



Pellegrini, José Luis

Instituto de Investigaciones Económicas, Escuela de Economía.

LOCALIZACIÓN DE LA POBREZA URBANA RELATIVA MEDIANTE TÉCNICAS ESPACIALES MULTIVARIADAS ¹

Resumen

Se propone e implementa un procedimiento multidimensional sencillo de análisis exploratorio de datos espaciales para localizar empíricamente zonas extensas relativamente pobres en áreas urbanas. El mismo emplea una versión local multivariada del coeficiente de Geary recientemente aparecida en la literatura, combinada con un análisis LISA univariado, con la finalidad de detectar conglomerados espaciales cuya intersección permita clasificar a ciertas zonas como pobres. Se aplica este procedimiento a datos censales de la ciudad de Rosario, Argentina, lográndose identificar zonas extensas a las que se puede calificar como las más pobres de la ciudad.

Palabras clave: Pobreza; Pobreza de los lugares; Análisis exploratorio de datos espaciales.

Abstract

A simple, multidimensional procedure of Exploratory Spatial Data Analysis is proposed and implemented in order to empirically locate large, relatively poor zones in urban areas. This paper uses a multivariate local version of the Geary coefficient recently published in the literature, combined with a univariate LISA analysis, in order to detect spatial conglomerates whose intersection allows to classify certain zones as poor. This procedure is applied to census data of the city of Rosario, Argentina, being able to identify broad areas which may be described as the poorest in the city.

Keywords. Poverty; Place Poverty; Exploratory Spatial Data Analysis (ESDA).

¹ Este trabajo se elaboró en el marco del Proyecto acreditado y financiado por la Universidad Nacional de Rosario ECO 166, denominado "Desigualdad: una aproximación desde el empleo, el ingreso y la distribución espacial de la actividad económica", dirigido por el autor.



1. Introducción

La noción de pobreza se aplica no solo a personas u hogares, sino también a lugares o zonas geográficas. Mientras las teorías y la investigación sobre la pobreza individual se centran en las características de las personas y los hogares, las que se ocupan de la pobreza de los lugares² (*place poverty*) se interesan no solo por las características de éstos, sino también por la manera en que la pobreza se distribuye de manera desigual a través de espacio (Cotter, 2002). La diferenciación entre ambas aproximaciones puede verse plenamente como una distinción intersecante (*intersecting distinction*, *ibidem*, p. 537) cuando se estudian lugares aislados o sin considerar sus relaciones geográficas con otros sitios, pero cuando se asigna importancia a la distribución geográfica o localización de la pobreza, se vuelve clave una dimensión que puede estar solo implícita en la intersección entre los dos enfoques: la dimensión espacial. La consideración explícita de la misma tiene implicancias importantes para el análisis, porque conlleva requerimientos metodológicos específicos.

Entendiendo pobreza como privación duradera en lo que respecta a necesidades cuya satisfacción requiere recursos económicos (*vide* Boltvinik, 1998; 1999), un lugar o zona pobre puede entenderse como aquél donde la insuficiencia de recursos productivos se traduce en privaciones de la población. Pero en un contexto urbano puede pensarse mejor como una zona en la que prevalece (y tal vez se concentra) una subpoblación pobre, lo cual puede inferirse de las elevadas proporciones en que se presentan, en la población total, características propias de la condición de pobreza³.

El estudio de la localización de la pobreza conlleva el problema de la identificación y delimitación empírica de las zonas pobres, el cual que se complica cuando se adoptan enfoques multivariados. Se trata de una dificultad común a muchas disciplinas científicas que deben tratar con numerosas variables distribuidas espacialmente. Las dificultades comunes se derivan de la necesidad de considerar varios atributos de los objetos o eventos estudiados y a la vez, la localización de las unidades de análisis en el espacio. Cuando se las estudia de manera exploratoria tratando de identificar, por ejemplo, conglomerados espaciales (*spatial clusters*), los métodos que se aplican son a veces de una complejidad considerable en cuanto a su implementación y pueden presentar dificultades de interpretación, así como limitaciones en cuanto a su precisión y sensibilidad (Demšar, Harris, Brunson, Fotheringham y McLoone, 2013; Grubestic, Wei y Murray, 2014).

Tal vez como consecuencia de ello, en el estudio multivariado de la distribución geográfica de la pobreza y de otros fenómenos socioeconómicos espacialmente localizados se recurre con alguna frecuencia a técnicas no espaciales, tales como la construcción de índices *ad hoc* o el análisis de conglomerados (*cluster analysis*) estándar, el cual que no toma en cuenta las relaciones espaciales y elude sus complicaciones (Demšar *et al.*, 2013).

² Esta traducción, que utiliza el plural, está tomada del Resumen en español de Powell, Boyne y Ashworth (2001).

³ Lo cual, obviamente, no implica dejar de lado la noción de privación de recursos geográficamente situados, privación que puede darse en varios planos: de los recursos materiales (tales como infraestructura urbana); del capital social (en el sentido de vínculos sociales económicamente valiosos); de la proximidad con sitios económicamente atractivos (como aquellos donde se encuentran puestos de trabajo o servicios comunitarios); etc.



El análisis no espacial se distingue por considerar los valores que toman las variables en cada localización sin tomar en cuenta sus relaciones espaciales, es decir, sin relación con los valores que las mismas toman en localizaciones vecinas. El análisis espacial, por el contrario, se caracteriza por centrarse en la consideración de estas relaciones.

Como ya se señaló, el Análisis Exploratorio de Datos Espaciales (AEDA, *ESDA* por sus siglas en inglés) suele ser arduo en el caso multivariado, pero las posibilidades de aplicar técnicas relativamente sencillas han aumentado gracias a la reciente extensión al caso multivariado del estadístico local *c* de Geary (Anselin, 2017), el cual puede utilizarse con fines exploratorios.

Esas posibilidades se abren, al menos, para un caso relativamente sencillo como es el que se aborda en este trabajo, donde se busca identificar, con una perspectiva multidimensional, emplazamientos de la pobreza en el interior de una ciudad. Se procura averiguar si existen uno o más conglomerados espaciales relativamente homogéneos en cuanto a varias características de pobreza, consideradas simultáneamente.

Se propone e implementa un procedimiento exploratorio espacial sencillo para localizar empíricamente, sobre la base de atributos de personas, hogares y viviendas, zonas pobres relativamente extensas en la ciudad de Rosario, siguiendo la sugerencia de Anselin (2017, p. 5) de combinar el empleo del nuevo estadístico local multivariado de Geary con el estadístico de Moran local univariado. Los datos utilizados fueron tomados del Censo Nacional de Población, Hogares y Viviendas 2010 (Censo 2010, en adelante). La condición de que las zonas sean relativamente extensas tiene por finalidad dejar de lado, en principio, a sitios pobres relativamente pequeños y aislados, como suelen ser los asentamientos irregulares situados en los intersticios de zonas no pobres, cuyo estudio detallado requeriría un abordaje diferente.

El lector podrá colegir fácilmente que la metodología propuesta está inspirada en la ilustración empírica de la aplicación del nuevo estadístico presentada por Anselin (2017), pero aquí se va algo más allá.

2. Importancia práctica de la distribución espacial de la pobreza urbana

La desigual distribución en el interior de las ciudades de la población pobre y su agrupamiento en ciertos sitios formando conglomerados espaciales es un fenómeno que desde un principio atrajo la atención de los investigadores. Domínguez Domínguez y Martín Caraballo (2006) recuerdan, citando a Atkinson, que Booth, quien habría sido el primero en intentar medir la magnitud de la pobreza entre 1892 y 1897, elaboró mapas de la misma en Londres.

La conformación de barrios (*neighborhoods*) de diferentes características socioeconómicas parece ser resultado de procesos endógenos propios de la conformación y desarrollo de las aglomeraciones urbanas, con secuelas que pueden afectar positiva o negativamente a las personas que residen en ellos. Existe una abundante literatura al respecto, que no es de interés para este trabajo.

Sean cuales fueran los orígenes y consecuencias de la conglomeración de población pobre en ciertas localizaciones, el conocimiento y delimitación de tales lugares tiene importancia práctica para las políticas públicas, especialmente aquellas de los gobiernos locales empeñados en la lucha contra la pobreza. Éstos normalmente no cuentan con medios para afectar significativamente las variables fundamentales que la determinan, pero tienen un alto grado de responsabilidad en la provisión de infraestructura urbana y pueden influir, cuando no los tienen



a su cargo directamente o concesionan, en la oferta de servicios públicos relacionados con carencias típicas de la población pobre, como abastecimiento de agua potable, saneamiento, atención primaria de la salud, etc. Con frecuencia están en condiciones de poner en práctica políticas de orientación selectiva de los mismos dentro de su jurisdicción territorial y por este medio, pueden contribuir a mitigar privaciones mejorando la posibilidad de satisfacer varias necesidades sin que se requiera el aumento previo del ingreso nominal individual o el subsidio monetario

3. Algunos antecedentes conceptuales y metodológicos

3.1. Pobreza: una condición compleja y controvertida. La pobreza es un fenómeno complejo que atraviesa todos o casi todos los aspectos de la vida. Utilizando una noción que Williamson (1991) aplicó a un asunto por completo diferente, puede decirse que la pobreza se presenta como un síndrome de características, cada una de las cuales puede tener diferente intensidad e incluso, no estar presente en cada caso particular. La apreciación del grado en que las características de pobreza se presentan individual o conjuntamente y alcanzan niveles críticos, es lo que lleva a calificar a una persona, hogar o lugar como pobre.

Indudablemente, se trata de un asunto en cuyo abordaje es prácticamente imposible desprenderse de juicios de valor. Éstos no necesariamente responden a consideraciones meramente subjetivas o son ideas preconcebidas del tipo de las que integran la visión en el sentido de Schumpeter (1949), sino que están implícitos en criterios que pretenden ser objetivos en tanto se esfuerzan por reflejar mínimos de satisfacción o estándares de vida socialmente aceptados. Pero parece difícil que tales criterios no resulten controversiales desde el punto de vista normativo. Incluso aquellos métodos que se definen como no normativos (Boltvinik, 1998; 1999) pueden llevar implícitos criterios que en realidad son normativos, ya que en ellos la identificación empírica de la pobreza supone la determinación apriorística de un nivel absoluto o de una posición relativa que se compara con las condiciones observadas.

La complejidad del problema y las diferentes estrategias de abordaje que son posibles, dan lugar a varias distinciones que no son intrínsecamente excluyentes ni estrictamente complementarias y que, al igual que aquella que hace Cotter (2002) entre pobreza de las personas y pobreza de los lugares, merecen ser descritas como intersecantes. A continuación, se pasa rápida revista a algunas de ellas, con el propósito de contextualizar criterios conceptuales y metodológicos adoptados más adelante en este trabajo.

3.2. Pobreza absoluta y relativa; unidimensional y multidimensional. En sentido amplio (Sen, 1992), la pobreza puede verse como un problema de subsistencia o de desigualdad y concomitantemente, aunque no exista una exacta correspondencia biunívoca, como una cuestión de privación absoluta o de privación relativa.

La noción de pobreza absoluta implica aceptar la existencia de un nivel por debajo del cual se experimenta una situación de privación, que puede determinarse por contraste con un estándar mínimo y sin necesidad de comparación interindividual alguna. La pobreza relativa entraña, en cambio, la imposibilidad de disfrutar de cierto nivel de vida usual en una sociedad dada, que puede cambiar entre un momento y otro (Sen, 1992; Boltvinik, 1998).

Empíricamente, la pobreza absoluta se puede identificar comparando una situación observada con cierto nivel de ingreso, consumo o gasto; por la insatisfacción de una o más necesidades básicas; o por deficiencias en el acceso a servicios básicos y en cuanto a aspectos claves de



los funcionamientos en educación, salud y niveles de vida. La pobreza relativa usualmente se identifica por una posición situada por debajo de cierto parámetro de una distribución, como media, mediana o algún percentil (ECLAC, 2009; 2013).

La pobreza puede abordarse de manera unidimensional o multidimensional. La primera manera recurre a medidas monetarias de variables como las ya nombradas: ingreso, consumo o gasto, las que permiten sintetizar diferentes situaciones de privación en una única dimensión. Sin embargo, la pobreza puede entrañar simultáneamente varias privaciones como desnutrición, falta de servicios básicos, bajo nivel educativo, etc. que tal vez no pueden ser bien descriptos enfocándose solamente en una variable (MPPN, s/f). El análisis multidimensional, a diferencia del unidimensional, supone que el bienestar tiene varios aspectos relevantes que no son captados por las variables monetarias.

3.3. Conceptos y metodologías usuales en Latinoamérica. En la práctica, en América Latina se utilizan tres conceptos de pobreza con sus correspondientes metodologías de medición: insuficiencia de ingresos, necesidades básicas insatisfechas y pobreza multidimensional. Los dos primeros pueden eventualmente combinarse.

La pobreza por insuficiencia de ingresos se identifica comparando el ingreso per cápita del hogar con los recursos requeridos para satisfacer las necesidades básicas de sus integrantes. Se trata de un método indirecto (esto es, de uno que procura evaluar la satisfacción solo potencial de las necesidades) que requiere para su implementación la determinación de niveles o líneas debajo de las cuales el hogar y sus integrantes se encuentran en la pobreza o en la indigencia o pobreza extrema (ECLAC, 2009).

Para Boltvinik (1998), la línea de pobreza es la quintaescencia del método unidimensional. En contraste, los indicadores no monetarios son por su misma naturaleza multidimensionales y generalmente son utilizados cuando se emplean métodos directos para la identificación de la pobreza. Estos procuran evaluar la satisfacción efectiva de necesidades específicas comparando la situación observada con un umbral normativo. Esta es la base del método las necesidades básicas insatisfechas (NBI). Existen dos variantes del mismo: la construcción de un índice compuesto mediante ponderadores que permiten combinar varios indicadores y la utilización de estos indicadores de manera dicotómica. En el primer caso, se considera pobres a los hogares que no superan cierto valor del índice y en el segundo, a los que no alcanzan el nivel de satisfacción en uno o más indicadores. Ésta última alternativa es la adoptada por el INDEC para la Argentina.

En los últimos tiempos se está generalizando el enfoque a la pobreza multidimensional, mediante la aplicación en un creciente número de países latinoamericanos del Índice de Pobreza Multidimensional (IPM), elaborado de acuerdo con la metodología Alkire Foster (MPPN, s/f). Se trata de un índice ponderado que combina indicadores correspondientes a tres dimensiones: educación, salud y condiciones de vida (Alkire, Conconi y Seth, 2014).

Varios indicadores componentes del IPM, correspondientes a las dimensiones educación y condiciones de vida, son del mismo tipo que los empleados para la determinación de NBI. De hecho, los índices de NBI se consideran precursores del IPM (ECLAC, 2013).



4. Metodologías de medición de la pobreza y análisis espacial

Los tres métodos para la identificación de la pobreza brindan diferentes posibilidades de llevar adelante un análisis espacial en una ciudad.

Los datos para evaluar la incidencia de la pobreza por insuficiencia de ingresos se toman normalmente de encuestas por muestreo, que en razón del tamaño de muestra no son representativas de áreas pequeñas. Las NBI, en cambio, han sido pensados para ser evaluadas sobre la base de datos censales que son específicos de áreas pequeñas (Hicks, 2000), tales como radios censales. El MPI, por su parte, se elabora con datos de encuestas⁴ (UNDP, s/f), aunque muchos de los datos necesarios podrían obtenerse de los censos de población.

Un análisis espacial que utilice los datos estadísticos disponibles debe recurrir a indicadores de pobreza que surjan de datos censales. De hecho, los datos de NBI de base censal han sido ampliamente utilizados en América Latina para construir mapas de pobreza (Hicks, 2000). Aquí se aprovecha el mismo tipo de fuente para el análisis exploratorio de datos espaciales orientado a la identificación de las áreas pobres.

5. Estadísticos para el análisis de datos espaciales⁵

5.1. Dos familias de índices. En AEDA se utilizan corrientemente dos familias de índices, las de Geary y de Moran, que como estadísticos conducen a resultados análogos, aunque suficientemente diferentes como para iluminar los mismos fenómenos con diferentes luces. A juicio del presente autor, puede decirse, también de ellos, que en cierto sentido son intersecantes.

Las principales similitudes y diferencias pueden apreciarse con facilidad en el caso global univariado, que fue aquél para el cual se definieron originalmente los estadísticos c de Geary e I de Moran. Ambos son casos de una clase muy general de índices de asociación basados en la suma de los productos cruzados de elementos apareados tomados de dos matrices de similitud. Las medidas de similitud espacial se obtienen cuando una de las dos matrices representa vecindad. Así, la matriz de productos cruzados es:

$$\Gamma = \sum_i \sum_j w_{ij} v_{ij} \quad (1)$$

donde w_{ij} es un elemento de W , la matriz cuadrada de orden n de los ponderadores espaciales, que captura las relaciones de vecindad entre los sitios i y j . Dichas relaciones

⁴ La fuente de datos para la estimación del IPM de la Argentina registrada por el PNUD (UNDP, s/f) es la Encuesta Nacional de Nutrición y Salud 2005, que a juicio del presente autor no puede considerarse estadísticamente representativa de toda la población, ya que la muestra fue deliberadamente sesgada hacia subgrupos de interés por cuestiones sanitarias (*vide* Dirección Nacional de Maternidad, Infancia y Adolescencia, s/f). Actualmente, el Ministerio de Energía y Minería de la Nación publica en su página Web (Ministerio de Energía y Minería, s/f) un mapa del IPM sin mencionar la fuente de los datos.

⁵ Esta sección está basada en Anselin (2017).



pueden entenderse como como contigüidad, distancia, conectividad, etc. Los v_{ij} son medidas de similitud o proximidad entre i y j en alguna otra dimensión (Sawada, s/f).

Sin embargo, c de Geary opera con comparaciones entre pares de localizaciones, mientras que I de Moran lo hace con covariaciones. Entonces, a partir de (1) se llega a los conceptos de Geary y de Moran haciendo, respectivamente $v_{ij} = (x_i - x_j)^2$ y $v_{ij} = x_i x_j$ (Anselin, 1995).

Las expresiones analíticas de ambos índices son:

$$c = \frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2 / 2 \sum_i \sum_j w_{ij}}{\sum_i (x_i - \bar{x})^2 / (n - 1)} \quad (2)$$

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (3)$$

donde w_{ij} son, como se dijo, los ponderadores espaciales, n el orden de la matriz cuadrada W ; x es la variable de interés medida en las localizaciones i y j ; y \bar{x} es la media de x . Como es usual, se supone que la variable x ha sido estandarizada con media 0 y variancia 1. Cuando las relaciones de vecindad se entienden como contigüidad, tal cual se hará más adelante en este trabajo, los ponderadores espaciales toman valores 1 si los sitios son contiguos y 0 si no lo son. Por convención, $w_{ij} = 0 \forall i = j$. En la práctica, esta matriz binaria suele presentarse estandarizada por filas.

El valor esperado de c bajo aleatorización, esto es, en ausencia de autocorrelación espacial, es 1. Valores entre 0 y 1 indican autocorrelación espacial positiva, mientras que valores superiores a 1 (normalmente menores que 2) indican autocorrelación espacial negativa. Los valores que puede tomar I varían entre -1 y 1, como lo hace el coeficiente de correlación convencional, no espacial (la expresión analítica de I es análoga a la de éste). Un valor 0 indica ausencia de autocorrelación espacial, mientras que valores positivos o negativos revelan autocorrelación espacial positiva o negativa. Empíricamente, ambos índices dan resultados similares, pero no idénticos y tienen entre sí una relación lineal inversa, aunque no perfecta. Los desvíos son atribuibles a que los índices, como ya se dijo, no miden lo mismo: diferencias entre pares, uno, y covariaciones, el otro (Sawada, s/f).

A partir de ambos índices globales univariados se desarrollaron estadísticos locales que los descomponen en la contribución de cada observación, así como bi o multivariados. Los que aquí interesan son los índices de Geary local multivariado y de Moran local univariado.

5.2. Estadístico de Geary local multivariado

El estadístico c_i de Geary (local univariado) se define como:



$$c_i = \sum_j w_{ij} (x_i - x_j)^2 \quad (4)$$

Como en la práctica la matriz W de ponderadores espaciales se estandariza por filas, el valor de la variable en cada localización se compara con el promedio de la misma en los sitios vecinos.

Bajo aleatorización, la esperanza de c_i , $E[c_i] = 2$. Un valor c_i estadísticamente significativo sugiere la conglomeración de valores similares cuando $c_i < 2$ (autocorrelación local positiva) y de valores diferentes (altos rodeados de bajos y viceversa) cuando $c_i > 2$ (autocorrelación local negativa). El significado de la autocorrelación local positiva es un tanto ambiguo, ya que los valores similares pueden ser ambos altos o bajos, pero también intermedios. Incluso, dos valores similares pueden estar uno por debajo y otro por encima de la media. La autocorrelación local negativa, por su parte, solo indica una gran diferencia entre valores vecinos: el estadístico no es sensible a patrones específicos, como valores altos (bajos) rodeados de bajos (altos).

Las propiedades de c_i mencionadas son heredadas por la versión multivariada, con una modificación debida a que cada sitio y sus vecinos deben computarse tantas veces como variables sean tomadas en consideración.

La diferencia al cuadrado entre los valores de las variables en diferentes localizaciones del espacio geográfico, que aparece en la expresión analítica de c_i (ecuación 4), puede interpretarse como una distancia al cuadrado en el espacio de los atributos, tal como se hace en el análisis factorial. En el caso univariado se trata de una distancia sobre la recta, pero puede generalizarse a un espacio de n -dimensional. Se demuestra con facilidad que el estadístico de Geary local es aditivo en los atributos.

Entonces, en el caso local multivariado el estadístico $c_{k,i}$ de Geary, para k atributos, es:

$$c_{k,i} = \sum_{v=1}^k c_{v,i} \quad (5)$$

donde $c_{v,i}$ es estadístico local de Geary para la variable v .

El estadístico $c_{k,i}$ se interpreta como el promedio ponderado de las distancias al cuadrado en el espacio multidimensional de los atributos entre los valores de cada variable en un sitio respecto de los valores que toma en sitios vecinos.

Para que los valores de $c_{k,i}$ se encuentren en el mismo rango que los valores de c_i , en lugar de la suma de los estadísticos locales puede tomarse el promedio simple de los mismos, dividiendo por un factor de escala k .



$$c_{k,i} = \sum_{v=1}^k c_{v,i}/k \quad (6)$$

Bajo aleatorización, $E[c_i] = 2k$ para la versión no escalada y $E[c_i] = 2$ para la escalada.

Para los contrastes de significación, en cada caso se genera por permutaciones condicionales⁶ una distribución empírica de referencia. El valor pseudo-p es la fracción del estadístico de referencia que es igual o más extremo que el valor observado. Tomando en cuenta las salvedades que hace Anslin (2017), estos contrastes deben tomarse con mucha precaución debido a los problemas que entrañan las comparaciones múltiples y los contrastes correlacionados, ya que el estadístico multivariado es una suma de estadísticos univariados. Según dicho autor, existiría cierta evidencia de que la cantidad de casos extremos crecería con el número de variables comprendidas, por lo que sería razonable dividir al valor pseudo-p tomado como criterio de significación por el número de variables, para evitar un número excesivo de sitios cuyos valores resulten significativos. Así, los contrastes de significación no deberían tomarse en sentido estricto, sino más bien como indicadores de localizaciones de interés en estudios exploratorios, a los que se debería someter a análisis de sensibilidad.

En síntesis, una correlación local positiva y estadísticamente significativa según el estadístico $c_{k,i}$ debe interpretarse como una gran posibilidad de que ciertos sitios que están próximos en el espacio geográfico se correspondan con valores de las variables que estén próximos en el espacio k -dimensional de los atributos. En otras palabras, como una gran posibilidad que algunos sitios que son vecinos también se parezcan en términos de las variables tomadas en consideración.

5.3. Estadístico de Moran local univariado

El estadístico I_i de Moran local se define como:

$$I_i = (x_i - \bar{x}) \sum_j w_{ij} (x_j - \bar{x}) \quad (7)$$

Al igual que el índice global puede variar entre -1 y 1, con la misma interpretación en cuanto a su significado en lo que refiere a autocorrelación. La interpretación del índice local se ve facilitada por el diagrama de Moran, que representa en la absisa los valores de la variable en cada localización y en la ordenada, el valor de la misma espacialmente desfasada, lo que significa que cuando W está estandarizada por filas, este valor es el promedio de los valores de la variable en localizaciones vecinas. Como la variable de interés está además estandarizada (con media 0), cada cuadrante representa combinaciones específicas de valores altos y bajos, lo cual conduce a la tipificación de conglomerados espaciales en aquellos que agrupan combinaciones de valores alto-alto y bajo-bajo de la variable, así como *outliers*

⁶ Este procedimiento consiste en generar una distribución para cada localización individual permutando aleatoriamente los valores observados en otras localizaciones, recomputando el estadístico en cada oportunidad.



espaciales: combinaciones alto-bajo y bajo-alto.

Anselin (1995) demostró que este estadístico, al igual que c_i , cumple las condiciones de un LISA (*Local Indicator of Spatial Association*), una de las cuales es que, para cada observación, dé una indicación de la extensión de un conglomerado espacial significativo de valores similares en torno de la misma. De hecho, I_i cumpliría una condición más restrictiva, ya que también es capaz de hacer lo mismo con valores disímiles al de la observación, aunque este aspecto no es de especial interés para el presente trabajo. La representación cartográfica da lugar a los mapas LISA.

También para I_i los contrastes estadísticos de significación (o pseudo-significación) se realizan frente a una distribución empírica de referencia generada mediante permutaciones condicionales.

6. Intersección de conglomerados

La información con que normalmente se cuenta para el estudio empírico la pobreza de los lugares proviene de datos de área, las cuales son unidades de agregación espacial definidas arbitrariamente. Se supone que el territorio estudiado está dividido en áreas continuas a las cuales se georreferencian las variables de interés.

Con las limitaciones discutidas más arriba, el estadístico $c_{k,i}$ de Geary puede emplearse para detectar correlaciones espaciales significativas entre varias variables. Permite identificar conglomerados espaciales (en adelante, conglomerados multivariados de Geary) donde es muy posible que dichas variables tomen valores semejantes. Sin embargo, no se puede saber si tales valores son altos, bajos o intermedios.

Para ello puede emplearse el estadístico I_i de Moran para cada variable individualmente considerada, el cual, en tanto que LISA, permite detectar los cuatro tipos de conglomerados espaciales significativos posibles que puede formar una variable, según las permutaciones entre valores altos y bajos de la misma (en adelante, conglomerados LISA).

Con éstos es posible identificar la parte del conglomerado multivariado de Geary que se corresponde con valores altos o bajos de las variables. Nótese que el problema que se procura resolver no es aquél de superposición de conglomerados (*clusters overlap*, Jaquez, 2008), donde se intenta averiguar si la asociación espacial entre diferentes conglomerados es estadísticamente significativa, sino solo a qué niveles de valores de las variables corresponden las localizaciones comprendidas por el conglomerado multivariado de Geary.

Para facilitar el análisis, conviene definir las variables utilizadas de manera que los valores altos (bajos) correspondan a mayor (menor) intensidad del fenómeno estudiado. En este caso, mayor (menor) pobreza. Eligiendo mayor ("cuanto mayor, peor"), para la identificación de las zonas pobres solo interesan los conglomerados LISA alto-alto, es decir, aquellos en que los valores altos de las variables en cada localización están asociados con valores también altos en sitios vecinos.

Las zonas pobres, o más precisamente, las zonas más pobres del lugar estudiado, pueden interpretarse como aquellas en las que un conglomerado multivariado de Geary se interseca



con los conglomerados LISA alto-alto de las variables individuales. Suponiendo, por cierto, que estos conglomerados existan. No debe olvidarse que se trata de un análisis exploratorio.

Si la representación espacial del conjunto intersección diera lugar a uno o más conglomerados, a los efectos prácticos quedarían aventadas las dudas que según se señala más arriba, suscitan los contrastes de significación de $c_{k,i}$. Al menos para las áreas que integran estos conglomerados

No obstante, la condición de que el conglomerado multivariado de Geary se superponga con conglomerados alto-alto de todas las variables puede parecer demasiado restrictiva. A fuer de analogía, cuando se define a la pobreza por NBI con un criterio dicotómico, para identificarla no se exige que todas las NBI consideradas estén presentes en cada caso particular y hasta puede bastar con solo una.

El inconveniente de adoptar aquí un enfoque más laxo, admitiendo como zonas pobres a partes del conglomerado multivariado de Geary que se corresponden solo con algunos de los conglomerados LISA, es que esto reavivaría las mismas dudas que se acaban de despejar. En efecto, las áreas que superaron el contraste de pseudo-significación multivariada habrían sido demasiadas. Esta objeción pierde importancia si se admite, siguiendo a Anselin (2017), que los contrastes de $c_{k,i}$ son útiles para dar indicación de que ciertos sitios son interesantes en un estudio exploratorio. Estos sitios pueden seguir siendo interesantes aún si la correspondencia con los conglomerados LISA no es total. En cualquier caso, existe un piso mínimo, tal que el análisis no deje de ser multivariado.

7. El estudio empírico.

Los datos analizados, correspondientes a la ciudad de Rosario, provienen como ya se dijo, del Censo 2010. La unidad de análisis espacial es el radio censal. En ese Censo se determinaron para Rosario 1.071 radios, de los cuales cuatro, situados sobre la costa del río Paraná, fueron excluidos del estudio por presentar muy pocos registros censales, seguramente a causa de que las áreas que abarcan están dedicadas exclusivamente a usos no residenciales.

Para efectuar el análisis se tomaron en cuenta tres dimensiones: educación, confort y hábitat, referidas, cada una, a uno de tres tipos de unidades de empadronamiento del Censo: personas, hogares y viviendas particulares. Para cada dimensión se seleccionan uno o dos indicadores de privación, del tipo de los utilizadas en los recientes estudios de pobreza multidimensional y en los más tradicionales de pobreza por NBI. Las variables se definen de modo que capturen el grado de privación de cada área, lo que se hace mediante la proporción de unidades de empadronamiento relevantes que están por debajo o por encima de cierto umbral del indicador.

Los elementos mencionados en el párrafo anterior se detallan en el cuadro 1. En el Apéndice, tabla A, aparecen las estadísticas sumarias de las variables.

Todas las variables se relacionan con aspectos críticos del bienestar de carácter estructural, por lo que la asociación espacial entre valores significativamente altos de todas o de la mayoría de ellas señalaría zonas extremadamente pobres, en el sentido de pobreza relativa.

La utilización de una cantidad relativamente pequeña de variables tiene el propósito de evitar las dificultades que plantea el estadístico local multivariado de Geary cuando aquellas son numerosas (*vide supra*).



Cuadro 1. Dimensiones de la pobreza, unidades de empadronamiento censal, indicadores, grados de privación y variables consideradas en el análisis.

Dimensiones de la pobreza y unidades de empadronamiento		Indicador de privación	Definiciones y nombres de las variables	
Dimensión	Unidad		Definición	Nombre
Educación	Persona	Bajo nivel educativo de la población en edad activa	Proporción de personas de 15 a 59 años que no completaron el nivel primario.	EDUCA
Confort	Hogar	Hacinamiento crítico	Proporción de hogares con más de tres personas por cuarto.	HACIN
		Equipamiento deficiente	Proporción de hogares sin heladera.	HELAD
Hábitat	Vivienda particular	Materiales poco resistentes y sólidos o de baja calidad	Proporción de viviendas de materiales de tipos III y IV en techo y pisos.	MATER
		Conexiones deficientes a servicios básicos	Proporción de viviendas con conexiones a servicios básicos de calidad insuficiente.	SERVI

Fuente: Elaborado sobre la base de INDEC e IPEC Santa Fe, Censo 2010.

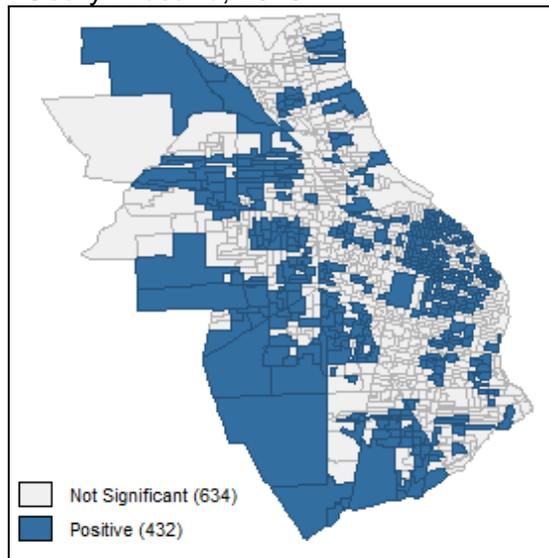
De acuerdo al procedimiento propuesto previamente, se llevó a cabo el análisis de la correlación espacial multivariada. Siguiendo la recomendación de Anselin (2017) ya referida, se adoptó para los contrastes el nivel crítico de pseudo-significación que resulta de dividir por el número de variables el nivel previamente adoptado de 0,05⁷, lo cual llevó dicho nivel al 0,01. El mapa de conglomerados (*cluster map*) resultante, luego de 999 permutaciones, se presenta en la ilustración 1.

El mapa muestra una cantidad muy considerable de radios censales (de hecho, más del 40%) con valores significativos según el estadístico local multivariado de Geary para las cinco variables utilizadas. Como se esperaba dada las variables elegidas, la correlación es positiva en todos los casos y las localizaciones significativas se agrupan en diferentes partes de la ciudad, principalmente en el centro y en la periferia.

⁷ Este es el empleado por Anselin (2017) y también, el utilizado corrientemente por el autor del presente trabajo cuando procura identificar conglomerados espaciales.



Ilustración 1. Mapas de conglomerados multivariados de Geary. Rosario, 2010.



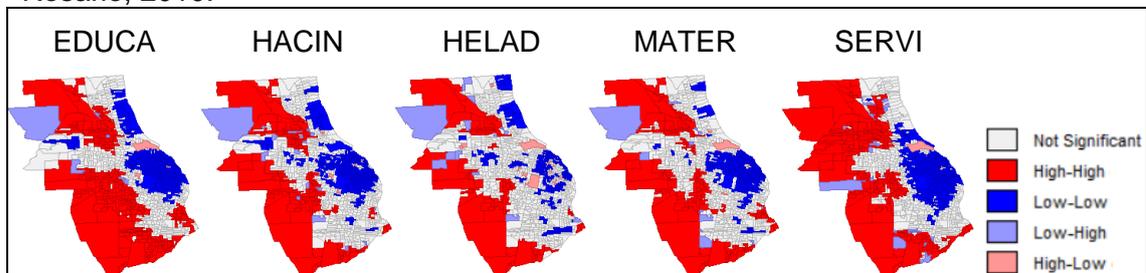
Seudo-p: 0,01

Permutaciones: 999

Fuente: Elaboración propia con datos de INDEC - IPEC Santa Fe, Censo 2010.

El siguiente paso es el análisis de la autocorrelación espacial y la elaboración de mapas LISA para cada una de las cinco variables consideradas individualmente. En este caso se adoptó el nivel de seudo-significación de 0,05. Dichos mapas se presentan en la ilustración 2, también luego de 999 permutaciones.

Ilustración 2. Mapas LISA de las variables individuales utilizadas en el análisis. Rosario, 2010.



Seudo-p: 0,05

Permutaciones: 999

Fuente: Elaboración propia con datos de INDEC - IPEC Santa Fe, Censo 2010.

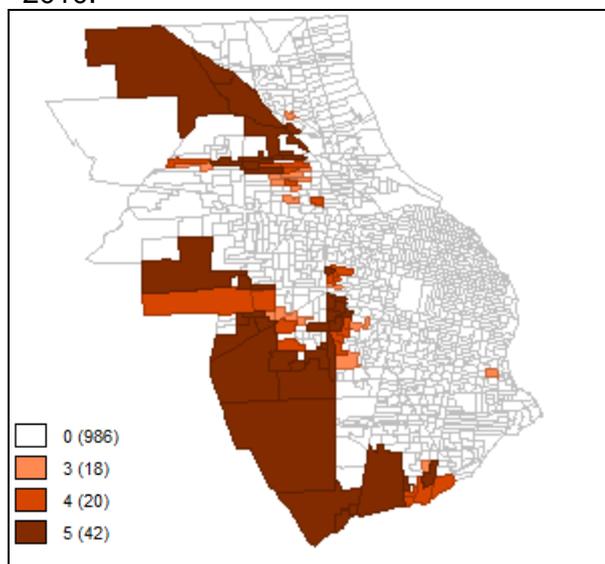


En los mapas LISA los valores altos y bajos de las variables forman grupos casi siempre compactos y en todos los casos, claramente contrapuestos en lo que se refiere a sus valores entre el centro de la ciudad y dos zonas de la periferia. Los grandes conglomerados alto-alto, que son los que aquí más interesan, se encuentran en la periferia y en gran parte, en los mismos lugares para todas las variables.

Finalmente, se lleva a cabo la intersección de los conglomerados multivariados de Geary con los conglomerados LISA alto-alto, para todas las variables individuales conjuntamente, así como para cuatro y para tres. Los resultados cartográficos de esta operación se presentan en el mapa de la ilustración 3, que muestra lo que puede interpretarse como las zonas más pobres de Rosario.

En el mismo puede verse que los conjuntos intersección forman, en el espacio, dos grandes zonas continuas situados en posición periférica, en el sudoeste y noroeste de la ciudad. También existe un pequeño conglomerado y unas pocas áreas aisladas, cercanos a las grandes zonas continuas. Solo se observa un área aislada y distante de ellas, situada en el sudeste de la ciudad. Cabe la salvedad de que en la conformación de ésta solo intervienen tres variables individuales.

Ilustración 3. Mapas de conglomerados resultante de intersecciones entre el conglomerado multivariado de Geary y conglomerados LISA alto-alto. Rosario, 2010.



Fuente: Elaboración propia con datos de INDEC - IPEC Santa Fe, Censo 2010.

Por otra parte, debe señalarse que la gran extensión geográfica de dichas zonas, especialmente aquellas que comprenden a los radios censales de mayor tamaño, puede resultar engañosa en lo que respecta a su importancia. Ello es así porque la densidad demográfica de los radios censales tiende a disminuir con su tamaño, debido a que la cantidad



de viviendas que comprende cada uno es aproximadamente la misma, excepto en los bordes de las localidades, donde esa cantidad puede ser menor⁸.

La cantidad de unidades de empadronamiento censal localizadas en las zonas pobres, de acuerdo a las tres definiciones utilizadas, se muestran en el cuadro 2. Naturalmente, esas zonas no comprenden a toda la población pobre de la ciudad. Ni toda la población de las zonas pobres es necesariamente pobre.

Cuadro 2. Proporciones de unidades de empadronamiento censal situadas en las zonas pobres, según la cantidad de variables individuales cuyos conglomerados LISA Alto-Alto resultaron intersecantes con el conglomerado multivariado de Geary, en %. Rosario, 2010.

Unidades	Cantidad de variables individuales que se intersecan con el conglomerado multivariado de Geary		
	Cinco	Cuatro	Tres
Personas	6,9	10,0	12,3
Hogares	5,5	8,0	10,0
Viviendas	4,6	6,6	8,4

Fuente: Elaboración propia con datos de INDEC - IPEC Santa Fe, Censo 2010.

Consideraciones finales

Los conglomerados representados en la Ilustración 3 fueron interpretados como las zonas más pobres de la ciudad de Rosario. Sin embargo, es posible que la cantidad de áreas que las componen haya resultado subestimada y dichas zonas sean realmente más extensas.

En efecto, existen 22 radios censales que pertenecen simultáneamente a los cinco conglomerados LISA alto-alto pero no al conglomerado multivariado de Geary, lo cual es llamativo. Esto puede ser consecuencia de haber aplicado un criterio muy restrictivo al adoptar un nivel crítico de pseudo-significación demasiado exigente en el contraste de $c_{k,i}$, o un criterio demasiado liberal en los contrastes de I_i . O ambas cosas a la vez. Estos son aspectos que deberán ser analizados en el futuro.

Indudablemente, los instrumentos de análisis espacial utilizados tienen sus limitaciones, pero también son muy sencillos en su implementación y en la interpretación de sus resultados.

Referencias bibliográficas

Alkire, S., Conconi, A. y Seth, S. (2014). *Multidimensional Poverty Index 2014: Brief Methodological Note and Results*. The Oxford Poverty and Human Development Initiative (OPHI), Oxford Department of International Development, University of Oxford.

⁸ El tamaño de los radios censales se determina según la cantidad de viviendas, en promedio 300, aunque en los bordes de las localidades puede bajar a 200 (INDEC, s/f).



- Anselin, L. (1995). Local Indicators of Spatial Association - LISA. *Geographical Analysis*, 27, (2), 93-115.
- Anselin, L. (2017). A Local Indicator of Multivariate Spatial Association: Extending Geary's c. Center for Spatial Data Science, University of Chicago. Disponible en https://s3.amazonaws.com/geoda/docs/LA_multivariateGeary1.pdf
- Boltvinik, J. (1998). *Poverty measurement methods: An overview*. SEPED Series on Poverty Reduction, UNDP, New York. Disponible en http://www.undp.org/poverty/publications/pov_red/Poverty_Measurement_Methods.pdf (January 25, 2005).
- Boltvinik, J. (1999). Métodos de medición de la pobreza. Conceptos y tipología. *Revista Socialis*, 1, 35-97.
- Cotter, D. A. (2002). Poor people in poor places: Local opportunity structures and household poverty. *Rural Sociology*, 67(4), 534-555.
- Demšar, U., Harris, P., Brunson, Ch., Fotheringham, A. S. y McLoone, S. (2013). Principal component analysis on spatial data: an overview. *Annals of the Association of American Geographers*, 103(1), 106-128.
- Dirección Nacional de Maternidad, Infancia y Adolescencia (s/f). *Encuesta Nacional de Nutrición y Salud*, Ministerio de Salud, Presidencia de la Nación, Argentina. <http://www.msal.gob.ar/dinami/index.php/publicaciones-para-equipos-de-salud/encuesta-nacional-nutricion-salud>
- Domínguez Domínguez, J. y Martín Caraballo, A. M. (2006). Medición de la pobreza: una revisión de los principales indicadores. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 2, 27-66.
- ECLAC - Economic Commission for Latin America and the Caribbean, Subregional Headquarters in Mexico (2009). *Income Poverty and Unsatisfied Basic Needs*. Disponible en http://repositorio.cepal.org/bitstream/handle/11362/25900/1/LCmexL949_en.pdf.
- ECLAC - Economic Commission for Latin America and the Caribbean (2013, April). *The Multidimensional Measurement of Poverty*. Twelfth meeting of the Executive Committee of the Statistical Conference of the Americas of the Economic Commission for Latin America and the Caribbean. Pucón, Chile. Disponible en <http://www.cepal.org/deype/noticias/documentosdetrabajo/0/49650/LCL3615i.pdf>
- Grubestic, T. H.; Wei, R. y Murray, A.T. (2014). Spatial Clustering Overview and Comparison: Accuracy, Sensitivity, and Computational Expense. *Annals of the Association of American Geographers*, 104(6), 1134-1156.
- Hicks, N. (2000). *An Analysis of the Index of Unsatisfied Basic Needs (NBI) for Argentina with Suggestions for Improvements*. In: Quinto Taller Regional sobre la Medición de la Pobreza: Métodos y Aplicaciones: documentos presentados, p. 101-110. Disponible en <http://repositorio.cepal.org/handle/11362/31528>



INDEC (s/f). *Unidades Geoestadísticas*. Disponible en <http://geoservicios.indec.gov.ar/codgeo/index.php?pagina=definiciones>.

Jacquez, G. M. (2008). Spatial Cluster Analysis. En S. Fotheringham and J. Wilson (Eds.), *The Handbook of Geographic Information Science*, capítulo 22, 395-416. Blackwell Publishing: Malden (MA), USA.

Ministerio de Energía y Minería (s/f). *Índice de Pobreza Multidimensional*. Proyectos. Presidencia de la Nación, Argentina. <https://sig.se.gob.ar/geoportal/index.php/proyectos/?id=22>

MPPN - Multidimensional Poverty Peer Network (s/f). *Multidimensional Poverty*. <http://www.mppn.org/multidimensional-poverty/what-is-multidimensional-poverty/>

Powell, M., Boyne, G. y Ashworth, R. (2001). Towards a geography of people poverty and place poverty. *Policy & Politics*, 29 (3), pp. 243-258. DOI: <https://doi.org/10.1332/0305573012501332>.

Sawada, M. (s/f). *Global Spatial Autocorrelation Indices - Moran's I, Geary's C and the General Cross-Product Statistic*. Department of Geography, University of Ottawa. Disponible en <http://www.lpc.uottawa.ca/publications/moransi/moran.htm>

Schumpeter, J. A. (1949). Science and ideology. *The American Economic Review*, 39(2), 346-359.

Sen, A. (1992). Sobre conceptos y medidas de pobreza. *Comercio exterior*, 42(4), 310-322.

UNDP – United Nations Development Program (s/f). *Multidimensional Poverty Index*. Human Development Report. Disponible en <http://hdr.undp.org/en/content/multidimensional-poverty-index-mpi>

Williamson, O. E. (1991). Comparative economic organization: The analysis of discrete structural alternatives. *Administrative Science Quarterly*, 36 (2), 269-296.



ANEXO

Tabla A. Estadísticas sumarias de las variables utilizadas

Variable	Media	Desvío Std.	Mediana	Mínimo	Máximo
EDUCA	0,133	0,115	0,093	0,000	0,441
HELAD	0,020	0,033	0,006	0,000	0,246
HACIN	0,024	0,044	0,008	0,000	0,412
MATER	0,095	0,145	0,033	0,000	0,908
SERVI	0,240	0,324	0,080	0,000	1,000

N = 1.066.

Fuente: Elaboración propia con datos de INDEC - IPEC Santa Fe, Censo 2010.