Third Meeting of the Expert Group
on Poverty Statistics
November 22-24, 1999
Lisbon, Portugal

INDEXES OF SOCIAL LAG IN THE COMMUNITIES OF MEXICO
AN APPLICATION

José Vences Rivera

Instituto Nacional de
Estadística, Geografía e
Informática, (INEGI)
México
e-mail: jvences@correoweb.com

1. Background

In the Threshold of the XXI century, the main challenge for Mexico is to decrease its poverty and to achieve a social justice, giving special attention to families living in extreme poverty, to excluded regions, to native communities, and to social groups with great disadvantages.

The unequal development of the country and its regions, the high levels of concentrated income, the manner that human settlements are distributed regarding the availability of natural resources - i. e. water-, as well as a high concentration and a high dissemination of the population, among other factors, have urged on the search for alternatives to alleviate the poverty conditions of millions of Mexicans, and at the same time to find better ways of development whit social justice.

Because of the above, the current social politics are planned through a serie of services and support programs that need the people from communities with a high social lag and exclusion, such as food, health, education, jobs, roads, sewerage, and other services as electrical and water systems.

It is important to mention that the census and survey information produced by the Instituto Nacional de Estadística, Geografía e Informática (INEGI), is widely used by Government Offices for the planning of politics about social aide.

2. Indexes of lag

On november 1996, the Instituto Nacional de Estadística, Geografía e Informática (INEGI) and the Government enterprise Distribuidora e Impulsora Comercial, S.A. (DICONSA) -which is responsible for the supply of food and basic products to the most lag sectors, managing a system of popular supply stores-, signed a cooperation agreement to develop a study about the social marginality and lag into rural localities and AGEBS*, through the construction of social lag indexes.

At the beginning of this survey, DICONSA had more than 20,000 small stores distributed along the Mexican territory. These establishments offered food and non-food products for general consumption at a cost under the cost of the private commerce; at the same time, DICONSA played a role as a price controller.

2.1. Objective

The main objective of this proyect was to establish criterions to re-distribute all the stores along the Mexican territory and to decide the openning of new popular supply stores where the population requires them the most.

---

* Group of 25-50 blocks perfectly defined by streets, avenues, alleyways or any other type of sing for identification.

Specifically, the study had the purpose of determining the degree of lag of the country's geographic units, grouping them into similar stratums with similar characteristics.

2.2. Requirements

Data from the 1990 Population and Housing General Census were used for this survey. This data was updated with the results of the 1995 Population and Housing Counting.

The selected variables were those that in a certain manner represents critical lacks related to life conditions of the population. Based on these variables, composed indexes were produced by means of the analysis of the robust principal components method. This method is described in issue 3 of this paper.

The construction of the indexes allowed to know the factors that define the deficiencies of the social-economic development, the unsatisfied basic social needs, and in general, levels of well-being the population living in the geographic units.

Index of Social Lag (ISL). This is a global index that allows to plan strategies and action lines to develop projects and social programs to improve the levels of well-being the population living with limited resources.

Index of Basic Lacks (IBL). It reflects the level of satisfaction of the most important basic needs of the population, allowing to elaborate policies to focus subsidies for basic goods.

Index of Infrastructure Lacks (IIL). It reflects the public services resources of the community. It is useful to plan social politics related to the requirements of social and physical infrastructure into the communities.

As an example, a classification table for the country's rural localities and urban areas with a population of more than 50 inhabitants, according to the 1990 Population Census data and classified according to the Social Lag Index results, is showed:

Table 1.

Stratum	**Number of rural localities** (50 and more inhabitants)	**Number of AGEB** (50 and more inhabitants)	**Social Lag**
1	9,150	9,173	Very low
2	14,110	10,218	Low
3	15,932	7,235	Medium
4	14,989	5,215	High
5	7,608	2,294	Very high
TOTAL	61,795	34,135	

From the total of rural localities, 36.7% were classified into the category of high and very high social lag, while into the urban areas (AGEB), 22% were under these conditions.

Impact of the survey:

On the basis of the performed survey, DICONSA carried out a purging and adjustment process into the national network of supply; reaching with this a reorientation of its development strategies, as well as to improve the network growth, focusing it fundamentally, to the country's rural zones. The plan is to move away, gradually, from the urban zones, because in there already exist other supply alternatives. In this way, DICONSA strenghts its presence into the weaker rural zones.

Table 2 is a comparative of the reorientation supply program impact taken from the performed survey.

Table 2.

Concept	**Before the Survey (1995)**		**Current Situation (1998)**	
Nomber of Stores:				
Total	21,614	(100.0%)	23,900	(100.0%)
Rural areas	20,553	(95.1%)	23,652	(99.0%)
Urban areas	1,061	(4.9%)	248	(1.0%)
Population being benefited (millons the inhabitants):	29.6	(100.0%)	29.5	(100.0%)
Total	26.5	(89.5%)	28.4	(96.3%)
Rural	3.1	(10.5%)	1.1	(3.7%)
Urban				

As showed in table 2, the number of stores increased in the 1995-1998 period; however the amount of beneficiaries was almost constant, mainly due to the relocation of stores from urban areas with less lag to the depressed rural areas in which a minor amount of inhabitants lives.

The reorientation of this program caused that beneficiaries from the rural zone were increased in absolute terms as much as in relative terms. In urban areas there was a decrease. Because of the above, the supply of the basic basket of food for general consumption, was strenghtened.

At the present, the Rural Supply Program is operating as follows:

- In 2,303 municipalities (95% of the national total), of which:
- 1,118 municipalities have a high and a very high lag (97% of the total in that situation)
- 1,048 have medium or low lag (93.7% of the total in that situation)

More than 30% of the indigenous pupulation is living into 803 of those municipalities, and there are 778 stores there. In total, exist 23,900 community stores which cover 71,232 rural localities, 33% in a direct manner and 6,790 in influence zones.

3. Description of the Methodology

This is a summary of the statistical procedures used to generate indexes for lag per locality and per group of blocks, as well as the stratification of geographic units. Also, the method to detect outliers is described. This method was useful to decide over the application of a robust procedure by which these indexes were obtained; it means, a method that was not affected by observations that became from the main structure of the data.

3.1. Constuction of indexes

The selected variables were taken as basis to construct the synthetic indexes that summarize in a great way the information contained into them in variance terms. To this purpose the multivariate statistical technique, known as *robust principal components*, were used. These are linear combinations of the original variables. There exist so many components (independent) as variables had been considered. Particularly, the first component is the one that absorbs more information from the group of variables. In the process of calculation of such components a complete weighting is given to the observations that probably are in the main structure of the data, and a reduced weighting is given to the outliers. The first "conventional" component was taken as initial estimation, to obtain -after a serie of iterations- the corresponding robustness; from this, the other robust components are produced in a progressive order.

In each case, the index produced by the first principal component take positive and negative values around its mean (it is zero when the matrix of correlations is used), it makes that the interpretation gets certain degree of abstraction, and not being too immediate. That was the reason to transform the index to a scale from 0 to 100, being able to express it in percentage terms and to locate the position in which an observation unit (OU) is found with respect to the others. The transformed index will take the minimum value of zero if there were any OU with minimum values in all variables. In the other extreme, this index takes the maximum value of one hundred when exists a OU with maximum values in all variables. It is important to mention that this transformation preserves the relative distances of the index original values.

3.2. Stratification of the units

A stratification of geographic units based on the mentioned lag indexes was made to form groups of homogenous units to easy the administration planning. For this case, the Centroide Method (mean in the univariated case) was used. It consists in minimize an objective function that guarantees the homogeneity of the units into each stratum. To make this method to function it is necessary to feed it with an initial classification, to obtain this classification the method of Dalenius and Hodges was used.

The definition of the number of stratums consisted in a serie of test, from 2 to 9 stratums; the one where the units into the stratums were

enough homogenous was chosen. In this case, the number 5 fulfilled this condition.

3.3. Outliers

In the present study, two procedures to detect multivariated outliers were used: a) Mahalanobis distance, which consists in weighting the observation distances at the mean by the matrix of covariances, and to compare it with the Chi-square distribution. In some cases, this method did not practically detected outliers, so it was decided to apply procedure b) Functions of Cambell influence, where the M-estimators are used. These could be considered as a modification of the classic estimators, then a complete weighting is assigned to the observations coming from the data main structure, and the influence of the observations located at the end of the contaminating distribution is reduced. In the final stage of the process, observations whose weighting is lower than one are extreme, and they become even more extreme when the weighting is moving closer to zero.

4. Final comment

In the past decade, the INEGI accumulated an infrastructure in human resources and informatic development that has allowed to expand options to generate basic statistics. Besides, through agreements with divers government agencies which are responsible of projects related to social policies, INEGI offers support on the application of methods that facilitate the study of the geography of the marginality and poverty,

In this sens, the study made jointly with DICONSA, has been as a detonator to make other applications under similar schemas.

5. References

Anderberg, M.R. (1973). **Cluster analysis for applications**. New York: Academic Press, Inc.

Campbell, N.A. (1980). **Robust Procedures in multivariate analysis**. I: Robust covariance estimation. Jour. of Ap. Statistics, V. 29, 231-237.

Hampbel, F. (1974). **The influence and its role in robust estimation**. Jour of de Am. Statistics Assoc. V .69, No. 346, 383-393.

Huber, P.J. (1981). **Robust Statistics**. New York. John Wiley & Sons.

Reyment, R.A., Blackith, R.E., and Campbell, N.A. (1984). **Multivariate Morphometry**. Second Ed. Academic Press, London.

Rodríguez, S. (1995). **Análisis de componentes principales robustos**. Tesis de Licenciatura, Facultad de Matemáticas, Universidad de Guanajuato.

Rosseeuw, P.J. y B.C. Van Zomeren (1990). **Unmasking multivariate Outliers and Leverage point**. Jour of the Am. statistical Assoc., V. 85, No. 411, 633-639.

Vences, J., Ramírez, I. y Flores, M.A. (1996). **Un sistema de análisis multivariado para grandes volúmenes de información**. Memoria del XI Foro Nacional de Estadística, 161-166, México.