


03

UV Universidad
Verdad 78

AUTOCORRELACIÓN ESPACIAL - ÍNDICES PARA DETERMINAR SU PRESENCIA EN DATOS GEOGRÁFICOS: BREVE REVISIÓN DE LA LITERATURA.

Spatial Autocorrelation – Indices for determining its presence in geographic data: A brief literature review.

 **Paul Esteban Bravo López**, Facultad de Ciencias de la Administración – Instituto de Estudios de Régimen Seccional del Ecuador – IERSE, Universidad del Azuay. (Ecuador) (pbravo@uazuay.edu.ec) (<https://orcid.org/0000-0003-1812-5613>)

Resumen

La importancia del espacio radica en que todos los eventos que ocurren en nuestro planeta están ligados a una ubicación, no solo espacial sino también temporal; es decir, todo ocurre en algún lugar y en algún momento dado. Ante esto, es necesario identificar las propiedades relevantes de estos eventos. Estas propiedades se relacionan tanto con objetos del mundo real como con las relaciones espaciales entre ellos. En el contexto del análisis de datos espaciales, la dependencia espacial en los valores de los atributos se considera como una propiedad fundamental que se determina por la autocorrelación espacial, que ha sido estudiada por varios autores, quienes han brindado diferentes definiciones de esta; sin embargo se puede decir que la autocorrelación espacial permite analizar la variabilidad de un fenómeno en el espacio para determinar patrones y verificar su comportamiento. Cuando la autocorrelación espacial es fuerte, los valores cercanos de una variable están estrechamente relacionados entre sí; pero cuando es débil, o incluso inexistente, los valores de una variable se distribuyen aleatoriamente en el espacio.

Existen varios índices formales que miden la autocorrelación espacial a nivel local o global, los cuales son aplicados en diversos ámbitos de estudio y que se describen en esta breve revisión de literatura.

Abstract

Importance of space lies in the fact that all events that occur on our planet are linked not only to a spatial but also to a temporal location, i.e. everything happens somewhere and at some point, in time. In view of this, it is necessary to identify the relevant properties of these events. These properties relate both to real-world objects and to spatial relationships between them. In context of spatial data analysis, spatial dependence in attribute values is considered as a fundamental property determined by spatial autocorrelation that has been studied by several authors, who have provided different definitions of it, however, spatial autocorrelation makes it possible to analyze the variability of a phenomenon in space in order to determine patterns and verify its behavior. When spatial autocorrelation is strong, nearby values of a variable are closely related to each other; but when it is weak, or even nonexistent, the values of a variable are randomly distributed in space.

There are several formal indices that measure spatial autocorrelation at local or global level, which are applied in various fields of study and are described in this paper.

Palabras clave

Autocorrelación espacial, índices, análisis espacial, información geográfica.

Keywords

Spatial autocorrelation, indices, spatial analysis, geographic information.

Es necesario entender que, en el espacio geográfico, los objetos pueden representarse de manera geométrica, usualmente como puntos, líneas o polígonos; pero también es posible analizar la dependencia espacial de acuerdo con los valores de sus atributos.

La presencia de dependencia espacial significa que los valores del mismo atributo, medidos en ubicaciones cercanas, tienden a ser similares entre sí, y tienden a ser más similares que los valores separados por distancias más grandes. Por “similar” se entiende que si el valor de un atributo es grande (o pequeño), entonces los valores cercanos del mismo atributo tendrán una tendencia a ser también grandes (o pequeños) (Haining y Haining, 2003).

1.

Introducción

La importancia del espacio radica en que todos los eventos que ocurren en nuestro planeta están ligados a una ubicación no solo espacial, sino también temporal; es decir, todo ocurre en algún lugar y en algún momento dado. Sin embargo, la representación de estos eventos no es simple y se debe aplicar una abstracción de ellos para analizarlos.

Haining y Haining (2003) mencionan que es necesario identificar las propiedades relevantes de estos eventos. Estas propiedades se relacionan tanto con entidades (objetos del mundo real) como con las relaciones espaciales entre ellas. En el contexto del análisis de datos espaciales, la dependencia espacial en los valores de los atributos se considera como una propiedad fundamental.

Según Lloyd (2010), la frase “dependencia espacial” se ha definido como la falta de independencia de los datos en lugares próximos entre sí; por lo tanto, una medida de autocorrelación espacial puede sugerir dependencia o independencia espacial.

En este sentido, es importante destacar la Primera Ley de la Geografía: “Todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes” (Tobler, 1970, p. 236).

Las cuestiones anteriores permiten entender que, para un análisis adecuado de la dependencia o independencia espacial, es necesario contar con elementos que permitan determinarla. Dichos elementos denominados índices, por lo general son conocidos en el ámbito del análisis espacial; sin embargo, existen otros que pueden ser importantes al momento de verificar esta variación.

La presente revisión de la literatura resume algunos índices que permiten determinar esta variabilidad en datos geográficos para aplicarlos en diversos campos de estudio como análisis urbanos y territoriales, análisis de patrones económicos o análisis de salubridad (Siabato y Manrique, 2019).

2.

Revisión de literatura

2.1. ¿Qué es la autocorrelación espacial?

La geografía experimentó un mayor desarrollo con el surgimiento de varias metodologías de análisis espacial, las cuales permitieron afrontar nuevas problemáticas. En este sentido, la denominación “autocorrelación espacial” surgió en una conferencia realizada en 1967 por Andrew Cliff y Keith Ord (Siabato y Manrique, 2019).

La autocorrelación espacial ha sido estudiada por diversos autores, quienes brindan diferentes definiciones de esta; por ejemplo, Burt *et al.* (2009) la definen como la correlación de una variable consigo misma en el espacio. Esta definición se basa en la emitida por Odland (1988):

La autocorrelación espacial existe cuando una variable exhibe un patrón regular sobre el espacio en el que sus valores en un conjunto de ubicaciones dependen de los valores de la misma variable en otras ubicaciones. (p. 7)

Cuando la autocorrelación espacial es fuerte, los valores cercanos de una variable están estrechamente relacionados entre sí. Cuando la autocorrelación espacial es débil, o incluso inexistente, los valores de una variable se distribuyen aleatoriamente en el espacio.

La autocorrelación espacial puede ser negativa o positiva. Si los valores similares de una variable tienden a agruparse en el espacio, la distribución geográfica de esa variable está positivamente autocorrelacionada espacialmente, mientras que, si valores muy diferentes de una variable tienden a agruparse, esa variable está negativamente autocorrelacionada espacialmente (Burt *et al.* 2009).

Lloyd (2010) define a la autocorrelación espacial como la correlación de una variable consigo misma, es decir los valores en lugares cercanos tienden a ser similares. Además, menciona que ciertas propiedades (topográficas, climáticas, etc.) tienden a variar suavemente y suelen estar positivamente autocorrelacionadas espacialmente al menos en algunas escalas de medición.

Camarero y Rozas (2006) definen a la autocorrelación espacial como una medida del grado de similitud para una variable cuantitativa o categórica Z entre puntos vecinos en el espacio (por ejemplo, i, j). Desde el punto de vista de la ecología, la autocorrelación espacial de una o varias variables distribuidas en el espacio puede describirse mediante funciones de estructura, como los correlogramas y los semivariogramas (Fortin, 1999; Legendre y Fortin, 1989 citados en Camarero y Rozas, 2006). Si la autocorrelación es positiva indica que los puntos vecinos muestran valores similares de la variable Z y si la autocorrelación es negativa indica que los puntos vecinos muestran valores muy diferentes. La presencia de autocorrelación invalida la suposición de independencia en la que se apoyan muchos de los test estadísticos clásicos.

Para contrastar lo mencionado hasta aquí, es importante destacar la revisión realizada por Griffith (1992) sobre distintas definiciones dadas a la autocorrelación espacial, las cuales son: a) autocorrelación atribuible a la posición geográfica de los datos; b) descriptor del tipo de patrón cartografiado; c) índice del contenido de información de datos georeferenciados no detectable mediante estadística clásica; d) herramienta diagnóstica para detectar modelos espaciales erróneos; e) problema generado al aplicar estadística clásica a datos espaciales; f) indicador de la validez de la división en subunidades

de un área; g) mecanismo de procesos espaciales y h) efecto espacial no deseado. Sin embargo, a pesar de estas definiciones, Fortin *et al.* (2002 citado en Camarero y Rozas, 2006) destaca la existencia de cuatro tipos de autocorrelación espacial: espuria, interpolada, verdadera e inducida. Son las dos últimas las que pueden brindar pistas de los procesos que generaron el patrón espacial.

Para Levine (2015), el concepto de autocorrelación espacial es uno de los más importantes en estadística espacial, ya que implica la falta de independencia espacial. La estadística clásica asume que las observaciones son elegidas independientemente y no están relacionadas espacialmente entre sí. En otras palabras, la autocorrelación espacial es un arreglo espacial en el que se ha alterado la independencia espacial. Cuando los eventos, personas o instalaciones se agrupan, se hace referencia a esta disposición como autocorrelación espacial positiva; por el contrario, cuando están extremadamente dispersos se denomina autocorrelación espacial negativa, que es una disposición más rara, pero existente.

De forma resumida, la autocorrelación espacial permite analizar la variabilidad de un fenómeno en el espacio para determinar patrones y verificar su comportamiento, es decir, comprender la distribución espacial del fenómeno analizado (Siabato y Manrique, 2019).

Existen índices formales que miden la autocorrelación espacial a nivel zonal o para datos de nivel individual con atributos de conteo o intervalo. Los índices simples intentan identificar si existe autocorrelación espacial para una sola variable mientras que los índices más complicados intentan estimar la variabilidad en la autocorrelación espacial en un área de estudio sobre una variable de atributo en particular.

2.1.1. Índices globales y locales

Los índices espaciales globales resumen, en una sola cantidad, el grado en que un patrón espacial observado se desvía de una hipótesis nula especí-

fica como la de la aleatoriedad espacial completa (Rogerson y Yamada, 2009). Estos índices no proporcionan información sobre si la incidencia puede aumentar significativamente alrededor de lugares particulares ni revelan información sobre el tamaño, la ubicación y la importancia de los conglomerados geográficos localizados. Cuando los índices globales demuestran ser significativos, es de interés saber cuáles son las regiones responsables de la significación. Por ello, es importante considerar los índices locales o específicos; algunos de ellos se han diseñado para probar la hipótesis nula “los conteos observados son elevados en las cercanías de un lugar en particular, en relación con los conteos esperados”, mientras que otros se desarrollaron para probar la hipótesis nula “los valores observados para una variable son espacialmente independientes en las proximidades de un lugar”.

Una propiedad deseable de los índices locales es que la suma de índices locales, en todas las regiones, sea igual a un múltiplo del índice global. Así, el índice global puede descomponerse en un conjunto de índices regionales y locales.

Levine (2015) recalca que, a veces no es posible analizar los datos a nivel individual; el investigador puede requerir agregar puntos de datos individuales a áreas espaciales (zonas) para comparar los eventos con los datos que sólo se obtienen para las zonas (por ejemplo, datos censales). Los datos zonales pueden incluir conteos de eventos por zona, e información complementaria, más otros datos que se agregan a las áreas pequeñas. En resumen, la zona se convierte en la unidad de análisis en lugar de los puntos de datos individuales.

Como las zonas no son eventos, deben ser analizadas espacialmente asumiendo que todos los datos residen en un solo punto dentro de ellas. Normalmente es el centroide (centro geográfico de la zona), pero a veces también se utiliza el centro de la distancia mínima (punto en el que se minimiza la suma de las distancias a todos los demás puntos), especialmente si la zona tiene forma irregular. Sin embargo, al asignar puntos de datos individuales a las zonas, se pierde información.

Los datos zonales también deben incluir una variable de atributo, una variable asociada con la zona. Este atributo puede ser un recuento o una variable continua para una propiedad distributiva de la zona o una variable binaria (es decir con valores 1 o 0).

Existen varios índices útiles para verificar el agrupamiento general donde la hipótesis nula es la aleatoriedad espacial en la distribución de los valores de los datos. Las pruebas pueden basarse en la distancia, cuando se dispone de datos a nivel individual o estar basadas en áreas que cuentan el número de eventos en subáreas definidas (Haining y Haining, 2003). Los índices analizados a nivel global son: I de Moran, C de Geary, Getis-Ord G; y a nivel local: LISA, Anselin Local Moran y Getis-Ord local; sin embargo, existen otros índices que también serán brevemente mencionados. Se debe acotar que los índices que se describen a continuación son globales porque representan un valor de resumen para todos los puntos (Levine, 2015).

2.1.1.1. I de Moran

El índice I de Moran es uno de los indicadores más antiguos. Se aplica a zonas o puntos que tienen variables de atributo asociadas a ellas. Para cualquier variable continua se puede calcular X_i , una media y también se puede calcular la desviación de cualquier observación de esa media. El índice compara el valor de la variable en cualquier lugar con el valor en todos los demás lugares. Matemáticamente, I de Moran se define como (Levine, 2015):

$$I = \frac{N \sum_i \sum_j W_{ij} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{ij}) \sum_i (X_i - \bar{X})^2} \quad (1)$$

Donde N es el número de casos, X_i es el valor de una variable en una ubicación particular, i , X_j es el valor de la misma variable en otra ubicación (donde i no es igual a j), \bar{X} es la media de la variable y W_{ij} es un peso aplicado a la comparación entre las ubicaciones i y j (Ecuación 1).

Al igual que un coeficiente de correlación estadístico, típicamente varía entre -1.0 y $+1.0$, sin embargo, esto no es absoluto. Un valor alto de I indica más autocorrelación espacial que un valor de I bajo (Levine, 2015). Cuando I es positivo, indica una agrupación de valores similares, mientras que cuando I es negativo, indica una agrupación de valores diferentes. Un valor de cero indica autocorrelación espacial ausente (Lloyd, 2010).

Rogerson y Yamada (2009) definen los valores de I de Moran cercanos a $+1$ como indicadores de un fuerte patrón espacial (los valores altos y bajos tienden a estar cerca uno del otro). Los valores cercanos a -1 indican una fuerte autocorrelación espacial negativa; es decir los valores altos tienden a estar cerca de los valores bajos. Si bien, los patrones espaciales con autocorrelación negativa son extremadamente raros; existen excepciones. Finalmente, los valores cercanos a cero indican una ausencia de patrón espacial. Cabe señalar que, según Rogerson y Yamada (2009) el rango de I de Moran no se ajusta exactamente al intervalo $[-1, 1]$, diferente del coeficiente de correlación ordinario.

Camarero y Rozas (2006) mencionan de acuerdo con la ecuación 1, que los pesos w_{ij} deciden la proximidad o “vecindario” de cada punto analizado. Por analogía con el ajedrez, se usan tres tipos de vecindad, según la posición de los puntos próximos: torre (cuatro vecinos), alfil (cuatro vecinos en posiciones diagonales) y reina (ocho vecinos). La definición del tipo de proximidad puede ser determinante para los resultados obtenidos (Cliff y Ord, 1981, citados en Camarero y Rozas, 2006).

En síntesis, para el caso de autocorrelación positiva, I tiende a $+1$; si la autocorrelación no existe $I = 0$ y para el caso de autocorrelación negativa I tiende a -1 .

El índice I de Moran es uno de los más utilizados y aplicados en varios y diversos estudios (Siabato y Manrique, 2019). Algunos casos puntuales de su aplicación son trabajos como el de Wilt et al. (2018), quienes realizaron un análisis espacio temporal de los efectos del huracán Sandy en pruebas de VIH

mediante I de Moran para detectar autocorrelación espacial en las zonas estadounidenses afectadas; o el de Yuan *et al.* (2018) quienes analizaron la contaminación con tierras raras de los suelos urbanos en Londres, detectando *hotspots* a través del índice en mención.

2.1.1.2. C de Geary

Es similar a I de Moran, sin embargo, la interacción no es el producto cruzado de las desviaciones de la media, sino la desviación de las intensidades de la localización de cada observación entre sí. Su definición matemática es (Levine, 2015):

$$C = \frac{(N-1) \sum_i \sum_j W_{ij} (X_i - X_j)^2}{2(\sum_i \sum_j W_{ij}) \sum_i (X_i - \bar{X})^2} \quad (2)$$

Donde N es el número de casos, X_i es el valor de una variable en una ubicación particular, i , X_j es el valor de la misma variable en otra ubicación (donde i no es igual a j), \bar{X} es la media de la variable y W_{ij} es un peso aplicado a la comparación entre las ubicaciones i y j (Ecuación 2).

Los valores de C varían típicamente entre 0 y 2, aunque 2 no es un límite superior estricto. El valor teórico de C es 1; es decir, si los valores de una zona no están relacionados espacialmente con ninguna otra zona, entonces el valor esperado de C sería 1. Los valores inferiores a 1 (entre 0 y 1) indican autocorrelación espacial positiva, mientras que los valores superiores a 1 indican autocorrelación espacial negativa. El índice de Moran proporciona un indicador más global, mientras que el coeficiente de Geary es más sensible a las diferencias en vecindarios pequeños. Se puede obtener una interpretación más intuitiva de C calculando un C ajustado = $1 - C$. En este caso, C ajustado estará en la misma escala que I de Moran. Un valor C ajustado que sea positivo indica una autocorrelación espacial positiva, mientras que un valor C ajustado que sea negativo indica una autocorrelación espacial negativa.

Por lo general, C de Geary será consistente con I de Moran, aunque existen ligeras diferencias. Debido a la naturaleza de la ponderación, C de Geary es más sensible a la agrupación local (o efectos de segundo orden) que I de Moran, que se considera mejor para medir la autocorrelación espacial de primer orden. Ambos índices deben utilizarse con cuidado y no pueden generalizarse por sí mismos (Levine, 2015). Cada uno enfatiza información ligeramente diferente con respecto a la autocorrelación espacial, pero ninguno de los dos es suficiente por sí solo.

C también se conoce como cociente de contigüidad de Geary y según Rogerson y Yamada (2009) los valores pequeños de C indican autocorrelación espacial positiva, mientras que los valores grandes de C indican autocorrelación espacial negativa. Tanto I como C pueden generalizarse fácilmente para permitir la medición de la autocorrelación espacial, mediante observaciones separadas por una determinada distancia o rango de distancias (Lloyd, 2010).

Camarero y Rozas (2006) indican que la interpretación de C es opuesta a la de I de Moran, ya que los valores de C que tienden a 0 ($c < 1$) indican autocorrelación positiva, mientras que cuando C tiende a 2 ($c > 1$) corresponde a autocorrelación negativa y, en ausencia de autocorrelación espacial significativa, $C = 1$. Es decir que altos valores de C equivalen a bajos valores de I y viceversa.

La elección entre I y C depende del investigador, ya que cada uno ofrece ventajas y desventajas. Cliff y Ord (1981 en Camarero y Rozas, 2006) señalan que I parece estar menos afectado por la distribución de los datos que C , y adjudican una ligera ventaja estadística a I respecto de C . Sin embargo, I parece ser más sensible a los valores extremos que C (Legendre y Fortin, 1989 citados en Camarero y Rozas, 2006).

En general, se considera que I de Moran tiene mayor poder estadístico en comparación con C de Geary; por lo tanto tiene mayor capacidad de detectar la autocorrelación espacial cuando existe (Cliff y Ord 1981; Haining 1990 citados en Rogerson y Yamada,

2009). De manera general, la comparación entre I y C sirve para enfatizar el hecho de que no existe un mejor índice, pues I de Moran lo hará mejor frente a algunas alternativas y C de Geary funcionará mejor con otras.

2.1.1.3. Getis-Ord G

Este es un índice de autocorrelación espacial global, pero para valores que se encuentran a una distancia especificada entre sí (Levine, 2015). Tiene una ventaja sobre los otros dos índices globales en el sentido de que puede distinguir entre *hotspots* y *coldspots*, lo que ni I de Moran ni C de Geary pueden hacer. G calcula la interacción espacial del valor de una variable particular en una zona con los valores de esa misma variable en zonas cercanas, similar a I y C; por lo tanto, también es una medida de asociación o interacción espacial. A diferencia de las otras dos medidas, solo identifica la autocorrelación espacial positiva, es decir, donde las zonas tienen valores similares a sus vecinos. No puede detectar autocorrelación espacial negativa donde las zonas tienen valores diferentes a sus vecinos. La identificación mencionada distingue entre autocorrelación espacial positiva en zonas con valores altos, cerca de otras zonas con valores altos (autocorrelación espacial positiva alta) y autocorrelación espacial positiva en zonas con valores bajos cerca de otras zonas también con valores bajos (autocorrelación espacial positiva baja). Además, el valor de G (Ecuación 3) se calcula con respecto a una distancia de búsqueda especificada (definida por el investigador) en lugar de una distancia inversa, como con I o C (Lee y Wong, 2005):

$$G(d) = \frac{\sum_i \sum_j W_j(d) X_i X_j}{\sum_i \sum_j X_i X_j} \quad (3)$$

Para una variable X, la ecuación anterior indica que el producto cruzado del valor de X en la ubicación i, y en otra zona, j, es ponderado por un peso de distancia, $W_j(d)$ que se define ya sea por “1” si las dos zonas son iguales o más cercanas que una distancia de umbral (d), o “0” en caso contrario. El producto cruzado es sumado para todas las demás zonas j sobre todas las zonas i. Por lo tanto, el numerador

es un subconjunto del denominador y puede variar entre 0 y 1. Si la distancia seleccionada es demasiado pequeña para que ninguna otra zona esté más cerca de esta distancia, entonces el peso será 0 para todos los productos cruzados de la variable X; por lo tanto, el valor de G(d) será 0. De manera similar, si la distancia seleccionada es demasiado grande para que todas las otras zonas estén más cerca de esta distancia, entonces el peso será 1 para todos los productos cruzados de la variable X; por lo tanto, el valor de G(d) será 1 (Lee y Wong, 2005).

Levine (2015) destaca la existencia de dos estadísticas G. La primera, G^* , incluye la interacción de una zona consigo misma; es decir, la zona ‘i’ y la zona ‘j’ pueden ser la misma zona. La segunda G no incluye la interacción de una zona consigo misma. Esencialmente, con G, el índice mide la interacción de una zona con zonas cercanas (un vecindario).

Es importante mencionar que G por sí solo no es muy significativo, ya que puede variar entre 0 y 1, a medida que aumenta la distancia de umbral, el índice siempre se aproxima a 1; en consecuencia, G se compara con un valor esperado de G bajo ninguna asociación espacial significativa.

En relación con el valor esperado de G, un valor Z positivo indica agrupación espacial de valores altos (autocorrelación espacial positiva alta o *hotspots*), mientras que un valor Z negativo indica agrupación espacial de valores bajos (autocorrelación espacial positiva baja o *coldspots*). Un valor de G alrededor de 0 indica que no hay autocorrelación espacial positiva, autocorrelación espacial negativa (la cual no puede detectar), o que el número de *hotspots* equilibra el número de *coldspots* (Levine, 2015).

Se debe considerar que el valor de esta prueba variará con la distancia de búsqueda seleccionada. Una distancia de búsqueda puede dar lugar a una asociación espacial significativa para G, mientras que otra no; es decir, el índice es útil para identificar las distancias a las que existe autocorrelación espacial, aunque en la práctica, se debe usar una pequeña distancia de búsqueda para identificar autocorrelación espacial local.

Una ventaja de G sobre los índices antes descritos es que puede distinguir entre *hotspots* y *coldspots*. Con I o C, un indicador de autocorrelación espacial positiva significa que las zonas tienen valores similares a sus vecinos. Sin embargo, esta autocorrelación podría ser causada por muchas zonas con valores bajos que también se concentran. En otras palabras, no se puede decir a partir de esos dos índices si la concentración es un punto caliente o un punto frío, mientras que con G esto es posible (Levine, 2015).

La principal limitación de G es que no puede detectar autocorrelación espacial negativa, una condición que, si bien es rara, sí ocurre y para ello deberían utilizarse los índices I o C, pero especialmente I (Levine, 2015).

Rogerson y Yamada (2009) mencionan que, para un patrón dominante de valores altos cerca de otros valores altos, el índice G es alto, junto con un valor Z positivo alto, y es bajo cuando hay una tendencia general de agrupación de valores bajos. En ejemplos empíricos, Getis y Ord enfatizan el uso conjunto del índice I con G. Si un patrón espacial de valores altos o bajos es dominante, I mostrará un valor altamente positivo. G es capaz de discriminar entre los dos patrones. En un patrón en el que predomina la agrupación de valores altos, la estadística G será alta; será baja si se agrupan valores bajos. Aunque el índice global I es una medida de autocorrelación espacial, G puede ser visto como una medida de concentración espacial.

Rogerson y Yamada (2009) mencionan brevemente algunos índices adicionales:

Estadístico Oden's I_{pop} : Oden (1995 citado en Rogerson y Yamada, 2009) explicó esta variación modificando I de Moran, ya que una de las características de I es que las variaciones dentro de la región pueden debilitar la validez de los supuestos de aleatorización o normalidad. Por ejemplo, se puede esperar que las regiones con poblaciones pequeñas muestren más variabilidad.

Estadística de Tango: Es un promedio ponderado de las covariaciones de desviaciones entre las frecuencias observadas y esperadas, para todos los pares de puntos.

Estadística espacial de chi-cuadrado R: Rogerson (1999) desarrolló y evaluó una estadística que se puede usar como una prueba global de agrupamiento. R será grande cuando existan grandes desviaciones entre los valores observados y los esperados dentro de las regiones, o cuando pares de regiones cercanas tengan desviaciones similares.

Cuzick-Edwards Test es una prueba global o general de *clustering*, útil con datos de caso - control (Epidemiología). La prueba estadística es un simple recuento de los vecinos más cercanos (*k-nearest*) de un caso (que también son casos), sumados sobre todos los casos.

Join Counts: Camarero y Rozas (2006) indican que las variables nominales más usadas en ecología son las binarias (0 refleja ausencia y 1 presencia), para las cuales se han desarrollado, sobre todo desde la geografía, índices como *join counts* que permiten estimar la autocorrelación espacial en unidades de muestreo adyacentes. Para el caso binario, la hipótesis nula considera que la probabilidad de que regiones vecinas sean del mismo tipo, por ejemplo 1-1 o 0-0, es mayor y por tanto se aleja de la aleatoriedad espacial completa (*Complete Spatial Randomness CSR*). Lloyd (2010) lo define brevemente como un método que evalúa el grado de agrupamiento/dispersión principalmente en variables categóricas (cualitativas).

2.1.2 Índices locales de autocorrelación espacial

Las medidas descritas anteriormente proporcionan resúmenes globales, sin embargo, se han realizado varias adaptaciones de estos enfoques para permitir la evaluación de la variación local en la autocorrelación espacial (Lloyd, 2010). A continuación, se describen algunos índices locales.

2.1.2.1 Local Indicator of Spatial Association (LISA)

Levine (2015) indica que el concepto básico detrás de una medida de autocorrelación espacial específica de una zona es el de un Indicador Local de Asociación Espacial (LISA) y ha sido discutido por varios investigadores como Mantel (1967) o Getis (1991) (citados en Anselin, 1995). LISA se define como cualquier estadística que satisfaga dos requisitos (Anselin, 1995):

1. LISA para cada observación indica hasta qué punto existe una agrupación espacial significativa de valores similares en torno a esa observación; y
2. La suma de los LISA para todas las observaciones es proporcional al indicador global de asociación espacial. Si la preocupación es evaluar el grado en que una estadística global es representativa del patrón promedio de asociación local, entonces este requisito es importante.

Además, LISA requiere dos condiciones: Primero, cada observación tiene un valor de una variable de atributo que se le puede asignar (es decir, un valor de intensidad o peso) a más de sus coordenadas X-Y. Segundo, se debe definir el vecindario. Puede tratarse de zonas adyacentes, de todas las demás zonas ponderadas negativamente por la distancia desde la zona de observación, o de todas las demás zonas ponderadas negativamente por la distancia desde la zona de observación hasta una cierta distancia a partir de la cual el peso es cero (ancho de banda). Una vez definidas, LISA indica el valor de la zona de observación en relación con su vecindario.

Anselin (1995) definió un conjunto de LISA que determina las formas locales de I de Moran y de C de Geary. Además, los índices de LISA permiten separar los indicadores globales (Lloyd, 2010).

Burt *et al.* (2009) se refieren a LISA como una estadística útil para comparar la similitud del valor de una variable X en un lugar dado A, con el valor de la misma variable en lugares vecinos. Si las ubicaciones vecinas tienen valores de la variable de interés

similares al valor encontrado en A, entonces se genera un LISA positivo; si los valores de la variable X son muy diferentes entre A y sus vecinos, se obtendrá un LISA negativo.

Un ejemplo puntual de aplicación de LISA, dentro de los múltiples existentes, es el trabajo de Darand *et al.* (2017) quienes utilizaron esta estadística para analizar la precipitación extrema en Irán, determinando la existencia de patrones espacio temporales.

2.1.2.2. Getis-Ord Local

Getis y Ord definieron G, como una medida global de asociación que permite evaluar la agrupación espacial en los valores de un atributo. La forma local de G se utiliza para medir patrones espaciales alrededor de cada área (Levine, 2015).

Getis-Ord Local aplica el índice G a zonas individuales para evaluar si una zona en particular está espacialmente relacionada con las zonas cercanas aplicándose a cada zona individual. Según Levine (2015), este índice es idóneo para identificar *hotspots* y *coldspots*, sin embargo, su mayor limitación es que no puede detectar una autocorrelación espacial negativa cuando una zona está rodeada de vecinos que son diferentes, ya sea por tener un valor alto rodeado de zonas con valores bajos o por tener un valor bajo y estar rodeada de zonas con valores altos.

Burt *et al.* (2009) mencionan, respecto a G, que no existe una comparación explícita del valor de una variable X entre una o más ubicaciones. Lo que la estadística G (o G*) identifica son áreas de una región de estudio donde existe concentración mayor que el promedio de la suma de todos los valores X para esa región (autocorrelación espacial positiva), o áreas de una región de estudio donde hay una concentración menor que el promedio de la suma de todos los valores X (autocorrelación espacial negativa). Se debe ser muy cuidadoso al interpretar zonas con valores G significativos y probablemente serviría mejor elegir solo aquellas zonas con el mayor o menor valor de G (Levine, 2015).

El creciente interés por investigar la asociación espacial se demuestra en la gran cantidad de literatura disponible (Siabato y Manrique, 2019) y aplicada a diversos ámbitos como por ejemplo el trabajo de Stankov y Dragičević (2015), quienes utilizaron los índices Getis-Ord local e I de Moran para determinar cambios en el patrón espacial de los valores de ingresos netos en Serbia.

2.1.2.3 Anselin Local Moran

Fue desarrollado por Luc Anselin y es el LISA más antiguo (Anselin, 1995). El procedimiento aplica I de Moran a zonas individuales, lo que permite identificarlas como similares o diferentes a su patrón cercano (Levine, 2015). En este sentido es importante destacar que, dado que I global, mide la similitud en observaciones sobre un área de estudio, Local Moran también indica la similitud de una zona en relación con sus vecinos. Así, en vecindarios donde tanto la zona como sus vecinos tienen valores de atributos altos, Local Moran será positivo indicando que la zona en particular es similar (es decir, también será alta). De manera similar, en vecindarios donde tanto la zona como sus vecinos tienen valores de atributo bajos, Local Moran también será positivo indicando que la zona es similar a sus vecinos (es decir, también será baja). Cuando Local Moran es positivo, se trata de un indicador de **similitud**; por el contrario, si una zona tiene un valor alto de la variable de intensidad mientras que sus vecinos tienen valores bajos o, alternativamente, tiene un valor bajo mientras que los vecinos tienen valores altos, entonces Local Moran será negativo. En síntesis, Local Moran indica si la zona es similar o disimilar a sus vecinos (Levine, 2015). Un requerimiento esencial para Local Moran, es resumir los datos en zonas a fin de producir el valor de intensidad o peso necesario, lo cual puede implicar trabajo adicional (Levine, 2015).

Otros índices destacados por Rogerson y Yamada (2009) para determinar autocorrelación espacial local son:

Score Statistic: Sirve para probar la hipótesis nula de que no hay incidencia aumentada alrededor de un sitio preespecificado i .

Tango C_F Statistic: Tango (1995 en Rogerson y Yamada, 2009) utilizó una estadística de puntuación modificada y generalizada para evaluar grupos alrededor de focos preespecificados. Su ventaja es permitir que se especifique más de un punto focal simultáneamente.

Getis G_i Statistic: La única diferencia entre las dos variantes de la estadística local G es si la región objetivo i está incluida o no en el cálculo de la estadística. Aunque Getis y Ord no mencionan ninguna especificidad sobre las situaciones en que una es más adecuada que la otra, G_i^* intuitivamente parece ser una medida más natural de la tendencia de agrupación local que G_i . Este último puede detectar como un *cluster* una región con un valor bajo, pero rodeada de regiones con valores altos, lo que a menudo es poco probable que sea el tipo de *cluster* que se quiere detectar. Sin embargo, G_i es útil cuando una región objetivo i tiene un valor sustancialmente alto en comparación con sus regiones circundantes porque el alto valor de la región objetivo, si se incluye, puede hacer que casi todo sea significativo.

Moving Window: Según Lloyd (2010), este es uno de los enfoques más utilizados para la adaptación local en el análisis espacial. Se puede utilizar *moving window* para estimar estadísticas basándose en una ponderación igual de los datos dentro de la ventana. Alternativamente, puede utilizarse un esquema de ponderación geográfica en el que las observaciones se ponderan en función de su distancia del lugar de interés (por ejemplo, el centro de *moving window*).

Son muy utilizadas en el procesamiento de imágenes y se conocen como “operadores focales”. Con un operador focal, los valores de salida son una función de las celdas vecinas. En otras palabras, si se utiliza un filtro de tres por tres píxeles, una forma de operador focal, entonces se calcula la media de las celdas en las ventanas y se escribe el valor

medio en la ubicación de la cuadrícula de salida que corresponde a la ubicación de la celda central en la ventana. Los tamaños diferentes de las ventanas pueden capturar la variación espacial a diferentes escalas.

3.

Conclusiones

Estudiar la autocorrelación espacial es importante porque permite conocer como varía un fenómeno espacialmente (Siabato y Manrique, 2019), lo cual es útil en distintos ámbitos del conocimiento, ya que todo lo que ocurre en nuestro planeta está ligado a un espacio. Una manera de conocer esta variación es mediante los índices de autocorrelación espacial descritos en este trabajo.

Debido a la cantidad de aplicaciones encontradas en la literatura, se podría decir que I y C son los índices más extendidos para determinar la existencia de autocorrelación espacial (Siabato y Manrique, 2019); sin embargo, se han investigado de forma sucinta una diversidad de estadísticas derivadas de estos, que permiten analizar el fenómeno mencionado y pueden aplicarse para fortalecer su hallazgo.

Si bien no existe una única manera de utilizar estos índices, es relevante conocer cómo usarlos, identificando primeramente si existe autocorrelación espacial positiva mediante I y C.

La autocorrelación espacial positiva indica que las zonas están situadas cerca de otras zonas con valores similares, ya sea que las zonas con valores altos en la variable estén situadas cerca de zonas también con valores altos o la condición opuesta (valores bajos cerca de otros valores bajos).

Si existe autocorrelación espacial positiva, es conveniente identificar el tipo utilizando G, el cual sólo se aplica para autocorrelación espacial positiva, pero puede distinguir un predominio de alto positivo, es decir, zonas con valores altos situadas cerca de otras zonas también con valores altos o bajo positivo, que significa lo contrario (zonas con valores bajos cerca de zonas con valores bajos). Como este índice requiere una distancia de búsqueda, el investigador debe identificar una distancia para la cual la autocorrelación espacial positiva es más distinguible; esto se logra ejecutando el correlograma Geis-Ord (Levine, 2015).

Si ambos índices son significativos, es una prueba contundente de que existe autocorrelación espacial considerable en los datos; sin embargo, no es fácil determinar a partir de los índices si la autocorrelación espacial se debe a factores globales (regionales) o locales. En este sentido, si I es significativo, pero C no lo es, esto podría indicar que el agrupamiento es una función de la concentración global más que de la concentración local, ya que I es más sensible a la variación regional. Si hay autocorrelación espacial negativa, que ocurre ocasionalmente, indica que las zonas con valores altos están ubicadas cerca de las zonas con valores bajos, o viceversa (Levine, 2015).

Finalmente, al ser la presente una revisión introductoria, es recomendable una investigación más profunda de los diferentes índices para que el investigador cuente con una conceptualización sólida y pueda aplicarlos correctamente en la medición de la distribución y variabilidad de eventos geográficos.

Referencias

- Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis*, 27(2), 93-115.
<https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Burt, J. E., Barber, G. M., y Rigby, D. L. (2009). *Elementary statistics for geographers*. Guilford Press.
- Camarero, J., y Rozas, V. (2006). Spatial surface-pattern analyses and boundary detection techniques applied in forest ecology. *Forest Systems*, 15(1), 66-87.
<http://dx.doi.org/10.5424/srf/2006151-00954>
- Darand, M., Dostkamyán, M., y Rehmani, M. I. A. (2017). Spatial autocorrelation analysis of extreme precipitation in Iran. *Russian Meteorology and Hydrology*, 42(6), 415-424.
<https://doi.org/10.3103/s1068373917060073>
- Griffith, D. A. (1992). What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *Espace Géographique*, 21(3), 265-280.
<https://doi.org/10.3406/spgeo.1992.3091>
- Haining, R. P., y Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. (p. 454). Cambridge University Press.
- Lee, J. y Wong, D. W. S. (2005). *Statistical Analysis with ArcView GIS and ArcGIS*. J. Wiley & Sons, Inc.: New York.
- Levine, N. (2015). Chapter 5: Spatial Autocorrelation Statistics. En N. Levine, *CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations, Version 4.0*.
- Lloyd, C. D. (2010). *Local models for spatial analysis*. CRC Press.
- Odland, J. (1988). *Spatial autocorrelation*. Reprint. Edited by Grant Ian Thrall. WVU Research Repository, 2020.
<https://researchrepository.wvu.edu/rri-web-book/20/>
- Rogerson, P.A. (1999), The Detection of Clusters Using a Spatial Version of the Chi-Square Goodness-of-Fit Statistic. *Geographical Analysis*, 31: 130-147.
<https://doi.org/10.1111/j.1538-4632.1999.tb00973.x>
- Rogerson, P., y Yamada, I. (2008). *Statistical detection and surveillance of geographic clusters*. CRC Press.
- Siabato, W., y Manrique, J. G. (2019). La autocorrelación espacial y el desarrollo de la geografía cuantitativa. *Cuadernos de Geografía: Revista Colombiana de Geografía*, 28(1), 1-22.
<http://dx.doi.org/10.15446/rcdg.v28n1.76919>
- Stankov, U., y Dragičević, V. (2015). Changes in the spatial pattern of net earnings: Evidence from Serbia. *Acta Oeconomica*, 65(3), 351-365.
<https://doi.org/10.1556/032.65.2015.3.1>
- Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234-240.
<https://doi.org/10.2307/143141>
- Wilt, G. E., Adams, E. E., Thomas, E., Ekperi, L., LeBlanc, T. T., Dunn, I., Molinari, N.A., y Carbone, E. G. (2018). A space time analysis evaluating the impact of hurricane sandy on HIV testing rates. *International Journal of Disaster Risk Reduction*, 28, 839-844.
<https://doi.org/10.1016/j.ijdrr.2018.04.003>

Yuan, Y., Cave, M., y Zhang, C. (2018). Using Local Moran's I to identify contamination hotspots of rare earth elements in urban soils of London. *Applied geochemistry*, 88, 167-178.
<https://doi.org/10.1016/j.apgeochem.2017.07.011>