generada en las encuestas demográficas y de salud y, por tanto, sólo permite estimaciones confiables para dominios definidos en el diseño de la muestra. En consecuencia, no es posible aplicar este método para áreas menores, como son los municipios. Para estimar el nivel de fecundidad de municipios se debe necesariamente recurrir a la información

generada en un censo, información que sólo permite aplicar el método indirecto, el cual, como se demostró en este trabajo, difiere considerablemente de la estimación correcta del nivel de fecundidad. La estimación correcta, o una muy aproximada a la correcta, puede derivarse a partir del modelo de red elástica estimado en este artículo.

A causa de que las encuestas demográficas y de salud se llevan a cabo cada cierto número de años, tampoco es posible disponer de estimaciones correctas del nivel de fecundidad para años intermedios. Con la finalidad de generar estimaciones confiables para esos años intermedios, se puede usar la información proporcionada por las encuestas de hogares, encuestas que son realizadas anualmente por el Instituto Nacional de Estadística y que proporcionan información para aplicar el método indirecto de estimación de la tasa global de fecundidad. Con esta estimación indirecta y el estimador preliminar de la fecundidad, se puede derivar la estimación ajustada o corregida de la tasa haciendo uso del modelo estimado de red elástica.

### **Conflicto de Intereses**

Los autores declaran 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](http://www.cs.ucl.ac.uk/staff/D.Barber/brml)
- Brass W. (1964). Uses of census or survey data for the estimation of vital rates. Economic Commission for Africa.
- 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.
- Coa R. (2023). Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección. Revista varianza No 22
- 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.
- Hastie T., Tibshirani R. and Friedman J. (2008). The elements of statistical learning. Data mining, inference and prediction. Springer, 2nd Ed.
- 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.

**Bolivia: Discrepancia entre las tasas globales de fecundidad estimadas con los métodos directo e indirecto. Un modelo para la corrección**

**ANEXO**

Cuadro 1A

Matriz de correlaciones entre estimadores de la TGF

	Estimador directo de TGF	1er estimador indirecto de TGF	2do estimador indirecto de TGF	Estimador preliminar de TGF
Estimador directo de TGF	1,00			
1er estimador indirecto de TGF	0,91	1,00		
2do estimador indirecto de TGF	0,94	0,98	1,00	
Estimador preliminar de TGF	0,97	0,89	0,90	1,00

Fuente: Elaboración propia

TGF: Tasa Global de Fecundidad