Statistical Classificatory Analysis Applied to Zonification

M^A BEATRIZ BERNÁBE LORANCA
Cuerpo Académico de Ingeniería de Software -Aplicaciones Estadísticas.
Facultad de Ciencias de la Computación, BUAP.
Calle 14 sur y Avenida San Claudio, Colonia San Manuel. Fax 015 222 2295500 ext. 7225
PUEBLA, MÉXICO

RAMIRO LÓPEZ SALES
Secretaría de Transportes y Vialidad del Distrito Federal
Álvaro Obregón 269, 9º. Piso, Colonia Roma CP 06700. Tel (55) 5209 9913 ext. 1114
MEXICO D.F

Abstract:- An important activity within the analysis of a city's dynamics is its division in zones. The purpose of a division in zones is to define areas with common characteristics (socioeconomic, land use, connection by roads, etc.) in order to simplify the analyses. Each discipline that studies the urban communities has its own criteria to describe the corresponding zones. In this paper we apply Principal Components Analysis (PCA) and Cluster Analysis to population data in order to obtain a classification of zones and at the same time to obtain an auxiliary tool for the definition of zones.
The study area is the Metropolitan Zone of Mexico's Valley (MZMV) and the data comes from the XII Population Census[8]. The unit observation used is the Basic Geo-statistical Area (BGSA) which is the one used defined by the National Institute of Geography, Statistics and Informatics of Mexico (INEGI).
The objective of this study is to use the resulting classification and variable groupings to validate, update or correct previously defined zones used for transportation planning and to provide additional means to analyze the transportation in the MZMV. Additionally, the results may prove useful for other disciplines that use population data to analyze the city's behavior.

Key-Words:- Zonification, Cluster Analysis, Principal Components Analysis, Population.

1 Introduction

When studying the population phenomenon it is difficult to properly identify all the indicators that characterize it. For research purposes, these indicators are generally translated into variables, which must be analyzed and selected considering the researcher's specific objectives. In this study the 170 variables from the XII Census were considered[8]. The geographic coverage is Mexico City and its Metropolitan Area: the data from the census refers to 4,974 census tracts that include the Federal District and 58 municipalities from the neighboring Mexico State, as shown in Figure 1.

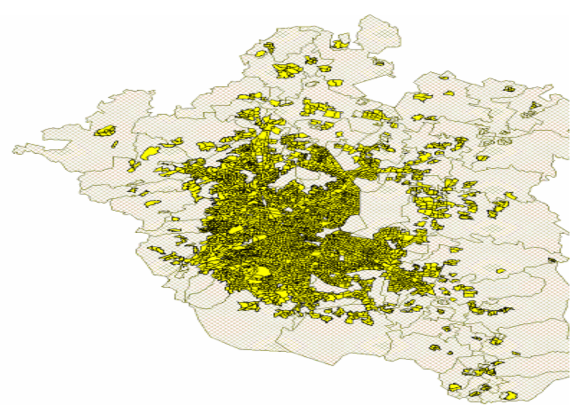

Fig. 1. Mexico City and its Metropolitan Area

Given the number of variables and the information they represent, a multivariate statistical approach is needed to reduce the dimensionality of the problem. Once the number of variables has been reduced, the resulting variables will be used to categorize the zones. In order to achieve this objective we will resort to cluster analysis, a statistical pattern recognition technique.

The general outline of this work is as follows: Section 1 corresponds to the introduction, which offers a short description of the problem at hand, and the proposed solution. The data analysis (section 2) explains the mechanics of variable selection and the relations among them which allows us to formulate an initial hypothesis about dependence and association between them. To analyze the problem with less variables, section 3 deals with the single correspondence analysis, the results given by this technique. Finally, section 4 gives the details of cluster analysis and the classes obtained by this process. Section 5 presents conclusions and further research lines and the appendix contains the variable table.

2 Data Analysis

The initial problem when considering census variables is the high correlation existing between them. In order to reduce the amount of variables used, a selection process was performed to select 57 of the 170 census variables [8]

Three main criteria were used for variable selection: data availability, data representation and grouping by correlation coefficients. In the first case, six variables present very few information or do not apply to urban zones like Mexico City (z98, z99, z137, z138, z147, z162), so they were discarded.

As to data representation, forty eight variables were not considered since they divide a previous variable into masculine and feminine (z2, z3, z5, z6, z8, z9, z11, z12, z14, z15, z17, z18, z20, z21, z23, z24, z25, z27, z28, z30, z31, z33, z34, z36, z37, z39, z40, z41, z43, z44, z46, z47, z62, z63, z65, z66, z68, z69, z85, z86, z88, z89, z91, z92, z166, z167, z169, z170). Another two variables expressing totals were discarded, since they represent combinations of other, more descriptive variables (z1, z119). A change in the representation of age ranges was required, with eight new variables (x1 to x8) representing non-overlapping age intervals instead of the thirteen age variables with overlapping intervals presented in the census data (z4, z7, z10, z13, z16, z19, z22, z26, z29, z32, z35, z38, z42). Lastly, a group of twenty one variables was discarded that either were not considered representative of the population phenomena under study or were notoriously dominated by a single feature (religion, literacy) and thus exhibited a nearly identical variation to the general population (z58 to z61, z64, z93 to z97, z100, z119, z121 to z125, z132 to z135).

After these selection stages, a group of ninety variables was obtained. With these variables a correlation matrix was obtained and groups of variables were formed with a correlation coefficient above 0.98 between each member of the group. This last step discarded 33 variables, leaving 57 for further analysis. See table A.1 in the appendix for the list of selected variables.

After obtaining the definitive group of variables, we proceed to further analyze the correlation between them. An immediate observation is that the variables tend to be related in groups with similar correlation between each pair of members of the group, pair-wise only correlated variables also appeared, but are not as common.

A correlation coefficient of at least 0.85 was considered to form the following groups; each group was given a distinctive label, for notation purposes only:

- Non-correlated variables
  x0, x2, z48, z49, z53, z054, z055, z083, z103, z109, z114, z115, z136, z142, z150, z163, z164
- 12 to 18 year old population
  x005, x006, x034, z051, z056, z075, z104, z105, z106, z113, z116, z148, z149
- Privately owned houses
  z148, z149
- 60 years or older population
  x008, z145
- Population with at least high-school education
  z050, z080, z107, z112, z117, z127, z129, z131
- Population with incomplete primary school education
  z070, z071, z077, z116, z126
- Population with college education
  z078, z082, z118, z140, z157, z159
- Self-employed population
  x006, z110, z111, z051, z056, z075, z107

- One-room only houses
  z126, z128, z130, z141
- Computer-owning households
  z160, z161

3 Principal Components Analysis

The multivariate techniques are a set of statistical methods aimed to simultaneously analyze several variables on each individual or object of a given data set. Their advantage lies in the better understanding that can be obtained of the phenomena under study, which uni-variated and bi-variated statistical methods are unable to achieve.

Among the multivariate techniques the one we'll use, Principal Components Analysis (PCA) is known as a inter-dependency method that identifies metric variables. PCA is used with metric variables in order to expose dependency or association relations in terms of a smaller set of variables called factors (if observable) or principal components (if not observable). In general terms, this technique allows a reduction in the *dimensionality* of the problem by decomposing the correlation matrix into the product of three matrices: eigenvalues, eigenvectors and transposed eigenvectors. [1, 3]

PCA is a valuable tool for the construction of a statistical population model oriented mainly to the reduction of its dimensions. The main feature that describes the resulting dimensions (components) and their related variables allows us to link together variables with similar composition, common features or regularities. Based on previous experiences [2, 6], the resulting components provide insight into the composition of the data. In this case the PCA produced 6 factors (or components), as shown in table 1

Table 1. Factor loadings, PCA

	f1	f2	f3	f4	f5	f6
X000	-0.4	-0.1	-0.5	0.4	-0.1	-0
X002	-0.7	0.25	0.35	0.31	0.23	-0.2
X005	-0.9	0.23	0.08	0.09	0.12	-0
X006	-1	0.17	0	0.02	0.09	0.01
X008	-0.8	-0.3	-0.1	-0.2	-0.1	0.1
X034	-0.9	0.35	0.09	0.11	0.09	-0.1
Z045	-0.7	-0.4	-0.1	-0.3	-0.2	0.09
Z048	0.22	0.7	0.36	0.05	-0.3	-0
Z049	0.14	0.73	0.32	0.1	-0.3	-0
Z050	-0.9	-0.2	0.03	0.2	-0	0.01
Z051	-0.9	0.37	0.01	-0.1	0.08	-0
Z053	-0.8	-0	-0.2	-0.1	-0.1	0.1
Z054	-0.8	0.06	-0.2	-0.2	-0.1	0.25
Z055	-0.7	0.11	0.25	0.35	0.18	-0.3
Z056	-1	0.09	-0	-0	0.01	0.04
Z070	-0.7	0.58	-0.1	-0.1	0.07	-0.1
Z071	-0.8	0.54	-0.1	-0.1	0.06	-0
Z075	-0.9	0.29	-0.1	0.1	0.01	0.05
Z077	-0.9	0.35	-0.1	0.04	0.01	0.06
Z078	-0.8	-0.6	0.13	-0	-0	0.04
Z080	-0.9	-0.3	0.02	0.17	-0.1	0.12
Z082	-0.5	-0.8	0.21	-0.2	0.01	-0.1
Z083	-0	-0.9	-0.1	-0.1	0.2	-0.1
Z103	-0.4	0.04	0.11	0.02	-0.3	0.04
Z104	-0.9	-0	0.11	0.14	0.1	0.06
Z105	-0.9	0.26	0.03	0.07	0.06	0
Z106	-0.9	0.38	-0	0.07	0.11	-0
Z107	-1	-0.2	-0	-0	-0.1	0.01
Z109	-0.2	0.51	0.26	-0.4	0.27	0.24
Z110	-0.9	0.09	-0.1	-0	-0	0.11
Z111	-0.9	-0.1	-0	-0.1	-0	0.13
Z112	-0.9	-0.4	-0	-0	-0.1	0.08
Z113	-0.9	0.2	-0	0.09	0.01	-0
Z114	-0.6	0.19	0.06	-0.2	0.12	0.32
Z115	-0.7	0.28	-0.2	-0.1	0.01	0.23
Z116	-0.9	0.36	-0.1	0.01	0	-0
Z117	-0.9	-0	-0	0.17	-0.1	0.02
Z118	-0.5	-0.8	0.26	-0.1	0.02	-0.1
Z126	-0.7	0.59	-0.1	-0.2	-0	-0.3
Z127	-0.9	-0.2	0.03	0.17	-0.1	0.04
Z128	-0.6	0.62	-0.1	-0.2	0.03	-0.3
Z129	-0.9	0.02	-0	0.13	-0.1	-0
Z130	-0.8	0.39	-0.2	-0.2	-0.1	-0.2
Z131	-0.9	-0.3	0.14	0.18	0.03	0.06
Z136	0.01	-0.1	-0.3	0.08	0.62	0.06
Z140	-0.8	-0.5	0.02	0.15	-0.1	0.07
Z141	-0.6	0.66	-0.1	-0.2	0.11	-0.2
Z142	-0.2	0.41	0.21	-0.3	0.24	0.01
Z148	-1	-0	0.13	0.09	0.1	0.02
Z149	-0.9	0.07	0.02	-0.1	0.09	0.04
Z150	-0.7	-0.2	-0.2	-0.1	-0.4	-0.2
Z157	-0.8	-0.5	0.04	0.04	-0.1	0
Z159	-0.7	-0.6	0.25	-0.1	0.05	-0.1
Z160	-0.4	-0.8	0.24	-0.2	0.01	-0.1
Z161	-0.3	-0.8	0.33	-0.2	0.08	-0.1
Z163	0.07	0.67	0.31	0.13	0.03	0.37
Z164	0.13	0.8	0.26	-0	-0.3	0.04
Expl.Var	32.1	11	1.75	1.56	1.45	1.08
Prp.Totl	0.56	0.19	0.03	0.03	0.03	0.02

Table 1 shows that the first factor accounts for 56% of the variance and the other 5 factors together account for 30% giving a total of 86% explanation of the variance.

As can be seen, a high proportion of the variables involve the first factor, with only eleven variables not having significant contribution (x0, z48, z49, z83, z103, z109, z136, z142, z160, z163, z164). The reason why these variables do not have a significant loading in the first factor is because they do not have significant correlation with any others or between them, as mentioned in the data analysis section.

Another important feature of the first factor loadings is that all the relevant ones are negative. Given the number of variables involved and the negative sign, we can assume that the first factor is inversely related to the general population.

The second factor, which explains 19% of the variance, presents two clearly defined groups of loadings, in contrast to the first one.

The first group is composed of all variables with negative loadings: z83, z82, z118, z160, z161, z78, z159, z140 and z157. Higher values in these variables are identified with higher income (z118, z160, z161, z159) and higher education levels (z83, z82, z78).

The second group is composed of all the variables with positive loadings: z164, z49, z48, z163, z141, z128, z126, z70, z71, z109. Higher values in these variables are identified with lower income (z109), lower education (z70, z71) and smaller, more populated housing (z164, z163, z128, z126).

With the above facts in mind, we can conclude that the second factor is inversely related to socioeconomic well-being.

The third factor contains two negative and eight positive significant loadings, and explains only 3% of the variance. The negative loadings (x0, z136) imply that for larger values of this factor there is lower population density and less houses connected to the public sewers. The positive loadings correspond to z48, x2, z161, z49, z163, z164, z118, z109, z55, z159; higher values of these variables are identified with large families (z48, x2, z49, z163, z164), high income (z118, z161, z55, z159) and in apparent contradiction, low income (z109). Given this data, and the amount of variance explained by this factor, we can conclude that it is directly related to the population in the outskirts of the city.

The fourth factor exhibits similar factor loadings to the third, but with opposite sign. The only significant difference being z50 (population with health insurance services)

The fifth factor has a significant positive loading on houses connected to the public sewers (z136). On the other hand, its negative loadings on indicate that higher values of this factor are associated with fewer rented houses, higher levels of employment and less dependency relationships (z150, z103, z048).

The sixth factor has six negative loadings and five positive loadings. The negative loadings stand for rented houses or with one or two rooms (z128, z126, z130, and z150), people born outside the state (z55) and houses with water service on-site (z141). Positive loadings correspond to high house occupancy, state-born people (z54) and jobs with low payment or none at all (z114, z115, z109). Even though the relation is not very clear, we can associate this factor with high occupancy houses.

4 Cluster Analysis

Since this paper tries to propose an objective, approximate model for the grouping of population variables and its corresponding zones, we briefly outline how cluster analysis fits with the tools used by this discipline.

4.1 Pattern Recognition

Cluster Analysis (CA) is identified as an unsupervised pattern recognition technique without learning. Even tough some authors do not recognize AC as a pattern recognition branch, the mathematical process that backs this technique up well justifies its inclusion in the field.

Things to note when using AC are

1. Population diversity is the raw material for quantitative analysis; however, sometimes an individual or isolated object is too small a recipient of diversity. So we must take care when choosing the observational unit, depending on the objective of the analysis.
2. It is convenient to form groups with the original subjects and focus analysis on them instead of the individual subjects.
3. If an already defined taxonomy exists it is preferable to use it, otherwise we should create one according to our needs.

The fundamental purpose of CA is to use the information contained in a number of variables for each subject to measure the similarity between them. Once the similarity has been measured, the subjects are assigned to groups homogeneous within them but different to other groups (clusters). This new "dimension" created by the clusters allows us to take a *segmented* approach to our analysis.

CA does not have inferential properties; instead, the resulting groups are usable only to the particular data set, sample, relevant variables, number of clusters, etc. that originated them. It is also important to remember that CA does not have much in common with Discriminant Analysis, since DA tries to explain a structure while CA tries to determine it [4].

The application of CA is appropriate when we have a multivariate data set and we want to study a set of interdependent relationships where there is no distinction between dependent and independent variables (as is the case with population zones). The main objective of CA is to reduce the number of *objects* by grouping them into a lower number of clusters; The set of clusters thus obtained minimizes the within-group variation and maximizes the between-group variation. CA is actually a set of different techniques that "group" or "classify" objects.

Regardless of the type of codification used for *p* qualitative variables, data for the subject *i* can be expressed as a p-dimensional vector x_i grouped in a data matrix X where each row is x_i :

$$X = [(x_{ij})] = \begin{bmatrix} x_1^t \\ \vdots \\ x_i^t \\ \vdots \\ x_n^t \end{bmatrix} = \left[x^1 \middle| \cdots \middle| x^j \middle| \cdots \middle| x^p \right]$$

The aim of CA is to present a classification scheme of all x_i into g clusters (groups, types, classes). This kind of analysis has many applications in other branches of knowledge, since it answers the question "How do we group the subjects? When an initial classification does not exists.

Cluster analysis can be applied to either variables or subjects. When applied to variables, data from each variable is compared to determine how "close" are the variables from each other. In this case the "closeness" criteria may be one of the following: Euclidean distance, Squared Euclidean distance, City-block distance, Chebychev distance, Power distance, Percent disagreement. Once the initial clusters are formed, another criteria are needed to link new variables; again there are different ones, depending on the nature of the data: Nearest neighbor, Furthest neighbor, Unweighted pair-group average, Weighted (by cluster size) pair-group average, Enweighted pair-group centroid, Eeighted pair-group centroid, or Ward's method.

In this case we used Euclidean distance and Ward's method, with the resulting dendogram in figure 2.

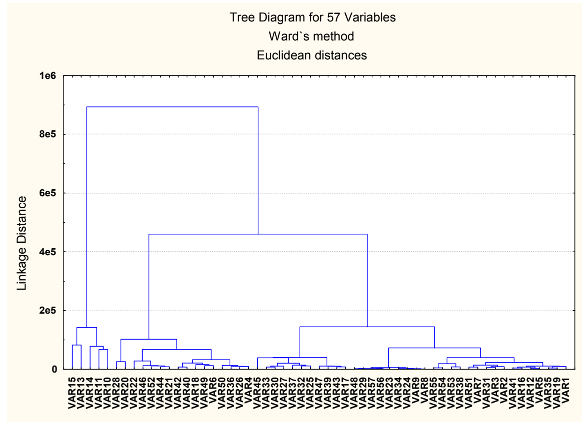

Fig. 2. Dendogram for 57 variables

Examining Fig. 2, we can see that 5 major clusters can be formed. We list them below together with the numbers of the variables that compose them:

Cluster 1
15, 13, 14, 11, 10
Cluster 2
28, 20, 22, 46, 52, 44, 21, 42, 40, 18, 49, 6, 50, 36, 26, 4
Cluster 3
45, 33, 30, 27, 37, 32, 25, 47, 39, 43, 17
Cluster 4
48, 29, 57, 56, 23, 34, 24, 9, 8
Cluster 5
55, 54, 53, 38, 51, 7, 31, 3, 2, 41, 16, 12, 5, 35, 19, 1

The resulting classification will be useful in further stages of research, together with the factors obtained in the previous section.

When applied to data, we used the CA technique known as k-means clustering, which is very different from the previous one. In this case the number of clusters is given beforehand and we start with k initial clusters and follow an iterative process; in each iteration we move objects between clusters trying to minimize variability within clusters and maximize variability between clusters.

In this case we used 16 clusters, with the resulting clustering of zones shown in

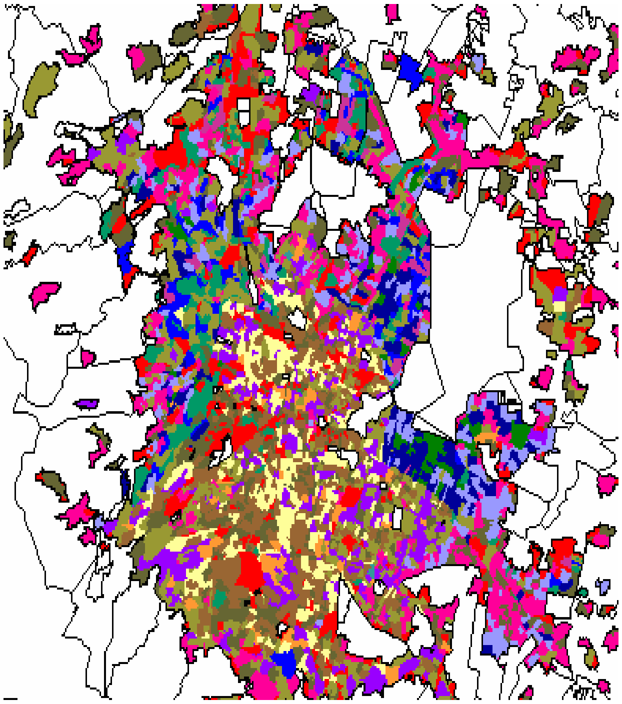

Fig. 3. Clusters of zones using the k-median method (central area only)

Figures 4 to 8 present the most relevant clusters. There we can see that clusters one and four represent the inner part of the city, while clusters two and eight correspond to the outer part. Cluster twelve deserves special attention because it corresponds to densely populated areas in the neighboring municipalities.

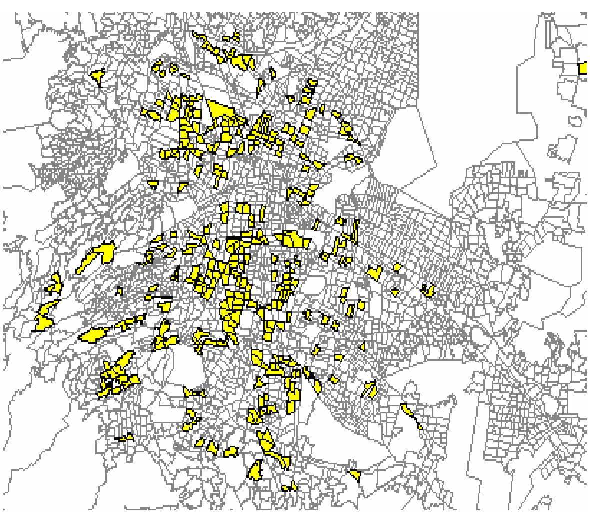

Fig. 4. Cluster 1

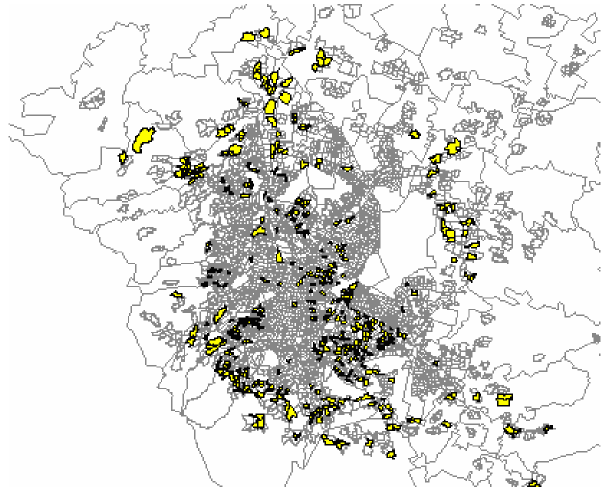

Fig. 5. Cluster 2

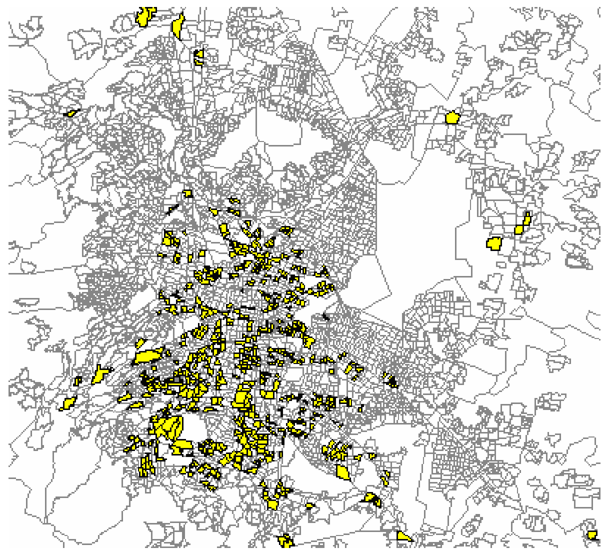

Fig. 6. Cluster 4

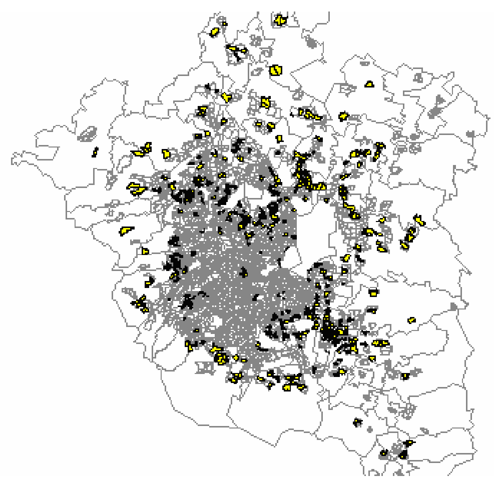

Fig. 7. Cluster 8

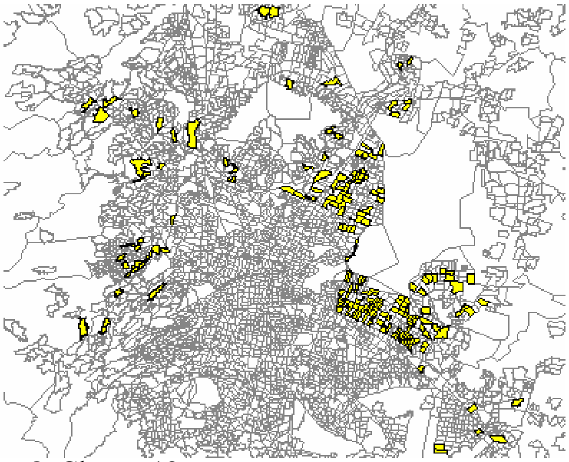

Fig. 8. Cluster 12

5 Conclusions

The results of this paper provide us some insight into the nature and structure of the population data for the MZVM. From the PCA we can see that the first factor is completely influenced by the general population, and thus is useful to represent the characteristics of the population as a whole. The second factor is quite interesting given its representation of economic well-being.

The remaining factors, in spite of the low variance they represent, provide us with indicators of zones with a significant deviation from the average.

The next step in this research topic is to use the resulting factors and the clusters of variables to provide a complete taxonomy that will help to set-up data for other pattern recognition techniques, like heuristical local models, simulated annealing or neural networks.

The clusters of zones, on the other hand, provide us with an initial classification useful for comparing with other classifications used for transportation planning. The ultimate objective of this process is to join both classifications in order to achieve a better understanding of the transportation dynamics of Mexico City.

Further research is needed to enhance the results provided by this work. Three main topics remain to be investigated: a complete taxonomy of the selected 57 variables, use of other patter recognition techniques to create zone's classifications, and comparison of the resulting classifications with transportation data.

References:

[1]. Anderson ,T.W, 1984, "An Introduction to Multivariate Statistical Analysis", 2nd Edition,. Wiley.

[2]. Bernábe L., B, Olsina, L, 2003 **"**Análisis Factorial de Correspondencias Simples en el procesamiento de datos cualitativos sobre funcionalidad en comercio electrónico", Novena Conferencia de Ingeniería Eléctrica 2003, CINVESTAV-IPN.

[3]. Chatfield, C.; Collins, A.J., 1991, "Introduction to Multivariate Analysis", Ed. Chapman & Hall.

[4]. Dillon, W.R. & Goldstein, M., 1984, "Multivariate Analysis: Methods and Applications", Wiley, New York.

[5]. Dixon,W.J , 1990, "BMDP Statistical Software Manual", Vol I, II., Dixon,W.J Eds, University of California Press, Berkeley, California.

[6]. Loranca, M.B. & Olsina, L.; 2003, "Técnicas Estadísticas para el Análisis de la Calidad de Sitios Web ", 6o Workshop Iberoamericano de Ingeniería de Requisitos y Ambientes Software (IDEAS 2003), Asunción Paraguay

[7]. Ruíz Shulcloper. J., Alba, E. y Lazo Cortés, M. "Reconocimiento de Patrones", (ediciones I, II, II y IV). México, D.F. (1994). Diplomado de titulación. Benemérita Universidad Autónoma de Puebla.

[8] XII Censo de Población y Vivienda. Resultados definitivos. INEGI. México 2001

[9] Ruiz Shulcloper, J., "Modelos matemáticos para el Reconocimiento de Patrones", Universidad Central de Las Villas, 1990.

Appendix A

Table A.1. Families and representative variables

Family	Rep.
x7, z56, z57, z81, z84, z87, z90 z101, z102, z108, z168	z56
z52, z50	z50
z67, x1, x3, x4	x34*
z71, z72, z74, z79	z71
z75, z76	z75
z73, z153, z155, z156, z157, z158	z157
z120, z139, z143, z144, z145, z146, z151, z152, z154, z165	z120
*x34 = x3+x4	

Table A.2 Variable table

Variable	Description	Selected
x000	Population density (inhabitants/hectare)	Yes
x002	Population from 4 to 6 years old	Yes
x005	Population from 15 to 18 years old	Yes
x006	Population from 18 to 24 years old	Yes
x008	Population from 60 to 65 years old	Yes
x034	Population from 6 to 15 years old	Yes
z045	Population of age 65 or older	Yes
z048	Index of economic dependecy relationship	Yes
z049	Children-female ratio	Yes
z050	Population with helth insurance services	Yes
z051	Population without health insurance services	Yes
z053	Handicaped population	Yes
z054	Population born in the state	Yes
z055	Population born outside the state	Yes
z056	Population 5 years old or older living in the municipality in 1995	Yes
z070	Population 15 years old or older without instruction	Yes
z071	Population 15 or older years old with complete primary education	Yes
z075	Population 15 years old or older with secondary education or technical/commerce education with complete primary education	Yes
z077	Population 15 years old or older with complete secondary education	Yes
z078	Population 15 years old or older with high-school or college education	Yes
z080	Population 18 years old or older with high-school education	Yes
z082	Population 18 years old or older with college education	Yes
z083	Average schooling level	Yes
z103	Unemployed population	Yes
z104	Population 12 years old or older student and economically inactive	Yes
z105	Population 12 years old or older economically inactive dedicated to housekeeping	Yes
z106	Population employed in the secondary sector	Yes
z107	Population employed in the tertiary sector	Yes
z109	Population employed as day labourer or unskilled labourer	Yes
z110	Self-employed population	Yes
z111	Employed population that worked up to 32 hours during the baseline week	Yes
z112	Employed population that worked from 33 to 40 hours during the baseline week	Yes
z113	Employed population that worked from 41 to 48 hours during the baseline week	Yes
z114	Employed population that does not receive payment	Yes
z115	Employed population that receives less than the minimum wage	Yes
z116	Employed population that receives 1 to 2 minimum wages monthly	Yes
z117	Employed population that receives 2 to 5 minimum wages monthly	Yes
z118	Employed population that receives more than 5 minimum wages monthly	Yes
z120	Occupied private houses	Yes
z126	Private houses with one room (Houses with two rooms, on of them kitchen)	Yes
z127	Private houses with 2 to 5 rooms (not including exclusive kitchen)	Yes
z128	Private houses with only one room	Yes
z129	Private houses with 2 to 5 rooms (including exclusive kitchen)	Yes
z130	Private houses with one bedroom	Yes
z131	Private houses with 2 to 4 bedrooms	Yes
z136	Private houses connected to the public sewers	Yes
z140	Private houses with drinking water in-house	Yes
z141	Private houses with drinking water on-site	Yes
z142	Private houses with carried drinking water (public tap and other houses)	Yes
z148	Privately owned houses	Yes
z149	Privately owned and payed houses	Yes
z150	Rented privated houses	Yes
z157	Private houses with telephone	Yes
z159	Private houses with private car or pick-up	Yes
z160	Private houses with computer	Yes
z161	Pribate houses with all goods	Yes
z163	Average occupancy of private houses	Yes
z164	Average room occupancy of private houses	Yes
z001	Total population	No
z002	Male population	No
z003	Female population	No

Variable	Description	Selected
z004	Population from 0 to 4 years old	No
z005	Male population from 0 to 4 years old	No
z006	Female population from 0 to 4 years old	No
z007	Population from 0 to 14 years old	No
z008	Male population from 0 to 14 years old	No
z009	Female population from 0 to 14 years old	No
z010	Population 5 years old or older	No
z011	Male population 5 years old or older	No
z012	Female population 5 years old or older	No
z013	Population 6 years old or older	No
z014	Male population 6 years old or older	No
z015	Female population 6 years old or older	No
z016	Population from 6 to 14 years old	No
z017	Male population from 6 to 14 years old	No
z018	Female population from 6 to 14 years old	No
z019	Population 12 years old or older	No
z020	Male population 12 years old or older	No
z021	Female population 12 years old or older	No
z022	Population 15 years old or older	No
z023	Male population 15 years old or older	No
z024	Female population 15 years old or older	No
z025	Female population from 15 to 49 years old	No
z026	Population from 15 to 64 years old	No
z027	Male population from 15 to 64 years old	No
z028	Female population from 15 to 64 years old	No
z029	Population from 15 to 19 years old	No
z030	Male population from 15 to 19 years old	No
z031	Female population from 15 to 19 years old	No
z032	Population 18 years old or older	No
z033	Male population 18 years old or older	No
z034	Female population 18 years old or older	No
z035	Population 20 years old or older	No
z036	Male population 20 years old or older	No
z037	Female population 20 years old or older	No
z038	Population from 20 to 24 years old	No
z039	Male population from 20 to 24 years old	No
z040	Female population from 20 to 24 years old	No
z041	Female population 50 years old or older	No
z042	Population 60 years old or older	No
z043	Male population 60 years old or older	No
z044	Female population 60 years old or older	No
z046	Male population 65 years old or older	No
z047	Female population 65 years old or older	No
z052	Population with health service from IMSS	No
z057	Population 5 years or older resident within the state by 1995	No
z058	Catholic population 5 years old or older	No
z059	Population 5 years old or older with a non-catholic religioin (includes without religion)	No
z060	Non-catholic 5 years old or older population (includes without religion)	No
z061	Literate population from 6 to 14 years old	No
z062	Male literate population from 6 to 14 years old	No
z063	Female literate population from 6 to 14 years old	No
z064	Literate population 15 years old or older	No
z065	Male literate population 15 years old or older	No
z066	Female literate population 15 years old or older	No
z067	School-going population from 6 to 14 years old	No
z068	Male school-going population from 6 to 14 years old	No
z069	Female school-going population from 6 to 14 years old	No
z072	15 years old or older population with complete primary education	No
z073	15 years old or older population with post-primary education	No
z074	15 years old or older population without post-primary education	No
z076	15 years old or older population with complete secondary education	No
z079	18 years old or older population without high-school education	No
z081	18 years old or older population without college education	No
z084	12 years old or older population single	No

Variable	Description	Selected
z085	Male 12 years old or older population single	No
z086	Female 12 years old or older population single	No
z087	15 to 24 years old population single	No
z088	Male 15 to 24 years old population single	No
z089	Female 15 to 24 years old population single	No
z090	12 years old or older population married	No
z091	Male 12 years old or older population married	No
z092	Female 12 years old or older population married	No
z093	12 years old or older population in free union	No
z094	Total living children born of 15 to 49 year old women	No
z095	Total living children born of 50 years old or older women	No
z096	Total deceased children of 15 to 49 year old women	No
z097	Total deceased children of 50 years old or older women	No
z098	Total living children born in 1999	No
z099	Total living children born from march 1999 to february 2000	No
z100	Average of living children born of 12 years old or older women	No
z101	Economically active population	No
z102	Economically inactive population	No
z108	Population employed as clerk or worker	No
z119	Total inhabitated houses	No
z121	Private houses with roofs made of light, natural or precarious materials	No
z122	Private houses with roofs made of concrete or bricks or with flat roofs	No
z123	Private houses with walls made of light, natural or precarious materials	No
z124	Private houses with walls made of bricks, blocks, stone, quarry or concrete	No
z125	Private houses with floor made of cement, mosaic, wood or other covering	No
z132	Private houses with exclusive kitchen	No
z133	Private houses with non-exclusive kitchen	No
z134	Private houses that use gas for cooking	No
z135	Private houses with exclusive sanitary service	No
z137	Private houses with sewer connected to septic tank, ravine, river, lake or sea	No
z138	Private houses without sewer	No
z139	Private houses with electricity	No
z143	Private houses with only sewerage and running water	No
z144	Private houses with only sewerage and electricity	No
z145	Private houses with only running water and electricity	No
z146	Private houses with only running water, sewerage and electricity	No
z147	Private houses without running water, sewerage, nor electricity	No
z151	Private houses with radio or tape recorder	No
z152	Private houses with TV	No
z153	Private houses with VCR	No
z154	Private houses with blender	No
z155	Private houses with refrigerator	No
z156	Private houses with washing machine	No
z158	Private houses with boiler	No
z162	Private houses without goods (none)	No
z165	Total of households	No
z166	Male-headed households	No
z167	Female-headed households	No
z168	Population in households	No
z169	Population in male-headed households	No
z170	Population in female-headed households	No