



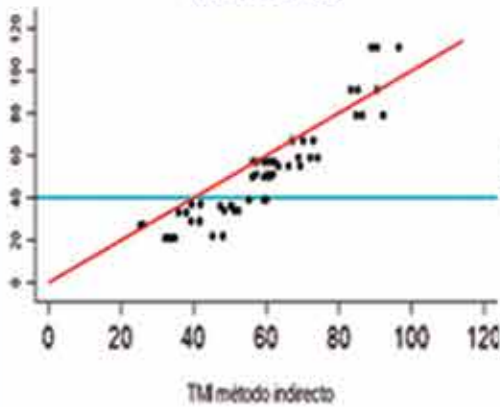
Universidad Mayor de San Andrés  
 Facultad de Ciencias Puras y Naturales  
 Carrera de Estadística

# Varianza No 22

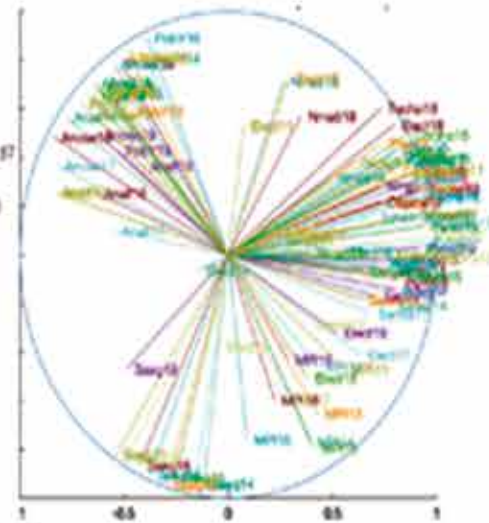
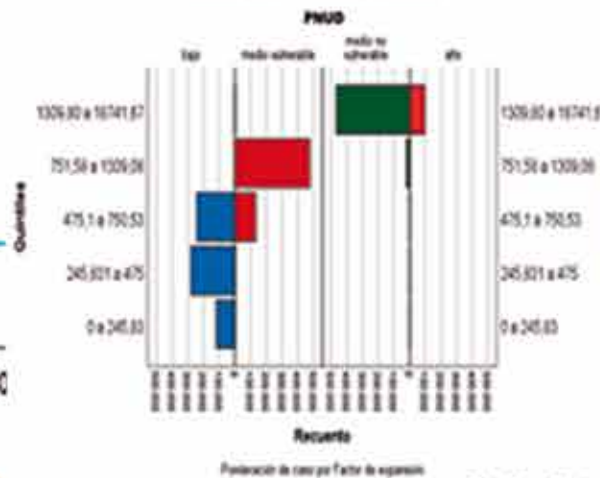
*Revista del Instituto de Estadística Teórica y Aplicada*

ISSN 2789-3510

Las tres familias



Grupo Mayor y Menor a lo nivel de juego: Grupo Mayor



UMSA  
 FCN  
 CARRERA  
 ESTADÍSTICA

**IETA**  
 Instituto de Estadística  
 Teórica y Aplicada

$$\begin{aligned}
 & E[(x-\mu)^2] \\
 &= E[x^2 - 2x\mu + \mu^2] \\
 &= E[x^2] - 2\mu E[x] + \mu^2 \\
 &= E[x^2] - 2\mu^2 + \mu^2 \\
 &= E[x^2] - \mu^2 \\
 &= E[x^2] - E[x]^2
 \end{aligned}$$



# *Varianza*

Revista de Investigación Científica del  
Instituto de Estadística Teórica y Aplicada

Número 22  
Octubre, 2023  
La Paz - Bolivia

Universidad Mayor de San Andrés  
Facultad de Ciencias Puras y Naturales  
Carrera de Estadística  
Instituto de Estadística Teórica y Aplicada (I.E.T.A.)

ISSN 2789-3510 VERSIÓN IMPRESA  
ISSN 2789-3529 VERSIÓN EN LÍNEA

**DEPÓSITO LEGAL**  
4-1-285-2021 P.O.

**REVISTA VARIANZA**  
Nº 22 - Octubre, 2023

**DIRECTOR CARRERA DE ESTADÍSTICA**  
M. Sc. Juan Carlos Flores López

**DIRECTOR INSTITUTO DE ESTADÍSTICA TEÓRICA Y APLICADA**  
Dr(c) Ramiro Coa Clemente

**DIAGRAMACIÓN Y DISEÑO**  
M. Zulema Vargas Cerrudo

*Los artículos presentados son entera responsabilidad de los autores*

VISIBILIDAD: REVISTAS BOLIVIANAS



La Paz - Bolivia  
Edificio Bloque FCPN - Campus Cota Cota  
Teléfonos: 2612824 -2612844  
Email: [ieta@umsa.bo](mailto:ieta@umsa.bo)  
Página web: <https://ojs.umsa.bo/ojs/index.php/revistavarianza>

# COMITÉ EDITORIAL

## EDITOR

**Ramiro Coa Clemente, Dr(c)**

(Estadístico, Demógrafo)

Universidad Mayor de San Andrés, Carrera de Estadística

Director del Instituto de Estadística Teórica y Aplicada

La Paz, Bolivia

E-mail: [rcoa@fcpn.edu.bo](mailto:rcoa@fcpn.edu.bo)

## COMITÉ CIENTÍFICO INTERNACIONAL

**Lizbeth Román Padilla, Ph.D.**

(Estadístico)

Universidad Anáhuac (Norte)

Ciudad de México, México

E-mail: [lizroman@hotmail.com](mailto:lizroman@hotmail.com)

**Yolanda M. Gómez Olmos, Dra.**

(Estadístico)

Universidad de Atacama

Atacama, Chile

E-mail: [yolanda.gomez@uda.cl](mailto:yolanda.gomez@uda.cl)

**Diego Gallardo Mateluna, Dr.**

(Estadístico)

Universidad de Atacama

Atacama, Chile

E-mail: [diego.gallardo@uda.cl](mailto:diego.gallardo@uda.cl)

**Luz Mery González García, Ph.D.**

(Estadístico)

Universidad Nacional de Colombia

Bogotá, Colombia

E-mail: [lgonzalezg@unal.edu.co](mailto:lgonzalezg@unal.edu.co)

**Martha Patricia Bohorquez Castañeda, Ph.D.**

(Estadístico)

Universidad Nacional de Colombia

Bogotá, Colombia

E-mail: [mpbohorquezc@unal.edu.co](mailto:mpbohorquezc@unal.edu.co)

**Adriana D'Amelio, Mg.**

(Estadístico)

Universidad Nacional de Cuyo

Mendoza, Argentina

E-mail: [estat06@hotmail.com](mailto:estat06@hotmail.com)

## **COMITÉ CIENTÍFICO NACIONAL**

**María Eugenia Aparicio Torrico, M.Sc.**

(Economista)

Universidad Autónoma Juan Misael Saracho UAJMS

Tarija – Bolivia

E-mail: [apariciouajms@gmail.com](mailto:apariciouajms@gmail.com)

**Arminda Flora Casso Lisarazu, M.Sc.**

(Economista)

Universidad Autónoma Juan Misael Saracho UAJMS

Tarija – Bolivia

E-mail: [armindacasso@gmail.com](mailto:armindacasso@gmail.com)



## PRESENTACIÓN

*Continuando con la edición semestral de nuestra Revista Científica, la Dirección del Instituto de Estadística Teórica y Aplicada (IETA), dependiente de la Carrera de Estadística de la Universidad Mayor de San Andrés, se complace en compartir con la comunidad universitaria, profesionales y población en general, la edición N° 22 de la Revista Varianza.*

*En esta nueva edición de la Revista se presentan tres artículos científicos originales. Ordenados alfabéticamente, en el primer artículo, luego de confirmar estadísticamente que las tasas de mortalidad infantil obtenidas para el país con el método indirecto se encuentran sistemáticamente sobrestimadas con relación a las tasas generadas con el método directo, consideradas como las tasas correctas, se procede a su corrección usando un modelo de regresión penalizado basado en algoritmos de machine learning; en el segundo artículo, con base en un conjunto de medidas de desigualdad de ingresos, se evidencia que en aquéllas poblaciones rurales del país con menor ingreso existe mayor desigualdad que en las poblaciones rurales de mayor ingreso; y en el tercer artículo, recurriendo al método multivariado para analizar tablas de tres vías, denominado STATIS, por su sigla en francés, se exploran las relaciones entre las diferentes dimensiones de la pobreza multidimensional urbana en el país y la evolución en el tiempo de esas relaciones.*

*Remarcar que los artículos científicos publicados en la presente edición fueron revisados y aprobados por el Comité Editorial. Es por ello que reiteramos nuestro enorme y sincero agradecimiento al grupo de profesionales que conforman el Comité Editorial de la Revista Varianza .*

*También nuestro respeto, valoración y agradecimiento a los autores por publicar sus artículos científicos en la Revista Varianza. Gracias a estos esfuerzos es que se puede concretar nuestro pequeño aporte al conocimiento científico. Finalmente, recordar que no difundir o publicar los resultados de una investigación es equivalente a no hacer investigación.*

Dr(c) Ramiro Coa Clemente

**DIRECTOR INSTITUTO DE ESTADÍSTICA TEÓRICA Y APLICADA**





# ÍNDICE

## Artículos originales

<b>Sobrestimación de la tasa de mortalidad infantil obtenida con el método indirecto. Un modelo para su corrección</b> <i>Autor: Ramiro Coa Clemente</i> .....	1
<b>Bolivia. Comparación de la desigualdad de ingresos en hogares rurales de bajos y altos ingresos</b> <i>Autores: Raúl León Delgado Álvarez y Jaime Tito Pinto Ajhuacho</i> .....	17
<b>Análisis estadístico multivariado exploratorio de estructuras temporales desde una óptica STATIS. Una aplicación a la pobreza urbana multidimensional en Bolivia</b> <i>Autor: Carlos Fernando Silva Viamonte</i> .....	33
<b>Instrucciones para autores</b> .....	49



# **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 Clemente<sup>1</sup>

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

✉ [clementecoa@gmail.com](mailto: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*

<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).

*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,  $\beta$  es un vector  $p * 1$  de coeficientes y  $\varepsilon_i$  es el error aleatorio asociado a la  $i$ -ésima observación, la función objetivo penalizada a ser minimizada es

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

## 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,  $\lambda$  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  $\phi$  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  $\lambda$ , 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  $\phi$  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  $\lambda > 0$ ,  $\phi$  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  $\phi$  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 salud<sup>2</sup>. 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 Sur<sup>3</sup>, 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

---

<sup>2</sup> 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.

<sup>3</sup> 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

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

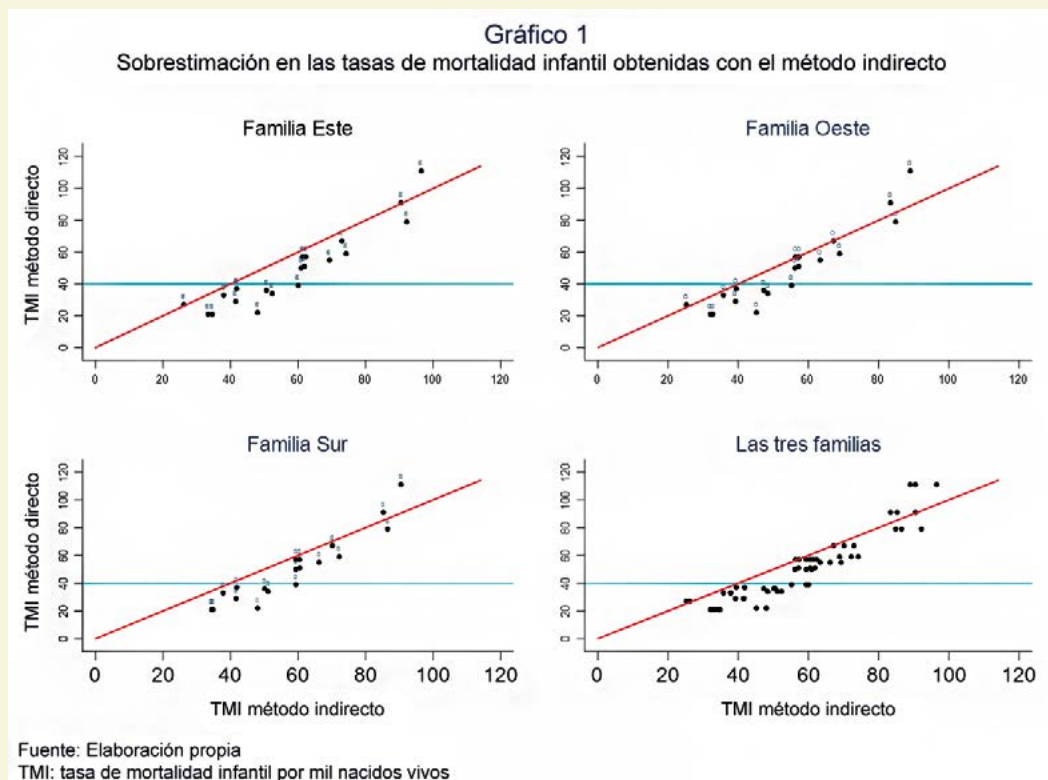
donde  $TMIDir$  es la tasa de mortalidad infantil con el método directo, y (ii) una diferencia relativa, definida como

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

## 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 indirecto<sup>4</sup> (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,

<sup>4</sup> 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

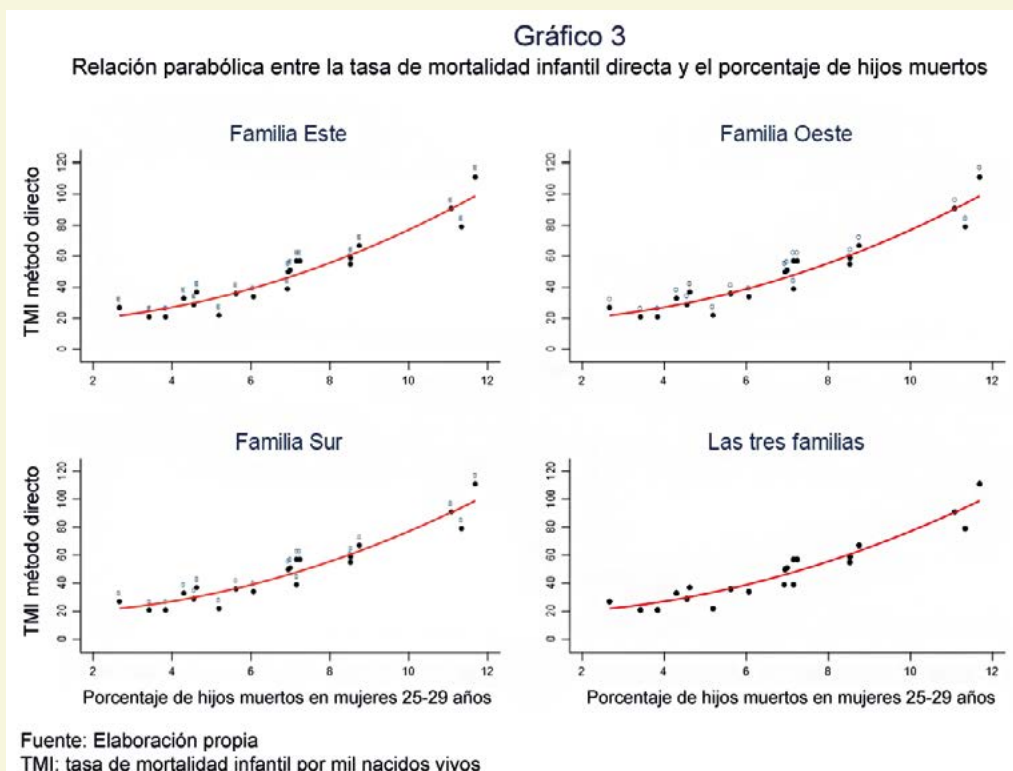
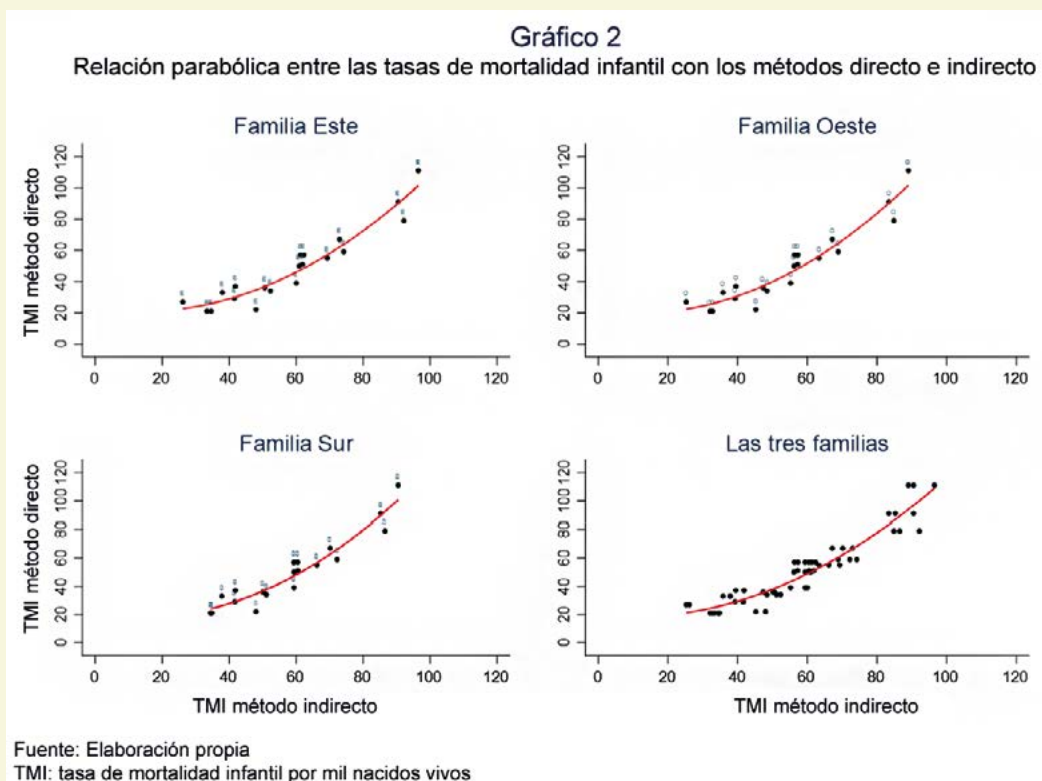
Hay una estrecha relación, en forma de parábola, entre las tasas de mortalidad calculadas con los métodos directo e indirecto<sup>5</sup> (Gráfico 2). Esta relación parabólica se observa para cada una de las tres familias de mortalidad (Este, Oeste y Sur) así como para el conjunto de datos. La misma 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.

<sup>5</sup> 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

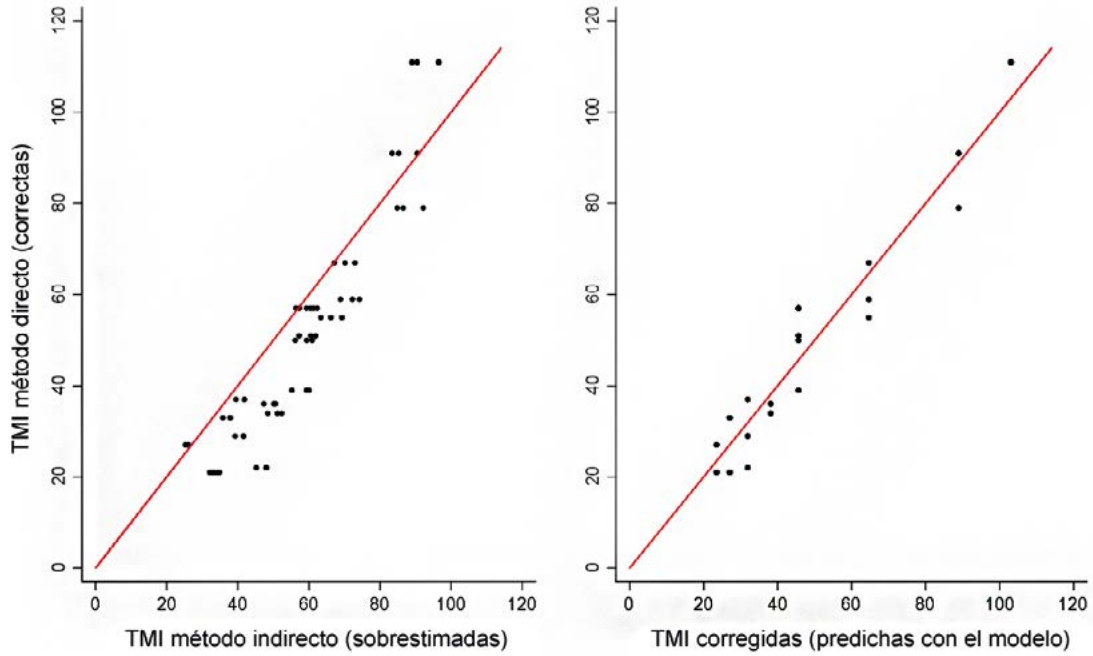
método directo y las tasas predichas con el modelo de regresión lasso. Idealmente, todos los puntos debieran estar sobre o muy próximos a la línea diagonal, lo que habría sugerido que el modelo de regresión genera muy buenas predicciones de las tasas, es decir, habría permitido eliminar totalmente el problema de sobrestimación. Las tasas de mortalidad infantil predichas, sin embargo, son razonables si se toma en cuenta que la única covariable predictorora 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

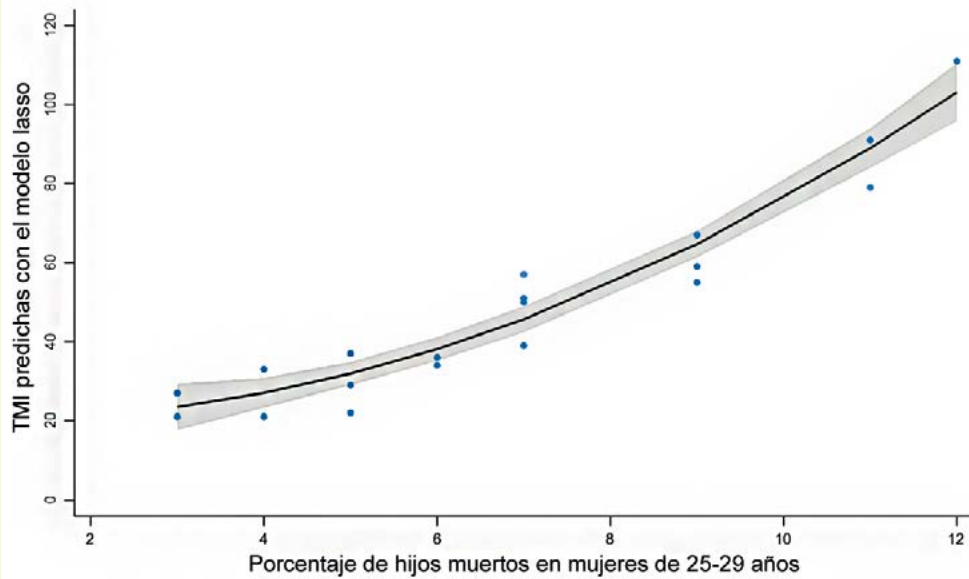
Relación de las tasas de mortalidad infantil correctas con las tasas sobrestimadas y con las tasas corregidas



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](http://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/>

# BOLIVIA. COMPARACIÓN DE LA DESIGUALDAD DE INGRESOS EN HOGARES RURALES DE BAJOS Y ALTOS INGRESOS

## BOLIVIA. COMPARISON OF INCOME INEQUALITY IN RURAL LOW-AND HIGH-INCOME HOUSEHOLDS, 2021

Raúl León Delgado Álvarez<sup>1</sup>

Universidad Mayor de San Andrés, La Paz - Bolivia

✉ [dea\\_5@hotmail.com](mailto:dea_5@hotmail.com)

Jaime Tito Pinto Ajhuacho<sup>2</sup>

Universidad Mayor de San Andrés, La Paz - Bolivia

✉ [titojaime\\_pinto@yahoo.com](mailto:titojaime_pinto@yahoo.com)

Artículo recibido: 01/08/2023

Artículo aceptado: 05/09/2023

---

### RESUMEN

*En base al análisis de la encuesta de ingresos de los hogares del año 2021, además de las consideraciones generales del estudio de la pobreza en Bolivia como los indicadores de la pobreza, y sobre la desigualdad, se analiza la comparación con la información mostrada en dicha encuesta, principalmente en los indicadores de Gini, Mehran y de Piesch, y también se propone un estudio más completo y de mayor eficacia con la de construcción de indicadores no solamente puntuales, sino por intervalos de confianza.*

*En la comparación de los ingresos del área rural se puede observar que los ingresos medios per cápita son mucho más dispares en la Región Menor (Pando, Chuquisaca, Cochabamba, Oruro y Potosí), que en la Región Mayor (Santa Cruz, La Paz, Tarija y Beni).*

**Palabras clave:** Desigualdad, pobreza, re muestreo.

### ABSTRACT

*Based on the analysis of the household income survey for year 2021, in addition to the general considerations based of the study of poverty in Bolivia as the indicators of poverty, and on inequality, the comparison with the information shown in said survey, mainly on the indicators of Gini, Mehran and Piesch, and a more complete and more effective study is also proposed, with that of not only punctual indicators, but by confidence intervals.*

*In the comparison of the income of the rural area, it can be observed that the average per capita income is much more disparate in the smaller region, (Pando, Chuquisaca, Cochabamba, Oruro y Potosí) than in the larger region (Santa Cruz, La Paz, Tarija y Beni).*

**Keywords:** Inequality, poverty, resampling.

---

## 1. INTRODUCCIÓN

La desigualdad es un término que se emplea cuando falta un equilibrio entre dos o más personas, presupone graves consecuencias en

un ámbito social, al encontrarse el problema de la desigualdad se ha convertido en el punto central de estudio dentro de la sociedad.

<sup>1</sup> Docente de la Carrera de Estadística de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés. ORCID: [0000-0001-7886-5264](https://orcid.org/0000-0001-7886-5264)

<sup>2</sup> Docente de la Carrera de Estadística de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés. ORCID: [0000-0002-1157-2165](https://orcid.org/0000-0002-1157-2165)

La desigualdad social, es un problema socioeconómico producto de la mala distribución de los ingresos en la sociedad.

La pobreza y la desigualdad son los mayores problemas de un país, y la migración tiene influencia en la inestabilidad social y económica.

Esta influencia se nota por el fortalecimiento del flujo migratorio interno a los centros urbanos mayores (ciudades) que aglomeran la mayor cantidad de población y concentran las actividades económicas.

Un estudio del Fondo Monetario Internacional (FMI) puso de relieve que una mayor igualdad de ingresos contribuía a prolongar la duración de las rachas de crecimiento económico de los países más que el libre comercio, la escasa corrupción gubernamental, la inversión extranjera o el bajo nivel de deuda externa (Berg y Ostry, 2011).

Tomando en cuenta el Desarrollo y la reducción de la pobreza a largo plazo, la desigualdad del tipo que sea, supone graves consecuencias, en toda sociedad, por lo cual es importante el estudio de la desigualdad.

En referencia a la desigualdad por ingreso, un informe advierte que las “desigualdades” excesivas podrían poner en peligro la sostenibilidad de economías, sociedades y comunidades, socavando esfuerzos para alcanzar los objetivos de Desarrollo Sostenible (ODSs).

En nuestro país, existen problemas en la economía como la pobreza, informalidad, desempleo y empleo precario, lo que provoca la disparidad de los ingresos; lo que nos muestra que la desigualdad económica afecta en el país, porque desalientan la formación profesional, obstruyen la movilidad económica y social y el Desarrollo humano

y en consecuencia inhabilitan el crecimiento económico.

El comportamiento o lo que está pasando en la sociedad, se pueden medir mediante indicadores sociales y económicos, que plantearan soluciones a problemas socioeconómicos, para reducir la brecha entre ricos y pobres.

Existen muchos estudios alrededor de estos conceptos como son pobreza y desigualdad en la misma, uno de ellos es el que desarrollo T.H. Marshall que introdujo el concepto de ciudadanía social en el año 1949 en su obra “Ciudadanía y Clase Social”, la cual analiza la evolución de los derechos ciudadanos en la historia y propone una definición de ciudadanía bajo tres elementos importantes: derechos civiles, derechos políticos y derechos sociales, dentro de estos últimos se desprenden los derechos sociales necesarios para garantizar una vida digna para todos los ciudadanos.

En el concepto de desigualdad en la pobreza entendiéndose como la disparidad económica y social que existe entre individuos que viven en condiciones de pobreza que se utilizara en el presente estudio está en concordancia con el axioma de monotonicidad y el axioma de monotonicidad de subgrupos, expresados en Sen (1976) “*Poverty and Ordinal Approach to Measurement*” y Foster Creer and Thorbecke (1984) “*Notes and Comments. A class of decomposable poverty measures*” respectivamente.

### **Análisis de la distribución de los ingresos económicos.**

El crecimiento y el equilibrio económico como proyectos opuestos para un país con menor desarrollo, en la actualidad proponen teorías que generan avances en el proceso económico, pero no la disminución de la desigualdad, por lo tanto, es importante

## **Bolivia. Comparación de la desigualdad de ingresos en hogares rurales de bajos y altos ingresos.**

---

mejorar la distribución de la riqueza, existen planteamientos concretos en programas de ayuda a la redistribución.

La determinación de indicadores de desigualdad dará una alarma, como el índice de Gini, para lo cual acudiremos a información de la Encuesta de Hogares de Bolivia del año 2021. En este trabajo se observará en regiones del área rural.

Bolivia está dividida administrativa y políticamente en 9 departamentos, y se tiene las áreas urbana y rural, dentro del área rural estudiaremos aquellas con mayor o menor desigualdad, con el fin de obtener una mayor significancia en la desigualdad de la pobreza, se verá la división o agrupación con ingresos per cápita en regiones mayores y menores a su media del área rural.

### **1.1 Objetivo General.**

El objetivo principal del estudio es comparar a través de varios índices como es la desigualdad de ingresos en hogares rurales de bajos y altos ingresos de Bolivia en el área rural, bajo la clasificación de menores y mayores que la media rural y mostrar no solo valores puntuales sino intervalos de confianza.

### **1.2 Objetivos Específicos.**

- 1.-Producir información de Indicadores sobre desigualdad del ingreso per cápita en hogares del área rural de Bolivia.
2. Medir indicadores de Gini, Mehran y Piesch para la el área rural de Bolivia.
- 3.- Cuantificar indicadores de desigualdad para hogares con menores y mayores ingresos en el área rural de Bolivia.
4. Identificar los departamentos de Bolivia donde más se dan los ingresos per cápita bajos y altos.

## **2. MÉTODO**

En el análisis que se realiza utilizamos información de la base de datos de la Encuesta de hogares 2021, realizada por el Instituto Nacional de Estadística (INE); quien periódicamente realiza encuestas para conocer variables de estudio, la unidad de observación es el hogar y el ámbito geográfico son los departamentos de Bolivia.

Para asimilar la desigualdad de ingresos podemos relacionar con las diferencias en la sociedad, es decir considerar que para algunos puede involucrar ventajas y desventajas para otros.

Las desventajas conducen a una muy limitada capacidad de ahorro, ante la posibilidad de que el hogar se quede sin ningún tipo de ingreso.

La persistencia de empleos precarios y bajos ingresos no solo suponen privaciones en el corto plazo, sino que tiene profundos impactos sobre el bienestar en las etapas más avanzadas del curso de vida, cuando se abandona el mercado de trabajo (Gallie, 2002).

### ***Medición de la distribución del ingreso per cápita.***

En una región se puede definir el ingreso relevante para medir el bienestar de las personas de una sociedad, y esta definición puede ser como la manera en que los ingresos de una familia son utilizados por la misma, según Ruiz Tagle y Contreras (1996), se puede pensar que al interior de los hogares se crea un monto de sus integrantes, para luego proceder a distribuirlos según las necesidades de sus miembros.

El **ingreso per cápita**, se obtiene dividiendo el ingreso total de la familia entre la cantidad de miembros de la misma, los estudios realizados con instituciones en torno a este tema para el análisis de la

desigualdad, trabajan con esta variable.

Buscando referencia al respecto se plantean diversos indicadores para utilizar conceptos similares en la descripción de las variables que indican desigualdad.

### *Instrumentos de análisis de desigualdad.*

Hay variedad de indicadores de desigualdad como el de Gini, y otros en relación a la entropía, estas medidas presentan distintas características, que pueden tener una sensibilidad mayor a los extremos de la distribución.

Otros son los de Mehran, Piesch, Theil, y los Coeficientes de concentración de los Percentiles superiores sobre los inferiores.

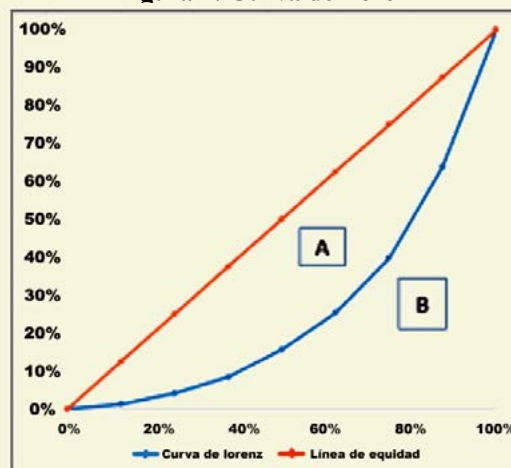
Entre las técnicas de Re muestreo se puede utilizar la de *Bootstrapping*, la cual permite medir la solidez de las variaciones de las medidas, en este caso, de distribución del ingreso.

### *Distribución del ingreso medido por el coeficiente de Gini.*

En los estudios de desigualdad se procede el análisis de la distribución del ingreso utilizando primeramente el coeficiente de Gini, que es escogida y utilizada por ser la medida de distribución de ingreso que puede tener un sesgo sólo en la parte media de la distribución y es la más utilizada en diversos trabajos y estudios.

El Indicador es el más usado para medir la desigualdad de la pobreza, Este coeficiente es igual a la proporción del área entre la Curva de Lorenz y la línea de equidistribución (área A del gráfico) con respecto a toda el área por debajo de la línea de equidistribución (correspondiente a A+B en el gráfico, que es igual a 0,5 dada la simetría de los rangos de la curva de Lorenz).

Figura 1: Curva de Lorenz



Fuente: <https://muyfinanciero.com/conceptos/indice-gini/>

Como resultado, el Coeficiente de Gini es un número entre 0 y 1, donde 0 es la perfecta igualdad (todos los individuos u hogares tienen el mismo ingreso) y 1 representa a la desigualdad (solo unos pocos hogares o individuos concentran todos los ingresos).

Gini (1912) definió su conocida medida de desigualdad en los términos siguientes:

$$IG = \frac{1}{x} \left( \frac{1}{2n^2} \sum_{i=1}^k \sum_{j=1}^k |x_i - x_j| n_i n_j \right)$$

Lo que se busca en la Curva de Lorenz, es medir los porcentajes entre pobreza y desigualdad, e indicar que una alta relación de ingresos bajos sobre el total puede conducir a que los estimadores de la desigualdad generen resultados mayores, sin que eso indique una variación en toda la distribución.

Un estudio con una connotación Estadística plantea la necesidad de ver cuan confiable pueden ser los indicadores de la pobreza y de desigualdad de la misma en el estudio del ingreso per cápita de los hogares así la inclusión de los indicadores de Mehran y Piesch permitirán una comparación de esos índices.

## Bolivia. Comparación de la desigualdad de ingresos en hogares rurales de bajos y altos ingresos.

### *Curva de Lorenz.*

Otro indicador de consistencia adicional para el análisis de distribución de ingreso corresponde a la comparación gráfica de las curvas de Lorenz, en su armado utiliza la función acumulada de la proporción de individuos, partiendo de los más pobres, en el eje de las abscisas, contra la proporción acumulada de los ingresos en el eje de las ordenadas.

La curva de Lorenz es una representación de la distribución del ingreso de una población (Lorenz, 1905). Esta curva tiene en el eje horizontal el porcentaje acumulado de individuos desde el más pobre hasta el más rico y en el eje vertical el porcentaje de ingreso perteneciente a cada porcentaje acumulado de la población.

En la curva de Lorenz se desprende una mayor concentración y desigualdad entre las familias con mayor ingreso, lo cual significaría en otros términos los de menores ingresos tienen una estructura equitativa entre sí, mientras que las familias con mayores recursos poseen grandes diferencias entre sí.

Por este motivo, es que también se recurre a los indicadores de la Entropía Generalizada, principalmente al índice de Theil, para buscar una mayor explicación de las causas de la desigualdad en cada región en particular.

### *Coefficientes de Entropía*

Los coeficientes de Entropía permiten, por un lado, medir la desigualdad por rangos de ingresos (dentro y entre grupos), y por otro lado, al permitir una descomposición de estos rangos, también hacen posible analizar variables o atributos que determinan la distribución del ingreso, como el peso específico de estas variables en la explicación de esta distribución.

La elaboración de estos coeficientes se inscribe dentro de lo que se han denominado las aproximaciones axiomáticas a la medición de la desigualdad de ingresos.

Dentro de estos coeficientes podemos mencionar al Índice de Theil, que tiene una sensibilidad constante a lo largo de la distribución.

Theil (1967) propuso un indicador para medir la desigualdad con base en el concepto de Entropía derivado de la teoría de la información. La Entropía describe cuanta aleatoriedad hay en un evento: el grado de Entropía de un evento es una función decreciente de su probabilidad de ocurrencia. Haciendo un paralelismo entre este concepto y la desigualdad distributiva. Theil, definió su Indicador como medida de desigualdad del ingreso:

$$\text{Índice de Theil} = \frac{1}{N} \sum \frac{x_i}{\mu} * \ln\left(\frac{x_i}{\mu}\right)$$

Donde N = número total de individuos o personas  $x_i$  = ingreso del individuo  $i$  y  $\mu$  = ingreso promedio de los individuos.

### **Índices para el análisis de la desigualdad.**

#### *Cocientes: razón entre cuantiles.*

Después de ordenar a la población según su nivel de ingreso per cápita, por ejemplo, se divide en grupos de igual tamaño, llamados cuantiles.

#### *Quintiles del ingreso*

Un método utilizado en la distribución del ingreso,  $Y = (y_1, y_2, \dots, y_n)$  es la descripción y análisis de los quintiles de los ingresos, en su procedimiento los ingresos son ordenados según su magnitud de menor a mayor.

El ingreso total se lo tiene sumando los



ingresos de los individuos que pertenecen al quintil y luego dividiendo esa suma por la suma total de los ingresos.

En cada quintil se puede determinar su porcentaje de participación, su notación es:

$$p_i = \frac{\text{Suma } Q_i}{\sum \text{Quintiles}} * 100$$

$$\sum p_i = 100$$

$Q_i$  es el quintil de orden  $i$ , Donde  $i = 1, 2, 3, 4, 5$

También se puede determinar la razón de la participación en el ingreso total del quintil más rico respecto del quintil más pobre, utilizando la expresión:

$$\text{Razon} = \frac{Q_5}{Q_1}$$

Esta razón indica el número de veces que el ingreso total del 20% más rico de la población supera al ingreso total del 20% más pobre.

El indicador consiste en calcular el cociente entre la media de la variable de análisis entre dos cuantiles, generalmente situados en lugares opuestos en la distribución.

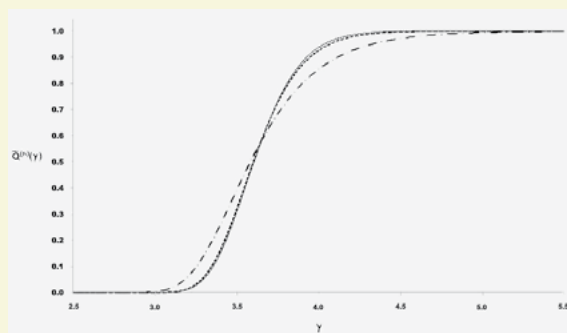
También el cociente se calcula para los cuantiles extremos, por ejemplo, entre los deciles 1 y 10, para dar cuenta del tamaño que pueden alcanzar las brechas sociales, algunos utilizan para estudiar cocientes de cuantiles intermedios, como por ejemplo el cociente entre el percentil 25 y el 75, con el fin de evitar valores extremos y de complementar el análisis estudiando varios puntos de la distribución.

### ***Función de distribución acumulada***

La función de distribución acumulada es otra manera de visualizar la distribución de una variable, indica la proporción acumulada de individuos (eje vertical)

para cada valor del rango de la variable (eje horizontal), y comienza en las coordenadas (0,0) para ir aumentando hasta llegar al valor máximo del rango de la variable, donde la proporción acumulada es igual a 1.

**Figura 2: Distribución acumulada**



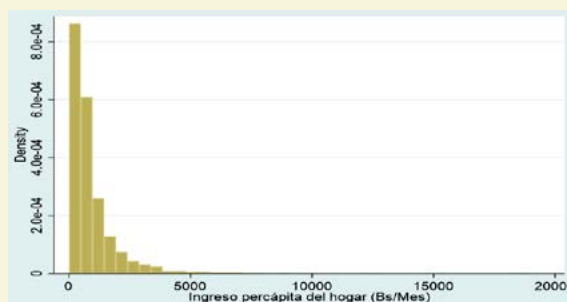
Fuente : [https://www.researchgate.net/figure/Funcion-de-distribucion-acumulada-evaluada-numericamente-tomando-constante-el-numero-de-fig5\\_216518896](https://www.researchgate.net/figure/Funcion-de-distribucion-acumulada-evaluada-numericamente-tomando-constante-el-numero-de-fig5_216518896)

## **3. RESULTADOS**

De acuerdo a lo planteado en la metodología se analiza la variable ingreso per cápita, al nivel del área rural por departamento de Bolivia, los datos de esta variable muestran un histograma con asimetría a la derecha.

**Gráfico 1.**

### **BOLIVIA - Área Rural: Histograma del ingreso percapita**



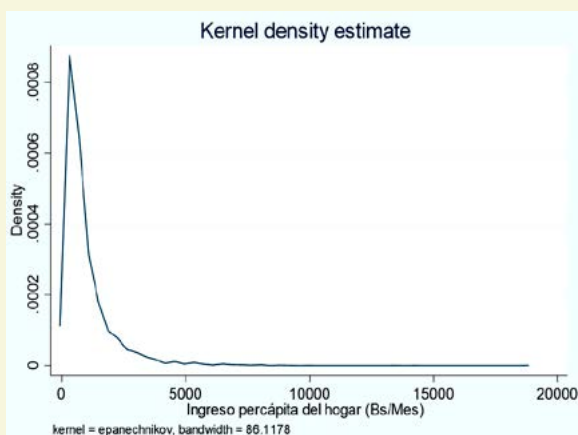
Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

Tomando en cuenta una estimación no paramétrica de la función de densidad de la variable ingreso per cápita por el método de kernel tendremos una representación suavizada del histograma y será una línea continua.

## Bolivia. Comparación de la desigualdad de ingresos en hogares rurales de bajos y altos ingresos.

**Gráfico 2**

**BOLIVIA - Área Rural: Kernel del ingreso per cápita**

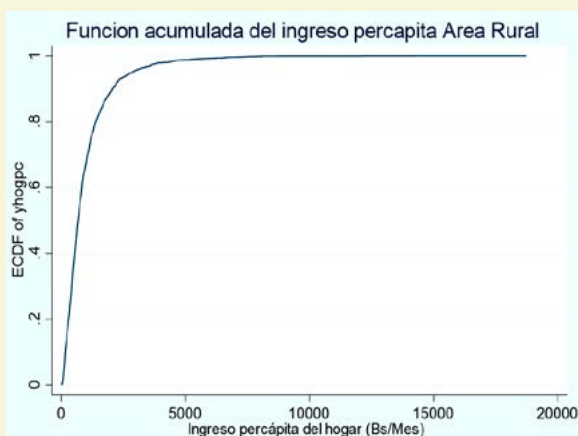


Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

De la misma variable tenemos su distribución acumulada.

**Gráfico 3**

**BOLIVIA - Área Rural: Función acumulada del ingreso per cápita**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

Lo que se quiere, es conocer que pasa al interior del área rural por departamento, como es el ingreso per cápita promedio.

Un criterio fue determinar el promedio a nivel rural nacional, que es 979 Bs., la cual nos permitió clasificar en dos regiones, la primera con información de hogares menores al promedio de ingreso per cápita, y el otro con ingresos mayores a la media, y en ella determinar los extremos para el análisis en los departamentos.

**Cuadro 1.**

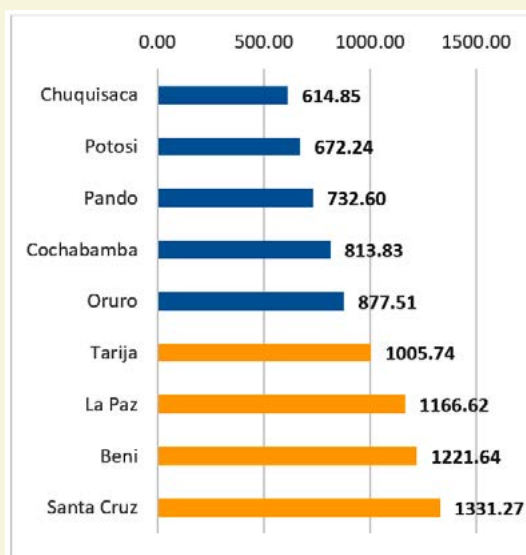
**BOLIVIA – Área Rural: Número de hogares y regiones menores y mayores a la media del ingreso per cápita rural**

Departamento	No. Hogares	Media
<b>Rural-Bolivia</b>	<b>3.516.671</b>	<b>979,0601</b>
Chuquisaca	317.708	614,8500
Potosí	515.210	672,2400
Pando	62.604	732,6000
Cochabamba	610.522	813,8300
Oruro	183.765	877,5100
Tarija	185.830	1.005,7400
La Paz	974.308	1.166,6200
Beni	125.588	1.221,6400
Santa Cruz	541.135	1.331,2700

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Gráfico 4**

**BOLIVIA - Área Rural: Media de ingresos per cápita por regiones menores y mayores**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

Es así que tomando como referencia el año 2021, se realiza un análisis utilizando la diversidad de instrumentos que ayuden a la interpretación de la realidad de las características en las áreas de estudio.

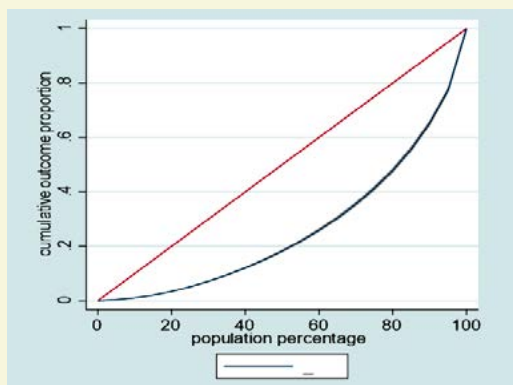
Se realizaron: estimaciones de la curva de Lorenz, coeficiente de Gini, Mehran, Piesch, índice de Theil y gráficos de la curva de Lorenz donde se considera el ingreso per cápita del hogar como variable para medir la desigualdad en el área rural de los nueve departamentos.

**Cuadro 2**  
**BOLIVIA - Área Rural: Coeficiente de Lorenz**

Percentil	Coeficiente Lorenz	Intervalo de Confianza (95%)	
		Límite Inferior	Límite Superior
0	0		
5	0,0037763	0,0035606	0,003993
10	0,0107013	0,0102173	0,0111853
15	0,0205508	0,01962	0,0214807
20	0,0339983	0,0324635	0,0355332
25	0,0512175	0,0491033	0,0533317
30	0,0717461	0,0690477	0,0744445
35	0,0947565	0,0915415	0,0979714
40	0,1207996	0,1169573	0,1246418
45	0,1501079	0,1456215	0,1545943
50	0,1825214	0,1773683	0,1876745
55	0,2187267	0,2128674	0,2245859
60	0,2593286	0,2527237	0,2659335
65	0,3041745	0,2968279	0,3115212
70	0,3552911	0,3471363	0,363446
75	0,4131055	0,4041525	0,4220586
80	0,479195	0,4695458	0,4888442
85	0,55741	0,5471861	0,5676339
90	0,6524112	0,6418109	0,6630115
95	0,7738436	0,7640303	0,7836568
100	1		

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Gráfico 5**  
**BOLIVIA - Área Rural: Coeficiente de Lorenz del ingreso per cápita.**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

La curva de Lorenz muestra que en el área rural el 25% de los hogares concentra el 5,12% de los ingresos per cápita con un intervalo de confianza de 4,91% y 5,33%,

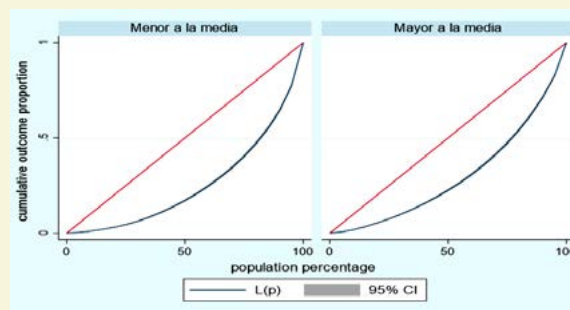
mientras que el 50% de los hogares acumula el 18,25% de los ingresos con un intervalo de confianza de 17,73% y 18,76%.

**Cuadro 3**  
**BOLIVIA - Área Rural: Coeficiente de Lorenz de las regiones menores y mayores al ingreso promedio del ingreso per cápita**

Percentil	Coeficiente de Lorenz	
	Región Menor	Región Mayor
0	0	0
5	0,0038284	0,0062723
10	0,0104228	0,0172109
15	0,0192085	0,0323582
20	0,0297872	0,0503733
25	0,0433085	0,0704016
30	0,0604699	0,093191
35	0,0817348	0,1185845
40	0,1063604	0,1467815
45	0,1343825	0,1780
50	0,1656943	0,2124265
55	0,2011839	0,2496007
60	0,2408239	0,2916865
65	0,2849263	0,3385923
70	0,3342258	0,3901566
75	0,3906625	0,4478998
80	0,4561185	0,5142347
85	0,5317139	0,5906473
90	0,6251435	0,679413
95	0,7532221	0,7930105
100	1	1

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Gráfico 6**  
**BOLIVIA - Área Rural: Coeficiente de Lorenz de ingreso per cápita.**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

En la gráfica de la curva de Lorenz, se ve que existe menos igualdad en la distribución

## Bolivia. Comparación de la desigualdad de ingresos en hogares rurales de bajos y altos ingresos.

del ingreso per cápita de los hogares de la Región con ingresos menores a la media, comparando con la distribución de los ingresos de los hogares de la Región Mayor con ingresos mayores, este último grupo está menos alejado de la igualdad  $G=0,5$ , del grupo con ingresos mayores.

Se puede observar que en el área rural (Región con ingresos menores) el 25% de los hogares concentra el 4,33% de los ingresos, mientras que el 50% de los hogares acumula el 16,56% de los ingresos; y el 75% acumula el 39,06%, en cambio en el área rural (con ingresos mayores) el 25% de los hogares concentra el 7,04% de los ingresos, el 50% de los hogares obtiene el 21,24% de los ingresos y el 75% acumula el 44,78% de los ingresos.

La grafica de Lorenz muestra que la distribución de los ingresos con ingresos menores (Chuquisaca, Cochabamba, Oruro, Potosí y Pando), está más alejada de la igualdad de 0,5 del grupo con ingresos mayores (Tarija, La Paz, Beni y Santa Cruz).

### Índices de desigualdad de ingresos

#### Razón de quintiles.

Analizando el ingreso medio per cápita de los hogares y su distribución porcentual por quintiles en el área rural, tenemos lo siguiente:

**Cuadro 4**

**BOLIVIA – Área Rural: Media, Suma del ingreso per cápita por quintiles y porcentaje de participación**

5 cuantiles de yhogpc	Media	Suma	Porcentaje de Participación
1	166,574769956	117.281.462,78	3,41
2	425,082122424	298.745.422,563	8,67
3	678,493142726	478.031.700,94	13,99
4	1.077,468136820	757.936.427,15	22,01
5	2.552,033087340	1.791.036.903,68	52,02
<b>Total</b>	<b>979,060114062</b>	<b>3.443.031.917,11</b>	<b>100</b>

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

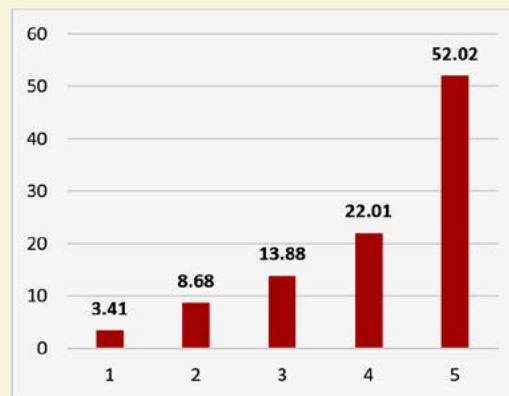
Para el área rural de Bolivia el primer quintil

(20% más pobre) genera el 3,41% de todo el ingreso que genera la población, el quinto quintil (20% más rico) genera el 52,02% de todo el ingreso que genera la población.

Los ingresos per cápita de los hogares del quinto quintil (20% más rico) ganan en promedio 15,25 veces más que el primer quintil (20% más pobre).

**Gráfico 7**

**BOLIVIA – Área Rural: Participación de quintiles del ingreso per cápita.**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Cuadro 5**

**BOLIVIA – Área Rural: Media, Suma del ingreso per cápita por quintiles y porcentaje de participación de región con ingresos menores a la media del ingreso per cápita**

5 cuantiles de yhogpc	Media	Suma	Porcentaje de Participación
1	154,6846759	81.811.905,24	6,65
2	416,359289	160.328.964,87	12,87
3	668,3450668	207.299.405,65	16,64
4	1.053,87	270.311.630,05	21,7
5	2.514,18	525.914.379,40	42,21
<b>Total</b>	<b>737,1638556</b>	<b>1.245.666.285,21</b>	<b>100</b>

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

Respecto a la Región Rural de ingresos menores, el primer quintil (20% más pobre) genera el 6,65% de todo el ingreso que genera la población, el quinto quintil (20% más rico) genera el 42,21% de todo el ingreso que genera la población.

Los ingresos per cápita de los hogares del quinto quintil (20% más rico) ganan en

promedio 6,34 veces más que el primer quintil (20% más pobre).

**Cuadro 6**  
**BOLIVIA – Área Rural: Media, Suma del ingreso per cápita por quintiles y porcentaje de participación de región con ingresos mayores a la media del ingreso per cápita**

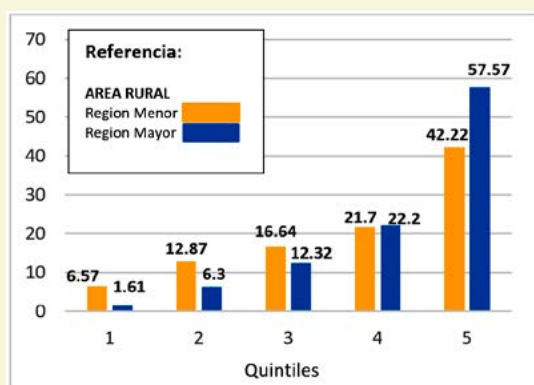
5 quintiles de yhogpc	Media	Suma	Porcentaje de Participación
1	202,4722783	35.469.557,54	1,61
2	435,6540789	138.416.457,69	6,29
3	686,4742882	270.732.295,29	12,32
4	1.091,01	487.624.797,10	22,19
5	2.568,11	1.265.122.524,28	57,57
<b>Total</b>	<b>1.202,81</b>	<b>2.197.365.631,91</b>	<b>100</b>

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

Respecto de la Región rural de ingresos mayores, el primer quintil (20% más pobre) genera el 1,61% de todo el ingreso que genera la población, el quinto quintil (20% más rico) genera el 57,57% de todo el ingreso que genera la población.

Los ingresos per cápita de los hogares del quinto quintil (20% más rico) gana en promedio 35,75 veces más que el primer quintil (20% más pobre).

**Gráfico 8**  
**BOLIVIA – Área Rural: Porcentaje de participación de quintiles por región menor y mayor**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

### Indicadores de Desigualdad.

Realizando las operaciones con datos de la Encuesta de hogares 2021 del INE, para el área rural se tiene el siguiente valor 0,4779 como se

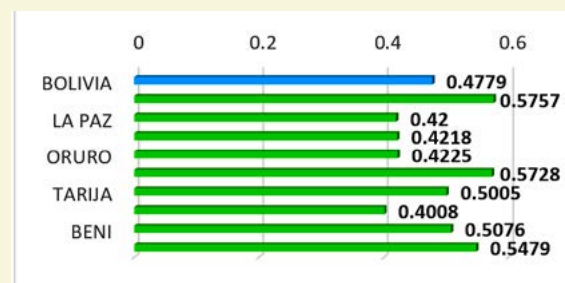
lo puede ver en la información proporcionada por el INE en sus informes estadísticos.

**Cuadro 7**  
**BOLIVIA: Índice de Gini para el ingreso per cápita mnsual área rural 2015-2021**

INDICADOR	2015	2016	2017	2018	2019	2020	2021
Bolivia	0,46	0,46	0,46	0,43	0,42	0,45	0,42
Área Rural	0,52	0,54	0,54	0,49	0,47	0,48	0,48

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Gráfico 9**  
**BOLIVIA – Área Rural: Índice de Gini por departamento**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Cuadro 8**  
**BOLIVIA - Área Rural (año 2021): Coeficiente de Gini, Mehran, Piesch y Theil**

	GINI	MEHRAN	PIESCH	THEIL
BOLIVIA	0,47797	0,62187	0,40602	0,40596
CHUQUISACA	0,57571	0,71512	0,506	0,6276
LA PAZ	0,42006	0,55006	0,35507	0,30753
COCHABAMBA	0,42185	0,55713	0,35422	0,31398
ORURO	0,42251	0,55482	0,35635	0,31828
POTOSI	0,57289	0,71304	0,50281	0,60514
TARIJA	0,50058	0,6396	0,43108	0,44865
SANTA CRUZ	0,40082	0,52875	0,33686	0,28605
BENI	0,5076	0,63961	0,4416	0,48791
PANDO	0,54792	0,6912	0,47628	0,61466

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

En el Cuadro se expone las estimaciones del coeficiente de Gini, Mehran, Piesch y Theil del área rural de Bolivia desglosado en los nueve departamentos, se ve que los índices están próximos al 0,50 con variaciones mínimas.

El índice de Gini para el área rural de Bolivia es 0,47797, el Índice de Mehran es 0,62187, el índice de Piesch 0,40602 y el índice de Theil es 0,40596.

## Bolivia. Comparación de la desigualdad de ingresos en hogares rurales de bajos y altos ingresos.

**Cuadro 9.**

**BOLIVIA: Área Rural región menor y mayor: Coeficiente de Gini Mehran, Piesch y Theil**

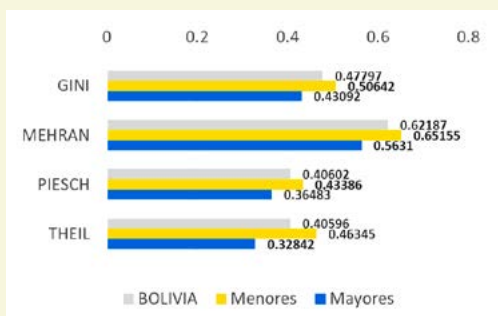
GRUPO	GINI	MEHRAN	PIESCH	THEIL
<b>BOLIVIA</b>	<b>0,47797</b>	<b>0,62187</b>	<b>0,40602</b>	<b>0,40596</b>
<b>Región Menor</b>	0,50642	0,65155	0,43386	0,46345
<b>Región Mayor</b>	0,43092	0,5631	0,36483	0,32842

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

Los índices mencionados son mayores para el grupo de ingresos menores, lo que nos indica que existe mayor desigualdad en los cuatro indicadores correspondientemente, se puede ver en la gráfica siguiente:

**Gráfico 10**

**BOLIVIA – Área Rural: (Región menor y mayor a la media del ingreso per cápita) Por índices de Gini, Mehran, Piesch y Theil**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Cuadro 10**

**BOLIVIA - Área Rural: Hogares más pobres en base al quintil 1 y 2**

BOLIVIA	HOGARES MÁS POBRES	
	QUINTIL 1	QUINTIL 2
Chuquisaca	2,46	8,18
La Paz	5,24	15,31
Cochabamba	4,8	14,85
Oruro	5	14,93
Potosí	2,68	8,03
Tarija	3,38	11,73
Santa Cruz	5,84	15,84
Beni	3,58	12
Pando	1,96	9,2

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

Tomando en cuenta la información del porcentaje acumulado de los ingresos per cápita de los hogares de los nueve departamentos, el análisis tomando en cuenta los quintiles,

nos indica que con mayor desigualdad de los ingresos es el departamento de Pando porque el 20% más pobre de los hogares recibe el 1,96% de los ingresos y el 40% de los hogares recibe el 9,20% de los ingresos, seguido por el departamento de Chuquisaca con 2,46% para el quintil uno y 8,18% para el segundo quintil, Potosí 2,68% en el primer quintil y 8,03% del segundo quintil, siendo los más bajos en comparación con los otros departamentos. Los departamentos que muestran más igualdad en la distribución de los ingresos son Santa Cruz y La Paz con 5,84% y 5,24% según indica su primer quintil respectivamente.

**Cuadro 11**

**BOLIVIA - Área Rural: Intervalo de confianza (95%) Índice de Gini (ingreso per cápita mensual) por departamento 2021**

Departamento	Puntual	Intervalo de Confianza (95%)	
		Límite Inferior	Límite Superior
Chuquisaca	0,5757	0,5515	0,5974
La Paz	0,4201	0,3980	0,4398
Cochabamba	0,4219	0,4004	0,4420
Oruro	0,4225	0,4007	0,4452
Potosí	0,5729	0,5534	0,5903
Tarija	0,5006	0,4724	0,5257
Santa Cruz	0,4008	0,3771	0,4245
Beni	0,5076	0,4611	0,5463
Pando	0,5479	0,5059	0,5879

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Cuadro 12.**

**BOLIVIA - Área Rural: Intervalo de confianza (95%) Índice de Mehran (ingreso per cápita mensual) por departamento 2021**

Departamento	Puntual	Intervalo de Confianza (95%)	
		Límite Inferior	Límite Superior
Chuquisaca	0,7151	0,6941	0,7341
La Paz	0,5501	0,5270	0,5708
Cochabamba	0,5571	0,5349	0,5779
Oruro	0,6678	0,6542	0,6804
Potosí	0,7130	0,6947	0,7293
Tarija	0,6396	0,6120	0,6652
Santa Cruz	0,5288	0,5058	0,5513
Beni	0,6396	0,5959	0,6762
Pando	0,6912	0,652189	0,7270784

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Cuadro 13**  
**BOLIVIA - Área Rural: Intervalo de confianza (95%) Índice de Piesch (Ingreso per cápita mensual) por departamento 2021**

Departamento	Puntual	Intervalo de Confianza (95%)	
		Límite Inferior	Límite Superior
Chuquisaca	0,5060	0,4800	0,5299
La Paz	0,3551	0,3330	0,3753
Cochabamba	0,3542	0,3330	0,3746
Oruro	0,3564	0,3338	0,3789
Potosí	0,5028	0,4822	0,5225
Tarija	0,4311	0,4025	0,4570
Santa Cruz	0,3369	0,3126	0,3615
Beni	0,4416	0,3939	0,4840
Pando	0,4763	0,4315	0,5192

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Cuadro 14**  
**BOLIVIA - Área Rural: Ingreso medio per cápita de hogares por región, según quintiles**

5 QUANTILES DE YHOGPC	INGRESO MEDIO PER CÁPITA DE HOGARES		
	Rural	Región Menor	Región Mayor
1	166,57	154,68	202,47
2	425,08	416,36	435,65
3	678,49	668,34	686,47
4	1.077,47	1.053,87	1.091,01
5	2.552,03	2.514,18	2.568,11

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Cuadro 15.**  
**BOLIVIA - Área Rural: Distribución porcentual del ingreso per cápita, según quintiles**

5 QUANTILES DE YHOGPC	DISTRIBUCIÓN PORCENTUAL DEL INGRESO		
	Rural	Región Menor	Región Mayor
1	3,41	6,56	1,61
2	8,68	12,87	6,3
3	13,88	16,64	12,32
4	22,01	21,7	22,19
5	52,02	42,22	57,57

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

En referencia a la medición de la pobreza y la desigualdad se conoce un indicador de “Polarización”, es la clasificación de “estratos de ingreso”, como indica los estudios que realiza el PNUD (Programa de las Naciones Unidas para el desarrollo).

El PNUD (2010) y López Calva y Juárez-Ortiz (2011) consideran clasificar a la población en estratos.

### Estrato de ingresos bajo

Comprende a todas las personas que poseen un ingreso del hogar per cápita inferior a la líneas de pobreza nacional.

### Estrato de ingreso medio vulnerable

Agrupar las personas que poseen un ingreso del hogar per cápita entre 1 y 2 veces la línea de pobreza nacional.

### Estrato de ingreso medio no vulnerable

Agrupar las personas que poseen un ingreso del hogar per cápita entre 2 y 5 veces la línea de pobreza nacional.

### Estrato de ingreso alto

Considera a las personas que poseen un ingreso del hogar per cápita mayor a 5 veces la línea de pobreza nacional.

Generando una variable que contenga las 4 categorías de estrato de ingreso per cápita, tenemos la siguiente clasificación.

**Cuadro 16.**  
**BOLIVIA - Área Rural: Número de hogares y porcentaje de hogares por estrato social PNUD.**

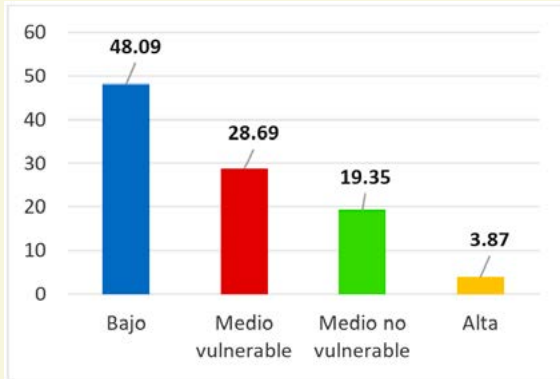
Estrato social PNUD Bolivia	Número de hogares	Porcentaje
Bajo	1.691.231	48,09
Medio vulnerable	1.008.840	28,69
Medio no vulnerable	680.631	19,35
Alta	135.969	3,87
<b>Total</b>	<b>3.516.671</b>	<b>100</b>

Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

En el estrato bajo hay 1,6 millones de hogares, que corresponde al 48,09%, y en el estrato alto hay 135.969 hogares que son el 3,87%.

## Bolivia. Comparación de la desigualdad de ingresos en hogares rurales de bajos y altos ingresos.

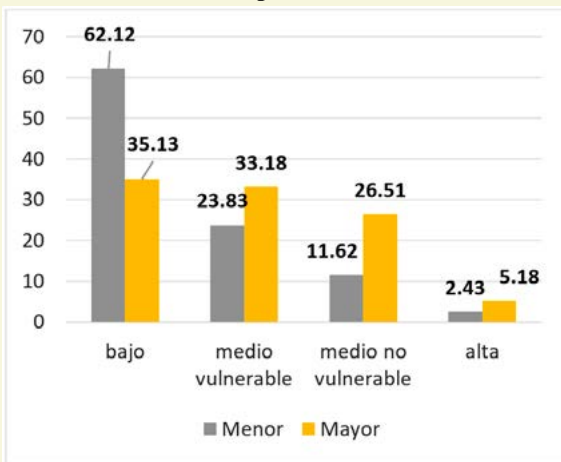
**Gráfico 11**  
BOLIVIA – Área Rural: Calificación del PNUD por estratos



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

Según los datos en el Área Rural en el estrato “bajo” en la Región de Menores ingresos los hogares tienden a ser los más afectados en su vida social y económica, la calificación del PNUD indica un 62,12%, lo cual lo hace muy significativo, y son más vulnerables en comparación de la Región Mayores ingresos (35,12%), un factor entre otros que ayuda a este último grupo posible sea la frontera agrícola que permite actividad económica y ayuda a las familias mejorar su ingreso económico.

**Gráfico 12**  
BOLIVIA – Área Rural: (Región menor y mayor) Por estratos pobreza de PNUD



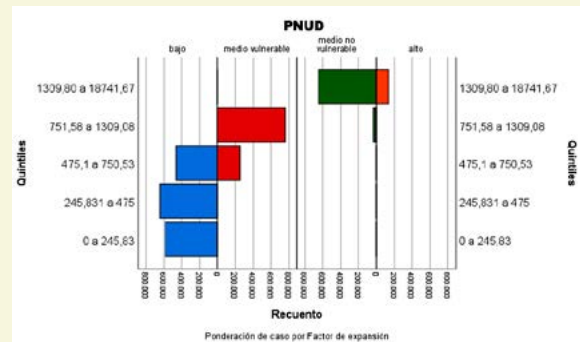
Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Gráfico 13**  
BOLIVIA – Área Rural: Región menor y mayor por quintiles del ingreso per cápita



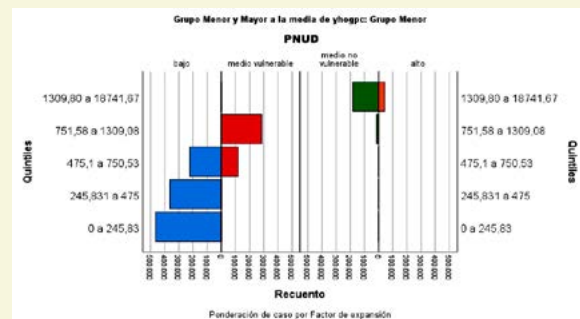
Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

**Gráfico 14**  
BOLIVIA – Área Rural: (Región menor y mayor) por clasificación del PNUD, según quintiles del ingreso per cápita



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

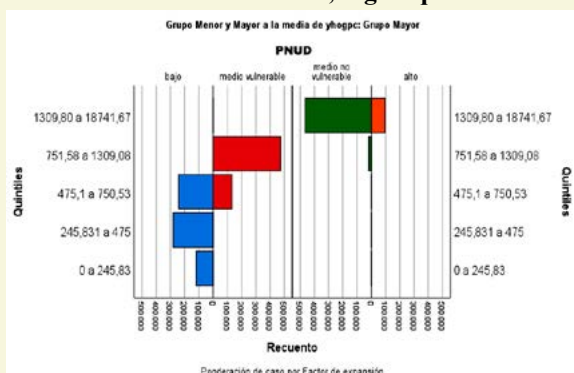
**Gráfico 15**  
BOLIVIA – Área Rural: Región menor por clasificación del PNUD, según quintiles.



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021



**Gráfico 16**  
**BOLIVIA – Área Rural: Región mayor por clasificación del PNUD, según quintiles.**



Fuente: Elaborado con datos del INE, Encuesta de Hogares 2021

#### 4.-CONCLUSIONES

En el área rural revisando la información se ve los ingresos medios per cápita son mucho más dispares en la Región Menor (Pando, Chuquisaca y Potosí), en cambio la Región Mayor, muestra hogares cuyos ingresos son aproximadamente el doble de la primera región.

Los hogares más desfavorables en sus ingresos se encuentran en la Región Menor con ingresos bajos que en lo posible, solo sea en su mayoría de subsistencia.

Al calcular las curvas de Lorenz hay menor igualdad de los ingresos per cápita en la Región Menor, y en la Región Mayor muestra una tendencia de igualdad cercana a 0,5.

Analizado el Ratio de quintiles en el área rural, se tiene que los ingresos per cápita de los hogares del quinto quintil (20% más rico) gana en promedio 15,25 veces más que el primer quintil (20% más pobre).

Se ve también que el 20% de los departamentos clasificados de pobreza menor reciben el 1,96 de los ingresos Pando seguido de Chuquisaca y Potosí. Los departamentos que muestran mayor distribución de sus ingresos son los de Santa Cruz y La Paz.

En comparación de los ingresos que muestra el PNUD, se ve que la mayoría de los hogares son calificados como “bajo” con un 48,09% y como “alto” en un 3,87%.

Una conclusión general nos indica que dentro del Área Rural los hogares considerados como pobres, según los indicadores de desigualdad resulta que sus ingresos per cápita son muy bajos y se deben plantear políticas públicas más efectivas que beneficien al desarrollo de los hogares considerados como más pobres, según los datos estas se encuentran en los departamentos de Pando, Chuquisaca y Potosí, por el otro lado en la Región de mayor ingreso son aceptables y mejores posible debido a actividades económicas que reporten mejores ingresos, como son los departamentos de Santa Cruz y La Paz.

El encarar y buscar soluciones a la desigualdad es invertir en cooperación y desarrollo, garantizar la protección de la población en emergencias humanitarias, creación de sistemas fiscales justos, mejorar el acceso a recursos públicos básicos, resumiendo no solo crear políticas públicas sino viabilizarlas.

#### 5.- DISCUSIÓN

La explicación del ingreso per cápita de los hogares, plantea un modelo para explicar el comportamiento de los ingresos, bibliografías.

Se tiene modelos como los de Gamma, exponencial y log normal, que profundizando explican el comportamiento de la variable de estudio.

Al momento se tienen instrumentos para medir la desigualdad que ayudan en la explicación del ingreso per cápita.

Trabajar en modelos estadísticos, ayudaran mucho a mejor la visualización de la realidad que se quiere conocer.

## Bolivia. Comparación de la desigualdad de ingresos en hogares rurales de bajos y altos ingresos.

La desigualdad en América Latina es siempre un tema de análisis, en los últimos años la economía ha crecido, tanto a nivel regional, como individual de los países, esa riqueza sigue concentrada en un sector de la sociedad, un rasgo que históricamente ha sido común en los países latinoamericanos.

En la comparación de los indicadores de desigualdad con otros países se tiene a nivel país, y el más utilizado es el Gini.

**Cuadro 17.**  
**LATINOAMERICA: Índices de Gini por países, año 2020**

<b>País</b>	<b>Gini</b>
Uruguay	40,2
Argentina	42,3
Bolivia	43,6
Paraguay	43,7
Perú	43,8
Chile	44,9
México	45,4
Nicaragua	46,2
Ecuador	47,3
Honduras	48,2
Guatemala	48,3
Brasil	48,9
Costa Rica	49,3
Panamá	49,8
Colombia	54,2

Fuente: Elaborado con datos del Banco Mundial 2020

Los indicadores utilizados son significativos, en sus escalas de medición coinciden en que la desigualdad es mayor o menor en referencia al ingreso per cápita.

En Bolivia está bien marcado esta desigualdad en el área rural, como ocurre en Pando, Chuquisaca y Potosí que muestran un índice de Gini mayor a 0,5, justificado mayor desigualdad entre la población más pobre, y los de ingresos mayores como son los departamentos de Santa Cruz y La Paz, existe menor desigualdad en la pobreza.

En el área rural existen ventajas y desventajas en el desarrollo económico, por ejemplo, en zonas del departamento de Potosí no se ven beneficiados con la tierra para una actividad agrícola, por lo cual migran a los centros poblados mayores en buscan de ingresos mejores para sus hogares, caso opuesto es en el departamento de Santa Cruz donde la tierra ofrece factores de ventaja en la agricultura.

### CONFLICTO DE INTERESES

Los autores declaran que no hay conflicto de intereses con respecto a la publicación de este documento.

### REFERENCIA BIBLIOGRÁFICA

- Chihara Laura, Hesterberg Tim, Mathematical Statistics with Resampling and R, 2019, Wiley & Sons.
- Carlos Gamero Buron, José Luid Iranzo Acosta, Modelos Probabilísticos Variables Aleatorias Continuas, Departamento de economía Aplicada, Universidad de Málaga.
- Jorge Muro Guerrero, (2019), Análisis de datos censurados: técnicas de estimación e inferencia no paramétricas y paramétricas, Facultad de Ciencias, Universidad de Zaragoza.
- Cochran, William G., (1996), Técnicas de muestreo, Ed. Continental, México.
- Reichmann, W. J., (1965), Uso y abuso de las estadísticas, Deusto, Bilbao.

- Robin Cavagnoud, Sophia Lewandowski y Cecilia Salazar, Pobreza, desigualdades y educación en Bolivia, (2005-2015), Boletín del Institut Français d' Etudes Andines.
- UDAPE, Unidad de Análisis de Políticas Sociales y Económicas, Análisis Económico, Volumen 20 (Junio 2005).
- Chantal Line Carpentier, Richar Kosul y Fabio David Passos., Objetivos 10 – Por qué es importante combatir la desigualdad, NACIONES UNIDAS Crónicas ONU (2012.)
- Minor Mora Salas, Juan Pablo Pérez Saenz, CSO Facultad Latinoamericana de Ciencias Sociales, Costa Rica.  
<http://bibliotecavirtual.clacso.org.ar/ar/libros/costar/flacso/cuad131.pdf>.
- Galindo Mariana y Viridiana Ríos (2015), Desigualdad, ¿cómo vamos? , Estudios Económicos, Vol 1, julio 2015, DF.México-CEPAL.
- Federico Stezano (2021), Enfoques, definiciones y estimaciones de pobreza y desigualdad en América Latina y el Caribe, Documentos de Proyectos NACIONES UNIDAS, CEPAL, FIDA-Invertir en la Población.
- Magdalena Cardemil Winkler (2022), Desigualdad económica, una amenaza para el Desarrollo sostenible, Biblioteca del Congreso Nacional de Chile Serie Minutas No. 03-22, 20/01/2022.
- Francisco H.G. Ferreira, Michael Welton (2004), La desigualdad en América Latina, Rompiendo con la historia?, Serie Desarrollo para todos-BANCO MUNDIAL.
- Silvia Escobar de Pabón, Walter Arteaga Aguilar, Giovanna Hurtado Aponte (2019), Desigualdades y pobreza en Bolivia: una perspectiva multidimensional CEDLA (Ed.) 184 p.
- Claudia C. Montaña Quiroga, Marcel Navia Mendoza (2021), Crecimiento económico, desigualdad y pobreza, evaluación empírica para el caso boliviano, Perspectiva Año 24 No. 48, pp 57-98, Universidad Católica Boliviana “San Pablo”, Unidad académica Regional Cochabamba, clasificación JEL: 015.
- Pobreza y desigualdad, Índice de Gini para el ingreso per cápita mensual según área 2005-2021, Estadísticas Sociales, Encuesta de Hogares, INSTITUTO NACIONAL DE ESTADÍSTICA-BOLIVIA.
- LAC Equity Lab: Desigualdad-Tendencias BANCO MUNDIAL BIRF-AIF, [www.bancomundial.org/es/topic/poverty/lac-equity-lab1/incoe-inequality/inequality-trends](http://www.bancomundial.org/es/topic/poverty/lac-equity-lab1/incoe-inequality/inequality-trends).
- Gerardo Esquivel, profesor en la UNAM y el Colegio de México, Indicadores de desigualdad, Conceptos y evidencia para México, (65D90886-291F-5A39-31A2-F928E9DF45EC).pdf
- Revista de Métodos Cuantitativos para la Economía y la Empresa, Universidad Pablo de Olavide, Sevilla, España.
- Cao y Fernández Casa; (2020), Técnicas de Simulación y Remuestreo, Capítulo 9 Métodos de remuestreo, rubenfcasal.github.io/simbook/bootstrap.html.
- Antonio Solanas, Vicenta Sierra, Bootstrap (1992): Fundamentos e introducción a sus aplicaciones, Facultad de Psicología Universidad de Barcelona.

# ANÁLISIS ESTADÍSTICO MULTIVARIADO EXPLORATORIO DE ESTRUCTURAS TEMPORALES DESDE UNA ÓPTICA STATIS. UNA APLICACIÓN A LA POBREZA URBANA MULTIDIMENSIONAL EN BOLIVIA

## EXPLORATORY MULTIVARIATE STATISTICAL ANALYSIS OF TEMPORAL STRUCTURES FROM A STATIS PERSPECTIVE. AN APPLICATION TO MULTIDIMENSIONAL URBAN POVERTY IN BOLIVIA

Carlos Fernando Silva Viamonte<sup>1</sup>  
Universidad de Salamanca, Salamanca, España,  
✉ [fersilvania@usal.es](mailto:fersilvania@usal.es)

Artículo recibido: 31/07/2023  
Artículo aceptado: 25/09/2023

### RESUMEN

*La medición de la pobreza desde una perspectiva multidimensional ha crecido en términos de aportes metodológicos que llevan a obtener medidas más robustas y confiables, pero principalmente, ya son muchos países que han adoptado la metodología ALKIRE-FOSTER apropiándose de la misma y de sus bondades técnicas. La Estadística Multivariante juega hoy en día un papel importante pues responde a esa necesidad, debido a que brinda un conjunto de técnicas que ayudan a tratar la enorme diversidad de problemáticas que hoy por hoy preocupan a las sociedades en términos de tratamiento y análisis de información objetiva. El presente artículo tiene el objetivo de presentar el STATIS: Structuration de Tableaux A Trois Indices de la Statistique aplicada a variables cuantitativas con el objetivo de encontrar una estructura común o representativa a todas las tablas o grupos de variables, en este caso años y las dimensiones de la pobreza multidimensional urbana por departamento. En el documento se muestra que a través del STATIS se pudo comparar y analizar las relaciones entre los diferentes años y las dimensiones de la pobreza multidimensional, se identificó la estructura intercorrelacional, a partir de la obtención de una tipología en los departamentos. Adicionalmente se observó la evolución de las dimensiones de la pobreza multidimensional en los departamentos a lo largo del periodo estudiado. Finalmente, se describió la estructura interna de las dimensiones estudiadas para los departamentos y su evolución en el tiempo.*

**Palabras clave:** Pobreza Multidimensional, STATIS, Análisis de Componentes Principales, Tablas Multivaria.

### ABSTRACT

*The measurement of poverty from a multidimensional perspective has grown in terms of methodological contributions that lead to obtaining more robust and reliable measures, but mainly, there are already many countries that have adopted the ALKIRE-FOSTER methodology, appropriating it and its technical benefits. Multivariate Statistics plays an important role today as it responds to this need, since it provides a set of techniques that help to deal with the enormous diversity of problems that currently concern societies in terms of treatment and analysis of objective information. This article aims to present the STATIS: Structuring of Tableaux A Trois Indices de la Statistique applied to quantitative variables with the aim of finding a common or representative structure for all tables or groups of variables, in this case years and the dimensions of urban multidimensional poverty by department. The document shows that*

<sup>1</sup> Coordinador de Gerencia General del Banco de Desarrollo Productivo BDP-SAM, Licenciado en Administración de Empresas, Máster en Estadística Aplicada, Máster en Economía, Doctor en Investigación. Estudiante de Doctorado en Estadística Multivariante Aplicada. ORCID: [0000-0001-5117-6565](https://orcid.org/0000-0001-5117-6565).

*through STATIS it was possible to compare and analyze the relationships between the different years and the dimensions of multidimensional poverty, the intercorrelational structure was identified, based on obtaining a typology in the departments. Additionally, the evolution of the dimensions of multidimensional poverty in the departments throughout the period studied was observed. Finally, the internal structure of the dimensions studied for the departments and their evolution over time were described.*

**Key words:** Multidimensional Poverty, STATIS, Principal Component Analysis, Multiway Tables.

---

## 1. INTRODUCCIÓN

El Objetivo 1 de la declaración de los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas, de los 17 desafíos del desarrollo con miras a cumplirlos hasta el 2030, es el Fin de la Pobreza, lo que demuestra que la reducción de la pobreza es uno de los objetivos más importantes de la agenda de políticas públicas del Estado Plurinacional de Bolivia y de muchos países del mundo. Este primer objetivo busca acabar con la pobreza en todas sus formas, en todo el mundo; pero fundamentalmente reconoce el cambio en la perspectiva teórica, conceptual y de medición de la pobreza, pasando de un enfoque univariado a una perspectiva multidimensional (Muñeton Santa et al., 2019).

En ese sentido, la medición de la pobreza desde una perspectiva multidimensional ha crecido en términos de aportes metodológicos que llevan a obtener medidas más robustas y confiables, pero principalmente, ya son muchos países que han adoptado la metodología ALKIRE-FOSTER apropiándose de la misma y sus bondades técnicas. Después del cálculo del índice y sus dimensiones, el siguiente paso es el uso de esta información para realizar un análisis más profundo del comportamiento de la pobreza multidimensional y su evolución a lo largo del tiempo bajo un enfoque distinto al que normalmente se hace (en un momento del tiempo).

Esta información, para convertirse en un

insumo importante, principalmente para el diseño de nuevas políticas públicas que ayuden a mejorar la calidad de vida de las personas, requiere de un análisis que use técnicas, métodos y herramientas más sofisticadas, y que lleven a cabo su cometido sin tener que entrar en importantes pérdidas de información, de confianza, precisión y certeza. La Estadística Multivariante juega hoy en día un papel importante pues responde a esa necesidad, debido a que brinda un conjunto de técnicas que ayudan a tratar la enorme diversidad de problemáticas que hoy por hoy preocupan a las sociedades en términos de tratamiento y análisis de información objetiva y con la ayuda de los enormes avances de la informática, que ha venido a proveer a la Estadística de los recursos más importantes para enfrentar grandes volúmenes de datos.

El presente artículo muestra un resumen de los aspectos más relevantes de una de las técnicas estadísticas desarrolladas para el tratamiento de grandes volúmenes de información como es el STATIS-ACT. En una primera parte, el artículo muestra los aspectos teóricos más relevantes del STATIS, como una técnica estadística de tratamiento de información de tablas múltiples. A continuación, se presenta una aplicación del STATIS a información sobre la pobreza multidimensional a nivel urbano. El objetivo es ilustrar el uso y utilidad de la técnica a través de un análisis multidimensional con el fin de valorar la utilidad de esta técnica multivariante y brindar una herramienta para el tratamiento de fenómenos sociales como es el de la pobreza multidimensional.

## 2. METODOLOGÍA

Los datos provienen del cálculo del índice de pobreza multidimensional y sus dimensiones para el Estado Plurinacional de Bolivia entre los años 2011 y 2019. Para ello se utilizó la metodología ALKIRE-FOSTER basado en las Encuestas de Hogares (EH), instrumento del Instituto Nacional de Estadística (INE), que tiene como objetivo suministrar información sobre las condiciones de vida de los hogares, a partir de la recopilación de información de variables socioeconómicas y demográficas de la población boliviana.

### 2.1 Dimensiones, indicadores y umbrales.

Para el cálculo del índice de pobreza multidimensional a nivel urbano y de acuerdo a la metodología Alkire-Foster se

inició con una etapa de identificación de la pobreza a nivel de hogar. A continuación, se definieron las dimensiones y sus respectivas ponderaciones, mismas que están dadas de acuerdo a dos premisas. Primero, una revisión exhaustiva de las experiencias de otros países y las recomendaciones de la OPHI, y segundo, la capacidad de la encuesta de dar la información necesaria para la construcción del indicador en los 9 años de estudio. El cálculo de la medida de pobreza multidimensional asume un conjunto de asunciones de carácter normativo acerca de las dimensiones, los indicadores, las líneas de corte y los pesos a usarse. La Tabla 1, muestra las dimensiones con sus respectivos indicadores y ponderación y la condición de privación que define a cada indicador que fueron utilizados para su respectivo cálculo.

**Tabla 1**  
**Dimensiones e Indicadores de la Pobreza Multidimensional propuesta.**

Dimensiones	Indicadores	Condición de Privación	Var
<b>1. Pobreza Monetaria. (1/4)</b>	1. Pobreza monetaria. (1/4)	La persona es pobre por ingresos.	PobY
	1. Analfabetismo. (1/12)	Persona de quince años o más que no saben leer ni escribir.	Analf
<b>2. Educación. (1/4)</b>	2. Logro educativo. (1/12)	Persona que no ha completado 6 años de escolaridad.	Anose
	3. Inasistencia. (1/12)	Niños y adolescentes no matriculados en edad escolar.	Nmati
<b>3. Salud. (1/4)</b>	1. Seguro médico. (1/4)	Personas que no cuentan con un seguro médico.	Sseg
	1. Electricidad. (1/28)	El hogar no cuenta con electricidad para alumbrar su casa.	Elect
	2. Saneamiento. (1/28)	El hogar no cuenta con saneamiento básico en su casa.	Sanea
<b>4. Entorno y calidad de vida. (1/4)</b>	3. Agua potable. (1/28)	El hogar no cuenta con agua potable proveniente de una red de suministro o pileta pública.	Agua
	4. combustible para cocinar. (1/28)	El hogar usa combustible “contaminante” para cocinar (estiércol, leña o carbón )	Cocina
	5. Pisos de la vivienda. (1/28)	El hogar tiene piso de tierra o arena	Piso
	6. Paredes de la vivienda. (1/28)	El material de las paredes de la vivienda (internamente y externamente) no tienen revoque y son de material adobe, tapial, piedra, madera, caña, palma y tronco expuesto.	Pared
	7. Techos de la vivienda. (1/28)	El material del techo de la vivienda es de paja, caña, palma o barro.	Techo

Fuente: Elaboración propia.

El siguiente paso fue la determinación del punto de corte denotado por  $\alpha$ , que señala cuándo un hogar o una persona tiene suficientes privaciones para ser considerado multidimensionalmente pobre. Hasta el momento no se ha establecido un método que defina el valor específico que debe tomar el parámetro  $\alpha$  (Muñeton Santa et al., 2019, p. 121). La elección de  $\alpha$  debería ser un política, es decir, debería definirse de manera normativa (Alkire & Foster, 2011). Siguiendo a (Alkire & Santos, 2014) un  $\alpha$  como una estimación internacional y debido a que existe una justificación normativa para el punto de corte de la pobreza con el fin de agrupar aquellas personas que no alcanzan el umbral mínimo.

## 2.2 Datos.

En ese sentido, se tiene los valores para cada uno de los años en términos del índice de pobreza multidimensional a nivel departamental y los valores de cada una de las dimensiones para la serie 2011 a 2019. Para los 9 departamentos, Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando (individuos) evaluados en 9 años en términos de 12 indicadores para 4 dimensiones (como se muestra en la Tabla 1).

## 2.3 Objetivos del estudio.

El objetivo fundamental es explorar la información para comprender las estructuras subyacentes a los datos, especialmente en términos de las relaciones entre las diferentes dimensiones de la pobreza multidimensional a nivel urbano y cómo estas explican su estructura de la misma, para cada departamento, así como su dinamismo a lo largo del periodo de tiempo de estudio. De esta forma se plantean los siguientes objetivos:

- Comparar y analizar las relaciones

entre los diferentes años y las dimensiones de la pobreza multidimensional.

- Identificar a los departamentos a partir de las diferentes dimensiones que definen la pobreza multidimensional, así como su estructura intercorrelacional, a partir de la obtención de una tipología en los departamentos.
- Estudiar y describir la evolución de las dimensiones de la pobreza multidimensional en los departamentos a lo largo del periodo estudiado.
- Explorar y describir la estructura interna de las dimensiones de la pobreza multidimensional en los departamentos, y su evolución en el tiempo.
- Explorar la existencia de “factores comunes” en los distintos años y departamentos, como también la existencia de “factores específicos”.

## 2.4 Métodos estadísticos multivariantes: STATIS-ACT.

El método STATIS del francés “Structuration des Tableaux À Trois Indices de la Statistique o Estructuración de tablas estadísticas de tres vías” es una técnica que también se conoce como ACT que significa en francés “Analyse Conjointe de Tableaux” (análisis conjunto de tablas) (Lavit et al., 1994) que en realidad es una generalización del análisis de componentes principales (PCA) cuyo objetivo es analizar varios conjuntos de datos de variables recopiladas en el mismo conjunto de observaciones o, como en su versión dual llamada dual-STATIS, varios conjuntos de observaciones medidas en el mismo conjunto

de variables (Abdi et al., 2012). Como tal, STATIS es parte de la multitabla (también llamada multibloque o análisis de consenso).

STATIS fue propuesto a partir del trabajo de Escoufier (Lavit et al., 1994) y fue descrito por primera vez por L'Hermier des Plantes. Los principales objetivos de esta técnica son (1) comparar y analizar las relaciones entre los diferentes conjuntos de datos, (2) integrar estos conjuntos de datos en otro conjunto con pesos promedio ponderado llamado compromiso (consenso) que luego se analiza a través de PCA para revelar la estructura común entre las observaciones y (3) para proyectar cada uno de los conjuntos de datos originales en el compromiso y de esa manera analizar las comunalidades y discrepancias existentes.

La metodología descriptiva de datos STATIS se usa frecuentemente para analizar múltiples conjuntos de variables medidas para las mismas observaciones y se ha utilizado recientemente en varios dominios, como la investigación sensorial y de ciencia del consumidor, quimiometría y monitoreo de procesos, ecología, visión por computadora, hidrología, ciencia de la información, neuroimágenes, medicina, control estadístico de calidad y biología molecular (Abdi et al., 2012).

#### ***2.4.1 Cuando se puede usar el método.***

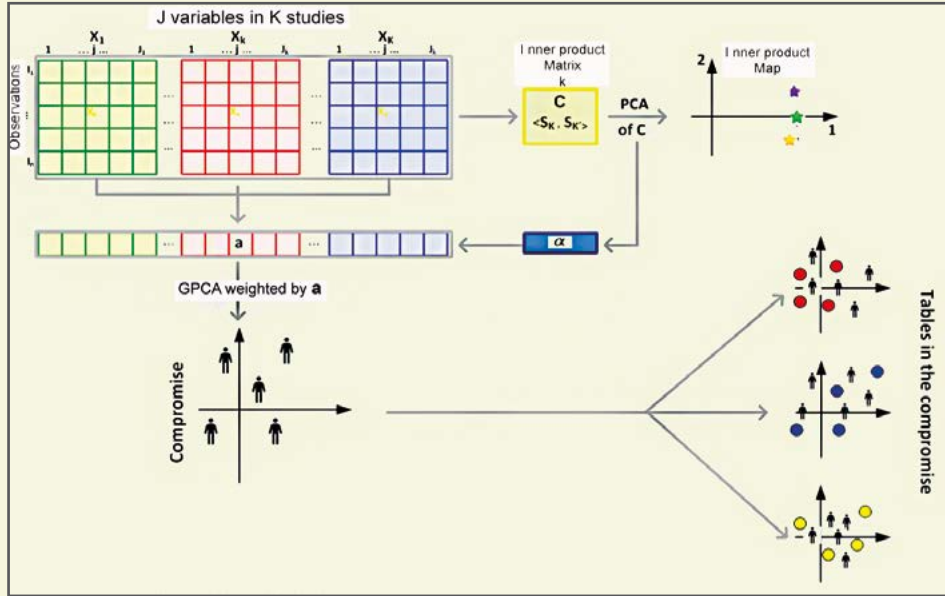
Este método se utiliza cuando se tiene la medición de varios conjuntos de variables en el mismo conjunto de observaciones. El número y/o la naturaleza de las variables utilizadas para describir las observaciones puede variar de un conjunto de variables a otro, pero las observaciones deben ser las mismas en todos los conjuntos de datos (Abdi et al., 2012). Por tanto, el análisis, mostrará una evaluación acerca de las posiciones de

las observaciones y sus cambios a lo largo del tiempo.

La idea central detrás de STATIS es analizar la estructura de los diferentes conjuntos de datos individuales (es decir, la relación entre los conjuntos de datos individuales) y derivar de esta estructura un conjunto óptimo de ponderaciones para calcular la mejor representación común de las observaciones. Para calcular este compromiso, los elementos de cada tabla se multiplican por el peso óptimo de esta tabla y el compromiso se obtiene sumando estas tablas K “ponderadas”. Estos pesos se eligen de modo que el compromiso proporcione la mejor representación (en un sentido de mínimos cuadrados) de todo el conjunto de tablas. El ACP del compromiso descompone la varianza del compromiso en un conjunto de nuevas variables ortogonales denominadas componentes principales ordenadas por la cantidad de varianza que explica cada componente. Las coordenadas de las observaciones en los componentes se denominan puntajes de factores y se pueden usar para trazar mapas de las observaciones en las que se representan como puntos de modo que las distancias en el mapa reflejen mejor las similitudes entre las observaciones. La posición de las observaciones de cada conjunto de datos también se puede representar como puntos en el compromiso. Como los componentes se obtienen combinando las variables originales, cada variable aporta una cierta cantidad a cada componente. Esta cantidad, llamada carga de una variable sobre un componente, refleja la importancia de esa variable para este componente y también se puede utilizar para trazar mapas de las variables que reflejan su asociación. Finalmente, como subproducto del cálculo de los pesos óptimos, los conjuntos de datos también se pueden representar como puntos en un espacio multidimensional. En la Figura 1 se proporciona un esquema de la técnica.



Figura 1  
Los diferentes pasos de STATIS (Abdi et al., 2012).



Fuente: Elaboración propia

#### 2.4.2 Los pasos del STATIS.

El método STATIS consta de dos pasos principales: primero se analiza la estructura de similitud del conjunto de tablas, etapa denominada análisis de la interestructura de las tablas (Lavit et al., 1994). Esto proporciona un conjunto de ponderaciones óptimas que se utilizan en el segundo paso. Este paso realiza un ACP generalizado (es decir, un GSVD) de  $X$  que integra los pesos como restricciones en las tablas y sus variables, este paso se denomina análisis de la intraestructura de las tablas.

##### Paso uno: análisis de la estructura entre tablas.

Se analiza la estructura de similitud del conjunto de las  $K$  tablas. La similitud entre dos matrices de datos  $X_{[k]}$  y  $X_{[k']}$  se denota. Esta similitud se evalúa transformando primero las matrices en matrices de productos cruzados  $I$  por  $I$  denominadas  $S_{[k]}$  y  $S_{[k']}$ , que se calculan como:

$$S_{[k]} = X_{[k]} X_{[k]}^T \text{ y } S_{[k']} = X_{[k']} X_{[k']}^T$$

La medida de similitud (similar a un coeficiente de correlación entre matrices cuadradas) se llama producto interior de Hilbert-Schmidt, producto interno o llamado también producto escalar, vector de covarianzas COVV o producto Frobenius de las matrices  $S_{[k]}$  y  $S_{[k']}$  y se denota por:

$$\begin{aligned} c_{k,k'} &= \langle S_{[k]}, S_{[k']} \rangle \\ &= \text{trace} \{ X_{[k]} X_{[k]}^T \times X_{[k']} X_{[k']}^T \} \\ &= \text{trace} \{ S_{[k]} \times S_{[k']} \} \\ &= \text{vec} \{ S_{[k]} \}^T \times \text{vec} \{ S_{[k']} \} \\ &= \sum_i \sum_j s_{i,j,k} s_{i,j,k'} \end{aligned}$$

Geoméricamente, el producto interno también se puede interpretar como un producto escalar entre dos matrices semidefinidas positivas  $S_{[k]}$  y  $S_{[k']}$  y, por lo tanto, es proporcional al coseno entre las matrices (porque el conjunto de matrices semidefinidas positivas es un espacio vectorial). El producto interno también se usa para definir la norma de una matriz como la raíz cuadrada del producto interno de esta

matriz consigo misma. Formalmente, la norma de una matriz de productos cruzados se denota  $\|S_{[k]}\|$  y se define como:

$$\|S_{[k]}\|^2 = \langle S_{[k]}, S_{[k]} \rangle$$

Debido a que todas las matrices  $S_{[k]}$  son semidefinidas positivas, el producto interno entre dos de estas matrices siempre es igual o mayor que cero. Cuando las matrices  $S_{[k]}$  y  $S_{[k']}$  se normalizan de manera que la suma de los cuadrados de sus elementos es uno, el producto interno entre estas dos matrices es igual al coseno entre estas matrices. Este coseno también se conoce como el coeficiente RV introducido por Escoufier como una medida de similitud entre matrices simétricas cuadradas (Abdi, n.d.).

Los productos internos entre todas las matrices  $K$  se recopilan en una matriz de productos internos  $K$  por  $K$  denominada  $C$ , donde  $c_{k,k'}$  da el valor del producto interno entre las tablas  $k$  y  $k'$ . La matriz  $C$  también se puede calcular desde  $k$  por  $I^2$  de la matriz  $Z$  definida como:

$$Z = [\text{vec}\{S_{[1]}\} | \dots | \text{vec}\{S_{[k]}\} | \dots | \text{vec}\{S_{[K]}\}]^T$$

Con esta matriz  $Z$ , calculamos  $C$  como:

$$C = ZZ^T$$

Esto demuestra que  $C$  es una matriz semidefinida positiva y que, por lo tanto, sus valores propios son todos positivos o nulos y sus vectores propios son reales y ortogonales entre sí. En consecuencia,  $C$  se puede descomponer en forma propia como:

$$C = U\Theta U^T \text{ con } U^T U = I$$

Entonces, la descomposición propia de  $C$  proporciona la PCA de la estructura de similitud entre las tablas (guardadas en  $X$ ) que ahora pueden representarse como puntos en un mapa PCA usando como coordenadas sus puntajes factoriales que se calculan como:

$$G = U\Theta^{\frac{1}{2}}$$

Tenga en cuenta que los vectores propios de  $C$  también podrían haberse obtenido de la SVD de  $Z$  como:

$$Z = U\Gamma V^T$$

Este enfoque, llamado RV-PCA, proporciona la estructura de similitud de las tablas e información sobre los valores de los pares de observaciones que contribuyen al patrón de similitud medido por  $\langle S_{[k]}, S_{[k']} \rangle$ .

Además de proporcionar una representación visual de la estructura de similitud, la descomposición propia de la matriz  $C$  también proporciona pesos óptimos para combinar las tablas en un compromiso. La descomposición propia de la matriz  $C$  siempre da un primer vector propio cuyos elementos tienen el mismo signo (que, por conveniencia, se considera positivo). Esta propiedad es una consecuencia del teorema de Perron-Frobenius que establece que las matrices semidefinidas positivas cuyos elementos son todos positivos siempre tienen un primer vector propio con todos los elementos que tienen el mismo signo.

De hecho, el valor de una tabla para el primer vector propio refleja su similitud general con todas las demás tablas (por ejemplo, su "comunidad"). Esto sugiere utilizar los valores de este primer vector propio para ponderar las tablas con el fin de dar más importancia a las tablas que representan bien al grupo y menos importancia a las tablas idiosincrásicas. Se puede demostrar, que el proceso proporciona una representación óptima del conjunto de tablas. Entonces, si el primer vector propio de  $C$  se denota  $u_1$ , el conjunto de pesos óptimos para las tablas se almacena en un vector  $K$  por 1 denotado y se calcula reescalando  $u_1$  de modo que la suma de los elementos de sea igual a uno:

$$\alpha = \mathbf{u} \mathbf{1} \times (\mathbf{u}^T \mathbf{1})^{-1}$$

Por conveniencia, los pesos  $\alpha$  se pueden agrupar en un vector  $\mathbf{J}$  por  $\mathbf{1}$  denotado por  $\mathbf{a}$  donde cada variable se le asigna el peso de la matriz a la que pertenece. Específicamente,  $\mathbf{a}$  se construye como:

$$\mathbf{a} = [\alpha_1 \mathbf{1}_{[1]}^T, \dots, \alpha_k \mathbf{1}_{[k]}^T, \dots, \alpha_K \mathbf{1}_{[K]}^T]$$

donde  $\mathbf{1}_{[k]}$  representa un vector  $\mathbf{J}_{[k]}$  de unos. Alternativamente, los pesos se pueden almacenar como elementos diagonales de una matriz diagonal denominada  $\mathbf{A}$  obtenida como:

$$\mathbf{A} = \text{diag}\{\mathbf{a}\} = \text{diag}\{[\alpha_1 \mathbf{1}_{[1]}^T, \dots, \alpha_k \mathbf{1}_{[k]}^T, \dots, \alpha_K \mathbf{1}_{[K]}^T]\}$$

**Paso dos: PCA generalizado GSVD de X.**

Una vez recopilados los pesos, se utilizan para calcular el GSVD de  $\mathbf{X}$  bajo las restricciones proporcionadas por  $\mathbf{M}$  (masas para las observaciones) y  $\mathbf{A}$  (pesos óptimos para las tablas  $\mathbf{K}$ ). Este GSVD se expresa como:

$$\mathbf{X} = \mathbf{P} \mathbf{A} \mathbf{Q}^T \text{ with } \mathbf{P}^T \mathbf{M} \mathbf{P} = \mathbf{Q}^T \mathbf{A} \mathbf{Q} = \mathbf{I}$$

Este GSVD corresponde a un GPCA de la matriz  $\mathbf{X}$  y, en consecuencia, proporcionará puntuaciones factoriales para describir las observaciones y cargas factoriales para describir las variables. Cada columna de  $\mathbf{P}$  y  $\mathbf{Q}$  se refiere a un componente principal también llamado dimensión (porque los números en estas columnas a menudo se usan como coordenadas para trazar mapas) (Abdi & Williams, 2010). En PCA, la anterior ecuación se puede reescribir como:

$$\mathbf{X} = \mathbf{F} \mathbf{Q}^T \text{ con } \mathbf{F} = \mathbf{P} \mathbf{A}$$

donde  $\mathbf{F}$  almacena las puntuaciones de los factores (que describen las observaciones) y  $\mathbf{Q}$  almacena las cargas (que describen las variables). En la notación de triple, STATIS es equivalente al análisis estadístico triple ( $\mathbf{X}$ ,

$\mathbf{A}$ ,  $\mathbf{M}$ ). Debido a que la matriz  $\mathbf{X}$  comprende  $\mathbf{K}$  tablas, cada una de las cuales comprende  $\mathbf{J}_{[k]}$  variables, la matriz  $\mathbf{Q}$  de los vectores singulares de la izquierda se puede dividir de la misma manera que  $\mathbf{X}$ . Específicamente,  $\mathbf{Q}$  se puede expresar como un bloque de columnas de la matriz como:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{[1]} \\ \vdots \\ \mathbf{Q}_{[k]} \\ \vdots \\ \mathbf{Q}_{[K]} \end{bmatrix} = [\mathbf{Q}_{[1]}^T | \dots | \mathbf{Q}_{[k]}^T | \dots | \mathbf{Q}_{[K]}^T]^T$$

donde  $\mathbf{Q}_{[k]}$  es una matriz  $\mathbf{J}_{[k]}$  por  $\mathbf{L}$  (siendo  $\mathbf{L}$  el rango de  $\mathbf{X}$ ) que almacena los vectores singulares derechos correspondientes a las variables de la matriz  $\mathbf{X}_{[k]}$ . De esta manera se tiene:

$$\begin{aligned} \mathbf{X} &= [\mathbf{X}_{[1]} | \dots | \mathbf{X}_{[k]} | \dots | \mathbf{X}_{[K]}] = \mathbf{P} \mathbf{A} \mathbf{Q}^T \\ &= \mathbf{P} \mathbf{A} ([\mathbf{Q}_{[1]}^T | \dots | \mathbf{Q}_{[k]}^T | \dots | \mathbf{Q}_{[K]}^T])^T \\ &= \mathbf{P} \mathbf{A} [\mathbf{Q}_{[1]}^T | \dots | \mathbf{Q}_{[k]}^T | \dots | \mathbf{Q}_{[K]}^T] \\ &= [\mathbf{P} \mathbf{A} \mathbf{Q}_{[1]}^T | \dots | \mathbf{P} \mathbf{A} \mathbf{Q}_{[k]}^T | \dots | \mathbf{P} \mathbf{A} \mathbf{Q}_{[K]}^T] \end{aligned}$$

Si definimos  $\mathbf{A}_{[k]}$  como:

$$\mathbf{A}_{[k]} = \alpha_k \mathbf{I},$$

de tenemos, en general,  $\mathbf{Q}^T$

$$\mathbf{Q}_{[k]}^T \mathbf{A}_{[k]} \mathbf{Q}_{[k]} \neq \mathbf{I}$$

**Puntuaciones factoriales.** Las puntuaciones factoriales para  $\mathbf{X}$  representan el mejor compromiso, la mejor representación común para el conjunto de matrices  $\mathbf{K}$  y se calculan como:

$$\mathbf{F} = \mathbf{P}$$

Las puntuaciones de los factores se pueden usar para trazar las observaciones como se hace en PCA estándar para el cual cada columna de  $\mathbf{F}$  representa una dimensión. Tenga en cuenta que la varianza de las puntuaciones de los factores de las observaciones se calcula

## Análisis estadístico multivariado exploratorio de estructuras temporales desde una óptica STATIS. Una aplicación a la pobreza urbana multidimensional en Bolivia

usando sus masas (almacenadas en la matriz  $\mathbf{M}$ ) y se puede encontrar como la diagonal de la matriz  $\mathbf{F}^T\mathbf{M}\mathbf{F}$ . Esta varianza es igual, para cada dimensión, al cuadrado del valor singular de esta dimensión como se muestra por:

$$\mathbf{F}^T\mathbf{M}\mathbf{F} = \mathbf{\Delta P}^T\mathbf{M}\mathbf{P}\mathbf{\Delta} = \mathbf{\Delta}^2$$

Como en el PCA estándar,  $\mathbf{F}$  se puede obtener de  $\mathbf{X}$ :

$$\mathbf{F} = \mathbf{P}\mathbf{\Delta} = \mathbf{X}\mathbf{A}\mathbf{Q}$$

Teniendo en cuenta la estructura de bloques de  $\mathbf{X}$ ,  $\mathbf{A}$  y  $\mathbf{Q}$ , también se puede escribir:

$$\mathbf{F} = \mathbf{X}\mathbf{A}\mathbf{Q} = \begin{bmatrix} | & & | \\ \mathbf{X}_{[1]} & \dots & \mathbf{X}_{[k]} & \dots & \mathbf{X}_{[K]} \\ | & & | \end{bmatrix} \times \mathbf{A} \times \begin{bmatrix} \mathbf{Q}_{[1]} \\ \vdots \\ \mathbf{Q}_{[k]} \\ \vdots \\ \mathbf{Q}_{[K]} \end{bmatrix}$$

$$= \sum_k \mathbf{X}_{[k]} \mathbf{A}_{[k]} \mathbf{Q}_{[k]} = \sum_k \alpha_k \mathbf{X}_{[k]} \mathbf{Q}_{[k]}$$

Esta ecuación sugiere que las puntuaciones de los factores parciales de una tabla pueden definirse como la proyección de esta tabla sobre sus vectores singulares derechos (es decir,  $\mathbf{Q}_{[k]}$ ). Específicamente, las puntuaciones de los factores parciales para el k-ésimo estudio se almacenan en una matriz denominada  $\mathbf{F}_{[k]}$  calculada como:

$$\mathbf{F}_{[k]} = \mathbf{X}_{[k]} \mathbf{Q}_{[k]}$$

Nótese que la matriz de puntajes de los factores de compromiso es el baricentro de los puntajes de los factores parciales porque es el promedio ponderado (donde los pesos están dados por los  $\alpha_k$ ) de los puntajes de los factores parciales.

$$\sum_k \alpha_k \mathbf{F}_{[k]} = \sum_k \alpha_k \mathbf{X}_{[k]} \mathbf{Q}_{[k]} = \mathbf{F}$$

También como en el PCA estándar, los

elementos de  $\mathbf{Q}$  son cargas y se pueden graficar solos o junto con las puntuaciones de los factores como un biplot (Greenacre, 2010). Como las cargas vienen en bloques (es decir, las cargas corresponden a las variables de una tabla), es práctico crear un biplot con las puntuaciones de los factores parciales (es decir,  $\mathbf{F}_{[k]}$ ) para un bloque y las cargas (es decir,  $\mathbf{Q}_{[k]}$ ) para este

bloque. Al hacerlo, a menudo es práctico normalizar las cargas de modo que su varianza sea conmensurable con la varianza de las puntuaciones de los factores. Esto se puede lograr, por ejemplo, normalizando, para cada dimensión, las cargas de un bloque de modo que su varianza sea igual al cuadrado del valor singular de la dimensión o incluso al valor singular mismo. Estos biplots son útiles para comprender la estructura estadística de cada bloque, aunque las posiciones relativas de las puntuaciones de los factores y las cargas no se pueden interpretar directamente porque solo las proyecciones de las observaciones en los vectores de carga se pueden interpretar significativamente en un biplot (Greenacre, 2010). Estas correlaciones se representan como mapas bidimensionales en los que también se traza un círculo de radio uno (llamado círculo de correlación) (Abdi & Williams, 2010). Cuanto más cerca del círculo esté una variable, mejor se “explicará” esta variable mediante los componentes utilizados para crear el gráfico (Abdi & Valentin, n.d.).

**Contribución de las observaciones, variables y tablas a una dimensión.** En STATIS, al igual que en PCA estándar, la importancia de una dimensión (es decir, el componente principal) se refleja en su valor propio, que indica cuánto de la inercia total (es decir, la varianza) de los datos son explicados por este componente.

Para comprender mejor las relaciones entre componentes y observaciones, variables y tablas, y también para ayudar a interpretar un componente, podemos evaluar cuánto contribuye una observación, una variable o una tabla completa a la inercia extraída por un componente. Para hacerlo, calculamos estadísticas descriptivas, llamadas contribuciones (Abdi & Williams, 2010). La estabilidad de estas estadísticas descriptivas también se puede evaluar mediante técnicas de validación cruzada, como Bootstrap, y este enfoque se puede usar para seleccionar los elementos relevantes para una dimensión.

**Contribución de una observación a una dimensión.** La varianza de las puntuaciones de los factores para una dimensión dada es igual a su valor propio (es decir, el cuadrado del valor singular) asociado con esta dimensión. Si denotamos  $\lambda_\ell$ , el valor propio de una dimensión dada, podemos reescribir la ecuación como:

$$\lambda_\ell = \sum_i m_i \times f_{i\ell}^2$$

(donde  $m_i$  y  $f_{i\ell}$  son, respectivamente, la masa de la  $i$ -ésima observación y la puntuación factorial de la  $i$ -ésima observación para la  $\ell$ -ésima dimensión).

Como todos los términos  $m_i \times f_{i\ell}^2$  son positivos o nulos, podemos evaluar la contribución de una observación a una dimensión como la relación entre la puntuación del factor ponderado al cuadrado y el valor propio de la dimensión. Formalmente, la contribución de la observación  $i$  al componente  $\ell$ , denotado  $\text{ctr}_{i,\ell}$ , se calcula como:

$$\text{ctr}_{i,\ell} = \frac{m_i \times f_{i\ell}^2}{\lambda_\ell}$$

Las contribuciones toman valores entre 0 y 1, y para un componente dado, la suma de las contribuciones de todas las observaciones es

igual a 1. Cuanto mayor es una contribución, más contribuye la observación al componente. Una heurística útil es basar la interpretación de un componente en las observaciones que tienen contribuciones mayores que la contribución promedio. Las observaciones con contribuciones altas y cuyos puntajes factoriales tienen signos diferentes pueden luego contrastarse para ayudar a interpretar el componente.

Alternativamente podemos usar relaciones de arranque para tener en cuenta las observaciones con relaciones de arranque grandes para una dimensión dada.

**Contribuciones de una variable a una dimensión.** De manera similar a las observaciones, podemos encontrar las variables importantes para una dimensión determinada calculando las contribuciones de las variables. La varianza de las cargas para las variables es igual a uno cuando se tienen en cuenta los pesos  $a_j$ . Entonces, si denotamos por  $a_j$  el peso alfa para la  $j$ -ésima variable, tenemos:

$$1 = \sum_j a_j \times q_{j,\ell}^2$$

(donde  $q_{j,\ell}$  es la carga de la variable  $j$ -ésima para la dimensión- $\ell$ ). Como todos los términos  $a_j \times q_{j,\ell}^2$  son positivos o nulos, se puede evaluar la contribución de una variable a una dimensión como su carga ponderada al cuadrado para esta dimensión. Formalmente, la contribución de la variable  $j$  al componente  $\ell$ , denotado por  $\text{ctr}_{j,\ell}$  se calcula como:

$$\text{ctr}_{j,\ell} = a_j \times q_{j,\ell}^2$$

Las contribuciones de las variables toman valores entre 0 y 1, y para un componente dado, las contribuciones de todas las variables suman 1. Cuanto mayor es la contribución de una variable a un componente, más

contribuye esta variable a este componente. Las variables con contribuciones altas y cuyas cargas tienen diferentes signos se pueden contrastar para ayudar a interpretar el componente.

**Contribución de una tabla a una dimensión.** La noción de contribución de una tabla es específica del análisis multibloque. Como una tabla comprende varias variables, la contribución de una tabla puede definirse simplemente como la suma de las contribuciones de sus variables (una simple consecuencia del teorema de Pitágoras que establece que las longitudes al cuadrado son aditivas). Entonces, la contribución de la tabla  $k$  al componente  $l$  se denota como  $ctr_{k,l}$  y se define como:

$$ctr_{k,l} = \sum_j^{J_{(k)}} ctr_{j,l}$$

Las contribuciones de la tabla toman valores entre 0 y 1, y para un componente dado, las contribuciones de todas las tablas suman 1. Cuanto mayor sea la contribución de una tabla a un componente, más contribuirá esta tabla a este componente. Las contribuciones de las tablas para una dimensión dada suman uno. Un enfoque alternativo vuelve a escalar las contribuciones para que la suma de las contribuciones de una dimensión sea ahora igual al valor propio de esta dimensión. Estas contribuciones reescaladas se denominan inercias parciales y se denotan como  $in_{k,l}$ . Las inercias parciales se obtienen a partir de las contribuciones multiplicando las contribuciones de una dimensión por el valor propio de la dimensión.

### 3. ANÁLISIS DE RESULTADOS

Dada la estructura de Tabla Múltiple de los datos disponibles, se aplicó la metodología STATIS a la base compuesta por las mediciones sobre pobreza multidimensional

urbana y sus dimensiones calculadas en base a la metodología Alkire-Foster, para este cometido se utilizó el programa MultBiplot una herramienta de análisis multivariante usando Biplots desarrollada por la Universidad de Salamanca (Vicente Villardon, 2015). En lo que sigue se presentan, analizan y discuten los principales resultados obtenidos.

En primer lugar, se verá la evolución global (interestructura) donde se comparan entre sí las 9 matrices que representan a cada año y se calcula la matriz compromiso (punto fundamental en la infraestructura), finalmente, se analiza la posición relativa media de los departamentos y sus trayectorias durante el periodo de estudio.

Inicialmente se comienza con el análisis de la interestructura con los datos del índice de pobreza multidimensional urbano y sus dimensiones para los nueve años escogidos. La tabla 2 muestra la matriz de correlación vectorial, como resultado del coeficiente RV. Se observa que la máxima correlación (0,959) aparece entre los años 2014 y 2015 lo que significa que la distribución del índice entre los departamentos es estable entre un año y otro. La mínima correlación vectorial (0,787) se presenta entre los años 2013 y 2019 reflejando una distribución dispereja entre los índices para cada departamento entre estos años. Se observa los productos escalares normados en la tabla 2 se aproximan a 1, lo que nos indica que existe una estructura común de los individuos (departamentos) en el interior de las tablas, que las posiciones de los individuos son estables y que están correctamente descritas por el compromiso.

**Tabla 2**  
**Matriz de correlación vectorial entre bloques (años).**

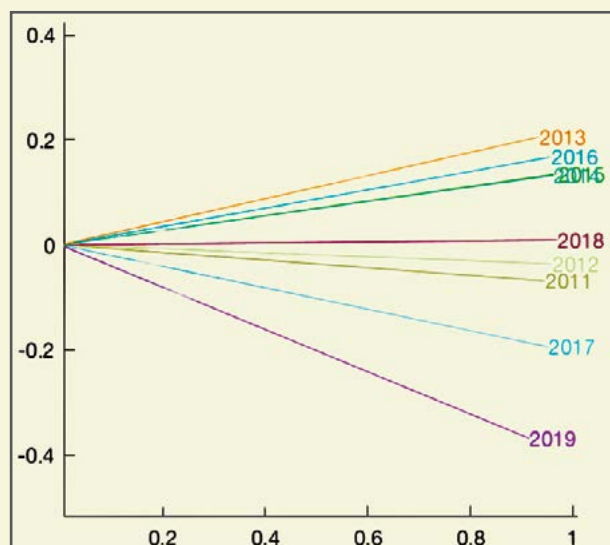
Años	2019	2018	2017	2016	2015	2014	2013	2012	2011
2019	1								
2018	0.876	1							
2017	0.929	0.913	1						
2016	0.830	0.926	0.886	1					
2015	0.840	0.932	0.899	0.952	1				
2014	0.829	0.903	0.900	0.943	0.959	1			
2013	0.787	0.915	0.854	0.916	0.921	0.919	1		
2012	0.881	0.924	0.900	0.880	0.944	0.923	0.873	1	
2011	0.872	0.922	0.884	0.878	0.893	0.897	0.865	0.900	1

Fuente: Elaboración propia.

En el plano de la interestructura (ver figura 2) se evidencian estos resultados, viéndose que los dos puntos más alejados corresponden a los años 2013 y 2019, y los que se encuentran más cercanos a los años 2014 y 2015. Particularmente, la representación euclídea de la figura 2 obtenida de los valores

propios de la matriz de correlación, capta el 94,169% de la inercia total de los años de estudio ( $\lambda_1=8.177$  y  $\lambda_2=0.299$ ). El año 2019 presenta un coseno al cuadrado (correlación con el eje 1) menor respecto de otros años, es decir, una menor correlación con el eje 1 correspondiente a la interestructura.

**Figura 2**  
**Plano de la interestructura.**



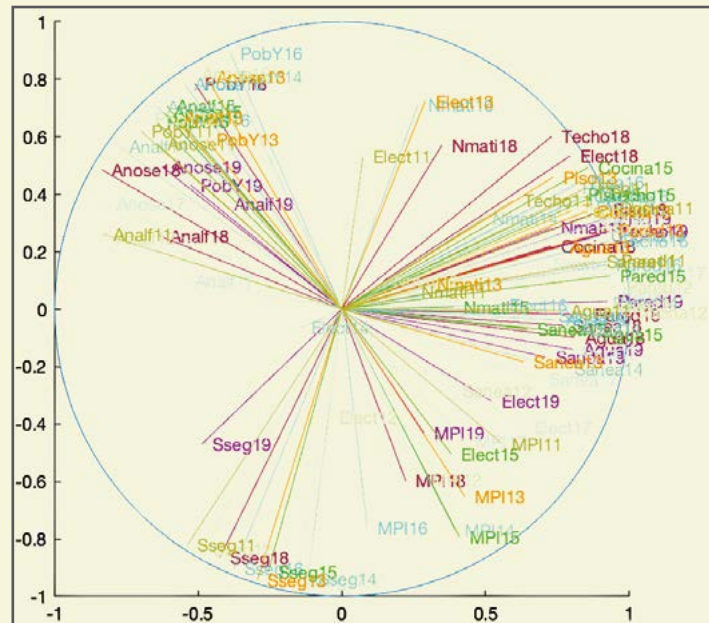
Fuente: Elaboración propia

En el caso de la intraestructura, la figura 3 muestra las proyecciones de las variables en cada uno de los años sobre el plano formado por las primeras componentes del compromiso,

la interpretación se facilita con la tabla 3 que contiene los coeficientes de correlación de las variables con cada uno de los ejes.

# Análisis estadístico multivariado exploratorio de estructuras temporales desde una óptica STATIS. Una aplicación a la pobreza urbana multidimensional en Bolivia

**Figura 3**  
Correlaciones entre variables y componentes de compromiso



Fuente: Elaboración propia

La interpretación de la figura 3 se facilita con la tabla 3 que contiene los coeficientes de correlación de las variables con los ejes factoriales. El primer eje explica el 90.852% de la varianza, dicho eje tiene una fuerte

correlación con la dimensión Entorno y calidad de vida, correlación negativa con el indicador Logro educativo, Analfabetismo y Pobreza monetaria.

**Tabla 3**  
Correlaciones entre variables observadas y factores de consenso.

Variable	MPI19	PobY19	Analf19	Anose19	Nmati19	Sseg19	Elect19	Sanea19	Agua19	Cocina19	Piso19	Pared19	Techo19
Axis 1	0,286	-0,526	-0,41	-0,625	0,728	-0,487	0,522	0,709	0,807	0,901	0,924	0,925	0,933
Axis 2	-0,432	0,432	0,368	0,495	0,284	-0,469	-0,321	-0,167	-0,141	0,259	0,318	0,027	0,272
Variable	MPI18	PobY18	Analf18	Anose18	Nmati18	Sseg18	Elect18	Sanea18	Agua18	Cocina18	Piso18	Pared18	Techo18
Axis 1	0,222	-0,511	-0,633	-0,835	0,349	-0,424	0,795	0,77	0,812	0,728	0,911	0,846	0,731
Axis 2	-0,598	0,785	0,254	0,485	0,571	-0,863	0,533	-0,057	-0,098	0,219	0,346	-0,019	0,602
Variable	MPI17	PobY17	Analf17	Anose17	Nmati17	Sseg17	Elect17	Sanea17	Agua17	Cocina17	Piso17	Pared17	Techo17
Axis 1	0,429	-0,608	-0,545	-0,819	0,701	-0,156	0,634	0,686	0,787	0,973	0,916	0,939	0,854
Axis 2	-0,46	0,666	0,098	0,368	0,158	-0,945	-0,411	-0,242	-0,152	0,129	0,313	0,055	0,322
Variable	MPI16	PobY16	Analf16	Anose16	Nmati16	Sseg16	Elect16	Sanea16	Agua16	Cocina16	Piso16	Pared16	Techo16
Axis 1	0,089	-0,389	-0,56	-0,547	0,265	-0,375	0,549	0,719	0,745	0,844	0,831	0,918	0,932
Axis 2	-0,761	0,892	0,658	0,784	0,711	-0,903	0,015	-0,023	-0,036	0,393	0,445	0,154	0,237
Variable	MPI15	PobY15	Analf15	Anose15	Nmati15	Sseg15	Elect15	Sanea15	Agua15	Cocina15	Piso15	Pared15	Techo15
Axis 1	0,41	-0,646	-0,609	-0,62	0,389	-0,257	0,381	0,645	0,874	0,856	0,821	0,935	0,889
Axis 2	-0,795	0,653	0,707	0,687	0,008	-0,916	-0,505	-0,069	-0,088	0,493	0,399	0,116	0,398
Variable	MPI14	PobY14	Analf14	Anose14	Nmati14	Sseg14	Elect14	Sanea14	Agua14	Cocina14	Piso14	Pared14	Techo14
Axis 1	0,393	-0,389	-0,776	-0,641	0,49	-0,118	-0,141	0,775	0,799	0,902	0,801	0,908	0,815
Axis 2	-0,769	0,81	0,566	0,706	0,316	-0,938	-0,064	-0,209	-0,068	0,254	0,422	0,023	0,377
Variable	MPI13	PobY13	Analf13	Anose13	Nmati13	Sseg13	Elect13	Sanea13	Agua13	Cocina13	Piso13	Pared13	Techo13
Axis 1	0,428	-0,47	-0,584	-0,471	0,293	-0,294	0,29	0,633	0,761	0,85	0,735	0,929	0,801
Axis 2	-0,654	0,59	0,666	0,808	0,086	-0,943	0,724	-0,184	0,217	0,34	0,459	0,275	0,344
Variable	MPI12	PobY12	Analf12	Anose12	Nmati12	Sseg12	Elect12	Sanea12	Agua12	Cocina12	Piso12	Pared12	Techo12
Axis 1	0,276	-0,686	-0,819	-0,521	0,144	-0,483	-0,045	0,386	0,902	0,977	0,852	0,959	0,801
Axis 2	-0,592	0,663	0,249	0,821	0,098	-0,834	-0,375	-0,283	0,009	-0,009	0,283	0,078	0,341
Variable	MPI11	PobY11	Analf11	Anose11	Nmati11	Sseg11	Elect11	Sanea11	Agua11	Cocina11	Piso11	Pared11	Techo11
Axis 1	0,558	-0,699	-0,83	-0,644	0,242	-0,538	0,073	0,886	0,763	0,931	0,853	0,937	0,597
Axis 2	-0,469	0,62	0,26	0,576	0,062	-0,822	0,531	0,167	-0,004	0,351	0,424	0,169	0,38

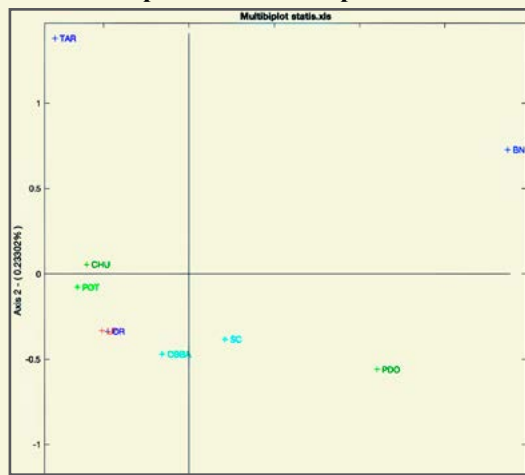
Fuente: Elaboración propia.



Para la infraestructura, las correlaciones de las modalidades con los ejes del compromiso en el primer plano factorial que recoge el 69.636% de la inercia total ( $\Lambda_1=0.818$  y  $\Lambda_2=0.415$ ). El análisis de los departamentos, la etapa de compromiso del STATIS construida a partir de la imagen euclidiana de los años de estudio permite observar el desempeño de los 9 departamentos y el índice de pobreza multidimensional y sus dimensiones. La ubicación espacial en el

plano (figura 4) muestra que, según el eje x, los departamentos se dividen en dos grupos; uno en el que los departamentos en general cuentan con bajos valores de las dimensiones en términos de porcentaje de cumplimiento, y otro donde los departamentos tienen mayor porcentaje de incumplimiento a las dimensiones que conforman el índice de pobreza multidimensional para los nueve años estudiados.

**Figure 4**  
**Plano compromiso de los departamentos**

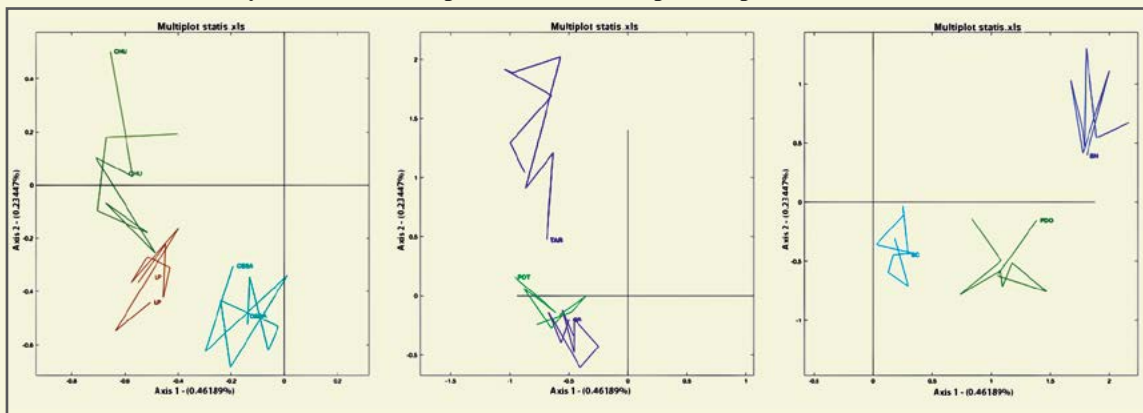


Fuente: Elaboración propia

En el análisis de las trayectorias de los departamentos, se relacionan las evoluciones horizontales de acuerdo a las modalidades que

caracterizan el primer eje y las evoluciones verticales, con aquellas que caracterizan el segundo eje como se puede ver en la figura 5.

**Figure 5**  
**Trayectoria de los departamentos en el primer plano factorial**



Fuente: Elaboración propia

#### **4. DISCUSIÓN**

En el presente artículo se mostró una aplicación del método STATIS en el estudio de la pobreza multidimensional, sus dimensiones y su evolución en el tiempo para los años antes mencionados, siendo esta forma de análisis una alternativa a emplear cuando se tienen períodos de tiempo.

De los resultados es posible proponer una clasificación de las observaciones basadas en las posiciones de los datos sobre los ejes factoriales. El análisis de las diferentes formas de las trayectorias nos refleja el comportamiento de cada uno de los departamentos, en cada una de las diferentes ocasiones en que se midieron. El análisis de los datos con el método STATIS produce resultados plausibles y lógicos que permite estimar la similitud en los nueve años de estudio, las trayectorias de las dimensiones que componen la pobreza multidimensional y su correlación con los ejes, así como las asociaciones de observaciones particulares que tengan variaciones similares, la posición y las trayectorias individuales de los departamentos, de una ocasión a otra.

El STATIS es un método de análisis de tablas de tres vías, de gran importancia en el análisis estadístico multivariante. Este tipo de tablas se caracterizan porque presentan una primera entrada, vía o índice, que identifica a los individuos que son objeto de estudio en este caso los departamentos, una segunda entrada para las variables que se han medido sobre esos individuos (las dimensiones), y una tercera entrada para las diversas situaciones (condiciones u ocasiones) en las que se realizaron mediciones (años). Se analizó las diferencias y semejanzas entre los diferentes escenarios a través de las configuraciones de los individuos y de las relaciones entre los diferentes grupos de dimensiones.

El STATIS se ha usado en diversos campos, como la evaluación sensorial, imágenes moleculares, imágenes cerebrales, economía, sostenibilidad y quimiometría, entre otras muchas áreas, pero su aparición inicial fue en estudios de tipo ecológico. En el caso de análisis de la pobreza y métodos estadísticos de análisis exploratorio multi vía, después de una revisión exhaustiva en diferentes bases bibliográficas como SCOPUS y EMBASE no se identificaron la implementación de estos métodos al campo de la pobreza económica social.

#### **5. CONCLUSIONES**

El presente artículo ha tomado como datos la medición de la pobreza multidimensional para los años 2011 al 2019 que fue calculado con la metodología Alkire-Foster a partir de la información en la Encuesta de Hogares, debido a la semejanza de las variables que se utilizaron en la encuesta para todos los años. Los datos consolidados fueron analizados a través de un STATIS.

El trabajo ha sido desarrollado bajo un caso de estudio particular, con el objetivo de contribuir con distintas herramientas estadísticas descriptivas que ayuden a desarrollar un análisis más objetivo de los principales fenómenos económicos sociales. Por otro lado, se confirma la premisa que estas metodologías de análisis de tablas de tres entradas pueden replicarse en otras aplicaciones de la misma encuesta y otras áreas.

Se realizó un estudio exploratorio de la información mencionada para comprender las estructuras subyacentes a los datos referente a las dimensiones de la pobreza multidimensional a nivel urbano, para cada departamento, así como su dinamismo a lo largo del período de tiempo de estudio. A través del STATIS se pudo comparar y analizar

las relaciones entre los diferentes años y las dimensiones de la pobreza multidimensional, se identificó la estructura intercorrelacional, a partir de la obtención de una tipología en los departamentos. Adicionalmente se observó la evolución de las dimensiones de la pobreza multidimensional en los departamentos a lo largo del periodo estudiado. Finalmente,

se describió la estructura interna de las dimensiones estudiadas para los departamentos y su evolución en el tiempo.

## CONFLICTO DE INTERESES

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

## REFERENCIA BIBLIOGRÁFICA

- Abdi, H. (n.d.). *RV Coefficient and Congruence Coefficient*. <http://www.utd.edu/>
- Abdi, H., & Valentin, D. (n.d.). *The STATIS Method*. <http://www.utd.edu/>
- Abdi, H., & Williams, L. J. (2010). Principal component analysis. In *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 2, Issue 4, pp. 433–459). <https://doi.org/10.1002/wics.101>
- Abdi, H., Williams, L. J., Valentin, D., & Bannani-Dosse, M. (2012). STATIS and DISTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. In *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 4, Issue 2, pp.124–167). <https://doi.org/10.1002/wics.198>
- Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7–8), 476–487. <https://doi.org/10.1016/j.jpubeco.2010.11.006>
- Alkire, S., & Santos, M. E. (2014). Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index. *World Development*, 59, 251–274. <https://doi.org/10.1016/j.worlddev.2014.01.026>
- Greenacre, M. (2010). *Biplots in Practice* (Fundación BBVA, Ed.).
- Lavit, C., Escoufier, Y., Sabatier, R., & Traissac, P. (1994). The ACT (STATIS method). In *Computational Statistics & Data Analysis* (Vol. 18).
- Muñeton Santa, G., Pineda Varela, L., & Keep Buitrago, J. P. (2019). *Medición de la pobreza multidimensional para la ciudad de Medellín, Colombia*. <https://orcid.org/0000-0002-5194-1914>
- Vicente Villardon, J. L. (2015). *MultBiplot: A package for Multivariate Analysis using Biplots*. Universidad de Salamanca. <http://biplot.usal.es/ClassicalBiplot/index.html>

# INSTRUCCIONES PARA AUTORES

## REVISTA VARIANZA

Revista Científica del Instituto de Estadística Teórica y Aplicada (IETA),  
Carrera de Estadística, Facultad de Ciencias Puras y Naturales  
Universidad Mayor de San Andrés  
La Paz, Bolivia

ISSN 2789-3510, versión impresa

ISSN 2789-3529, versión en línea

<https://ojs.umsa.bo/ojs/index.php/revistavarianza/>

<https://ieta.umsa.bo/>

## MISIÓN

Difundir principalmente artículos originales de investigación científica en diferentes ámbitos de la vida, basados en el uso de métodos y técnicas estadísticas. También difundir artículos de naturaleza teórica en el campo de la Estadística. Todo ello con el propósito de contribuir al desarrollo de nuestra sociedad.

## VISIÓN

Llegar a ser la revista científica nacional de mayor calidad e impacto en el campo de la estadística aplicada y teórica, así como ser el principal referente para el contexto internacional.

## TIPOS DE MANUSCRITOS

En la Revista Varianza se publican principalmente **artículos originales**, aquéllos que resultan de una investigación científica y que contribuyen, en alguna medida, al conocimiento científico y/o solución de alguna problemática. Los artículos originales pueden ser de naturaleza teórica o práctica. Los de naturaleza práctica se enfocan en dar respuesta, con base en el uso de métodos y/o técnicas estadísticas apropiadas, a problemas o preguntas de investigación en distintos campos de la vida; mientras los de naturaleza teórica presentan un nuevo método o técnica estadística, o pueden ofrecer una versión mejorada de uno existente.

Además, si bien no son artículos originales de investigación, en la Revista Varianza también se publican otro tipo de manuscritos, como ser:

**De revisión**, que constituyen básicamente informes sobre avances o estado del arte de un tema particular, con base en la recopilación y selección de artículos científicos originales,

**Comunicaciones breves**, manuscritos que comunican de manera breve algunos datos de una investigación original que el editor cree que serán interesantes para muchos investigadores y que probablemente estimularán más la investigación en esa área,

**Estudios de caso**, informan los resultados sobre casos específicos de fenómenos interesantes. Su propósito es hacer que otros investigadores conozcan la posibilidad de que un fenómeno específico pueda ocurrir,

**Reseñas**, consisten en resúmenes concisos generalmente sobre libros recientemente publicados en el campo de la Estadística,

**Notas científicas**, presentan observaciones y descripciones científicas breves de métodos o resultados, comunican resultados de estudios pequeños, avances de trabajos de investigación o noticias de interés científico,

**De enseñanza**, son manuscritos sobre temas relacionados a la enseñanza de la Estadística, por ejemplo la comprensión de un teorema o un método de estimación. Tiene el propósito de clarificar y complementar los conocimientos estadísticos de los estudiantes y los docentes.

## **PROCESO DE REVISIÓN DE MANUSCRITOS**

Luego de haber recibido el manuscrito, se inicia el proceso de su revisión, el cual tiene tres etapas:

### **Primera etapa: Revisión por el editor**

En esta etapa, el editor revisa si el tema del manuscrito es apropiado para la Revista y si cumple con las instrucciones para los autores. Se revisa la pertinencia del manuscrito para la Revista, los aspectos de forma del manuscrito y el cumplimiento de requisitos básicos exigidos en las normas de la Revista Varianza. El autor será contactado para informarle si su manuscrito es apto para pasar a la revisión por pares a doble ciego o si requiere mejorar algunos detalles del manuscrito sugeridos por el editor o si es rechazado (por no presentarse en el formato exigido en las normas, por tener errores metodológicos importantes, porque el manuscrito ha sido publicado previamente o porque el aporte no es nuevo, entre otros). En caso de no existir faltas o errores, el manuscrito pasa a la segunda etapa.

### **Segunda etapa: Revisión por evaluadores externos**

Cada manuscrito que llega a esta etapa es sometido al proceso de revisión por pares a “doble ciego”. Esta modalidad significa que cada manuscrito es revisado por dos evaluadores externos a nuestra institución, ambos miembros del comité científico, con la restricción de que ni el evaluador sabe el nombre del autor del manuscrito y ni el autor sabe quiénes son sus evaluadores. Para la asignación del manuscrito a los dos evaluadores se toma en cuenta el vínculo entre el tema del manuscrito y la especialidad o experiencia de los evaluadores.

En esta etapa se evalúa rigurosamente el contenido del manuscrito, poniendo énfasis en los aspectos metodológicos. A través de una ficha de evaluación, el dictamen de la revisión es una de las siguientes cuatro alternativas: (i) aceptado sin modificaciones, (ii) aceptado con observaciones leves, (iii) aceptado con observaciones profundas o (iv) rechazado. Se comunica al evaluador tanto el dictamen como las observaciones, en caso de existir. Una vez que el autor subsana las observaciones, nuevamente el manuscrito corregido es derivado

al revisor para su evaluación, y así sucesivamente hasta que el manuscrito es aceptado o rechazado. En caso de no corregirse las observaciones, el manuscrito es rechazado.

Algunos aspectos que se toman en cuenta en la evaluación, son:

- a) Claridad en el planteamiento del objetivo principal y/o la hipótesis a probar.
- b) Pertinencia del método estadístico para el cumplimiento del objetivo y/o prueba de hipótesis.
- c) Datos apropiados para el cumplimiento del objetivo y/o prueba de hipótesis.
- d) Grado de profundidad de la investigación.
- e) Coherencia en el análisis cuantitativo, incluido la correcta interpretación de los resultados.
- f) Generación de nuevo conocimiento científico y aporte para la formulación de políticas, programas y proyectos.
- g) Pertinencia de la referencia bibliográfica usada.
- h) Calidad de la redacción, incluido el resumen (síntesis del manuscrito en forma clara y adecuada).
- i) Ajuste del formato a las normas APA.

### **Tercera etapa: Diagramación**

Los manuscritos aprobados por el comité científico pasan a diagramación, a una última revisión de redacción y a maquetación. Esta tarea es realizada por un profesional en diagramación y el editor, en coordinación con el autor. Se trata de una última revisión del manuscrito, sólo de forma. No se acepta ninguna modificación con relación al contenido del manuscrito ya aceptado, sea de texto, tabla o gráfico, como tampoco se acepta la inclusión de un texto adicional, una tabla o un gráfico.

## **FORMATO PARA ARTÍCULOS ORIGINALES DE INVESTIGACIÓN**

### **Título**

El título debe estar en español y en inglés, ambos en mayúscula, en no más de 15 palabras cada uno, por lo que debe ser lo más corto posible y debe reflejar plenamente el contenido del manuscrito.

### **Autor(es)**

Debajo del título deben colocarse el nombre completo del autor, su afiliación institucional durante la realización del manuscrito y su correo electrónico. Si son dos o más autores, colocar el nombre completo, la afiliación institucional y el correo electrónico de cada uno. Para el caso de dos o más autores, el principal autor es el que más ha contribuido a la investigación, y deberá encabezar la lista. En pie de página, y con la numeración correspondiente para cada autor, se debe informar la profesión, un breve curriculum vitae si se desea, y el número ORCID.

## **Resumen**

El resumen debe estar en español y en inglés, con títulos en mayúscula (RESUMEN y ABSTRACT, respectivamente), y muy bien redactado en no más de 250 palabras. Debe incluir con absoluta claridad y precisión el objetivo, la metodología, los datos usados, el principal resultado, alguna idea de discusión y la principal conclusión.

## **Palabras claves**

Debajo del resumen en español deben incluirse las palabras clave (Palabras clave), las cuales sirven para identificar el artículo por parte de usuarios nacionales e internacionales. Incluir de cuatro a siete palabras clave que no formen parte del título del artículo, ordenadas alfabéticamente y separadas por comas. Igualmente, debajo del resumen en inglés incluir las mismas palabras clave, pero en inglés (Key words).

## **Introducción**

Debe presentar el problema dentro de un marco teórico y/o revisión bibliográfica que acompañe a la hipótesis y/o objetivo principal del trabajo. El problema debe describir claramente lo que se resolverá con la investigación; la revisión bibliográfica expone el marco de referencia que da sustento al trabajo de investigación, resalta, a través de citas, estudios previos que se relacionan específicamente con la investigación; el objetivo principal debe indicar claramente, en forma precisa y sin ambigüedad, la finalidad de la investigación; mientras la hipótesis debe plantear lo que trata de probar la investigación. El objetivo y la hipótesis deben estar al final de esta sección.

## **Materiales y métodos**

En esta sección se describe minuciosamente la forma en que se realizó la investigación, de modo que un lector que acceda al artículo pueda comprenderlo plenamente y reproducirlo a fin de determinar la confiabilidad y la validez de los resultados. Esta sección debe describir claramente los aspectos principales respecto de los datos, los instrumentos, y los métodos y técnicas estadísticas usados en la investigación.

## **Resultados**

En esta sección se expone el principal o los principales hallazgos obtenidos con la investigación, todos ellos en estricta consonancia con el objetivo principal y/o con la hipótesis de la investigación. La correcta interpretación de los resultados es de suma importancia en esta sección. Para transmitir los principales hallazgos pueden incluirse, si es necesario, cuadros, gráficos y diagramas, evitando la redundancia, evitando un número excesivo de datos y manteniendo la objetividad (imparcialidad y honestidad). Todos los cuadros, gráficos y diagramas deben enumerarse en el orden que aparecen en el texto.

## **Discusión**

Esta sección está orientada a interpretar los resultados de la investigación en relación con el

objetivo principal, la hipótesis y el estado de conocimiento actual del tema de la investigación, esto es, se debe indicar qué significan los hallazgos encontrados y cómo estos se relacionan con el conocimiento actual sobre el tema.

Además de compararlos y contrastarlos con los resultados de otros estudios relevantes, resaltando sus limitaciones y ventajas tanto conceptuales como metodológicas, argumentar las implicaciones de los resultados para la formulación de políticas, programas y/o proyectos, y argumentar las implicaciones para futuras investigaciones.

### **Conclusiones**

Las conclusiones son derivadas de los resultados y de la discusión, y responden al objetivo y/o la hipótesis de la investigación. Constituyen los aportes y las innovaciones de la investigación

### **Agradecimiento**

En esta sección no se incluye ningún elemento científico, sólo se trata de ser cortés con quienes colaboraron en la investigación. Se puede reconocer la contribución de personas o instituciones que ayudaron realmente en la investigación, pero no se las puede considerar como coautores.

### **Conflicto de intereses**

Acá, el autor o autores deben declarar no tener ningún conflicto de intereses con su artículo científico.

### **Referencias bibliográficas**

Esta sección contiene la referencia de libros y artículos citados en las diferentes secciones del manuscrito, en formato APA. Debe existir siempre una correspondencia entre las citas que se hace en el trabajo y las que se lista en las referencias bibliográficas, ya que normalmente los lectores estarán interesados en verificar los datos que efectivamente se utilizaron para la investigación.

Ejemplos de cómo enunciar las referencias bibliográficas en formato APA se pueden encontrar en la siguiente dirección (URL): <https://normas-apa.org/referencias/>

## **ESPECIFICACIONES PARA LA PRESENTACIÓN DE MANUSCRITOS**

Los manuscritos que se presentan deben estar escritos en Word, hoja tamaño carta, doble columna, letra Times New Román de tamaño 11, espacio simple, margen izquierdo de 2.5 cm., y los demás márgenes de 2.0 cm. Todo el manuscrito, incluido texto, gráficos, cuadros, diagramas y otros, debe contener entre 10 y 20 páginas, con títulos y subtítulos enumerados. Los gráficos, cuadros y diagramas no deben exceder el 30 por ciento del manuscrito. Adicionalmente, todas las notas y referencias deben ir acorde al formato APA.

Para la presentación del manuscrito debe acompañarse una carta en la que se indique el tipo



de manuscrito que se está enviando (artículo original, revisión, estudio de caso, reseña, nota científica o manuscrito de enseñanza) a la siguiente dirección: [ieta@umsa.bo](mailto:ieta@umsa.bo). Los autores pueden enviar sus manuscritos en cualquier momento del año.

## **PERIODICIDAD DE LA PUBLICACIÓN**

La versión impresa de la Revista Varianza se publicó por primera vez el año 2001, desde ese año hasta el 2020 se publicó anualmente, si bien no se pudo editar en algunos años. Sin embargo, a partir del segundo semestre de 2021 la publicación es semestral, en los meses de abril y octubre.

En cambio, la versión digital (on line) de la Revista se publica desde el segundo semestre del año 2021, también en los meses de abril y octubre.

Con el propósito de incrementar la visibilidad de la Revista Varianza y facilitar la búsqueda de artículos por parte de los lectores, desde el año 2023 la Revista Varianza también se publica junto a las revistas científicas de otras unidades de la Universidad Mayor de San Andrés. Se puede acceder a la página a través de la dirección <https://ojs.umsa.bo/ojs/>.

## **CONFLICTO DE INTERESES**

La Revista Varianza tiene la política de evitar cualquier conflicto de interés de los autores, del comité editorial y del propio editor responsable. Se recomienda a cada autor evitar cualquier conflicto de interés relacionado con su artículo, debiendo comunicar oportunamente al editor responsable, como también se pide al comité editorial impedir cualquier conflicto de interés en el proceso editorial.

## **ÉTICA DE PUBLICACIÓN**

La revista Varianza tiene compromiso con la ética de la investigación, por ello, promueve los siguientes aspectos:

- a) Evitar conflictos de intereses
- b) Evaluar objetivamente los manuscritos
- c) Respetar los criterios de evaluación de los evaluadores
- d) Conservar la confidencialidad de los autores y evaluadores, durante todo el proceso de revisión.

## **FINANCIAMIENTO DE LA REVISTA**

La Universidad Mayor de San Andrés (UMSA) asigna anualmente recursos financieros al Instituto de Estadística Teórica y Aplicada (IETA) para impresión y difusión de la Revista Varianza. La publicación en la revista es gratuita bajo la modalidad Open Access.

## **PROPIEDAD INTELECTUAL**

Para los manuscritos aceptados para su publicación, el o los autores deben autorizar formalmente al editor, a través de un documento firmado, su publicación en la Revista Varianza. En el documento firmado, el lector también afirma ser legítimo propietario del manuscrito a publicar y que no existe problemas de derechos de autor con terceros y/u otros conflictos de naturaleza ética. Todo el contenido de la Revista, excepto aquéllo que expresamente sea identificado, está bajo la licencia Creative Commons.

## **LICENCIAMIENTO**

La Revista Varianza se encuentra bajo licenciamiento Creative Commons atribución CC BY <https://creativecommons.org/licenses/by/4.0/>. La licencia permite que otros distribuyan, mezclen, adapten y construyan sobre su trabajo, incluso comercialmente, siempre que reconozcan la creación original. Esta es la licencia más complaciente que se ofrece. Recomendado para la máxima difusión y uso de materiales con licencia.

**Dirección: Calle 27 de Cota Cota  
Bloque F.C.P.N. - Primer Piso  
Email: [ieta@umsa.bo](mailto:ieta@umsa.bo)  
Página web: <https://ojs.umsa.bo/ojs/index.php/revistavarianza>**

**La Paz - Bolivia**