

PARTICULARIDADES DE LA INFORMACIÓN ESPACIAL EN LOS RADIOS CENSALES

RESUMEN

La creciente disponibilidad de datos georreferenciados en ámbitos urbanos en los últimos años ha potenciado los análisis a escala de radio censal. Se los puede elaborar a través de sistemas de información geográfica o acceder a ellos a través de geoportales. No obstante, los datos agregados que conforman los radios censales presentan ciertas dos particularidades que deben ser tenidas en cuenta para no llegar a conclusiones erróneas: el Problema de la Unidad Espacial Modificable que ocurre cuando al promediar los datos al momento de pasar de una escala mayor a una menor; y la autocorrelación espacial en donde

Celemín, Juan Pablo

IGEHCS-UNCPBA.

jpcelemin@conci.et.gov.ar

las unidades espaciales vecinas tienden a mostrar un comportamiento similar. Por medio de distintos ejemplos estadísticos y el uso de sistemas de información geográfica se muestra el alcance de estas propiedades de la información georreferenciada a escala intra-urbana para conocer su impacto en los resultados de trabajos relacionados con la planificación urbana y en la toma de decisiones.

Palabras clave: información georreferenciada; Problema de la Unidad Espacial Modificable; autocorrelación espacial; radio censal.

INTRODUCCIÓN

La mayoría de los datos que se utilizan para los estudios urbanos poseen un sustento territorial, es decir, que los podemos ubicar en el espacio a través de coordenadas. Otras disciplinas utilizan indicadores que no se ubican en el territorio (por ejemplo, la psicología recurre a indicadores para estudiar el comportamiento de las personas. Para ello no necesita saber dónde se ubican espacialmente). Cuando una variable se localiza en el espacio adopta ciertas particularidades que, si no son conocidas, pueden llevar al investigador a la obtención de conclusiones erróneas. En el presente trabajo se considerarán las implicancias del Problema de la Unidad Espacial Modificable (PUEM) y la Autocorrelación Espacial (AE) en los datos espaciales a través de tablas, mapas y gráficos en diferentes niveles geográficos.

DESARROLLO

Problema de la Unidad Espacial Modificable

Existe un problema al utilizar unidades espaciales artificiales establecidas administrativamente por organismos oficiales que no suelen reflejar necesariamente de modo estricto los fenómenos sociales y territoriales existentes. A este fenómeno, típico de la Geografía y de los SIG se lo conoce como el Problema de la Unidad Espacial Modificable (PUEM). Es decir, que la división del territorio y los agrupamientos resultantes no son neutros (Velázquez, 2006). Consecuentemente, como unidad espacial conviene una mínima celdilla o cuadrícula espacial de tamaño tan pequeño como resulte conveniente y no

grandes polígonos, que agravarían el PUEM (Moreno Jiménez, 2003). Más conocida es la falacia ecológica que consiste en

Provincias	Generación de RSU	
	Per cápita	Total (miles ton/año)
	(kg / hab.día)	
Misiones	0.44	26
Entre Ríos	0.6	63
Chaco	0.61	88
Tierra del Fuego	0.64	90
Formosa	0.65	111
Catamarca	0.69	122
Jujuy	0.71	148
Tucumán	0.73	161
Salta	0.76	163
La Rioja	0.77	166
Santa Cruz	0.82	169
Santiago del Estero	0.83	178
Buenos Aires	0.83	226
Río Negro	0.86	232
Corrientes	0.87	255
Neuquén	0.92	261
Chubut	0.95	306
San Juan	0.96	316
La Pampa	0.98	369
Córdoba	1.05	678
Santa Fe	1.11	1204
San Luis	1.12	1235
Mendoza	1.15	1493
Ciudad de Buenos Aires	1.52	4268
Promedio	0.8570	513.6667

Tabla 1. Generación de residuos sólidos urbanos por provincia. *Fuente:* Estrategia Nacional para la Gestión Integral de Residuos Sólidos Urbanos (ENGIRSU) 2004.

un tipo de error en la interpretación de datos estadísticos, en el que se infiere la naturaleza de los individuos a partir de las estadísticas agregadas del grupo al que dichos individuos pertenecen. Es prácticamente lo mismo que el PUEM pero sin el componente espacial.

La mejor manera de conocer las consecuencias del PUEM es a través de los siguientes ejemplos que van de lo regional (provincias y regiones) a lo local (radios censales). Es común encontrar que los datos obtenidos están delimitados administrativamente (a nivel país, provincia, departamento). En algunos otros casos también se puede obtener información demarcada de manera más "natural" como puede ser una cuenca

hidrográfica.

A continuación hay dos tablas. En la primera está la generación de residuos sólidos urbanos (RSU, Per cápita y totales) por provincia:

A continuación lo que se hace es agregar los datos para llevarlos a una escala más chica (menor nivel de detalle), en este caso por región de acuerdo a los criterios de agrupación de provincias del INDEC. Por agregar refiero al procedimiento de promediar los datos para pasar de una escala más grande (provincias) a otra más chica (región):

Al comparar los promedios de ambas tablas se observa que aumentan levemente en la segunda en ambos casos (columnas Per cápita kg/hab. día y Total miles tn/año) a pesar de que se parte de los mismos datos. Esta particularidad va totalmente en contra de nuestra intuición ya que uno supondría que los valores deberían mantenerse siempre igual. Es por ello que este fenómeno es considerado un problema. Sin embargo rara vez es mencionado cuando se trabaja con datos ubicados espacialmente por dos motivos: por desconocimiento o porque hasta ahora no se le ha encontrado solución. En el ámbito de los Sistemas de Información Geográfica este problema es conocido pero no ocurre en otras disciplinas que trabajan con datos espaciales.

Otro ejemplo recurriendo otra vez al área urbana de Mar del Plata. A partir de imágenes satelitales se puede obtener un índice de vegetación (NDVI) que distingue las áreas con mayor cantidad de espacios verdes en el interior de la localidad. Los datos se presentan a escala de píxel. Pero si se desea hacer los resultados comparables

Región	Generación de RSU	
	Per cápita (kg / hab.día)	Total (miles ton /año)
Centro	1.02	1134.67
Patagonia	0.84	211.6
NEA	0.64	120.0
NOA	0.75	156.33
Cuyo	1.08	1014.67
Promedio	0.8641	527.4533

Tabla 2. Generación de residuos sólidos urbanos por región. *Fuente:* elaboración personal en base a datos de la Estrategia Nacional para la Gestión Integral de Residuos Sólidos Urbanos (ENGI RSU) 2004.

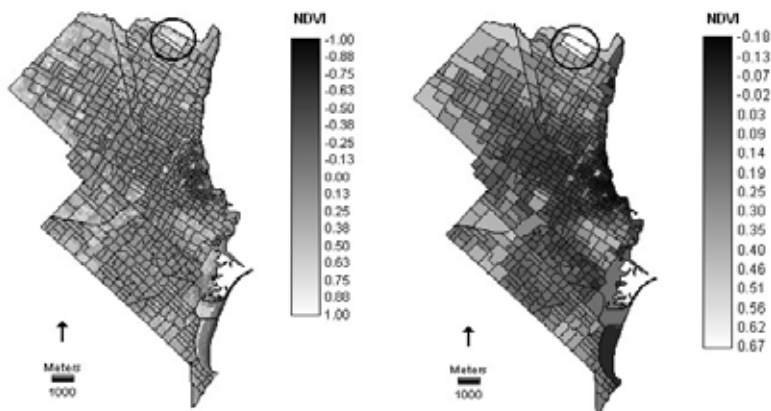


Figura 1. NDVI para la ciudad de Mar del Plata. *Fuente:* elaboración propia.

con otros datos es necesario convertirlos a escala de radio censal. El procedimiento consiste en pasar del formato *raster* propio de las imágenes satelitales a vectoriales, utilizados por los SIG. Para ello es necesario promediar todos los píxeles presentes dentro de cada radio para que éste adquiera un valor único al momento de hacer la conversión raster-vectorial.

En la imagen de la izquierda está el índice de vegetación (NDVI) a escala de píxel para luego ser agregados a escala de radio censal en la imagen de la derecha. Obsérvese como cambia el rango de valores una vez agregados los píxeles. Al ser promediados el rango es menor (-0.18 a 0.67). Cada vez que se promedian los datos extremos se pierden (-1 a 1 en este caso) y la distribución se vuelve más compacta. En el círculo negro se encuentra una zona residencial boscosa que, como es esperable, registra los valores mayores del índice de vegetación.

La principal implicancia estadística el fenómeno implica que si correlacionamos dos indicadores a medida que agregamos

los datos de manera sucesiva (por ejemplo de radio censal a fracción censal, de ésta a departamento, luego a provincia y finalmente a todo el país) la correlación entre ambos tenderá a ir en aumento. Si no tenemos en cuenta al PUEM puede suceder que los datos conduzcan a la obtención de conclusiones erróneas.

Por lo tanto, lo ideal es trabajar (si se consiguen) con los datos lo más desagregados posible. Entonces, si existe interés en el estudio de la generación de residuos sólidos en la Argentina lo mejor es recolectar los datos por departamento, y si están disponibles a una escala mayor (radio o fracción censal).

A continuación otro ejemplo a través de un gráfico para demostrar como la agregación de datos puede llevarnos a obtener conclusiones erróneas al momento de trabajar con variables cuantitativas de cualquier índole. Si tenemos dos variables, A y B, existe relación entre ellas si al aumentar los valores de A también lo hacen los de B, o por el contrario si al aumentar los valores de A disminuyen los de B. En

otras palabras, se considera que dos variables cuantitativas están relacionadas entre sí cuando los valores de una de ellas varían de forma sistemática con respecto a los valores homónimos de la otra.

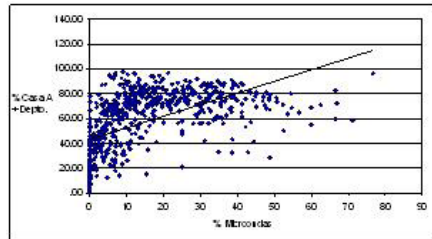
Existen varias maneras de medir esa relación. La más utilizada es por medio de un Coeficiente de Correlación Lineal de Pearson es un índice estadístico que permite medir la fuerza de la relación lineal entre dos variables. Su resultado es un valor que fluctúa entre -1 (correlación perfecta de sentido negativo) y +1 (correlación perfecta de sentido positivo). Cuanto más cercanos al 0 sean los valores, indican una mayor debilidad de la relación o incluso ausencia de correlación entre las dos variables. Por ejemplo: si correlacionamos las variables población con título universitario e ingreso veremos que nos da un valor de Pearson cercano a 1 ya que, como es esperable, las personas con mayor nivel educativo tienen la capacidad de acceder a mejores trabajos mejores pagos. En este caso la correlación es positiva porque al aumentar una variable también lo hace la otra. La situación opuesta ocurre con población con necesidades básicas insatisfechas e ingreso. Lógicamente a mayor nivel de ingreso menor va a ser las necesidades básicas insatisfechas que va tener un individuo o un hogar. En este caso la relación es negativa (con un valor cercano a -1) ya que a mayor incremento del ingreso menor son las necesidades básicas insatisfechas.

Una de las particularidades de la agregación es que al agregarlas (promediarlas) las variables tienden a parecerse cada vez más entre ellas. Esto se debe al efecto "suavizante" que ocurre al promediarse

los valores.

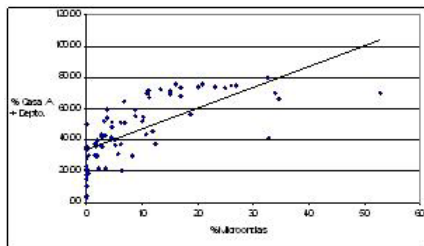
En este caso se tomaron los radios censales, fracciones y departamentos de la provincia del Chubut. El ejemplo parte de la suposición de que los hogares que poseen un horno microondas se encuentran en viviendas apropiadas, sin ningún tipo de déficit (en el censo de denominan técnicamente Casa Tipo A y Departamentos). Dado que el censo no interroga sobre el ingreso de los habitantes una manera efectiva de inferirlo es a través del equipamiento del hogar (tenencia de horno microondas, PC, heladera con freezer...)

A pesar de que el universo de análisis es siempre el mismo percibiremos como la correlación entre ambas variables aumenta



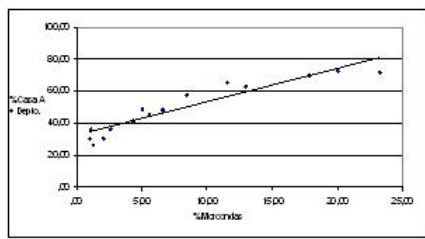
$r = 0,550$

Figura 2. Correlación a escala de radio censal (634 unidades de análisis).



$r = 0,706$

Figura 3. Correlación a escala de fracción censal (74 unidades de análisis).



$r = 0,941$

Figura 4. Correlación a escala departamental (15 unidades de análisis).

considerablemente a medida que se reduce la escala de análisis. En el eje de las Y está el % de viviendas apropiadas y en el de las X el % de hogares que poseen horno microondas. Mientras que r es el resultado de Pearson que nos muestra el grado de asociación entre las variables.

Llamativamente se observa como la correlación pasa de 0,550 cuando trabajamos con radios censales a 0,941 cuando lo hacemos a nivel departamental. Lo que sucede es que la famosa agregación. Una mala interpretación nos conduciría a suponer que porque la correlación es muy alta a escala departamental casi todas las viviendas apropiadas poseen horno microondas, pero si desagregamos la información a nivel censal la correlación es media. Esta ilusión estadística nos obliga a tomar recaudos al momento de analizar la información en diferentes escalas de análisis, que ha sido agregada de forma sucesiva.

AUTOCORRELACIÓN ESPACIAL.

Por su parte la **autocorrelación espacial (AE)** es la concentración o dispersión de los valores de una variable en el espacio. La AE refleja el grado en que objetos o actividades en una unidad geográfica son similares a

otros objetos o actividades en unidades geográficas próximas (Goodchild, 1987). Este tipo de autocorrelación prueba la primera ley geográfica de Tobler (1970) de que *"todo está relacionado con todo lo demás, pero que las cosas cercanas están más relacionadas que las cosas distantes"*.

Funciona de la siguiente manera: es muy raro encontrar cortes bruscos en la distribución espacial de una variable o indicador. Pero si sucede nos encontramos ante dos posibles escenarios.

Estamos en presencia de una observación atípica (también llamada *outlier*) que requiere de mayor atención por parte nuestra.

El dato ha sido mal ingresado

Dentro del círculo amarillo detectamos una radio censal de color blanco, con un valor muy bajo en relación a los de sus vecinos. En ese lugar hay un asentamiento precario que justamente bordea una de las zonas más acomodadas de la ciudad. Entonces tenemos un escenario tipo a. Los mapas y los gráficos son ideales para detectar este tipo de observaciones atípicas mientras que en las tablas pueden perderse entre la gran cantidad de datos. Exceptuando el *outlier* observamos una configuración espacial de la variable muy clara ya que desciende suave y progresivamente desde el centro de la ciudad a la periferia. No hay cortes abruptos en su distribución. Responde muy bien a la autocorrelación espacial.

Como se mencionó con anterioridad en el espacio todas las cosas están relacionadas, pero las cercanas están más relacionadas que las distantes. En otras palabras los valores de las unidades espaciales vecinas (radios censales en los ejemplos anteriores)

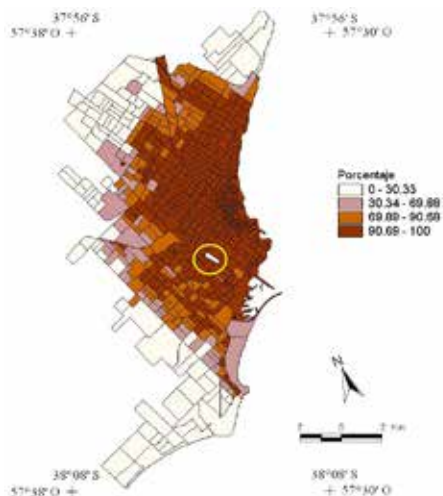


Figura 5. Porcentaje de población que reside en hogares con inodoro condescarga de agua y desagüe a red pública. Escala radio censal. Mar del Plata. *Fuente:* elaboración propia.

tenderán a ser similares. A pesar de la simplicidad del enunciado, éste tiene importantes implicancias estadísticas. En particular se destaca por no responder a los principios convencionales de la estadística inferencial en la cual a partir de una muestra se infiere el comportamiento del total de la población. Es un procedimiento muy utilizado cuando hay elecciones. Para saber cómo van a resultar el encuestador entrevista a personas (técnicamente llamadas unidades de observación) elegidas al azar para saber su intención de voto. Lo que cada individuo piensa no es influenciado por los otros ya que ni siquiera se conocen por ser elegidos aleatoriamente. Es decir que las unidades de observación son independientes entre sí. Pero en el territorio esto no ocurre ya que lo que acontece en un lugar, también repercutirá en el resto del área, especialmente en su entorno más cercano. Las características de cualquier unidad

espacial (radio censal por ejemplo) seguramente serán similares a la de sus radios censales vecinos. En otras palabras hay autocorrelación espacial. Esta particularidad también es conocida como un problema de dependencia espacial. Y es un problema justamente porque es contraria a lo que la estadística inferencial postula acerca de la independencia de las observaciones.

Para analizar esta propiedad específica de los datos espaciales se han desarrollado métodos para estudiarlos. Así es como la AE ha pasado de ser únicamente una particularidad de la información espacial a ser un índice estadístico. Existe varios de este tipo entre los que se destaca la I de Moran que se encuentra en la de los SIG actuales. Técnicamente, la I de Moran es el más conocido coeficiente de correlación de Pearson al que se le agrega una matriz de pesos definida por el usuario. Cabe destacar que la AE no es un procedimiento

estadístico más y es, quizás, después del promedio y la varianza la propiedad más importante de cualquier variable geográfica y, a diferencia de las anteriores, está explícitamente vinculada con patrones espaciales (Goodchild en Kemp, 2008)

CONCLUSIONES

El uso de datos georreferenciados está en auge a partir del creciente uso de tecnologías asociadas a su uso, representación cartográfica y difusión. Los geoportales y los sistemas de información geográfica han contribuido a su exposición tanto en el ámbito académico como en el resto de la sociedad. No obstante, es poco conocido que este tipo de datos posee ciertas particularidades que los diferencian de la información cuantitativa carente de componente espacial. En este trabajo a través del uso de tablas, gráficos y mapas se muestra sus dos principales particularidades: el Problema de la Unidad Espacial Modificable (PUEM) y la Autocorrelación Espacial (AE).

En el PUEM sucede que a medida que las unidades espaciales se agrupan para formar otras de mayor tamaño (de radios a fracciones, de éstas a municipios, y de éstos a provincias, por ejemplo), las correlaciones entre las variables estudiadas tienden a aumentar aunque el universo de observación sea siempre el mismo. Esto se debe a que la información al promediarse para pasar de un agrupamiento a otro se vuelve más homogénea. Lo particular del problema es que solamente sucede cuando el agrupamiento se produce entre unidades espaciales vecinas y no juntadas al azar.

Se han elaborado diferentes propuestas que abordan posibles soluciones, pero en

general existe cierto consenso acerca de que el PUEM no tiene solución a menos que se disponga de los datos individuales (Anselin 2000). No obstante, esta postura presenta dos limitaciones: la primera se refiere a la confidencialidad y privacidad de los datos y, en segundo lugar, se encuentra la dificultad de manejar una cantidad descomunal de registros que dificultaría el procesamiento estadístico en un potencial trabajo de investigación. Lo que sí está claro es que cualquier solución debe encontrar la forma de eludir la pérdida de información resultante de la agregación de datos.

Como detalla con precisión Bosque Sendra, (1995:15-16) cuando se trabaja con datos provenientes de los censos la mayoría de las unidades de observación censales poseen carácter artificial y arbitrario debido a que su trazado no se basa, en muchas ocasiones, en ningún criterio objetivo sobre la situación de la población en el espacio ya que se establecen por razones logísticas para facilitar la actividad de los agentes censales que recogen los cuestionarios del censo. En otros casos, como en la delimitación de los sectores urbanos se utiliza un criterio definido pero que tampoco asegura que la delimitación obtenida sea acorde a la distribución espacial de los datos censales. Consiguientemente, se puede llegar a decir que no existen, verdaderamente, fronteras naturales en el espacio, y es por ello difícil, en cualquier caso, establecer unidades de observación espaciales totalmente apropiadas para los hechos censales.

Otra particularidad de la información que se localiza especialmente y que rara vez es mencionada a pesar de que afecta a

los indicadores socio- ambientales. En el espacio nada se encuentra distribuido al azar. De hecho todas las cosas están relacionadas, pero las cercanas están más relacionadas que las distantes. Este principio que suena bastante obvio tiene sus bases en una de los fundamentos de la estadística espacial denominado autocorrelación espacial y tiene profundas implicancias en el estudio de los datos que se localizan en el territorio.

La utilidad de la AE está en su capacidad para estudiar la forma en que un fenómeno se irradia a través de las unidades espaciales, y si tal conducta corresponde a algún modelo de difusión conocido o bien registra la segregación espacial de alguna característica. En definitiva, refleja el grado en que objetos o actividades en una unidad geográfica son similares a los objetos o actividades en unidades geográficas próximas (Vilalta y Perdomo, 2005: 326) y es, quizás, después del promedio y la varianza la propiedad más importante de cualquier variable geográfica y, a diferencia de las anteriores, está explícitamente vinculada con patrones espaciales (Goodchild, 2008).

Resumiendo, los datos que se localizan en el espacio (entre ellos los indicadores ambientales) poseen particularidades que deben ser tenidas en cuenta al momento de realizar nuestras investigaciones. El PUEM nos obliga a prestarle atención a los datos que viene en escalas muy chicas (provincia, país), y que están muy agregados (es decir que sufrieron promedios de manera sucesiva). La mejor manera de enfrentarlo es conseguir datos que se encuentren lo más desagregados posibles. La autocorrelación espacial nos resalta la importancia de estudiar la distribución

de los datos para distinguir anomalías o situaciones particulares.



BIBLIOGRAFÍA

- ANSELIN, L.** (2000): *The Alchemy of Statistics, or Creating Data Where No Data Exist. Annals of the Association of American Geographers*. Vol. 90, nº 3, pp. 586-592.
- BOSQUE SENDRA, J.** (1995): "Organización Territorial de la Población de la Comunidad Autónoma de Madrid Informe monográfico sobre el Tomo X de las publicaciones del Censo de Población y Vivienda de 1991", en: *Estudios y Análisis*. Madrid: Consejería de Economía de la Comunidad de Madrid, 116 pp.
- GOODCHILD, M.** (1987): *Spatial Analytical Perspective on Geographical Information Systems. International Journal of Geographical Information Systems*, 1: 327-334.
- GOODCHILD, M.** (2008): "Geographic information science: the grand challenges", in: **WILSON, J. & FOTHERINGHAM, A.** (Editors): *The Handbook of Geographic Information Science*. Malden, MA, Blackwell, 596-608.
- KEMP, K.** (2008): *Encyclopedia of geographic information science*. Sage.
- MORENO JIMÉNEZ, A.** (2003): *Modelado y representación cartográfica de la competencia espacial entre establecimientos minoristas. Boletín de la Asociación de Geógrafos Españoles*, (35): 55-78.
- SECRETARÍA DE AMBIENTE Y DESARROLLO SUSTENTABLE** (2004): *Estrategia Nacional para la Gestión Integral de Residuos Sólidos Urbanos (ENGIRSU)*.
- TOBLER, W.** (1970): *A Computer Movie Simulation Urban Growth in the Detroit Region. Economic Geography*, 46 (2): 234-240.
- VELÁZQUEZ, G.** (2006): *Calidad de vida y escala urbana en la Argentina (2001). Revista Universitaria de Geografía [online]*. 2006, vol.15, n.1 [citado 2016-05-25], pp.37-61. Disponible en: <http://bibliotecadigital.uns.edu.ar/scielo.php?script=sci_arttext&pid=S1852-42652006001100003&lng=es&nrm=i50>.
- VILALTA y PERDOMO, C.** (2005): *Como enseñar autocorrelación espacial. Economía, Sociedad y Territorio*. Vol. V, nº 18, pp. 323-333.