

MEDICIÓN DE LA POBREZA EN EL PERÚ: UN ANÁLISIS MULTIVARIANTE DESDE UN ENFOQUE MULTIDIMENSIONAL

CARLOS TEODORO CAYCHO CHUMPITAZ

<https://orcid.org/0000-0002-8087-2807>

Universidad Nacional del Santa, Escuela de Posgrado,
doctorado de Estadística Matemática, Chimbote, Perú
Universidad de Lima, Facultad de Ciencias Empresariales y Económicas,
carrera de Economía, Lima, Perú
ccaycho@ulima.edu.pe

Recibido: 7 de noviembre del 2023 / Aceptado: 19 de enero del 2024

doi: <https://doi.org/10.26439/ddee2024.n005.6741>

RESUMEN. Este artículo contribuye al logro del objetivo de desarrollo sostenible 1, fin de la pobreza (ODS 1), a través de la aplicación de un modelo de análisis multivariante para determinar si un hogar es pobre o no pobre, a partir de veintiún variables obtenidas de la base de datos de la Encuesta Nacional de Hogares (ENAH0). Al respecto, se usó el enfoque de medición multidimensional de la pobreza y la técnica de análisis multivariante árboles de decisión. Asimismo, se usaron como dimensiones de estudio las características de los miembros del hogar, las características de la vivienda, la escolaridad de los miembros del hogar, el combustible para cocinar, los activos del hogar y los ingresos y gastos del hogar. El modelo obtenido consideró como determinantes, si un hogar es pobre o no pobre, a las variables gasto monetario per cápita mensual, número de miembros del hogar y si la vivienda del hogar tiene piso de tierra.

PALABRAS CLAVE: pobreza multidimensional / análisis multivariante / ODS 1 / Perú

Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

A MULTIVARIATE ANALYSIS MODEL FOR THE MEASUREMENT OF POVERTY IN PERU

ABSTRACT. This article contributes to the research related to the Sustainable Development Goal 1: End Poverty (SDG 1) by applying a multivariate analysis model to determine whether a household is poor or non-poor based on 21 variables obtained from the ENAHO database. In this regard, the multidimensional poverty measurement approach and the multivariate analysis technique of decision trees were used. Likewise, the following study dimensions were the characteristics of the household members, the characteristics of the home, the education of the household members, cooking fuel, household assets, and household income and expenses. The model obtained considered as determinants whether a household is poor or not poor the variables: monthly per capita monetary expenditure, number of household members and whether the household's home has a dirt floor.

KEYWORDS: multidimensional poverty / multivariate analysis / SDG 1 / Peru

Códigos JEL: I32, O54

1. INTRODUCCIÓN

La pobreza es un fenómeno social, complejo y multidimensional. Por ello, el enfoque de medición multidimensional es el que permite una mejor comprensión de este fenómeno, porque brinda una visión más detallada, dado que se puede identificar a la pobreza desde aspectos relacionados con la educación, la salud, la vivienda, entre otros. No obstante, en el Perú, el Instituto Nacional de Estadística e Informática (INEI), ente oficial que mide la pobreza, no cuenta con indicadores de pobreza bajo el enfoque de medición multidimensional. Por ello, en este trabajo se utilizó un modelo multidimensional para la medición de la pobreza en el Perú propuesto por Alkire y Foster (2008) y la técnica de análisis multivariante árboles de decisión, que considera veintiún variables independientes y la condición de pobreza como variable dependiente que tiene las categorías de los hogares (1 = pobre y 2 = no pobre) de la base de datos sumaria de la encuesta nacional de hogares (ENAH0) del 2018 (INEI, 2018a). Sin embargo, al igual que otras investigaciones, para medir la pobreza bajo este enfoque, existen limitaciones en los datos dada la complejidad de conseguir información sobre educación, salud, seguridad y nutrición de la misma población. Por ejemplo, la base de datos sumaria de la ENAH0 2018 no cubre aspectos sobre salud, seguridad ni nutrición (INEI, 2018a).

De un tiempo a esta parte, se ha incrementado la preocupación por la sostenibilidad. Por ello, el enfoque de medición de la pobreza multidimensional contribuye con el ODS 1 (fin de la pobreza) y, sobre todo, con el objetivo 1.2, el cual busca erradicar al menos a la mitad de la proporción de la población que vive en situaciones de pobreza en todas sus dimensiones. De igual manera, también contribuye a otros ODS en base a la dimensión en que se encuentran, tales como salud (ODS 3), educación (ODS 4) y vivienda (ODS 11). Por todo ello, este trabajo aporta en el análisis de la pobreza con el uso de análisis estadístico multivariante para determinar las variables necesarias y reducir las tasas de pobreza de una manera sostenible y multidimensional; además, permite el estudio simultáneo de datos y múltiples variables de cada individuo o unidad de investigación. Asimismo, la técnica multivariante que se usa son los árboles de decisión, que son algoritmos que se aplican sobre un conjunto de datos a fin de obtener un modelo (Berlanga et al., 2013; Pérez & Santín, 2007).

2. REVISIÓN DE LITERATURA

En la revisión de literatura para la medición de la pobreza, existen diversos enfoques: multidimensional, de necesidades básicas insatisfechas (NBI), subjetivo, relativo y mixto integrado (Pérez-Campos & Rodríguez, 2015). El más usado en los últimos tiempos es el enfoque de medición multidimensional.

Según Clausen (2019), desde fines del siglo xx a nivel global, ha crecido el interés por el concepto de pobreza multidimensional, así como los esfuerzos para comprenderla,

medirla y eliminarla. Este interés ha motivado la propuesta del índice de pobreza humana (IPH) y los lineamientos de las Naciones Unidas (ONU), como los objetivos del milenio (ODM), en el 2000, y los objetivos de desarrollo sostenible (ODS), en el 2015.

El Programa de las Naciones Unidas para el Desarrollo (PNUD), en su interés por analizar y estudiar de manera multidimensional el desarrollo humano, propuso el índice de desarrollo humano (IDH). Luego, en el 2010, desarrolló con Oxford Poverty and Human Development Initiative (OPHI) el índice de pobreza multidimensional (IPM), indicador que tiene como objetivo consolidarse como una medida de pobreza aguda no monetaria que permita hacer un análisis comparativo entre economías de ingresos bajos y medios (Alkire & Santos, 2014). La revisión de la literatura indica que, a inicios del siglo XXI, las investigaciones como las de Bourguignon y Chakravarty (2003) y Atkinson (2003), sobre la medición multidimensional de la pobreza, sentaron una base para medir las interrelaciones entre las diferentes dimensiones de la pobreza, utilizando el índice de pobreza Foster-Greer-Thorbecke (FGT) y el enfoque de conteo de pobreza que se implementó en Europa.

Muchos países como Armenia, Chile, Colombia, Ecuador, Filipinas, México, Mozambique, Pakistán y Vietnam han desarrollado e implementado medidas multidimensionales de la pobreza en el marco de los ODS (Zavaleta et al., 2018). Asimismo, otras divisiones de la ONU también tuvieron otros planteamientos. La Comisión Económica para América Latina y el Caribe (2014) propuso un índice de pobreza multidimensional, específicamente para la realidad latinoamericana, a través de información de las NBI de dieciocho países de la región. De igual forma, el Fondo de las Naciones Unidas para la Infancia (UNICEF, 2015) midió la pobreza multidimensional en las poblaciones más jóvenes a nivel internacional de un grupo de más de cuarenta países con base en el *multiple overlapping deprivation analysis* (MODA).

La perspectiva para analizar y medir la pobreza, en estos últimos tiempos, vincula el uso del enfoque multidimensional de la pobreza con el análisis estadístico multivariante. Al respecto, se cuenta con los trabajos de Aguilar Estrada et al. (2018ab), quienes investigaron utilizando métodos estadísticos multivariantes, tales como el análisis de conglomerados. En la investigación de Colafranceschi et al. (2011), se trabajó con un grupo poblacional, al cual se identificó como pobre multidimensional a través de distintas dimensiones y se aplicó el análisis de correspondencias múltiples y el análisis de conglomerados. Otro trabajo fue el de Poza y Fernández (2010), quienes realizaron un análisis factorial de segundo orden con el objetivo de construir un indicador de pobreza multidimensional y detectar los principales focos de riesgo de pobreza en España.

En esta investigación se utilizaron los árboles de decisión, porque es una técnica fácil de comprender y analizar, además porque permite un análisis rápido que cuantifica el costo y las probabilidades de los diferentes resultados, lo que brinda una ayuda en la toma de decisiones (Zuniga & Abgar, 2011). Sobre ello, se puede indicar que los árboles

de decisión, de acuerdo con Hair et al. (2004), son parte de una técnica de análisis multivariante, basada en reglas que consisten en nodos (puntos de decisión) y de ramas (conexiones entre nodos) que alcanzan múltiples resultados al estar de paso entre dos o más nodos. En esta técnica, se realizan particiones secuenciales en el conjunto o matriz de datos con el objeto de maximizar las diferencias en la variable dependiente.

3. POBLACIÓN Y MUESTRA

La población¹ bajo estudio está constituida por todos los hogares que residen en las viviendas particulares del área urbana y rural de los departamentos del país y de la provincia constitucional del Callao. En esta investigación, la unidad de análisis es el hogar. De acuerdo con los datos levantados por los censos nacionales del 2017 (XII de Población, VII de Vivienda y III de Comunidades Indígenas) (INEI, 2018b), el tamaño de la población es de 8 252 284 hogares residentes en viviendas particulares con ocupantes presentes.

Por otro lado, según el INEI (2016, p. 147), la muestra es probabilística, de áreas, estratificada, multietápica e independiente en cada departamento de estudio. El tamaño de muestra se obtuvo de la ENAHO 2018 y es de 37 462 hogares en vivienda particulares con ocupantes entrevistados (véase la Tabla 1).

Tabla 1
Cantidad de hogares entrevistados por departamento, según área geográfica

		Área geográfica		
		Área urbana	Área rural	Total
Departamento	Amazonas	555	670	1225
	Áncash	1133	801	1934
	Apurímac	383	567	950
	Arequipa	1807	620	2427
	Ayacucho	623	569	1192
	Cajamarca	555	912	1467
	Callao	1011	0	1011
	Cusco	608	685	1293
	Huancavelica	351	688	1039
	Huánuco	553	735	1288
	Ica	1297	257	1554
	Junín	899	668	1567

(continúa)

1 La población considerada en este trabajo proviene de los resultados de los censos nacionales del 2017 publicados en la nota de prensa 155 del 7 de septiembre del 2018 del INEI (2018b). https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1539/

(continuación)

		Área geográfica		
		Área urbana	Área rural	Total
Departamento	La Libertad	1151	415	1566
	Lambayeque	1084	353	1437
	Lima	4163	356	4519
	Loreto	924	564	1488
	Madre de Dios	674	311	985
	Moquegua	837	160	997
	Pasco	669	610	1279
	Piura	1188	473	1661
	Puno	753	1085	1838
	San Martín	869	473	1342
	Tacna	1178	193	1371
	Tumbes	700	164	864
	Ucayali	813	355	1168
	Total	24 778	12 684	37 462

Nota. Elaboración propia con base en la información de INEI (2018).

4. VARIABLES E INDICADORES

La base de datos sumaria de la ENAHO 2018 tiene la variable original pobreza, la cual es resultado del procesamiento INEI-ENAHO 2018 que tiene tres categorías que determinan si un hogar es “1 = pobre extremo”, “2 = pobre no extremo” o “3 = no pobre” (INEI, 2018a). Para la aplicación de la técnica de árboles de decisión en el presente trabajo de investigación, se definió la variable dependiente pobreza_2 mediante una recodificación a dos grupos: un hogar es “1 = pobre” o “2 = no pobre”. Asimismo, se trabajaron con 21 variables independientes, las cuales fueron obtenidas de la mencionada base de datos (ENAHO 2018a) (véase la Tabla 2).

Tabla 2
Variables independientes

Variab	Nombre de la variable
1. Percep	Número total de perceptores de ingresos en el hogar
2. Mierper	Número total de miembros en el hogar
3. Estrsocial	Estrato socioeconómico
4. luz1	En el hogar no se tiene electricidad
5. telef2	En el hogar se tiene teléfono fijo
6. potable2	En el hogar no se tiene acceso al agua potable

(continúa)

(continuación)

Variables	Nombre de la variable
7. combusti	En el hogar se cocina con bosta, estiércol, leña o carbón
8. radio	En el hogar se tiene radio
9. tvcolor	En el hogar se tiene TV a color
10. bici	En el hogar se tiene bicicleta
11. refri	En el hogar se tiene refrigeradora/congeladora
12. auto	En el hogar se tiene auto/camioneta
13. moto	En el hogar se tiene moto
14. camion	En el hogar se tiene camión
15. miem_esco	Número de miembros del hogar en edad escolar
16. toestud1	Número de miembros del hogar que no asisten al colegio
17. anoestu	Número total de años de escolaridad de los miembros del hogar
18. pisost	La vivienda del hogar tiene piso de tierra o arena
19. ingmonepercap	Ingreso monetario per cápita del hogar
20. gasmonepercap	Gasto monetario per cápita del hogar
21. Area	Área geográfica

5. TÉCNICAS DE PROCESAMIENTO Y ANÁLISIS DE LOS DATOS

Para este trabajo se utilizaron los siguientes *softwares* estadísticos: R y RStudio y el IBM® SPSS® Statistics versión 26. Por un lado, con el *software* estadístico R, se revisó las diferentes formas de aplicación de la técnica de los árboles de decisión, y se eligió el análisis CHAID. Luego, se aplicó el IBM® SPSS® Statistics, porque la base de datos tenía el formato SAV del SPSS y dicho *software* cuenta con la técnica del análisis CHAID.

En el procesamiento se usó el análisis CHAID, que construye un árbol que examina las relaciones entre las veintinueve variables independientes y la variable dependiente. Este análisis crea un diagrama de árbol que identifica las variables independientes que predicen en forma significativa a la condición de “pobre” de un hogar. En nuestro caso, contamos con la base de datos sumaria de la ENAHO 2018, la cual es un resumen a nivel de hogar de esta encuesta anual de hogares que han sido clasificados como “pobres” o “no pobres”. La técnica generó una serie de tablas de contingencia con las que puede relacionar a la variable dependiente con cada una de las veintinueve variables independientes.

El análisis CHAID es el más usado de los cuatro métodos que proporciona el IBM® SPSS® Statistics versión 26 (CHAID, CHAID Exhaustivo, CRT y QUEST). Este análisis puede producir más de dos grupos en cualquier nivel de crecimiento del árbol, por lo cual, crea un árbol más amplio que los métodos crecientes binarios. Por ello, se considera que es aplicable para todo tipo de variable. De ese modo, dicho análisis inicia aplicando la prueba de independencia chi-cuadrado con cada una de las variables independientes y la variable dependiente. Así selecciona a la variable independiente que tiene una

asociación más significativa con la variable dependiente. En el caso de que una variable independiente tenga más de dos categorías, las compara, luego une las categorías que no muestran ninguna diferencia y muestra resultados en forma conjunta. Este proceso de unir categorías finaliza cuando todas las categorías restantes difieren a un nivel de significación específico. De esta manera, y en forma sucesiva, crea sus grupos y los divide por la variable independiente más relacionada con la variable dependiente. Este proceso finaliza cuando no encuentra variables independientes que estén relacionadas con la variable dependiente en los grupos formados. Luego, se obtienen los denominados nodos terminales de los grupos finales que ya no se puede subdividir.

6. RESULTADOS Y DISCUSIÓN

En la Figura 1 y la Tabla 3 se muestra el árbol de decisión y su resumen. Cada caja del diagrama representa un grupo o subgrupo de casos que se denominan nodo. El primero (nodo 0) representa la muestra entera de los casos; esto es, los 6749 hogares “pobres” ($\text{pobreza_2} = 1$) que representan el 18 % del total de hogares del archivo de datos y los 30 713 hogares “no pobres” ($\text{pobreza_2} = 2$) que corresponden al 82 % del total de hogares del archivo de datos. A este nodo raíz se le denomina nodo parental o padre.

De las veintiún variables independientes que se utilizaron, la variable gasto monetario per cápita (*gasmonepercap*) es la que tiene la más alta asociación con la variable dependiente y se puede considerar como la mejor predictora para considerar a un hogar como pobre. Por esta variable, la muestra se ha clasificado en cuatro grupos de hogares:

- $\text{gasmonepercap} \leq 2047,926$ (20% de los hogares [nodo 1])
- $2047,926 < \text{gasmonepercap} \leq 3564,337$ (20% de los hogares [nodo 2])
- $3564,337 < \text{gasmonepercap} \leq 5488,693$ (20% de los hogares [nodo 3])
- $\text{gasmonepercap} > 5488,693$ (40% de los hogares [nodo 4])

A estos nodos (1, 2, 3 y 4), se les denomina nodos filiales.

Luego, para el nodo 1, cuando $\text{gasmonepercap} \leq 2047,926$ se consideró a la variable “total de miembros del hogar” (*mieperho*) por ser la siguiente variable con la más alta asociación con la variable dependiente, se obtuvo 7492 hogares (20 %) clasificados en cinco grupos:

- $\text{mieperho} \leq 1,0$ (1042 hogares [2,8 %, nodo 5])
- $1,0 < \text{mieperho} \leq 2,0$ (1552 hogares [4,1 %, nodo 6])
- $2,0 < \text{mieperho} \leq 3,0$, (1062 hogares [2,8 %, nodo 7])
- $3,0 < \text{mieperho} \leq 5,0$ (2308 hogares [6,2 %, nodo 8])
- $\text{mieperho} > 5,0$ (1528 hogares [4,1 %, nodo 9]).

Figura 1
Árbol de decisión

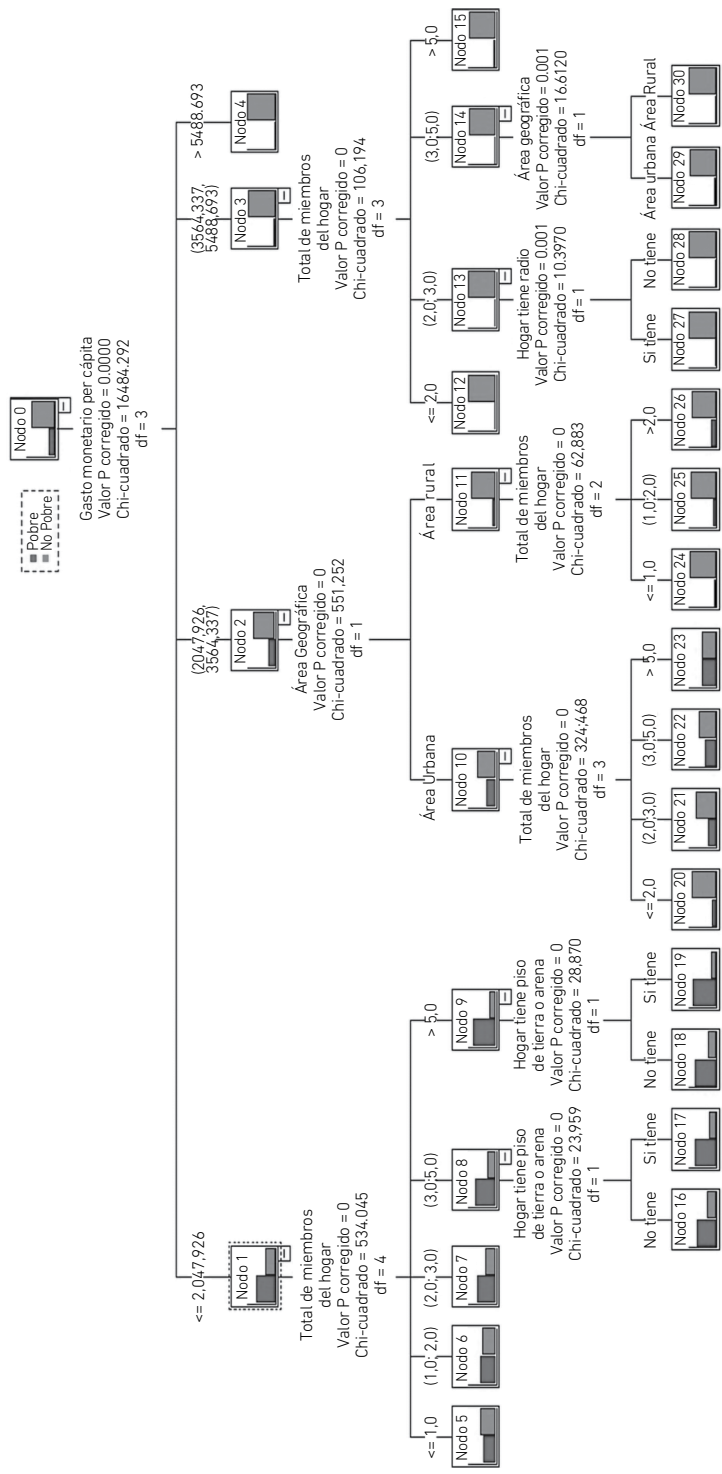


Tabla 3
Resumen del árbol de decisión

Nodo	1		2		Total	
	Número	Porcentaje	Número	Porcentaje	Número	Porcentaje
0	6749	18,0 %	30 713	82,0 %	37 462	100,0 %
1	4962	66,2 %	2530	33,8 %	7492	20,0 %
2	1646	22,0 %	5847	78,0 %	7493	20,0 %
3	141	1,9 %	7351	98,1 %	7492	20,0 %
4	0	0,0 %	14 985	100,0 %	14 985	40,0 %
5	461	44,2 %	581	55,8 %	1042	2,8 %
6	849	54,7 %	703	45,3 %	1552	4,1 %
7	701	66,0 %	361	34,0 %	1062	2,8 %
8	1711	74,1 %	597	25,9 %	2308	6,2 %
9	1240	81,2 %	288	18,8 %	1528	4,1 %
10	1273	32,8 %	2608	67,2 %	3881	10,4 %
11	373	10,3 %	3239	89,7 %	3612	9,6 %
12	3	0,1 %	2041	99,9 %	2044	5,5 %
13	14	1,0 %	1455	99,0 %	1469	3,9 %
14	63	2,3 %	2690	97,7 %	2753	7,3 %
15	61	5,0 %	1165	95,0 %	1226	3,3 %
16	523	67,8 %	248	32,2 %	771	2,1 %
17	1188	77,3 %	349	22,7 %	1537	4,1 %
18	442	74,4 %	152	25,6 %	594	1,6 %
19	798	85,4 %	136	14,6 %	934	2,5 %
20	134	13,1 %	890	86,9 %	1024	2,7 %
21	179	28,3 %	454	71,7 %	633	1,7 %
22	564	39,2 %	876	60,8 %	1440	3,8 %
23	396	50,5 %	388	49,5 %	784	2,1 %
24	22	3,7 %	578	96,3 %	600	1,6 %
25	53	6,7 %	738	93,3 %	791	2,1 %
26	298	13,4 %	1923	86,6 %	2221	5,9 %
27	2	0,2 %	832	99,8 %	834	2,2 %
28	12	1,9 %	623	98,1 %	635	1,7 %
29	63	2,9 %	2126	97,1 %	2189	5,8 %
30	0	0,0 %	564	100,0 %	564	1,5 %

Nota. Método de crecimiento: CHAID. Variable dependiente: pobreza_2. Bonferroni ajustado.

(continúa)

(continuación)

Categoría pronosticada	Nodo padre	Variable independiente primaria	Sig. ^a	Chi-cuadrado	gl	Valores de división
2						
1	0	gasmonepercap	0,000	16 484,292	3	< = 2047,926
2	0	gasmonepercap	0,000	16 484,292	3	(2047,926, 3564,337]
2	0	gasmonepercap	0,000	16 484,292	3	(3564,337, 5488,693]
2	0	gasmonepercap	0,000	16 484,292	3	> 5488,693
2	1	mieperho	0,000	534,045	4	< = 1,0
1	1	mieperho	0,000	534,045	4	(1,0, 2,0]
1	1	mieperho	0,000	534,045	4	(2,0, 3,0]
1	1	mieperho	0,000	534,045	4	(3,0, 5,0]
1	1	mieperho	0,000	534,045	4	> 5,0
2	2	Area	0,000	551,252	1	Área urbana
2	2	Area	0,000	551,252	1	Área rural
2	3	mieperho	0,000	106,194	3	< = 2,0
2	3	mieperho	0,000	106,194	3	(2,0, 3,0]
2	3	mieperho	0,000	106,194	3	(3,0, 5,0]
2	3	mieperho	0,000	106,194	3	> 5,0
1	8	pisost	0,000	23,959	1	No tiene
1	8	pisost	0,000	23,959	1	Sí tiene
1	9	pisost	0,000	28,870	1	No tiene
1	9	pisost	0,000	28,870	1	Sí tiene
2	10	mieperho	0,000	324,468	3	< = 2,0
2	10	mieperho	0,000	324,468	3	(2,0, 3,0]
2	10	mieperho	0,000	324,468	3	(3,0, 5,0]
1	10	mieperho	0,000	324,468	3	> 5,0
2	11	mieperho	0,000	62,883	2	< = 1,0
2	11	mieperho	0,000	62,883	2	(1,0, 2,0]
2	11	mieperho	0,000	62,883	2	> 2,0
2	13	radio	0,001	10,397	1	Sí tiene
2	13	radio	0,001	10,397	1	No tiene
2	14	Area	0,000	16,612	1	Área urbana
2	14	Area	0,000	16,612	1	Área rural

Nota. Método de crecimiento: CHAID. Variable dependiente: pobreza_2. Bonferroni ajustado.

En el nodo 9 se consideró a la variable hogar tiene piso de tierra o arena (*pisost*) por ser la siguiente variable con la más alta asociación con la variable dependiente. En este caso, a los 1528 hogares (4,1 %) se les clasificó en dos grupos:

- *pisost* = 2, 594 hogares (1,6 %, nodo 18)
- *pisost* = 1, 934 hogares (2,5 %, nodo 19)

Estos últimos nodos (el 18 y 19) son considerados como nodos terminales.

Al analizar la Figura 1 y la Tabla 3, se puede observar que, en los subgrupos formados por *gasmonepercap* $\leq$ 2047,926, la proporción de hogares pobres es del 66,2 %. Dentro de este subgrupo, se puede observar que cuando se pone atención a los subgrupos formados por *mieperho* $>$ 5,0, la proporción de hogares pobres es del 81,2 %. Si se observa dentro de este penúltimo subgrupo, al analizar los subgrupos formados por hogar si tiene piso de tierra o arena (*pisost* = 1), la proporción de hogares pobres es del 85,4 %. Esta proporción de hogares pobres es mayor que el 18 % obtenido en la muestra global.

Luego, se tiene que el árbol de decisión tiene una profundidad de tres niveles, porque se han considerado tres particiones por las variables:

- Gasto monetario per cápita: *gasmonepercap* $\leq$ 2047,926
- Total de miembros del hogar: *mieperho* $>$ 5,0
- Hogar tiene piso de tierra o arena: *pisost* = 1

En la Tabla 4, se muestran las ganancias en cada uno de los veintinueve nodos terminales que tiene el diagrama de árbol que hemos construido. Esta tabla presenta el resumen de cada uno de estos nodos terminales, los cuales se pueden identificar con el número del nodo en la primera columna de la izquierda de la tabla. Los nodos terminales están ordenados en forma descendente por la columna respuesta.

Tabla 4
Ganancias por nodos del modelo

Nodo	Nodo		Ganancia		Respuesta	Índice
	Número	Porcentaje	Número	Porcentaje		
19	934	2,5	798	11,8	85,4 %	474,3
17	1537	4,1	1188	17,6	77,3 %	429,0
18	594	1,6	442	6,5	74,4 %	413,0
16	771	2,1	523	7,7	67,8 %	376,5
7	1062	2,8	701	10,4	66,0 %	366,4
6	1552	4,1	849	12,6	54,7 %	303,6
23	784	2,1	396	5,9	50,5 %	280,4
5	1042	2,8	461	6,8	44,2 %	245,6
22	1440	3,8	564	8,4	39,2 %	217,4
21	633	1,7	179	2,7	28,3 %	157,0
26	2221	5,9	298	4,4	13,4 %	74,5
20	1024	2,7	134	2,0	13,1 %	72,6
25	791	2,1	53	0,8	6,7 %	37,2
15	1226	3,3	61	0,9	5,0 %	27,6
24	600	1,6	22	0,3	3,7 %	20,4
29	2189	5,8	63	0,9	2,9 %	16,0
28	635	1,7	12	0,2	1,9 %	10,5
27	834	2,2	2	0,0	0,2 %	1,3
12	2044	5,5	3	0,0	0,1 %	0,8
4	14 985	40,0	0	0,0	0,0 %	0,0
30	564	1,5	0	0,0	0,0 %	0,0

Nota. Método de crecimiento: CHAID. Variable dependiente: pobreza_2.

Se puede observar que el nodo 19 tiene el mejor desempeño y representa, aproximadamente, el 85,4 % de los 934 hogares —el 2,5 % de los 37 462 hogares observados—. Además, hay 798 hogares que son pobres; es decir, $(798 / 934) * 100 \% = 85,4 \%$. Esto significa que hay una mayor proporción de hogares pobres que en la muestra total; esto es $(6749 / 37462) * 100 \% = 18,0 \%$. El índice para este nodo indica $(85,4 \% / 18,0 \%) = 474,3 \%$, lo cual nos permite entender que esperamos tener casi cinco veces más hogares pobres que si hubiéramos utilizado la selección de casos aleatoria en la muestra completa.

En la Tabla 5 se muestra la proporción de error para nuestro modelo. Con esta estimación, se puede conocer la exactitud predictiva del árbol. Acompañando a esta medida se muestra el error típico de esta proporción de error.

Tabla 5
Riesgo del modelo

Estimación	Desviación del error
0,112	0,002

Nota. Método de crecimiento: CHAID. Variable dependiente: pobreza_2.

Se han clasificado incorrectamente a 1852 hogares pobres y a 2337 hogares no pobres; es decir, a $1852 + 2337 = 4189$ hogares. Esto indica que hay una proporción de error del $((4189 / 37462) * 100 \% = 11,2 \%)$ de los hogares observados en la muestra.

Con este error típico, podríamos estimar la proporción de error π en toda la población, al aplicar un nivel de confianza del 95 %. Es decir, obtener el intervalo de confianza del 95 % ($Z_{0,975} = 1,96$) para la proporción de error en la población mediante la siguiente expresión: $IC(\pi) = (\hat{p} \pm Z_{1-\alpha/2} \hat{\sigma}_{\hat{p}}) = (0,10808; 0,11592)$. Esto significa, que nuestra proporción de error estaría entre 10,8 % y 11,6 %, con un error de estimación del 0,392 %.

En la Tabla 6 se muestra una matriz que indica el número de clasificaciones correctas e incorrectas que se ha generado con este modelo. Para cada nodo se registra la cantidad de hogares pobres y hogares no pobres que fueron correctamente clasificados y cuantos no se asignaron correctamente.

Tabla 6
Clasificación del modelo

Observado	Pronosticado		
	1	2	Porcentaje correcto
1	4897	1852	72,6 %
2	2337	28 376	92,4 %
Porcentaje global	19,3 %	80,7 %	88,8 %

Nota. Método de crecimiento: CHAID. Variable dependiente: pobreza_2.

Se puede observar que, de los 6749 hogares pobres, el modelo clasificó correctamente a 4897; es decir, la proporción de acierto fue del $((4897 / 6749) * 100 \% = 72,6 \%)$. En tanto que, de los 30 713 hogares no pobres, el modelo clasificó correctamente a 28 376, cuya proporción de acierto es del $((28376 / 30713) * 100 \% = 92,4 \%)$. A nivel global, la proporción de acierto del modelo es del $((4897 + 28376) / (6749 + 30713)) * 100 \% = 88,8 \%$. Lo que significa que el modelo es muy bueno, porque tiene casi 90 % de acierto.

7. CONCLUSIONES

De acuerdo con la Figura 1, el nodo 19 es el que tiene mejor desempeño, ya que hay una mayor proporción de hogares pobres que en la muestra total, el cual está representado por el 85,4 %. El índice para este nodo indica $(85,4 \% / 18,0 \%) = 474,3 \%$, lo cual nos permite entender que esperamos tener casi cinco veces más hogares pobres que si hubiéramos utilizado la selección de casos aleatoria en la muestra completa. Por lo tanto, se considera pobres a aquellos hogares que tienen un gasto per cápita de 2047,926 soles como máximo al mes, que cuentan con más de cinco miembros del hogar y que la vivienda del hogar tenga un piso de tierra.

El modelo tiene una proporción de error que estaría entre 10,8 % y 11,6 %, con un error de estimación del 0,392 % a un nivel de confianza del 95 %. A nivel global, la proporción de acierto del modelo es del 88,8 %, lo que significa que el modelo es muy bueno, porque tiene casi 90 % de acierto.

De este modo, para la medición de la pobreza bajo el enfoque multidimensional, existen limitaciones en los datos, dada la complejidad de conseguir información sobre educación, salud, seguridad, nutrición, ingresos, gastos, vivienda, entre otros, de la misma población. En nuestro caso, la base de datos de la sumaria de la ENAHO 2018 no cubre aspectos sobre salud, seguridad y nutrición. Se recomienda que se levanten encuestas que logren cubrir todos los aspectos antes mencionados para la misma unidad de análisis.

Créditos de autoría

Carlos Teodoro Caycho Chumpitaz: Conceptualización, metodología, *software*, análisis de datos, investigación, redacción; preparación del primer borrador, redacción; revisión y edición.

REFERENCIAS

- Aguilar Estrada, A., Caamal C. I., & Portillo, V. M. (2018a). Intensidades de pobreza multidimensional en México a nivel municipal. *Revista Mexicana de Ciencias Agrícolas*, 9(1), 251-258. <https://doi.org/10.29312/remexca.v9i1.864>
- Aguilar Estrada, A., Caamal, I., & Ortiz-Rosales, M. (2018b). Pobreza multidimensional en Chiapas: generalizada pero heterogénea. *Revista LiminaR. Estudios Sociales y Humanísticos*, 16(2), 105-117. <https://doi.org/10.29043/liminar.v16i2.600>
- Alkire, S., & Foster, J. (2008). Counting and multidimensional poverty measurement. *Oxford Poverty and Human Development Initiative*.
- Alkire, S., & Santos, M. (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>

- Atkinson, A. (2003). Multidimensional deprivation: Contrasting social welfare and counting approaches. *Journal of Economic Inequality*, 1, 51-65. <https://doi.org/10.1023/A:1023903525276>
- Berlanga, V., Rubio, H. M., & Vilà, B. R. (2013). Cómo aplicar árboles de decisión en SPSS. *Revista d'Innovació i Recerca en Educació*, 6(1), 65-79. <https://doi.org/10.1344/reire2013.6.1615>
- Bourguignon, F., & Chakravarty, S. (2003). The measurement of multidimensional poverty. *The Journal of Economic Inequality*, 1, 25-49. <https://doi.org/10.1023/A:1023913831342>
- Clausen, L. J. (2019). *Posibilidades y desafíos de la medición de la pobreza multidimensional en el Perú: elementos para potenciar el debate*. Instituto Nacional de Estadística e Informática. https://www.inei.gob.pe/media/MenuRecursivo/investigaciones/pobreza_multidimensional.pdf
- Colafranceschi, M., Peyrou, M., & Sanguinetti, M. (2011). Pobreza multidimensional en Uruguay: una aplicación de técnicas multivariantes. *Quantum. Revista de Administración, Contabilidad y Economía*, 6(1), 28-55. <https://dialnet.unirioja.es/servlet/articulo?codigo=7330423>
- Comisión Económica para América Latina y el Caribe. (2014). *Panorama social de América Latina 2014*. <https://hdl.handle.net/11362/37626>
- Fondo de las Naciones Unidas para la Infancia. (2015). *Multidimensional overlapping deprivation analysis (MODA)*. <https://www.unicef-irc.org/research/multidimensional-overlapping-deprivation-analysis-moda-201-2015/>
- Hair, J. F., Anderson, R., Tatham, R., & Black, W. (2004). *Análisis multivariante* (5.ª ed.). Pearson Prentice Hall.
- Instituto Nacional de Estadística e Informática. (2016). *Evolución de la pobreza monetaria 2009-2015*. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1347/
- Instituto Nacional de Estadística e Informática. (2018a). Perú - *Encuesta nacional de hogares sobre condiciones de vida y pobreza 2018*. Sistema de Documentación Virtual de Investigaciones Estadísticas. http://webinei.inei.gob.pe/anda_inei/index.php/catalog/672/study-description
- Instituto Nacional de Estadística e Informática. (2018b). Perú: *perfil sociodemográfico. Informe nacional. Censos nacionales 2017: XII de Población, VII de Vivienda y III de Comunidades Indígenas*. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1539/

- Pérez, C., & Santín, D. (2007). *Minería de datos. Técnicas y herramientas*. Thomson.
- Peréz-Campos, P., & Rodríguez, A. (2015). *El ejercicio de medir la pobreza en el Perú: evidencia para una política de inversión en el talento*. Ministerio de Educación.
- Poza, L. C., & Fernández, C. J. (2010). Una aproximación a la construcción de un indicador de pobreza multidimensional: ¿cuáles son los focos de riesgo en España? *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 10, 43-72. <https://doi.org/10.46661/revmetodoscuanteconempresa.2164>
- Zavaleta, D., Moreno, C., & Santos, M. (2018). *Introducción al enfoque de las capacidades: aportes para el desarrollo humano en América Latina*. Manantial.
- Zuniga, C., & Abgar, N. (2011). *Breve aproximación a la técnica de árbol de decisiones*. <https://niefcz.files.wordpress.com/2011/07/breve-aproximacion-a-la-tecnica-de-arbol-de-decisiones.pdf>

