The importance of health spending in the Base of the Pyramid (BoP) population from Norte de Santander, Colombia

La importancia del gasto en salud en la población de la Base de la Pirámide (BoP) de Norte de Santander, Colombia

Julio C. Contreras-Velásquez[1]; julio.contreras@tdea.edu.co Jorge Isaac García-Navarro[2]; jigarcian@gmail.com Zaida Rocío Contreras Velasquez[3]; José Alban Londoño Arias[1]; email: jlondono5@tdea.edu.co Neida Albornoz[4]; e-mail: Neida.albornoz@unisimon.edu.co Miguel Rosario-Cohen[4,5] Miguelrosariocohen@gmail.com Maritza Torres[6]; e-mail: maritza.torres@saludzona6.gob.ec Lissé Angarita[7]; e-mail: lisse.angarita@unab.cl Henry Garcia-Pacheco[8,9]; Email: Henrygarciapacheco@gmail.com Valmore Bermúdez[10] Email: valmore.bermudez@unisimon.edu.co

[1]Administrative and Economic Sciences, Tecnológico de Antioquia Institución Universitaria, Medellín, Antioquia, Colombia.
[2]State management of the National Institute of Statistics (INE), Táchira, Venezuela. Universidad Católica, San Cristóbal, Venezuela.
[3]Universidad Francisco de Paula Santander, Cúcuta, Norte de Santander, Colombia.
[4]Facultad de Ciencias Jurídicas y Sociales, Universidad Simón Bolívar, Cúcuta, Colombia.
[5]Dirección de Creación intelectual, Doctorado en Ciencias Gerenciales, Universidad Sur del Lago. Santa Barbara del Zulia, Venezuela.
[6]Ministerio de Salud Pública del Ecuador Centro de Salud Baños
[7]Escuela de Nutrición y Dietética, Facultad de Medicina, Universidad Andres Bello, Sede Concepción 4260000, Chile
[8]Universidad del Zulia, Facultad de Medicina, Departamento de Cirugía. Hospital General del Sur "Dr. Pedro Iturbe". Maracaibo, Venezuela.
[9]Unidad de Cirugía para la Obesidad y Metabolismo (UCOM). Maracaibo, Venezuela.
[10]Universidad Simón Bolívar, Facultad de Ciencias de la Salud, Barranquilla, Colombia.
Correspondence author: julio.contreras@tdea.edu.co; jcontrerasv16@gmail.com

Received: 06/24/2022 Accepted: 10/19/2022 Published: 11/25/2022 DOI: https://doi.org/10.5281/zenodo.7607716

Abstract

Introduction. The article analyzes whether health is among the three main expenditures that a household allocates from its income in the population at the base of the pyramid (BoP) and whether or not there is an association with people belonging to different socioeconomic levels in the department of Norte de Santander, Colombia. **Methods**. The data analyzed come from a questionnaire applied to a sample of 2,394 households in the Norte de Santander region (Colombia), which were classified by socio-economic levels through the construction of a global synthetic index that, through scales, allows households to be segmented by poverty conditions, from the most intense poverty condition to the most favorable non-poverty condition respectively, resulting in 1,521 households at the Base of the Pyramid. **Results**. The findings show that for BoP households, health expenditures are presented in the highest proportion of the total sample, with a value of 15.51%, while in the middle class only 10.64% of households included health expenditures among the top three and the upper class 8.96%. Likewise, the categories of expenditure that are associated with the different socio-economic levels are housing, public services, health, and education. **Conclusions**. Although the income level of the study population is low, it can be inferred that the BoP households surveyed in the region of Norte de Santander, Colombia, health expenditures are presented in the highest proportion of the total sample, with a value of 15.51%, while in the middle class only 10.64% of households included health expenditures among the top three, and the upper class 8.96%.

Keywords: base of the pyramid, socioeconomic level, health spending, Health, Poverty.

Resumen

Introducción. El artículo analiza si la salud se encuentra entre los tres principales gastos que un hogar destina de sus ingresos en la población de la base de la pirámide (BdP) y si existe o no una asociación con personas pertenecientes a diferentes niveles socioeconómicos en el departamento de Norte de Santander, Colombia. **Métodos**. Los datos analizados provienen de un cuestionario aplicado a una muestra de 2.394 hogares de la región de Norte de Santander (Colombia), los cuales fueron clasificados p or n iveles s ocioeconómicos mediante la construcción de un índice sintético global que, a través de escalas, permitió segmentar los hogares por condiciones de pobreza, desde la condición de pobreza más intensa hasta la condición de no pobreza más favorable. **Resultados**. 1.521 hogares se ubicaron en la Base de la Pirámide. Los hallazgos muestran que, para los hogares de la BdP, los gastos en salud se presentan en la mayor proporción del total de la muestra, con un valor de 15.51%, mientras que en la clase media sólo el 10.64% de los hogares incluyó los gastos en salud entre los tres principales, y la clase alta el 8.96%. Asimismo, las categorías de gasto que se asocian a los distintos niveles socioeconómicos son vivienda, servicios públicos, salud y educación. **Conclusiones**. A pesar que el nivel de ingresos de la población de estudio es bajo, se puede inferir que los hogares de la BoP encuestados en la región de Norte de Santander, Colombia, los gastos en salud se presentan en mayor proporción del total de la muestra, con un valor de 15,51%, mientras que en la clase media sólo el 10,64% de los hogares incluyeron los gastos de salud entre los tres principales, y la clase alta el 8,96%.

Palabras clave: base de la pirámide, nivel socioeconómico, gasto sanitario, Salud, Pobreza

Introduction

Poverty is not evenly distributed around the world. In regions such as sub-Saharan Africa, South Asia and Latin America, millions of poor people struggle to survive[1]. The unfair distribution of wealth and income has become a structural problem in most economies[2,3], mainly affecting people at the bottom of the pyramid (BoP), a population largely neglected in terms of solvent and stable solutions over time.

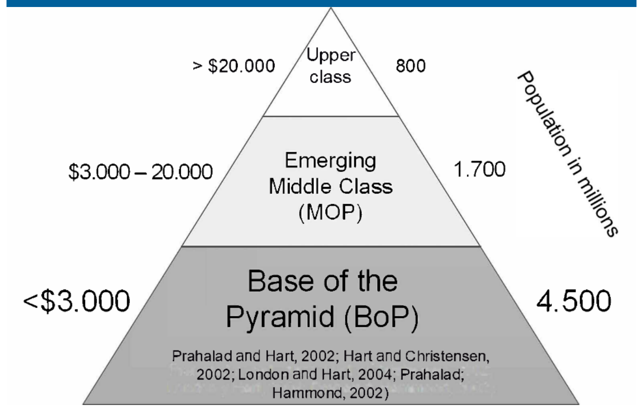
Latin America has 360 million people at the base of the pyramid and corresponds to the second largest market in terms of income, with a figure close to $510 billion dollars. In addition, it is the region with the largest population-income gap, since 70% of the 360 million people together earn only 28% of the total regional income, demonstrating the inequality in income distribution that exists in the countries of the region[4].

Starting from the need stated in the previous paragraph, the article analyzes whether health is among the three main expenditures that a household allocates from its income in the population at the base of the pyramid and whether or not there is an association with people belonging to different socioeconomic levels in the department of Norte de Santander, Colombia.

Population of the BoP

From a demographic point of view, the BoP is a population living in several, mainly developing, countries, their common denominator being their low-income levels[5]. If one classifies the world's population by its annual or daily disposable income as well as by its quantity, one obtains a pyramid-shaped distribution with very few rich people at the top and a huge base of poor people at the bottom (See Figure 1). These poor people represent the lowest level of the overall income pyramid, and are therefore called the base of the pyramid[6,7]. According to Prahalad[8] it is a population of 4.5 billion people worldwide who subsist on less than 3,000 US dollars a year.

Figure 1. World economic pyramid



Source: Own elaboration

The BoP population varies between regions and countries (9), has characteristics such as cultural complexity and difficulty of access to products and services, especially in rural areas where there is physical isolation caused by inefficient investment in infrastructure[10].

Latin America represents 9.1% of the world's BoP population, comprising 360 million people with an estimated total income of 509 billion and a total expenditure of about 632 billion USD (11). According to studies by Guesalaga & Marshall[11], the category in which the BoP spends the most income is food, which in the case of Latin America has a percentage of 41% of total expenditures, followed by 14.5% of household expenditures (See Table 1).

Table 1. Distribution of BoP expenditure by geographic area and product category

Category	Africa	Asia	Europe	Latin America	Total
Food	47.1	55.3	56.8	41	53.2
Housing	9.4	10.2	8.8	14.5	10.4
Water	1.6	0.3	0.8	0.9	0.5
Energy	5.9	6.2	7.2	5.5	6.2
Household goods	10.4	6.6	9.3	9.8	7.6
Health	4.5	2.4	3.7	4.3	2.9
Transportation	6.5	4.2	2.8	6.6	4.5
ICT	1.2	1.3	1.3	2.2	1.4
Education	2.8	2.0	1.2	1.6	1.9
Other	10.6	11.5	8.1	13.6	11.4

Source: Guesalaga & Marshall (2008)

The lack of access to food products or the poor quality of those that are available to this population, produces poor eating habits and inadequate nutrient intake that affect physical health[12], as found by Daly & Leonard[13], where in 75% of low-income households, at least one member was in poor health; similarly, feelings of powerlessness can be reflected in psychological problems[14]; For example, in a study carried out in Zimbabwe by Chikweche & Fletcher[15], the people interviewed claimed that they must use the bar of washing soap as bath soap, even though it does not adapt to the sensitivity of the skin; cooking oil produced from by-products of the waste of the slaughtering process of animals for export instead of vegetable oil; as well as consuming pieces of soy, instead of meat, sacrificing the content of proteins for an adequate nutrition.

The context of the Norte de Santander region (Colombia)

The Norte de Santander region is located in the northeastern part of Colombia; its extension is 21 648 km², which represents 1% of the national territory. It limits to the west with the departments of Cesar and Santander, with whom it also borders to the south, along with Boyacá; to the east and north its limit is the border with Venezuela[16].

This Department is made up of 40 municipalities organized into six sub-regions: metropolitan, made up of six municipalities including San José de Cúcuta (capital city), Los Patios, Villa del Rosario, El Zulia, San Cayetano and Puerto Santander. The Western sub- region, also known as the Province of Ocaña, made up of ten municipalities: Ábrego, Cáchira, Convención, El Carmen, La Esperanza, Hacarí, La Playa de Belén, Ocaña, San Calixto and Teorama; the Central sub-region, made up of seven municipalities: Arboledas, Cucutilla, Arboledas, Cucutilla, Cáchira, El Carmen, La Esperanza, Hacarí, La Playa de Belén, Ocaña, San Calixto and Teorama: Arboledas, Cucutilla, Gramalote, Lourdes, Salazar de Las Palmas, Santiago and Villa Caro; the Southwest subregion made up of six municipalities: Cácota, Chitagá, Mutiscua, Pamplona, Pamplonita and Silos; Subregion North is the one with the smallest number of municipalities, with four: Bucarasica, El Tarra, Sardinata, Tibú. And the Southeast Subregion is made up of six municipalities: Bochalema, Chinácota, Duranía, Herrán, Labateca, Ragonvalia and Toledo.

According to the projection of the National Administrative Department of Statistics of Colombia (DANE) to 2015, the study region has a population of 1,355,787 inhabitants, representing 2.8% of the total national population[17]. The population is distributed in 78.5% in urban areas and 21.5% in rural areas, 47.9% live in the city of Cúcuta and of these, 96.6% occupy the urban area of the city, in the other 39 municipalities of the Department, the figures show that only in seven municipalities, more than 50% of the population lives in municipal capitals. In the other 32 municipalities, the population is mostly rural[16].

Figure 2 shows the location of the Norte de Santander region in Colombia, as well as the geographical distribution of its sub-regions.

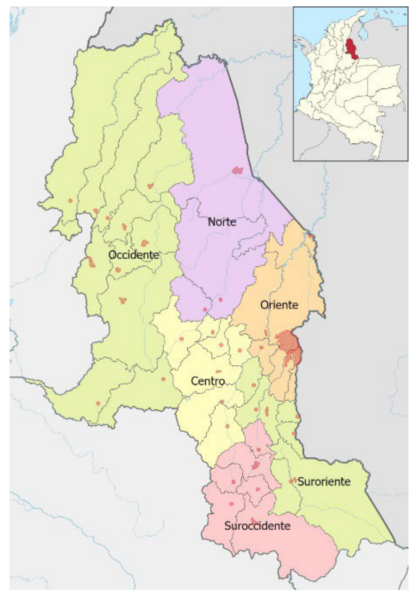


Figure 2. Map of the Norte de Santander region (Colombia)

Population and sample

The statistical unit of analysis of the data collected is the household, distributed in the geographic zones of the department of Norte de Santander. To determine the size of the sample, the population distribution of the six geographic zones of the department under study was considered. The type of sampling used is probabilistic stratified multistage with systematic random selection in the primary sampling units (segments); followed by secondary units (blocks) and, finally, the final observation unit (households).

The strata are represented by the geographic zones defined by each municipality, which is weighted according to its population size, which in turn contain the clusters of the first selection (segments) within these, the clusters of the second selection (blocks) that contain the observation units (households).

For the calculation of the sample size, it is estimated through the sampling formula of proportions in finite populations:

$$(Z\_(\alpha/2)^2\ PQN)/(\varepsilon^2\ (N-1)+Z^2\ PQ)$$

Where, N represents the number of households 338,580 households, with a confidence interval of 95%, a maximum admissible error of 2% and considering the maximum dispersion of the key variable P and Q, both with a value of 0.5. The size of the resulting initial sample n = 2,384 households.

In order to ensure the efficiency, sufficiency and precision of the sample, an oversampling system is adopted that takes into account an adjustment by design, which guarantees the execution of the sample with the sizes estimated as minimum within the levels of confidence and maximum permissible error.

The system consists of applying a non-coverage factor (t) to the sample sizes (n) estimated for each of the zones, with which the final operational selection size was determined to be 2,394 households. Table 2 presents the sample size for each geographic zone of the department of Norte de Santander, obtained after adjusting the sample design.

Table 2. Sample distributed by geographical areas

Geographical area	Housing 2015	Sample	Sample adjusted by design
Center	10,041		69
North	16,525		
West	56,936	401	404
East	219,184	1,544	1,550
Southwest	21,138		
Southeast	15,025	104	104
Norte de Santander	338,850	2,384	2,394

Source: Own elaboration

The instrument used is a structured questionnaire that has a reliability measured by Cronbach's Alpha of 0.76 (See Table 3).

Table 3. Item statistics of the Questionnaire Instrument of 2,394 households

Indicator	Media	Standard deviation	Cronbach's Alpha
Area	1.85	0.36	0.75
Gender	1.57	0.50	0.77
Age Groups	2.24	1.02	0.78
Educational Level	3.79	1.22	0.76
Can read and write	1.01	0.12	0.77
Working Arrangement	1.76	0.43	0.77
Family	2.50	0.98	0.78
Opinion Leader	1.80	1.07	0.78
Floor material	2.69	0.52	0.75
Wall material	3.78	0.50	0.75
Roof material	3.84	1.07	0.74
Bedrooms	2.94	0.91	0.74
Number of beds	2.72	0.69	0.75
Overcrowding	1.99	0.07	0.76
Number of bathrooms	2.54	0.67	0.75
Number of bathrooms with shower	2.44	0.65	0.74
Drinking water service	3.48	0.60	0.76
Excreta disposal	4.79	0.60	0.75
Urban Cleaning Service	4.69	0.83	0.74
Housing Tenure	5.49	0.85	0.76
Kitchen Tenure	1.97	0.17	0.76
Cooking fuel	2.93	0.35	0.76
Electricity Service	2.00	0.04	0.76
Fridge Holding	1.98	0.15	0.76
Ownership of a landline telephone at home	1.32	0.47	0.75
Cell phone number	2.46	0.66	0.77
TV ownership	1.99	0.11	0.76
Number of TVs	2.53	0.53	0.75
Cable TV Ownership	1.72	0.45	0.75
Vehicle ownership	1.18	0.38	0.76
Number of vehicles	1.19	0.42	0.76
Motorcycle ownership	2.00	0.00	0.76
Number of motorcycles	1.12	0.42	0.76
Internet ownership	1.49	0.50	0.75

Source: Own elaboration based on SPSS

BoP segmentation method

People of the same social class are roughly equal in terms of their income and social status, work in roughly equal occupations, and tend to have similar tastes in some products and services. They also tend to socialize with each other and share many ideas and values[18].

The unidimensional method of classifying households by socioeconomic levels through income has been the most widely used; however, it presents difficulties that limit and bias the analysis, due to measurement errors such as underdeclaration and non-declaration of income information by source[19;20,21,22,23]; the difficulty in quantifying income by period of time when people have an informal job[24,25]; at the same time, if we want to make comparisons between different developing countries, using the economic parity technique, the volatility in macroeconomic indicators and exchange rate policies, as in the case of Venezuela, make the analysis difficult[26].

This study uses a methodology that consists of constructing a global synthetic index that, through scales, allows households to be segmented by poverty conditions, from the most intense poverty condition to the most favorable non-poverty condition respectively. To arrive at the index, an Optimal Quantitative Valuation of a set of variables associated with the living conditions of the households in the Norte de Santander sample is carried out[27].

Given that we have ordinal categorical variables, we assign values to each of the categories of the variables and then observe which households have similar qualities with respect to a set of variables. The technique used is that of optimal scaling by alternating least squares, using the technique of principal components analysis for categorical variables, Categorical Principal Components Analysis (CATPCA); which allows obtaining the quantification results of score coefficients of tenure or deprivation assigned to each one of the households[28].

To quantify the categorical variables, the codes of the categories are replaced by optimal numerical values, to be able to determine the existing relations between them[29]. The process of quantitative valuation of the categories is carried out in pairs, using the optimal scaling method according to alternating least squares, where it is iterated in two stages, in the first one the model is estimated and in the second stage, the optimal scaling is carried out. These two stages are alternated iteratively until a certain convergence is achieved. The optimal scaling level of the variables is ordinal, so that the transformed values represent ordered categories[27,28].

After obtaining the score coefficients, the second multivariate technique, k-means cluster analysis, is applied, where households are classified into socio-economic levels by groups with similar characteristics ranging from the most intense poverty condition to the most favorable non-poverty condition respectively. The procedure is explained in detail below.

Data analysis and results

Segmentation of the households at the Base of the Pyramid
Selection of variables associated with living conditions

Based on the studies carried out by Camardiel et al.,[28]; Vyas and Kumaranayake[30], for the construction of a synthetic poverty index, the variables associated with the selected living conditions and the indicators of each one is presented:

(1) Structural characteristics of the household: Floor material, Wall material, Ceiling material, Number of bathrooms, Number of bathrooms with shower, Sleeping rooms.

(2) Access to household services: Electricity service, Cooking fuel, Urban sanitation service, Excreta disposal (sewage), Drinking water service, Fixed telephone at home.

(3) Household equipment ownership: Refrigerator ownership, Kitchen ownership, Number of TV sets, Cable TV ownership, Internet ownership.

(4) Assets: Home Ownership, Number of cell phones, Number of Vehicles, Number of motorcycles.

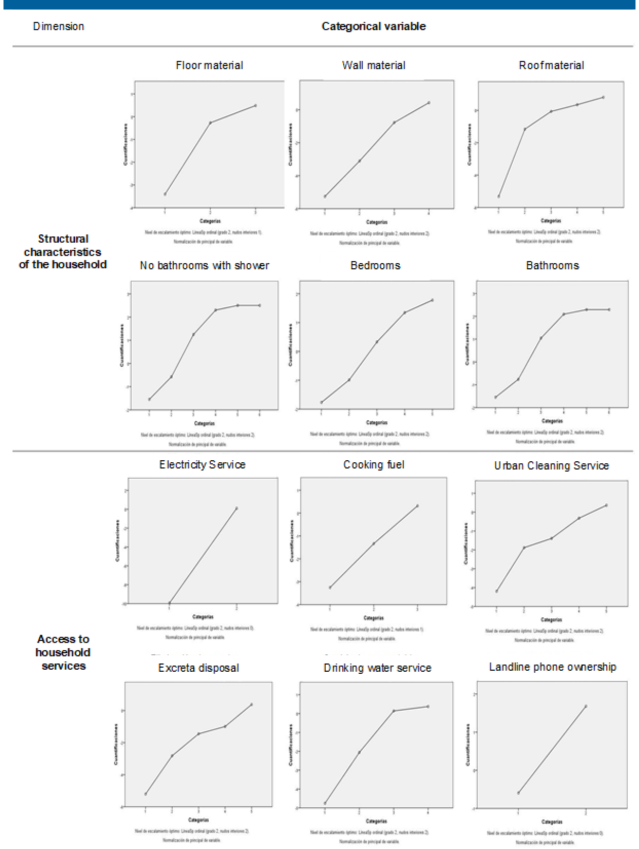
(5) Socioeconomic Characteristics: Level of Occupation (Employed or Unemployed), Educational Level of head of household.

Transformation of Original Variables by Optimal Scaling

The procedure is based on the assignment of numerical values to the categories of each variable, in order to subsequently analyze the scale of measurement of the variables with the optimal scaling technique. The final value of the quantified variables depends on the scaling level. In order to preserve the ordinal value of the original variables, the transformations of each variable are represented as category points on a vector passing through the origin.

The relationship between the quantifications and the original categories resulting from the selected optimal scaling is shown by the ordinal scaling transformation plots (See Figure 3), where the horizontal axis is the original category code, and the vertical axis represents the optimal quantifications, the greater the slope between categories, the greater the importance.

Figure 3. Ordinal Categorical Variables Transformation Plots



Source: Own elaboration based on SPSS results.

Quantification of qualitative variables and assignment of scoring coefficients to households

Given the qualitative nature of the variables under study, measured on a nominal and ordinal scale, conducive to the construction of a synthetic index capable of measuring and segmenting the socioeconomic level of the households, the method of the optimal scaling level algorithm by alternating least squares is applied, through the technique of principal components analysis for categorical variables, Categorical Principal Components Analysis (CATPCA), which allows obtaining in numerical value the score coefficients of tenure or lack assigned to each one of the households[31,32].

To quantify the categorical variables, the codes of the categories are replaced by optimal numerical values to determine the existing relations between them (29). The process of quantitative valuation of the categories is carried out in pairs, using the optimal scaling method according to alternating least squares, where it is iterated in two stages. In the first stage, the model is estimated and in the second stage, the optimal scaling is performed. These two stages are alternated iteratively until a certain convergence is achieved. The optimal scaling level of the variables is ordinal, so that the transformed values represent ordered categories[27,28].

The number of principal components to extract is defined by the researcher and a common rule is to select those components whose associated eigenvalue is greater than one. However, for the construction of socioeconomic levels, it is assumed that the first principal component captures the greatest variability.

The eigenvalue of each principal component indicates the percentage of variability in the total data that is explained. In other studies, the first component explains between 12% and 27%[33], between 13% and 16% (30) and 26% (34). Although these percentages are not high, they indicate the complexity of the correlations between the variables.

The estimation of the model obtained is presented in Table 4, where the total percentage of variance explained is 42.29%, that is, it is the total information retained or explained by the two dimensions of the model, where the variables of dimension 1 contribute 32.537% and dimension 2 retains 9.761%. The Cronbach's alpha coefficient of both dimensions is positive, which allows us to observe accumulation of information above the average of the model in other studies.

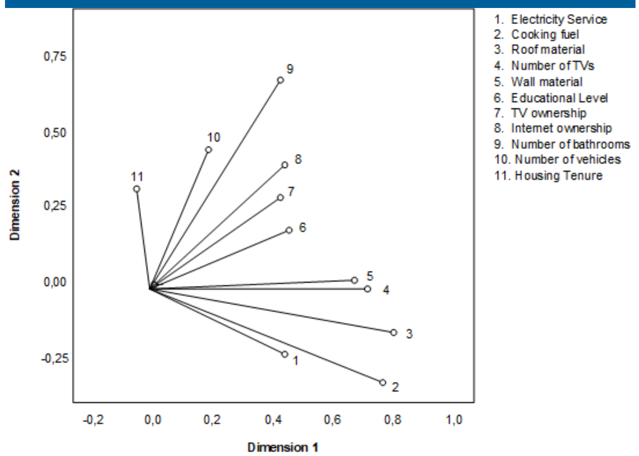
Table 4. Model estimation of the optimal scaling method

		Posted variance	
Dimension	Cronbach's Alpha	Total (eigenvalue)	% variance
1	0.906	7.483	32.537
	0.580	2.245	9.761
Total	.938[a]	9.728	42.297

a. The total Cronbach's alpha is used in the total eigenvalue. Source: Own elaboration based on SPSS results.

Once the consistency of the optimal scaling model has been verified, we proceed to determine the two dimensions or principal components, which are new variables, in this case two, that allow us to reduce the dimensionality of all the variables. By analyzing Figure 4, it can be inferred that dimension 2 represents asset holding variables, while dimension 1 represents variables of structural conditions and access to household services.

Figure 4. Saturation graph of variables by dimension



Source: Own elaboration based on SPSS results.

Figure 5 shows the quantification of the categories of the variables on a map. The interpretation of the plane with the vectors of each category is carried out based on the following guidelines: 1. The close categories imply the same combination of responses; 2. The distant categories imply different patterns of responses; 3.

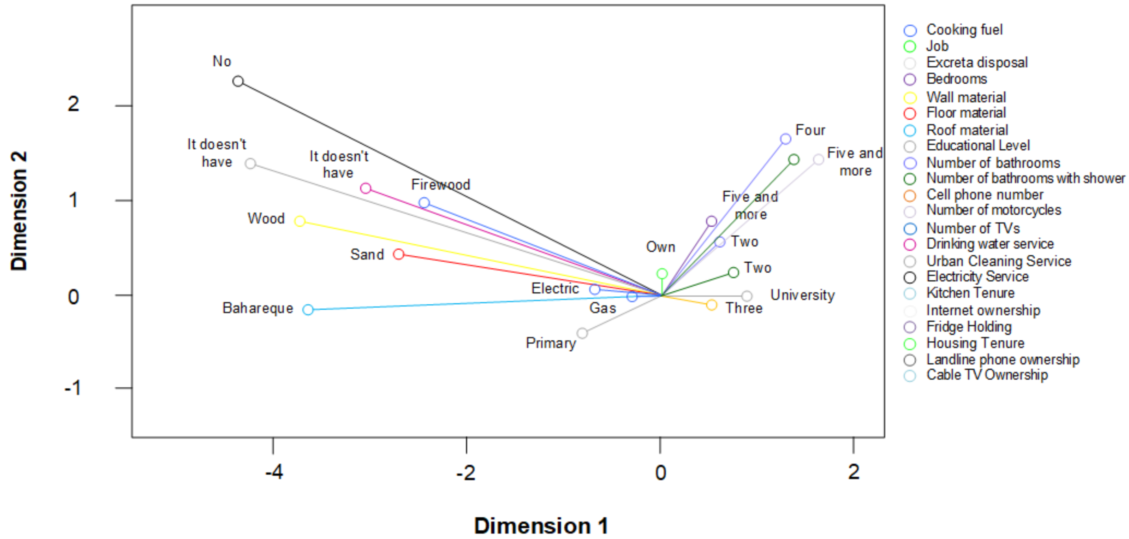
The graph of category points represented by the coordinates of the vectors in the two- dimensional plane allows for the analysis of possible associations or patterns of relationships. Figure 5 shows on the upper left side, clearly differentiated relationship patterns between the absence of services such as electricity, urban sanitation, excreta disposal, while the other upper right side of dimension 1 shows patterns of association between the possession of goods.

Construction and standardization of the Socio-Economic Level Index (INSE)

The construction of a synthetic index that represents the socioeconomic score of households is based on the optimal quantifications of the evaluated variables resulting from the categorical principal components analysis. The index can be defined as a dependent variable, from the linear combination of the vectors that define each of the categories within each variable.

Let $\hat{Y}_{jk}$ be the optimal categorical quantifications of the J variables which is equivalent to 23 variables, with j = 1, ..., m and of the k-th category within each variable. For example, $\hat{Y}_{23}$ is the quantification of variable 2 corresponding to Floor material of the third category which is Cerm/gran/parq.

Figure 5. Category Point Charts



Source: Own elaboration based on SPSS results.

In this way we obtain a vector Ŷ which is formed by the quantifications $\hat{Y}_{1k}$, ... $\hat{Y}_{jk}$, ... $\hat{Y}_{mk}$, represented in Equation 1.

$$\hat{Y} = (\hat{Y}\_1k, \ldots \hat{Y}\_jk, \ldots \hat{Y}\_mk \ ) \qquad \textbf{Equation 1}$$

For the case of the first variable $\hat{Y}_{1k}$ which has s1 categories, i.e., the vector of quantifications of the first variable is shown in Equation 2.

$$\hat{Y}\_1k = (\hat{Y}\_11, \ldots \hat{Y}\_1k, \ldots \hat{Y}\_1s1 \ ) \qquad \textbf{Equation 2}$$

Within each variable a minimum of the optimal quantifications is defined to order the weights (See Equation 3).

$$[\min (\hat{Y}\_jk) + \delta\_i =] \quad [[\mu]\_j] + \delta\_i = [\mu]\_(j+1) \ with \ j \in J$$

$$\min (\hat{Y}_{jk}) = \mu_j \ con \ j \in J \qquad \textbf{Equation 3}$$

$$\min (\hat{Y}_{jk}) = \mu_j \ con \ j \in J$$

The value of the next quantization to the minimum will be given by Equation 4, where □i is a scalar that measures the difference between a smaller and a larger quantization with i = 1, ..., k-1, i.e., for example □1 measures the difference that exists between the minimum and the second quantization with respect to the values given by the algorithm.

$$\min (\hat{Y}_{jk}) + \delta_i = \mu_j + \delta_i = \mu_{j+1} \ con \ j \in J \qquad \textbf{Equation 4}$$

$$\min (\hat{Y}_{jk}) + \delta_i = \mu_j + \delta_i = \mu_{j+1} \ con \ j \in J$$

When obtaining the quantifications on the scale from zero to a maximum value within the

variable, the sum of these maximums is performed as shown in Equation 5.

$$\textstyle\sum_j \max (\hat{Y}_{jk}) = \sum_j (\mu_{j+(s_j-1)} - \mu_j) \ con \ j \in J \qquad \textbf{Equation 5}$$

$$\textstyle\sum_j \max (\hat{Y}_{jk}) = \sum_j (\mu_{j+(s_j-1)} - \mu_j) \ con \ j \in J$$

In order to guarantee the ordinality from lowest to highest of the categorical quantifications of the components of the vector Ŷjk and to facilitate the reading of the values of the global index and of the partial indexes of socioeconomic level, a typification is carried out so that the global index takes values between 0 and 100, where the lower limit represents the most intense poverty condition and the upper limit, the best non-poverty condition (28).

On knowing the highest possible value that a household can have with respect to the variables studied, a scaling factor fe is obtained, which is the result of dividing between the desired scale, in this case 100, and the highest possible value, as shown in Equation 6. The data obtained are presented in Table 5.

$$fe = 100/(\textstyle\sum\_j \max (\hat{Y}\_jk)) \qquad \textbf{Ecuación 6}$$

Table 5. Scaling Factor Calculation

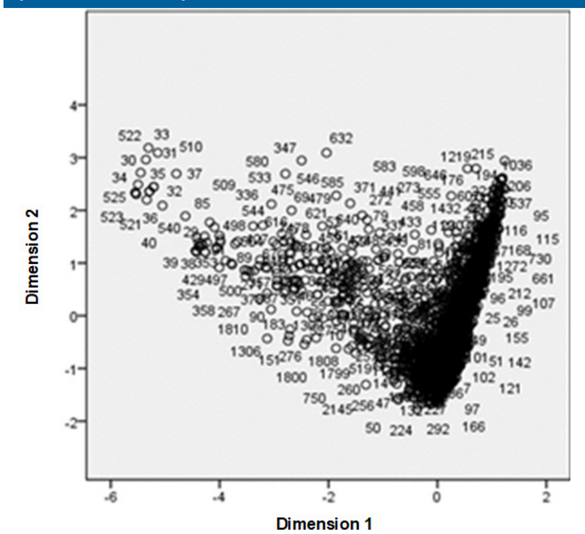
Variable	Min	Max Transf
Wall material	-5.26	5.69
Floor material	-3.41	3.90
Roof material	-5.32	6.13
Bedrooms	-1.77	3.55
Number of bathrooms	-4.76	3.84
Number of bathrooms with shower	-1.55	4.05
Drinking water service	-1.55	5.13
Excreta disposal	-5.21	5.59
Urban Cleaning Service	-4.19	3.90
Electricity Service	-9.94	10.04
Cooking fuel	-3.26	3.57
Housing Tenure	-1.42	2.12
Kitchen Tenure	-3.31	3.61
Fridge Holding	-3.54	3.82
Landline phone ownership	-0.60	2.28
Cell phone number	-2.63	3.31
Number of TVs	-3.81	4.96
Cable TV Ownership	-1.29	2.07
Vehicle ownership	-0.43	2.12
Number of motorcycles	-0.67	10.15
Internet ownership	-0.72	2.11
Educational Level	-1.85	3.78
Job	-1.41	2.12
		97.84
		1.02

Finally, the Socioeconomic Level Index (INSE) is obtained imputing to the categories the corresponding valuation by adding the scores in each household, applying Equation 6.

$$INSE = fe \sum_{i} \max \left(\hat{Y}_{jk}\right) \qquad \textbf{Equation 6}$$

After the construction of the INSE, it is used in each variable of each household to obtain a score. Figure 6 shows the location of the registration code of each household in the map, after assigning the scores.

Figure 6. Graph of location of household registration points in the map



Source: Own elaboration based on SPSS results.

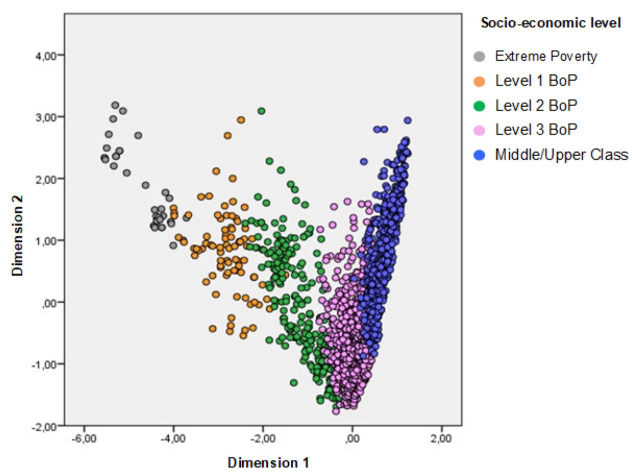
Classification of households into socio-economic groups

The classification method used is cluster analysis, which aims to conglomerate the observations into groups that are internally homogeneous and heterogeneous among them. The technique used is cluster k-means, where the grouping is done by minimizing the sum of quadratic distances between each observation and a centroid of its group or cluster. The number of groups or centroids is defined by the researcher, for this study each group represents a socioeconomic level.

The number of socio-economic levels or groups defined is four, ranging from the most intense poverty condition to the most favorable non-poverty condition respectively. A first level that represents extreme poverty, the second, third and fourth levels represent three levels of the BoP as proposed by Guesalaga and Marshall (11); Hammond et al. (35) in their studies, and the fifth level that groups the middle and upper class households; this is how the four levels are obtained: 1. Extreme Poverty; 2. BoP Level 1; 3. BoP Level 2; 4. BoP Level 3; and 5. Middle and upper class.

After applying the k-means clustering technique, with k = 5, Figure 7 is obtained, where the codes for each observation that were located on the map (see Figure 6) are grouped into each of the defined socioeconomic groups or levels.

Figure 7. Households by socio-economic level in Norte de Santander



Source: Own elaboration based on SPSS results.

The quantitative interpretation of Figure 7 is summarized in Table 6, which shows the ranges of the INSE obtained for each socioeconomic level of the households in the Norte de Santander region, after applying the k-means cluster technique. If the household has an SESI between 0 and 25.02 points it is classified as Extreme Poverty; between 25.03 and 46.30 points it is considered Level 1 BoP; between 46.31 and 62.47 points it is considered Level 2 BoP; between 62.48 and 74.29 points it is classified as Level 3 BoP; and between 74.30 and 100 points it is considered Middle and Upper Class.

Table 6. Classification of households by socioeconomic level in Norte de Santander

INSE Rank	Socioeconomic level	Frequency	Percentage	Cumulative percentage
Between 0.00 and 25.02	Extreme poverty	37	1.55%	1.55%
Between 25.03 and 46.30	Level 1 BoP	85	3.55%	5.10%
Between 46.31 and 62.47	Level 2 BoP	263	10.99%	16.08%
Between 62.48 and 74.29	Level 3 BoP	1.173	49.00%	65.08%
Between 74.30 and 100	Middle and Upper Class	836	34.92%	100.00%
	Total	2.394	100.00%	

Source: Own elaboration based on SPSS results.

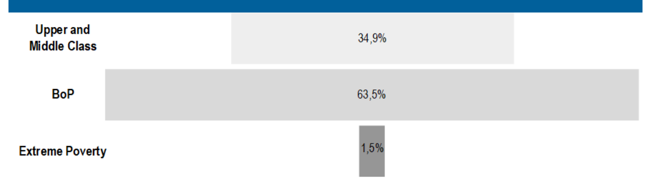
Figure 8 graphically shows the socioeconomic pyramid of households in the Norte de Santander region, where each socioeconomic level is distributed as a percentage, after applying the multivariate method of living conditions. It can be seen that the BoP has the highest proportion with 63.5%.

Income expenditure by category of services or products in BoP households in Norte de Santander, Colombia

With respect to the destination of household income expenditure, Table 7 shows which categories of products or services are among the three most important for households, depending on their socioeconomic level. It is evident that for the BoP households, health expenditures are presented in the highest proportion of the total sample, with a value of 15.51%, while in the middle class only 10.64% of the households included health expenditures among the top three, and in the upper class 8.96%.

Table 8 shows whether there is an association between the variables of expenditure category and socioeconomic status. The values of the same row and sub-table that do not share the same subscript are significantly different at $p< 0.05$ in the bilateral test of equality for column proportions, that is, it is inferred that there is an association between the two variables. Thus, the categories that have an association between the two variables are housing, public services, health, and education.

Figure 8. Socioeconomic pyramid of households in Norte de Santander



Source: Own elaboration

Table 7. Spending by category by socio-economic level

Expenditure Category	Socioeconomic Levels									
	Extreme Poverty		BoP		Middle Class		Upper Class		Total	
	Absolute Frequency(n)	Relative Frequency %	Absolute Frequency(n)	Relative Frequency %	Absolute Frequency(n)	Relative Frequency %	Absolute Frequency(n)	Relative Frequency %	Absolute Frequency(n)	Relative Frequency %
Housing	0	0.00	637	42.41	146	29.32	69	20.60	852	35.92
Food	37	100.00	1482	98.67	492	98.80	329	98.21	2340	98.65
Public Services	3	8.11	1298	86.42	478	95.98	313	93.43	2092	88.20
Clothing	22	59.46	146	9.72	42	8.43	37	11.04	247	10.41
Health	2	5.41	233	15.51	53	10.64	30	8.96	318	13.41
Transportation	35	94.59	282	18.77	77	15.46	46	13.73	440	18.55
Education	4	10.81	325	21.64	177	35.54	139	41.49	645	27.19

Table 8. Association between Expenditure Category and Socioeconomic Status

Category	Socioeconomic Level			
	Bop	Middle Class	Upper Class	Total
	%	%	%	%
Housing	42.41_a	29.32_b	20.60_c	36.49
Food	98.67_a	98.80_a	98.21_a	98.63
Public Services	86.42_a	95.98_b	93.43_b	89.46
Clothing	9.72_a	8.43_a	11.04_a	9.64
Health	15.51_a	10.64_b	8.96_b	13.53
Transportation	18.77_a	15.46_a	13.73_a	17.34
Education	21.64_a	35.54_b	41.49_b	27.45

Discussion

This study discussed whether the health expenditure of the population at the base of the pyramid is among the main destinations of their income and whether there is an association with people belonging to different socioeconomic levels in the department of Norte de Santander, Colombia. The results provide relevant information about a population that represents the majority in developing countries. The first finding confirms the above. 63.5% of the households belong to the BoP after the application of the multivariate segmentation technique. In addition, the use of this technique facilitates the standardization to classify the target population of the study, allowing to mitigate the difficulties of the traditional univariate method by income below 2 USD proposed by Prahalad, thus facilitating the comparative analysis between different developing countries.

Although the income level of the study population is low, it can be inferred that the BoP households surveyed in the region of Norte de Santander, Colombia, health expenditures are presented in the highest proportion of the total sample, with a value of 15.51%, while in the middle class only 10.64% of households included health expenditures among the top three, and the upper class 8.96%.

It is hoped that this study will motivate the academic community to continue researching the population that most represents the countries of the region, so that the results can serve as a basis for building public policies that are adapted to the reality of the context and are not copied from models of developed countries that differ greatly from the complex problems of the region.

Acknowledgements
Article generated by the research Project: Comparative analysis of the social, politics, economic and psychological elements of smuggling in the North of Santander and Táchira state, Universidad Simón Bolívar - Cúcuta campus. Research group of "Altos Estudios de Frontera" -ALEF-. Started on August 01st 2016 and finished December 31st 2016. Funded by the Universidad Simón Bolívar - Cúcuta campus, Colombia.

References

1. Yunus M. Creating a world without poverty: Social business and the future of capitalism. PublicAffairs; 2007.
2. Counts A. Small loans, big dreams: How Nobel prize winner Muhammed Yunus and microfinance are changing the world. 1st ed. Chichester, England: John Wiley & Sons; 2008.
3. Kliksberg B. América Latina: La Región más Desigual de Todas. Revista de ciencias sociales. 2009;11(3). Available at: http://dx.doi.org/10.31876/rcs.v11i3.25295
4. Pineda-Escobar MA. Negocios y sostenibilidad en la Base de la Pirámide. Punto Vista [Internet]. 2015;5(9). Available at: http://dx.doi.org/10.15765/pdv.v5i9.566
5. Praceus S. Consumer innovation at the base of the pyramid: Emerging patterns of user innovation in a resource-scarce setting. Springer Gabler; 2014.
6. Prahalad CK, Hart SL. The fortune at the bottom of the pyramid. Rev Eletrônica Estratég Neg [Internet]. 2010;1(2):1. Available at: http://dx.doi.org/10.19177/reen.v1e220081-23
7. Prahalad CK. Fortune at the bottom of the pyramid, the: Eradicating poverty through profits: Revised and updated 5th anniversary edition. Wharton School Publishing; 2010.
8. Prahalad CK. La Oportunidad de Negocios en la Base de la Pirámide: Un Modelo de Negocio Rentable, Que Sirve A las Comunidades Mas Pobres. Grupo Editorial Norma; 2005.
9. Thakur R. Community marketing: serving the base of the economic pyramid sustainably. J Bus Strategy [Internet]. 2015;36(4):40–7. Available at: http://dx.doi.org/10.1108/jbs-04-2014-0041
10. Contreras Velásquez JC, Wilches Duran SY, Delgado Rangel M, Cerda Carrasco MR. Mercado base de pirámide urbano y rural en Norte de Santander, Colombia. Revista venezolana de gerencia. 2017;21(76).
11. Guesalaga R, Marshall P. Purchasing power at the bottom of the pyramid: differences across geographic regions and income tiers. J Consum Mark [Internet]. 2008;25(7):413–8. Available at: http://dx.doi.org/10.1108/07363760810915626
12. Kempson E. Life on a low income. York, England: Joseph Rowntree Foundation; 1996.
13. Daly M, Leonard M. Against All Odds: Family Life on a Low Income in Ireland. Dublin: Combat Poverty Agency; 2002.
14. Andreasen AR. The Disadvantaged Consumer. New York, NY, Estados Unidos de América: Free Press; 1975.
15. Chikweche T, Fletcher R. Understanding factors that influence purchases in subsistence markets. J Bus Res [Internet]. 2010;63(6):643–50. Available at: http://dx.doi.org/10.1016/j.jbusres.2009.04.024
16. ANH, PNUD. Diagnóstico Socioeconómico del departamento de Norte de Santander. www.anh.gov.co. 2016.
17. Albornoz NC, Gallardo H, R. M, Numa N, Orbegoso L, Ramírez C, et al. Análisis del contrabando en el Norte de Santander, desde la percepción de los constructos de la institucionalidad. Cúcuta: Universidad Simón Bolívar; 2016.
18. Coleman RP. The continuing significance of social class to marketing. J Consum Res [Internet]. 1983;10(3):265. Available at: http://dx.doi.org/10.1086/208966
19. Camelo H. Subdeclaración de ingresos medios en las encuestas de hogares, según quintiles de hogares y fuente del ingreso. Segundo Taller Regional sobre Medición del Ingreso en las Encuestas de Hogares-LC/R. :459 73.
20. Donza E. Incidencia de la no respuesta a las preguntas de ingresos en la Encuesta Permanente de Hogares, consideraciones teóricas y efectos. Gran Buenos Aires 1990-2010. IX Jornadas de Sociología. 2011;8 12.
21. Feres JC. Falta de respuesta a las preguntas sobre el ingreso: su magnitud y efectos en las encuestas de hogares de América Latina. Segundo Taller Regional sobre Medición del Ingreso en las Encuestas de Hogares-LC/R 1886-1998, BID. 1998;399–422.
22. Llach J, Montoya S. En pos de la equidad: la pobreza y la distribución

del ingreso en el área metropolitana de Buenos Aires, diagnóstico y alternativas de políticas. Buenos Aires: IERAL; 1999.

23. Roca E, Pena H. La Declaración de Ingresos en las Encuestas de Hogares. En: 5° Congreso Nacional de Estudios del Trabajo. Buenos Aires: ASET; 2001. p. 1 24.

24. Gómez PS. Construcción de estratos socioeconómicos en Paraguay: aplicación con análisis de componentes principales en R. Astrolabio. 2014;(12):422–47.

25. Minujin A, Bang JH. Indicadores de inequidad social. Acerca del uso del" índice de bienes" para la distribución de los hogares. Desarrollo Económico. 2002;42(165):129 46.

26. Loayza NV, Rancière R, Servén L, Ventura J. Macroeconomic volatility and welfare in developing countries: An introduction. World Bank Econ Rev [Internet]. 2007;21(3):343–57. Available at: http://dx.doi.org/10.1093/wber/lhm017

27. Vélez EC, Moreno H. Metodología estadística del modelo de ponderaciones del sistema de selección de beneficiarios de programas sociales SISBEN. Bogotá: Departamento Nacional de Planeación; 1994.

28. Camardiel A, Vásquez M, Ramírez G. Una propuesta para la construcción de un Índice Sintético de Pobreza. Revista venezolana de análisis de coyuntura. 2000;121 42.

29. Van der Burg E, de Leeuw J, Verdegaal R. Homogeneity analysis with sets of variables: An alternating least squares method with optimal scaling features. Psychometrika [Internet]. 1988;53(2):177–97. Available at: http://dx.doi.org/10.1007/bf02294131

30. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan [Internet]. 2006;21(6):459–68. Available at: http://dx.doi.org/10.1093/heapol/czl029

31. Kolenikov S, Angeles G. Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? Rev Income Wealth [Internet]. 2009;55(1):128–65. Available at: http://dx.doi.org/10.1111/j.1475-4991.2008.00309.x

32. Khudri MM, Chowdhury F. Evaluation of socio-economic status of households and identifying key determinants of poverty in Bangladesh. European Journal of Social Sciences. 2013;27(3):377–87.

33. McKenzie DJ. Measuring inequality with asset indicators. J Popul Econ [Internet]. 2005;18(2):229–60. Available at: http://dx.doi.org/10.1007/s00148-005-0224-7

34. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data-or tears: An application to educational enrollments in states of India. Demography [Internet]. 2001;38(1):115. Available at: http://dx.doi.org/10.2307/3088292

35. Hammond AL, Kramer WJ, Katz RS, Tran JT, Walker C. The next 4 billion: Market size and business strategy at the base of the pyramid. Washington D.C., DC, United States: World Resources Institute; 2007.