Journal of Namibian Studies, 40 (2024): 280-293 ISSN: 2197-5523 (online)

Multivariate Cluster-Ann Method For The Analysis Of Multidimensional Poverty In The Control And Public Management Of Poverty And Inequality In Colombia

Efrain De la Hoz-Granadillo[1], Valentina De La Hoz-Lopez[2], Tomas Fontalvo-Herrera[3]

[1]Post Ph.D., Ph.D. in Industrial Engineering, Senior researcher of Minciencias. Full Professor of the Industrial Administration Program, Faculty of Economic Sciences, Universidad de Cartagena, Cartagena, Colombia. H-Index: 21. https://orcid.org/0000-0002-5196-813X E-mail: edelahozg@unicartagena.edu.co

[2]Electronic Engineering, Universidad del Norte, Barranquilla, Colombia. IEEE SB Uninorte Board Member. https://orcid.org/0009-0008-0038-7838 E-mail: vndelahoz@uninorte.edu.co

[3]Post Ph.D., Ph.D. in Administration, Ph.D. in Social Sciences. Senior researcher of Minciencias. Full Professor of the Industrial Administration Program, Faculty of Economic Sciences, Universidad de Cartagena, Cartagena, Colombia. H-Index: 34. https://orcid.org/0000-0003-4642-9251. E-mail: tfontalvoh@unicartagena.edu.co

Abstract

The measurement and monitoring of poverty variables are crucial for evaluating government management, serving as key indicators of the quality of life and the success of public policies. This research introduces a multivariate Cluster-ANN method to analyze multidimensional poverty, aiding in public control and management of poverty and inequality in Colombia. The methodology involves cluster analysis to identify multidimensional poverty profiles across dimensions such as education, childhood and youth conditions, employment, health, and housing in all 33 departments of Colombia. Subsequently, an Artificial Neural Network model forecasts a department's membership in a multidimensional poverty profile, revealing characteristic profiles and three poverty levels. This approach allows for the identification of departments facing critical situations. The artificial neural network model exhibits 100% accuracy in predicting department membership based on profiles identified in the cluster analysis.

Keywords: Multidimensional Poverty, Poverty Profiles, Data Analytics, Clustering, Artificial Neural Networks.

Introduction

Poverty transcends mere economic constraints or diminished income; it is increasingly acknowledged as a complex social phenomenon, encompassing deprivation across various dimensions. This state denotes not only significant deficits in formal employment and nutritional aspects but also in access to education and healthcare services (Gamboa et al., 2020). This factor significantly influences the economies of numerous Latin American countries (Munevar, 2019).

In Colombia, as reported by the National Administrative Department of Statistics (Departamento Administrativo Nacional de Estadística-DANE) in 2022, the incidence rate of multidimensional poverty increased by 0.6 percentage points from 2019 to 2020. This led to more than 489,000 individuals falling below the threshold of multidimensional poverty, predominantly concentrated in densely populated urban centers and dispersed rural areas. Conversely, the National Development Plan (Plan Nacional de Desarrollo) aspires to achieve a breakthrough in extreme poverty, reducing the population in conditions of extreme poverty from 12.2% in 2021 to 9.6% in 2026. Moreover, it endeavors to reduce poverty to promote peace, narrowing the gap of multidimensional poverty in municipalities through the Territorial Development Program (PDET) (National Planning Department-DNP, 2023).

These statistics underscore the social challenges associated with poverty, underscoring the imperative of constructing objective and robust knowledge structures. Such structures are vital for informed decision-making in the formulation, management, and oversight of programs and public policies that contribute positively to bridging social gaps and reducing inequality. Against this backdrop, this research addresses key questions: How can multidimensional poverty profiles in the departments of Colombia be identified? What are the distinctive characteristics of multidimensional poverty profiles in Colombia? And, how can a department's affiliation with a multidimensional poverty profile be projected?

Reference Framework

Poverty and Multidimensional Poverty

Multidimensional poverty, defined as the simultaneous and prolonged deprivation experienced by individuals across various dimensions of their lives, has gained prominence in the analysis of the socioeconomic circumstances of diverse populations. As highlighted by Gimenez and Valente (2016), a comprehensive study of poverty necessitates focusing not only on monetary aspects and purchasing power but also delving deeper into the deprivation of fundamental capabilities across various domains, including education, living conditions, and access to health services. This approach enables a more nuanced and comprehensive understanding of the intricate facets of poverty.

According to Manly and Navarro (2016), a pivotal factor in poverty reduction is economic growth geared toward equitable distribution of its benefits. Hence, indices such as the Multidimensional Poverty Index (MPI), introduced by Alkire and Foster in 2011, play a crucial role in quantifying an individual's level of poverty in a more encompassing manner. Furthermore, the enhancement of poverty indices stands as a fundamental determinant for the overall growth of nations and the mitigation of economic inequality (Yanxi & Tiangxuan, 2024).

In more specific terms, as of 2022, DANE reported a national Gini coefficient for Colombia of 0.556. This Gini coefficient, ranging between 0 and 1, serves as a key measure of inequality, where 1 denotes a completely unequal distribution, and 0 signifies an ideologically egalitarian society with uniform income distribution.

Cluster and ANN analysis

Cluster analysis stands out as an unsupervised multivariate statistical analysis technique essential for grouping observations into characteristic profiles within a population. This grouping aids in better understanding the studied population and proves particularly valuable for analyzing business sectors based on performance or management indicators. De La Hoz and López

(2017) emphasize that cluster analysis is instrumental in evaluating business sectors or population groups, serving as a crucial tool in assessing their performance.

Furthermore, Mathai, Provost, and Haubold (2022) assert that cluster analysis is a multivariate exploratory technique designed to classify observations into groups, maximizing their similarity within each group. This technique encompasses both hierarchical and non-hierarchical clustering methods, offering a comprehensive approach to analyzing patterns and relationships within datasets.

Fontalvo, De La Hoz, and Morelos (2018) demonstrated the successful application of cluster analysis in studying financial efficiency within various business contexts. Peña (2002) highlights that cluster analysis discerns patterns or behaviors within a sample, identifying groups characterized by similar intragroup features while exhibiting differences between groups. Moreover, it proves valuable in reducing dimensions or variables in a study problem. The amalgamation process involves a proximity analysis to determine the level of similarity or dissimilarity among observations and establish grouping criteria. Similarity measures how alike two or more observations are, while dissimilarity gauges the differences between samples, with the squared Euclidean distance being a commonly used metric (De la Garza et al., 2013; Hair and Babin, 2018; Manly, Navarro 2016, Manzano and Uriel, 2017).

Markos, Moschidis, and Chadjipantelis (2020) delve into clustering observations using sequential dimension reduction and clustering of mixed-type data. They employ principal component analysis and a hierarchical clustering algorithm or partitioning to scores. This proves to be an effective strategy, especially when categorical variables offer more informative insights than continuous variables regarding the underlying clustering structure.

An Artificial Neural Network (ANN) is a statistical technique that is conceptualized as a structure that, through an algorithm, simulates the behavior of the human brain. It begins with a fundamental unit called a neuron, which responds to an external stimulus through adjustment or activation functions that transform the stimulus into an output (Restrepo, Viloria, and Robles, 2021). As Perez (2017) outlined, artificial neural networks comprise a collection of highly interconnected elements capable of processing information and learning from the data presented to them. They possess significant potential for application across diverse fields, addressing complex problems and contributing to the development of theoretical models. ANNs are applicable in problems in which it is necessary to make forecasts or projections based on the analysis of previous information that allows the development of learning processes in the behavior of the data. In addition, it has potential applications in classificatory analysis.

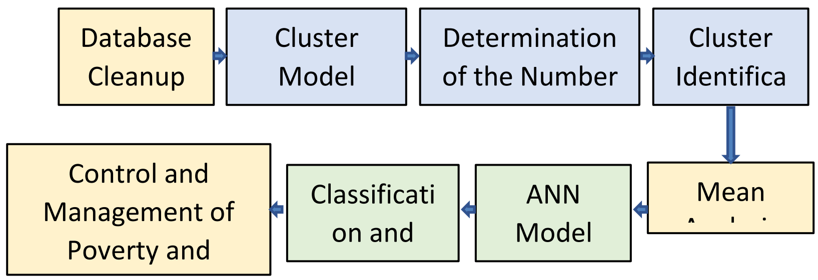
Methodology

This research conducted a quantitative, cross-sectional, and systematic analysis of multidimensional poverty indicators sourced from DANE, derived from the National Quality of Life Survey (ECV). The survey was administered to a sample of 89,203 households across 33 departments in Colombia during 2021. Examined variables span five dimensions and fifteen indicators: Educational conditions (Illiteracy and Low educational achievement), Childhood and Youth conditions (School absenteeism, School lag, Barriers to access early childhood care services, and Child labor), Employment (Informal employment and Long-term unemployment), Health (Lack of health insurance and Barriers to access health when needed), and Housing conditions and public services (Lack access to improved water sources, Inadequate disposal of excreta, Inadequate flooring material, Inadequate wall material, and Critical overcrowding). All indicators were expressed as a percentage of the population experiencing the poverty indicator.

Initially, the information was consolidated and organized. Subsequently, to identify poverty level profiles in Colombia's departments, twelve models were evaluated in the cluster analysis using Minitab software. These models combined four linkage methods (ward, average, centroid, and complete) and three distance metrics (Euclidean, squared Euclidean, and squared Pearson), enabling the establishment of multidimensional poverty clusters.

In the artificial neural network analysis using SPSS software, the poverty indicators from clustering served as factors, and the resulting clusters were the dependent variable. The artificial neural network modeling process partitioned records with 50% for training, 30% for testing, and 20% for validation. Six architectures were evaluated, combining activation functions like Hyperbolic Tangent and Sigmoid in the hidden layers (two layers) and Sigmoid, Hyperbolic Tangent, and Identity in the output layer, along with batch training. This facilitated projecting a department's affiliation to a specific profile. Figure 1 illustrates the adopted methodological design in this research.

Figure 1. Cluster-ANN Method for Analyzing Multidimensional Poverty Profiles



Results

Initially, the database underwent a thorough cleansing process to ensure its validity. Subsequently, a cluster analysis was performed to identify characteristic groups of multidimensional poverty. Employing Minitab 19 software, 12 clustering models were assessed, combining 4 linkage methods (Ward, Centroid, Average, and Complete) and 4 distance measurement methods (Euclidean, squared Euclidean, and Pearson). Information from the multidimensional poverty indicators recorded in Colombia's 33 departments was utilized for this purpose.

The selection of the cluster model was based on the level of homogeneity determined by the average distance from the centroid, the level of heterogeneity based on the distances between cluster centroids, and the dendrogram structure. The chosen model employed the Ward linkage method and squared Euclidean distance measurement, demonstrating superior indicators of homogeneity and heterogeneity in relation to the average distance from the centroid and the distance between cluster centroids.

Furthermore, the elbow diagram criterion, considering the level of distance between observations (De la Garza, et al., 201), was employed to determine the existence of three characteristic profiles or clusters of multidimensional poverty in Colombia (refer to Figure 2). This conclusion finds support in Figure 3 of the dendrogram, illustrating the formation of the three clusters or profiles of multidimensional poverty.

Figure 2. Elbow Plot of Distance Level

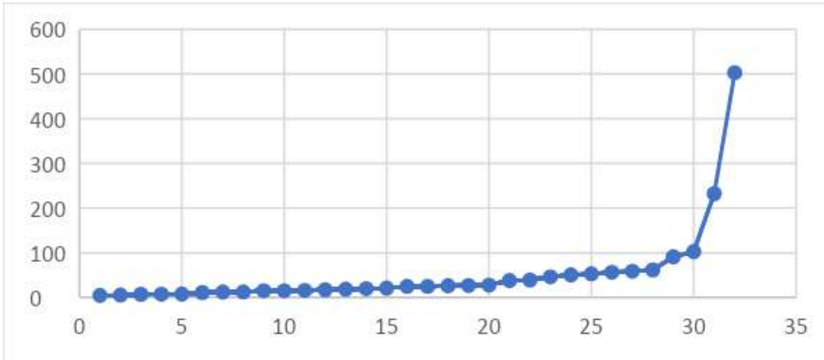


Figure 3. Dendrogram: Multidimensional Poverty Profiles

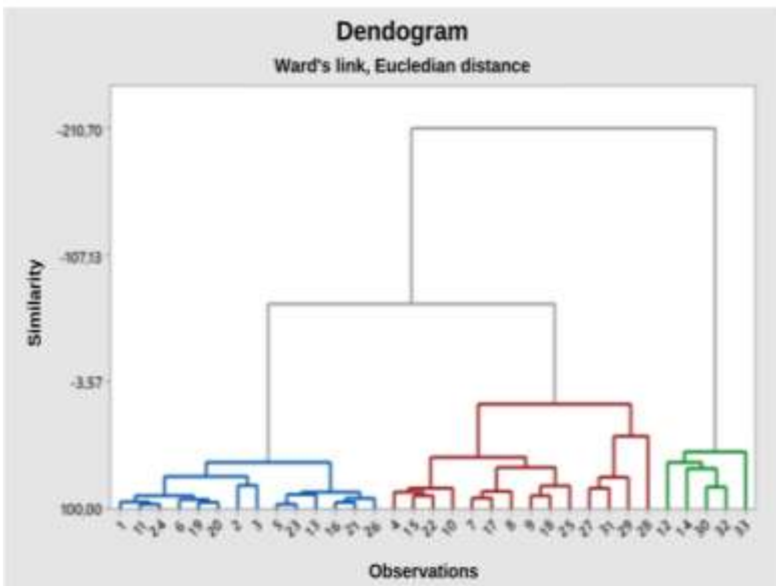


Table 1 presents the final partition achieved for multidimensional poverty profiles, with 14 departments classified in Cluster 1, 14 in Cluster 2, and 5 in Cluster 3. The average distance from the centroid for each cluster is 23.375.

Table 1. Final Partition

	Number of observations	Within the sum of squares of the conglomerate	Average distance from centroid	Maximum distance from centroid
Conglomerate 1	14	2571,9	11,9186	29,0759
Conglomerate 2	14	13313,7	27,4692	73,8245
Conglomerate 3	5	5404,4	30,7372	43,6397

In Table 2, the average distance between clusters is presented, with a mean value of 67.9376, reflecting the level of heterogeneity among the clusters.

Table 2. Distance between Cluster Centroids

	Conglomerate 1	Conglomerate 2	Conglomerate 3
Conglomerate 1	0,000	34,7093	100,678
Conglomerate 2	34,709	0,0000	68,426
Conglomerate 3	100,678	68,4259	0,000

Next, Table 3 exhibits the centroids of multidimensional poverty variables as representative indicators for each cluster group.

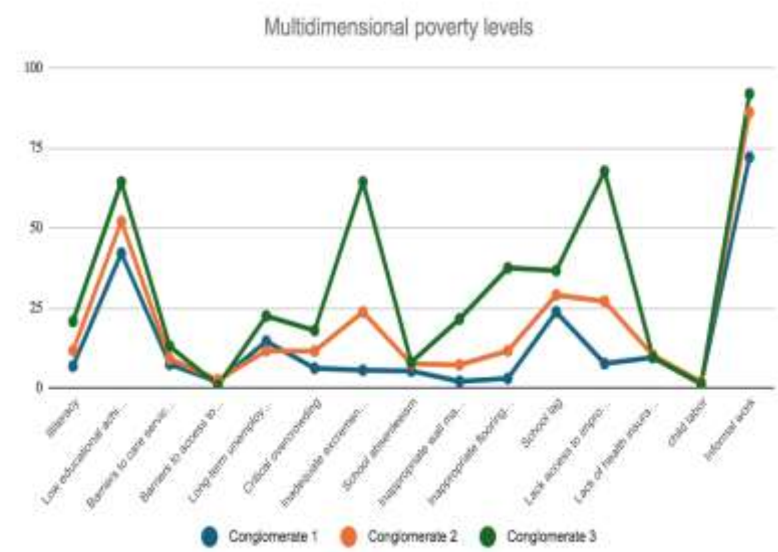
Table 3. Group Centroids

Variable	Conglomerate 1	Conglomerate 2	Conglomerate 3
Illiteracy	6.75	11.6	20.76
Low educational achievement	42.0643	51.9786	64.28
Barriers to childcare services	7.3357	9.0643	13
Barriers to access to services	1.9357	2.4429	0.78
Long-term unemployment	14.5357	11.5714	22.44
Critical overcrowding	6.1214	11.4571	17.9
Inadequate excrement disposal	5.4786	23.7071	64.32
School absenteeism	5.25	7.4643	7.96
Inappropriate wall material	2.0071	7.1571	21.48
Inappropriate flooring material	2.9071	11.55	37.46

School lag	23.7143	28.9857	36.62
Lack access to improved water sources	7.5786	27.0357	67.66
Lack of health insurance	9.45	10.1857	9.56
child labor	1.3214	1.8714	1.22
Informal work	72.05	86.0786	91.86

In Figure 4, a line graph of centroids for each cluster is depicted. Three levels of multidimensional poverty are distinguished, with Cluster 3 representing the most critical level of multidimensional poverty. The prioritized factors in an improvement program for this cluster include “Informal work” at 91.86%, Lack of access to improved water sources (67.66%), Inadequate excrement disposal (64.32%), and Low educational achievement (64.28%). Similarly, Cluster 2 establishes the prioritized poverty factors as Informal work at 86.0786%, Low educational achievement (51.9786%), School lag (28.9857%), and Lack of access to improved water sources (27.0357%). Cluster 1, representing the level with better indicators, identifies the prioritized factors as Informal work (72.05%), Low educational achievement (42.0643%), and School lag (23.7143%).

Figure 4. Levels of Multidimensional Poverty



From these results, Table 4 showcases the departments of Colombia based on the level of multidimensional poverty. Accordingly, level 3 of multidimensional poverty constitutes 15.15% of the departments in Colombia, with a poverty rate of 31.82%. Level 2 encompasses 42.42% of the departments, with a poverty rate of 20.17%, while level 1 includes 42.42% of the departments, with a poverty rate of 13.89%.

Table 4. Classification of Departments by Level of Multidimensional Poverty

Level 1	Level 2	Level 3
Antioquia	Bolívar	Chocó
Atlántico	Caquetá	La Guajira
Bogotá	Cauca	Guainía
Boyacá	Cesar	Vaupés
Caldas	Córdoba	Vichada
Cundinamarca	Magdalena	
Huila	Nariño	
Meta	Norte de Santander	
Quindío	Sucre	
Risaralda	Arauca	
Santander	Putumayo	
Tolima	San Andrés	
Valle del Cauca	Amazonas	
Casanare	Guaviare	

Subsequently, in this study, architectures of Artificial Neural Networks were examined using duplicated data to ensure ample information during the network learning process. A structured architecture was established with a hidden layer, a hyperbolic tangent activation function in the hidden layer, an Identity activation function in the output layer, and a batch training type. The resulting model can predict the affiliation of a department to one of the dimensional poverty levels identified in the cluster analysis. Table 5 presents the classification model results of the ANN, demonstrating a 100% correct classification for the reserve sample in poverty levels 1 and 2. For level 3, no observations were recorded in the reserve sample. These findings underscore the capability of the proposed ANN model to project or forecast a department's affiliation to the identified multidimensional poverty levels.

Table 5. Forecast of Correct Classification by the ANN Model

Ranking

Sample	Observed	Forecasted			
		1,00	2,00	3,00	Percent correct
Training	1,00	14	0	0	100,0%
	2,00	0	12	0	100,0%
	3,00	0	0	2	100,0%
	Overall percentage	50,0%	42,9%	7,1%	100,0%
Sample	1,00	4	0	0	100,0%
	2,00	0	5	0	100,0%
	3,00	0	0	0	0,0%
	Overall percentage	44,4%	55,6%	0,0%	100,0%
Reservation	1,00	4	0	0	100,0%
	2,00	0	3	0	100,0%
	3,00	0	0	0	0,0%
	Overall percentage	57,1%	42,9%	0,0%	100,0%

Dependent variable: cluster

Following the ANN analysis, a study was conducted on the importance of independent variables determining the levels of multidimensional poverty. Table 6 presents, in order of importance, the determining variables for poverty, with Inadequate wall material being the most crucial variable for classification, and Illiteracy being the least significant for the model's discrimination.

Table 6. Importance of Independent Variables

Variable	Importance	Normalized Importance
Inappropriate wall material	,084	100,0%
Inappropriate flooring material	,077	92,5%
Lack of health insurance	,077	92,1%
child labor	,076	90,7%
Critical overcrowding	,067	80,0%
Lack access to improved water sources	,066	79,1%
Long-term unemployment	,065	77,2%
Informal work	,065	77,2%
School absenteeism	,065	77,2%
child labor	,062	73,6%
Low educational achievement	,062	73,5%

Barriers to access to services	,061	73,1%
Inadequate excrement disposal	,060	72,1%
Barriers to care services	,060	71,8%
Illiteracy	,054	64,6%

Conclusions and Contributions

The application of advanced multivariate analysis techniques to comprehend complex social phenomena provides a robust foundation for the systematic design of public policies, rendering a valuable instrument for the precise monitoring and oversight of public management. This approach enables a rigorous evaluation of the attained outcomes in the interventions addressing multifaceted societal challenges. This research significantly contributes to the theoretical framework of public management control by establishing and validating quantitative criteria that enhance the effectiveness of control and management processes.

As a methodological and practical advancement, this study introduces a methodology adept at quantifying and categorizing characteristic profiles of multidimensional poverty conditions across various departments in Colombia. Additionally, it offers a predictive capability regarding departmental affiliations with specific poverty clusters. The robustness of the cluster analysis in identifying characteristic groups related to multidimensional poverty and the efficacy of the proposed methodology in establishing a nuanced ranking of multidimensional poverty levels are emphasized. This nuanced ranking is instrumental in strategically prioritizing resource allocation for the control and management of multidimensional poverty.

Furthermore, the analysis employing artificial neural networks has demonstrated high efficiency in projecting departments into distinct poverty levels, achieving a perfect 100% classification accuracy. Grounded in the model presented and its outcomes, this research equips decision-makers with informed insights for crafting development plans, orchestrating cohesive programs and public policies, and judiciously allocating resources at both national and departmental levels.

It is crucial to underscore that the proposed methodology is versatile, with the potential for application to diverse problematic contexts where the identification of characteristic population groups is essential for comprehensive understanding and targeted interventions.

References

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

DANE (2022). En 2022, en el total nacional, la pobreza monetaria fue del 36,6 % y la pobreza monetaria extrema fue del 13,8 %. Retrieved from https://www.dane.gov.co/files/operaciones/PM/cp-PM-2022.pdf

De la Garza, J., Morales, B., & González, B. (2013). Análisis Estadístico Multivariante, Un Enfoque Teórico y Práctico (1st ed., pp. 150-178). Mc.Graw Hill.

De La Hoz, E., & López Polo, L. (2017). Application of Cluster Analysis Techniques and Artificial Neural Networks for the Evaluation of the Exporting Capability of a Company. Información Tecnológica, 28(4), 67–74. https://doi.org/10.4067/S0718-07642017000400009

Departamento Nacional de Planeación-DNP. (2023). Colombia, Potencia mundial de la vida. Bases del Plan Nacional de Desarrollo 2022-2026. Retrieved from https://colaboracion.dnp.gov.co/CDT/portalDNP/PND-2023/2023-03-17-bases-plan-nacional-desarrollo-web.pdf

Departamento Administrativo Nacional de Estadística-DANE. (2022). Pobreza multidimensional. Resultados 2021. Retrieved from https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2021/presentacion-rueda-de-prensa-pobreza-multidimensional-21.pdf

Fontalvo, T., De La Hoz, E., & Morelos, J. (2018). Método Combinado Análisis de Conglomerado y Análisis Discriminante Multivariado para Identificar y Evaluar Perfiles de Eficiencia Financiera en Empresas Exportadoras. Información Tecnológica, 29(5), 227–234. http://dx.doi.org/10.4067/S0718-07642018000500227

Gamboa, G., Mingorría, S., & Scheidel, A. (2020). The meaning of poverty matters: Trade-offs in poverty reduction programmes. Ecological Economics, 169, 106450. https://doi.org/10.1016/j.ecolecon.2019.106450

Gimenez, C., & Valente, X. (2016). An approach to poverty from the capabilities approach Amartya Sen. Provincia, núm. 35, pp. 99-149, 2016. Universidad de los Andes. https://www.redalyc.org/journal/555/55548904005/html/

Hair, J., & Babin, B. (2018). Multivariate Data Analysis. Cengage. ISBN: 978-1-4737-5654-0

Manly, B., & Navarro, J. (2016). Multivariate Statistical Methods: A Primer. CRC Press.https://doi.org/10.1201/9781315382135

Manzano, A., & Uriel, E. (2017). Análisis multivariante aplicado con R. 2ª ed. Ediciones Paraninfo. ISBN 13: 9788428329699

Markos, A., Moschidis, O., & Chadjipantelis, T. (2020). Sequential dimension reduction and clustering of mixed-type data. International Journal of Data Analysis Techniques and Strategies, 12(3), 228–246. https://doi.org/10.1504/IJDATS.2020.108043

Mathai, A., Provost, S., & Haubold, H. (2022). Multivariate statistical analysis in the real and complex domains. Springer Nature. https://doi.org/10.1007/978-3-030-95864-0

Munévar, W. (2019). Pobreza, exclusión social y comportamiento económico de los países latinoamericanos. Revista Observatorio de la Economía Latinoamericana, (2019), 1-16. https://www.eumed.net/rev/oel/2019/02/pobreza-exclusion-social.html

Peña, D. (2002). Análisis de Datos Multivariantes. Mc.Graw Hill. Madrid, España.

Pérez, C. (2017). Redes Neuronales a través de ejemplo. Aplicaciones con Matlab. ISBN; 9781978319974.

Restrepo, D., Viloria, J., & Robles, C. (2021). El camino a las Redes Neuronales Artificiales. Editorial Unimagdalena.

Yanxi, B., & Tiangxuan, L. (2024). Multidimensional poverty and growth: Evidence from India 1998–2021. Economic Modelling, 130(106586). https://doi.org/10.1016/j.econmod.2023.106586