

Artículo Original / Original Article

Modelo LASSO para comparar indicadores de desarrollo social y bienestar en Perú y la región suramericana

LASSO model to compare indicators of social development and well-being in Peru and the South American region

Roberto Alejandro Pacheco-Robles^{1*} ; Santiago Vela-Del-Águila¹ ; Oscar Tuesta-Hidalgo¹ ; Juan Carlos Tuesta-Hidalgo¹ ; Marco Antonio Nureña-Hidalgo¹ ; Juan Manuel Vela-Lozano¹ 

¹Universidad Nacional Autónoma de Alto Amazonas, Yurimaguas, Perú

RESUMEN

Medir la pobreza es un tema pendiente de atender en Sudamérica. En el presente estudio se tuvo por objetivo describir, explicar y comparar los niveles de desarrollo social y bienestar de los ciudadanos de Perú frente a Sudamérica. La investigación fue no experimental, método cuantitativo, explicativa, con diseño longitudinal; empleamos el modelo de regresión LASSO y PLS con datos relacionados a indicadores de desarrollo y bienestar. Identificamos tres indicadores que diferencian a los países considerados: Argentina, Brasil, Chile, Colombia, Perú y Uruguay; estos reflejan características asociadas al género, grupos etarios, acceso a la seguridad social y condición laboral; observamos la presencia de dos ciclos estacionales a lo largo de toda la serie: el primero asociado a las dos primeras décadas de estudio (1986-2000) y el segundo que inicia el año 2000 y muestra tendencia a la estabilidad. Evidenciamos la presencia de grupos de países en Sudamérica con características similares relacionadas con el género, el acceso a la seguridad social y la condición laboral.

Palabras clave: comparación; grupo etario; indicadores de género; seguridad social

ABSTRACT

Measuring poverty is a pending issue in South America. The objective was to describe, explain and compare the levels of social development and well-being of the citizens of Peru compared to South America. The research was non-experimental, quantitative method, explanatory, with longitudinal design; We use the LASSO and PLS regression model with data related to development and well-being indicators. We identify three indicators that differentiate the countries considered: Argentina, Brazil, Chile, Colombia, Peru and Uruguay; These reflect characteristics associated with gender, age groups, access to social security and employment status; we observe the presence of two seasonal cycles throughout the entire series, the first associated with the first two decades of study (1986-2000) and the second that begins in the year 2000 and shows a tendency towards stability. We show the presence of groups of countries in South America with similar characteristics related to gender, access to social security and employment status.

Keywords: comparison; age group; gender indicators; social security

Cómo citar / Citation: Pacheco-Robles, R. A., Vela-Del-Águila, S., Tuesta-Hidalgo, O., Tuesta-Hidalgo, J. C., Nureña-Hidalgo, M. A., & Vela-Lozano, J. M. (2023). Modelo LASSO para comparar indicadores de desarrollo social y bienestar en Perú y la región suramericana. *UNAAACIENCIA-PERÚ*, 2 (2), e29. <https://doi.org/10.56926/unaaaciencia.v2i2.29>

Recibido: 05/06/2023

Aceptado: 15/09/2023

Publicado: 10/10/2023

*Roberto Alejandro Pacheco-Robles - apacheco@unaaa.edu.pe (autor de correspondencia)



©Los autores. Este es un artículo de acceso abierto, distribuido bajo los términos de la Licencia Creative Commons Atribución 4.0 Internacional

1. INTRODUCCIÓN

La población mundial se estima en más de 7 500 millones de habitantes según el informe Perspectivas de la Población Mundial (ONU DAES, 2022). En ese reporte, la mayoría en países en vías de desarrollo tiene graves problemas como falta de agua potable y sistemas de saneamiento con grandes asentamientos informales, situación que provoca hambre y necesidad, hecho que se expone más a consecuencia de la pandemia de COVID-19 (UNESCO, 2019).

La mayoría de las personas que viven por debajo del umbral de pobreza se encuentran en Asia meridional y África subsahariana, con un aproximado de 800 millones de pobres en el mundo, representado casi el 13% de la población (ONU, 2022). En un nuevo informe anual, la CEPAL estima que el total de personas pobres ascendió a 209 millones a finales del 2020, 22 millones de personas más que el año anterior; la pobreza y la pobreza extrema alcanzaron en el 2020 en América Latina niveles que no se han observado en los últimos 12 y 20 años (CEPAL, 2021).

En ese sentido, instituciones como la ONU (2022) propusieron diecisiete Objetivos de Desarrollo Sostenible (ODS), en ellos se analiza las condiciones locales en torno a estos temas y aporta nuevos conocimientos para entender, comprender y cuantificar el problema y ayudar a desarrollar políticas acordes con la realidad. De hecho, tiene como primer objetivo la eliminación de la pobreza, seguido de mejorar la salud y de incrementar el bienestar.

Ahora bien, para instituciones como el Banco Interamericano de Desarrollo (BID), la salud se ve amenazada por ciertas condiciones, entre ellas las que viven las familias, como las deficiencias en viviendas o con varios ocupantes, con poco o ningún acceso a agua potable, deficiente servicio de saneamiento y dificultades para la adquisición de alimentos nutritivos; otra limitante se denota en el deficiente acceso al bono otorgado por el gobierno peruano, y en su mayoría se trata de familias pobres con nulo acceso a tecnologías de información y comunicaciones.

Diversos estudios muestran que la raíz de este gran problema es la corrupción, y en la mayoría de los casos es provocada por el propio gobierno que tiene como gestores a operadores políticos sin perfil técnico, además del crecimiento demográfico de las ciudades sin planificación y la llegada de inmigrantes generada por migraciones de carácter político social, afectando el empleo y la convivencia social, ocasionando desigualdad social y restricciones a la libertad a nivel de la región suramericana.

Se infiere además que, un factor importante que mantiene a la región en esta situación es la educación, pues se basa únicamente en resultados cuantitativos, dejando de lado el desarrollo actitudinal basado en valores. Otros factores que se consideran importantes son el cambio climático, los conflictos armados, los modelos de negocio, entre otros. De manera similar, se dice que ser pobre es una condición humana natural, que el hombre nace pobre; por tanto, la riqueza tiene causas y es importante estudiar las razones por las cuales se mantiene el estado natural de pobreza en la población.

Mencionamos algunas de las consecuencias de este problema: afecta los derechos fundamentales de las personas, la salud, educación, vivienda digna, agua potable, empleo, ingresos, justicia, etc. Además, el medir apropiadamente la pobreza es un tema pendiente de atender en los países de

Sudamérica, los parámetros más usados son: la pobreza monetaria, en algunos casos se toma otros parámetros como carencias ecológicas, reducción de la esperanza de vida, incremento del analfabetismo, incremento de la anemia, reducción del ingreso mínimo vital, deficiente atención sanitaria, incremento de la mortalidad infantil, deficiente acceso a los servicios de electricidad, servicios de saneamiento, agua potable; así como artefactos eléctricos, internet, etc.

En cuanto al tema de medir la pobreza, el filósofo y economista Sen (1985), es quien aporta el Índice de Desarrollo Humano y 3 métodos para medirla: el método de línea de pobreza, la metodología de las necesidades básicas insatisfechas y el método integrado de medición de la pobreza.

Acercas de la pobreza en Perú, datos del INEI (2021) muestran que, en el año 2020, la pobreza monetaria afectó al 30,1% de la población del país, incrementándose en 9,9 puntos porcentuales en comparación con el año 2019 y esto probablemente a las consecuencias del aislamiento social obligatorio impuesto por la pandemia de Covid-19. En los últimos cinco años la población en situación de pobreza aumentó en 8,3 puntos porcentuales y en la última década disminuyó en 0,7 puntos porcentuales.

El índice de avance contra la desigualdad en el Perú 2017-2018 maneja datos oficiales que se utilizan como base para el análisis y responde a la idea de que la desigualdad es un grave problema que afecta el bienestar y la estabilidad de la sociedad y que amenaza los derechos esenciales de las personas; por ejemplo, justicia tributaria, inversión a las personas, gasto en programas sociales, empleo digno, derecho económico de la mujer, gobernanza socioambiental y de los recursos naturales, entre otros (Armando Mendoza, 2019).

En Perú se vive un estado de democracia institucional y se cuentan con programas sociales que permiten trabajar temas encaminados a erradicar la pobreza extrema; en el presente estudio se busca analizar la pobreza con los instrumentos ya existentes, y aportar con información a tomarse en cuenta para las causas teóricas, sociales, culturales, que inciden en los niveles de desarrollo social y bienestar. Así, nos proponemos abordar el siguiente problema, ¿Cuáles son los significados, las privaciones y los logros asociados al bienestar y la calidad de vida de los ciudadanos en Perú en comparación con los de la región suramericana en el período 1985-2020?

En el Perú, se crea la Mesa de Concertación para la Lucha contra la Pobreza como parte del acuerdo político entre el Estado y la Sociedad Civil de trabajar en conjunto. Se formaliza mediante el Decreto Supremo N° 001-2001-PROMUDEH, modificado por el Decreto Supremo N° 014-2001-PROMUDEH. En la Ley 27867 Ley de Gobiernos Regionales.

En cuanto a sus objetivos están el generar procesos de articulación y concertación al interior de la sociedad civil en la que se establece condiciones y metas para lograr el desarrollo de capacidades, además se da énfasis al control, promueve los partidos políticos; establece el aporte de los movimientos sociales y de la sociedad civil. En cuanto a su definición, la ONU puntualiza en su propuesta que Desarrollo Social son las etapas en las que una sociedad evoluciona, adquiriendo mejores condiciones de vida, que se vinculan con los progresos económicos de la nación. Si mejora la economía impulsa a la nación y mejora considerablemente el bienestar con respecto a la calidad de vida, ya que tenemos mejor acceso a los servicios, mejorando así la competitividad de las empresas, y se crean nuevos proyectos de estado.

Lopez Flores (2013) en su investigación determina que, las limitaciones económicas de las familias y del estado afectan las condiciones terribles de hambre, insalubridad, falta de educación, vestuario, trabajo digno, migración permanente. Mattei (2017) describe en su investigación que un grupo de investigadores determina que la política de la reforma agraria en Brasil no combate la pobreza y las desigualdades sociales y regionales, sin embargo, otros investigadores evalúan como un instrumento necesario la reforma en la agricultura.

A continuación, citamos algunas teorías que explican nuestras variables, desarrollo social y bienestar; Teoría de la Modernización propuesta por Alvin So, Teoría de la Dependencia representada por Raúl Prebisch, Teoría de los Sistemas Mundiales liderada por Immanuel Wallerstein, Teoría de la Globalización propuesta por Kaplan, B., Teoría del Desarrollo Humano liderado por Amartya Sen.

Ahora bien, Londoño Vásquez & Castañeda Naranjo (2010) manifiestan que un tema importante de atender dentro de las ciencias sociales es la comprensión, expresada también por Max Weber como pionero de la llamada Sociología Comprensiva. Respetando los hechos sociales, culturales, históricos que los investigadores buscan sentido y su consecuente vinculación con las implicancias metodológicas; estas implicancias buscan reconocer el esfuerzo para lograr que la explicación de la fundamentación de una explicación causal correcta desde el punto de vista científico.

Este estudio se justifica ya que los resultados serán aprovechados por investigadores e instituciones públicas para tomar los resultados en su nueva propuesta de políticas públicas. En ese sentido, el objetivo de la investigación fue describir, explicar y comparar los niveles de desarrollo social y bienestar de los ciudadanos de Perú frente a la región suramericana. Dentro de los objetivos específicos se propone: describir mediante una serie de tiempo el comportamiento de los indicadores de desarrollo social y bienestar en Perú y la región suramericana, explicar y comparar mediante regresión LASSO (Least Absolute Shrinkage and Selection Operator) los indicadores de desarrollo social y bienestar en Perú y la región suramericana.

2. MATERIALES Y MÉTODOS

La investigación fue no experimental, con el método cuantitativo, explicativo, con diseño longitudinal (Hernández Sampieri et al., 2014). Empleamos el modelo de regresión LASSO y PLS (Partial Least Squares) con el fin de disminuir la dimensionalidad de la matriz de datos X, además de identificar aquellos indicadores que determinan diferencias o similitudes estadísticas entre países. Consideramos esas dos metodologías (PLS y LASSO) para el análisis de los datos ya que estamos en presencia de un caso que involucra muchas más variables que observaciones ($p \geq n$), y, en consecuencia, el estimador de los parámetros del modelo no está bien definido (el sistema de ecuaciones normales es indeterminado). Esta situación es cada vez más frecuente en diversos ámbitos de la ciencia (Genética, bioinformática, procesamiento de señales, econometría, ciencias sociales, etc.), con lo cual se ha convertido en un área de investigación muy dinámica en los últimos años (Fan & Li, 2006).

Las variables de estudio fueron:

Desarrollo social: Son las condiciones en las que deben vivir las familias y no solo se refiere a las condiciones monetarias, también a las carencias de tipo emocionales y ambientales (Fernández & Gutiérrez, 2013).

Bienestar: La noción de bienestar, es consecuencia de los efectos del desarrollo sobre el individuo, las familias y la sociedad (Uribe Mallarino, 2004).

La población estuvo conformada por los indicadores de desarrollo y bienestar de los ciudadanos residentes en Perú y la región suramericana; la muestra fue censal, utilizando la base de datos relacionada con indicadores de desarrollo, denominada Números para el Desarrollo (BID, 2020).

La técnica utilizada fue documental. Consideramos la base de datos relacionada con indicadores de desarrollo, denominada BID. A partir de esta base de datos generamos una matriz de datos X que abarca un periodo comprendido entre 1986-2017, la cual agrupa ocho países de Sudamérica (Argentina, Brasil, Chile, Colombia, Perú y Uruguay), y además incluye 253 indicadores seleccionados con base en un criterio de consistencia de la información a lo largo de la serie temporal considerada en el estudio.

Realizamos análisis de regresión LASSO y PLS con el fin de disminuir la dimensionalidad de la matriz de datos X , además de identificar aquellos indicadores que determinan diferencias o similitudes estadísticas entre países. Consideramos esas dos metodologías (PLS y LASSO) para el análisis de los datos ya que estamos en presencia de un caso que involucra muchas más variables que observaciones ($p \geq n$), y en consecuencia, el estimador de los parámetros del modelo no está bien definido (el sistema de ecuaciones normales es indeterminado).

Para el análisis de datos cuantitativos se utilizó la estadística descriptiva y el análisis multivariado (Regresión PLS y LASSO). A continuación, se describen brevemente estas dos metodologías según lo señalado por Bjorn-Helge & Wehrens (2016).

2.1. Regresión PLS

Los métodos de regresión multivariante como la regresión de componentes principales (PCR) y la regresión de mínimos cuadrados parciales (PLSR) han adquirido gran popularidad en una amplia gama de campos, incluidas las ciencias sociales. La razón principal es que han sido diseñados para abordar el problema de que hay muchas variables predictoras, posiblemente correlacionadas y relativamente pocas muestras, una situación que es común, especialmente en ciencias sociales, donde se dispone de un gran número de indicadores, pero pocos registros de los mismos.

En el contexto habitual de la regresión lineal múltiple (MLR), la solución de mínimos cuadrados para

$$Y = XB + \varepsilon$$

es dada por

$$B = (X^T X)^{-1} X^T Y$$

El problema a menudo es que $X^T X$ es singular, ya sea porque el número de variables (columnas) en X excede el número de objetos (filas), o debido a colinealidades. PCR y PLSR evitan esto, descomponiendo X en puntuaciones ortogonales T y cargas P .

$$X = TP$$

y la regresión de Y no sobre X en sí, sino en las primeras columnas de las puntuaciones T . En la PCR, las puntuaciones están dadas por los vectores singulares izquierdos de X , multiplicados por los valores singulares correspondientes, y las cargas son los vectores singulares derechos de X .

Sin embargo, esto solo toma en cuenta la información sobre X y, por lo tanto, puede ser desestimada para fines de predicción. PLSR tiene como objetivo incorporar información sobre X e Y en la definición de las puntuaciones y las cargas. De hecho, para una versión específica de PLSR, llamada SIMPLS (De Jong, 1993), se puede demostrar que las puntuaciones y las cargas se eligen de tal manera que describan lo más posible la covarianza entre X e Y , donde la PCR se concentra en la varianza de X . Otros algoritmos PLSR dan resultados idénticos a SIMPLS en el caso de una variable Y , pero se desvían ligeramente para el caso Y multivariado.

2.2. Algoritmos

En PCR, aproximamos la matriz X por los primeros Componentes principales (PC), generalmente obtenidos de la descomposición de valores singulares (SVD):

$$X = \tilde{X}_{(a)} + \varepsilon_X = (U_{(a)} D_{(a)} V_{(a)}^T) + \varepsilon_X = T_{(a)} P_{(a)}^T + \varepsilon_X$$

A continuación, haciendo la regresión de Y sobre las puntuaciones, lo que conduce a coeficientes de regresión:

$$B = P(T^T T)^{-1} T^T Y = V D^{-1} U^T Y$$

donde se han eliminado los subíndices a para mayor claridad.

Para PLSR, los componentes, llamados variables latentes (LVs) en este contexto, se obtienen de forma iterativa. Se comienza con la SVD de la matriz de productos cruzados $S = X^T Y$, que incluye información sobre la variación en X e Y , y sobre la correlación entre ellas. Los primeros vectores singulares izquierdo y derecho, w y q , se utilizan como vectores de peso para X e Y , respectivamente, para obtener puntuaciones t y u :

$$t = Xw = Ew$$

$$u = Yq = Fq$$

donde E y F se inicializan como X e Y , respectivamente. Las puntuaciones t de X a menudo se normalizan:

$$t = t / \sqrt{t^T t}$$

Las puntuaciones Y no son realmente necesarias en la regresión, pero a menudo se guardan para fines de interpretación. A continuación, las cargas X e Y se obtienen retrocediendo contra el mismo vector t :

$$p = E^T t$$

$$q = F^T t$$

Finalmente, las matrices de datos se "desinflan": la información relacionada con esta variable latente, en forma de los productos externos tp^T y tq^T , se resta de las matrices de datos (actuales) E y F .

$$E_{n+1} = E_n - tp^T$$

$$F_{n+1} = F_n - tq^T$$

La estimación del siguiente componente puede comenzar desde la SVD de la matriz de productos cruzados $E_{n+1}^T F_{n+1}$. Después de cada iteración, los vectores w , t , p y q se guardan como columnas en las matrices W , T , P y Q , respectivamente. Una complicación es que las columnas de la matriz W no se pueden comparar directamente: se derivan de matrices sucesivamente eliminadas E y F .

Se ha demostrado que una forma alternativa de representar los pesos, de tal manera que todas las luces se relacionen con la matriz X original, está dada por

$$R = W(P^T W)^{-1}$$

Ahora, estamos en la misma posición que en el caso de la PCR: en lugar de hacer la regresión de Y sobre X , usamos las puntuaciones T para calcular los coeficientes de regresión, y luego los convertimos nuevamente a las variables originales multiplicando previamente con la matriz R (ya que $T = XR$).

De nuevo, aquí, solo se utilizan los primeros componentes. Se debe determinar cuántos componentes son óptimos, generalmente mediante validación cruzada.

Se pueden encontrar muchas formulaciones alternativas en la literatura. Se ha demostrado, por ejemplo, que solo uno de X e Y necesita ser desinflado; alternativamente, se puede desinflar directamente la matriz de productos cruzados S (como se hace en SIMPLS, por ejemplo). Además, hay muchas formas equivalentes de escalado.

En teoría, PLSR debería tener una ventaja sobre la PCR. Uno podría imaginar una situación en la que un componente menor en X está altamente correlacionado con Y ; no seleccionar suficientes componentes conduciría a predicciones muy malas. En PLSR, dicho componente estaría automáticamente presente en el primer LV. En la práctica, sin embargo, casi no hay diferencia entre el uso de PLSR y PCR; En la mayoría de las situaciones, los métodos alcanzan precisiones de

predicción similares, aunque PLSR generalmente necesita menos variables latentes que la PCR. Dicho de otra manera, con el mismo número de variables latentes, PLSR cubrirá más de la variación en Y y PCR cubrirá más de X. A su vez, ambos se comportan de manera muy similar a la regresión ridge (Frank & Friedman, 1993).

También se puede demostrar que tanto la PCR como la PLSR se comportan como métodos de contracción (Mevik & Wehrens, 2007), aunque en algunos casos la PLSR parece aumentar la varianza de los coeficientes de regresión individuales, una posible explicación de por qué la PLSR no siempre es mejor que la PCR.

2.3. Regresión LASSO

Motivado por el objetivo de encontrar una técnica de regresión lineal que fuera estable pero que realizara selección de variables, Tibshirani (2011) propuso LASSO (Least Absolute Shrinkage and Selection Operator). En ese sentido, LASSO es una técnica de regresión lineal regularizada, como Ridge, con una leve diferencia en la penalización (norma L1 en lugar de L2) que trae consecuencias importantes. El auge en los últimos años en la investigación y aplicación de técnicas tipo LASSO, se debe principalmente a la existencia de problemas donde $p \geq n$ y al desarrollo paralelo de algoritmos eficientes (Tibshirani, 2011).

LASSO resuelve el problema de mínimos cuadrados con restricción sobre la Norma L1 del vector de coeficientes:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 \right\}, \text{ sujeto a } \sum_{j=1}^p |\beta_j| \leq s$$

O en forma equivalente, minimizando:

$$\sum_{i=1}^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Siendo $\lambda \geq 0$ los respectivos parámetros de penalización por complejidad.

A diferencia de la regresión Ridge y mínimos cuadrados ordinarios, el estimador $\hat{\beta}^{lasso}$ es no lineal en el vector de respuesta Y, y no existe una expresión en forma "cerrada" del mismo (salvo en el caso de un diseño ortogonal $X^T X = I_p$).

Para valores crecientes de λ o decrecientes de s , los coeficientes β_j se contraen hacia cero como en Ridge (shrinkage), con la diferencia de que algunos de ellos se anulan.

Esto es, LASSO produce estimación y selección de variables en forma continua y simultánea, siendo especialmente útil en el caso $p \geq n$.

Los avances en los algoritmos para implementar Regresión LASSO en forma eficiente han sido muy importantes. En sus comienzos, la estimación se realizaba resolviendo para cada valor de s el problema de programación cuadrática. El método no es eficiente para un número grande de variables. Posteriormente, surgieron los algoritmos LARS (Efron et al., 2004) y de coordenada

descendente (Friedman et al., 2010) que permitieron reducir enormemente el costo computacional (Tibshirani, 2011).

Una formulación amplia de las técnicas de penalización/regularización puede plantearse como:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \phi_{\lambda}(\beta) \right\}$$

Donde $\beta = (\beta_1, \dots, \beta_p)$, $\lambda \geq 0$ y ϕ es una función de penalización sobre el "tamaño" de β , en general de la forma $\phi_{\lambda}(\beta) = \lambda \sum_{j=1}^p \phi_j(|\beta_j|)$ con ϕ_j creciente en β_j .

Una familia de funciones de penalización muy utilizada es la correspondiente a la norma-Lq, dada por:

$$\phi_{\lambda}(\beta) = \lambda (\|\beta\|_q)^q = \lambda \sum_{j=1}^p |\beta_j|^q, q > 0$$

Los estimadores resultantes en este caso son también conocidos como estimadores Bridge (Fu, 1998). Algunos casos particulares importantes son: $q = 1$ (LASSO) y $q = 2$ (Ridge). Además, los métodos que penalizan por el número de variables pueden ser vistos como el caso límite $q \rightarrow 0$. Para $q > 1$ el estimador no realiza selección de variables (Fan & Li, 2001). Por otro lado, LaSSO corresponde al valor de q más pequeño que produce una región factible convexa.

2.4. Ajuste de modelos polinómicos

Con el fin de comparar el comportamiento de los indicadores entre el grupo de países considerados en la presente investigación, se ajustaron modelos polinómicos de grado 4 como se describe a continuación:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 x_1^3 + \hat{\beta}_4 x_1^4$$

Donde:

Y_i representa el valor del indicador, x_1 el año.

2.5. Análisis no paramétrico

Sobre los indicadores seleccionados mediante las técnicas de regresión PLS y LASSO se realizó un análisis no paramétrico que incluyó la prueba de Kruskal-Wallis y una comparación múltiple basada en rangos con el fin de comparar el grupo de países considerados en el presente estudio en función de cada uno de los indicadores antes señalados. Finalmente, cabe destacar que, los análisis estadísticos antes señalados se realizaron con la ayuda del entorno de programación de libre acceso R-3.6.2 para Windows.

El análisis de regresión PLS y LASSO de datos relacionados con indicadores de desarrollo y bienestar antes descrito se realizó empleando códigos en el entorno del software R (R Core Team, 2021) específicamente con los paquetes "corr", "corrplot", "psych" y "stats".

3. RESULTADOS

En la Tabla 1 se observan los resultados del análisis de regresión LASSO y por mínimos cuadrados parciales (PLS) sobre datos de indicadores sociales en países de Sudamérica. Además, se muestra que la dimensionalidad de la matriz de datos de indicadores de desarrollo se redujo en la regresión PLS. No obstante, el modelo LASSO presenta una ventaja frente al modelo PLS, ya que solo identificó tres indicadores, mientras que el modelo PLS seleccionó 29 componentes principales (Figura 1), que si bien es cierto representa una reducción considerable de la dimensionalidad del conjunto total de indicadores, está muy por encima de la dimensionalidad del modelo LASSO.

Tabla 1.

Análisis de regresión LASSO y PLS sobre datos de indicadores sociales en países de Sudamérica

Modelo de regresión	Indicador	Coefficientes del modelo	Varianza explicada (%)	Error cuadrático medio	λ óptimo
LASSO	% de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación.	-5,224378e-18	18	1,61963e-29	1,00
	% de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad.	1,267632e-17	40		
	% de hombres trabajadores por cuenta propia entre 15 y 24 años de edad.	3,992421e-17	42		
PLS	Número de componentes principales.			29	

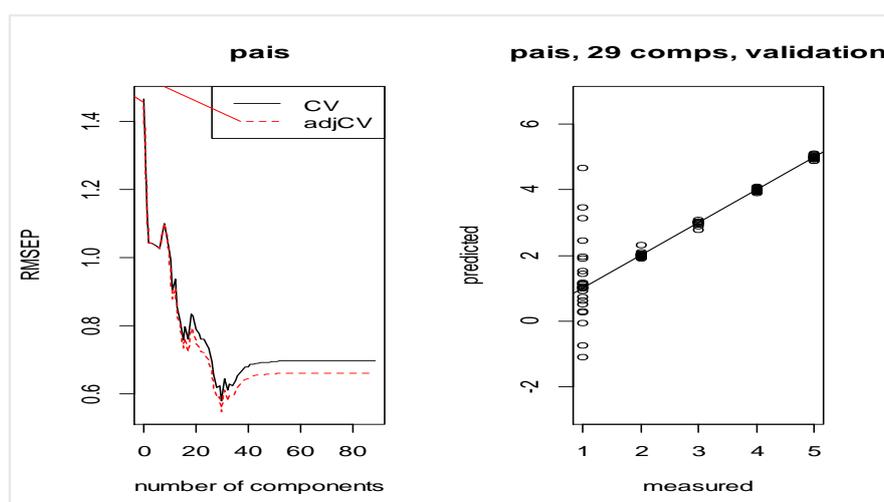


Figura 1. Curva de la raíz cuadrada del error cuadrático medio de predicción (RMSEP) y predicciones de validación cruzada para datos de indicadores sociales en países de Sudamérica

En ese orden, la regresión LASSO consideró tres indicadores que permiten diferenciar a los países de Sudamérica considerados en esta investigación, entre ellos; Argentina, Brasil, Chile, Colombia, Perú y Uruguay. Estos indicadores reflejan características asociadas al género, grupos etarios, acceso a la seguridad social y condición laboral.

En ese sentido, los indicadores asociados a la condición laboral, es decir, % de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad y % de hombres trabajadores por cuenta propia entre 15 y 24 años de edad son los que mejor explican la variabilidad en el modelo de regresión LASSO que permite discriminar entre países, ya que ambos indicadores exhiben la mayor proporción de varianza explicada (40% y 42%), respectivamente, mientras que el indicador % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación muestra una proporción de varianza menor (18%).

Por otro lado, en la Figura 2 se muestra una representación gráfica de las puntuaciones discriminantes de la regresión LASSO sobre indicadores sociales de países de Sudamérica. Allí se observa una clara diferenciación que permite identificar tres grupos de países en función de los indicadores seleccionados.

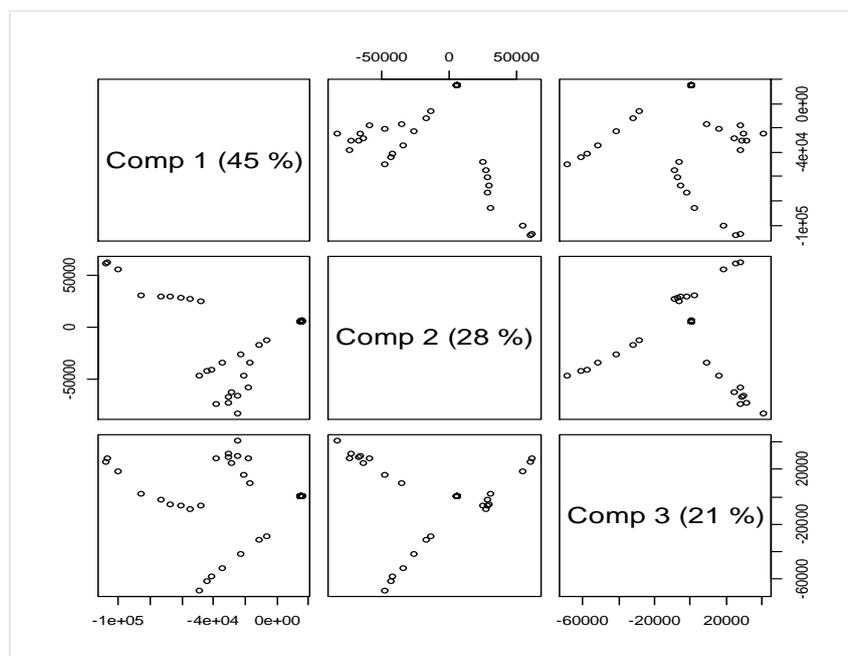


Figura 2. Gráfico de puntuaciones para datos de indicadores sociales en países de Sudamérica

En las Figuras 3, 4 y 5 se muestra la evolución temporal del % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación, % de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad y el % de hombres trabajadores por cuenta propia entre 15 y 24 años de edad en países de Sudamérica, respectivamente.

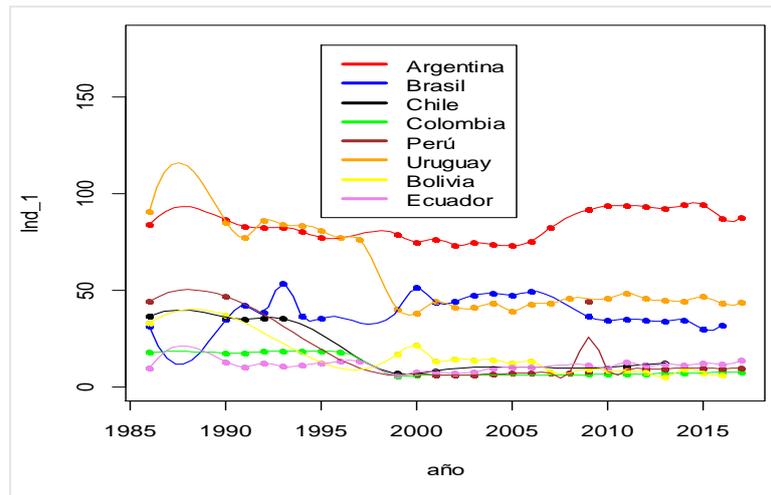


Figura 3. *Tendencia histórica del % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación en países de Sudamérica*

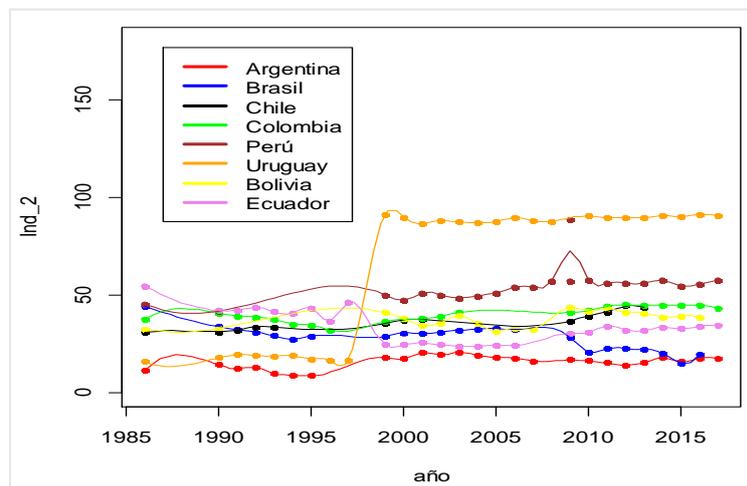


Figura 4. *Tendencia histórica del % de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad en países de Sudamérica*

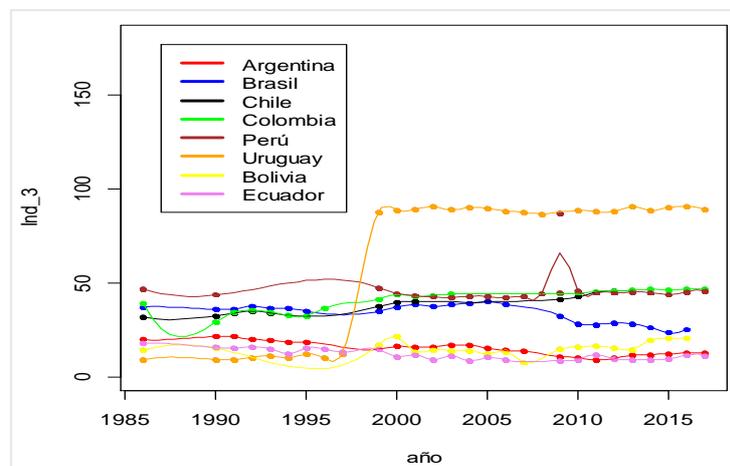


Figura 5. *Tendencia histórica del % de hombres trabajadores por cuenta propia entre 15 y 24 años de edad en países de Sudamérica*

En ese orden, en relación al indicador que mide el % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación en países de Sudamérica (Figura 3), se observa la presencia de dos ciclos estacionales a lo largo de toda la serie, un primer ciclo asociado a las dos primeras décadas de estudio (1986-2000), con una tendencia que muestra un descenso de este indicador.

No obstante, en este primer ciclo de la serie se observa una diferencia entre dos grupos de países, a saber, Uruguay y Argentina quienes exhiben valores altos y similares entre sí en relación a este indicador frente a otro bloque de países como Bolivia, Brasil, Chile, Colombia, Ecuador y Perú, quienes reportan los valores más bajos en el % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación.

Por otro lado, la serie muestra otro ciclo estacional que inicia a partir del año 2000 con valores asociados al indicador que reflejan una tendencia a la estabilidad en países como Bolivia, Colombia, Chile, Ecuador y Perú, quienes se agrupan en un bloque de países con los índices más bajos en relación al % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación.

Lo mismo ocurre con otro bloque intermedio de países, integrado por Brasil y Uruguay, mientras que en este período Argentina muestra un comportamiento diferente al resto de los países, que incluso, además de exhibir los valores más altos para este indicador, llega a evidenciar una tendencia que sugiere un incremento en el % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación.

Seguidamente, los indicadores que miden el % de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad y el % de hombres trabajadores por cuenta propia entre 15 y 24 años de edad en países de Sudamérica (Figuras 4 y 5), respectivamente, muestran un comportamiento similar entre sí, pero distinto al indicador anterior. Cabe destacar que las series asociadas a estos dos indicadores muestran dos ciclos estacionales claramente definidos como en el caso de la serie anterior, un primer ciclo (1986-2000), caracterizado por un único bloque que agrupa a todos los países incluidos en el estudio, quienes muestran una tendencia a la estabilidad de ambos indicadores.

No obstante, las series asociadas a los indicadores que miden el % de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad y el % de hombres trabajadores por cuenta propia entre 15 y 24 años de edad muestran un segundo ciclo definido a partir del año 2000. En ese sentido, se observa un comportamiento similar de estos indicadores en un bloque de países conformados por Argentina, Bolivia, Brasil, Chile, Colombia, Ecuador y Perú, quienes exhiben además de los valores más bajos para estos dos indicadores, una tendencia a la estabilidad a lo largo del periodo que abarca este segundo ciclo, mientras que Uruguay muestra los valores más altos de estos indicadores y un incremento considerable en la segunda mitad de la década de los 90, el cual se estabiliza a partir del año 2000.

Por otro lado, en las Figuras 6, 7 y 8 y en especial, en la Tabla 2, se muestran los resultados de un análisis más exhaustivo de la evolución temporal de estos tres indicadores, relacionados con el ajuste de modelos polinómicos de grado 4. Allí se observa que, en el caso del indicador que mide el % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación,

los datos asociados a países como Bolivia, Chile, Colombia y Uruguay muestran un mejor ajuste ($R^2= 0,9319; 0,9157; 0,8048$ y $0,8181$), respectivamente, mientras que el resto de los países, a saber, Argentina, Brasil, Perú y Ecuador, presentan valores más bajos en el coeficiente de ajuste ($R^2= 0,5916; 0,5760; 0,5734$ y $0,3423$), respectivamente. Estos resultados sugieren un comportamiento similar de este indicador en el caso del primer grupo de países mencionados (Bolivia, Chile, Colombia y Uruguay), que los diferencia del resto de países considerados en el estudio.

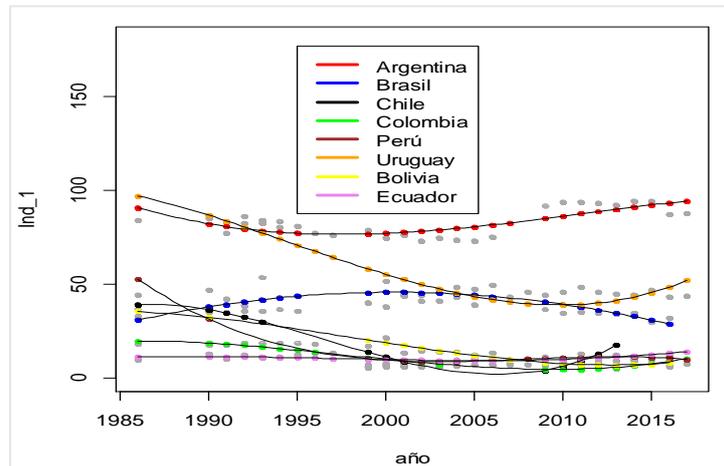


Figura 6. Ajuste de modelos polinómicos sobre el % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación en países de Sudamérica

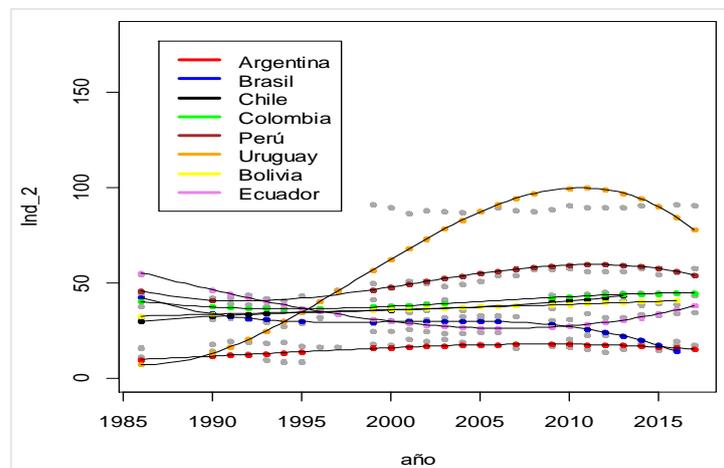


Figura 7. Ajuste de modelos polinómicos sobre el % de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad en países de Sudamérica

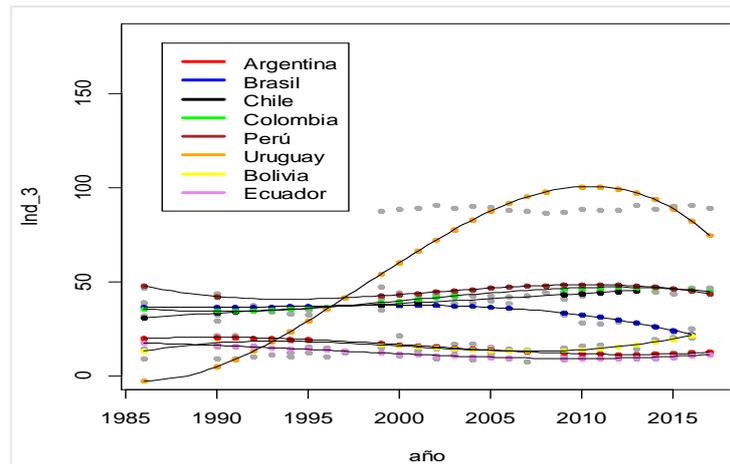


Figura 8. Ajuste de modelos polinómicos sobre el % de hombres trabajadores por cuenta propia entre 15 y 24 años de edad en países de Sudamérica

Tabla 2.

Criterios de selección de modelos polinómicos de grado 4 sobre características asociadas al género, acceso a la seguridad social y condición laboral en países de Sudamérica

Indicador	País	Normalidad de los errores		Criterio de selección del modelo			Error estándar residual	
		W	P valor	Autocorrelación de errores		R ²		AIC
				Durbin-Watson	P valor			
Mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación (%)	Argentina	0,883	0,000	0,27152	0,0000	0,5916	158,8712	5,184
	Brasil			1,3127	0,0267	0,576	144,4239	4,947
	Chile			0,88061	0,0037	0,9157	81,10276	4,48
	Colombia			0,51234	0,0000	0,8048	111,8653	2,709
	Perú			0,35824	0,0000	0,5734	166,2588	9,325
	Uruguay			0,44714	0,0000	0,8181	198,6979	8,633
	Bolivia			1,5515	0,1147	0,9319	93,6553	2,461
	Ecuador			0,9965	0,0012	0,3423	110,4114	1,814
Mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad (%)	Argentina	0,874	0,000	0,3761	0,0000	0,4522	126,5251	2,715
	Brasil			0,60937	0,0000	0,8518	114,8783	2,603
	Chile			0,93672	0,0057	0,9006	55,13688	1,65
	Colombia			0,37914	0,0000	0,7670	98,2695	1,989
	Perú			1,1688	0,0134	0,3699	157,0509	7,564
	Uruguay			0,41043	0,0000	0,8298	228,8716	15,1
	Bolivia			0,93122	0,0026	0,3011	108,7915	3,666
	Ecuador			0,62281	0,0000	0,7496	157,7232	4,506
Hombres trabajadores por cuenta propia entre 15 y 24 años de edad (%)	Argentina	0,852	0,000	0,86529	0,0003	0,8854	90,93463	1,332
	Brasil			0,53966	0,0000	0,8521	105,5805	2,126
	Chile			1,1678	0,0242	0,9605	46,35338	1,177
	Colombia			1,3355	0,0319	0,8343	107,9848	2,48
	Perú			1,0093	0,0037	0,0519	168,1224	9,729
	Uruguay			0,37325	0,0000	0,8351	233,0818	16,32

Bolivia	1,3043	0,0366	0,5267	93,65733	2,456
Ecuador	1,2189	0,0097	0,8252	90,28022	1,232

Seguidamente, en relación al indicador que mide el % de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad, se observa un grupo de países integrado por Brasil, Chile, Colombia, Uruguay y Ecuador, quienes muestran un mejor ajuste ($R^2 = 0,8518; 0,9006, 0,7670; 0,8298$ y $0,7496$), respectivamente, lo que presume una evolución temporal similar de este indicador en este grupo de países, que los diferencia del resto.

No obstante, respecto al indicador que mide el % de hombres trabajadores por cuenta propia entre 15 y 24 años de edad, se observa que Perú muestra el valor más bajo en relación al coeficiente de ajuste para este indicador ($R^2 = 0,0519$), lo que lo diferencia del resto de países, quienes presentan coeficientes de ajuste que van desde 0,8252 hasta 0,9605, lo que se traduce para el caso peruano en una evolución temporal de este indicador muy distinta al resto de países, lo que sugiere un análisis más exhaustivo del mismo.

Además de los resultados anteriores, hay que resaltar los resultados de la prueba de autocorrelación serial de los errores (Durbin-Watson), la cual sugiere que los tres indicadores muestran errores correlacionados en el tiempo, característica típica de las series temporales. Así mismo, se verifica el cumplimiento del supuesto de normalidad de los errores mediante la prueba de Wilk-Shapiro, la cual sugiere que para cada uno de los indicadores los errores muestran una tendencia a distribuirse normalmente ($p \leq 0,05$).

En las Tablas 3 y 4 se muestran los resultados del análisis no paramétrico sobre características asociadas al género, acceso a la seguridad social y condición laboral entre países de Sudamérica. Allí se observa que la prueba de Kruskal-Wallis evidencia diferencias significativas ($p < 0,05$) en el comportamiento de los tres indicadores evaluados entre algunos de los países considerados en este estudio (Tabla 3).

Tabla 3.

Análisis no paramétrico de Kruskal-Wallis sobre características asociadas al género, acceso a la seguridad social y condición laboral en países de Sudamérica

Indicador	Estadístico de prueba y su significación asintótica	
	H	P valor
% de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación	134,6802	0,0000
% de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad	101,7446	0,0000
% de hombres trabajadores por cuenta propia entre 15 y 24 años de edad	57,4565	0,0000

Nota: P valor $\leq 0,05$ sugiere diferencias significativas entre al menos dos países en relación al indicador.

Tabla 4.

Comparación no paramétrica sobre características asociadas al género, acceso a la seguridad social y condición laboral entre países de Sudamérica

País	Indicador		
	% de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación	% de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad	% de hombres trabajadores por cuenta propia entre 15 y 24 años de edad
Argentina	83,48176 A	15,56688 E	15,16184 D
Brasil	39,81587 C	28,11757 D	33,87996 B
Chile	19,546 D	36,63177 C	38,79092 BC
Colombia	10,70414 E	40,06236 BC	41,0295 AB
Perú	12,82868 E	54,34982 A	46,20973 A
Uruguay	56,37852 B	65,386 A	62,80978 AB
Bolivia	13,56137 E	37,53184 C	15,45232 D
Ecuador	10,57781 E	33,76304 C	11,91027 D

Nota: Países con letras iguales sugieren similitudes estadísticas en relación al indicador.

En ese orden, al comparar desde el punto de vista no paramétrico los promedios (basados en rangos) de estos tres indicadores (Tabla 4), se puede evidenciar lo siguiente: en primera instancia, en relación al % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación, se puede observar que solo se identifican similitudes estadísticas en el comportamiento de este indicador entre Ecuador, Colombia, Perú y Bolivia, quienes reportan los valores más bajos de este indicador (10,57%, 10,70%, 12,82% y 13,56%), respectivamente, mientras que los demás países del grupo reportan diferencias estadísticas entre ellos.

No obstante, hay que resaltar el hecho de que Argentina encabeza el grupo con el % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación más alto (83,48%), seguido de Uruguay con 56,37%, Brasil con 39,81% y Chile con 19,54%.

Seguidamente, respecto al indicador que mide el % de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad se observa que, existen similitudes estadísticas entre Perú y Uruguay, quienes reportan los valores más altos para este indicador (54,34% y 65,38%), respectivamente, mientras que Colombia con 40,06% reporta similitudes con Bolivia (37,53%), Chile (36,63%) y Ecuador (33,76%).

No obstante, países como Brasil (28,11%) y Argentina (15,56%), que además de ubicarse como los países con los % más bajo de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación, no exhiben ninguna similitud con el resto de los países considerados.

Finalmente, al referirnos al indicador que mide el % de hombres trabajadores por cuenta propia entre 15 y 24 años de edad, se observa que Perú (46,20%) junto con Uruguay (62,80%), Colombia (41,03%), Chile (38,79%) y Brasil (33,87%) se ubican en bloque de países que se caracteriza por reportar los valores más altos para este indicador.

Sin embargo, países como Ecuador, Argentina y Bolivia se diferencian del anterior bloque de países al presentar los % más bajos (11,91%, 15,16% y 15,45%), respectivamente, de hombres trabajadores por cuenta propia entre 15 y 24 años de edad. Estos resultados, confirman la presencia de grupos de

países con características similares relacionadas con el género, el acceso a la seguridad social y la condición laboral en países de Sudamérica, como lo evidenciaron las metodologías basadas en regresión PLS y LASSO.

CONCLUSIONES

El modelo LASSO presentó ventaja frente al modelo PLS, ya que solo identificó tres indicadores que permitieron diferenciar a los países de Sudamérica considerados en esta investigación, entre ellos; Argentina, Brasil, Chile, Colombia, Perú y Uruguay. Estos indicadores reflejaron características asociadas al género, grupos etarios, acceso a la seguridad social y condición laboral. Los indicadores asociados a la condición laboral (% de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad y porcentaje de hombres trabajadores por cuenta propia entre 15 y 24 años de edad) fueron los que permiten una mejor discriminación entre países.

En relación al porcentaje de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación en países de Sudamérica, se observó la presencia de dos ciclos estacionales a lo largo de toda la serie, un primer ciclo asociado a las dos primeras décadas de estudio (1986-2000), con una tendencia que muestra un descenso de este indicador.

En este primer ciclo de la serie se observó una diferencia entre dos grupos de países; Uruguay y Argentina quienes exhiben valores altos y similares entre sí en relación a este indicador frente a otro bloque de países como Bolivia, Brasil, Chile, Colombia, Ecuador y Perú. El segundo ciclo estacional que inicia a partir del año 2000 muestra una tendencia a la estabilidad en países como Bolivia, Colombia, Chile, Ecuador y Perú, quienes se agrupan en un bloque de países con los índices más bajos en relación al % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación. Lo mismo ocurre con otro bloque intermedio de países, integrado por Brasil y Uruguay, mientras que en este período Argentina muestra un comportamiento diferente al resto de los países, que incluso, además de exhibir los valores más altos para este indicador, llega a evidenciar una tendencia que sugiere un incremento en el % de mujeres de 65 o más años de edad con bajo nivel educativo que reciben una pensión por jubilación.

Los indicadores que miden el porcentaje de mujeres trabajadoras por cuenta propia entre 15 y 64 años de edad y el porcentaje de hombres trabajadores por cuenta propia entre 15 y 24 años de edad, mostraron un comportamiento similar entre sí, con dos ciclos estacionales claramente definidos; un primer ciclo (1986-2000), caracterizado por un único bloque que agrupa a todos los países incluidos en el estudio, quienes muestran una tendencia a la estabilidad de ambos indicadores. Finalmente, estos resultados, confirmaron la presencia de grupos de países con características similares relacionadas con el género, el acceso a la seguridad social y la condición laboral en países de Sudamérica, como lo evidenciaron las metodologías basadas en regresión PLS y LASSO.

FINANCIAMIENTO

Esta investigación recibió apoyo financiero por la Universidad Nacional Autónoma de Alto Amazonas mediante Resolución de Comisión Organizadora N° 131-2021-UNAAA/CO.

CONFLICTO DE INTERESES

No existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Pacheco-Robles, R. A., Vela-Del-Águila, S. y Nureña-Hidalgo, M. A.

Curación de datos: Tuesta-Hidalgo, O., Tuesta-Hidalgo, J. C. y Vela-Lozano, J. M.

Análisis formal: Tuesta-Hidalgo, J. C., Nureña-Hidalgo, M. A., Vela-Del-Águila, S. y Vela-Lozano, J. M.

Investigación: Pacheco-Robles, R. A., Vela-Del-Águila, S., Tuesta-Hidalgo, O., Tuesta-Hidalgo, J. C., Nureña-Hidalgo, M. A. y Vela-Lozano, J. M.

Metodología: Pacheco-Robles, R. A., Vela-Del-Águila, S., Tuesta-Hidalgo, O. y Tuesta-Hidalgo, J. C.

Supervisión: Pacheco-Robles, R. A., Tuesta-Hidalgo, O. y Tuesta-Hidalgo, J. C.

Redacción-borrador original: Pacheco-Robles, R. A., Vela-Del-Águila, S., Nureña-Hidalgo, M. A. y Vela-Lozano, J. M.

Redacción-revisión y edición: Pacheco-Robles, R. A., Vela-Del-Águila, S., Tuesta-Hidalgo, O., Tuesta-Hidalgo, J. C. y Vela-Lozano, J. M.

REFERENCIAS BIBLIOGRÁFICAS

Armando Mendoza, N. (2019). *Brechas Latentes-índice de avance contra la desigualdad en el Perú 2017-2018 (1ta ed.)*. OXFAM.

BID. (2020). *Números para el Desarrollo*. Banco Interamericano de Desarrollo.
<https://code.iadb.org/es/herramientas/numeros-para-el-desarrollo>

Bjorn-Helge, M., & Wehrens, R. (2016). *Introduction to the pls Package*. Cran R. <https://cran.r-project.org/web/packages/pls/vignettes/pls-manual.pdf>

CEPAL. (2021). *Pandemia provoca aumento en los niveles de pobreza sin precedentes en las últimas décadas e impacta fuertemente en la desigualdad y el empleo*. Organización de Las Naciones Unidas. <https://www.cepal.org/es/comunicados/pandemia-provoca-aumento-niveles-pobreza-sin-precedentes-ultimas-decadas-impacta>

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2). <https://doi.org/10.1214/009053604000000067>

Fan, J., & Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96(456), 1348–1360.
<https://doi.org/10.1198/016214501753382273>

Fan, J., & Li, R. (2006). Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery. *Statistics Theory*, 1(3).
<https://doi.org/https://doi.org/10.48550/arXiv.math/0602133>

Fernández, L., & Gutiérrez, M. (2013). Bienestar Social, Económico y Ambiental para las Presentes y Futuras Generaciones. *Información Tecnológica*, 24(2), 121–130.
<https://doi.org/10.4067/S0718-07642013000200013>

- Frank, I. E., & Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, 35(2), 109. <https://doi.org/10.2307/1269656>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1). <https://doi.org/10.18637/jss.v033.i01>
- Fu, W. J. (1998). Penalized Regressions: The Bridge versus the Lasso. *Journal of Computational and Graphical Statistics*, 7(3), 397. <https://doi.org/10.2307/1390712>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación (6ta ed.)*. McGraw-Hill Education.
- INEI. (2021). *Pobreza monetaria alcanzó al 30,1% de la población del país durante el año 2020*. <https://www.inei.gob.pe/prensa/noticias/pobreza-monetaria-alcanzo-al-301-de-la-poblacion-del-pais-durante-el-ano-2020-12875/>
- Londoño Vásquez, D. A., & Castañeda Naranjo, L. S. (2010). La comprensión como método en las ciencias sociales. *Revista Virtual Universidad Católica Del Norte*, 31. <https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/43>
- Mattei, L. (2017a). O debate sobre a reforma agrária no contexto do Brasil rural atual. *Política & Sociedade*, 15, 234. <https://doi.org/10.5007/2175-7984.2016v15nesp1p234>
- Mattei, L. (2017b). O debate sobre a reforma agrária no contexto do Brasil rural atual. *Política & Sociedade*, 15, 234. <https://doi.org/10.5007/2175-7984.2016v15nesp1p234>
- Mevik, B.-H., & Wehrens, R. (2007). The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software*, 18(2). <https://doi.org/10.18637/jss.v018.i02>
- ONU. (2022). *Objetivos del Desarrollo Sostenible*. Organización de Las Naciones Unidas. <https://www.un.org/sustainabledevelopment/es/>
- ONU DAES. (2022). *World Population Prospects 2022*. <https://www.un.org/development/desa/pd/es/content/World-Population-Prospects-2022>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Sen, A. (1985). Resources, Values and Development. *The Economic Journal*, 95(379), 822. <https://doi.org/10.2307/2233061>
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
- UNESCO. (2019). *Informe Mundial de las Naciones Unidas sobre el Desarrollo de los Recursos Hídricos 2019*. <https://www.acnur.org/5c93e4c34.pdf>
- Uribe Mallarino, C. (2004). Desarrollo social y bienestar. *Universitas Humanística*, 58(58). <https://revistas.javeriana.edu.co/index.php/univhumanistica/article/view/9509>