

PS-1059

REGIONALISATION OF THE PROVINCE OF CORDOBA BASED ON SOCIO-DEMOGRAPHIC VARIABLES USING THE DATAMINING TECHNIQUE, THE CLUSTER ANALYSIS

Cecilia Beatriz Díaz (Facultad de Ciencias Económicas, Universidad Nacional de Córdoba, Córdoba – Argentina) - cdiaz@eco.unc.edu.ar

The aim of the research was the regionalisation of the province of Cordoba through the use of one of the classification techniques of data mining called cluster analysis. The methodology whose main purpose is to group objects based on their characteristics was used and it let us identify two big types: rich and poor municipalities. Then, they were subdivided and, as a result, more homogeneous groups were found. To reach the aim of achieving geographical areas, the existence of evidence about spatial dependence was investigated using indicators of “unsatisfied basic necessities” and “the home-chief’s low level of education”. The geographical place in the province determines poverty. That is to say, that the socio-economical conditions of a municipality depend on its neighbours’ characteristics. The results acquire great importance because they let us achieve an adequate focus on policies of distribution of resources and of tax collection. They will let us improve the planification of public investments to reorientate the division of the budget according to the growth necessities of each regions.

Keywords: Regionalisation – census - classification – cluster – spatial dependence

REGIONALIZACIÓN DE LA PROVINCIA DE CÓRDOBA A PARTIR DE VARIABLES SOCIO-DEMOGRÁFICAS CON LA APLICACIÓN DE UNA TÉCNICA DE DATAMINING, EL CLUSTER ANALYSIS

La investigación tuvo como objetivo la regionalización de la Provincia de Córdoba mediante la aplicación de una de las técnicas de clasificación de datamining denominada análisis de cluster. La metodología utilizada, cuyo principal propósito es agrupar objetos basándose en las características que poseen, permitió identificar dos grandes clases: municipios *ricos* y *pobres*. Luego se subdividieron logrando grupos más homogéneos. Para alcanzar el propósito de lograr áreas geográficas, se indagó sobre la existencia de evidencias acerca de la dependencia espacial utilizando los indicadores de “necesidades básicas insatisfechas” y el “bajo nivel de instrucción del jefe de hogar”. La ubicación geográfica en la Provincia es determinante de la pobreza, es decir, las condiciones socio-económicas de un municipio dependen de las características que tengan sus vecinos. Los resultados adquieren gran importancia, ya que permiten lograr una adecuada focalización de políticas de distribución de recursos y de recaudación impositiva. Permitirá mejorar la planificación de las inversiones públicas, para la reorientación de las partidas presupuestarias de acuerdo a las necesidades de crecimiento de cada región.

Palabras claves: regionalización - censo - clasificación - cluster - autocorrelación espacial

I. Introducción al concepto de desarrollo económico y social

Desde 1990 el Programa de las Naciones Unidas para el Desarrollo (PNUD) elabora anualmente un Informe Mundial sobre el Desarrollo Humano de los países miembros. En 1995 se presentó el primer Informe Argentino elaborado por una comisión especial del Senado de la Nación, que siguió la misma metodología que el organismo internacional para calcular el Índice de Desarrollo Humano. Esa primera estimación realizada para nuestro país revela la desigual distribución del indicador y se analizan las desigualdades provinciales.

Argentina ocupa el puesto 39 de Desarrollo Humano en el orden mundial, entre 174 países, ocupando los primeros puestos Canadá, Noruega y los Estados Unidos.

El Informe advierte la existencia de dos Argentinas. Suponiendo que cada provincia fuera un Estado independiente, la Ciudad de Buenos Aires ocuparía el puesto 20, pero Jujuy estaría en el puesto 85. Esta clasificación reproduce dentro del país la misma distancia que en el orden mundial separa a Irlanda de Túnez o a España de Paraguay.

Provincias	IDH 1991		IDH 1996	
	Valor	Rango	Valor	Rango
Ciudad de Bs. As.	0.896	1	0.892	1
Córdoba	0.844	2	0.831	4
Neuquén	0.838	3	0.817	10
Tierra del Fuego	0.838	3	0.854	2
Santa Cruz	0.835	5	0.843	3
La Pampa	0.832	6	0.821	5
Chubut	0.832	6	0.819	9
Provincia de Bs. As.	0.828	8	0.820	8
Santa Fe	0.820	9	0.821	5
Mendoza	0.795	10	0.820	7
La Rioja	0.771	11	0.804	12
Río Negro	0.767	12	0.810	11
Tucumán	0.761	13	0.790	18
Catamarca	0.757	14	0.799	15
San Luis	0.756	15	0.801	13
Entre Ríos	0.738	16	0.801	13
Salta	0.713	17	0.788	17
San Juan	0.703	18	0.798	16
Misiones	0.703	18	0.772	20
Santiago del Estero	0.691	20	0.774	19
Chaco	0.690	21	0.755	24
Formosa	0.683	22	0.764	23
Corrientes	0.680	23	0.772	21
Jujuy	0.676	24	0.772	21

Fuente: Informe Argentino de Desarrollo Humano, 1999

De la comparación entre los años 1991 y 1996 se advierte que la mayoría de las jurisdicciones han mantenido su posición relativa. Los saltos más significativos corresponden a Neuquén, Chubut y Tucumán que descienden en su posición, mientras que Santa Fe, Mendoza y San Juan ascienden.

Planteo del problema de estudio

En la Provincia de Córdoba existen una diversidad de factores (sociales, económicos, culturales, geográficos, etc.) que han determinado que dentro del territorio de la Provincia se produzcan diferencias en la estructura socio-demográfica de los aglomerados urbanos, que pueden ser agrupados según características comunes u homogéneas. **El objetivo de la investigación es detectar grupos de localidades con características similares y analizar si estos pueden asimilarse a regiones de la Provincia de Córdoba.**

El régimen municipal de la Provincia de Córdoba es el más autónomo de la República Argentina. La Constitución provincial de 1987 reconoce a los municipios como comunidades naturales fundados en la convivencia y les asegura la autonomía política, administrativa, económica y financiera (a las ciudades también les asegura la autonomía institucional, es decir, la facultad de dictar sus cartas orgánicas), adelantándose, de esa manera, a la reforma de la Constitución nacional de 1994. En ella establece el título: “REGIONALIZACION” y en su Artículo 175 dispone: “*Una ley especial establece la regionalización de la Provincia a los fines de facilitar la desconcentración administrativa, la más eficiente prestación de los servicios públicos, y unificar los diversos criterios de división territorial*”. A su vez, la Ley de Regionalización Provincial N° 9206, reconoce en cada departamento a una Comunidad Regional integrado por los municipios y comunas en forma voluntaria.

Por su parte la Ley Provincial de Educación en el Artículo 15, establece que “*El sistema educativo se organizará, en términos de política educacional, conforme con los principios de democratización, centralización política y normativa, regionalización, desconcentración operativa y administrativa y participación social, a cuyo efecto las autoridades provinciales:*

Garantizarán la equidad en los servicios educativos, a fin de alcanzar igualdad en las oportunidades y posibilidades de acceso, permanencia y logro educativos, y ofrecerán una EDUCACIÓN que asegure la democrática distribución de los conocimientos personal y socialmente relevantes.

Establecerán las grandes líneas de política educativa y asegurarán el cumplimiento de los objetivos del sistema, articulando orgánicamente su gobierno y administración, con el objeto de lograr la mayor eficacia de las acciones propuestas.

Adecuarán la oferta educativa a las características de las distintas regiones de la Provincia, respetando sus pautas socioculturales y promoviendo el despliegue de sus potencialidades.

Asegurarán que los organismos regionales y locales de administración educativa, ejerzan las competencias que esta norma les confiera.

Reconocerán la capacidad y la responsabilidad de la sociedad para intervenir en la toma de decisiones educativas y en el control de su ejecución, y promoverán su participación activa”.

La provincia se encuentra dividida en 26 departamentos, cada uno subdividido en pedanías. Indudablemente, las autonomías locales son una fortaleza del régimen municipal cordobés. Institucionalmente, de los 427 gobiernos locales existentes (más del 20 por ciento del total nacional), 249 son municipios (más de dos mil habitantes) y 178 son comunas (menos de dos mil habitantes). Sin embargo, desde un punto de vista poblacional, hay 108 municipios que tienen menos de dos mil habitantes (deberían ser comunas pero

siguen siendo municipios en virtud de una cláusula transitoria de la Constitución provincial) y hay comunas que tienen más de dos mil habitantes (cinco, que deberían ser municipios).

II. Objetivos de la investigación

Los objetivos planteados en la presente investigación, son:

- 1 Seleccionar variables que permitan diferenciar los localidades de la Provincia de Córdoba a partir de diferentes fuentes de datos.
- 2 Clasificar las unidades de análisis, es decir, detectar grupos homogéneos utilizando los índices construidos ad-hoc.
- 3 Describir cada uno de los grupos que se identifiquen.
- 4 Representar gráficamente la composición socio-demográfico de la provincia de Córdoba.

III. Fuente de Datos

En esta investigación se utilizó la información del Censo de Población y Vivienda de 2001 de la Provincia de Córdoba, cuyos datos fueron proporcionados por la Dirección de Estadística de la Provincia de Córdoba. El relevamiento se llevó a cabo entre los días sábado 17 y domingo 18 de noviembre, y consistió en censar a las personas presentes en la vivienda, en ese momento, incluyendo a los que no residen habitualmente en el hogar, pero que pasaron la noche allí.

La metodología que se utilizó fue la de entrevista directa, no se aplicó muestra es decir se utilizó un mismo y único cuestionario para relevar la información, sólo se diferenciaron los cuestionarios censales que se aplicaron según sean *instituciones colectivas* u *hogares particulares*.

La **unidad de análisis es la localidad**, definida como áreas urbanas dentro de la estructura censal, para cada uno de los 26 departamentos en que se divide la Provincia de Córdoba. En el lenguaje común el término *localidad* se usa en varios sentidos. Se suele nombrar de esta forma a un “municipio”, a la llamada “zona urbanizada” de un pueblo o ciudad.

La definición técnica adoptada por el censo de 2001, conceptualiza la localidad como porción de la superficie de la tierra caracterizada por la forma, cantidad, tamaño y proximidad entre sí de ciertos objetos físicos artificiales fijos (edificios) y por ciertas modificaciones artificiales del suelo (calles), necesarias para conectar aquellos entre sí. Brevemente, una localidad se define como **concentración espacial de edificios conectados entre sí por calles**. Es idea central de esta definición la proximidad entre edificios, estén o no destinados a vivienda: a partir de la zona reconocidamente céntrica de una determinada localidad, la envolvente llega en cualquier dirección hasta donde la continuidad de edificación se interrumpe por largo trecho. Para delinear la envolvente es indiferente que la edificación se extienda en forma regular (sobre un trazado catastral amanzanado) o en forma caótica (viviendas precarias en terrenos invadidos, edificación salpicada sobre laderas montañosas o a lo largo de rutas, etc.).

Para cada unidad de análisis se calcularon indicadores que se crearon ad-hoc para esta investigación. La “Base Usuaría” (tabla de datos agregados) publica datos de población, vivienda y hogar de 550 localidades, desagregando para cada departamento la población rural dispersa a los fines de la zonificación.

Adicionalmente, se incluyen variables que surgen de otras fuentes de información:

- ✓ Distribución de “locales” del Censo Económico del año 2005
- ✓ Distribución del Gasto Público Provincial
 - Gasto Social
 - Gasto en Salud
 - Gasto en Educación
 - Gasto en Viviendas
- ✓ Distribución de la Inversión Pública Provincial
 - Inversión en caminos
 - Inversión en edificios públicos
 - Inversión en obras hidráulicas
 - Inversión en otras construcciones
 - Inversión en viviendas
- ✓ Distribución del Producto Bruto Geográfico

Estas variables son tratadas como ilustrativas o secundarias ya que están medidas a nivel de departamento, pero es importante su inclusión ya que permite enriquecer la descripción de los grupos que conforman las “zonas” de la Provincia.

Indicadores utilizados

Los indicadores expresan el porcentaje de personas u hogares que poseen ciertas características en una localidad determinada. Por lo tanto, el valor que asume la variable indica el grado de satisfacción de una necesidad, la capacidad de subsistencia, las condiciones de vida y el acceso a servicios sanitarios, etc.

De acuerdo al análisis bibliográfico, el desarrollo está relacionado con variables que permitan medir las condiciones de vida de la población, tales como, salud, educación, vivienda, vulnerabilidad, seguridad social, empleo, salarios, pobreza, distribución del ingreso, entre otros.

Indicadores de comportamiento social

- *Nivel ocupacional*

Distribución de la población ocupada de 14 o más años por categoría ocupacional:

- 1) Porcentaje de obreros o empleados del sector público
- 2) Porcentaje de obreros o empleados del sector privado
- 3) Porcentaje de patronos
- 4) Porcentaje de trabajadores por cuenta propia.
- 5) Porcentaje de trabajadores familiares con o sin salario

Distribución de la población ocupada de 14 o más años por categoría ocupacional:

- 6) Porcentaje de ocupados
- 7) Porcentaje de desocupados o subocupados
- 8) Porcentaje de jubilados o pensionados
- 9) Porcentaje de estudiantes
- 10) Porcentaje de otros inactivos

- *Máximo nivel de escolaridad alcanzado por el jefe del hogar*

Se calcularon seis indicadores:

- 1) Porcentaje de jefes o jefas de hogar sin instrucción, es decir, sin asistencia escolar.
- 2) Porcentaje de jefes o jefas de hogar con nivel de instrucción primaria incompleta.
- 3) Porcentaje de jefes o jefas de hogar con nivel de instrucción primaria completa o secundaria incompleta.
- 4) Porcentaje de jefes o jefas de hogar con nivel de instrucción secundario completo o terciario o universitario incompleto.
- 5) Porcentaje de jefes o jefas de hogar con nivel de instrucción terciario o universitario completo.

- *Nivel socioeconómico*

Las *condiciones de la vivienda y de los servicios sanitarios* expresan distintos grados de privación, o no, referidos al tipo de vivienda, al hacinamiento por cuarto, a la procedencia y provisión del agua, a las características de los servicios sanitarios y el material predominante en los pisos.

Se utilizaron dos indicadores:

- 1) Porcentaje de población en viviendas deficitarias, tales como, ranchos, casillas o viviendas precarias, inquilinatos, hoteles y pensiones.
- 2) Porcentaje de hogares con más de tres personas por cuarto (se refiere al hacinamiento por cuarto).
- 3) Porcentaje de hogares en vivienda sin retrete con descarga de agua.

Las viviendas sin acceso a la red pública de agua, sin retrete con descarga de agua o con piso de tierra son consideradas precarias y, en general, son características que presentan las definidas como deficitarias, aunque, estas fueron calificadas de esa manera teniendo en cuenta el *tipo de vivienda*. Por tal motivo se incluyó este último indicador que refleja mayor precariedad y resumiría las condiciones antes mencionadas.

El censo no indaga sobre niveles de ingresos, entonces, para lograr una aproximación al mismo se calcularon indicadores que tienen en cuenta el régimen de tenencia de la vivienda y la cobertura de salud, ya que según la modalidad que asuma indicaría la posibilidad económica para hacer frente a la inversión o gasto. Los indicadores calculados son los siguientes:

- 4) Porcentaje de hogares propietarios de la vivienda.
- 5) Porcentaje de hogares inquilinos, ocupantes en relación de dependencia o prestada.
- 6) Porcentaje de hogares ocupantes de hecho de la vivienda. Este último indicador permite detectar los sectores más pobres.

Con respecto a la cobertura de salud, poseerla o no tiene fundamental importancia para determinar el tipo de institución a la que se recurre en caso de requerir atención médica. Se calcularon los siguientes indicadores:

- 7) Porcentaje de población que tiene obra social o plan médico
- 8) Porcentaje de población que no tiene obra social o ni plan médico

El siguiente conjunto de indicadores muestran el grado de confort o bienestar de la población en la vida cotidiana.

- 9) Porcentaje de hogares particulares con cocina con pileta e instalación de agua
- 10) Porcentaje de hogares particulares con cocina con pileta sin instalación de agua
- 11) Porcentaje de hogares particulares sin cocina
- 12) Porcentaje de hogares que tienen PC con conexión a Internet
- 13) Porcentaje de hogares que tienen PC sin conexión a Internet
- 14) Porcentaje de hogares que tienen heladera con freezer o sólo freezer
- 15) Porcentaje de hogares que tienen heladera sin freezer
- 16) Porcentaje de hogares que tienen horno microondas
- 17) Porcentaje de hogares que tienen lavarropa automático
- 18) Porcentaje de hogares que tienen lavarropa común
- 19) Porcentaje de hogares que tienen teléfono fijo y celular
- 20) Porcentaje de hogares que sólo tienen teléfono celular
- 21) Porcentaje de hogares que sólo tienen teléfono fijo
- 22) Porcentaje de hogares que tienen televisión por cable
- 23) Porcentaje de hogares que tienen video casetera

Otro indicador que se utilizó fue *la privación material de los hogares*, definida según la metodología del Índice de Privación Material de los Hogares (INDEC, "El estudio de la pobreza según el Censo Nacional de Población, Hogares y Viviendas 2001. El Índice de Privación Material de los Hogares", 2004, inédito) de la siguiente manera¹:

- Privación sólo de recursos corrientes: hogares con ingresos insuficientes, estimados a partir de los años de escolarización de los miembros del hogar ocupados o jubilados/pensionados y la cantidad de personas que componen el hogar.
- Privación sólo de recursos patrimoniales: hogares con deficiencias en las viviendas que habitan, tanto en las características de los materiales como en las instalaciones sanitarias.
- Privación convergente: Hogares que presentan privación conjunta de recursos corrientes y patrimoniales.

- 24) Porcentaje de hogares particulares con privación de recursos corrientes
- 25) Porcentaje de hogares particulares con privación de recursos patrimoniales
- 26) Porcentaje de hogares particulares con privación convergente

La estratificación social muestra desigualdades en la permanencia y utilización del sistema educativo. Se calcularon los porcentajes, para el total de la población de 3 años y más, de las siguientes categorías:

- 27) Porcentaje de población sin instrucción (nunca asistió).
- 28) Porcentaje de población con primario incompleto, incluye los que asisten de 1° a 6° grado; y los que asistieron a la escuela primaria/1° del EGB o de

¹ Fuente: Página Web de la Dirección de Estadísticas de la Provincia de Córdoba:
http://web2.cba.gov.ar/actual_web/estadisticas/censo2001/resultados/index.htm

- 1° a 6° grado. Se incluyeron en esta categoría los que asisten al nivel inicial por su escasa representación y muestran el inicio de la educación formal.
- 29) Porcentaje de población con secundario incompleto, incluye los que declaran tener la primaria completa; los que asisten a 7°, 8° y 9° del EGB. Los que asistieron y aprobaron 7° y 8° del EGB; los que señalaron que asistieron a la secundaria pero no aprobó ningún nivel.
- 30) Porcentaje de población con secundario incompleto, incluye los que declaran tener la primaria completa; los que asisten a 7°, 8° y 9° del EGB. Los que asistieron y aprobaron 7° y 8° del EGB; los que señalaron que asistieron a la secundaria pero no aprobó ningún nivel.
- 31) Porcentaje de población con terciario o universitario incompleto, incluye los que declaran que aprobaron el polimodal o el secundario, es decir, completaron ese nivel. Además los que asisten o asistieron al terciario o universitario.
- 32) Porcentaje de población con terciario o universitario completo, incluye los que declaran que aprobaron y completaron esos niveles.

Indicadores de comportamiento demográfico

Cuando se realizan estudios con fines económicos y sociales es conveniente disponer de una clasificación resumida de grandes grupos de edades, para lo que se calcularon los siguientes indicadores:

- 33) Porcentaje de población menores de 5 años
- 34) Porcentaje de población de 5 a 12 años
- 35) Porcentaje de población de 13 a 17 años
- 36) Porcentaje de población de 18 a 24 años
- 37) Porcentaje de población de 25 a 35 años
- 38) Porcentaje de población de 36 a 49 años
- 39) Porcentaje de población de 50 a 64 años
- 40) Porcentaje de población de más de 64 años

También se utiliza la *relación de dependencia* que es la razón entre el total de personas de edades *dependientes*, menores de 15 y mayores de 64 años, y las personas en las edades *económicamente productivas*, 14 a 64 años.

- 41) Relación de dependencia de la población de 0-14 años = $\frac{\text{Pobl. 0-14 años}}{\text{Población 15-64 años}} \times 100$
- 42) Rel. de dep. de la población de 65 años y más = $\frac{\text{Pobl. 65 años y más}}{\text{Población 15-64 años}} \times 100$

Para reflejar la distribución por sexo se calculó la relación de masculinidad, que expresa el número de varones por cada 100 mujeres.

$$43) \text{ Razón de masculinidad} = \frac{\text{Total de varones}}{\text{Total de mujeres}} \times 100$$

44) Promedio de hijos por mujer como un estimador de la fecundidad.

IV. Metodología

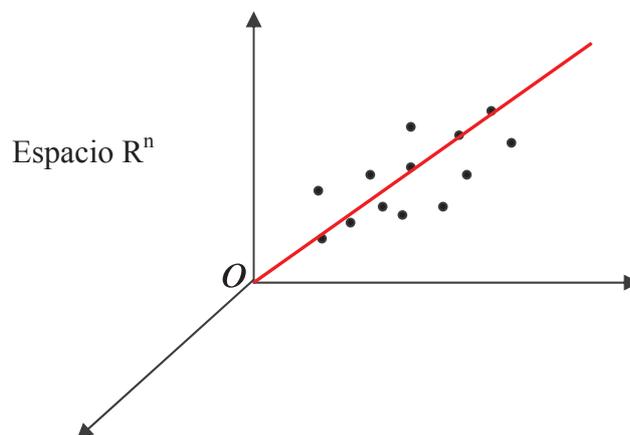
Se utilizó una de las técnicas de análisis multivariado, es decir, la consideración simultánea de más de dos variables medidas en cada individuo u objeto de investigación.

La ventaja principal de este tipo de tratamiento de la información es su capacidad de “encuadrar” las variables utilizadas en una investigación, con el fin de comprender las relaciones complejas que no se observan si se utilizan métodos univariados o bivariados.

Si se cuentan con n unidades de análisis, a quienes se les han medido p características, la información se organiza generalmente en matriz X de n filas (individuos) y p columnas (variables).

Geoméricamente los individuos se pueden representar como una nube de puntos en el espacio de p dimensiones. La siguiente figura representa los individuos en un espacio de tres dimensiones ($p=3$).

Figura N°1: Representación de la nube de puntos en el espacio de tres dimensiones



Existen dos grandes familias de métodos que permiten analizar esta nube tabla de datos:

1. los métodos factoriales, entre los que se destaca el de componentes principales, y
2. los de cluster.

Estos métodos son complementarios ya que se comienza determinando los ejes factoriales que sirven de input en el cluster.

Análisis de componentes principales

El propósito general del método de componentes principales es condensar la información contenida en una serie de variables originales en un número más pequeño de dimensiones complejas, denominadas factores o componentes, con mínima pérdida de información.

Por este método se reemplazan las p variables originales por k combinaciones lineales de las mismas (donde k es mucho menor que p), las que explicarán una importante proporción de la dispersión total.

El objetivo entonces, es tomar p variables, X_1, X_2, \dots, X_p , y encontrar combinaciones lineales, Z_1, Z_2, \dots, Z_p , no correlacionadas. Cada una de estas variables Z , denominadas componentes principales, están midiendo distintas dimensiones de los datos y es posible ordenarlas de manera que:

$$\text{var}(Z_1) \geq \text{var}(Z_2) \geq \dots \geq \text{var}(Z_p)$$

Si la suma de las varianzas de las componentes principales es igual a la varianza total, lo que se espera es que la varianza de la mayoría de las variables Z sean tan bajos que resulten insignificantes, por lo que la dispersión del conjunto de datos puede ser adecuadamente descrito por pocas variables Z con varianzas importantes.

Método de Cluster

El método de cluster es la denominación de un conjunto de técnicas multivariadas cuyo principal propósito es agrupar objetos basándose en las características que poseen. Los objetos se distribuyen de tal forma que cada uno es muy parecido a los que hay en el conglomerado de acuerdo a algún criterio de clasificación.

Las *clases* de objetos resultantes deberían mostrar un alto grado de homogeneidad interna (dentro del grupo) y un alto grado de heterogeneidad externa (entre grupos). Por lo tanto, cuando se representen gráficamente, si la clasificación es acertada, los objetos dentro de los conglomerados estarán muy próximos y los grupos diferentes estarán alejados.

Complementariedad entre el análisis factorial y el de cluster

Los métodos factoriales se adaptan particularmente bien para la exploración de las grandes bases de datos, pero no siempre alcanzan a dar una visión satisfactoria del conjunto de la información. La visualización de los gráficos representa sólo una parte del análisis, pero a veces son demasiado complejas para ser interpretados con facilidad.

Por ello, las técnicas de clasificación, entre las que se encuentra el análisis de cluster, pueden complementar y matizar los resultados de los análisis factoriales, ayudando a la interpretación.

Si hay demasiados puntos que representan individuos sobre un plano factorial es de utilidad proceder a reagruparlos en familias homogéneas. Las

clases son utilizadas para ayudar a la interpretación de los planos factoriales identificando así zonas bien descriptas.

Para describir una tabla de datos, la puesta en práctica conjunta de estas técnicas se opera de la siguiente manera:

Etapa 1: El análisis factorial

El análisis factorial está utilizado como una etapa previa a la clasificación por dos razones: por su poder de descripción y por su poder de reducción de las dimensiones, ya que permite trabajar con coordenadas factoriales menos numerosas que las variables de partida.

Etapa 2: Clasificación a partir de los factores

Es equivalente a efectuar una clasificación de los individuos en el conjunto de p variables o en el de p factores. Pero también se puede tener en cuenta un subespacio factorial de dimensión q ($q < p$) y realizar una clasificación sobre los q primeros ejes. El hecho de abandonar los últimos factores genera una especie de suavización de los datos, lo que en general mejora la partición produciendo clases más homogéneas. La dificultad reside, a veces, en la elección del número de ejes a retener.

Etapa 3: Descripción automática de las clases

Una vez reagrupados los individuos en clases se procede a describirlas. Para ello, se calculan estadígrafos de comparación entre los valores internos a la clase y los valores globales, tanto para las variables numéricas como para las nominales. Finalmente, estos valores test permiten ordenar las variables por orden de significación.

Autocorrelación espacial

Las técnicas de análisis espacial (Exploratory Spatial Data Analysis) permiten establecer si una variable se distribuye en forma aleatoria o, si por el contrario, existe evidencia de algún tipo de dependencia espacial entre sus medidas. Esta metodología es una extensión del análisis exploratorio de datos, que se limita a realizar una descripción gráfica y estadística con el fin de establecer la existencia de patrones de comportamiento en los datos utilizados, pero sin tener en cuenta factores espaciales.

La autocorrelación espacial es la concentración o dispersión de los valores de una variable en un mapa. Dicho de otra manera, refleja el grado en que objetos en una unidad geográfica son similares a otros objetos en unidades geográficas próximas. Este tipo de autocorrelación prueba la primera ley geográfica de Tobler (1970) que afirma que *todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las distantes*.

La técnica más antigua y típica para la detección y medición de la autocorrelación es el coeficiente I de Moran (1950). El diseño es similar al coeficiente de correlación de Pearson. Sus valores varían entre +1 y -1, donde el primer valor significa una autocorrelación positiva perfecta (perfecta concentración), y el segundo una autocorrelación negativa perfecta (perfecta dispersión); el cero significa un patrón espacial totalmente aleatorio. La diferencia entre los dos coeficientes, de Moran y Pearson, se basa en que en el primer caso la asociación de valores en el conjunto de datos está determinada por una matriz de distancias o contigüidad que predefine los valores vecinos.

La fórmula del coeficiente I de Moran es:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \cdot \frac{\sum \sum W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

En esta fórmula, n significa el número de las unidades o puntos en el mapa, W_{ij} es la matriz de distancias que define si las áreas o puntos geográficos, i y j , son o no vecinos. Este coeficiente I se sujeta a una prueba de significancia estadística de valores Z , es decir, con el supuesto de una distribución normal.

V. Resultados

En el análisis factorial, las variables pueden ser activas o ilustrativas. Las primeras se utilizan para construir los ejes factoriales, mientras que las segundas, también llamadas suplementarias, aportan información útil para la interpretación de los ejes factoriales, pero no se emplean para generar los factores.

En general la selección de las variables activas se hace teniendo en cuenta la información que es relevante, que resuma la información original, y además, tengan un significado claro. Por otra parte, la información que contienen las variables ilustrativas no tiene sentido considerarlas a la hora de construir los ejes.

La distribución de las variables fue la siguiente:

- CONTINUAS ACTIVAS: 71 variables. Medidas de población y hogares
- CONTINUAS ILUSTRATIVAS: 18 variables. Medidas departamentales

El primer resultado numérico que puede ser interpretado es la lista de los valores propios con los correspondientes porcentajes de varianza que los mismos representan. Los diez primeros ejes, considerados para la construcción de los clusters, alcanzan el 62% de la varianza total, lo que constituye una buena representación para este tipo de investigaciones.

Si se analizan los tres primeros ejes, se concluye que son perfectamente descriptibles. El *factor 1* se caracteriza por variables que tienen que ver con el estándar de vida y el nivel socioeconómico. En el eje negativo se ubican los indicadores que muestran no disponer de artefactos del hogar indispensables y niveles bajos de educación alcanzados. Mientras que en el positivo se encuentran las variables que representan una mayor instrucción, y mayor confort.

El *factor 2* describe composición demográfica de la población (estructura etaria) y régimen de tenencia de la vivienda. En el eje negativo se concentran la población joven que no es propietaria de la vivienda. Mientras que en el positivo, se localiza una población más envejecida, propietarios de la vivienda.

El *factor 3* describe la condición de actividad. En el eje negativo se concentran desocupados e inactivos, formado por estudiantes y personas adultas. Mientras que en el positivo, se encuentran los ocupados y los que tienen un rango de edad económicamente activa.

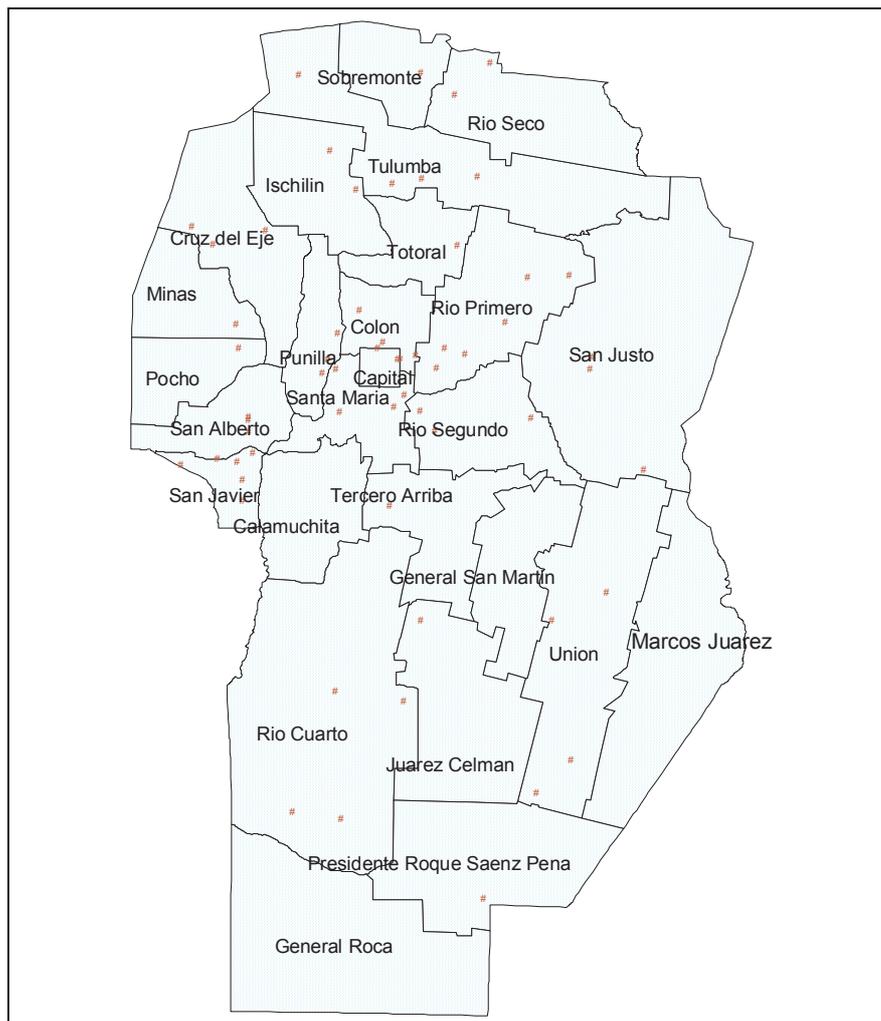
Análisis de Cluster

Se realizó el procesamiento para obtener el análisis de clusters, sobre los diez primeros factores para formar las clases o grupos. Las variables que caracterizan a cada uno de ellos, se ordenan según el grado de significación del test de diferencias de medias entre la media general versus la media del grupo (ver Anexo IV)

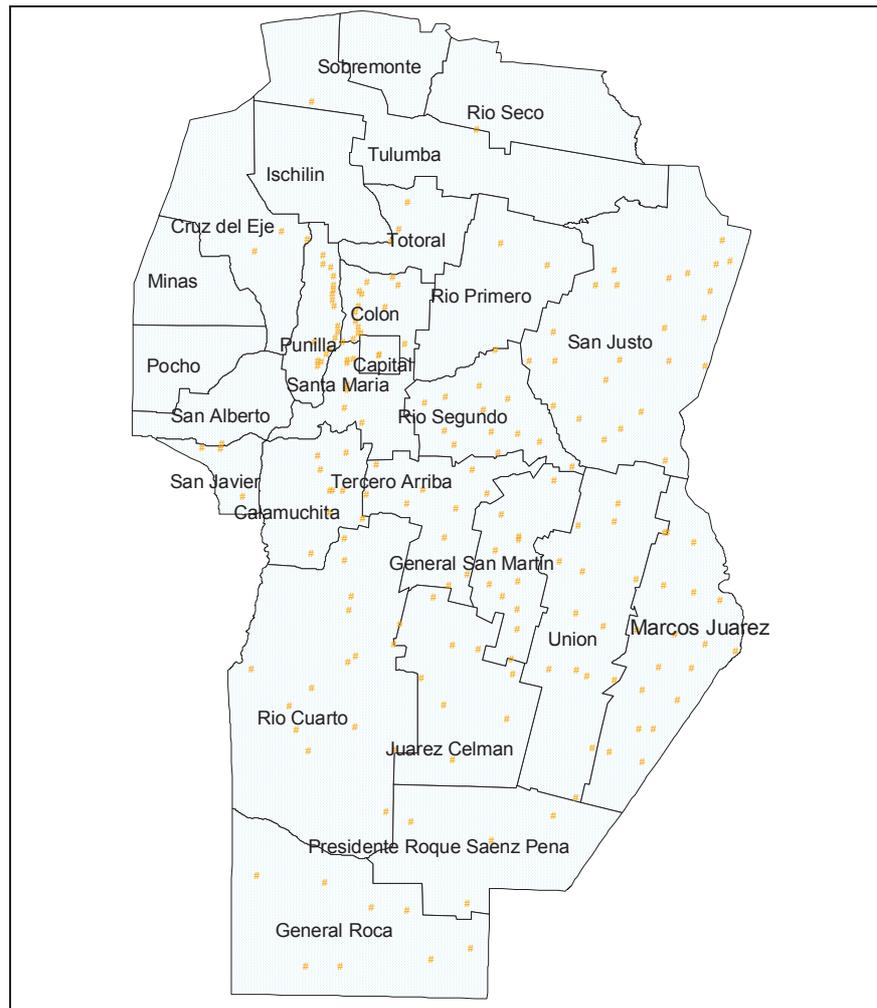
La primera partición es en municipios *pobres* y *ricos*. Cuya descripción es:

Municipios <i>pobres</i> 181 puntos	Municipios <i>ricos</i> 369 puntos
<p>Condición de vida</p> <ul style="list-style-type: none"> ✓ No tienen elementos de confort en el hogar, como por ejemplo, teléfono, lavarropa o heladera; ✓ Es significativo el porcentaje con privación convergente, es decir, recursos patrimoniales y corrientes; ✓ No tienen cobertura de salud; ✓ Elevado de NBI de hacinamiento, condiciones sanitarias y vivienda; <p>Nivel de instrucción</p> <ul style="list-style-type: none"> ✓ Nivel primario <p>Condición de Actividad</p> <ul style="list-style-type: none"> ✓ Desocupado <p>Estructura etaria</p> <ul style="list-style-type: none"> ✓ Población joven, dada la mayor proporción de niños <p>Variables departamentales</p> <ul style="list-style-type: none"> ✓ Gas'to público por cápita 	<ul style="list-style-type: none"> ✓ Mayor proporción de hogares con artefactos electrodomésticos que mejoran la calidad de vida, tales como, lavarropa automático, heladera con o sin freezer, teléfono fijo y celular, TV por cable, videocasetera, horno microondas, PC con y sin internet, entre otros. <ul style="list-style-type: none"> ✓ Nivel terciario o universitario <ul style="list-style-type: none"> ✓ Ocupado – Jubilados y pensionados <ul style="list-style-type: none"> ✓ Población envejecida, dada la proporción de varones y mujeres mayores de 65 años <ul style="list-style-type: none"> ✓ Participación en el Producto Bruto Geográfico ✓ Distribución de locales relevados en el Censo Económico ✓ Inversión pública, en viviendas, en caminos, obras hidráulicas ✓ Seguridad: cantidad de delitos por policía

A continuación se ilustran los mapas en donde se representa la distribución de las localidades que conforman cada grupo.



Mapa N° 1: Provincia de Córdoba con la distribución de localidades del grupo *pobres*



Mapa N° 2: Provincia de Córdoba con la distribución de localidades del grupo ricas

<i>Departamento</i>	<i>Localidades</i>		<i>Departamento</i>	<i>Localidades</i>	
	<i>Pobres</i>	<i>Ricas</i>		<i>Pobres</i>	<i>Ricas</i>
Calamuchita	8	29	Río Cuarto	5	28
Capital	2	1	Río Primero	20	8
Colón	10	23	Río Seco	12	2
Cruz del Eje	15	6	Río Segundo	3	19
General Roca	3	12	San Alberto	15	7
Gral. San Martín	2	16	San Javier	12	9
Ischilín	12	2	San Justo	7	38
Juárez Celman	2	16	Santa María	10	29
Marcos Juárez	2	22	Sobremonte	2	4
Minas	7	2	Tercero Arriba	1	17
Presidente Roque Saenz Peña	1	10	Totoral	6	7
Pocho	8	1	Tulumba	9	3
Punilla	4	29	Unión	2	26

Tabla N° 1: Tabla resumen de la distribución de localidades de cada grupo.

Estos dos grandes grupos pueden seguir dividiéndose en busca de una mayor homogeneidad en las particiones. A continuación se describen seis conglomerados.

Tipo de Municipios según la condición socioeconómica de la población		
<i>Sin patrimonio</i> 64 puntos	<i>Sin recursos corrientes</i> 118 puntos	<i>Sin recursos</i> 69 puntos
<p>Condición de vida</p> <ul style="list-style-type: none"> ✓ Con privación de recursos <i>patrimoniales</i> ✓ Sin cobertura de salud <p>Instrucción del Jefe</p> <ul style="list-style-type: none"> ✓ Sin instrucción - Nivel primario <p>Condición de Actividad</p> <ul style="list-style-type: none"> ✓ Trabajador familiar con sueldo <p>Estructura etaria</p> <ul style="list-style-type: none"> ✓ Población envejecida, varones de 50 a 64 años <p>Var. departamentales</p> <ul style="list-style-type: none"> ✓ No caracterizan el grupo 	<p>Condición de vida</p> <ul style="list-style-type: none"> ✓ Con privación de recursos <i>corrientes</i> ✓ Sin cobertura de salud <p>Instrucción del Jefe</p> <ul style="list-style-type: none"> ✓ Nivel primario completo; secundario incompleto - completo <p>Condición de Actividad</p> <ul style="list-style-type: none"> ✓ Desocupado <p>Estructura etaria</p> <ul style="list-style-type: none"> ✓ Población joven, alta relación niños población activa <p>Var. departamentales</p> <ul style="list-style-type: none"> ✓ No caracterizan el grupo 	<p>Condición de vida</p> <ul style="list-style-type: none"> ✓ Con privación de recursos <i>convergente</i> ✓ Sin cobertura de salud - NBI <p>Instrucción del Jefe</p> <ul style="list-style-type: none"> ✓ Nivel primario <p>Condición de Actividad</p> <ul style="list-style-type: none"> ✓ Desocupado – Empleado público <p>Estructura etaria</p> <ul style="list-style-type: none"> ✓ Población joven, alta relación niños población activa <p>Variables departamentales</p> <ul style="list-style-type: none"> ✓ Gasto: social, educación y público per cápita.
<i>Nivel medio bajo</i> 37 puntos	<i>Nivel medio</i> 185 puntos	<i>Nivel medio alto</i> 71 puntos
<p>Condición de vida</p> <ul style="list-style-type: none"> ✓ Tenencia de la vivienda “no propietarios” ✓ Con cobertura de salud ✓ Con artefactos de confort <i>estandar</i> <p>Instrucción del Jefe</p> <ul style="list-style-type: none"> ✓ Nivel primario completo; secundario incompleto <p>Condición de Actividad</p> <ul style="list-style-type: none"> ✓ Ocupado – Empleado s. privado <p>Estructura etaria</p> <ul style="list-style-type: none"> ✓ Población joven, niños menores de 12 años y de 24 a 35 años <p>Var. departamentales</p> <ul style="list-style-type: none"> ✓ Gasto en educación per cápita y en caminos 	<p>Condición de vida</p> <ul style="list-style-type: none"> ✓ Tenencia de la vivienda “propietarios” ✓ Con cobertura de salud ✓ Con artefactos de <i>mayor</i> confort <p>Instrucción del Jefe</p> <ul style="list-style-type: none"> ✓ Nivel terciario – universitario incompleto <p>Condición de Actividad</p> <ul style="list-style-type: none"> ✓ Ocupado – Jubilado - Estudiante <p>Estructura etaria</p> <ul style="list-style-type: none"> ✓ Población envejecida, alta relación de adultos de 65 y más años respecto de la población activa <p>Var. departamentales</p> <ul style="list-style-type: none"> ✓ Cantidad de delitos por policía ✓ Participación en el PBG ✓ Inversión en caminos, viviendas, pública, obras hidráulicas ✓ Gasto en vivienda y educación per cápita 	<p>Condición de vida</p> <ul style="list-style-type: none"> ✓ PC con y sin Internet ✓ TE fijo y celular ✓ Con cobertura de salud <p>Instrucción del Jefe</p> <ul style="list-style-type: none"> ✓ Nivel secundario – terciario/universitario incompleto o completo <p>Condición de Actividad</p> <ul style="list-style-type: none"> ✓ Ocupado – Estudiante <p>Estructura etaria</p> <ul style="list-style-type: none"> ✓ Mayor proporción de mujeres de 36 a 49 años y de 50 a 64 años <p>Variables departamentales</p> <ul style="list-style-type: none"> ✓ Cantidad de delitos por policía ✓ Cantidad de locales de producción de bienes y servicios; hasta cinco empleados ✓ Inversión en edificios públicos

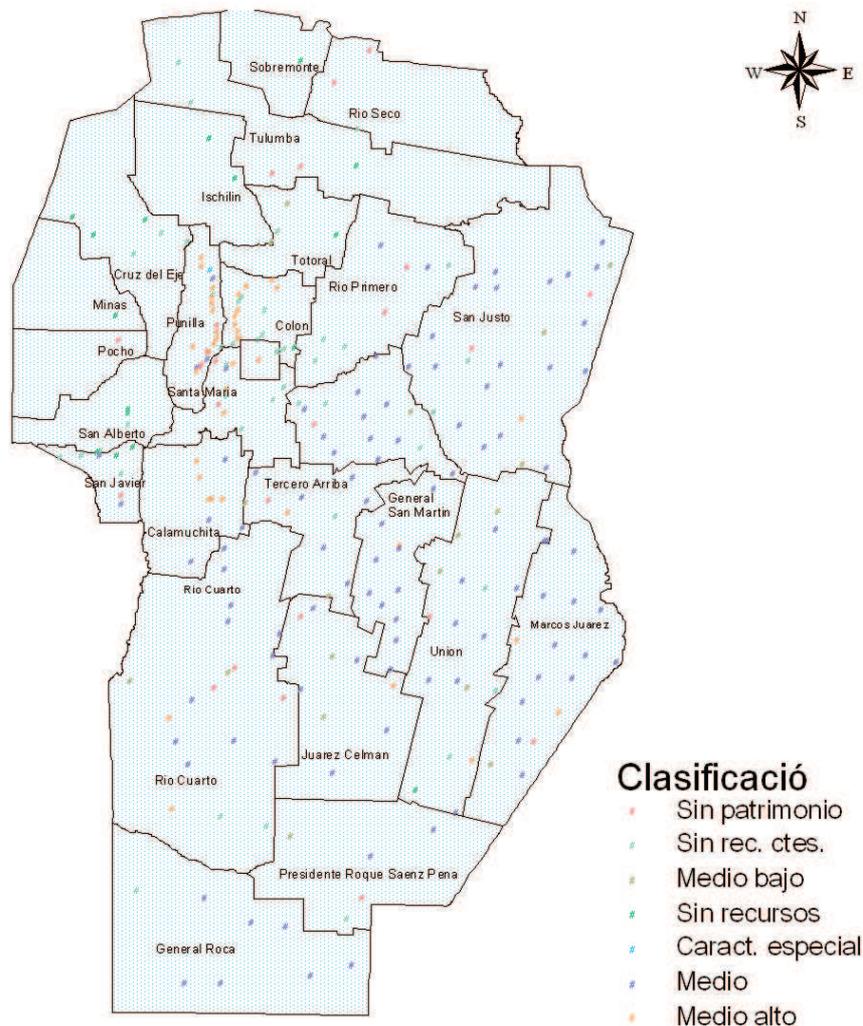
A las seis clases anteriores deben agregarse dos particiones, compuestas por sólo tres localidades cada una, con características muy particulares.

Villa San Miguel – Calamuchita Villa Elisa – Marcos Juárez Las Oscuras – San Alberto	Arroyo San Antonio – Calamuchita Villa El Tala – Calamuchita Las Jarillas – Punilla
<ul style="list-style-type: none"> ✓ Población muy envejecida: muy alta la proporción de adultos mayores de 65 años. ✓ No propietarios de la vivienda ✓ Con NBI por condiciones sanitarias 	<ul style="list-style-type: none"> ✓ Muy alta relación de masculinidad ✓ Nivel de instrucción terciario – universitario completo ✓ Ocupados ✓ Trabajadores del sector privado

Departamento	S/Patrimonio	S/R. Ctes	S/Recursos	Medio bajo	Medio	Medio Alto	Caract. Especial	Total de localidades
Calamuchita	7	8		1	8	10	3	37
Capital		2				1		3
Colón	2	14	2		1	14		33
Cruz del Eje		3	15	1		2		21
General Roca	2	1	1	3	8			15
Gral. San Martín	1	1		1	15	2		20
Ischilín	2	6	6					14
Juárez Celman	1	1		3	12	1		18
Marcos Juárez	2	1		1	17	2	1	24
Minas	1	3	5					9
Pocho	6	1	2					9
P. R. Sáenz Peña	1	3		1	6			11
Punilla	5	5			5	17	1	33
Río Cuarto	3	6		2	19	3		33
Río Primero	10	7	3	2	6			28
Río Seco	3	5	6					14
Río Segundo	1	4		1	16			22
San Alberto		10	7	1		3	1	22
San Javier	1	6	8	1	5			21
San Justo	3	4	1	7	29	1		45
Santa María	4	14	2	2	6	11		39
Sobremonte		2	2	1		1		6
Tercero Arriba	1	1		2	13	1		18
Totoral	3	4	3	2	1			13
Tulumba	3	3	5			1		12
Unión	2	3	1	5	18	1		30
Total por grupo	64	118	69	37	185	71	3	550

Tabla N° 2: Tabla resumen de la distribución de localidades de cada grupo.

La siguiente figura ilustra el mapa de la Provincia de Córdoba, donde se representa la distribución de las localidades que conforman cada grupo.



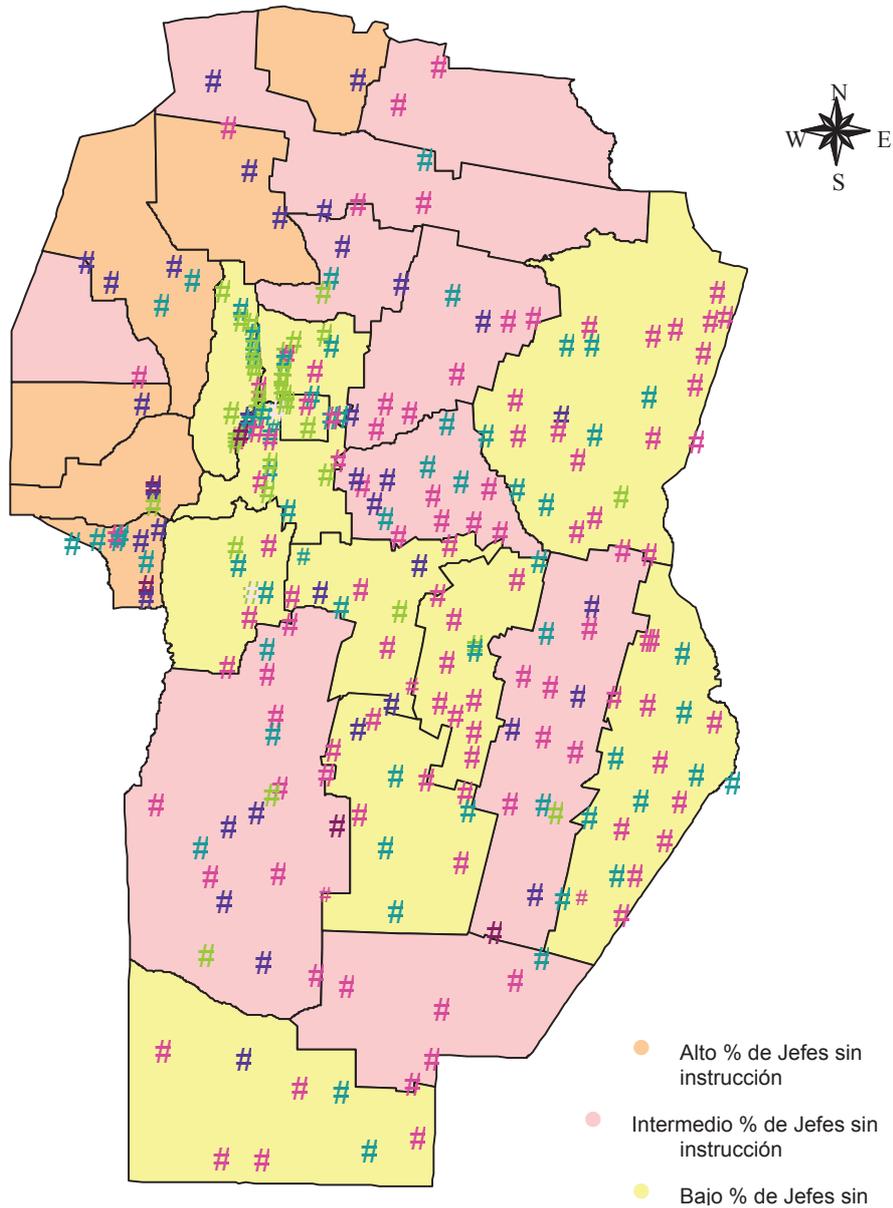
Mapa N° 3: Provincia de Córdoba con la distribución de localidades de cada grupo.

A continuación se realiza un análisis de asociación espacial, que evalúa la hipótesis de aleatoriedad espacial comparando los valores de los indicadores “NBI por condiciones sanitarias”, “NBI privación convergente” y “Nivel de instrucción Bajo del Jefe de Hogar” (sin instrucción o primario incompleto) en cada municipio con los valores correspondientes a los municipios vecinos. El resultado indica que no hay un comportamiento aleatorio, sino por el contrario existe algún tipo de dependencia.

	NBI por condiciones sanitarias	NBI por privación convergente	Nivel de Instrucción de Jefe “Bajo”
I de Moran	0.244	0,112	0.393
z-Normal I	3,395	1,588	5,450

Tabla N° 8: Resultados de la prueba de autocorrelación espacial

El cuadro anterior muestra una autocorrelación positiva y estadísticamente significativa, es decir una tendencia a la concentración espacial de los municipios. Los valores de z superiores a 2,58 indican que con un nivel de confianza del 99%, la concentración no es aleatoria, sino que por el contrario el nivel de los indicadores NBI de cada municipio depende no solamente de sus condiciones propias sino además de la situación de sus vecinos inmediatos.



Mapa N° 4: Provincia de Córdoba con la distribución de los departamentos según el porcentaje de jefes con nivel de instrucción "Bajo"

VI. Conclusiones

De acuerdo a los resultados obtenidos se puede afirmar que existe dependencia espacial de los indicadores de pobreza, como son las necesidades básicas insatisfechas y el bajo nivel de instrucción del jefe de hogar en la Provincia de Córdoba. Es posible afirmar que la ubicación geográfica juega un papel fundamental, es decir, la pobreza de un municipio depende en forma importante del nivel de pobreza de los municipios vecinos.

Otro resultado, no menos importante, fue la identificación de clusters o conjunto de municipios con características similares. La primera partición se logra identificando localidades que denominamos *pobres* y *ricas*. A partir de estas dos grandes *clases* se pueden generar divisiones que logran mayor homogeneidad en la conformación de los grupos. Un análisis detallado de los municipios que pertenecen a cada uno de los departamentos de la provincia permitió establecer regiones, según la categorización de cada uno de ellos.

La metodología utilizada en el análisis de los datos, constituye un aporte importante en el procesamiento de información para la toma de decisiones. Por ejemplo, en términos de política económica, estos resultados adquieren gran importancia, ya que permiten lograr una adecuada focalización de políticas de distribución de recursos y mejorar la recaudación impositiva. Un municipio pobre rodeado de otros municipios pobres requiere de un esfuerzo adicional para superar la situación, que si sus vecinos fueran municipios ricos.

Es necesario tomar conciencia de que la situación de un municipio no sólo depende de sus propios indicadores económicos y sociales, sino que además está afectado por la situación de sus vecinos.

VII. Bibliografía

- (1) BRYAN, Manly. Multivariate statistical methods: A primer, Chapman and Hall, Londres, 1986.
- (2) CAMISA, Zulma. Introducción al estudio de la Fecundidad, CELADE. 1982.
- (3) CAMPOY, Luis. Estratificación ocupacional de la población censada (1960) de la Argentina, por medio de variables ocupacionales y el complemento del nivel educacional – U.N. Cuyo - 1988
- (4) CEPAL-CELADE-BID. Impacto de las tendencias demográficas sobre los sectores sociales en América Latina, Santiago, Chile, 1996.
- (5) CONSEJO FEDERAL DE INVERSIONES (CFI). Estructura social de la Argentina. Indicadores de la Estratificación Social y de las Condiciones de Vida de la Población en base al Censo de Población y Vivienda de 1980, Buenos Aires, 1989.
- (6) CONSEJO FEDERAL DE INVERSIONES (CFI). Díaz, Cecilia. Estratificación socio-demográfica aplicando técnicas de análisis multivariadas”. 2005.
- (7) DIAZ, Margarita. Discusión y aplicación de métodos multivariados en investigaciones sociales. Tesis doctoral, Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba. Córdoba, 1998.
- (8) HAIR Joseph y otros. Análisis multivariante, Prentice Hall Iberia, 5ª edición en español. Madrid, 1999.

- (9) INDEC. Mapas de la pobreza en Argentina – Comité Ejecutivo para el Estudio de la Pobreza en Argentina (CEPA) – Documento de trabajo de trabajo N° 4 – Marzo 1994
- (10) LEBART, L. Tratamiento estadístico de datos. Marcombo, Barcelona, 1985.
- (11) LEBART, L. Statistique exploratoire multidimensionnelle, Dunod. París, 1995.
- (12) LISTA, Carlos A. Cuadernos de Sociología, Atenea, Córdoba, 1992.
- (13) PASSANANTE, María Inés. La pobreza en 27 ciudades argentinas. Boletín de Lecturas Sociales y Económicas, Año 6, N° 29, 1999.
- (14) PNUD. Informe de Desarrollo Humano, 1996
- (15) SHARMA, Subhash, Applied Multivariate Techniques, J. Wiley & Sons. Nueva York, 1996.
- (16) TORRADO, Susana. Información e investigación sociodemográfica en América Latina, CLACSO en Santiago de Chile, 1978.
- (17) TORRADO, Susana. Estructura Social Argentina:1945-1983, Ediciones de la Flor, Buenos Aires, 1992.
- (18) TORRADO, Susana. Procreación en la Argentina. Hechos e ideas, Ediciones de la Flor, Buenos Aires, 1993.
- (19) TORRADO, Susana. Población y desarrollo en la Argentina – En busca de la relación perdida, Comisión Familia y Minoridad del Honorable Senado de la Nación Argentina, 1999.

Autora

Currículum Vitae

- Contadora Público Nacional
- Magíster en Demografía
- Directora Centro de Computación y Tecnologías de Información de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba
- Profesora Adjunta Cátedra Sistemas de Información y Procesamiento de Datos I de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba
- Investigadora del Instituto de Estadística y Demografía de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba