MULTIVARIATE METHODS IN DEMOGRAPHY STUDIES: AN APPLICATION TO MUNICIPAL MIGRATORY BEHAVIOR FOR SETTLEMENT CATEGORY IN CUBA[1]

Ana D. Boquet, Instituto de Planificación Física, Cuba
Lydia de la C. Lera y Minerva Montero, Instituto de Cibernética, Matemática y Física, Cuba

ABSTRACT
In this paper a study of the municipal migrations by means of Principal Component Analysis and STATIS is presented. The data used for this purpose are migratory rates calculated at the Office of Statistics of Cuba. The rates are calculated for three periods (1989-1991, 1992-1994 and 1995-1996) for each settlement category (municipality center*, small urban and small rural settlements) and for two kinds of movements: the migrations within the province and into other provinces of the country. Due to not all municipalities have settlements of all categories, we use a multivariate multilevel model for estimates efficiently the variance and covariance matrix of a vector of multivariate response with missing data.

Key words: Principal component, multilevel model, common structure, STATIS.

RESUMEN
En el presente trabajo se hace un estudio de las migraciones a nivel municipal, a través del empleo del Análisis de Componentes Principales y el método STATIS. Se utilizó como fuente las tasas migratorias de saldo, calculadas a partir de las bases de datos de la Oficina Nacional de Estadística. Las tasas se calcularon por trienios (1989-1991, 1992-1994 y 1995-1996), para cada categoría de asentamiento (cabecera municipal, urbano de base y rural de base) y para dos tipos de movimientos, las migraciones dentro de la provincia y hacia otras provincias del país. Debido a que no todos los municipios tienen asentamientos de todas las categorías, se utilizó un modelo multinivel multivariado para poder estimar eficientemente la matriz de varianzas y covarianzas de un vector de respuesta multivariada con datos perdidos.

1. INTRODUCTION

Migratory movements of population are of interest for a wide range of specialist, because its sensibility to socio-economics changes in the territories. Our country has maintained a systematic study on the theme for the last forty years, with data obtain from different sources. Present work is based on the individual reports of changes of address to the Office of Identity Card and Registry of the Population, that they deliver to the National Office of Statistics (ONE), where the permanent changes of residence are selected to elaborate the national statistics upon the theme.

In our research we worked with internal migration movements from 1989 to 1996, when 2,378.6 thousands persons changed permanently their municipality of residence, of them, 38% continued living in the same province, resting 62% also changed province of residence. Additionally, other 695.1 thousands of people changed residence in Havana City, but those persons don't count as migrants because of the territorial compactness of the city (Boquet, 1997).

At the beginning of the decade of the 90's the most complex moment in the history of Cuba as independent nation occurred due to the loss of 85% of its market, more of the half of the petroleum supplies and more than 70% of the importation. It suffices to say as an indicator of the situation confronted that the Gross Domestic Product decreased in almost 35% between 1989 and 1994.

In the period 1989 to 1996 we can differentiate three stages: 1989 to 1991, the beginning of the mentioned crisis; 1992 to 1994 the worst years of the crisis; and 1995 to 1996, where the first signals of recovery are observed. The researchers on internal migrations were interested in knowing the repercussion of the crisis phenomenon on the internal migratory movement of the population in the country.

[1]Este trabajo fue presentado en el X Encuentro de Estadística Cuba-México.
*Municipality center is the town in which is located municipality government.

For this research we used this information divided into the three stages explained above for all municipalities of the country (except Havana City). Havana City was omitted because it is the capital of the country and the first receiver of migrants. Three kinds of settlement categories were differentiated: municipality centers (MC), small urban settlements (SUS) (without political-administrative category) and small rural settlements (SRS). We also analyzed two kinds of movements, the migrations within the province (intra-province) (I) and the migrations into other provinces of the country (inter-province) (E).

The selected indicator to measure the movements was the migratory rate (MR) of balance in each one of the periods indicated and for each zone regarding each movement, so that for each phase we have 6 variables and 154 observations (the municipalities). With these ones we conformed the three data matrices. The six studied variables are:

- MR.I.MC: migratory rate intra-province of the municipality center.
- MR.I.SUS: migratory rate intra-province of the urban settlement.
- MR.I.SRS: migratory rate intra-province of the rural settlement.
- MR.E.MC: migratory rate inter.-province of the municipality center.
- MR.E.SUS: migratory rate inter- province of the urban settlement.
- MR.E.SRS: migratory rate inter- province of the rural settlement.

2. MATERIAL AND METHODS

In order to determine if the singularity of the municipalities is due to an isolated phenomenon or a different behavior, we use a procedure that permits to classify the territories into typical or atypical upon the basis of the multivariate information data. We propose a Principal Component Analysis (PCA) for each evaluated period where the elements of each matrix represent the 6 different migratory rates for the 154 municipalities.

Due to the geographical characteristics of certain municipalities, some of the values of their variables are "impossible". Then each row of the matrix can be seen as a response multivariate vector with missing values. As the PCA is based upon particular functions of the variance-covariance matrix, it is difficult to estimate this matrix with missing data. The majority of the statistics packages ignores the observation with a missing response or, in the best case, uses special proceedings to manage missing data. However, it is possible to estimate efficiently the variance-covariance matrix of a multivariate response vector with missing data using a multivariate multilevel model (Goldstein, 1995).

We consider the model as follow: the municipalities are the level 2 units and the measurements of each municipality are the level 1 unit. Each observation measured in the level 1 has a response that can be the migratory rate. The basic explanatory variables are a set of dummy variables that indicate which response variable is present.

The level 1 model is given by:

$$y_{ij} = \beta_{1i} z_{1ij} + \beta_{2i} z_{2ij} + \beta_{3i} z_{3ij} + \beta_{4i} z_{4ij} + \beta_{5i} z_{5ij} + \beta_{6i} z_{6ij}, \quad (1)$$

where:

$$z_{1ij} = \begin{cases} 1 & \text{if MR.I.MC} \\ 0 & \text{otherwise} \end{cases} \qquad z_{2ij} = \begin{cases} 1 & \text{if MR.I.SUS} \\ 0 & \text{otherwise} \end{cases} \qquad \ldots\ldots\ldots\ldots \qquad z_{6ij} = \begin{cases} 1 & \text{if MR.E.SRS} \\ 0 & \text{otherwise} \end{cases}$$

The level 2 models, for the random coefficients of the 1-level are given by:

$$\begin{aligned} \beta_{1i} &= \gamma_1 + u_{1i} \\ \beta_{2i} &= \gamma_2 + u_{2i} \\ \beta_{3i} &= \gamma_3 + u_{3i} \\ \beta_{4i} &= \gamma_4 + u_{4i} \\ \beta_{5i} &= \gamma_5 + u_{5i} \end{aligned} \quad (2)$$

$$\beta_{6i} = \gamma_6 + u_{6i}$$

Substituting (1) in (2) yields the combined model:

$$y_{ij} = \gamma_1 z_{1ij} + \gamma_2 z_{2ij} + \gamma_3 z_{3ij} + \gamma_4 z_{4ij} + \gamma_5 z_{5ij} + \gamma_6 z_{6ij} + u_{1i} z_{1ij} + u_{2i} z_{2ij} + u_{3i} z_{3ij} + u_{4i} z_{4ij} + u_{5i} z_{5ij} + u_{6i} z_{6ij}$$

The variance-covariance matrix of the errors u_i is:

$$\Omega = \begin{bmatrix} \sigma^2_{u_{11}} & \sigma_{u_{12}} & \cdots & \sigma_{u_{16}} \\ \sigma_{u_{21}} & \sigma^2_{u_{22}} & \cdots & \sigma_{u_{26}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u_{61}} & \sigma_{u_{u_{62}}} & & \sigma^2_{u_{66}} \end{bmatrix}$$

Note that there is no level-1 variation specified, because level 1 exists solely to define the multivariate structure. The level 2 variances-covariances are the (residuals) between territories variances. When there are several missing answers, the multivariate model can be advantageous, because the individuals with missing data are not ignored because the multivariate procedure uses the available information of the answers instead of ignoring the individual with a lost answer, the multivariate procedure uses of the available information of the answers that are presented for this individual Therefore, the multivariate model can be very useful for techniques like PCA when there are missing observations, because the variance-covariance (or correlation) matrix Ω would be the residual matrix after regressing on explanatory variables. In forming the derived variable (as the "scores") we would use the level 2 estimates of the observed variables, i.e., level 2 residual: $\hat{u}_{1i}, \hat{u}_{2i}, ..., \hat{u}_{6i}$, and not the observed variables.

The classical PCA does not permit to do a joint analysis of the periods, for this reason we used STATIS (Structuration des Tableaux A Trois Indices de la Statistique) (Lavit, 1988; GERI, 1996) method in order to analyze the global evolution of the migration in the three periods. STATIS is an exploratory method of Data Analysis that permits to study simultaneously several data groups. It is used when the data are quantitative and the main objective is to search a common structure among k groups. These groups can be the same observation measures in the time or different observations with the same variables. The method permits graphical representations of the evolution of the phenomenon under study

3. RESULTS AND DISCUSSION

Tables 1, 2 and 3 show the principal components for the data of the three periods. As the variables are measures in the same scale we used the variance-covariance matrix.

Table 1. Principal Components in the period 1989-1991.

	Factor 1	Factor 2	Factor 3
MR.I.MC	**-.799613**	.395685	.224433
MR.I.SUS	.071529	-.427767	**.641575**
MR.I.SRS	.040568	.611911	-.436746
MR.E.MC	**-.789045**	.265831	.369406
MR.E.SUS	-.516582	-.498141	-.400118
MR.E.SRS	**.645913**	.425121	.365483
Expl. Var	1.952798	1.213526	1.082868
Prp.Total	.325466	.202254	.180478

Table 2. Principal Components in the period 1992-1994.

	Factor 1	Factor 2	Factor 3
MR.I.MC	**.808825**	-.216785	-.354435
MR.I.SUS	.178091	**.817819**	.258033
MR.I.SRS	-.501784	-.595418	-.048830
MR.E.MC	**.819949**	.007266	-.391620
MR.E.SUS	.387142	-.193726	**.740516**
MR.E.SRS	-.542065	.369025	-.462386
Expl. Var	2.053731	1.244109	1.110120
Prp. Total	.342289	.207351	.185020

Table 3. Principal Components in the period 1995-1996.

	Factor 1	Factor 2	Factor 3
MR.I.MC	-.127284	-.851257	-.067254
MR.I.SUS	.817728	-.266003	.413445
MR.I.SRS	.678643	.194167	-.161954
MR.E.MC	.129955	-.739894	-.516940
MR.E.SUS	.790507	-.187141	.198919
MR.E.SRS	.526535	.420646	-.620669
Expl. Var	2.064466	1.592506	.893715
Prp. Total	.344078	.265418	.148952

The two first principal components together explain 52.7 %, 54.9 % and 60.9 %, respectively, of the total variance. In Tables 1 and 2 we observe that in the first component the migratory situation of the municipal center, in the intra-province and in the inter-province movements, is the most noticeable element. For the second component, in Table 1 the movement of the small rural settlements (intra-province and inter-province) is more important, always with sign opposite to the urban and the rural ones, i.e., in some territories the migrants arrive at urban settlements and in others to the rural ones, being somewhat more important the arrival to the rural settlements. In the second period (Table 2), associated to the worst moment of the crisis, the movement among provinces is not significant for the settlements of this zone, having greater importance the situation of the urban settlements. In the last period (Table 3), associated to the post crisis, the migratory situation in the zone appears important in the first factor, now with the same sign to the urban and to the rural settlements, i.e., the territories receiving migrants in this zone do it so much in urban settlements as rural ones, while the emitting of this zone are also in both categories of settlements. The movements to municipal centers are important only in the second factor; it shows a change in the migratory behavior of the territories in relation with the two first periods. In these places occur the majority of movements but they are emitting to the same places generally; the province capitals and Havana City present a greater homogeneity in the migratory behavior.

Figures 1, 2 and 3 show the distribution of the municipalities in the Cartesian plane of the two first principal components, we can see that the graph corresponding to the second period is a rotation and contraction from the first one, and how upon initiating the recuperation occurs a true change, when many municipalities move away of the values of equilibrium. By means of these graphs it is possible to classify the municipalities according to its migratory behavior. Thus, it is identified with the municipalities that are found around the origin as the municipalities in migratory equilibrium for all their zones and in the extremes are located those in which predominate one or other movement.

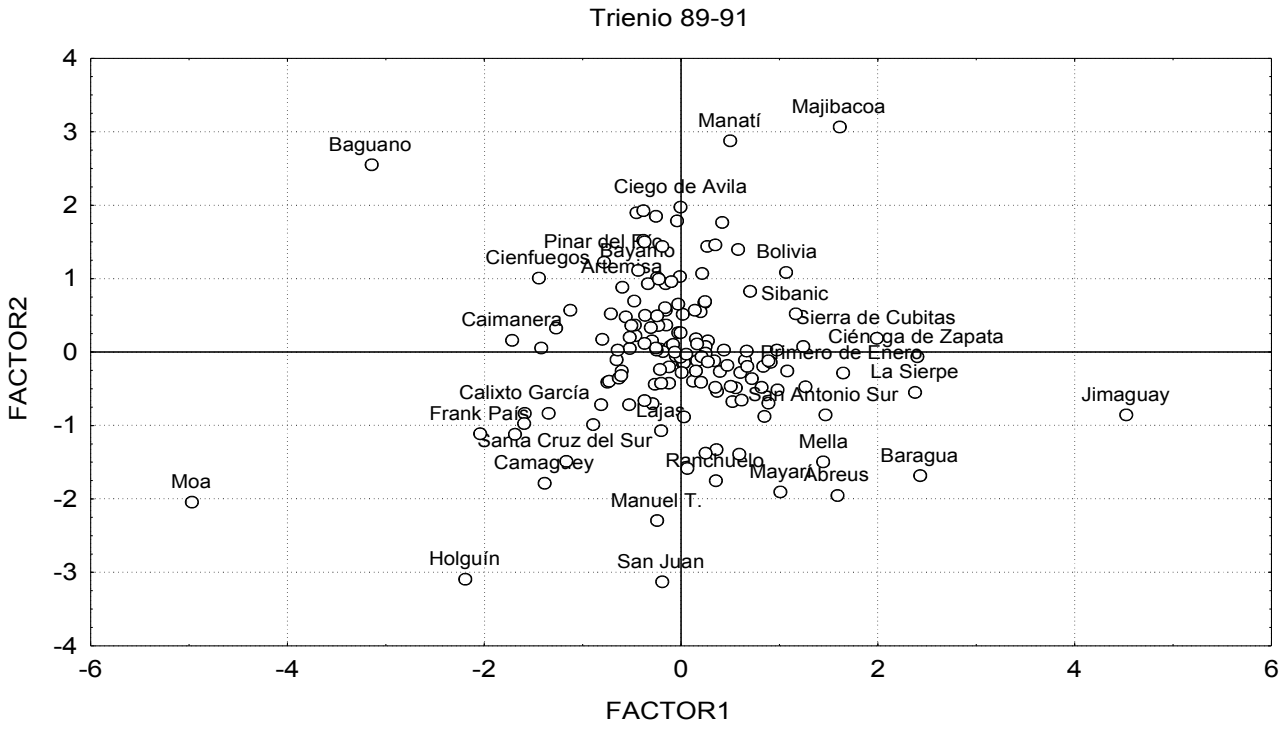


Figure 1. PCA of the period 89-91.

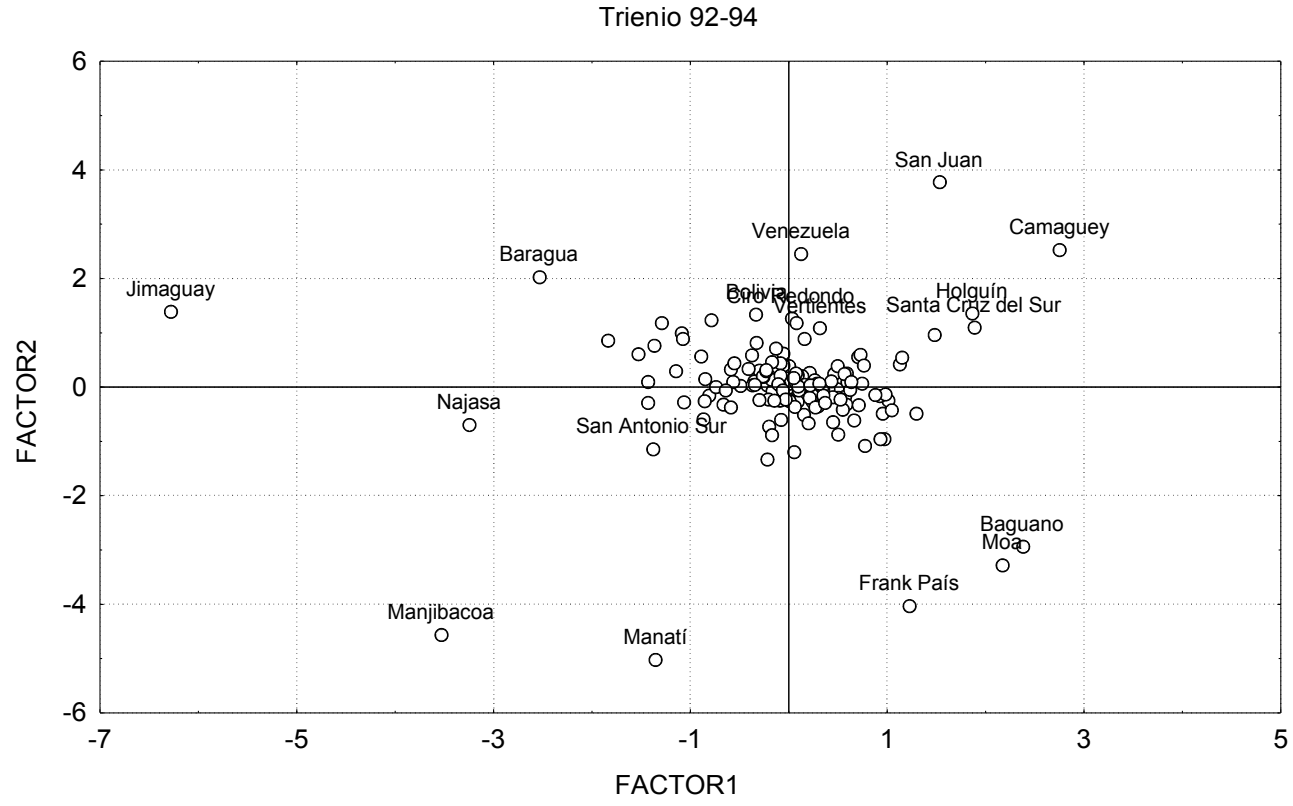


Figure 2. PCA of the period 92-94.

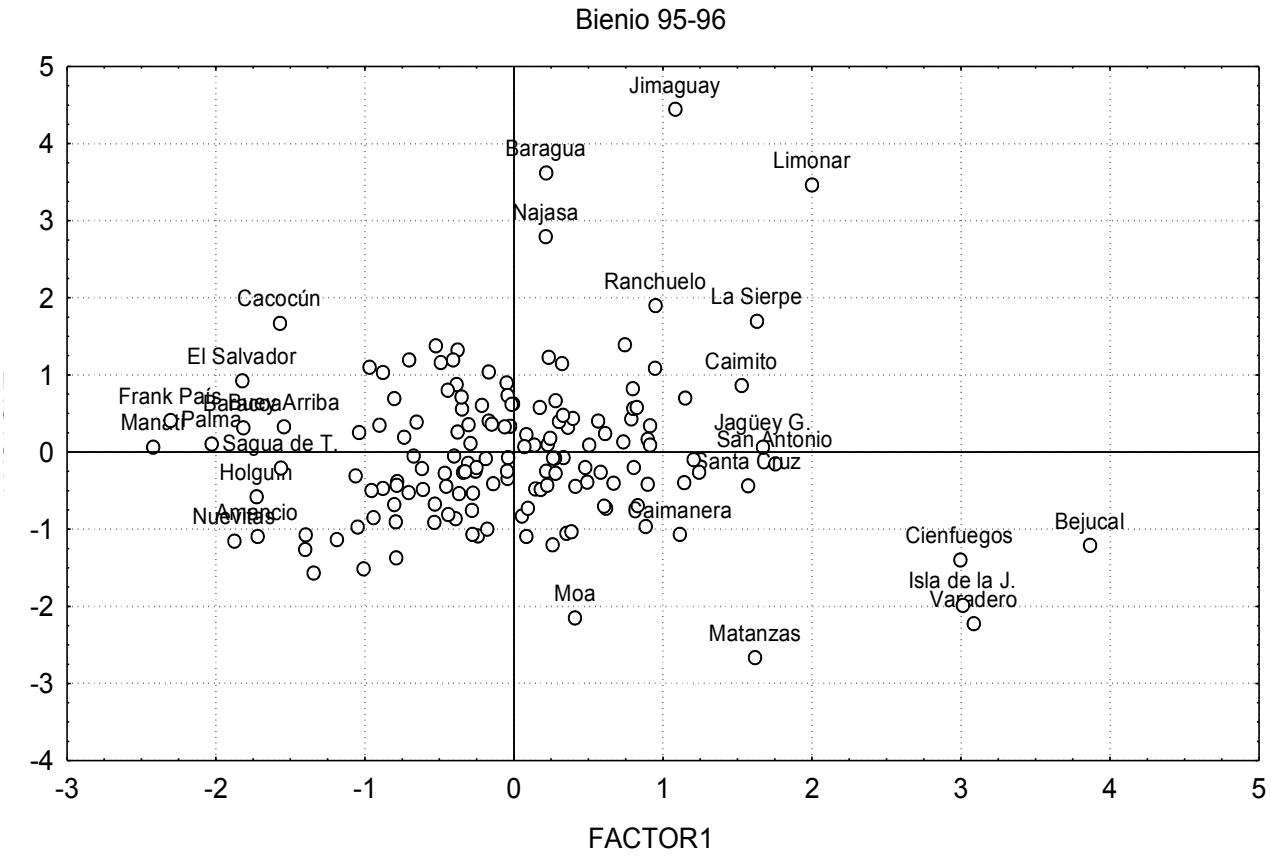
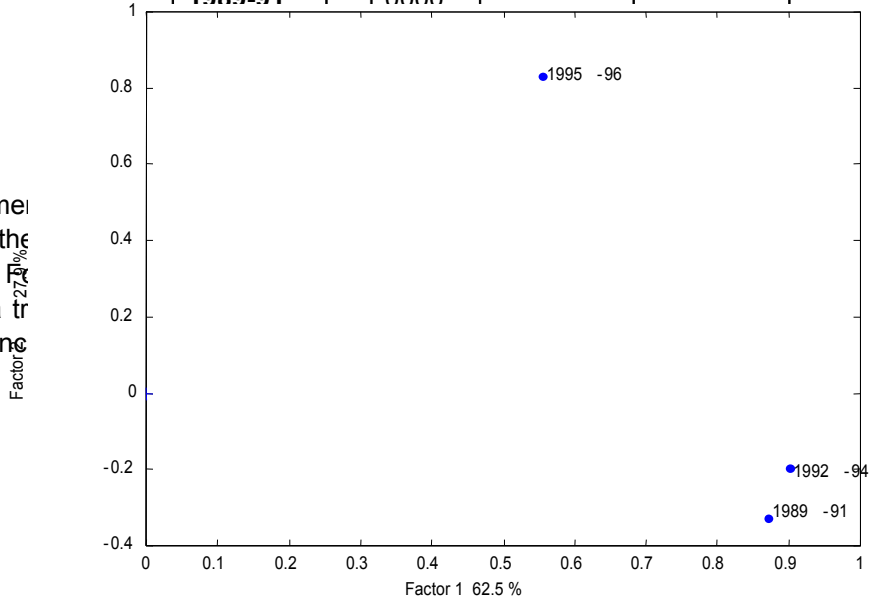


Figure 3. PCA of the period 1995-96.

Table 4 shows the RV coefficients of Rayleigh, the norm scale products among the municipalities grouped in the different periods. It is observed that the coefficient for the two first periods is relatively large and that decreases when the time increases. This signifies that the two first periods are relatively stable regarding its structure, corroborating the results observed in the previous graphics, where it is emphasized, as we have said, the mobility to the municipality centers.

Table 4. RV Matrix.

	1989-91	1992-94	1995-96
1989-91	1.0000		



As compleme ds, it shows a separation of the nicipalities are more stables. F and 1992-1994 form part of a tr a break of the existing tendenc

Figure 4. Euclidean Image of the Tables.

Figure 5 is a representation of the municipalities in the compromise and Figure 6 shows some significant trajectories associated to places characterized by a different phenomenon that took place in the analyzed periods

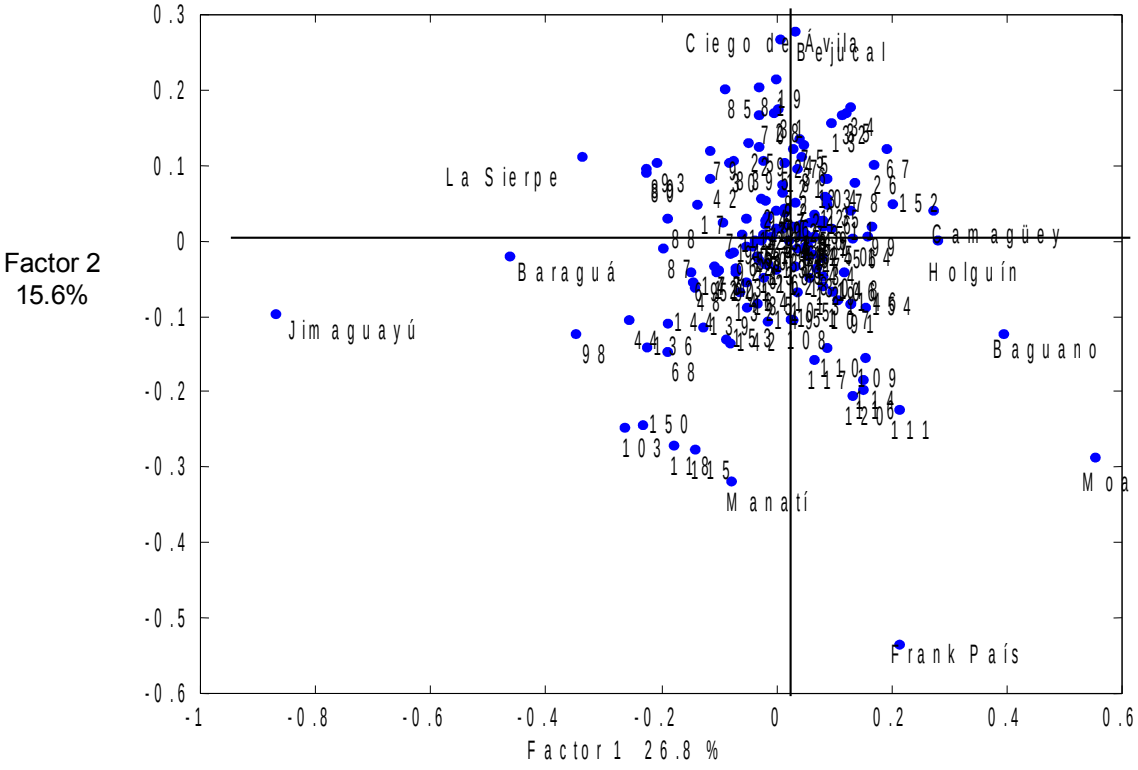


Figure 5. Euclidean Image of the compromise (municipalities).

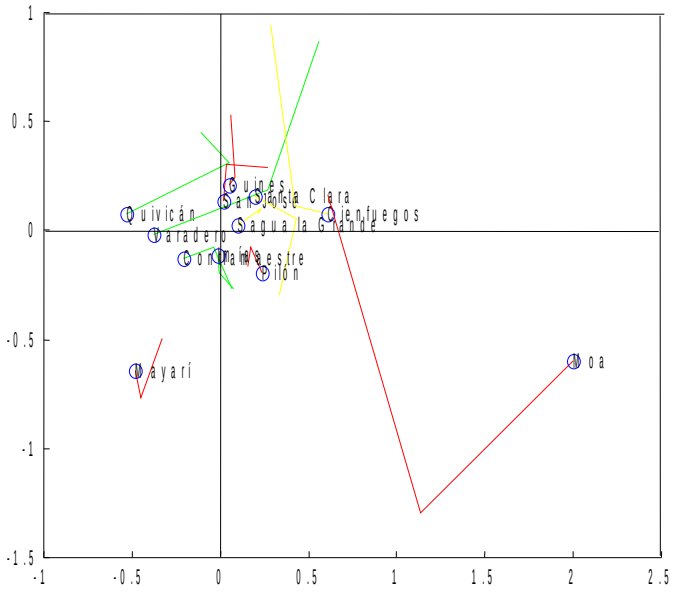


Figure 6. Some important trajectories.

Analyzing the correlations among the variables with the axes (Table 5) we can see that the variables associated to the migrations in the municipal centers and rural migrations inside of the provinces are more correlated with axis 1 and explains 26.8 % of the variance.

Table 5. Correlations of the variables with the axes in the compromise.

Period 1989-91

	Axis 1	Axis 2
MR.I.MC	**0.7901**	0.1266
MR.I.SUS	-0.0280	**0.4596**
MR.I.SRS	-0.0679	-0.0236
MR.E.MC	**0.7996**	0.3619
MR.E.SUS	0.4241	-0.3618
MR.E.SRS	**-0.5507**	0.6620

Period 1992-94

	Axis 1	Axis 2
MR.I.MC	**0.8214**	0.2115
MR.I.SUS	0.0157	**0.5212**
MR.I.SRS	-0.2631	-0.1771
MR.E.MC	**0.7958**	0.4185
MR.E.SUS	0.3835	-0.2805
MR.E.SRS	**-0.6133**	0.6333

Period 1995-96

	Axis 1	Axis 2
MR.I.MC	0.5653	0.1152
MR.I.SUS	0.1660	0.4545
MR.I.SRS	-0.0193	0.2102
MR.E.MC	0.6217	0.3850
MR.E.SUS	-0.0239	0.4189
MR.E.SRS	-0.4089	0.5773

4. CONCLUSIONS

The use of these methods permits the analysis of the migratory behavior in the set of municipalities of the country, because it was possible to incorporate municipalities without settlements in recognized zones for the System of Settlement of Populations.

The combined use of these methods allowed the description of the studied phenomenon on each stage in order to identify the real nature of the migrations and their evolution in the analyzed periods.

The results of this paper show the actual influence of national economy phenomenon on the migratory behavior of the country. The migratory behavior in the municipal centers is homogeneous, while there are different migratory situations among the urban settlements of base. That is why the latter occupies a first plane of attention because of its importance for the study of tendencies of the population migratory behavior.

ACKNOWLEDGMENT

We acknowledge to Dr. Jesús E. Sánchez for his help in reviewing the manuscript.

REFERENCES

BOQUET, A.D. (1997): **Estudio descriptivo de las migraciones internas en Cuba de 1989 a 1996**. Informe Técnico del Instituto de Planificación Física.

GROUPE GERI (1996): **L'analyse des donnés évolutives**. Methods et applications. Editions TECHNIP.

GOLDSTEIN, H. (1995): **Multilevel Statistical Models**. (2nd edition). New York, Halsted Press.

LAVIT, CH. (1988**): Analyse Conjointe de Tableaux Quantitatifs**. Masson, Paris.

MARDIA, K.V.; J.T. KENT and J.M. BIBBY (1979): **Multivariate Analysis**, Academic Press, London.