

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 Viamonte¹
Universidad de Salamanca, Salamanca, España,
✉ 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

¹ 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
1. Pobreza Monetaria. (1/4)	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
2. Educación. (1/4)	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
3. Salud. (1/4)	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
4. Entorno y calidad de vida. (1/4)	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 α , 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 α (Muñeton Santa et al., 2019, p. 121). La elección de α debería ser un política, es decir, debería definirse de manera normativa (Alkire & Foster, 2011). Siguiendo a (Alkire & Santos, 2014) un α 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).

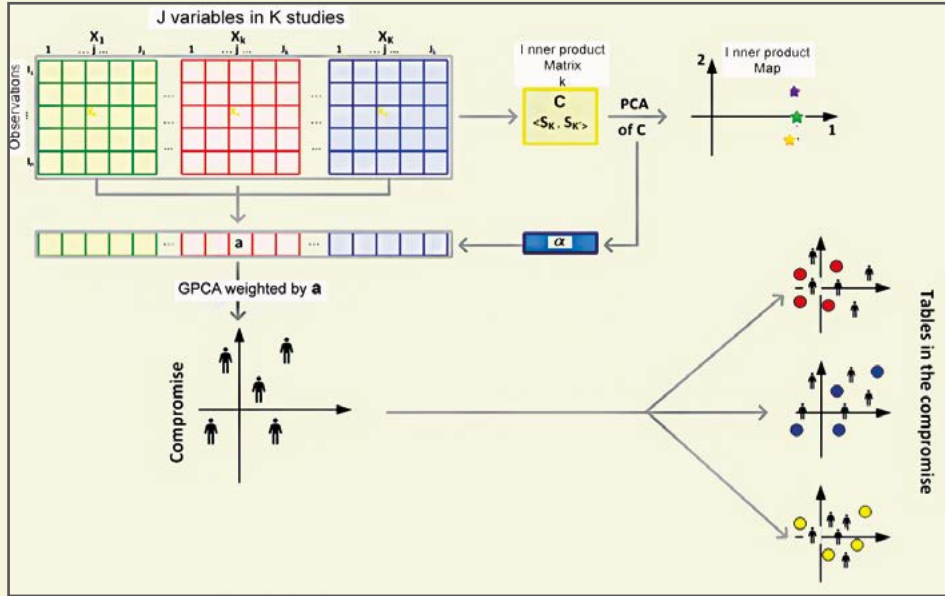
2.4.1 Cuando se puede usar el método.

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 α 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 α_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 λ_ℓ , 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 ℓ -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 ℓ , 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- ℓ). 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 ℓ , 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:

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

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 descrita por el compromiso.

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

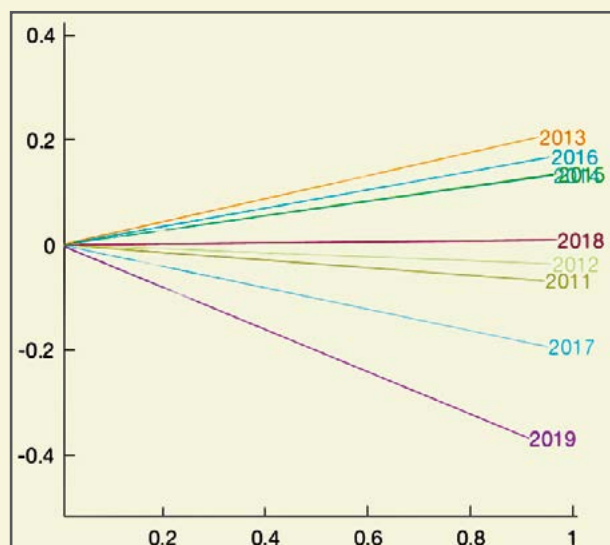
Años	2019	2018	2017	2016	2015	2014	2013	2012	2011
2019	1								
2018	0.876	1							
2017	0.929	0.913	1						
2016	0.830	0.926	0.886	1					
2015	0.840	0.932	0.899	0.952	1				
2014	0.829	0.903	0.900	0.943	0.959	1			
2013	0.787	0.915	0.854	0.916	0.921	0.919	1		
2012	0.881	0.924	0.900	0.880	0.944	0.923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

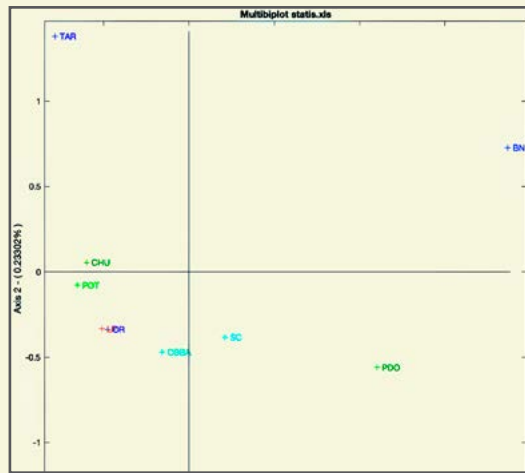
En el caso de la intraestructura, la figura 3 muestra las proyecciones de las variables en cada uno de los años sobre el plano formado por las primeras componentes del compromiso,

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

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

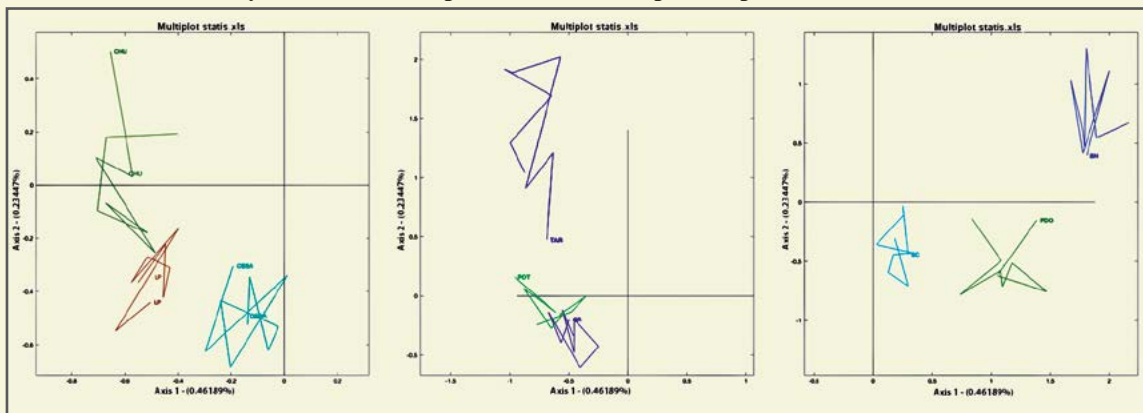


Fuente: Elaboración propia

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>