

### ANEXO III METODOLOGÍA PARA LA MEDICIÓN DE LA AUTOCORRELACIÓN ESPACIAL Y LA HETEROGENEIDAD ESPACIAL DE LA POBREZA INFANTIL EN AMÉRICA LATINA Y EL CARIBE

Cuando se hace un análisis exploratorio de datos que incluye el componente geográfico con información sobre la localización de los fenómenos a estudiar, hay que considerar de forma explícita las características propias de esta información: georreferenciación, multidireccionalidad y multidimensionalidad. La primera ley de geografía de Tobler señala que "todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes" (Tobler, 1970, en Miller, 2004). Es decir, es probable que una unidad territorial que presenta altos índices de pobreza infantil también tenga una relación con los índices de pobreza de sus vecinos. Los nuevos métodos estadísticos bajo el análisis exploratorio de datos espaciales permiten considerar la localización del fenómeno en estudio, tanto en términos absolutos (dónde) como relativos (distribución espacial, distancia). Además, permiten no solo la fácil visualización de la pobreza infantil, sino también el explorar cómo se comporta en el espacio y qué patrón sigue. Sobre la base de este concepto se habla de conglomerado espacial de determinadas características (*spatial clustering*), lo que significa que el grado de correlación que existe entre indicadores está en función de la contigüidad y distancia entre las unidades territoriales.

Muchas veces este conglomerado espacial es examinado solo por el hecho de ser un factor desconcertante en los modelos estadísticos probabilísticos tradicionales, los que asumen que las variables que se calculan sobre la base de observaciones independientes están distribuidas de manera idéntica. Pero el conglomerado espacial en

sí –sea medido por la autocorrelación espacial o por la heterogeneidad espacial que lo subyace– también es una fuente de información relevante sobre la relación entre el territorio y las características sociodemográficas. La identificación de estos conglomerados espaciales de pobreza infantil permite la exploración de las causas, mecanismos y efectos de esta, lo que no necesariamente se lograría con una mera clasificación estadística de unidades territoriales ordenadas según el nivel de incidencia de este fenómeno. Si bien el conglomerado espacial no es prueba definitiva del efecto del territorio en la pobreza que experimentan los niños y adolescentes, sí lo es la localización, que ofrece evidencias sobre la posible relación y causalidad entre variables, que debería ser examinada con mayor profundidad (Miller, 2004).

Vilalta y Perdomo (2005) proponen una definición simple para la autocorrelación espacial y la heterogeneidad espacial: la primera es "la concentración o dispersión de los valores de una variable en un mapa", mientras que la segunda es "la variación de las relaciones entre las variables en el espacio" (pág. 325). En un artículo fundamental sobre la identificación de patrones de conglomerado espacial, Anselin (1995) presenta una profunda discusión sobre las medidas de autocorrelación y heterogeneidad espacial. En esta sección se exponen los elementos esenciales de su artículo.

Como uno de los primeros pasos para la caracterización de la distribución espacial en las ciencias sociales, es necesario hacer un análisis exploratorio de datos espaciales (*Exploratory Spatial*

*Data Analysis, ESDA*). En palabras de Anselin (1999), ESDA puede definirse como el conjunto de técnicas que describen y visualizan las distribuciones espaciales, identifican localizaciones atípicas o "atípicos espaciales" (*spatial outliers*), identifican patrones de asociación espacial, conglomerados (*clusters*) o enclaves calientes (*hot spots*) e indican estructuras espaciales u otras formas de heterogeneidad espacial. Por tanto, estos métodos tienen un carácter descriptivo (estadístico) más que confirmatorio (econométrico).

Tradicionalmente las herramientas relacionadas con un proceso de análisis exploratorio de datos espaciales permitían calcular un solo indicador para la totalidad de las unidades territoriales de una determinada área de estudio. Estos indicadores de autocorrelación espacial, tales como el I de Moran global –uno de los más antiguos (Moran, 1950), ofrecían una medida de resumen de la intensidad del conglomerado espacial de las unidades territoriales consideradas, pero carecían de una indicación de cuáles unidades formaban parte de tal o cual conglomerado espacial y de dónde se localizaban estos conglomerados. En suma, los indicadores de autocorrelación global solo señalan el grado de la relación espacial entre valores de las unidades territoriales, pero no entregan una indicación del patrón de esta relación (agrupada o dispersa).

Por otra parte, el conjunto de indicadores de asociación espacial local (*Local Indicators of Spatial Association, LISA*) son derivados de los indicadores globales de autocorrelación espacial, tradicionalmente calculados en ejercicios de análisis exploratorio de datos espaciales. Según Anselin (1995), para que un indicador sea considerado de asociación espacial local, es necesario que cumpla con dos condiciones: la primera es que la suma de los valores del indicador de asociación espacial local para cada unidad territorial sea proporcional a un indicador global de autocorrelación espacial; y la segunda es que identifique los conglomerados espaciales de unidades territoriales. Por lo tanto, la ventaja adicional de los indicadores de asociación espacial local respecto de los indicadores globales de autocorrelación es su identificación de heterogeneidad espacial, que asigna a cada unidad territorial un valor que indica el tipo de relación que tiene con sus unidades vecinas. En consecuencia, los indicadores de asociación espacial local no solo identifican patrones de asociación espacial, sino también los valores extremos del comportamiento territorial de las variables –aquellas unidades territoriales con valores atípicos, donde su relación espacial cae fuera de la distribución de las demás unidades.

Existen también indicadores que identifican conglomerados espaciales, pero sin un correspondiente indicador global de asociación espacial, como el  $G_i^*$  de Getis-Ord (Getis y Ord, 1992).

Dada la valiosa información que brindan tanto la medida global de la autocorrelación espacial como la medida del patrón de esta autocorrelación, en este documento el conglomerado espacial de unidades territoriales se mide por medio del I de Moran, que es a la vez una medida de autocorrelación y de heterogeneidad espacial y representa el indicador utilizado con mayor frecuencia en este tipo de indagación. Una ventaja adicional de usar el I de Moran es la facilidad con que se pueden calcular y ejecutar, al mismo tiempo, análisis espaciales descriptivos con el *software* GeoDa, diseñado para estos fines<sup>1</sup>.

### El I de Moran global y local

Como se señaló anteriormente, el I de Moran puede ser presentado en dos formas distintas: el índice global y el índice local. Tanto el primero como el segundo varían entre -1 y 1, donde los valores negativos indican un conglomerado espacial de unidades territoriales con valores de análisis distintos y valores positivos indican un conglomerado espacial de unidades territoriales con valores de análisis similares –sean estos altos o bajos. Los valores del I de Moran cercanos a cero indican la falta de una relación espacial entre los valores de análisis. Además, como se destacó en la sección previa, el I de Moran global representa una estadística resumen que indica la intensidad de autocorrelación entre grupos de unidades territoriales, pero no identifica el patrón de estas relaciones espaciales. No obstante, esta estadística es de gran utilidad como primer paso para identificar la presencia o no de una relación espacial entre los indicadores de análisis de unidades territoriales, permitiendo rechazar o afirmar la hipótesis de la no existencia de autocorrelación espacial.

La otra forma de presentar el I de Moran es el índice local, que representa el grado de correlación del indicador de una unidad territorial con los indicadores de sus vecinas.

El I de Moran local se calcula de la siguiente forma, para cada caso  $i$ :

$$(1) \quad I_i = z_i \sum_j w_{ij} z_j$$

donde:

$z_i, z_j$  corresponden a los desvíos del promedio de las observaciones y  $w_{ij}$  es la matriz de vecindad, que se discutirá más adelante con mayor detalle.

Como resultado, el I de Moran local identifica unidades territoriales donde valores de análisis altos o bajos se agrupan espacialmente,

<sup>1</sup> Este software se puede descargar gratuitamente desde <<http://geodacenter.asu.edu>>.

así como también unidades territoriales con valores muy distintos a los de las áreas circundantes. Por tanto, este índice puede reconocer cinco tipos de conglomerados espaciales, que son los siguientes:

i) alto-alto (*high-high*): una unidad territorial con un valor de análisis por encima del promedio, rodeada significativamente por áreas vecinas que también se encuentran por sobre la media con respecto a la variable de interés. Estas unidades territoriales corresponden a los denominados enclaves calientes (*hot spots*), identificados por el índice  $G^*$  de Getis-Ord;

ii) bajo-bajo (*low-low*): una unidad territorial con un valor de análisis inferior al promedio, rodeada por áreas vecinas que también se encuentran bajo la media en relación con la variable de interés. Estas unidades territoriales corresponden a los denominados enclaves fríos (*cold spots*), identificados por el índice  $G^*$  de Getis-Ord;

iii) bajo-alto (*low-high*): presencia de una unidad territorial con un valor de análisis bajo, rodeada significativamente por áreas vecinas con valores que se encuentran por sobre la media de la variable de interés;

iv) alto-bajo (*high-low*): presencia de una unidad territorial con un valor de análisis alto, rodeada significativamente por áreas vecinas con valores que se encuentran bajo la media de la variable de interés; y

v) relación no significativa (*not significant*): presencia de unidades territoriales donde el valor de análisis de la variable de interés no se relaciona significativamente con los valores que presentan sus vecinos.

Cuando se consideran indicadores de pobreza donde un alto valor indica una mayor incidencia del fenómeno en una determinada unidad territorial, es fácil pensar en casos donde existan conglomerados de unidades territoriales contiguas pobres (*hot spots*) o prósperas (*cold spots*), mientras los casos de conglomerados espaciales combinados bajo-alto y alto-bajo son menos intuitivos. Ejemplos de estos últimos tipos de patrones espaciales podrían darse donde existen condominios cerrados de altos ingresos, construidos en medio de lugares predominantemente pobres o vulnerables –para el caso de un conglomerado espacial alto-bajo–, o donde existen favelas o barriadas pobres que se asientan en medio de barrios de altos ingresos –para el caso de un conglomerado espacial bajo-alto. Asimismo, para los fines de delimitar las áreas prioritarias para la pobreza infantil, lo que produce más interés son los enclaves calientes, o sea, grupos de unidades territoriales contiguas con valores de pobreza infantil significativamente altos.

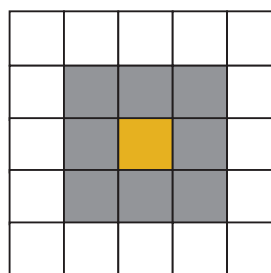
### Consideraciones metodológicas finales en el análisis espacial de la pobreza infantil

Regresando a la primera ley de Tobler, donde “todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes”, es lícito preguntarse qué se considera como cercano. De hecho, la definición de vecindad que se utilice con los polígonos que conforman las unidades territoriales es muy importante, ya que tendrá un alto peso en el cálculo del I de Moran. Este cálculo requiere crear un ponderador espacial, que se origina en la forma de una matriz de contigüidad, que toma valores de uno para unidades territoriales adyacentes y de cero para los casos contrarios (estas matrices pueden estandarizarse por fila para facilitar su interpretación). Para nuestros fines, la vecindad entre unidades territoriales está definida en primer orden de contigüidad por un criterio denominado “dama”, que significa que si cualquier parte de los bordes o vértices que conforman las unidades territoriales se tocan, estas son consideradas como vecinas. Se descartó uno de los criterios tradicionalmente utilizado, pero que a su vez es más rígido, conocido como “torre”, que exige que todo un borde sea común entre ambas unidades para ser consideradas vecinas (véase el gráfico 1).

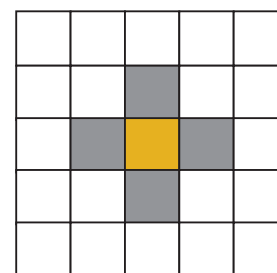
Existen otras dos consideraciones metodológicas que pueden afectar el cálculo de los indicadores de análisis espacial a nivel de unidades territoriales (y no por observaciones individuales): la unidad de análisis y la escala geográfica. En efecto, se trata de la distorsión potencial que existe al imponer una definición de los límites territoriales sobre fenómenos espaciales continuos, ya que existe la posibilidad de que la definición de la unidad territorial y la escala con la que se mide la agrupación espacial introduzcan una parcialidad estadística. Por una parte, existe el problema de las unidades espaciales modificables

### Gráfico A3-1 DEFINICIÓN DE VECINDAD EN UNIDADES TERRITORIALES HIPOTÉTICAS POR CRITERIO DE DAMA Y DE TORRE

A. Vecindad por criterio de dama, 1er orden de contigüidad



B. Vecindad por criterio de torre, 1er orden de contigüidad



Fuente: Elaboración propia.

(*Modifiable Areal Unit Problem*, MAUP), que se encuentra con frecuencia en el análisis espacial a causa de la arbitrariedad con que sus límites políticos u operativos –por ejemplo, sectores censales– se han definido. Por lo general, las unidades operativas espaciales utilizadas en el manejo de los datos georreferenciados están definidas en función de facilitar metodológica y operacionalmente la captura de datos en terreno (por ejemplo, en el caso del operativo censal se asigna un área de enumeración por cada empadronador), o corresponden a una división político-administrativa impuesta sobre el territorio y, por tanto, no reflejan los espacios de vida de las personas, como tampoco representan unidades espaciales con poblaciones homogéneas (por ejemplo, límites territoriales políticos que dividen pueblos originarios o poblaciones indígenas). Es así como existe el riesgo de que el fenómeno poblacional estudiado no se ajuste a las unidades geográficas para las que contamos con la información. Otra faceta de las unidades espaciales modificables tiene relación con la escala, donde el nivel de agregación de los datos que se utilizan afecta en gran medida los resultados de los estudios de autocorrelación espacial. Por ejemplo, el nivel de autocorrelación espacial varía si se calcula a nivel de división administrativa mayor (DAM), o a nivel de estado, región, departamento, provincia o municipio, por tratarse de un cambio en la escala del análisis.

Estando conscientes de estas precisiones estadísticas, para la presente investigación es razonable trabajar con los datos censales agregados a nivel de Divisiones Administrativas Menores (DAME). Como el fin de esta indagación es proveer información a nivel local, la DAME, que en la mayoría de los países corresponde al nivel municipal, es una unidad político-administrativa que se ajusta como beneficiaria de políticas y programas sociales. Esta decisión también responde a una cuestión práctica, pues son pocos los países donde existe y se ha divulgado la cartografía digital de los censos nacionales a un nivel menor de DAME.

Cabe terminar esta sección con algunas reflexiones sobre el uso del I de Moran, particularmente en este estudio. Aunque en el cálculo del índice se toman como observaciones los desvíos del promedio de cada una de las unidades territoriales, para un análisis a nivel regional (América Latina y el Caribe), previamente, se procede a homologar las distribuciones de los indicadores calculados por municipio para cada país, lo que se hace para poder comparar porcentajes de pobreza infantil que vienen de distintos contextos y realidades. Esta homologación es imprescindible, dado que el objetivo de la investigación no es cuantificar las diferencias nacionales de un país a otro, sino describir y estudiar la distribución espacial de las diferencias subnacionales a lo largo de la región. Sin esta estandarización, los I de Moran identificarían enclaves calientes de pobreza infantil que abarcarían la mayor parte del territorio en los países pobres, o puntos fríos de pobreza infantil que abarcarían la mayor parte del territorio en los países prósperos, sin identificar las

disparidades en la incidencia de la pobreza infantil a nivel subnacional, que seguramente existen.

La mayoría de las veces, el proceso de estandarización no hace sino buscar una escala común, donde la distribución transformada se asemeja a la forma de una distribución conocida –que en el caso de la estandarización por el método *z-score* corresponde a una distribución normal. Es decir, la estandarización transforma todas las distribuciones normales en una sola y esta sigue siendo normal. Sin embargo, y en particular para este estudio, se requiere homologar el conjunto de las distribuciones regionales entre sí, para que sean similares entre ellas, independiente de si se asemejan o no a una distribución normal o a otra conocida. Por ello no se habla de estandarización, sino de homologación.

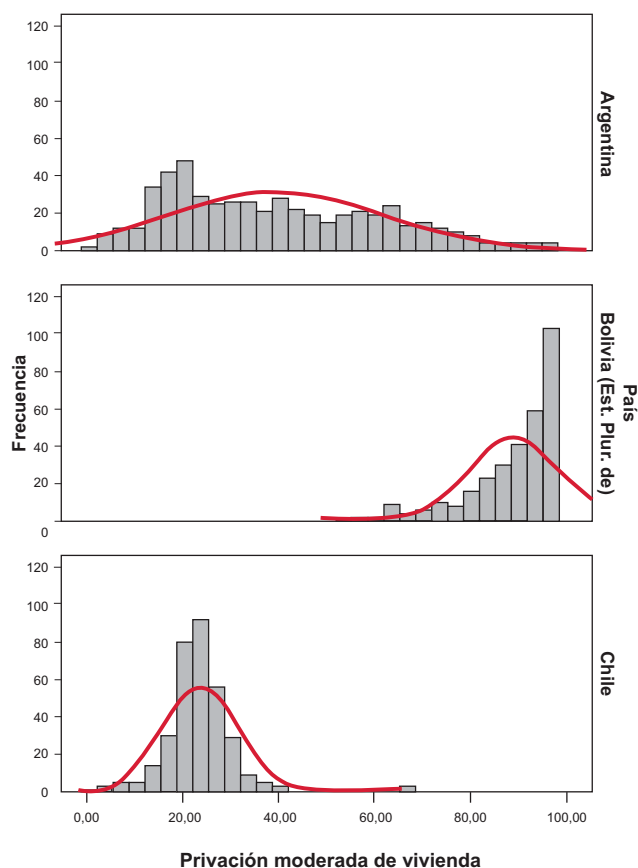
Al hacer esta homologación, se observa la existencia de una gran heterogeneidad en los rangos de las distribuciones de valores en los indicadores de privación de cada país, como se ilustra en el gráfico 2. Esta heterogeneidad se presenta sobre todo en los extremos de las distribuciones, por lo que se hace necesario eliminar los valores atípicos para poder hacer la comparación a nivel regional. Esto se puede conseguir utilizando el método de winsorización, consistente en estilizar los extremos, evitando tener en ellos valores que sean demasiado dominantes y que puedan causar problemas en el cálculo del I de Moran. El proceso de homologación considera entonces dos pasos:

i) centrar y escalar (estandarización): al centrar se desplazan las distribuciones para que tengan un valor medio común y al escalar se ajustan para que compartan una varianza común. La estandarización se puede llevar a cabo con cualquiera de los métodos conocidos, como por ejemplo el *z-score* o el de máximos y mínimos. En este estudio se ha optado por el segundo, que se detalla más adelante; y

ii) eliminar los valores atípicos (winsorización): para cada variable utilizada en el estudio se empleó la estrategia de winsorizar las distribuciones, lo que significa truncarlas en un punto específico, correspondiente a un par de percentiles determinado, que en conjunto suman un porcentaje de winsorización. Por ejemplo, si dicho porcentaje es del 10%, implica que todas las observaciones con un valor inferior al 5o percentil se desplazan hacia arriba del valor de este y todas las observaciones con un valor por sobre el 95o percentil se desplazan hacia abajo de dicho valor.

La estandarización y homologación se pueden hacer en un solo paso, para poder aplicar una metodología de homologación que considere los máximos y mínimos sustituidos por el valor de los percentiles respectivos (95o y 5o), de acuerdo al porcentaje especificado de los datos, a fin de evitar que los valores extremos distorsionen los resultados.

**Gráfico A3-2**  
**ARGENTINA, ESTADO PLURINACIONAL**  
**DE BOLIVIA Y CHILE:**  
**DISTRIBUCIÓN DE LA POBLACIÓN MENOR**  
**DE 18 AÑOS EN SITUACIÓN DE POBREZA**  
**INFANTIL CON PRIVACION EN VIVIENDA**  
**A NIVEL SUBNACIONAL, ALREDEDOR DE 2000**  
*(En porcentajes)*



Fuente: Centro Latinoamericano y Caribeño de Demografía (CELADE) – División de Población de la CEPAL sobre la base de los censos de población y vivienda de la ronda 2000 procesados con el programa REDATAM.

La homologación se calcula como:

$$(2) \quad Zs_i = \frac{X_i - Vp_5}{Vp_{95} - Vp_5}$$

donde:

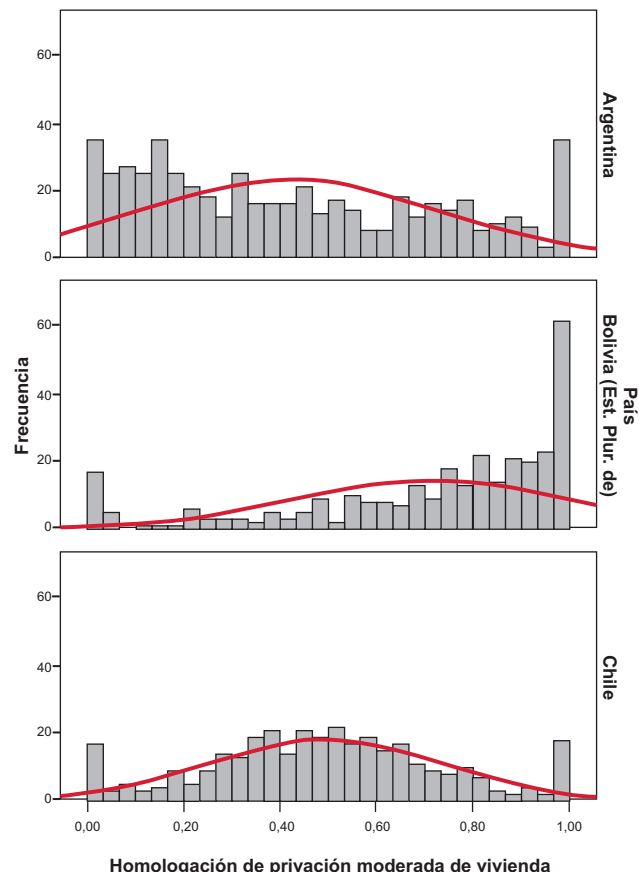
$Zs_i$  = valor de la variable estandarizada

$X_i$  = valor de la variable

$Vp_5$  = valor alcanzado del percentil (5°)

$Vp_{95}$  = valor alcanzado del percentil (95°)

**Gráfico A3-3**  
**ARGENTINA, ESTADO PLURINACIONAL**  
**DE BOLIVIA Y CHILE:**  
**HOMOLOGACIÓN DE LA DISTRIBUCIÓN DE LA**  
**POBLACIÓN MENOR DE 18 AÑOS EN SITUACIÓN**  
**DE POBREZA INFANTIL CON PRIVACION EN**  
**VIVIENDA A NIVEL SUBNACIONAL, ALREDEDOR**  
**DE 2000**  
*(En porcentajes)*



Fuente: Centro Latinoamericano y Caribeño de Demografía (CELADE) – División de Población de la CEPAL sobre la base de los censos de población y vivienda de la ronda 2000 procesados con el programa REDATAM.

Si el valor resultante es menor que 0, se recodifica a 0, y si es mayor que 1, se recodifica a 1, logrando así un valor en el rango de 0 a 1. Los indicadores utilizados para calcular el I de Moran a nivel regional fueron homologados y transformados siguiendo esta metodología y utilizando el programa estadístico SPSS.

Este procedimiento permite el cálculo de un I de Moran que destaca las disparidades subnacionales en la incidencia de la pobreza infantil y permite, a la vez, la identificación de enclaves calientes de este fenómeno distribuidos en toda la región, principalmente conglomerados espaciales de unidades territoriales que traspasan las fronteras de un país.