Social segregation Profiles Based on *FANNY* e *GoM* methods: an application for Belo Horizonte, Brazil, 1996.

Ricardo Alexandrino Garcia
Adriana de Miranda-Ribeiro

1) Introduction

The purpose of this paper is to compare the results of the application of two different methodologies, *FANNY* – Fuzzy Analysis – and *GoM* – Grade of Membership, on the construction of a social segregation index. The principal difference between the methodologies is that, while FANNY is applicable to continuous data, GoM work whit discrete variables.

The same variables, from the Demographic Brazilian Census of 1996, were used in the application of both methodologies. The *census sectors* were the unit of analysis, because they are the small spatial unit to which data is available and also because, in Belo Horizonte, even in close areas, there are significant socioeconomic differences. The variables available are those related to: sex, age, education, family relationship, internal migration, place of residence and home type. Thus, the analyses were restricted to those variables.

An index can be understood as a synthetic measure, that reduces the dimensionality of a group of variables, but maintain the variability of the original information. In this paper, we searched for a measure that could best reproduce the social segregation definition. Two profiles were created, trying to represent the two socioeconomic opposite conditions what, in a general way, represent sectors with good and bad socioeconomic conditions.

The literature points out that social segregation is anchored in two factors: inequality and poverty. Besides that and the available variables, some aspects, that were direct or indirectly connected to the definition, could be performed. Other variables were incorporated to the database, to better characterize and differentiate the census sectors.

The variables were divided in three groups.[1]

Diagram 1: Social Segregation Index Variables

- SOCIAL SEGREGATION
  - Social Aspects
    - Domestic Servants (%)
    - Household head sex ratio
  - Demographic Aspects
    - Sector resident population (%)
    - Sex ratio
    - Migrant male population (%)
    - Migrant female population (%)
  - Educational Aspects
    - Population that frequents school, by age (%)
    - Years of study average, by sex, age and total

[1] The division does not alter the results, just serves as an exhibition instrument.

2) Methodology

The method *FANNY* that is rendered to the continuous data multidimensional modeling, has been developed from the theory of the hazy groups, by Zadeh (1965).[2] The method *GoM*, developed from the same theory of groups, is rendered to the of discreet data modeling. The great advantage in the use of such methodologies consists on the capacity of the analysis of heterogeneous data.

In a general way, the theory of hazy groups allows different elements to possess degrees of pertinence to several groups, in way to turn possible the mathematical representation of vague and imprecise concepts. The degree of pertinence of an element of an universe to a certain group or hazy partition, is represented by a real number in the [0,1] interval, which represents how true it is the statement that that element belongs to that partition. While in certain methods of cluster analysis (crisp sets), each analyzed element belongs to an only cluster, generating a clear division of the elements, the methods *FANNY* and *GoM* allow the estimation of the association degree between each element and the several clusters. In other words, the methods *FANNY* and *GoM* associate an object the several clusters.

In the method FANNY, for each object i and cluster v, there is an association, $u_{i,v}$, which indicates the degree of pertinence of the object to the cluster. The associations are defined by iterative processes, that minimize the function *f*:

$$f = \sum_{v=1}^{k} \frac{\sum_{i,j=1}^{n} u_{iv}^2 u_{jv}^2 d(i,j)}{2\sum_{j=1}^{n} u_{jv}^2} \qquad (1),$$

where *d(i, j)* represents the distances (ou dissimilarities) between the objects *i* e *j*; and $u_{i,v}$ is the unknown association of the objects *i* to the cluster *v*. The functions are subjected of the conditions:

$$u_{i,v} \mathbf{>= 0}, \text{ for i= 1, ..., n; v= 1, ..., k}$$

$$\sum_{v} u_{i,v} \mathbf{=1}, \text{ for i=1, ..., n ,}$$

which show that the associations between each object and the several clusters have to be necessarily have null or positive, and that the sum of the associations between each object and the several clusters is constant and equal to 1.

The method GoM estimates, by a model of multinomial probability, two types of parameters: one of association, g_{ik}, and another of structure, λ_{kjl}; in other words, the degrees of pertinence (g) of each element (i) to each profile, or type (k); and the probabilities of each category (l) of each variable (j) in each profile (k) that define those profiles. According to Manton (1994), the model of multinomial probability is given by:

$$L_{(y)} = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{i=1}^{Lj}(\sum g_{ik}\lambda_{kjl})^{Yijl}$$

Where: Y_{ijl} represents the value (zero or unit) that each element (*i*) assumes in the variabel (*j*) category (*l*). For all *i* and *j* combinations, Y_{ijl} will always be equal to zero, except for *l*, that will be equal to the unit.

By the determination of the number of clusters, “ideal profiles” are created. Thus, for each cluster, each variable has an ideal value. The association degree of an object to a cluster will be

[2] ZADEH, L. A. (1965) Fuzzy Sets, **Inf. Control 8**, 338-353.

closer to 1, as much closer as the values of its variables are to the established values for the " ideal " profile. It's possible to happen, but not necessary, that, among the objects, there are pure types or, in other words, objects that totally belong to an only cluster.

3) Results

The results show that the two profile present opposite characteristics in all variables, what mean the capacity to represent the desired opposites. The association degree or degree of pertinence to the bad conditions profile determine the Social Segregation Index (SSI).

Diagram 2: Social Segregation Index characterization

SOCIAL SEGREGATION INDEX

- Higher household sex ratio
- Less porcentage of household servants
- Higher sex ratio
- Less porcentage of female and male migrants
- Less percentage of people at school for all age groups
- Less years of study average for all age groups and sex
- Higher population density

For the analysis of the results obtained for SSI - SSI_{FANNY} and SSI_{GoM} - two strategies were adopted: the first was a comparison of the absolute values of SSI_{FANNY} and SSI_{GoM} - although it is known that the absolute values should not be considered, without care; the second was the spatial distribution analysis. Although in the same direction, the results were quite differentiated.

In absolute terms, SSI_{GoM} presented many sectors with extreme values, in other words, with degrees of pertinence equal to zero or one, what indicates that the section totally belongs to one or other profile. In other words, it can be considered as a "pure" type. From the whole sectors, 809 sections were considered totally of 'high level', while 580 were considered totally of 'low level'.

The results of SSI_{FANNY} showed that no sector was considered as a "pure" type. The values varied from 0,1774 to 0,7823, what indicates that the higher segregated sector has 78% of the characteristics of the bad conditions profile and the less segregated has 78% of the characteristics of the good conditions profile. These two sectors, in the SSI_{GoM} results, had the values 0 and 1, respectively, what does not differentiate them from the other sectors with the same result.

4) Final Considerations

The results indicate that both methodologies - *FANNY* e *GoM* – are powerful instruments of profile composition, for a great set of variables and heterogeneous objects. In developing countries, where the socioeconomic differences are intense, the exploitation of these tools can be fundamental, in terms of urban planning and application of public policies.

Although some differences were observed, the results go on the same direction, what validate them. In the case of the database used in this paper, the *FANNY* methodology was more adequate, as presented a higher differentiation of the degrees of pertinence of the 2108 census sectors of Belo Horizonte to the two profiles. It does not mean that *FANNY* is better than GoM, but can indicate that *FANNY* is more efficient for continuous data. By the other side, the results can be related to the fact that the original variables are continuous and not heterogeneous enough for a good categorization.