

Uso de SIG para la toma de decisiones

NANCY STANECKA ¹ y CECILIA DÍAZ ¹²

nstanec@gmail.com

RESUMEN

El presente trabajo tiene por objetivo mostrar el avance del Proyecto de Investigación “Observatorio Geo-Estadístico de la Ciudad de Córdoba” subsidiado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Córdoba, utilizando herramientas de software libre, con datos georreferenciados, también de dominio público disponibles en el portal web del Gobierno de la Provincia de Córdoba en la Dirección de Estadísticas y Censos.

En una primera etapa se realiza un análisis exploratorio a través del mapeo de los datos del Censo de población del año 2008 de la ciudad de Córdoba, usando un esquema de clasificación y colores que permiten una visualización clara de algunas de las características importantes que los mapas presentan. Luego de construir mapas temáticos, en una segunda etapa, esta investigación se propone analizar la auto-correlación espacial de variables socio-económicas y demográficas.

Se utilizaron dos herramientas de software libre: 1) para la construcción de mapas gvSIG (<http://www.gvsig.org/web/>); y 2) se indaga qué variables evidencian correlación espacial a través del análisis Geo-estadístico usando R, un lenguaje y entorno de programación de acceso libre para análisis estadístico y gráfico (<http://www.r-project.org/>).

Una conclusión importante de esta investigación es haber encontrado una metodología que ayudará en el proceso de diseño de políticas públicas y de toma de decisiones empresariales, que en estos tiempos modernos se vuelven cruciales.

Palabras Claves: gvSIG; Lenguaje R; Correlación espacial; Datos socio-demográficos

¹ Instituto de Estadística y Demografía, Facultad de Ciencias Económicas, UNC

² Centro De Cómputos, Facultad de Ciencias Económicas, UNC

INTRODUCCIÓN

La distribución espacial de la población constituye una temática siempre vigente para diferentes disciplinas de las ciencias sociales. Conocer con alguna certeza la distribución de la población en el espacio físico es de vital importancia tanto para la elaboración de políticas sectoriales y globales de desarrollo económico y social –en este sentido, es imprescindible contar con información acerca de la distribución de las actividades productivas, la cobertura de servicios básicos de salud y educación, la estructura del empleo, la localización de unidades productivas, de servicios, de comercio, etc.–, como para comprender el rol que puede desempeñar la distribución de los grupos sociales en la reproducción de las desigualdades sociales (Toro Labe, 1996).

Un Sistema de Información Geográfica (SIG) es una herramienta informática que sirve para ingresar, almacenar, gestionar, recuperar, actualizar, analizar y producir información, asociada a lugares o zonas geográficas. En otras palabras, un SIG permite responder preguntas acerca de dónde se encuentran ciertas cosas o acerca de qué cosas se encuentran en una ubicación dada.

Más precisamente, un SIG funciona como una base de datos con información geográfica (datos alfanuméricos), que se encuentra asociada por un identificador común a los objetos gráficos de un mapa digital. De esta forma, señalando un objeto se conocen sus atributos, y a la inversa, preguntando por un registro de la base de datos se puede saber su localización en la cartografía. De esta forma se hace posible mostrar la información en capas temáticas para realizar análisis multicriterio complejos.

El presente trabajo tiene por objetivo mostrar el avance del Proyecto de Investigación “Observatorio Geo-Estadístico de la Ciudad de Córdoba” subsidiado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Córdoba, utilizando herramientas de software SIG de acceso libre, con datos georreferenciados, también de dominio público disponibles en el portal web del Gobierno de la Provincia de Córdoba en la Dirección de Estadísticas y Censos. Los resultados obtenidos constituyen un instrumento de gran valor para el análisis de múltiples problemáticas y la elaboración de políticas públicas, pues por un lado las estadísticas demográficas, sociales y económicas constituyen la base de la planificación y la gestión de políticas públicas, y por otro lado la distribución espacial de los indicadores socioeconómicos sirve de guía cuando se toman decisiones en materia de desarrollo regional, la prestación de servicios y en muchas otras esferas.

En una primera etapa se realiza un análisis exploratorio a través del mapeo de los datos del Censo de población de 2008 de la ciudad de Córdoba, usando un esquema de clasificación y colores que permiten una visualización clara de algunas de las características importantes que los mapas presentan. Luego de construir mapas temáticos, en una segunda etapa, esta investigación se propone analizar la auto-correlación espacial de variables socio-económicas y demográficas.

Se utilizaron dos herramientas de software libre: 1) para la construcción de mapas gvSIG (<http://www.gvsig.org/web/>); y 2) se indaga qué variables evidencian correlación espacial a través del análisis Geo-estadístico usando R, un lenguaje y entorno de programación de acceso libre para análisis estadístico y gráfico (<http://www.r-project.org/>).

2. MATERIALES Y METODOS

Datos:

La unidad de análisis es el radio censal, definido como subdivisiones territoriales de la ciudad de Córdoba, cuya función es catastral.

Los datos corresponden al censo provincial de población y vivienda del año 2008, a partir del cual se diseñó una “Base Usuaría” con información agregada referida a los radios censales.

VARIABLES

Las variables se definen a nivel de hogares o de población. A continuación se mencionan las seleccionadas:

- Porcentaje de hogares NBI.
- Porcentaje de hogares con privación material patrimonial.
- Porcentaje de hogares con privación material recursos corrientes.
- Porcentaje de mujeres desocupadas.
- Porcentaje de hombres desocupados.
- Porcentaje de población sin cobertura de salud

Análisis de datos

Aquí se trabajará con los denominados “datos espaciales” que son aquellos que surgen de mediciones u observaciones realizadas en localizaciones o regiones determinadas, siendo dicha situación geográfica parte de la información (Anselin, 1994). En función de la variable respuesta y del tipo de localización de los datos (puntos o áreas), los *datos espaciales* se clasifican en tres grandes grupos: datos de área (*lattice data*), datos continuos espaciales (*geostatistical data*) y proceso aleatorio espacial (Cressie, 1993).

En particular, los *lattice data*, provienen de un proceso espacial $\{Z(s_i) : s_i \in R\}$, donde R es una colección contable de regiones espaciales, que pueden estar distribuidas en

forma regular o irregular, e incorpora además la información sobre los vecinos cercanos (Faes y Abrahantes, 2007). A esta estructura responde la información de este trabajo, con tasas a nivel de radio censal.

En estos datos el supuesto de independencia entre mediciones que se asume en el modelo lineal clásico es cuestionable debido a la presencia de observaciones que pueden estar auto-correlacionadas, es decir que existe un patrón sistemático en la distribución espacial de la variable en estudio. Si la correlación espacial existe, es posible diferenciar entre dos componentes de variación: la variación a gran escala, que informa acerca de la tendencia global y la variación a pequeña escala, que representa la correlación entre los términos de error (Cressie, 1993; Lawson, 2001). Para cuantificar esa correlación se han desarrollado métodos estadísticos específicos que incluyen las coordenadas de la ubicación geográfica de cada observación. Una de las herramientas más usadas para medir esta dependencia, en un espacio discreto, es el Índice de Moran, que mide en forma global la presencia o ausencia de autocorrelación espacial en un conjunto de datos distribuidos sobre una superficie (Moran, 1948). Este índice se calcula como:

$$I = \frac{n \sum_i \sum_j W_{ij} (Z_i - \hat{Z})(Z_j - \hat{Z})}{(\sum_i \sum_j W_{ij}) \sum_i (Z_i - \hat{Z})^2}$$

donde:

Z_i es la respuesta en la localización i -ésima ($i: 1, 2, \dots, n$)

W_{ij} medida de la proximidad espacial de las localizaciones i y j .

En este trabajo se plantea que la intensidad de la relación es inversamente proporcional a la distancia euclídea base en latitud y longitud de cada radio censal, indicando la no constancia de dicha intensidad entre distintos radios pedanías. Esto es,

$$W_{ij} = \begin{cases} 0 & \text{si } i = j \\ \frac{1}{d_{ij}} & \text{si } i \neq j \end{cases}$$

A los fines de remover el *efecto de la tendencia global (variación a gran escala)* se aplica un modelo de regresión, considerando como covariables latitud y longitud.

El modelo se presenta como:

$$Z(s) = \mu(s) + \varepsilon(s)$$

donde el primer término representa el efecto general más el efecto de latitud y longitud y el segundo, el residuo. Luego de remover la tendencia global, se calcula nuevamente el Índice de Moran sobre los residuos obtenidos a través del modelo de regresión a fin de determinar si existe o no *dependencia espacial a pequeña escala*.

3. RESULTADOS

Para la construcción de mapas temáticos se utilizó la herramienta gvSIG, un potente Sistema de Información Geográfica (SIG) libre diseñado para dar solución a todas las necesidades relacionadas con el manejo de información geográfica. Se caracteriza por ser una solución completa, fácil de usar y que se adapta a las necesidades de cualquier usuario de SIG. Es capaz de acceder a los formatos más comunes, tanto vectoriales como raster, , integra estándares OGC, y cuenta con un amplio número de herramientas para trabajar con información de naturaleza geográfica (consulta, creación de mapas, geoprocetamiento, redes, etc.) que lo convierten en una herramienta ideal para usuarios que trabajen con la componente territorial.

gvSIG es utilizado por una comunidad internacional creciente, en los más diversos sectores y aplicaciones.

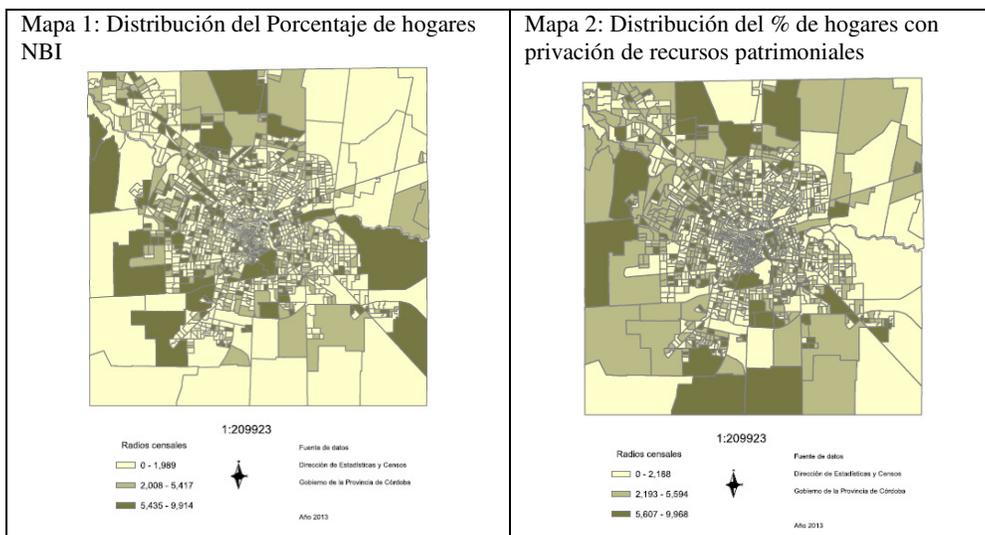
Algunas de sus **características más destacadas** son:

- **Portable:** funciona en distintas plataformas hardware / software, Linux, Windows y Mac OS. El lenguaje de programación es Java.
- **Modular:** es ampliable con nuevas funcionalidades mediante el desarrollo de extensiones, permitiendo una mejora continua de la aplicación, así como el desarrollo de soluciones a medida.

- **De código abierto:** licencia GNU/GPL, lo que permite su libre uso, distribución, estudio y mejora.
- **Interoperable** con las soluciones ya implantadas: es capaz de acceder a los datos de otros programas privativos, como ArcView, AutoCAD o Microstation sin necesidad de cambiarlos de formato.
- **Internacionalizable:** está disponible en más de una veintena de idiomas (castellano, inglés, alemán, italiano) y permite la incorporación de nuevos idiomas con facilidad.
- **Sujeto a estándares:** sigue las directrices marcadas por el Open Geospatial Consortium (OGC).

Esta herramienta se puede descargar gratuitamente de la página <http://www.gvsig.org/web/>.

A continuación se presentan los mapas que muestran la distribución espacial de cada una de las variables:



Mapa 3: Distribución del Porcentaje de Hogares con Privación Recursos Corrientes



1:209923

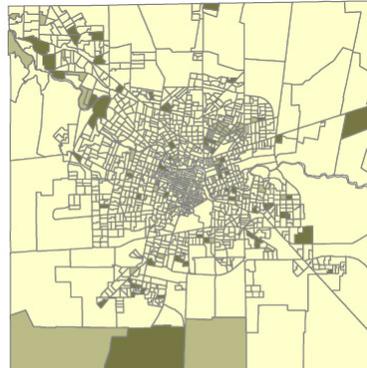
Radiocensales

- 0 - 2,113
- 2,116 - 5,195
- 5,213 - 9,964

Fuente de datos
Dirección de Estadísticas y Censos
Gobierno de la Provincia de Córdoba

Año 2013

Mapa 4: Distribución del Porcentaje de Población SIN Cobertura de Salud



1:209923

Radiocensales

- 0 - 2,78
- 3,226 - 7,5
- 7,635 - 9,979

Fuente de datos
Dirección de Estadísticas y Censos
Gobierno de la Provincia de Córdoba

Año 2013

Mapa 5: Distribución del Porcentaje de Hombres Desocupados



1:209923

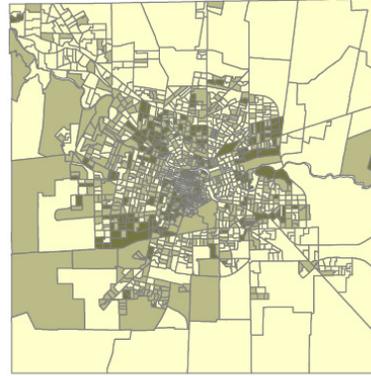
Radiocensales

- 0 - 3,462
- 3,468 - 6,073
- 6,081 - 9,865

Fuente de datos
Dirección de Estadísticas y Censos
Gobierno de la Provincia de Córdoba

Año 2013

Mapa 6: Distribución del Porcentaje de Mujeres Desocupadas.



1:209923

Radiocensales

- 0 - 8
- 9 - 14
- 15 - 32

Fuente de datos
Dirección de Estadísticas y Censos
Gobierno de la Provincia de Córdoba

Año 2013

Los porcentajes de hogares con NBI y con privación de recursos patrimoniales tienen una distribución similar, los niveles más altos se insertan en sectores de nivel socioeconómico medio o alto, en forma de puntos o pequeñas áreas tal como puede observarse en el Mapa N° 1 y N° 2. En general, los individuos sin recursos económicos viven en lugares cercanos a las fuentes de trabajo.

En el Mapa N° 4 puede observarse la distribución de la población sin cobertura de salud. Casi la totalidad de la ciudad tiene un nivel bajo de individuos sin acceso a la salud, sólo en sectores puntuales se manifiesta esta problemática.

En tanto que en el Mapa N° 5 y N° 6 se distingue una distribución diferente del fenómeno de la desocupación según el género.

Luego de la construcción y análisis de los mapas temáticos se procedió a calcular la correlación espacial, a través del índice de Moran. En este caso se utilizó el lenguaje **R** para análisis estadístico y gráfico, que se distribuye bajo la licencia GNU GPL. El código fuente está escrito principalmente en C y algunas rutinas en Fortran, y está disponible para máquinas Unix y Linux, o como archivos binarios pre-compilados para Windows, Linux (Debian, Mandrake, RedHat, SuSe), Macintosh y Alpha Unix.

Esta herramienta tiene una naturaleza doble de programa y lenguaje de programación, además permite cargar diferentes bibliotecas o paquetes con finalidades específicas de cálculo o gráfico. En este caso se utilizó la librería “*ape*”.

En la Tabla 1 se presentan las estimaciones del Índice de Moran, donde se da cuenta de la dependencia espacial global en los porcentajes de hogares con NBI, con privación de recursos, tanto patrimoniales como corrientes, al igual que en los porcentajes de hombres y mujeres desocupadas, y la proporción de población sin cobertura de salud.

Tabla 1: Estimaciones del Índice de Moran para las variables seleccionadas

| Variable | Indice de Moran | Varianza | Valor <i>p</i> |
|---|-----------------|----------|----------------|
| % de hogares NBI | 0,079 | 0,000007 | 0.000 |
| % hog. c/ privación recursos patrimonial | 0,094 | 0,000007 | 0.000 |
| % de hog. c/ privación recursos corriente | 0,130 | 0,000007 | 0.000 |
| % de mujeres desocupadas | 0,113 | 0,000007 | 0.000 |
| % de hombres desocupados | 0,025 | 0,000007 | 0.000 |
| % de población sin cobertura de salud | 0,155 | 0,000007 | 0.000 |

Detectada la autocorrelación espacial, se procedió al planteo de una regresión donde las variables independientes consideradas son latitud y longitud. Debido a que estas últimas asumen valores muy elevados, se realizó un cambio de origen de las mismas, lo que no

cambia la significatividad de las variables en la regresión. Analizando los resultados presentados en la Tabla 2, se observa que, la longitud resulta significativa para todas las variables menos para el % de hogares con privación patrimonial. A medida que nos desplazamos hacia el este de la ciudad disminuye el % de hogares con NBI, el % de hogares con privaciones de recursos corrientes, el % de desocupados y el % de población sin cobertura de salud. Por su parte la latitud resulta significativa en el porcentaje de hogares con privación patrimonial y en el % de mujeres desocupadas. (considerando un nivel de significación del 0.10, las cuales se destacan con (*)).

Tabla 2: Coeficientes de la regresión y p-value para cada variable considerada

| Variable | Ordenada al origen | Latitud | Longitud |
|---|--------------------|------------|------------|
| % de hogares NBI | 6,128 | 0,101 | -0,361 (*) |
| % hog. c/ privación recursos patrimonial | 7,301 | 0,633 (*) | -0,132 |
| % de hog. c/ privación recursos corriente | 5,665 | -0,237 | -1,196 (*) |
| % de mujeres desocupadas | 9,930 | -0,227 (*) | -0,624 (*) |
| % de hombres desocupados | 4,604 | -0,054 | -0,360 (*) |
| % de población sin cobertura de salud | 28,154 | 0,016 | -0,991 (*) |

En el Tabla 3 se muestra las estimaciones del Índice de Moran obtenidas a partir de los residuos de la regresiones. Puede observarse que habiendo descontado la variación a gran escala, aún persiste dependencia espacial a pequeña escala (efecto vecino) para todas las variables en estudio.

Tabla 3: Estimaciones de los Índices de Moran para residuos de regresión y p-value para las variables seleccionadas.

| Variable | Índice de Moran | Varianza | Valor p |
|---|-----------------|----------|-----------|
| % de hogares NBI | 0,027 | 0,001179 | 0,000 |
| % hog. c/ privación recursos patrimonial | 0,032 | 0,001179 | 0,000 |
| % de hog. c/ privación recursos corriente | 0,066 | 0,001179 | 0,000 |
| % de mujeres desocupadas | 0,044 | 0,001179 | 0,000 |
| % de hombres desocupados | 0,012 | 0,001179 | 0,000 |
| % de población sin cobertura de salud | 0,059 | 0,001179 | 0,000 |

4. CONCLUSIONES

En los últimos años, dada la disponibilidad de información georeferenciada y de herramientas SIG, se han desarrollado métodos que tienen por objetivo incluir la dependencia espacial en la modelación. En este trabajo, se utilizaron herramientas de software libre: primero, para el análisis exploratorio se utilizó gvSIG que permitió la construcción de mapas temáticos; segundo, a través de R se calculó el Índice de Moran con dos conjuntos de datos: variables originales y residuos de regresión sobre latitud y longitud. Los resultados muestran dependencia espacial en gran escala y en pequeña escala en todas las variables sociodemográficas analizadas, en la ciudad de Córdoba.

Creemos en la importancia de poder reproducir este esquema metodológico para aplicarlo a otras variables y/ o regiones pudiendo incluso extender la aplicación de la misma a toda la provincia.

El aporte fundamental es haber encontrado una metodología que ayudará en el proceso de diseño de políticas públicas y de toma de decisiones, que en estos tiempos modernos se vuelven cruciales.

5. BIBLIOGRAFIA

Anselin, L. (1995) *Local indicators of spatial associations-LISA*. Geographical Analysis, Vol. 27:93-115.

Burrough, P. (1988): Principles of Geographical Information Systems for land resources assessment. Oxford University Press, Oxford. P 80

Cressie, N, (1993) *Statistics for spatial data*, New York: Wiley.

Druck, S., Carvalho M., Câmara G. and Monteiro A., 2004. *Spatial Analysis of Geographical Data*. Embrapa Publisher, Atibaia, Brazil, pp: 209.

Faes, C, y Abrahantes J, (2007) *Desease Mapping*, material de estudio Master en Biostatistics

López Hernández, F. y Palacios Sánchez, M (2000) *Distintos modelos de dependencia espacial. Análisis de autocorrelación*. XIV Reunión Asepelt. España.

Moran, P.A.P. The interpretation of Statistical Maps, *Biometrika* 35, 255-260. 1948.
Ripley, B (1981) *Spatial Statistics*, New York: Wiley,