Multiple correspondence analysis method for evaluating the standard of living. Case study Albania

Evgjeni Xhafaj
Aleksander Moisiu University, genaxhafa@yahoo.com

Daniela Halidini
Aleksander Moisiu University, daniela_qendraj@hotmail.com

Neime Gjikaj
Aleksander Moisiu University, gjika.memi7@gmail.com

Multiple correspondence analysis method for evaluating the standard of living. Case study Albania

Evgjeni Xhafaj[1], Daniela Halidini [2], Neime Gjikaj[3]

[1,2,3,] Department of Mathematics, Faculty of Information Technology, University "Aleksandër Moisiu" Durrës, Albania.
genaxhafa@yahoo.com, daniela_qendraj@hotmail.com
gjika.memi7@gmail.com

Abstract. The main purpose of this study is to see the implementation of the multiply statistical method MCA (Multiple Correspondence Analysis) in large databases. Objectives of this study are to find the connections between qualitative variables categorized, concretely in our paper the connections between socioeconomic indicators and the standard of living. Otherwise this is called the concept of the multidimensional poverty. Different variables are being extracted like indicators that have a connection with the measuring of the multidimensional poverty. The data were obtained from the survey of measuring the life standard (LSMS) in 2012, which includes in the study about 6671 household, that are representative for the entire country. MCA results reflect several dimensions, where the first dimension, which is the more important, is called the welfare dimension. We represent the different discrimination measure of this dimension in connection with all the variables, where it is noted that from all the variables the biggest discrimination measure has the educational attainment of the household head. So the latter is the variable that has the greatest importance in relation to the welfare dimension. The results of this study help in the orientation of social policy to increase the standard of living, where the main role in this has the educational level.

Keywords: Multiple Correspondence Analysis, LSMS, welfare, educational level.

1. Introduction

Measuring the standard of living has historically been problematic because of the difficulty of defining an aggregate measure that captures the notion of welfare [1]. The evaluation of economic status methods are basically: the first is based in the incomes or the expenditure of consumption and the second, that's not based in the mentioned elements, is called the non monetary poverty. According to INSTAT [2], the non monetary poverty is composed of some indicators that have no relation with the monetary aspect. Non-monetary poverty includes indicators that are not related to monetary terms but with access to basic services such as water supply, sanitation infrastructure, power supply etc, as well as their quality. Often, to fulfill the monetary aspect of poverty with the non-monetary index is used the index of unmet basic needs (NBP). This index is comprised of five indicators located together and show: inadequate water supply and toilet (lack of running water and toilets with sewage in flat), inadequate housing conditions (as perceived by own family), inadequacy of power supply (outage for

6 hours or more per day), overcrowding of dwelling (3 or more persons per room) and the inadequacy of the level of education of the household head (with primary education or lower). Kabudula et al (2016) [3] show us three different indices for socioeconomic status (SES) which are constructed from asset-based indicators of households. One of these indices is based on the MCA method. From the results it turns out that the index, which is obtained from MCA, is such that the positive weights are assigned with the asset expected to be associated with higher SES, and negative weights are assigned to items expected to be associated eith lower SES. In Ndjanyou paper (2007) [4] is used MCA method which take into account 37 variables, which aims to identify the indicators that describe the real situation of poverty, where the first dimension is called the dimension of welfare. A positive weight to the variable modality in relation to this dimension contributes in growing the welfare, and a negative weight reduces the level of welfare. Based on the MCA method is obtain an index for each household. From the results is obtain a result equally the same with our paper: educational attainment of the household head has a great importance in explaining welfare dimension. Noglo (2014) [5] uses the MCA technique to build a composite poverty indicator. The results of the multidimensional poverty from the two approaches MCA and FGT showed that the households consisting of a large number of members living in rural areas , which are directed from the head with age greater than 51 years old, and who are less educated are the poorest. Njong et al (2008) [6] uses techniques PCA, MCA and fuzzy set to evaluate the multidimensional poverty index (Multidimensional Poverty). The results show that the index created on the basis of PCA dominates in relation to indexes created by MCA and fuzzy set. From the results it is concluded that the index created by the PCA shows a lower level of poverty than the indices generated by MCA and fuzzy set.

2. Methods

The data used to analyze the poverty are taken from the 2012 Living Standards Measurement Study (LSMS) for Albania. The survey covered both rural and urban populations. MCA is the generalization of correspondence analysis (CA) to more than the categorical variables. CA and MCA can be viewed as an adaptation to categorical data of principal component analysis (PCA), Jolliffe (2002) [7]. As PCA, MCA aims to identify a reduced set of synthetic dimensions maximizing the explained variability of the categorical data set in question. The advantages in using MCA to study associations of categorical data are then to obtain a simplified representation of the multiple associations characterizing attributes as well as to remove noise and redundancies in data. The exploratory and visualization based approach characterizing MCA provides immediate interpretation of the results. The indicators of asset ownership of all households are organized into a matrix **X** of ones and zeros called the ‘‘indicator matrix’’. In the indicator matrix, each categorical asset indicator is decomposed into a set of mutually exclusive and exhaustive binary categories that each take only the value 0 or 1 such that every household has a ‘1’ in exactly one of each asset’s set of categories and a “0” in the rest of the asset’s categories. Second, a matrix **S** is calculated by taking the X^2 metric on roe/column profiles of **X.** Greenacre (2007) [8] gives the calculation of S by the following formula:

$$S = D_r^{-\frac{1}{2}}(P - rc^t)D_c^{-\frac{1}{2}} \qquad (1)$$

Where **P** is the matrix formed by dividing each element of the matrix **X** by the sum of its elements, **r** is a vector whose elements are the sums of the row elements of the matrix **P**, **c** is a vector whose elements are the sums of the column elements of the matrix **P**, and **Dr** and **Dc** are diagonal matrices formed from **r** and **c** respectively. Singular value decomposition (SVD) is then performed on the matrix **S** to decompose it into three matrices such that:

$$S = UD_{\alpha}V^{t} \quad (2)$$

The columns of the matrices **U** and **V** referred to as left and right singular vectors, are respectively the eigenvectors of the matrices $\mathbf{SS^t}$ and $\mathbf{S^tS}$ and the columns of the diagonal matrix $\mathbf{D_\alpha}$ known as singular values. In the following table appear categorical variables that are used in this study. These variables are being considered having in view the concept of non-monetary poverty given by INSTAT for Albania.

1. Outside appearance of building: (Plastered, partially plastered, not plastered)
2. Condition of dwelling unit: (under construction, inappropriate for living, appropriate for living, very good condition)
3. Area of dwelling: (Less than 40 square meters, 40-69, 70-79, 100-130, more than 130 square meters)
4. Type of toilet : (wc outside without pipping, wc outside with pipping, wc outside the dwelling but inside building, wc inside the house, two or more wc inside)
5. Inadequate heating
6. Leaking roof
7. Damp walls, floor or basement
8. Father educational level : (Completed primary 4 years, Completed primary 8 years, Completed secondary, Completed university degree)

3. Results

The results of the Multiple Correspondence Analysis are listed in table 1 with the discrimination measure for each variable.

Table 1. Discrimination Measures

Variables	1- Dimensions
Outside appearance of building	,005
Condition of dwelling unit	,001
Area of dwelling	,009
Type of toilet	,002
Inadequate heating	,004
Leaking roof	,002
Damp walls, floor or basement	,003
Father educational level	7.248

Table 2. Results

Dimension	Total (eigenvalue)	Inertia
1	7.5	0.937

Table 1 and table 2 show the information of MCA solution: discrimination measures, eigenvalues and inertia. Discrimination measures as shown in table 1 indicate the relationship between (latent) dimensions and manifest variables. The first axis with higher explanatory power is the basis to explain the non-monetary welfare and for this reason, it is also named the welfare axis. Discriminating measures of variables are numbers corresponding to the variance of factorial scores of the variable. As the measurement of discrimination is a variance, it gives an account of the importance of the variable according to the phenomenon. Its values on the first dimension (axis) which is welfare dimension show that, when many indicators are taken together, those which permit the best to distinguish the poor from the non-poor. It is noted that the variable which have a high rate of discrimination is the education of the household head.

The main use of inertia is as an indicator of the number of axes to retain for further analysis. To define the number of dimensions to retain, the following criteria/considerations were employed: a) scree test b) eigenvalue (inclusion of MCA dimensions with inertia above 0.2 [9]. It is noted that in our case the inertia which has showed in Table 2 is 0.937. We have chosen only one dimension (axis) to interpret because it is difficult to interpret other dimensions.

Conclusion

The objective of this study was to investigate the links between qualitative variables, namely the links between socio-economic indicators and welfare. This aim is achieved by using the method MCA. This method aims to reduce dimensionality with the least possible loss of information, whereas the interpretation of the extracted dimensions (latent) is based on the categories of the variables that have been analysed. The relational structure between categories (of manifest variables) and the (latent) dimensions provides the information which category-clusters are explained by which latent dimensions. From our results dimension (latent) is interpreted as a dimension of welfare, and from all of the variables, the education of the household head has a higher discrimination measure. This variable has the greatest importance in explaining the welfare dimension. The results of this study help in the orientation of social policy to increase the standard of living, where the main role in this has the educational level.

References

1. Mazumdar, K. :Measuring the well-being of the developing countries. Achievement and Improvement indices, Social Indicators Research.(1999) 89-96
2. www.instat.gov.al
3. Chodzieadziea E. Kabudula et al.:Assessing Changes in Household Socioeconomic Status in Rural South Africa, 2001–2013: A Distributional Analysis Using Household Asset Indicators, Social Indicators Research, ISSN: 0303-8300 (Print) 1573-0921 (Online) (2016)
4. Ndjanyou, L.:Multidimensional Poverty in Cameroon: its determinants and spatial distribution, the African Economic Research Consortium (2007), 100-121
5. Noglo, Y.:Measuring and analyzing the non-monetary approach of multidimensional poverty by the basic needs in Togo, Social Science Research Network(2014), 36-55
6. Njong, A.: Characterizing weights in the measurement of multidimensional poverty: An application of data-driven approaches to Cameroonian data, OPHI Eorking Paper No. 21(2008)
7. Jollifee, T.: Principal Components Analysis Second Edition, ISBN 0-387-95442-2, Springer-Verlag New York (2002), 230-240
8. Greenacre, M.: Correspondence analysis in practice. Boca Raton: CRC Press (2007), 12-45
9. J. F. Hair, R. L. Tatham, R. E. Anderson, and E. Black.: Multivariate Data Analysis, Prentice-Hall, Upper Saddle River, NJ, USA, 5th edition. (1998), 34-78