



Desarrollo y Sociedad

ISSN: 0120-3584

ISSN: 1900-7760

revistadesarrolloysociedad@uniandes.edu.co

Universidad de Los Andes

Colombia

Morales, Katherine; Flores, Miguel; Salazar Méndez, Yasmín  
Analysis of Principal Nonlinear Components for the  
Construction of a Socioeconomic Stratification Index in Ecuador  
Desarrollo y Sociedad, núm. 88, 2021, Julio-, pp. 43-82  
Universidad de Los Andes  
Bogotá, Colombia

DOI: <https://doi.org/10.13043/DYS.88.2>

Disponible en: <https://www.redalyc.org/articulo.oa?id=169167260003>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso  
abierto

## Analysis of Principal Nonlinear Components for the Construction of a Socioeconomic Stratification Index in Ecuador

Katherine Morales<sup>1</sup>, Miguel Flores<sup>2</sup>, Yasmin Salazar Méndez<sup>3</sup>

DOI: 10.13043/DYS.88.2

### Abstract

Socio-economic stratification classifies people or groups of people within a society. Although social stratification is a universal characteristic of human societies, the criteria considered to classify individuals is not unique and some methodological approaches are distinguished. In this article, we build an indicator of socioeconomic stratification for Ecuador through a Nonlinear Principal Components Analysis using data from the 2010 Census. This methodology allows the incorporation of numerical and categorical variables, and nonlinear relationships. The main results suggest that the households located in the urban area show better conditions and greater access to basic services. Also, education positively affects social and economic conditions of both individuals and the households. In light of these results, public policy should target education and public investment in the provision of basic services in rural areas.

*Key words:* Statistical analysis; social class; social inequality; Ecuador.

*JEL classification:* C18; D63; I31.

- 
- 1 Departamento CITI, Laboratorio Samovar de Telecom SudParis – Institut Polytechnique de Paris. París – Francia. Email: katherine.morales\_quinga@telecom-sudparis.eu.
  - 2 MODES, SIGTI, Department of Mathematics, Escuela Politécnica Nacional. Quito – Ecuador. E-mail: miguel.flores@epn.edu.ec
  - 3 Department of Quantitative Economics, Escuela Politécnica Nacional. Quito – Ecuador. E-mail: yasmin.salazar@epn.edu.ec

This paper was received on September 23, 2020, revised on February 11, 2021, and finally accepted on April 11, 2021.

## Análisis de componentes principales no lineales para la construcción de un índice de estratificación socioeconómica para el Ecuador

Katherine Morales<sup>4</sup>, Miguel Flores<sup>5</sup>, Yasmín Salazar Méndez<sup>6</sup>

DOI: 10.13043/DYS.88.2

### Resumen

La estratificación socioeconómica clasifica a las personas o grupos de personas dentro de una sociedad. Si bien la estratificación social es una característica universal de las sociedades humanas, el criterio considerado para clasificar a los individuos no es único y se distinguen algunos enfoques metodológicos. En este artículo, construimos un indicador de estratificación socioeconómica para Ecuador a través de un Análisis de Componentes Principales no Lineales utilizando datos del Censo de 2010. Esta metodología permite la incorporación de variables numéricas, categóricas y relaciones no lineales. Los principales resultados sugieren que los hogares ubicados en el área urbana presentan mejores condiciones y mayor acceso a los servicios básicos. Además, la educación afecta positivamente las condiciones sociales y económicas tanto de las personas como de los hogares. A la luz de estos resultados, la política pública debe apuntar a la educación y a la inversión pública para la provisión de servicios básicos en las zonas rurales.

*Palabras clave:* análisis estadístico; clase social; desigualdad social; Ecuador.

*Clasificación JEL:* C18; D63; I31.

---

4 Departamento CITI, Laboratorio Samovar de Telecom SudParis – Institut Polytechnique de Paris. París – Francia. E-mail: katherine.morales\_quinga@telecom-sudparis.eu.

5 MODES, SIGTI, Department of Mathematics, Escuela Politécnica Nacional. Quito – Ecuador. E-mail: miguel.flores@epn.edu.ec

6 Department of Quantitative Economics, Escuela Politécnica Nacional. Quito – Ecuador. E-mail: yasmín.salazar@epn.edu.ec

Este artículo fue recibido el 23 de septiembre del 2020, revisado el 11 de febrero del 2021 y finalmente aceptado el 11 de abril del 2021.

## 1. Introduction

The existence of socioeconomic differences between the members of a society are inherent to all modern societies in fields such as education, occupation, income, and prestige, among others (Rodríguez et al. 2020; Leo et al. 2016). Socio-economic stratification plays a fundamental role in public policy for the planning, design and implementation of specific programs to reduce social and economic inequality (Zhou and Wodtke, 2019).

Socio-economic stratification is not without controversy and disagreements (Fujihara, 2020; Tang, 2017; Haer, 1957), since proposing valid indicators implies connecting theories and quantitative or qualitative methods, which allow theories to be operationalized. According to Haug (1977, p. 51), clear theoretical proposals are the main "obstacle" in the measurement of social stratification, since these are "complex, contradictory, diverse". Thus, the criterion considered to classify individuals is not unique and some methodological approaches are distinguished. In general, three approaches are prominent: quantitative, qualitative and mixed approach (Silva, 1981). In the first approach, which is also used in this article, indicators or classifiers are built using statistical methods in order to classify individuals, households or housing based on variables that collect individual, household or housing characteristics. For example, these variables are considered to be related to the education and occupation of the head of household, housing characteristics, income level and purchasing power.

The Principal Components Analysis (PCA) is commonly considered a suitable method for the construction of indicators as it reduces a large number of variables to a smaller number of uncorrelated linear combinations of these variables. These are called principal components and they represent the data as closely as possible (Mori et al., 2016). However, despite the benefits of the PCA and its broad-ranging use, this methodology presents a limitation when it comes to building indicators as the variables must be linearly related and numerical (Linting et al., 2007).

Given the qualitative and quantitative nature of the variables normally used for social stratification, the PCA assumptions are not met. Linting et al. (2007) propose an alternative named Nonlinear Principal Components Analysis (nonlinear PCA) which allows the incorporation of numerical and categorical variables

and nonlinear relationships. Variables are quantified by the optimal scaling process, so that variance is optimized. Optimal scaling is based on the alternating least squares procedure, an iterative process that uses the previous quantifications to estimate the following quantifications, until converging to the solution (Krijnen, 2006; Kuroda et al., 2013).

In non-linear PCA, it is important to evaluate the stability of the results obtained. However, defining stability is a controversial issue already under discussion. Ferrari and Manzi (2010) argue that stability is an unfinished issue and that until now no conclusive and definitive results have been presented. For example in Gi (1990), the stability of an analysis method is defined as the analysis' degree of sensitivity to variations in the data or parameters of the model. According to the same author, a solution is stable if a small change that is unimportant in the data, the model or technique leads to a small and unimportant change in the results. In this article, stability is the degree of sensitivity of the results against changes in the data. Small changes in the data should lead to changes in the level of the results of the analysis. Efron and Tibshirani (1993); Linting and Van der Kooij (2007) present an option for the verification of solution stability by means of non-parametric bootstrap.

In this context, this article builds an indicator for socio-economic stratification for Ecuador at provincial level<sup>7</sup>. In this country, socioeconomic stratification plays a fundamental role in public policy, for the planning, design and implementation of specific programs, to exemplify, the identification of the beneficiaries of conditional income transfer programs. It is also a tool used in demography, sociology, political science and marketing, among other areas of knowledge.

The contribution of this article, alongside the socio-economic stratification indicator, is that it is obtained through the non-linear PCA method. To the best of our knowledge, there are no studies using this methodology and our aim is mainly empirical.

The data used correspond to the Population and Housing Census (2010). The selected variables are related to characteristics of the head of household, the pos-

---

7 In Ecuador, a province is a political-administrative division made up of two or more cantons. A canton is the second level administrative division. The country has 24 provinces and 221 cantons.

session of household goods and the characteristics of the housing. For the stability analysis of the results obtained from the non-linear PCA, the bootstrap method mentioned in the previous paragraph is applied.

The article is organized as follows. After this introductory part, Section 2 contains a brief literature review. Section 3 describes the database and methodology used. In Section 4 the results are presented and discussed. Finally, Section 5 shows the conclusions of the work.

## 2. Brief Literature Review

Social stratification consists in classifying people or groups of people in a society (Kerbo, 2017). Although, academically, European social scientists such as Max Weber, Karl Marx, Wilhelm Dilthey, Émile Durkheim are recognized as the pioneers in the study of stratification and social class (Lemos, 2012), the idea of classifying individuals dates back to advanced ancient societies. For example, Kamakura and Mazzon (2013) mention the social classes of Athens (citizens, , and slaves), Sparta (spartiates, perioecis and helots) and Ancient Rome (patricians, noble plebeians and plebeian gentes) to argue that the existence of well-established social classes with clearly defined characteristics to differentiate them extends to antiquity. The same authors state that this classificatory notion is still used today but with different terminology.

Although the construction of groups implicitly implies the legitimacy of the differences in the distribution —of wealth, goods or income—, ignoring individual heterogeneity would create a false illusion of equality and temporal stability of socio-economic conditions (Kerbo, 2017). It is an undeniable fact that the unequal distribution of goods favors a group that concentrates not only on wealth, but also on power, and leaves those who are not as important a part in this division at a disadvantage.

Even though from a normative point of view the acceptable level of inequality, as well as the deliberate delimitation of social and economic groups, are debatable, these facts are usually associated with social and political conflicts, which can even put democratic systems at risk (Acemoglu et al., 2015). In some way, the study of socio-economic stratification ratifies the existence of these differences (see Fujihara, 2020; Wu, 2019; Wright, 2005; Featherman

and Hauser, 1976), however, the transparency of the imbalances in the distribution of goods and wealth provides information that allows reflection on the desirable structure of a society in terms of social and economic inequality. It also, depending on its members' level of aversion to inequity, offers instruments to reduce inequality, to a greater or lesser extent.

### 3. Data and Methodology

#### 3.1 Data

##### 3.1.1. Description of the database

The data used in this article correspond to the Population and Housing Census (2010) conducted by the INEC, which provides the following general information: Ecuador, at the census date, had 14,483,499 inhabitants. At household level there were 3,815,527 households and 3,810,548 heads of household. The difference between household and house is that the word household is used to designate a person or a group of people who live under the same roof and share the food expenses, while the house is the space where people live, which is separated by walls or other elements covered by a roof (INEC, 2012).

The information used in this paper relates to 3,802,566 households, in conjunction with information on housing and the heads of households. Variables at individual, household and housing level had to be used to make sure that the segmentation conducted collects as much information about Ecuadorian households as possible. In terms of provinces, Table 1 shows that the province with the most households in Ecuador is Guayas, with 958,965 households (25.22%), followed by Pichincha with 727,838 households (19.14%).

For the selection of variables, the initial criterion consisted of considering those variables<sup>8</sup> that present the least amount of missing data, between 0% and 10%. However, in these cases, a simple imputation process was carried out;

---

8 The variables of the initial stage will not necessarily be part of the definitive model. In the following sections, the variables selected for the execution of the algorithm and creation of the indicator are presented.

**Table 1.** Number of households and participation of Ecuador's provinces in the 2010 census

Province	Households	Participation
Galápagos	7236	0.19%
Pastaza	19818	0.52%
Zamora Chinchipe	21371	0.56%
Napo	22462	0.59%
Orellana	31495	0.83%
Morona Santiago	33352	0.88%
Sucumbíos	43056	1.13%
Carchi	44136	1.16%
Bolívar	47723	1.25%
Cañar	58627	1.54%
Santa Elena	76194	2.00%
Santo Domingo de los Tsáchilas	95221	2.50%
Imbabura	103009	2.71%
Cotopaxi	103137	2.71%
Loja	116892	3.07%
Chimborazo	125407	3.30%
Esmeraldas	129539	3.41%
Tungurahua	140536	3.70%
El Oro	163290	4.29%
Azuay	188331	4.95%
Los Ríos	201933	5.31%
Manabí	343088	9.02%
Pichincha	727838	19.14%
Guayas	958965	25.22%

Source: Authors based on Data from the Population and Housing Census (2010)

that is, the missing values of a variable were replaced by the average value of the units observed in the variable.

3.1.2. Categorization of variables

The nonlinear PCA method makes it possible to analyze a set of data with qualitative and quantitative variables; however, as the objective of this article is to obtain an indicator of households' social stratification in the provinces, the variables are



categorized in an ordinal way, which refers to changing the original scale of the variables in the Table 2, by turning them into ordinal variables. The categories are classified under the criteria of possession of goods and access to services that each category represents; i.e., the first category "1" is assigned to the lowest value categories under the criteria mentioned above, while the last category is assigned to the ones that present the best characteristics.

Table 2. Selected preliminary variables

Variable	Abbreviation	Scale	Categories
Population			
Education of the head of household	P23	Nominal	11
Type of educational center	P22	Nominal	4
Private insurance	P07	Nominal	2
Can read and write	P19	Nominal	2
Households			
Type of hygienic service of household	H03	Nominal	3
Shower	H04	Nominal	3
Landline telephone	H07	Nominal	2
Cell phone	H08	Nominal	2
Internet	H09	Nominal	2
Computer	H10	Nominal	2
Cable television	H11	Nominal	2
House possession	H15	Nominal	7
Number of exclusive bedrooms for sleeping	H01	Numerical	-
Number of people in the house	TP1	Numerical	-
Last payment of electricity	H12	Numerical	-
The way they drink water	H06	Nominal	5
Housing			
Area	URH	Nominal	2
Ceiling material	V01	Nominal	6
Wall material	V03	Nominal	7
Floor covering material	V05	Nominal	7
Source of water	V07	Nominal	5
The way they get water	V08	Nominal	4
WC	V09	Nominal	5

(Continued)

**Table 2.** Selected preliminary variables

Variable	Abbreviation	Scale	Categories
Garbage disposal	V13	Nominal	5
Main entrance path	VAP	Nominal	6
Electrical power service	V10	Nominal	4
Type of housing	VTV	Nominal	7
Ceiling condition	V02	Nominal	3
Condition of the walls	V04	Nominal	3
Condition of the floor covering	V06	Nominal	3
Electric meter	V11	Nominal	3
Number of light bulbs in house	V12A, V12B	Numerical	-

Source: Authors based on Data from the Population and Housing Census (2010)

For example, if we take the variable area of Table 2, we assign "1" to rural area and "2" to urban area, since the rural area generally has less access to services than the urban area. In tables 3 and 4, the preliminary variables are presented with their respective categories.

Furthermore, in some cases, it was necessary to create the variables based on the census information; for example, the case of variable overcrowding, which was constructed from the general information of the household and the head of the household. Overcrowding refers to the relationship between the number of people in a house (TP1) and the space or number of available rooms (H01) (See Table 2). Since poor people's access to resources is limited, the housing facilities they occupy tend to be less appropriate than those available for non-poor people, "1" is assigned to yes and "2" to no.

### 3.2. Methodology

The first version of the Nonlinear Principal Components Analysis (nonlinear PCA) method was described by Guttman (1941). Other contributions to this methodology were made in later years by Kruskal (1964), Shepard (1966) and Kruskal and Shepard (1974).

The nonlinear PCA is presented as a generalization of the PCA, since it makes it possible to incorporate qualitative variables, with ordered and unordered

Table 3. Preliminary Variables

Electrical power service		Condition of the walls	
1	None	1	Poor
2	Power generator	2	Good
3	Electric network company	3	Regular
Electric meter		Condition of the floor covering	
1	None	1	Poor
2	Shared	2	Good
3	Exclusive use	3	Regular
Type of housing		Can read and write	
1	Choza (Construction with adobe or straw walls, dirt floor and thatched roof)	1	No
2	Covacha (Construction with rustic materials such as branches, cardboard, asbestos, cans or plastic, with wood or dirt floors)	2	Yes
3	Rancho (Rustic construction covered with palm or straw, with cane walls and with cane or earth wood floor)	Private insurance	
4	Room in tenement house	1	No
5	Apartments in a house or a building	2	Yes
6	House-Villa	Overcrowding	
7	Other atypical housing	1	No
Ceiling condition		2	Yes
1	Poor	Last payment of electricity	
2	Good	1	Lower than 5
3	Regular	2	From 5 to 9
Type of educational centre		3	From 10 to 14
1	Mainly state-run	4	From 15 to 20
2	Nobody studies	5	From 21 to 29
3	Mainly private	6	From 30 to 39
Private insurance		7	From 40 to 49
1	Nobody in household	8	From 50 to 59
2	Some people	9	From 60 to 69
3	All	10	More than 70

**Table 4.** Preliminary Variables

Education of the head of household		Floor covering material	
1	None	1	Other materials
2	Literacy center	2	Dirt
3	Pre-school	3	Cane
4	Primary	4	Brick or concrete
5	Secondary	5	Ceramics, tile, marble
6	Higher (non-university)	6	Untreated plank
7	Higher (university)	7	Boarding, parquet or floating floor
8	Postgraduate	Source of water	
Type of WC		1	Other (rain)
1	Exclusive	2	Delivery
2	Shared	3	River, slope, ditch
3	None	4	Well
Shower		5	Public network
1	Exclusive	How do they get water	
2	Shared	1	Not by pipeline
3	None	2	By pipeline outside the building
Landline telephone		3	By pipeline outside the house and inside the building
1	Yes	4	By pipeline inside the house
2	No	WC	
Cell phone		1	None
1	Yes	2	Latrine
2	No	3	Direct discharge into the sea or river
Internet		4	Connected to septic tank
1	Yes	5	Connected to cesspit
2	No	6	Connected to public network
Computer		Garbage disposal	
1	Yes	1	Other
2	No	2	Burying
Cable television		3	They burn it
1	Yes	4	They throw it into the river or stream
2	No	5	They throw it into vacant or irregular lands
Area		6	By garbage collection vehicle
1	Urban		
2	Rural		

(Continued)

Table 4. Preliminary Variables

Ceiling material		Main entrance path	
1	Other materials	1	Other
2	Straw, thatch	2	River, sea, lake road
3	Tile or shingle	3	Pathway, chaquiñan (trails in Ecuador)
4	Zinc sheet	4	Dirt road or ballast
5	Asbestos	5	Paved stone
6	Concrete	6	Cobbled, paved or concrete
Wall material		Number of light bulbs	
1	Other materials	1	5 or less
2	Uncoated cane	2	From 6 to 10
3	Coated cane	3	From 11 to 15
4	Wood	4	from 16 to 20
5	Adobe or clay	5	From 21 to 25
6	Brick	6	From 26 to 30
7	Concrete	7	From 31 to 35
		8	From 36 to 40
		9	More than 41

categories (ordinal, nominal), and to discover and treat nonlinear relationships between variables. Mori et al. (2016) present nonlinear PCA as the result of minimizing two loss functions, rank and homogeneity analysis, both of which are the subject to restrictions, in conjunction with the optimal scaling. In order to find the solution, one has to minimize several parameters of the loss function simultaneously using the alternating least squares technique (Mori et al., 2016).

Young (1981) expresses that optimal scaling is a quantification technique that assigns numerical values to the categories of the variables, under the restrictions of the analysis level of the variable (numerical, ordinal and nominal) and turns them into a vector of optimal scale. On the other hand, while minimizing the loss function, the solution must be found with several parameters simultaneously. The Alternating Least Squares (ALS) algorithm is used as a tool to solve this minimization problem. The ALS algorithm is an iterative process based on the alternation, where first the loss function for the first parameter is minimized while the remainder is kept fixed, and then the loss function for

the second parameter is minimized, keeping the remainder fixed, etc. (Krijnen, 2006; Kuroda et al., 2013).

### 3.2.1. Bootstrap Method

In the context of the nonlinear PCA, the nonparametric bootstrap procedure is employed to evaluate the stability of the nonlinear PCA solution (Linting and Van der Kooij, 2007). This procedure incorporates the random bootstrap with replacement, which is based on the original sample called principal sample and consists of obtaining B bootstrap samples from the principal sample (Efron, 1997; Efron and Tibshirani, 1993). The bootstrap samples obtained are balanced; that is, it is guaranteed that the initial observations appear exactly B times in the B samples (Linting and Van der Kooij, 2007; Davison et al., 1986).

Subsequently, the analysis (CATPCA algorithm) is performed for each of the bootstrap samples, which results in B values for the quantifications of categories of the variables. B bootstrap values for each of the nonlinear PCA results are obtained, forming a bootstrap distribution, from which a confidence interval can be calculated. These same confidence intervals are used to analyze the stability of the results (Linting and Van der Kooij, 2007; Markus, 1994; Efron and Tibshirani, 1993). In this work, the confidence intervals will be obtained for the outputs of the algorithm, i.e. for the optimal quantifications for the categories of the variables.

The quantifications of the categories of selected variables and the bootstrap samples along with the quantifications of the principal sample are recorded in order to create the confidence intervals (a confidence interval is associated with a level of confidence at (1) 100%). Linting and Van der Kooij (2007) present the construction of the confidence intervals for the quantifications obtained through percentiles.

In the context of nonlinear PCA, the quantifications with small intervals, obtained from bootstrap samples indicate how stable the optimal quantifications are; in other words, if the quantifications of the bootstraps samples are not very different, the solution is considered stable (Linting and Van der Kooij, 2007).

This study employed the CATPCA algorithm, included in the SPSS statistical program, which was introduced in version 10 of SPSS in mid-1999 (Meulman et al., 1999). The patch used in this work is IBM-SPSS 23 and the complete algorithm is followed by Mori et al. (2016); Washed (2004); Fernando (2014).

4. Results and Discussion

4.1. Results

4.1.1. Selection of variables

The CATPCA algorithm is applied for each province with the preliminary variables defined in the previous section (See Table 3 and Table 4). This is done in order to maximize the variance explained by each principal component within each province. Using this criterion, several similarities are identified when analyzing the variables that contribute most to the variance of the first component, and a pattern is found within the 24 provinces. Table 5 presents some selected variables.

Table 5. Selection of variables

Variable	Abbreviation	Level of analysis	Categories
Education of the head of household	P23	Ordinal	8
Landline telephone	H07	Ordinal	2
Computer	H10	Ordinal	2
Area	URH	Ordinal	2
Wall material	V03	Ordinal	7
Floor covering material	V05	Ordinal	7
Main entrance path	VAP	Ordinal	6
Number of light bulbs in House	FO	Ordinal	9
Condition of the walls	V04	Ordinal	3
Condition of the floor covering	V06	Ordinal	3

4.1.2. Result analysis

Based on the selection of variables described in the previous section and once the sample design was obtained for each of the analyzed provinces, the CATPCA algorithm was performed to acquire the optimal quantifications of the

categories of variables and to analyze the variance explained by the components. At this point, the bootstrap procedure is applied with  $B=1000$ , followed by the presentation of a stability analysis of the results.

### *Variance analysis*

The results obtained include the variance that is recorded for the first (% Variance C1) and second (% Variance C2) components, along with the total variance (% Total) (See Table 6). It was also concluded that the variance explained by the two principal components ranges between 45% and 59%. The province whose first component explains the highest percentage of the data is Loja, with 47.73%, and the province that has the lowest variance is Galápagos, with 27.99%. For Pichincha, the first component has a variance of 35.73% and for Guayas, a value of 39.63%. The variance calculated for the second component is observed to be greater than 10% in all provinces.

To determine the reliability in the internal consistency of components, the Cronbach alpha coefficient<sup>9</sup> of the first ( $ACC_1$ ) and second component ( $ACC_2$ ) is used (See table 6).

The value of the Cronbach alpha coefficient in all provinces is greater than 0.7; thus, it can be concluded that the first component is reliable, depending on how well the information is represented. The Cronbach alpha coefficient of the eigenvalue associated with the total eigenvalue presents values greater than 0.8; that is, when the two components are used, the total reliability increases. It is also important to mention that the total variance explained in the first results of the execution of the CATPCA algorithm with the preliminary variables (See Tables 4 and 3), is lower in all provinces with respect to the results obtained in this section with the final variables. This is due to the selection of the final variables under the criterion of greatest contribution.

Table 7 presents the analysis of the explained variance for selected variables in the first component for each province. For example, the education variable for

---

9 The Cronbach alpha coefficient is a value that allows to measure internal consistency, in other words, it indicates how well the information of multiple variables is represented in a single indicator. The coefficient takes values between 0 and 1. The value of the coefficient will be greater the greater the correlation between the variable is. The closer the index gets to end 1, the better the reliability of proposed variables selection, considering an acceptable reliability from 0,70 (Cortina, 1993)



Table 6. Cronbach's Alpha-Variance

Province	% Variance C <sub>1</sub>	% Variance C <sub>2</sub>	% Total	ACC <sub>1</sub>	ACC <sub>2</sub>
Azuay	43.21	10.84	54.05	0.85	0.09
Bolívar	41.85	12.23	54.08	0.85	0.20
Cañar	37.90	13.73	51.63	0.82	0.30
Carchi	38.45	12.87	51.32	0.82	0.24
Cotopaxi	41.92	12.33	54.25	0.85	0.21
Chimborazo	46.31	13.07	59.37	0.87	0.26
El Oro	32.98	12.42	45.40	0.77	0.22
Esmeraldas	37.46	12.60	50.06	0.81	0.23
Guayas	39.63	12.71	52.34	0.83	0.24
Imbabura	42.12	11.99	54.11	0.85	0.19
Loja	47.73	10.90	58.62	0.88	0.09
Los Ríos	34.90	13.95	48.86	0.79	0.32
Manabí	39.85	12.88	52.73	0.83	0.25
Morona Santiago	38.07	14.73	52.80	0.82	0.36
Napo	39.83	12.93	52.76	0.83	0.25
Pastaza	36.19	15.22	51.41	0.80	0.38
Pichincha	35.73	12.27	48.00	0.80	0.21
Tungurahua	40.43	13.43	53.86	0.84	0.28
Zamora Chinchipe	38.30	13.01	51.31	0.82	0.26
Galápagos	27.99	14.14	42.13	0.71	0.33
Sucumbíos	36.97	12.92	49.89	0.81	0.25
Orellana	36.22	12.92	49.15	0.80	0.25
Santo Domingo de los Tsáchilas	33.33	14.80	48.13	0.78	0.36
Santa Elena	33.23	14.05	47.28	0.78	0.32

Source: Authors based on Data from the Population and Housing Census (2010)

the head of the household (P23) in the province of Chimborazo shows greater variance, with a value of 52%, in Morona Santiago, this variable has a value of 25.09%. On the other hand, in Galápagos, the landline telephone variable (H07) has the lowest recorded variance with 12.09%, while in Bolívar, it represents 47.40%. In the province of Sucumbíos, the landline telephone variable (H10) presents the lowest value compared to the rest of the provinces which can be explained by the fact that access to a telephone network in the Eastern provinces is rather poor- with a value of 38.79%, and in Loja 46.75%. The

**Table 7.** Explanation of variance of the variables in the first component

Province	P23	H07	H10	URH	V03	V05	VAP	FO	V04	V06
Azuay	0.4566	0.2118	0.4193	0.4763	0.4431	0.4456	0.4782	0.4386	0.4616	0.4895
Bolívar	0.4680	0.4740	0.3620	0.5040	0.2970	0.3810	0.4920	0.4320	0.3810	0.4040
Cañar	0.3975	0.3181	0.3576	0.2941	0.3475	0.4170	0.3516	0.4069	0.4210	0.4787
Carchi	0.4066	0.3490	0.3675	0.4211	0.1903	0.5018	0.3818	0.4317	0.3458	0.4494
Cotopaxi	0.4854	0.4137	0.4262	0.4180	0.1889	0.4935	0.4443	0.4559	0.3961	0.4698
Chimborazo	0.5248	0.4077	0.4611	0.5988	0.1884	0.6426	0.5474	0.4981	0.3375	0.4241
El Oro	0.3322	0.3286	0.3855	0.1391	0.3197	0.2787	0.2176	0.4155	0.4215	0.4598
Esmeraldas	0.3588	0.3940	0.3821	0.3079	0.5456	0.3589	0.3170	0.3271	0.3974	0.3570
Guayas	0.3838	0.4012	0.3992	0.2434	0.4505	0.2847	0.4098	0.4356	0.4865	0.4682
Imbabura	0.4784	0.3617	0.4047	0.3961	0.2241	0.5918	0.4100	0.4481	0.4151	0.4822
Loja	0.5122	0.3671	0.4676	0.5455	0.5195	0.5645	0.4867	0.4093	0.4228	0.4773
Los Ríos	0.3500	0.2972	0.3153	0.2808	0.5009	0.2973	0.3299	0.3110	0.4181	0.3897
Manabí	0.4294	0.3168	0.3936	0.3903	0.5204	0.3295	0.3700	0.3642	0.4517	0.4194
Morona Santiago	0.2509	0.3843	0.4169	0.5012	0.5613	0.3040	0.4430	0.4593	0.2290	0.2568
Napo	0.3319	0.3535	0.3891	0.4107	0.5581	0.4382	0.2757	0.3254	0.4455	0.4551
Pastaza	0.2957	0.3373	0.3850	0.3260	0.4830	0.3256	0.4119	0.4219	0.2947	0.3381
Pichincha	0.4441	0.3746	0.4271	0.1298	0.1985	0.5013	0.2787	0.4783	0.3586	0.3819
Tungurahua	0.4798	0.4112	0.4218	0.4586	0.1093	0.5704	0.4172	0.4423	0.3308	0.4013
Zamora Chinchipe	0.3411	0.3312	0.3633	0.4216	0.5772	0.4170	0.3557	0.3294	0.3535	0.3400
Galápagos	0.3479	0.1209	0.3608	0.1106	0.2193	0.4282	0.0689	0.2973	0.3963	0.4485
Sucumbíos	0.2733	0.2928	0.3079	0.3840	0.6103	0.5308	0.3066	0.2305	0.3883	0.3727
Orellana	0.2971	0.2675	0.3510	0.3491	0.5475	0.4543	0.2798	0.2773	0.4046	0.3939
Santo Domingo de los Tsáchilas	0.2949	0.2588	0.3224	0.2736	0.4975	0.3854	0.2697	0.2864	0.3690	0.3755
Santa Elena	0.3140	0.3716	0.3941	0.0832	0.3085	0.2664	0.1858	0.4495	0.4421	0.5077

Source: Authors based on Data from the Population and Housing Census (2010)

area (URH) of house location in Santa Elena has a value of 8.3% and 59.88% in Chimborazo. The wall material (V03) in Sucumbíos has the highest value (61.03%), while Tungurahua has the lowest (10.92%). Santa Elena presents a variance of 26.63% and Chimborazo a 64.26%. In Galápagos, the main entrance path variable (VAP) presents a variance of 6.89%; owing to the fact that in this province, the main entrance road for over 65% of the dwellings

is paved, cobbled or concrete. In the characteristics related to wall (V04) and floor (V06) conditions, Morona Santiago presents the lowest values at province level: 22.89% and 25.67% respectively. The highest values are in Guayas (48.74%) and Santa Elena (50.77%). A similar analysis can be performed of the results of the second component.

### *Quantification analysis*

The CATPCA algorithm applied for each province with the selected variables (See Table 5) makes it possible to obtain the optimal quantifications of the categories of variables (See table 8). Using these results, the outcomes can be analyzed. The variables describing the materials of the walls (V03) and floor (V05) of the house, along with those that describe their condition (V04, V06 respectively) present the lowest value quantifications for the first categories when compared to the quantification of the first categories for the rest of the variables.

We can also see that the quantifications obtained are similar (not equal) for the different provinces, with the exception of the number of light bulbs variable (F0), which in each province takes values that are quite different from each other. For example, in Pichincha the category of the number of light-bulbs variable take values in the range of -1.104 to 1.959, while in Orellana it ranges between -0.501 to 7.230. This effect is due to the difference in frequency of the categories of this variable in each province.

Another exception is seen in the adobe or clay category in the wall material variable (V03), since in the provinces of Los Ríos, Manabí, Morona Santiago, Napo, Pastaza, Sucumbíos, Orellana, Santo Domingo de los Tsáchilas and Santa Elena the quantifications are positive, while in the rest of the provinces they are negative, thanks to the low frequencies these provinces present for this category. Finally, the categories with null frequency have a null quantization. This characteristic is presented in the second category of the variable, main entrance path to housing (VAP), with alternatives such as river, sea or lake, since some provinces might not display housing with this characteristic due to their geographical location. The variable appears mostly for households belonging to the provinces of the Coast or the Amazon Regions, such as: El Oro, Esmeraldas, Guayas, Los Ríos, Manabí, Morona Santiago, Napo, Pastaza, Zamora Chinchipe, Sucumbíos, and Orellana.

Table 8. Quantifications by province

Abbrev.	Lev.	Azuay	Bolívar	Cañar	Carchi	Cotopaxi	Chimborazo	El Oro	Esmeraldas
P23	1	-1.517	-1.054	-1.216	-1.275	-1.152	-1.174	-1.113	-0.901
P23	2	-1.189	-1.054	-.791	-1.275	-1.003	-1.174	-1.113	-0.901
P23	3	-.591	-.354	-.340	-.603	-0.368	-0.351	-0.702	-0.644
P23	4	-.591	-.354	-.340	-.540	-0.368	-0.351	-0.702	-0.644
P23	5	.623	.621	.783	.845	0.757	0.715	0.195	0.405
P23	6	1.231	1.420	1.587	1.554	1.490	1.306	1.148	1.152
P23	7	1.603	2.102	2.210	2.167	2.177	1.762	2.037	2.264
P23	8	2.120	3.259	3.309	2.926	2.999	2.328	3.249	3.142
H07	1	-.826	-0.534	-0.640	-.779	-0.599	-0.627	-0.537	-0.540
H07	2	1.211	1.871	1.561	1.284	1.668	1.595	1.861	1.851
H10	1	-.742	-.362	-0.483	-.502	-0.456	-0.507	-0.540	-0.406
H10	2	1.347	2.761	2.071	1.992	2.193	1.974	1.852	2.466
URH	1	-1.100	-.673	-0.83	-1.038	-0.682	-0.825	-1.841	-1.019
URH	2	.909	1.485	1.204	.963	1.466	1.212	0.543	0.981
V03	1	-2.379	-2.763	-2.712	-2.450	-3.324	-2.852	-2.970	-1.340
V03	2	-2.379	-2.385	-2.111	-2.450	-2.836	-2.852	-2.462	-1.340
V03	3	-2.233	-1.780	-1.725	-2.450	-2.712	-2.760	-2.012	-1.242
V03	4	-1.630	-1.122	-1.725	-1.382	-2.109	-2.128	-1.981	-1.242
V03	5	-1.529	-1.068	-1.725	-1.382	-2.109	-2.128	-1.981	-0.132
V03	6	.589	.712	.501	.641	0.354	0.352	0.118	0.590
V03	7	1.059	2.126	1.176	1.718	1.878	1.797	1.566	1.629
V05	1	-2.507	-1.505	-2.230	-1.693	-1.385	-1.258	-3.137	-0.826
V05	2	-2.507	-1.505	-2.230	-1.693	-1.385	-1.258	-3.137	-0.826
V05	3	-2.001	-1.505	-2.230	-1.693	-1.156	-0.893	-2.871	-0.826
V05	4	-.455	0.182	-.025	-.561	-0.355	-0.279	-0.218	-0.826
V05	5	.297	0.182	.514	.857	0.749	0.759	0.758	-0.826
V05	6	.297	0.182	.514	.857	0.749	0.759	0.758	1.210
V05	7	.895	2.465	.922	1.261	2.128	1.602	1.162	1.210
VAP	1	-1.569	-1.047	-1.322	-1.398	-1.109	-1.135	-2.053	-1.750
VAP	2							-2.053	-1.750
VAP	3	-1.569	-1.047	-1.322	-1.398	-1.109	-1.135	-2.053	-1.488
VAP	4	-.610	-.583	-.389	-.850	-0.526	-0.668	-0.826	-0.348
VAP	5	-.187	.092	.072	-.850	-0.035	0.174	-0.538	-0.116

(Continued)

Table 8. Quantifications by province

Abbrev.	Lev.	Azuay	Bolívar	Cañar	Carchi	Cotopaxi	Chimborazo	El Oro	Esmeraldas
VAP	6	.975	1.339	1.339	1.009	1.351	1.146	1.063	1.315
FO	1	-1.009	-556	-.804	-.728	-0.634	-0.688	-0.654	-0.537
FO	2	.291	1.348	.639	.961	1.073	1.101	1.097	1.512
FO	3	1.238	2.806	1.729	2.189	2.329	1.991	2.312	2.680
FO	4	1.594	3.280	2.240	2.386	2.799	2.298	2.817	3.267
FO	5	1.796	3.500	2.692	2.599	3.087	2.451	3.121	3.547
FO	6	2.020	3.500	2.964	2.599	3.087	2.451	3.233	3.547
FO	7	2.217	3.500	3.135	3.089	3.087	2.451	3.341	3.585
FO	8	2.251	3.500	3.135	3.089	3.499	2.476	3.580	4.076
FO	9	2.251	4.042	3.185	3.089	3.894	2.476	3.580	4.076
V04	1	-2.216	-1.549	-2.001	-1.907	-1.982	-2.021	-1.937	-1.662
V04	2	-.580	-0.31	-.342	-.406	-0.390	-0.486	-0.564	-0.390
V04	3	.829	1.28	.956	1.133	1.060	1.042	0.974	1.190
V06	1	-2.084	-1.374	-1.856	-1.621	-1.601	-1.684	-1.776	-1.631
V06	2	-0.517	-.229	-.274	-.391	-0.312	-0.393	-0.479	-0.377
P23	1	-0.943	-1.471	-1.230	-0.847	-1.059	-1.330	-0.991	-1.258
P23	2	-0.943	-1.471	-1.230	-0.813	-1.059	-0.867	-0.916	-1.258
P23	3	-0.821	-0.527	-0.641	-0.553	-0.534	-0.462	-0.618	-0.943
P23	4	-0.762	-0.463	-0.641	-0.553	-0.534	-0.462	-0.618	-0.717
P23	5	-0.014	0.671	0.513	0.363	0.452	0.186	0.081	0.183
P23	6	0.803	1.020	0.934	1.380	1.467	0.676	0.806	0.709
P23	7	1.824	1.748	1.669	2.484	2.047	2.306	2.162	1.728
P23	8	3.167	2.289	2.282	4.045	3.249	3.381	3.327	2.976
H07	1	-0.675	-0.781	-0.621	-0.385	-0.421	-0.636	-0.542	-0.643
H07	2	1.480	1.280	1.610	2.598	2.375	1.572	1.846	1.556
H10	1	-0.545	-0.634	-0.597	-0.344	-0.421	-0.490	-0.511	-0.600
H10	2	1.835	1.578	1.676	2.904	2.378	2.042	1.956	1.666
URH	1	-2.334	-1.114	-1.124	-1.070	-1.144	-0.800	-0.834	-0.990
URH	2	0.428	0.897	0.890	0.935	0.874	1.250	1.199	1.010
V03	1	-2.360	-3.616	-1.863	-1.760	-1.380	-2.148	-1.082	-4.293
V03	2	-2.360	-3.616	-1.863	-1.760	-1.380	-2.148	-1.082	-3.753
V03	3	-1.884	-3.616	-1.721	-1.542	-1.263	-2.093	-1.082	-3.148
V03	4	-1.884	-1.572	-1.132	-1.394	-1.263	-0.368	-1.007	-0.243

(Continued)

Table 8. Quantifications by province

Abbrev.	Lev.	Azuay	Bolívar	Cañar	Carchi	Cotopaxi	Chimborazo	El Oro	Esmeraldas
V03	5	0.055	-1.572	-1.132	0.547	0.442	0.289	0.530	-0.059
V03	6	0.406	0.511	0.731	0.547	0.629	1.060	0.912	0.569
V03	7	0.745	1.599	1.473	1.305	1.554	1.273	1.268	0.679
V05	1	-0.445	-1.754	-1.587	-0.590	-0.752	-2.597	-0.908	-3.780
V05	2	-0.445	-1.754	-1.587	-0.590	-0.752	-2.597	-0.908	-3.780
V05	3	-0.445	-1.649	-1.587	-0.590	-0.752	-2.484	-0.908	-3.273
V05	4	-0.445	-0.473	-0.156	-0.590	-0.752	0.233	-0.908	0.217
V05	5	-0.445	0.911	0.626	-0.590	-0.752	0.233	-0.908	0.217
V05	6	2.247	0.911	0.626	1.696	1.329	0.233	1.102	0.217
V05	7	2.247	1.297	1.378	1.696	1.329	1.530	1.102	0.554
VAP	1	-2.165	-1.547	-1.389	-1.604	-1.672	-1.945	-2.116	-3.745
VAP	2	-2.165			-1.604	-1.672	-1.945	-2.116	-3.745
VAP	3	-2.165	-1.547	-1.389	-1.604	-1.672	-1.495	-1.899	-1.571
VAP	4	-1.041	-0.980	-0.532	-0.704	-0.439	0.022	-0.288	0.155
VAP	5	-0.487	-0.200	0.187	-0.171	0.160	0.322	0.317	0.445
VAP	6	0.876	1.074	1.117	1.294	1.105	1.167	1.094	0.659
FO	1	-0.633	-0.827	-0.754	-0.439	-0.588	-0.645	-0.601	-0.709
FO	2	0.987	0.764	0.704	1.807	1.156	1.209	1.137	1.002
FO	3	2.183	1.728	1.915	3.633	2.714	2.063	2.451	2.037
FO	4	2.762	2.054	2.257	4.304	3.309	2.410	2.906	2.388
FO	5	2.973	2.192	2.484	4.608	3.393	2.589	3.444	2.527
FO	6	3.336	2.192	2.484	4.903	3.666	2.717	3.444	2.527
FO	7	3.336	2.192	2.601	5.159	3.995	2.717	3.444	2.527
FO	8	3.633	2.255	2.725	5.159	3.995	2.874	3.444	2.633
FO	9	3.668	2.255	2.725	5.159	3.995	3.081	3.444	2.739
V04	1	-1.999	-2.118	-1.702	-1.887	-1.263	-1.598	-1.510	-1.692
V04	2	-0.674	-0.548	-0.469	-0.150	-0.472	-0.536	-0.450	-0.593
V04	3	0.952	0.925	1.038	1.063	1.336	1.122	1.248	1.061
V06	1	-1.667	-1.896	-1.537	-1.807	-1.247	-1.603	-1.429	-1.636
V06	2	-0.681	-0.551	-0.325	-0.197	-0.453	-0.574	-0.464	-0.683
V06	3	1.036	0.948	1.109	1.138	1.375	1.095	1.279	1.017

(Continued)

Table 8. Quantifications by province

Abbrev.	Lev.	Pichincha	Tungurahua	Zamora Ch	Galápagos	Sucumbíos	Orellana	Sto D	S.Elena
P23	1	-1.676	-1.463	-1.154	-1.381	-0.962	-1.014	-1.170	-0.870
P23	2	-1.626	-1.463	-0.684	-1.381	-0.837	-0.872	-1.170	-0.870
P23	3	-1.030	-0.836	-0.597	-1.135	-0.589	-0.831	-0.618	-0.705
P23	4	-0.990	-0.535	-0.597	-1.135	-0.589	-0.647	-0.618	-0.661
P23	5	-0.003	0.603	0.428	-0.032	0.378	0.546	0.403	0.455
P23	6	0.657	1.178	1.131	0.834	1.309	1.571	1.324	1.352
P23	7	1.248	1.768	2.177	1.443	2.657	2.364	2.231	2.343
P23	8	1.877	2.304	2.926	1.858	3.621	3.859	3.360	3.795
H07	1	-1.216	-0.733	-0.637	-1.483	-0.478	-0.387	-0.661	-0.447
H07	2	0.822	1.365	1.569	0.675	2.091	2.581	1.512	2.235
H10	1	-0.968	-0.594	-0.490	-0.927	-0.444	-0.447	-0.523	-0.393
H10	2	1.033	1.682	2.041	1.079	2.253	2.236	1.912	2.542
URH	1	-1.541	-0.847	-0.850	-2.351	-0.952	-0.958	-1.698	-1.120
URH	2	0.649	1.181	1.177	0.425	1.050	1.044	0.589	0.893
V03	1	-3.953	-3.982	-1.205	-4.128	-1.163	-0.836	-2.575	-2.083
V03	2	-3.953	-3.982	-1.205	-4.128	-1.163	-0.836	-2.575	-2.083
V03	3	-3.953	-3.982	-1.205	-4.026	-1.163	-0.836	-2.575	-1.653
V03	4	-3.953	-3.592	-1.205	-4.026	-1.085	-0.769	-2.575	-0.775
V03	5	-2.621	-3.592	-0.755	-1.769	0.273	0.060	0.044	0.456
V03	6	0.006	0.124	0.787	-0.040	0.875	1.211	0.351	0.456
V03	7	1.577	1.482	1.118	1.401	1.181	1.736	0.687	1.259
V05	1	-2.631	-1.678	-0.999	-2.947	-0.873	-1.205	-0.384	-2.081
V05	2	-2.631	-1.678	-0.999	-2.947	-0.873	-1.205	-0.384	-2.081
V05	3	-2.631	-1.678	-0.999		-0.873	-1.205	-0.384	-2.081
V05	4	-1.258	-0.697	-0.999	-1.002	-0.873	-1.205	-0.384	0.481
V05	5	0.105	0.371	-0.999	0.896	-0.873	-1.205	-0.384	0.481
V05	6	0.105	0.371	1.001	0.896	1.145	0.830	2.606	0.481
V05	7	1.117	1.421	1.001	0.896	1.145	0.830	2.606	0.481
VAP	1	-1.906	-1.268	-1.645	-1.667	-2.179	-1.743	-3.307	-1.437
VAP	2			-1.645		-2.179	-1.743		
VAP	3	-1.906	-1.268	-1.645	-1.667	-1.932	-1.743	-3.307	-1.437
VAP	4	-1.638	-0.913	-0.162	-1.667	-0.602	-0.337	-0.414	-0.704
VAP	5	-1.541	-0.430	0.083	-0.262	0.260	0.284	0.190	0.038

(Continued)

Table 8. Quantifications by province

Abbrev.	Lev.	Pichincha	Tungurahua	Zamora Ch	Galápagos	Sucumbíos	Orellana	Sto D	S.Elena
VAP	6	0.605	1.082	1.277	0.671	1.071	1.247	0.805	1.604
FO	1	-1.104	-0.796	-0.625	-0.869	-0.514	-0.501	-0.693	-0.523
FO	2	0.390	0.737	1.207	0.783	1.496	1.495	0.919	1.352
FO	3	1.183	1.799	2.391	1.574	2.985	2.940	2.292	3.007
FO	4	1.521	2.323	2.581	1.832	3.512	3.920	2.930	3.649
FO	5	1.708	2.491	3.117	2.351	4.396	3.920	3.256	3.802
FO	6	1.818	2.529	3.117	2.414	4.412	4.668	3.256	3.802
FO	7	1.916	2.729	3.117	2.747	4.412	4.668	3.256	3.889
FO	8	1.916	2.773	3.343	2.747	4.412	4.805	3.256	3.889
FO	9	1.959	2.885	3.519	2.747	4.412	7.230	3.583	3.889
V04	1	-2.583	-2.479	-1.524	-3.371	-1.630	-1.374	-1.901	-1.583
V04	2	-1.278	-0.626	-0.485	-1.486	-0.351	-0.482	-0.619	-0.386
V04	3	0.672	0.860	1.196	0.582	1.292	1.336	1.057	1.383
V06	1	-2.400	-2.016	-1.449	-3.161	-1.506	-1.301	-1.719	-1.461
V06	2	-1.299	-0.546	-0.422	-1.411	-0.424	-0.543	-0.682	-0.183
V06	3	0.664	0.935	1.260	0.596	1.326	1.325	1.059	1.447

Source: Authors based on Data from the Population and Housing Census (2010)

Stability analysis (Bootstrap)

A stability analysis is shown below to complement the results obtained in the previous sections and to identify whether the quantifications are stable. In other words, the analysis must be robust to changes in the selection of data and given that the quantification process is considered an integral part of the analysis, the algorithm is applied for each bootstrap sample. If the quantifications are substantially different from the quantifications of the principal sample in the analysis of the bootstrap samples, the quantifications will be unstable. If the quantifications have small confidence intervals, this means that they are stable. Table 9 presents the quantifications (Quantification) obtained from the CATPCA algorithm, along with the lower (Lower) and upper (Upper) limits of each category obtained using the bootstrap method. Specifically, the stability analysis of the results obtained for the province of Guayas is shown. The study considers a total number of bootstrap samples within each province, with a 95% confidence level and the solutions are obtained in two dimensions. Considering the aforementioned criteria, it can be concluded that the



quantifications obtained in the province of Guayas are stable, since the confidence intervals are small.

**Table 9.** Guayas - Quantifications and confidence intervals

Abbreviation	Category	Level	Quantification	Lower	Higher
P23	None	1	-0.94	-0.99	-0.89
P23	Literacy center	2	-0.94	-0.99	-0.89
P23	Pre-school	3	-0.82	-0.97	-0.75
P23	Primary	4	-0.76	-0.78	-0.74
P23	Secondary	5	-0.01	-0.04	0.01
P23	Higher (non-university)	6	0.80	0.65	0.97
P23	Higher (university)	7	1.82	1.79	1.86
P23	Postgraduate	8	3.17	3.04	3.30
H07	No	1	-0.68	-0.68	-0.67
H07	Yes	2	1.48	1.46	1.50
H10	No	1	-0.54	-0.55	-0.54
H10	Yes	2	1.84	1.81	1.86
URH	Rural area	1	-2.33	-2.37	-2.30
URH	Urban area	2	0.43	0.42	0.43
V03	Other materials	1	-2.36	-2.41	-2.31
V03	Uncoated cane	2	-2.36	-2.41	-2.31
V03	Coated cane	3	-1.88	-1.95	-1.84
V03	Wood	4	-1.88	-1.93	-1.78
V03	Adobe or clay	5	0.06	-0.16	0.27
V03	Brick or block	6	0.41	0.40	0.42
V03	Concrete	7	0.75	0.71	0.78
V05	Other materials	1	-0.44	-0.45	-0.44
V05	Dirt	2	-0.44	-0.45	-0.44
V05	Cane	3	-0.44	-0.45	-0.44
V05	Brick or concrete	4	-0.44	-0.45	-0.44
V05	Ceramic. tile. vinyl or marble	5	-0.44	-0.45	-0.44
V05	Untreated plank	6	2.25	2.22	2.28
V05	Boarding, parquet or floating floor	7	2.25	2.22	2.28
VAP	Other	1	-2.16	-2.25	-2.07
VAP	River / sea / lake	2	-2.16	-2.25	-2.07

(Continued)

Table 9. Guayas – Quantifications and confidence intervals

Abbreviation	Category	Level	Quantification	Lower	Higher
VAP	Road, path, chaquiñán (trail)	3	-2.16	-2.25	-2.07
VAP	Dirt road	4	-1.04	-1.08	-1.00
VAP	Paved stone	5	-0.49	-0.54	-0.44
VAP	Cobbled. p c	6	0.88	0.86	0.89
FO	5 or less	1	-0.63	-0.64	-0.62
FO	From 6 to 10	2	0.99	0.96	1.02
FO	From 11 to 15	3	2.18	2.11	2.25
FO	from 16 to 20	4	2.76	2.66	2.87
FO	From 21 to 25	5	2.97	2.81	3.14
FO	From 26 to 30	6	3.34	3.18	3.49
FO	From 31 to 35	7	3.34	3.18	3.50
FO	From 36 to 40	8	3.63	3.41	3.78
FO	More than 41	9	3.67	3.51	3.87
V04	Poor	1	-2.00	-2.07	-1.93
V04	Regular	2	-0.67	-0.70	-0.64
V04	Good	3	0.95	0.94	0.97
V06	Poor	1	-1.67	-1.73	-1.61
V06	Regular	2	-0.68	-0.70	-0.66
V06	Good	3	1.04	1.02	1.05

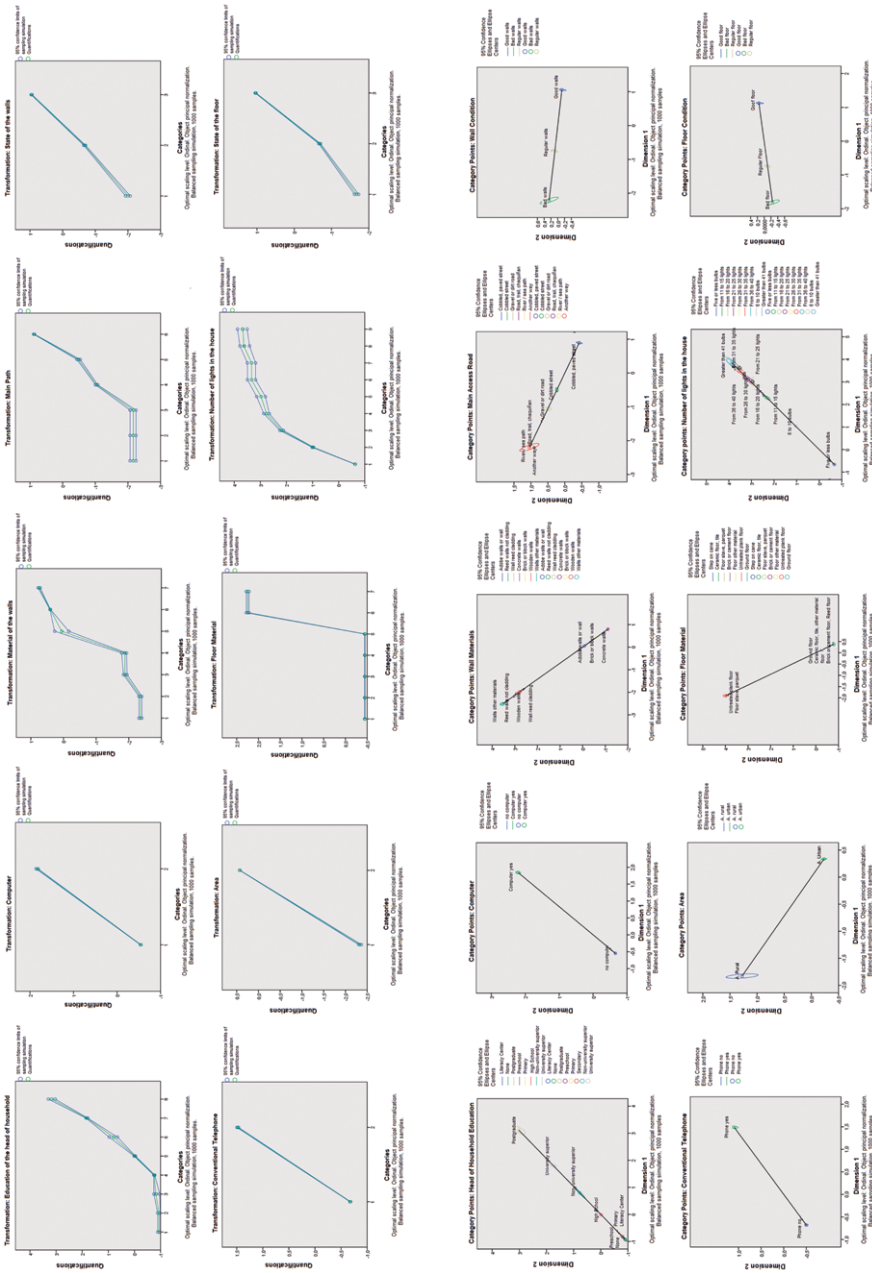
Source: Authors based on Data from the Population and Housing Census (2010)

4.2. Discussion

4.2.1 Analysis of national behavior

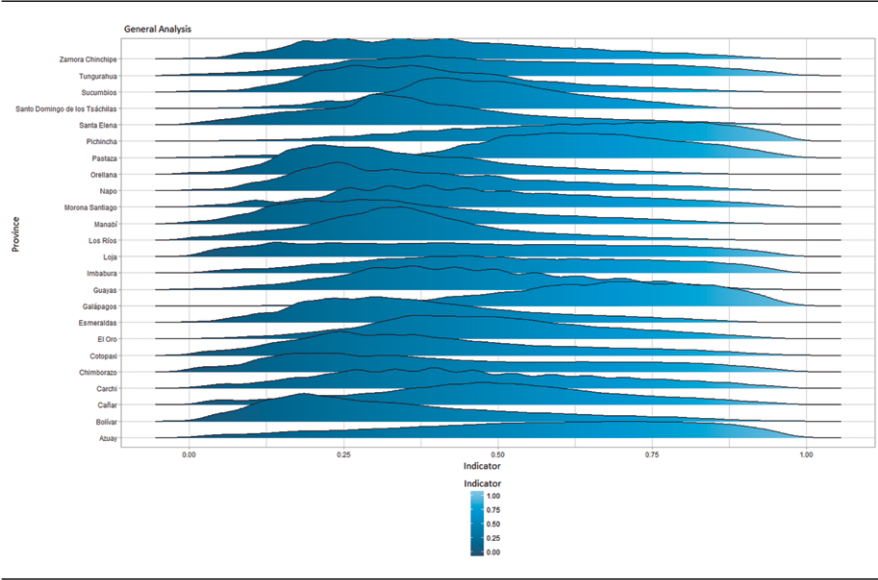
The indicators for each province are constructed based on the quantifications of the categories of variables in Table 5. Figure 2 displays the distributions of the indicators for Ecuador’s different provinces. The indicator values fall within the range of [0; 1] due to the rescaling performed; however, it is important to note that this does not mean that the indicators are comparable, as the quantifications of categories obtained for each province and with which the indicator is constructed with are different. To exemplify, for the variable head of the household’s education level (P23), the quantification of the first category –no education– in Pichincha is equal to -1.73, while in Sucumbíos it is equal

Figure 1. Guayas – Bootstrap



to  $-0.96$ . The quantification of the last category, postgraduate, in Pichincha is equal to 1.88 and in Sucumbíos, it is equal to 3.63.

Figure 2. Indicator for the national level



Once the socioeconomic indicators of each province have been constructed, a descriptive statistics analysis for each of these indicators is carried out. See Table 10. For the reasons stated above, the analysis to follow is not performed in a comparative context, but a descriptive context within each province. The average value for the socioeconomic indicator of households in the provinces of: Azuay, Pichincha, Pastaza and Galápagos shows high figures, whereas the lowest values appear for the provinces of: Orellana, Bolívar, Los Ríos and Manabí. At national level, the average values are in the range of 0.3350 to 0.673. On the other hand, the dispersion of the socioeconomic indicators of the provinces of Azuay, Chimborazo and Loja are high; while in Santo Domingo and Los Ríos the dispersion is low. The degree of dispersion for all provinces varies between 0.146 and 0.251 and, in general, it is high in all provinces. Moreover, Pastaza, Galápagos, Pichincha, Azuay and Imbabura have a negative asymmetry value; that is, the values of the household indicator tend to be located more on the right side of the average. In the rest of the provinces (See Figure 2), this coefficient is positive, so their distribution is concentrated on the left. This frame demonstrates that within each province a large percentage of households are

below average. The result reflects the country's unequal living conditions in the country and also suggests that there is a minority group of households with good socioeconomic conditions.

Table 10. Summary of indicators

Province	Average	Standard Deviation	Coefficient of asymmetry	Kurtosis	Coefficient of variation
Azuay	0.572	0.230	-0.355	2,260	40,243
Bolívar	0.337	0.200	0.764	2,840	59,348
Cañar	0.481	0.204	0.016	2,560	42,370
Carchi	0.458	0.216	0.235	2,260	47,131
Cotopaxi	0.395	0.210	0.547	2,570	53,048
Chimborazo	0.427	0.246	0.363	2,200	57,618
El Oro	0.484	0.181	0.219	2,620	37,317
Esmeraldas	0.376	0.182	0.617	2,820	48,418
Guayas	0.450	0.182	0.225	2,440	40,452
Imbabura	0.523	0.223	-0.034	2,130	42,653
Loja	0.460	0.251	0.125	1,920	54,565
Los Ríos	0.351	0.158	0.487	3,370	44,994
Manabí	0.352	0.187	0.596	2,730	53,050
Morona Santiago	0.464	0.209	0.263	2,360	45,014
Napo	0.386	0.194	0.613	2,640	50,299
Pastaza	0.627	0.178	-0.706	3,880	28,323
Pichincha	0.628	0.202	-0.418	2,440	32,206
Tungurahua	0.493	0.218	0.146	2,190	44,302
Zamora Chinchipe	0.420	0.200	0.393	2,420	47,702
Galápagos	0.673	0.163	-0.484	2,860	24,159
Sucumbíos	0.391	0.167	0.556	2,920	42,780
Orellana	0.335	0.164	0.693	3,670	48,922
Santo Domingo de los Tsáchilas	0.473	0.146	0.098	2,850	30,945
Santa Elena	0.377	0.185	0.597	3,120	49,160

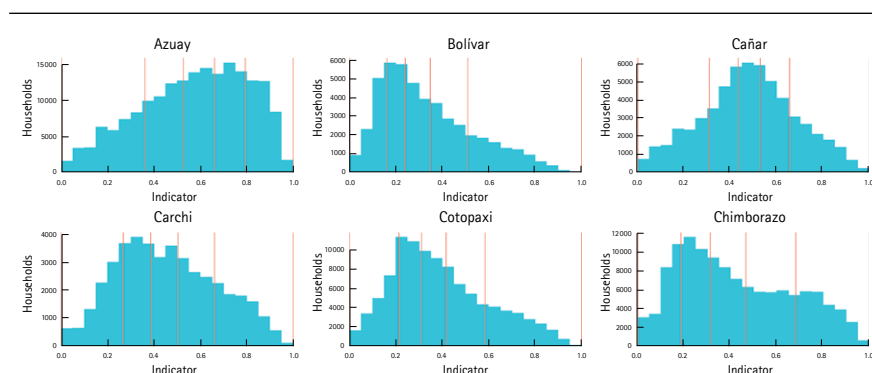
Source: Authors based on Data from the Population and Housing Census (2010)

The kurtosis coefficient for Los Ríos, Orellana and Pastaza is greater than 3, which implies that the distributions of the indicator are leptokurtic; that is, households are concentrated in the average. In the rest of the provinces this coefficient takes values of less than 3, indicating that there is a small concentration of households

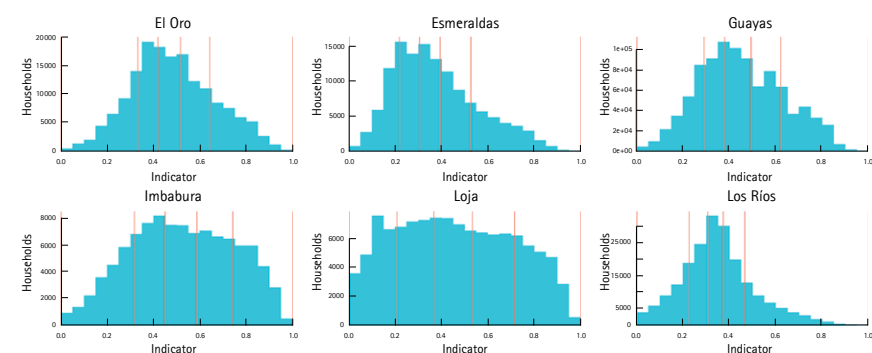
around the average. As the coefficient of variation makes it possible to compare the relative dispersion of the socioeconomic indicators of provinces, the greatest dispersion of the socioeconomic indicator is observed in the provinces of Bolívar, Chimborazo, Cotopaxi, Loja and Manabí. This reflects that households within these provinces present greater variability than in the rest of the provinces. Finally, in Pichincha, Pastaza, Galápagos and Santo Domingo de los Tsáchilas there is less variability among households.

Thus, the households within each province are grouped over quintiles, where the first quintile represents the households in worse conditions in terms of goods and services, while the fifth quintile represents the households in better conditions. In the Figure 3 and 4, the cut-off point of the quintiles is observed for each of the country's provinces.

Figure 3. Indicator - Quintiles by provinces

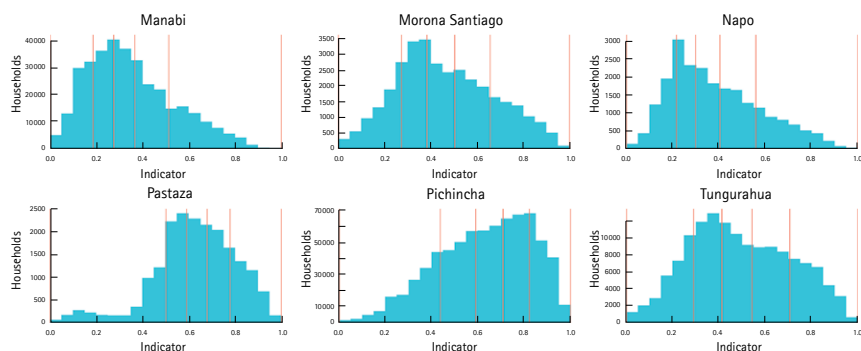


(a) Azuay, Bolívar, Cañar, Carchi, Cotopaxi, Chimborazo

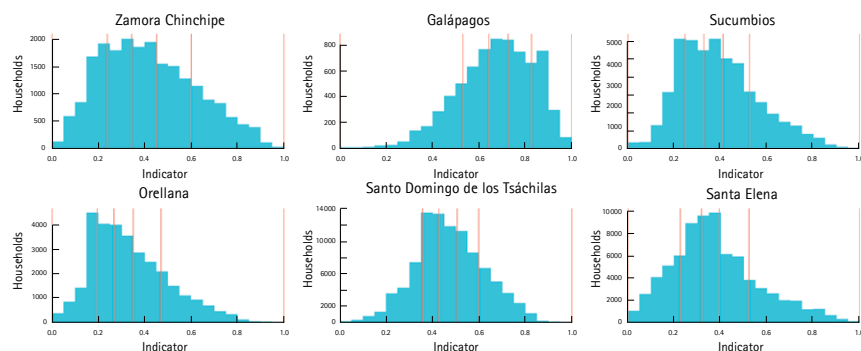


(b) El Oro, Esmeraldas, Guayas, Imbabura, Loja, Los Ríos

Figure 4. Indicator - Quintiles by provinces



(a) Manabí, M. Santiago, Napo, Pastaza, Pichincha, Tungurahua



(b) Z. Chinchipe, Galápagos, Sucumbios, Orellana, S. Domingo, S. Elena

It is important to mention that the province of Galápagos is a nature reserve that is isolated to a certain extent thanks to its geographical position. Also, its main economic activity is tourism, which indirectly affects all sectors of the island's local economy (Taylor et al., 2007). The above, together with the fact that it presents the lowest percentage of variance explained by the components (See Table 6), the general results for Galápagos are different from the rest of the country's provinces.

#### 4.2.2 Distribution analysis of the indicator over quintiles

Given the importance of having a detailed description of the different quintiles within each province, their characteristics are presented for the particular case of the province of Pichincha.

*Pichincha*

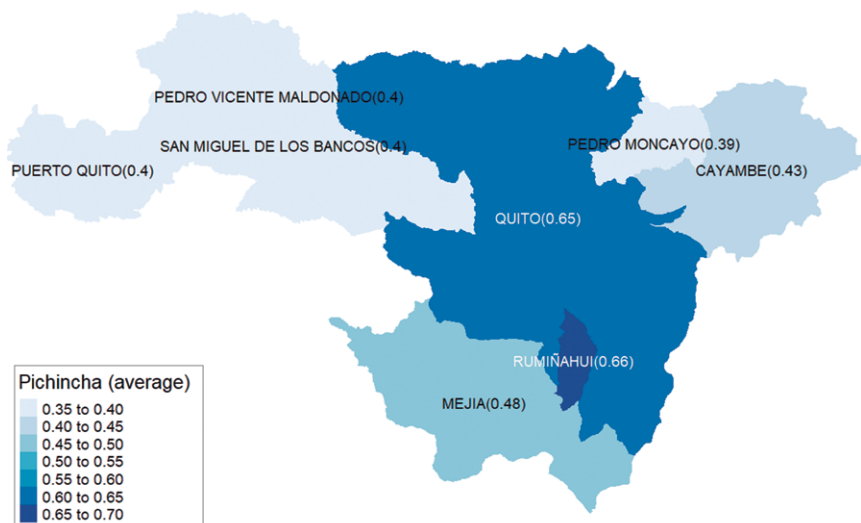
The province of Pichincha contains about 19.14% of the country's households distributed in 8 cantons. Figure 5(a) shows that the Rumiñahui canton has the highest average indicator value (0,66), followed by Quito (0,65). This implies that, in general, the households of these cantons have better living conditions. This could be explained by the fact that Quito, the capital of Ecuador, and Rumiñahui are strongly interconnected by geographical location, labor interests and economic activity between these two cities. On the other hand, Pedro Moncayo has the lowest average value (0.39). The households in this canton have the worst living conditions in the province of Pichincha. When analyzing Figure 5(b), the largest indicator dispersion is in the Cayambe canton (0.22), while the smallest dispersion occurs in Pedro Vicente Maldonado and San Miguel de los Bancos (0.16).

Figure 6 displays an analysis of the quintiles of the households belonging to Pichincha cantons. 64.66% of Pedro Vicente Maldonado's households belong to the first quintile, which makes it the canton with the highest percentage of households belonging to this quintile, followed by San Miguel de los Bancos and Pedro Moncayo with 63.39% and 63.09% respectively. The main activities in these cantons are agriculture and tourism. At the other extreme, Quito and Rumiñahui have the lowest percentage of households within the quintile, with 16.65% and 15.57% respectively and the dispersion in each canton ranges between 0.08 and 0.12. Some 26.35% of Mejía households are in the second quintile, followed by Pedro Vicente Maldonado with 21.64%. The cantons with the lowest percentage of households in this quintile are Quito and Rumiñahui with 19.74% and 19,5%, respectively. The analysis of the following quintiles is similar, since as of the third quintile, Quito and Rumiñahui have a higher percentage of households. Pedro Moncayo and Pedro Vicente Maldonado have the lowest percentage of households within these quintiles. The dispersion of Pichincha households within each quintile is low, except for the one belonging to the first quintile that has dispersions of up to 0.12. In the other quintiles, the dispersion fluctuates between 0.03 and 0.04.

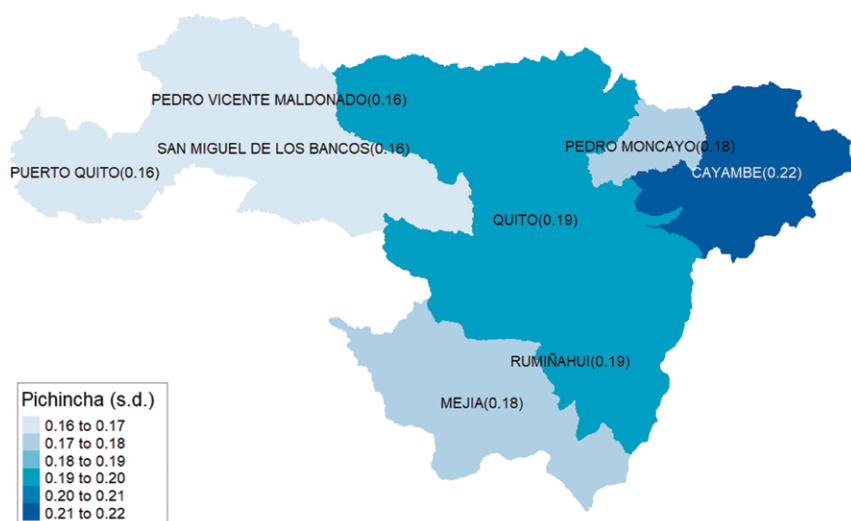
Table 11 presents the profile of the households belonging to each quintile. The characteristics of the first quintile consist of rented, lent or ceded housing or type of rooms in tenement house—which constitutes the majority—and those located in the rural area. The housing materials are tile, wood, adobe, and brick with poor conditions. They do not have basic services such as electric power,



Figure 5. Pichincha - Indicator



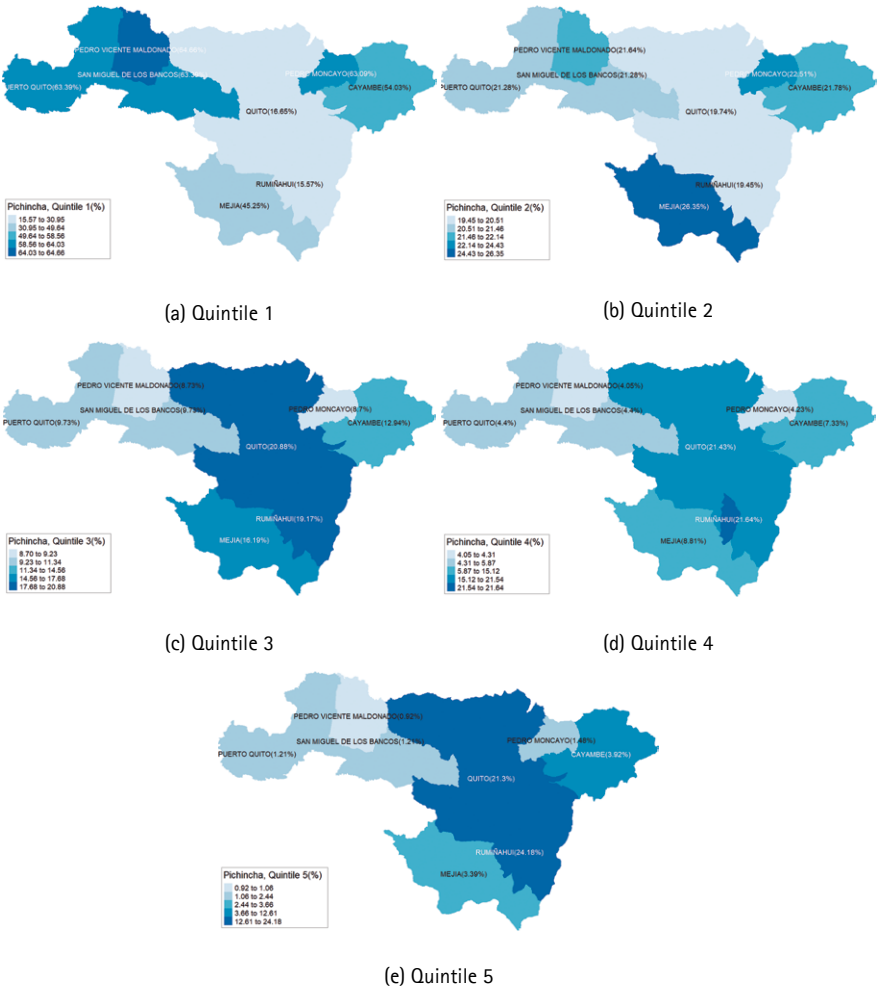
(a) Average



(b) Standard deviation

Fuente: Elaboración propia

Figure 6. Pichincha - Quintiles



Fuente: Elaboración propia

WC, garbage collection, landline telephone, internet, computer, cell phone and cable television. The head of household has a low level of schooling, none of the household members has private insurance, and they study in state-run educational centers. The households belonging to the second quintile are characterized by being in the rural area, with their own housing but inherited or ceded, the materials of the roof, walls and floor are zinc, adobe and brick or cement, respectively. These tend not to be in very good condition. They don't have the

basic services; the shower and WC is shared with other households and the level of education of the head of household is primary. The members of the households of this quintile conduct their studies in public educational centers and no one has private insurance. In the third, fourth and fifth quintiles, the households have a WC and shower for private use, as well as the basic services mentioned above. The housing of these households in these quintiles are their own, and tend to be apartments or houses that are in good condition, built with brick, block, concrete, ceramic tile or parquet materials. The main entrance road is paved or concrete. Household members of the third quintile study in public centers and do not have private insurance. Those in the fourth and fifth quintiles study in private centers and some or all members of the household have private insurance.

**Table 11.** Analysis of Pichincha Individuals

Variable	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Education of the head of household	Literacy center None Preschool	Preschool Primary	Secondary	Higher (non-university)	Higher (university) Postgraduate
Type of WC	None	Shared	Exclusive use	Exclusive use	Exclusive use
Shower	None	Shared	Exclusive use	Exclusive use	Exclusive use
Telephone land-Line	No	No	Yes	Yes	Yes
Cell phone	No	No	Yes	Yes	Yes
Internet	No	No	No	Yes	Yes
Computer	No	No	No	Yes	Yes
Cable TV	No	No	No	Yes	Yes
Possession of housing	Rented, lent, or ceded (not paid)	Own (gifted, donated, inherited or in possession)	For service	Own and being-Paid	Own and totally
Area	A. Rural	A. Rural	A. Urban	A. Urban	A. Urban
Ceiling material	Tile or shingle	Zinc sheet	Asbestos	Concrete	Concrete
Wall material	Wood Adobe or clay	Adobe or clay	Brick or block	Concrete	Concrete

(Continued)

Table 11. Analysis of Pichincha Individuals

Variable	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Floor covering material	Dirt Brick or concrete	Brick or concrete	Ceramic, ceramic tile, vinyl or marble	Ceramic, ceramic tile, vinyl or marble Boarding, parquet, Plank	Boarding, parquet, plank floating floor
Water pro-water	From river, slope, canal from well	By public network	By public network	By public network	By public network
How do they the water	Does not receive water pipeline By pipeline outside	By pipeline outside The housing, inside the Building	By pipeline inside the housing	By pipeline inside the housing	By pipeline inside the housing
WC	None Connected to septic tank or cesspit	Connected to public Sewage Network	Connected to public sewage network	Connected to public Sewage Network	Connected to public sewage network
Garbage the disposal	Other They burn it Throwing into river, ditch or channel	By garbage collection vehicle	By garbage collection vehicle	By garbage collection vehicle	By garbage collection vehicle
Main entrance Path	Other Road, path, chaquiñan Ballast or dirt	Paved stone Ballast or dirt	Cobbled, paved, concrete	Cobbled, paved, concrete	Cobbled, paved, concrete
Electric energy service	None Electric company network	Electric company network for public service	Electric company network for public service	Electric company network for public service	Electric company network for public service
Electric meter	None	Shared	Exclusive use	Exclusive use	Exclusive use
Type of housing	Room(s) in tenement House	Apartment in house or building	Apartment in house or building	House/Villa	House/Villa
Ceiling condition	Poor	Regular	Good	Good	Good
Wall condition	Poor	Regular	Good	Good	Good
Condition of the floor covering	Poor	Regular	Good	Good	Good

(Continued)

Table 11. Analysis of Pichincha Individuals

Variable	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Number of light bulbs	5 or less	From 6 to 10	From 6 to 10	From 6 to 10	More than 10
Last payment of electricity	5 or less	From 5 to 9	From 10 to 14	From 10 to 20	More than 20
Type of education	Mostly	Mostly	Mostly	Mostly	Mostly
center	State-run	State-run	State-run	Private	private
Private insurance	Nobody	Nobody	Nobody	Some people	All

## 5. Conclusions

This article built socioeconomic stratification indicators for provinces in Ecuador. The results of this analysis constitute a tool for the fields of economics, political science, demography, sociology and marketing. In economics, for example, the information on social strata is essential for designing and implementing public policies. We constructed indicators of socio-economic stratification at the provincial level, which reduce the variability of the indicators obtained with respect to those obtained in similar studies, also conducted in Ecuador. Another contribution of this article is the application of the nonlinear PCA that allows the incorporation of quantitative and qualitative variables.

Our results were subjected to a sensitivity analysis using the bootstrap method and in general, showed that they are robust. However, some categories presented a low frequency which ultimately lead to large confidence intervals. This issue could be solved by appropriately clustering these low frequency categories. Another alternative is to modify the scaling of the abovementioned categories by using different scale levels, such as ordinal spline. On the other hand, the main advantage of using nonlinear PCA, compared to other similar techniques, is that the optimal quantifications obtained from ordinal variables retain their ordinal property, which allows for a socio-economic stratification interpretation of indicators.

Regarding the socio-economic conditions, it can be concluded that the households located in the urban area display better conditions and greater access to

basic services. Likewise, the education level of the head of household is a key factor when characterizing household quintiles and the results suggest that, in general, education positively affects social and economic conditions of both individuals and the households. The results also suggest high inequality among the provinces. In Guayas, for example, there is huge inequality between Samborondón and Colimes. A similar situation is observed in Pichincha with cantons like Quito and Pedro Vicente Maldonado. In light of these results, public policy should be focused on the education, both in quantity and quality, of the school-age inhabitants of the cantons in worse social and economic conditions. Likewise, public investment in the provision of basic services such as drinking water, sewerage, electricity, access roads and garbage collection should focus on the cantons that, as explained by their conditions, have been the most neglected.

## Acknowledgements

The authors did not receive funding from any institution to carry out this research. The authors would like to thank the anonymous reviewers for their suggestions and comments to the original version of this paper.

## References

1. Acemoglu D., Naidu S., Restrepo, P. and Robinson, J. (2015) Democracy, Redistribution, and inequality. *In Handbook of Income Distribution*, 2, 1885-1966.
2. Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
3. Davison, A. C., Hinkley, D. V., & Schechtman, E. (1986). Efficient bootstrap simulation. *Biometrika*, 73(3), 555-566, <https://doi.org/10.1093/biomet/73.3.555>
4. Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*, 7, 1-26.

5. Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
6. Ferrari, P. A., & Manzi, G. (2010). Nonlinear principal component analysis as a tool for the evaluation of customer satisfaction. *Quality technology & quantitative management*, 7(2), 117-132.
7. Featherman, D., Hauser, R. (1975). Design for a Replicate Study of Social Mobility in the United States. In *Social Indicator Models*, edited by Kenneth C. Land and Seymour Spilerman. New York: Russell Sage.
8. Fujihara, S. (2020). Socio-Economic Standing and Social Status in Contemporary Japan: Scale Constructions and Their Applications, *European Sociological Review*, 36 (4), 548-561.
9. Gifi, A. (1990). *Nonlinear multivariate analysis*. Wiley.
10. Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. The prediction of personal adjustment. New York: Social Science Research Council.
11. Haug, M. (1977). Measurement in Social Stratification. *Annual Review of Sociology*, 3(1), 51-77.
12. Haer, J. (1957). Predictive Utility of Five Indices of Social Stratification. *American Sociological Review*, 22(5), 541-546.
13. Kamakura, W., & Mazzon, J. (2013). Socioeconomic status and consumption in an emerging economy. *International Journal of Research in Marketing*, 30(1):4-18.
14. Kerbo, H. (2017). Social Stratification. In *The Wiley-Blackwell Encyclopedia of Social Theory*, B.S. Turner (Ed.). <https://doi.org/10.1002/9781118430873.est0761>
15. Krijnen, W. P. (2006). Convergence of the sequence of parameters generated by alternating least squares algorithms. *Computational statistics & data analysis*, 51(2), 481-489.

16. Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29(2), 115-129.
17. Kruskal, J. B., & Shepard, R. N. (1974). A nonmetric variety of linear factor analysis. *Psychometrika*, 39(2), 123-157.
18. Kuroda, M., Mori, Y., Masaya, I., & Sakakihara, M. (2013). Alternating least squares in nonlinear principal components. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6), 456-464.
19. Linting, M., Meulman, J. J., Groenen, P. J., & van der Kooij, A. J. (2007). Nonlinear principal components analysis: introduction and application. *Psychological methods*, 12(3), 336.
20. Linting, M., Meulman, J. J., Groenen, P. J., & Van der Kooij, A. J. (2007). Stability of nonlinear principal components analysis: An empirical study using the balanced bootstrap. *Psychological methods*, 12(3), 359.
21. Markus, M. T. (1994). Bootstrap confidence regions in nonlinear multivariate analysis, 28. DSWO Press, Leiden University.
22. McLeod J.D., Nonnemaker J.M. (1999). Social Stratification and Inequality. In: Aneshensel C.S., Phelan J.C. (eds) *Handbook of the Sociology of Mental Health*. Springer, Boston, MA. [https://doi.org/10.1007/0-387-36223-1\\_16](https://doi.org/10.1007/0-387-36223-1_16)
23. Meulman, J. J., Heiser, W. J. (1999). *SPSS Categories 10.0*. Chicago: SPSS incorporated.
24. Mori, Y., Kuroda, M., & Makino, N. (2016). *Nonlinear principal component analysis and its applications*. New York, USA: Springer.
25. Rodrigues, M. (2012) Estratificação social na teoria de Max Weber: Considerações em torno do tema. *Revista Iluminart*, 1(9).
26. Rodríguez, L.; Maza, O.; Macías, J.; Ortiz, D. (2020). Analysis of inequality via social stratification. *Entreciencias: Diálogos en la Sociedad del*



*Conocimiento*, 8(22). <https://doi.org/10.22201/enesl.20078064e.2020.2276859e22.76859>

27. Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, 3(2), 287–315.
28. Silva, G. (1981). Critérios de Estratificação Social. *Revista de Saúde Pública*, 15, pages: 38 – 45.
29. Tang, Y. (2017). Social Class Index: A New Measurement for Gender Social Stratification. *Sociology and Anthropology*, 5, 388–398.
30. Taylor, E., Stewart M., Hardner, J. (2007). Estimación de la importancia del turismo y la pesca en la economía de Galápagos. En Galápagos: migración, economía, cultura, conflicto y acuerdos, compilado por Pablo Ospina y Cecilia Falconí, 115–130. Quito: Corporación editora Nacional, Universidad Andina Simón Bolívar y Programa Naciones Unidas para el Desarrollo.
31. Wu, X. (2019), Inequality and Social Stratification in Postsocialist China, *Annual Review of Sociology*, 45 (1), 363–382.
32. Young, F. (1981) Quantitative analysis of qualitative data. *Psychometrika*, 46(4):357–388.
33. Wright EO (2005) Approaches to Class Analysis. Cambridge: Cambridge University Press.
34. Zhou, X; Wodtke, G. (2019). Income Stratification among Occupational Classes in the United States, *Social Forces*, 97(3), 945–972.