



Detección y caracterización de zonas marginales en la ciudad de Medellín mediante el análisis exploratorio de datos espaciales

Monografía

Autor

Miguel Noreña

Código 200317001003

Asesor

Juan Carlos Duque

Universidad EAFIT

Escuela de Administración

Medellín

2009

Contenido

1	Introducción	1
2	Planteamiento y Justificación del Problema	2
3	Objetivos	4
3.1	General	4
3.2	Específicos	4
4	Revisión de Literatura	4
4.1	Definición de Slum	4
4.2	Literatura sobre el AEDE	6
4.3	El AEDE en Estudios de Pobreza	7
4.4	Construcción de Regiones Analíticas	9
5	Metodología	11
5.1	Suavización de Tasa Espacial	12
5.2	Modelo Factorial de Componentes Principales	13
5.3	Max-P-Region	14
5.4	Índice Slum	16
5.5	Clusters	16
6	Resultados	18
6.1	Suavización de Tasa Espacial	18
6.2	Modelo Factorial de Componentes Principales	24
6.3	Max-P-Region	27

6.4	Índice Slum	29
6.5	Variables Slum	31
6.6	Clusters	36
7	Conclusiones	37
8	Anexos	42

Lista de Tablas

1	Preguntas de carácter socioeconómico de la ECV Medellín 2007	19
2	Valores de la <i>I de Moran</i> global para las tasas suavizadas	24
3	Varianza de las variables explicada por los componentes	25
4	Cargas de los componentes rotados ortogonalmente	26
5	Caracterización de los hogares de Medellín por medio de las variables <i>slum</i>	30
6	Situación de las regiones analíticas según variables <i>slum</i>	31
7	Clasificación de las regiones analíticas más afectadas por las variables <i>slum</i>	34

Lista de Figuras

1	Flujo metodológico	11
2	Distribución espacial de las variables de la ECV Medellín 2007	21
3	Distribución espacial de las variables de la ECV Medellín 2007	22
4	Distribución espacial de las variables de la ECV Medellín 2007	23
5	Comportamiento espacial de las puntuaciones de los componentes	27
6	Puntuaciones de los componentes según regiones analíticas	28
7	Comparación entre número de hogares por barrio y número de hogares por región analítica	29
8	Comportamiento de las regiones analíticas en el diagrama de dispersión de Moran	30
9	Distribución espacial de los componentes del índice <i>slum</i> a nivel de regiones analíticas	33
10	Detección de <i>clusters</i> espaciales mediante el estadístico local <i>I de Moran</i>	38
11	Flujo de información a nivel de personas, hogares y barrios	42
12	Flujo de información de las variables socioeconómicas de la ECV Medellín 2007	43
13	Flujo de información de las variables socioeconómicas de la ECV Medellín 2007	44
14	Flujo de información de las regiones analíticas	44
15	Flujo de información georreferenciada	45

1 Introducción

Las zonas marginales, mejor conocidas en la literatura internacional como *slums*, son el nombre bajo el cual se denomina a aquellas porciones geográficas intra-urbanas que presentan un alto nivel de pobreza y condiciones de vida muy por debajo del promedio poblacional. Durante los últimos cinco años, el estudio de las zonas marginales ha llamado la atención de investigadores de diferentes áreas del conocimiento (psicología, epidemiología, sociología, salud pública, ambiental y ocupacional, demografía, geografía, entre otros). De hecho, el número de publicaciones en revistas indexadas en la *ISI Web of Knowledge* que contienen la palabra *slum* entre sus palabras clave, se ha multiplicado por 40 entre 2005 y 2008.

El estudio de las zonas marginales adquiere especial relevancia en América Latina si se tiene en cuenta que en dicha región, según las Naciones Unidas, un 32% de la población urbana vive en zonas marginales. Dicho porcentaje es aún más importante si se considera que un 76% de la población vive en zonas urbanas.¹

En un contexto más cercano, según declaraciones del profesor Jorge Giraldo Ramírez durante el foro Pobreza y Responsabilidad Pública en el Valle de Aburrá (Medellín, 20 de septiembre de 2005),² el Área Metropolitana de Medellín ostenta el mayor porcentaje de pobres, comparada con otras áreas metropolitanas de Colombia. En el mismo foro, la Alcaldía de Medellín reportó que más del 50% de la población de Medellín se encuentra en los niveles más bajos del índice de desarrollo humano. Buena parte de esta problemática está relacionada con el rápido crecimiento demográfico que experimentó la ciudad de Medellín a partir de los años 50, el cual desbordó la capacidad de las autoridades públicas para asegurar una completa cobertura de servicios públicos en las nuevas áreas pobladas, mayoritariamente ubicadas en las laderas de la ciudad.

Desde el punto de vista de formulación de políticas públicas, la correcta identificación y delimitación de las zonas marginales en una ciudad es un punto clave para asegurar una correcta y eficiente asignación de recursos destinados a mejorar las condiciones de vida de las poblaciones menos favorecidas.

El presente trabajo de grado busca aportar nuevas evidencias en el estudio de las zonas marginales de la ciudad de Medellín por medio de la aplicación de los últimos desarrollos metodológicos dentro de las técnicas del análisis exploratorio de datos espaciales (AEDE). Así, a partir de los datos disponibles en la Encuesta de Calidad de Vida Medellín 2007 (ECV Medellín 2007), se buscará identificar y caracterizar las zonas de la ciudad que califican como zonas marginales. Estas metodologías permitirán crear un punto de corte en el continuo del nivel de pobreza que permita, de una forma rigurosa,

¹En regiones como África subsahariana, el porcentaje de población que vive en zonas marginales es de 71,9%, pero sólo un 35% de su población vive en zonas urbanas.

²<http://www.veeduriamedellin.org.co/biblioteca2.shtml?x=269>

clasificar zonas de la ciudad como marginales y no marginales.

Los resultados de esta investigación permitirán profundizar en el conocimiento sobre las desigualdades intra-urbanas en Medellín y, en el mejor de los casos, aportarán nuevas evidencias para que los hacedores de política cuenten con más elementos de juicio que les permitan formular nuevas y mejores políticas para la reducción de la pobreza en Medellín y para la mejor asignación de los recursos.

El resto del documento está estructurado como sigue: la sección 2 presenta el planteamiento y la justificación del problema; los objetivos se encuentran en la sección 3; en la sección 4 se hace una revisión de la literatura sobre la identificación de zonas marginales y sobre las herramientas para el AEDE; en la sección 5 se describen las técnicas metodológicas utilizadas a lo largo de la investigación; la sección 6 contiene los resultados arrojados al implementar dichas técnicas metodológicas; en la sección 7 se hacen unos comentarios finales que resumen los hallazgos de esta investigación; y, por último, en la sección 8 se incluye información complementaria de suma utilidad en el desarrollo de la monografía.

2 Planteamiento y Justificación del Problema

De acuerdo con el demógrafo John Weeks,³ en ciudades en las que las tasas de nacimiento superan el nivel de reemplazo y en las que además existe un alto nivel de migración que generan un crecimiento demográfico descontrolado, es necesario reorientar la atención en el estudio de las diferencias entre zonas urbanas y rurales hacia el estudio de las desigualdades intra-urbanas (Weeks et al., 2006).

En este sentido, aunque los datos demográficos de la ciudad de Medellín muestran que la tasa de reemplazo está por encima de la de nacimiento,⁴ el nivel de migración ha generado una explosión demográfica similar a la descrita por Weeks. Esto hace que el estudio de las desigualdades en dicha ciudad adquiera una especial relevancia.

Dichas desigualdades ya se hacen evidentes en Medellín. En la ciudad se encuentran barrios donde hay elevados niveles de vida y que cuentan con excelentes prestaciones de servicios públicos, infraestructuras, desarrollo urbanístico y comercial. Estos barrios contrastan con zonas de la ciudad, generalmente ubicadas en las laderas, cuyos habitantes viven en una precaria situación socioeconómica. Como resultado, la ciudad presenta una polarización marcada que se convierte en una amenaza para su progreso ya que la marginación de un porcentaje significativo de la población obstaculiza el

³John Weeks es el autor de uno de los textos más relevantes en temas de demografía: *“Population: An Introduction to Concepts and Issues”* (Weeks, 2007).

⁴De acuerdo con Rosero-Bixby (2004), en las ciudades de México D.F., Monterrey, Río de Janeiro y Medellín, la tasa de nacimiento es y será inferior a la de reemplazo.

desarrollo.

Un valor agregado de esta investigación consiste no sólo en la aproximación metodológica, sino también en la escala geográfica en la cual se lleva a cabo. De hecho, hasta lo que se conoce, esta será la primera investigación en temas de identificación de zonas marginales que trabaje con la definición de regiones analíticas. Las regiones analíticas dividen la ciudad en áreas geográficas que no coinciden con las divisiones administrativas (arbitrarias o normativas) y que están directamente relacionadas con el fenómeno objeto de estudio.⁵ La utilización de dichas regiones permitirá una mejor delimitación geográfica de las zonas que sean clasificadas como zonas marginales, lo cual proporcionará mayores elementos para una asignación más eficiente y efectiva de los recursos públicos y para el diseño de políticas sociales más focalizadas tendientes a sacar a la población más vulnerable de su condición de pobreza.⁶

En el contexto espacial también destaca el hecho de que considerar las interacciones entre áreas geográficas próximas⁷ resulta un importante aporte en la literatura existente sobre la pobreza en Medellín, ya que al contemplar dichas interacciones, se estarán teniendo en cuenta efectos de propagación entre las diferentes áreas de la ciudad; es decir, la pobreza sufrida en un área puede trasladarse a otras áreas como resultado de las interrelaciones existentes. Determinar esto es importante porque puede dar un giro al modo como se diseñan los planes de ayuda dirigidos a la población más vulnerable, teniendo en cuenta que los esfuerzos realizados en un determinado barrio repercutirán en la condición social de los demás.

Por último, cabe resaltar que el tema de la pobreza merece estar dentro de los principales temas de investigación de los economistas en Colombia, debido a su incidencia en la población y a los modestos avances que se han logrado en este aspecto a pesar del crecimiento económico de los últimos años. Es una problemática que se agudizó a finales de la década de los noventa con la crisis económica atravesada por el país y que, no obstante la expansión del producto experimentada desde 2004, sigue representando grandes retos debido a los preocupantes indicadores sociales aún presentes.

⁵Para una definición más detallada sobre las diferencias entre regiones analíticas y normativas ver Duque et al. (2007).

⁶Cabe también señalar que el uso de regiones analíticas es una alternativa metodológica para minimizar los inconvenientes derivados del problema de la unidad espacial modificable. Para más detalles sobre este problema ver Openshaw and Taylor (1981).

⁷Dichas interacciones son un *input* importante dentro de las técnicas para el AEDE.

3 Objetivos

3.1 General

1. Identificar y caracterizar las zonas marginales de la ciudad de Medellín, mediante la utilización de los más recientes desarrollos para el AEDE, para aportar nueva evidencia que ayude a reducir los problemas de pobreza en la ciudad.

3.2 Específicos

1. Establecer un nuevo sistema de áreas geográficas (regiones analíticas) especialmente diseñadas para el estudio de temas de pobreza en la ciudad.
2. Aplicar técnicas para el AEDE para, de una forma estadística, poder identificar porciones geográficas de la ciudad que califiquen como zonas marginales.
3. Caracterizar la situación socioeconómica de las zonas marginales con el fin de identificar las necesidades más apremiantes de la población de bajos ingresos que deben ser atendidas por las autoridades gubernamentales.

4 Revisión de Literatura

Tomando en cuenta el enfoque de la presente investigación, es clave empezar por una revisión de la literatura sobre el concepto de *slum*. Para abordar el AEDE, se repasarán investigaciones que hayan implementado esta metodología, enfatizando en el análisis derivado de los estudios espaciales y en el alcance de sus conclusiones. Por último, es importante hacer un repaso de los trabajos realizados sobre la distribución espacial de la pobreza que permitan visualizar los avances logrados en la identificación de las regiones más pobres.

4.1 Definición de Slum

En la literatura no se encuentra una única definición para el término *slum* puesto que su utilización varía mucho entre países y suele agrupar una amplia gama de características. Tradicionalmente, este término se utilizaba para referirse a aquellas zonas en las ciudades y pueblos que alguna vez habían sido habitadas por individuos pudientes quienes se mudaron a otros lugares dejando sus viviendas en manos de grupos de personas de menores ingresos. Como resultado, estas zonas sufrieron deterioro y quedaron

marginadas frente a los nuevos centros de desarrollo (United Nations Human Settlements Programme, 2003). La palabra *slum* se empleó para marcar una división entre las diferentes áreas de una ciudad, especialmente para referirse a lugares con condiciones sanitarias deplorables y viviendas precarias donde se unían el vicio y la delincuencia. Sin embargo, esta connotación comenzó a prevalecer en los países desarrollados, ya que en los países de ingresos bajos la palabra *slum* no tenía un significado peyorativo.

En la literatura castellana, el término *slum* tiene diferentes traducciones pues no existe un consenso en la forma como debe llamarse a las zonas caracterizadas por la presencia de invasores, personas de bajos ingresos, viviendas de materiales precarios y epidemias. Algunos autores han empleado el término “gueto” aunque este presenta un inconveniente al ser más específico si se tiene en cuenta que un gueto es un sitio habitado por individuos de una misma raza. Igualmente puede encontrarse la palabra “tugurio” para hacer alusión a las zonas marginales, pero en ocasiones se ha tratado de evitar su utilización ya que acarrea una connotación negativa en países donde las diferencias sociales son generadoras de conflictos internos. En general, pueden encontrarse múltiples equivalencias para *slum* en español, las cuales varían entre países haciendo que sea complejo encontrar una sola definición.

Por otro lado, también existe una falta de consenso con respecto a la métrica que debe utilizarse para definir si un área geográfica califica, o no, como *slum*. Generalmente, los censos poblacionales y las encuestas realizadas por las autoridades gubernamentales no incluyen cuantificaciones de personas viviendo en áreas con condiciones inapropiadas de habitabilidad. Según United Nations Human Settlements Programme (2003), la dificultad en la medición radica en lo siguiente:

1. Los *slums* son muy complejos para intentar definirlos en un solo parámetro.
2. Son un concepto relativo debido a que lo que se considera como *slum* en una ciudad puede no serlo en otra ciudad, incluso tratándose de un mismo país.
3. Sus variaciones son muy amplias como para tratar de definir criterios universales.
4. Están en constante cambio lo cual impide que un criterio sea válido por un período prolongado de tiempo.
5. La naturaleza espacial de los *slums* hace que su tamaño esté sujeto a cambios jurisdiccionales o en la agregación espacial.

En la presente investigación se utilizará la palabra “zona marginal” para hacer referencia a un *slum* y, siguiendo los lineamientos de la ONU-HABITAT, se considerará una zona como marginal cuando se presente una o más de las siguientes situaciones: falta de acceso a agua potable, falta de conexión a alcantarillado, espacio inadecuado (definido

como más de dos personas por habitación), material precario de las viviendas y no propiedad de las viviendas. Debe resaltarse que esta definición no abarca la totalidad de las zonas pobres dentro de un área de estudio. Esto se ha evidenciado en estudios previos que han identificado en las ciudades asentamientos con condiciones de pobreza que no califican dentro de la definición de la ONU-HABITAT (Weeks et al., 2006).

4.2 Literatura sobre el AEDE

Comenzar con una definición del AEDE es clave para entender en qué consiste esta metodología. De acuerdo con Vayá y Moreno,

“...el análisis exploratorio de datos espaciales (AEDE) se centra de forma explícita en los efectos espaciales y consiste en el conjunto de técnicas que permiten describir distribuciones espaciales, identificar localizaciones atípicas (*outliers* espaciales), descubrir esquemas de asociación espacial (*cluster* espacial) y sugerir diferentes regímenes espaciales u otras formas de inestabilidad espacial. El centro de este concepto lo ocupa la noción de autocorrelación espacial, es decir, el fenómeno por el cual la similitud locacional (observaciones con proximidad espacial) se une con la similitud de valores (correlación de atributos)”.

(Vayá and Moreno, 2001, pp. 29)

El AEDE se convierte en una herramienta de mucha utilidad para el caso de estudios sociales o económicos en los cuales la ubicación geográfica de los agentes es un factor importante a tener en cuenta. La estadística descriptiva deja a un lado la localización geográfica de los datos y, al hacer esto, la especificación de los modelos econométricos planteados puede ser incorrecta (Acevedo and Velásquez, 2008). Por tal razón, el AEDE es una técnica que permite corregir este problema. Entre sus ventajas se encuentran: detección de patrones espaciales en los datos, identificación geográfica de conjuntos de datos atípicos, y, para los casos en los que se dispone de datos de panel, visualización de tendencias en los patrones espaciales.

El autor más importante en la literatura del AEDE es Luc Anselin quien ha implementado esta metodología en diversos temas que van desde el comercio mundial y la competitividad de los países hasta indicadores sociales, como el número de homicidios en Estados Unidos. Precisamente un estudio de criminalidad es un punto de referencia para los objetivos de esta investigación, ya que tanto la pobreza como los homicidios son variables sociales cuyo análisis debe incorporar el factor geográfico. De acuerdo con un estudio de Anselin et al. (2001), la criminalidad en los condados de Estados Unidos

durante el período 1960-1990 no presentó aleatoriedad espacial sino que presentó autocorrelación espacial, debido a la aparición de *clusters* regionales con altos niveles de criminalidad. Este resultado, al tratarse de una variable social, es un buen indicador de que el AEDE es adecuado para realizar un estudio de pobreza en Medellín en el cual se determine la existencia de relaciones espaciales entre zonas marginales de la ciudad. No obstante, el AEDE presenta algunos inconvenientes puesto que una falla en el nivel de desagregación de las unidades espaciales puede conducir a conclusiones espurias. Es difícil hacer frente a este problema porque, en ocasiones, la información no está disponible para las unidades que la teoría indica deben ser analizadas (Anselin et al., 2001).

Al igual que en el estudio de pobreza se pasa de un nivel regional a un nivel intra-urbano, las técnicas del AEDE pasan de la utilización de indicadores globales a indicadores locales. Los primeros analizan las regiones de una muestra de forma conjunta y son incapaces de detectar aquellas áreas donde se concentran valores atípicos. Son medidas que no capturan la inestabilidad en la distribución espacial de las variables, ignorando la no estacionariedad en el espacio de algunos procesos (Getis and Ord, 1992). Sin embargo, este problema fue resuelto con la disponibilidad de datos para unidades con alto grado de desagregación y con los avances en los sistemas de computación, lo cual condujo a que desde principios de la década de los noventa el interés se centre en los indicadores locales capaces de identificar patrones espaciales en áreas de menor tamaño (Fotheringham et al., 2000).

4.3 El AEDE en Estudios de Pobreza

Con respecto a la implementación del AEDE para identificar zonas marginales, una investigación realizada por Weeks et al. (2007) se basa en la identificación de estas zonas en Accra (Ghana) a través del cálculo del índice *slum* compuesto por cinco variables: conexión a acueducto, conexión a alcantarillado, hacinamiento, materiales de construcción de las viviendas y propiedad de las viviendas. Apoyados en información de un censo poblacional e implementando técnicas de agregación espacial, los investigadores concluyen que es posible identificar la mayor parte de los barrios más pobres con el uso de imágenes satelitales, ya que estos lugares suelen estar caracterizados por poca vegetación y por la prevalencia de material impermeable. No obstante, las características de estos barrios pueden variar en una misma área metropolitana, por lo cual hay que ser muy cauteloso antes de hacer generalizaciones (Weeks et al., 2007).

Con relación a la utilización de las técnicas del AEDE, mediante el estadístico *I de Moran*, en el estudio se identifican patrones espaciales en las zonas marginales de Accra, lo cual lleva a concluir que, a pesar de que las zonas marginales estén dispersas por todo el territorio de la ciudad, existen agrupaciones con valores más críticos del índice *slum*.

También se recurre a la construcción de regiones analíticas debido a la disponibilidad de información sólo para pequeñas unidades geográficas. Basados en la metodología de agregación de Duque (2004), los investigadores construyen unas regiones analíticas representativas de las características de los barrios de Accra. Obtienen resultados no significativos de la *I de Moran*, lo cual indica que capturaron los patrones espaciales en las regiones analíticas y que dieron una solución al problema de la unidad espacial modificable.

En Colombia, algunos estudios han subrayado la importancia de incluir factores geográficos en el análisis de la pobreza debido a que la distribución de la misma ha mostrado patrones de autocorrelación espacial. Pérez (2005) realizó una investigación de la dimensión espacial de la pobreza en este país utilizando el índice de necesidades básicas insatisfechas (NBI) y el índice de calidad de vida (ICV). Su análisis departamental lo llevó a concluir que en los años 1985 y 1993 la pobreza en el país presentó dependencia espacial reflejada en la existencia de *clusters* espaciales de departamentos pobres. Por su parte, las disparidades en los índices de NBI de municipios de un mismo departamento reflejaron las relaciones espaciales existentes con municipios de departamentos vecinos. Un inconveniente de este estudio fue el hecho de haber desarrollado el análisis a una escala geográfica muy agregada como son los departamentos. Es claro que en Colombia existe una heterogeneidad entre las zonas rurales y urbanas de los departamentos registrada en los mayores niveles de pobreza de las primeras. Incluso a nivel municipal las diferencias intra-urbanas son marcadas, lo cual impide una correcta apreciación de los patrones espaciales y puede conducir a interpretaciones equivocadas.

En Chile también se han utilizado herramientas espaciales para realizar un diagnóstico de la pobreza en el país. No obstante los altos niveles de pobreza que se presentan en América Latina, Chile es un ejemplo exitoso de mejoramiento de la calidad de vida de sus habitantes, ya que en menos de dos décadas logró reducir la pobreza de 38,6% en 1990 a 13,7% en 2006. Agostini et al. (2008) examinan detalladamente los avances sociales logrados en el país sudamericano con la implementación de mapas de pobreza que permiten visualizar una distribución espacial a nivel rural y urbano. Una de las novedades de este trabajo es la utilización de datos censales para complementar los datos disponibles en las encuestas de ingresos. Esta combinación permitió contar con un mayor volumen de información para el estudio dado que la sola utilización de las encuestas de ingresos no hubiera permitido la realización del estudio al nivel de desagregación espacial deseado. Con esta metodología los investigadores llegan a la conclusión de que la pobreza en Chile presenta disparidades a nivel rural y urbano y destacan la utilidad de los mapas en el diagnóstico de esta problemática y en el diseño de políticas.

En esta misma línea de utilización de mapas de pobreza se encuentra el trabajo de Petrucci et al. (2003) en Ecuador. Sirviéndose de modelos de regresión espacial y de mapas, los investigadores descubren que la omisión de indicadores geográficos adecuados

conduce a una asignación ineficiente de los recursos destinados a ayudar a la población de bajos ingresos. Esto se da porque muchas veces se utiliza una unidad espacial con alto grado de agregación - como los departamentos - para estudiar la problemática de las personas de bajos ingresos, lo cual conduce a resultados que no reflejan la verdadera condición de pobreza. Incluso cuando se está examinando el desarrollo de una ciudad puede encontrarse el problema de unidades muy agregadas puesto que es común encontrar diferencias marcadas a nivel intra-urbano (Haddad and Nedovic-Budic, 2003).

A manera de conclusión, puede afirmarse que al revisar la literatura del AEDE queda claro que uno de sus principales problemas es la definición de las unidades geográficas. Por ejemplo, Pérez (2005), en su estudio de la dimensión espacial de la pobreza en Colombia mencionado previamente, encontró que algunos resultados arrojados al realizar un análisis departamental no concordaban con aquellos obtenidos en su análisis municipal. Las correlaciones espaciales eran mucho más débiles en el primer caso, lo cual da mayor fortaleza a la creación de regiones analíticas que capturen las interrelaciones entre diferentes unidades geográficas. En el presente trabajo se crearon esas regiones utilizando la metodología propuesta por Duque (2007) y que fue aplicada por Weeks et al. (2007) mostrando mejores resultados que los obtenidos al usar unidades espaciales administrativas. Además, como lo sugieren Weeks et al. (2006), en el estudio de problemas sociales es sumamente importante desarrollar análisis a nivel urbano porque cada vez las disparidades son más marcadas en áreas urbanas dadas las altas tasas de migración y de expansión territorial.

4.4 Construcción de Regiones Analíticas

La construcción de regiones analíticas es una alternativa para el estudio espacial al proporcionar áreas de estudio diferentes a las divisiones administrativas sobre las que comúnmente se desarrolla una investigación. Una opción es la agregación de unidades geográficas de tal manera que se cuente con áreas de mayor magnitud donde las diferencias entre las unidades de una misma región se minimicen al tiempo que las diferencias interregionales se maximicen. El proceso es complejo porque generalmente las áreas de estudio presentan similitudes en diferentes aspectos, por lo cual la agregación varía dependiendo de los criterios elegidos por el investigador. Lo ideal sería probar todas las combinaciones posibles y elegir la que mejor se acomode a los objetivos de estudio, pero ante un número elevado de unidades geográficas en cuestión esto se vuelve inviable y se debe recurrir a diferentes herramientas de agregación. Otro punto a considerar es la decisión ante la cual se enfrenta el investigador quien debe determinar si su prioridad es recoger todas las áreas de estudio en un número reducido de regiones, lo cual presenta el inconveniente de sacrificar detalle por simpleza, o si, por el contrario, debe maximizar las regiones con el propósito de no perder detalles esenciales en los análisis.

En ocasiones se confunde el concepto de construcción de regiones (*region-building*) con el de agrupación general de unidades (*general grouping*). El primero hace alusión a la creación de regiones coherentes donde las fronteras desempeñan un papel fundamental, mientras que la agrupación general se encarga de juntar áreas geográficas de acuerdo con una similitud particular que no considera el factor geográfico. Sin embargo, como lo plantean Byfuglien and Nordgard (1973), los resultados de ambos procesos son en ocasiones muy similares debido a que un área suele presentar mayor grado de semejanza con sus vecinos espaciales que con áreas ubicadas a grandes distancias. Lo anterior también encuentra un apoyo teórico en la “Primera Ley de Geografía” de Tobler según la cual todas las cosas están relacionadas entre sí pero las cosas más próximas tienen una relación más estrecha (Tobler, 2004).

Byfuglien and Nordgard (1973) definen cuatro problemas principales que deben ser resueltos para poder llevar a cabo un adecuado proceso de agregación de unidades geográficas:

1. Escogencia de las unidades básicas a partir de las cuales comienza la agregación: Es importante empezar por las menores unidades posibles con el fin de no perder información. Sin embargo, generalmente la selección de estas unidades está sujeta a las divisiones administrativas. También está sujeta al juicio del investigador quien debe decidir cuáles áreas deben ser consideradas para el análisis inicial, tratando de mantener un equilibrio entre la conservación de la mayor cantidad posible de información y la creación de regiones uniformes en tamaño. Es por esto que se recomienda un conocimiento previo de las áreas de estudio para evitar errores en el proceso de agregación.
2. Descripción de las unidades básicas mediante pocas variables relevantes e independientes: Las características de las unidades básicas elegidas deben ser recogidas por variables que sean ortogonales, o sea, que no estén correlacionadas. La ortogonalidad evita la utilización de variables redundantes ya que se puede presentar una situación en la cual más de una variable esté describiendo una determinada característica. Usualmente se utilizan los métodos de análisis de componentes principales y de análisis de factores para capturar las características de las unidades geográficas. En la sección 5.2 se encuentra una explicación del primero de estos métodos.
3. Utilización de una medida que cuantifique las diferencias entre las regiones tomando en cuenta las variables mencionadas en el numeral anterior: Una de las medidas más utilizadas es el coeficiente de correlación aunque generalmente se requieren múltiples variables para que su interpretación sea significativa. Si se ha optado por reducir las variables a unos cuantos componentes o factores su utilización no es recomendable. También se recurre frecuentemente al cuadrado de la distancia euclidiana como medida de la disimilitud entre regiones. Su utilización suele hac-

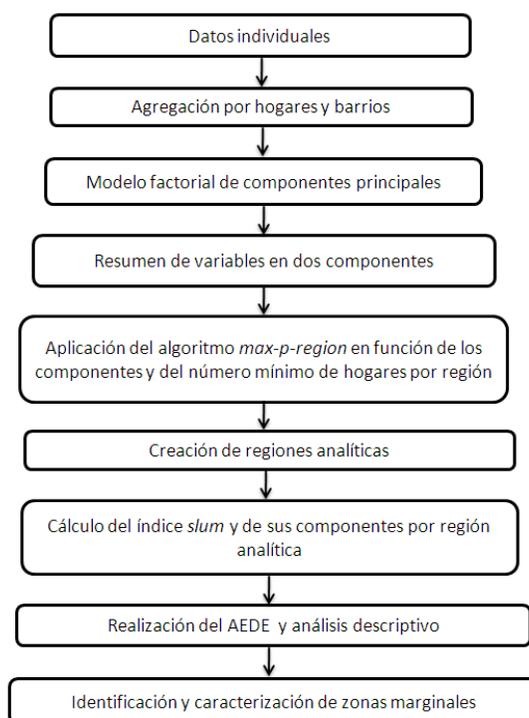
erse en estudios que han implementado el análisis de componentes principales o el análisis de factores.

4. Elección de un método que logre el objetivo de crear regiones homogéneas en su interior pero heterogéneas en comparación con las demás: Existen diversos métodos de agregación pero en esta investigación se implementó el algoritmo *max-p-region* del cual se encuentra una explicación en la sección 5.3.⁸

5 Metodología

A continuación se ilustra y explica el procedimiento que se siguió en el desarrollo de la investigación:

Figura 1: Flujo metodológico



La información a partir de la cual se desarrolló el estudio de pobreza en la ciudad de Medellín fue proporcionada por la ECV Medellín 2007. Esta es una encuesta realizada

⁸Duque et al. (2007) incluye un resumen de las diferentes técnicas de construcción de regiones analíticas. Por su parte, en Byfuglien and Nordgard (1973) se encuentra una comparación extensa de diferentes métodos de agregación de unidades geográficas.

anualmente por la Alcaldía de Medellín en colaboración con otras entidades gubernamentales. Su cobertura abarca la totalidad del área urbana de Medellín además de sus cinco corregimientos: San Cristóbal, Altavista, Santa Elena, Palmitas y San Antonio de Prado. En esta investigación se omitieron los corregimientos por tratarse de un análisis únicamente del área urbana. La ECV Medellín 2007 está compuesta por 184 preguntas divididas en las siguientes nueve secciones: viviendas, hogares, demografía y educación, seguridad social, ingresos y empleo, egresos, instancias de participación, violencia sexual e intra-familiar, y nutrición. La muestra incluye 79.912 personas, agrupadas en 21.861 hogares, quienes respondieron a cada una de las preguntas, bien fuera a nivel personal o en representación del jefe de hogar cuando la información fuera la misma para todos los encuestados.

Con el propósito de comprender la metodología, la figura 1 se puede dividir en dos grandes componentes: definición y creación de regiones analíticas y análisis de pobreza. La definición y creación de regiones analíticas comenzó por la adecuación de los datos de la ECV Medellín 2007 al agregar la información por niveles en el *software Stata*. Primero, los datos individuales se agruparon en hogares mediante la creación de una variable de identificación para cada encuestado de tal manera que la persona se asignara a su hogar correspondiente y haciendo la distinción entre hogares que habitaban una misma vivienda. Luego se procedió a realizar una agregación de la información por barrios separando los hogares de acuerdo con su respectivo barrio de residencia. Antes de la construcción de regiones analíticas fue necesario resumir la información disponible para lo cual se inició la fase de reducción de datos a través del modelo factorial de componentes principales. Una vez la información quedó reducida, se implementó el algoritmo *max-p-region* cuyo resultado fue la creación de las regiones analíticas.

El otro componente de la metodología, el análisis de pobreza, incluyó el cálculo del índice *slum* y de sus elementos para luego proceder al AEDE y a un análisis descriptivo. Las regiones analíticas identificadas como las más vulnerables se analizaron a fondo con el fin de caracterizar su situación de pobreza.

5.1 Suavización de Tasa Espacial

Los métodos de suavización de datos son útiles cuando se cuenta con diferentes tamaños de muestras para las áreas geográficas estudiadas. Uno de estos, la suavización de tasa espacial (*spatial rate smoother*), se caracteriza por arrojar amplios patrones espaciales en los datos y por disminuir en gran medida el número de valores atípicos. Esto último es el resultado de que las tasas que se calculan a través de este método en realidad reflejan un promedio para una región de unidades geográficas, ya que para la unidad geográfica objeto de estudio se utiliza información de las poblaciones de sus vecinos (Anselin et al., 2004). Lo anterior puede visualizarse en la ecuación del método de

suavización de tasa espacial:

$$\hat{\pi}_i = \frac{\sum_{j=1}^N w_{ij} O_j}{\sum_{j=1}^N w_{ij} P_j} \quad (1)$$

donde w_{ij} es la relación entre las unidades geográficas i y j , j son los vecinos de i , O_j es el número de eventos de j y P_j su población en riesgo.

Esta es una prueba no paramétrica, es decir, no se asume una determinada distribución de las observaciones. Además, las ponderaciones espaciales se aplican de forma independiente al numerador y al denominador. También es importante resaltar que la suavización obtenida varía de acuerdo con los criterios que se utilicen en la definición de los vecinos y del peso que se dé a la relación entre las unidades geográficas. Entre mayor sea el número de vecinos, mayor valor tendrá el denominador. Así se atenúa el problema de las unidades geográficas con muestras pequeñas, las cuales pueden presentar porcentajes artificialmente elevados debido a la falta de datos (Carlin and Louis, 2000).

5.2 Modelo Factorial de Componentes Principales

El modelo factorial de componentes principales (*principal components factor model*) es una técnica mediante la cual un grupo de variables se resume en un menor número de componentes que retiene la mayor varianza posible de estas. A través de esta metodología se logra una simplificación de los datos, lo cual permite construir modelos que corrigen algunos de los problemas encontrados en econometría, como la multicolinealidad y las regresiones con múltiples parámetros.

El modelo factorial de componentes principales se deriva del análisis de componentes principales y su diferencia primordial radica en que K variables se expresan en J ($J < K$) y no en K componentes. Matemáticamente, el análisis de componentes principales está representado por:

$$Z_k = l_{k1}F_1 + l_{k2}F_2 + l_{k3}F_3 + \dots + l_{kK}F_K \quad (2)$$

donde Z_k representa las variables observadas, F_K son los componentes (por definición no están correlacionados entre sí) y l_{kK} son las cargas de los componentes, es decir, expresan el peso de la variable Z_k en el componente estandarizado F_K (Hamilton, 1992). Vale la pena mencionar que el número de componentes equivale al número de variables observadas, por lo cual este tipo de análisis no reduce la cantidad de información disponible. Sin embargo, cuando el objetivo es disminuir el número de

variables, se recurre al modelo factorial de componentes principales cuya representación matemática es la siguiente:

$$Z_k = l_{k1}F_1 + l_{k2}F_2 + l_{k3}F_3 + \dots + l_{kJ}F_J + v_k \quad (3)$$

$$v_k = l_{k,J+1}F_{J+1} + l_{k,J+2}F_{J+2} + \dots + l_{kK}F_K \quad (4)$$

donde v_k es un residual que recoge los componentes que no son relevantes en la explicación de la varianza de las variables observadas, ya que los primeros J componentes explican la mayor parte de la varianza de estas. La ausencia de este residual en la ecuación (2) viene explicada por el hecho de que los componentes explican en su totalidad la varianza de las variables observadas.

En la implementación del modelo factorial de componentes principales también se calculan las puntuaciones de los componentes, lo cual se realiza corriendo una regresión donde las variables observadas son las variables explicativas, así:

$$\tilde{F}_j = c_{1j}Z_1 + c_{2j}Z_2 + \dots + c_{kj}Z_k \quad (5)$$

\tilde{F}_j son las estimaciones de los valores desconocidos de F_j y c_{kj} son los coeficientes de las puntuaciones de los componentes (Hamilton, 1992).

5.3 Max-P-Region

Como lo definen Duque et al. (2008), el modelo *max-p-region* busca, a través de un algoritmo, la agrupación de un conjunto de n áreas geográficas en un número máximo de p regiones que cumplan los siguientes requisitos:

1. Las áreas pertenecientes a una misma región deben estar geográficamente conectadas.
2. Se define una restricción que corresponde a una característica particular y cada región debe cumplir con el valor mínimo definido para esta característica.
3. Cada una de las n áreas puede pertenecer solamente a una región.
4. Cada región debe contener por lo menos un área geográfica.

Una de las ventajas de este procedimiento es que permite crear *clusters* dado que las áreas que se asignan a una región en particular poseen características similares. El modelo está compuesto por unas variables de agregación que se caracterizan por marcar las diferencias entre las regiones analíticas, conduciendo a la construcción de regiones con unidades similares en su interior. Sumado a estas variables, el *max-p-region* utiliza otras entradas, pues no basta con establecer similitudes entre unidades geográficas para fijar su asignación a una región. El factor espacial es determinante y por esta razón debe establecerse una estructura de vecindad. Por último, antes de correr el algoritmo, se fija una restricción que debe ser cumplida por cada región analítica. Aunque el objetivo del *max-p-region* es la creación del máximo número de regiones, lo cual se obtendría cuando $n = p$, esto no siempre es posible porque no todas las n áreas geográficas satisfacen el umbral fijado para la restricción.

La solución al problema *max-p-region* consta de dos fases previas a la implementación del algoritmo matemático: la fase de construcción y la fase de búsqueda local (Duque et al., 2008). La primera comienza con la selección aleatoria de una de las áreas geográficas para iniciar el proceso de construcción de una región. Si el área escogida cumple la restricción impuesta, esta se convierte en región pero si, por el contrario, el atributo del área seleccionada se encuentra por debajo del umbral fijado, se lleva a cabo un proceso iterativo el cual se encarga de agregar áreas vecinas no asignadas hasta que se satisfaga la restricción. Este proceso de selección se restringe a aquellas unidades que compartan lado con al menos una de las áreas incluidas en la región en cuestión. Una vez el proceso de construcción de una región culmine, se repiten los pasos mencionados previamente hasta que todas las áreas geográficas hayan sido incluidas o hasta que se presente una situación en la cual las unidades restantes no cumplan los requisitos para formar una nueva región. Si sucede esto último, las áreas no asignadas, a las cuales se les denomina “enclaves”, son agregadas a una de las p regiones creadas.

Las regiones que resultan de esta fase presentan algunos atributos que vale la pena destacar. En primer lugar, no existe restricción en relación a su forma, permitiendo diversidad en la configuración regional. Además, las regiones son tomadas por igual en cuanto a su importancia, a diferencia de otras metodologías que buscan identificar una unidad geográfica central a la cual se adhieren las demás.

La fase de búsqueda local tiene como objetivo minimizar la medida global de heterogeneidad que indica la disimilitud existente entre las áreas asignadas a una misma región. De modo iterativo se ensayan soluciones factibles diferentes a las arrojadas por la fase de construcción, moviendo algunas áreas geográficas de una región a sus vecinas, pero siempre conservando el cumplimiento de las restricciones impuestas en la primera fase.

5.4 Índice Slum

Este índice, que mide la pobreza en un área de estudio determinada, consta de cinco características que de acuerdo con la ONU-HABITAT definen una zona como marginal: falta de acceso a agua potable, falta de conexión a alcantarillado, espacio inadecuado (definido como más de dos personas por habitación), material precario de las paredes de las viviendas y no propiedad de las viviendas. Weeks et al. (2007), en su estudio desarrollado para la ciudad de Accra, reunieron estas variables en un índice *slum* en el cual tomaron la información suministrada por un censo poblacional y repasando cada hogar, asignaron el valor de uno cuando se cumplía el evento de pobreza mientras que registraron el valor de cero en caso contrario. El cómputo del índice fue el siguiente:

$$S_h = \sum_{i=1}^5 slum_i \quad (6)$$

donde S_h era el índice de cada hogar y $slum_i$ cada una de las cinco variables. De acá se desprende que el valor máximo para el índice era cinco (todas las variables se cumplían) y su mínimo era cero (ninguna de las variables se cumplía).

Con miras a obtener resultados a nivel de las regiones analíticas estudiadas por los autores, el índice se calculó así:

$$S_{RA} = \frac{\sum_{h=1}^n S_h}{n} \quad (7)$$

en donde RA era cada una de las regiones analíticas, S_h el índice *slum* por hogar y n representaba el número de hogares por región.

5.5 Clusters

La agrupación en el espacio de variables con valores similares es conocida como *clustering*. Los *clusters* resultan de las no estacionariedades en la distribución espacial de algunos fenómenos y, en palabras de Knox (1989), se definen como “un grupo de ocurrencias limitadas geográficamente cuyo tamaño y concentración hacen que sea poco probable que hayan ocurrido por azar”. Cuando se presenta esta situación se puede contrastar si las variables analizadas tienen una distribución aleatoria en el espacio o si, por el contrario, su comportamiento obedece a la existencia de autocorrelación espacial. Estos contrastes se realizan preferiblemente con la utilización de indicadores locales en vez de indicadores globales, ya que estos últimos analizan las regiones de forma conjunta y son incapaces de detectar aquellas áreas donde se concentran valores atípicos (Vayá and Moreno, 2001).

Un indicador global muy utilizado es la *I de Moran* global:

$$I = N/S_0 \frac{\sum_{ij}^N w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}, i \neq j \quad (8)$$

donde x_i es el valor de la variable analizada en la región i , \bar{x} es la media muestral, w_{ij} es la matriz de pesos espaciales, N es el tamaño muestral y

$$S_0 = \sum_i \sum_j w_{ij} \quad (9)$$

Cuando se cuenta con muestras suficientemente grandes, la I estandarizada sigue una distribución $N(0,1)$. Así, un valor significativo con signo positivo (negativo) del estadístico estandarizado indica la presencia de autocorrelación espacial positiva (negativa).

Como se mencionó, el estadístico de la ecuación (8) no es adecuado para detectar *clusters*, razón por la cual los indicadores locales de asociación espacial, conocidos en la literatura como LISA (*Local Indicators of Spatial Association*), han acaparado la atención de los expertos en temas espaciales al lograr dos objetivos (Anselin, 1995):

1. Al calcular un valor del estadístico local para cada observación se proporciona información que permite determinar la relevancia de agrupaciones espaciales de valores similares en torno a la observación en cuestión.
2. La suma de los estadísticos de las observaciones equivalen a un indicador global.

Entre los estadísticos locales está el contraste local *I de Moran* desarrollado por Luc Anselin (Anselin, 1995). Este estadístico mide el grado de asociación existente entre un atributo localizado en un área i y sus vecinos. Su cálculo es el siguiente :

$$I_i = \frac{z_i}{\sum_i z_i^2 / N} \sum_{j \in J_i} w_{ij} z_j \quad (10)$$

donde z_i es el valor del atributo normalizado en la región i , J_i es el conjunto de regiones vecinas a i , w_{ij} es la matriz de pesos espaciales y N es el tamaño de la muestra.

Cuando se cuenta con muestras suficientemente grandes, por lo general se asume que la I_i estandarizada sigue una distribución $N(0,1)$. Así, un valor positivo (negativo) del estadístico estandarizado evidencia la presencia de un *cluster* de valores similares (disímiles) alrededor de la región i .

6 Resultados

6.1 Suavización de Tasa Espacial

Como se mencionó en la sección 5, los datos de la ECV Medellín 2007 fueron agregados por pasos hasta llegar al nivel de barrios. En esta etapa se identificó que los barrios no eran la unidad de estudio adecuada, por lo cual se determinó la construcción de regiones analíticas sobre las cuales pudiera realizarse un análisis que arrojara resultados útiles para el desarrollo de políticas públicas.

El primer paso en la construcción de las regiones analíticas fue la elección de un bloque de preguntas de carácter socioeconómico de la ECV Medellín 2007 que recogían diferentes aspectos de pobreza de los hogares. En cada una de las preguntas se hizo una división para separar las respuestas que indicaban la existencia de pobreza de aquellas que demostraban lo contrario. De esta manera, en el primer caso se otorgó el valor de uno y en el segundo, el valor de cero. A manera de ilustración, como se aprecia en la tabla 1, la pregunta cuatro de la encuesta es la siguiente:

Material predominante de los pisos de la vivienda habitada por el hogar:

1. Tierra o arena
2. Madera burda, tabla o tablón
3. Cemento o gravilla
4. Baldosa, vinilo, tableta o ladrillo
5. Alfombra o tapete de pared a pared, mármol, parqué, madera pulida

En esta pregunta se hizo una división entre la respuesta 1. y las demás respuestas puesto que, de acuerdo con los criterios utilizados a nivel internacional, existe déficit cualitativo de vivienda cuando el material del piso es en tierra.

Para cada barrio se obtuvo la frecuencia de pobreza y su respectivo número de hogares con el objetivo de calcular las tasas porcentuales, pero las muestras para los barrios presentaban una gran disparidad. Por ejemplo, el barrio 705 arrojó una tasa de 100% en la variable cuartos para dormir y otros usos pero solamente se encuestaron seis hogares en dicho barrio. Acá se evidenció lo expresado en la sección 5.1 en cuanto a que las muestras muy pequeñas en ocasiones presentan valores muy elevados. Por el contrario, el barrio 103, con 213 hogares encuestados, tuvo una tasa de 0,47% para la misma variable.

Tabla 1: Preguntas de carácter socioeconómico de la ECV Medellín 2007

Número de la pregunta	Descripción
4	¿Cuál es el material predominante de los pisos?
5	¿Cuál es la fuente de donde la vivienda toma el agua?
6	Servicios públicos que posee la vivienda: ¿Posee acueducto?
6	Servicios públicos que posee la vivienda: ¿Posee alcantarillado?
7	¿La vivienda se encuentra ubicada en zona de alto riesgo?
8	Cuartos de la vivienda: ¿Tiene cocina?
8	Cuartos de la vivienda: ¿Tiene cuartos para dormir y otros usos?
10	¿Cuál es el servicio sanitario utilizado por este hogar?
13	Electrodomésticos propiedad del hogar: ¿Posee nevera o enfriador?
14	¿Dónde preparan los alimentos en este hogar?
15	¿Con qué energía o combustible cocinan en este hogar?
35	¿Sabe leer y escribir más de un párrafo? (Aplica sólo para personas de 5 y más años)
51	¿Tiene nivel de Sisbén?
136-138	¿Cuánto recibió el mes pasado en todos sus ingresos?
175	¿Ha oído usted hablar o conoce sobre los métodos de planificación familiar?
181	¿En el último mes por falta de dinero se presentó alguna de las siguientes situaciones de desnutrición?

Fuente: Tabla construida con base en la ECV Medellín 2007

Para hacer frente a los inconvenientes enfrentados cuando se tienen muestras dispares, se aplicó la suavización de tasa espacial en un *software* llamado *GeoDa* (Anselin, 2003). Este programa facilitó la georreferenciación de las variables socioeconómicas de la ECV Medellín 2007 permitiendo obtener el primer diagnóstico de la situación de pobreza de los barrios, tomando como base las tasas suavizadas. Después se utilizó un sistema de información geográfica (SIG) llamado *ARC GIS* donde se introdujeron los valores de las tasas y mediante los mapas de quiebres naturales (*natural breaks*) se pudieron detectar de forma preliminar las zonas aparentemente más vulnerables de la ciudad. De las dieciséis preguntas seleccionadas inicialmente, por tener valores iguales a cero, se descartaron las concernientes a la fuente de agua, presencia de cocina en la vivienda, acueducto y energía o combustible utilizado para cocinar.

Las figuras 2, 3 y 4 corresponden a los valores de las tasas suavizadas divididos en cinco rangos cada uno con su respectivo color. Conforme se va oscureciendo el color, más crítica es la situación de pobreza. Salta a la vista el amarillo predominante en los mapas de las variables alcantarillado, material de los pisos, servicio sanitario y vivienda ubicada en zona de riesgo. Este es un indicador de buenos resultados de Medellín en materia de infraestructura. Por el contrario, la tonalidad oscura que domina el mapa de ingresos totales puede confundir al lector, pues un gran porcentaje de barrios tiene valores cercanos a uno. Investigando al respecto, se observó en las respuestas de la ECV Medellín 2007 que un gran número de personas se rehusó a revelar sus ingresos, quizá por miedo a seguimientos por parte de las autoridades fiscales, aunque la encuesta dice explícitamente que sólo se utiliza para fines estadísticos. En las demás variables se observa en general una división entre la zona norte y la zona sur de la ciudad, con la primera exhibiendo mayores niveles de pobreza y la segunda arrojando mejores indicadores de calidad de vida. Se destacan las partes nordeste y oeste en las cuales se ubican barrios con los mayores índices de pobreza según las variables escogidas de la ECV Medellín 2007. Estos se concentran principalmente en las comunas Popular y San Javier que cuentan con un historial de problemas en materia socioeconómica.

En los mapas también se aprecian zonas en color blanco que corresponden a lugares de la ciudad que no eran relevantes para el análisis de pobreza llevado a cabo en esta investigación. Estas áreas se removieron siguiendo los criterios utilizados por el Departamento Administrativo de Planeación de la Alcaldía de Medellín que las clasifica como áreas institucionales o áreas de expansión. Algunos de los sitios removidos fueron: Ecoparque Cerro El Volador, Unidad Deportiva Atanasio Girardot, Jardín Botánico, Parque Norte, Oleoducto, Terminal de Transporte y Centro Administrativo La Alpujarra.⁹ Adicionalmente, por falta de datos, se quitaron seis barrios: Corazón de Jesús, Guayaquil, Calle Nueva, Perpetuo Socorro, Barrio Colombia y Tenche. Su ubicación corresponde a la franja blanca de mayor tamaño que puede observarse en el centro de los mapas.

⁹Para más información al respecto, acceder a: <http://www.medellin.gov.co/ReportesSite/index.jsp>

Figura 2: Distribución espacial de las variables de la ECV Medellín 2007

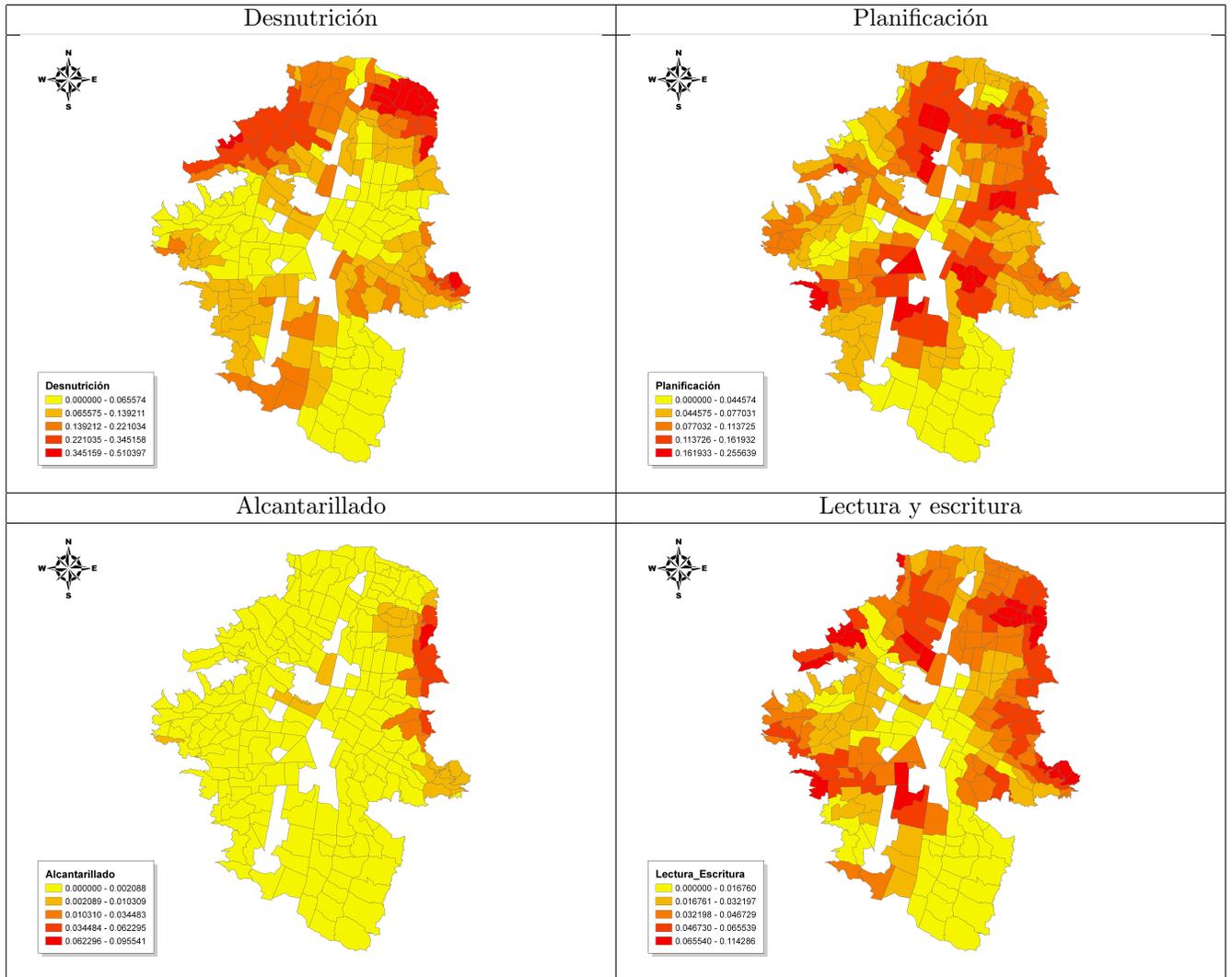


Figura 3: Distribución espacial de las variables de la ECV Medellín 2007

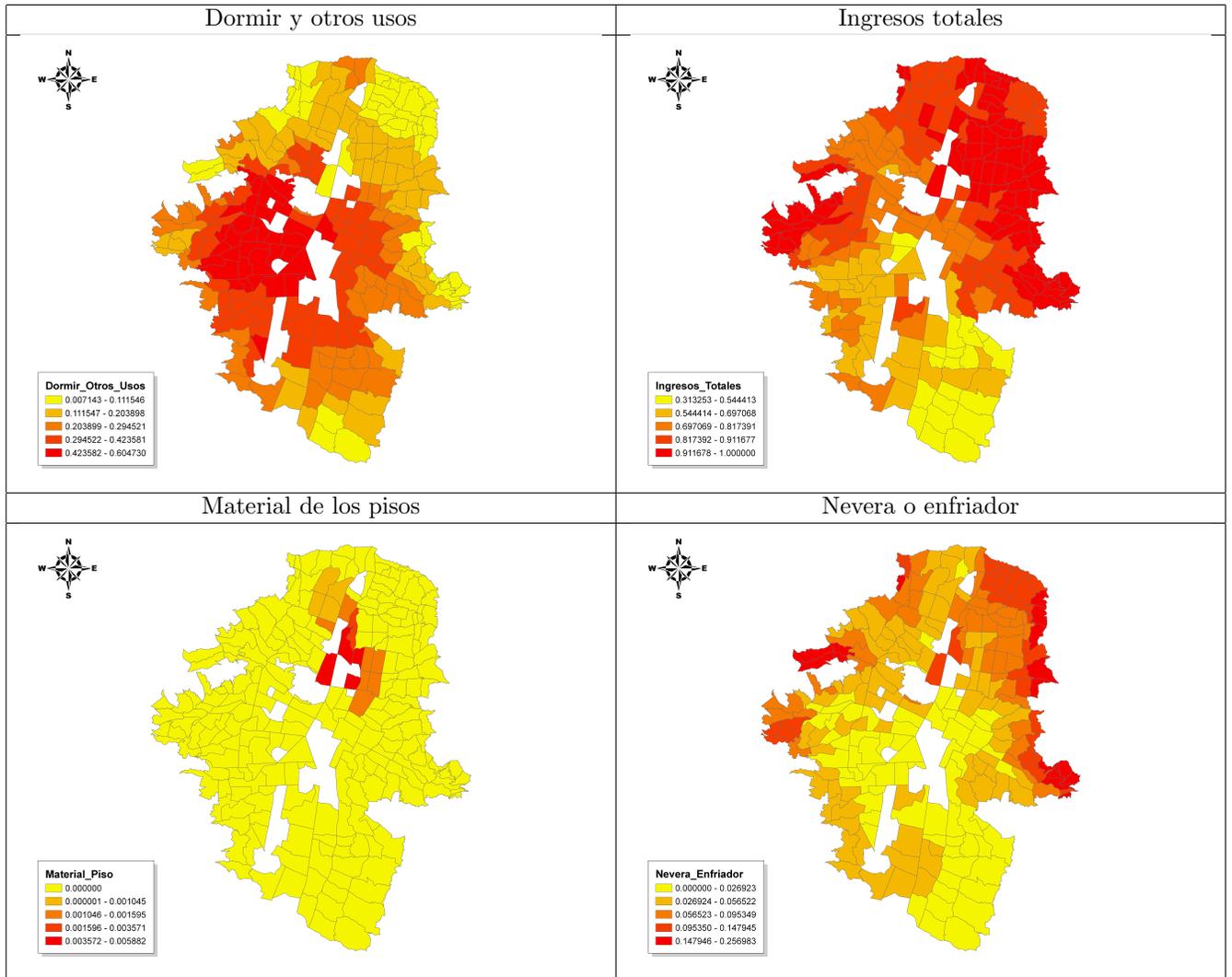


Figura 4: Distribución espacial de las variables de la ECV Medellín 2007

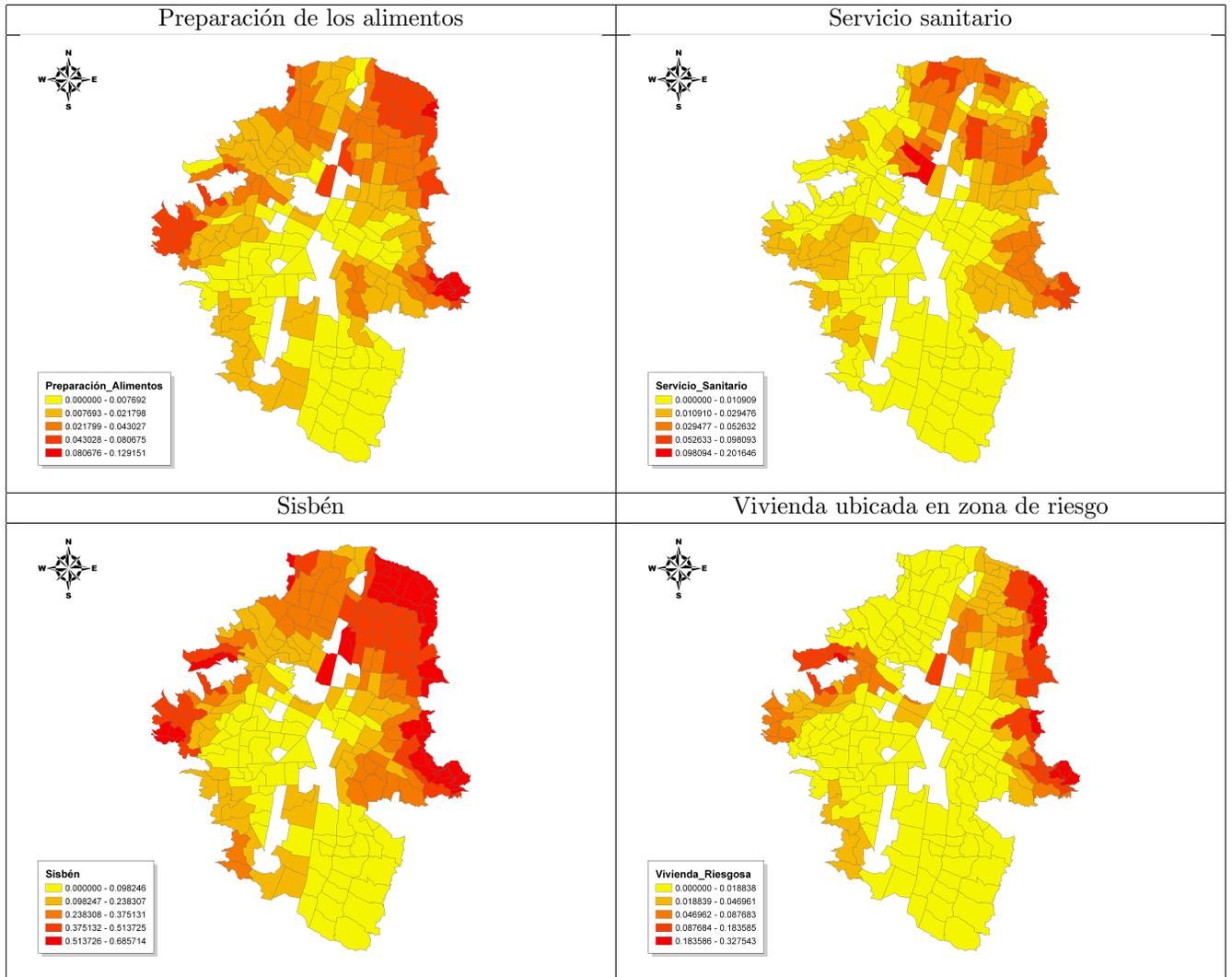


Tabla 2: Valores de la *I de Moran* global para las tasas suavizadas

Variable	<i>I de Moran</i>
spatrate desnutrición	0,7707
spatrate planificación	0,6609
spatrate alcantarillado	0,6023
spatrate lectura escritura	0,6811
spatrate dormir otros usos	0,8811
spatrate ingresos totales	0,8948
spatrate material piso	0,4984
spatrate nevera enfriador	0,8045
spatrate preparación alimentos	0,8284
spatrate servicio sanitario	0,6610
spatrate sisbén	0,8855
spatrate vivienda riesgosa	0,7411

*En todos los casos, al realizar 999 permutaciones el $p - value < 0,01$

En el caso de todas las variables existe autocorrelación espacial a un nivel de significación del 1%, lo cual prueba que la pobreza en Medellín exhibe patrones definidos en el espacio. Esto se observa en la tabla 2 donde todos los valores de la *I de Moran* global resultan estadísticamente significativos tras realizar 999 permutaciones. Los resultados del estadístico apoyan lo apreciado en los mapas, ya que por lo general las variables se caracterizan por la concentración de valores altos en ciertas zonas de la ciudad y valores bajos en otras áreas.

6.2 Modelo Factorial de Componentes Principales

La tabla 3 incluye información de seis de los componentes que resultaron al correr el modelo factorial de componentes principales en *Stata*. Es claro que el primero fue el que mayor varianza total capturó, lo cual se observa bajo el encabezado proporción. Su diferencia es muy notoria con relación a los demás componentes pero se decidió conservar el segundo porque tenía una interpretación coherente que facilitaba el análisis de las variables socioeconómicas. Se menciona también que el criterio del valor propio mayor que uno para elegir los componentes no siempre es adecuado porque según este se debería conservar el componente tres y descartar el cuatro cuando en realidad la diferencia en la varianza que recogen no es significativa. Esto se constató en esta investigación pues inicialmente, siguiendo este criterio, se retuvieron tres componentes. Sin embargo, se optó por descartar el tercero porque no se le encontró una interpretación clara y, además, la varianza que recogía no era significativa. Los dos componentes

Tabla 3: Varianza de las variables explicada por los componentes

Componente	Valor propio	Diferencia	Proporción	Acumulado
Componente 1	5,71435	4,42998	0,4762	0,4762
Componente 2	1,28437	0,15974	0,1070	0,5832
Componente 3	1,12463	0,14324	0,0937	0,6769
Componente 4	0,98139	0,23583	0,0818	0,7587
Componente 5	0,74556	0,05570	0,0621	0,8209
Componente 6	0,68986	0,22419	0,0575	0,8783

retenidos fueron rotados ortogonalmente para que cada variable tuviera una relación más fuerte con uno de estos y en búsqueda de resultados que tuvieran una interpretación razonable. Esta clase de rotación garantizó mantener la correlación de cero entre los componentes.

La tabla 4 muestra la relación estrecha que guarda el primer componente con la mayoría de las variables tras la rotación. Estas variables corresponden a las tasas suavizadas de las preguntas de la ECV Medellín 2007 que generalmente aplican por igual a todos los miembros del hogar. Tal fue el caso de material de los pisos, alcantarillado, vivienda ubicada en zona de riesgo, cuartos para dormir y otros usos, nevera o enfriador, preparación de los alimentos, ingresos totales, Sisbén y desnutrición. Estas variables, con excepción de dormir y otros usos, tienen cargas con signo positivo, lo cual significa que los barrios con valores altos para el componente uno sufren condiciones de precariedad en las variables mencionadas. Por su parte, el segundo componente guarda una relación más estrecha con las tasas de servicio sanitario, planificación y lectura y escritura. Exceptuando servicio sanitario, las otras dos tienen la característica de ser preguntas personales debido a que miembros de un mismo hogar pueden dar respuestas diferentes. Aunque este componente es más difícil de interpretar, puede decirse que valores altos representan barrios afectados por la falta de educación (por el lado de la planificación y el alfabetismo) y servicio sanitario de mala calidad.

La figura 5 presenta las puntuaciones de los componentes por barrios y en ella puede apreciarse que los componentes capturan adecuadamente el comportamiento de las variables de la encuesta al conservar el patrón de disparidad entre el norte y el sur de la ciudad. En el caso del componente uno se observa claramente cómo las condiciones socioeconómicas de los barrios de Medellín van empeorando conforme estos se ubican más lejos del centro. La gran mayoría de los barrios periféricos, incluso en el sur, tienen las puntuaciones más altas. Las zonas nordeste y oeste mantienen su tonalidad oscura tal y como se observó anteriormente en el comportamiento individual de las variables socioeconómicas.

Tabla 4: Cargas de los componentes rotados ortogonalmente

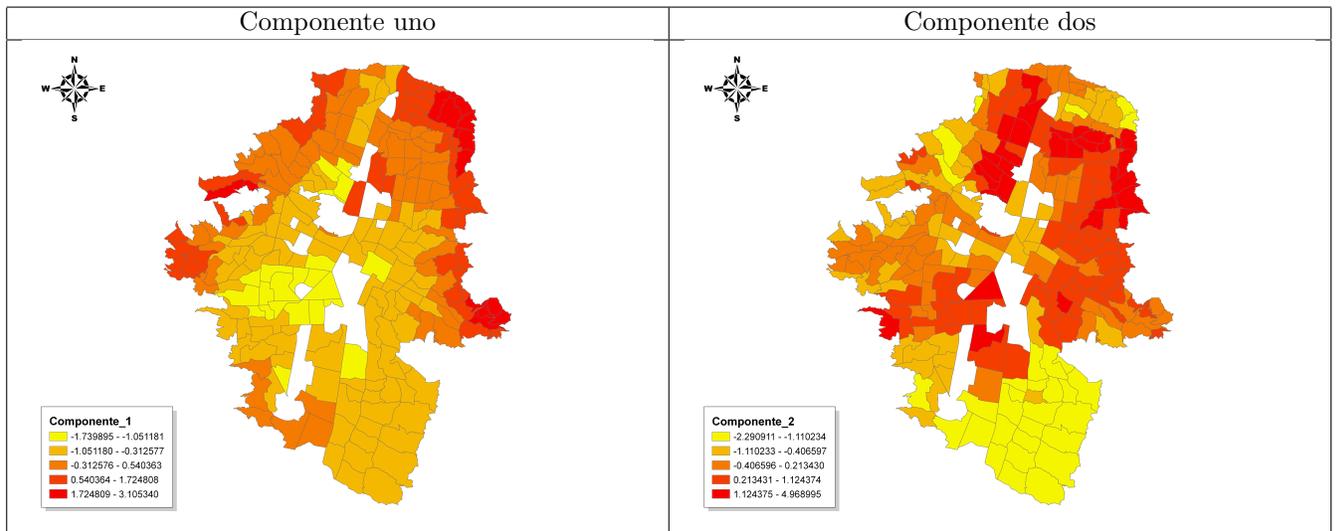
Variable	Componente 1	Componente 2	Unicidad
spatrate matpiso	0,0933	0,0056	0,9913
spatrate alcant	0,3995	0,2973	0,7520
spatrate viviriesgo	0,8045	0,1114	0,3404
spatrate dormotros	-0,7907	-0,0114	0,3747
spatrate servsanit	0,2433	0,6421	0,5285
spatrate nevenf	0,9250	0,1514	0,1215
spatrate prepalim	0,8904	0,1184	0,1931
spatrate lecescl	0,5680	0,6064	0,3097
spatrate ingtotales	0,5736	0,4741	0,4462
spatrate sisben	0,9153	0,2617	0,0938
spatrate planif	0,0330	0,8317	0,3072
spatrate desnut1	0,6741	0,0524	0,5428

Lo mismo no puede afirmarse al analizar el mapa del componente dos, el cual exhibe valores más bajos para estas zonas. Teniendo en cuenta que este componente captura los niveles de educación de los barrios, la explicación para los mejores indicadores se encuentra en el programa de construcción de Parques Biblioteca diseñado por la Alcaldía de Medellín. El efecto más notable está relacionado con la comuna Popular donde el impacto del Parque Biblioteca España se manifiesta especialmente en los barrios Santo Domingo Savio No. 2 y La Avanzada ubicados en la parte nororiental e identificados con color amarillo. La zona de influencia de este Parque Biblioteca abarca catorce barrios que se benefician, entre otros, de 108 computadores con acceso gratuito a internet, un auditorio con capacidad para 179 personas y una ludoteca para los niños hasta diez años de edad.¹⁰ Los resultados positivos en estos barrios quedan evidenciados en el color naranja claro que se aprecia en la parte nororiental del mapa. En el caso puntual del Parque Biblioteca San Javier, su influencia se extiende a una población de 341 mil habitantes en la zona centro-occidente de Medellín y cuenta con acceso gratuito a internet. También ofrece salas de libros para adultos y niños con una colección que asciende a los 15 mil volúmenes.

Para culminar la fase del modelo factorial de componentes principales, en una base de datos se juntaron las puntuaciones de los componentes para cada barrio y el número de hogares por barrio. Esta información sirvió de entrada para el siguiente paso que fue la utilización del algoritmo *max-p-region*, el cual definió las nuevas regiones analíticas.

¹⁰Información obtenida de: <http://www.reddebibliotecas.org.co>

Figura 5: Comportamiento espacial de las puntuaciones de los componentes

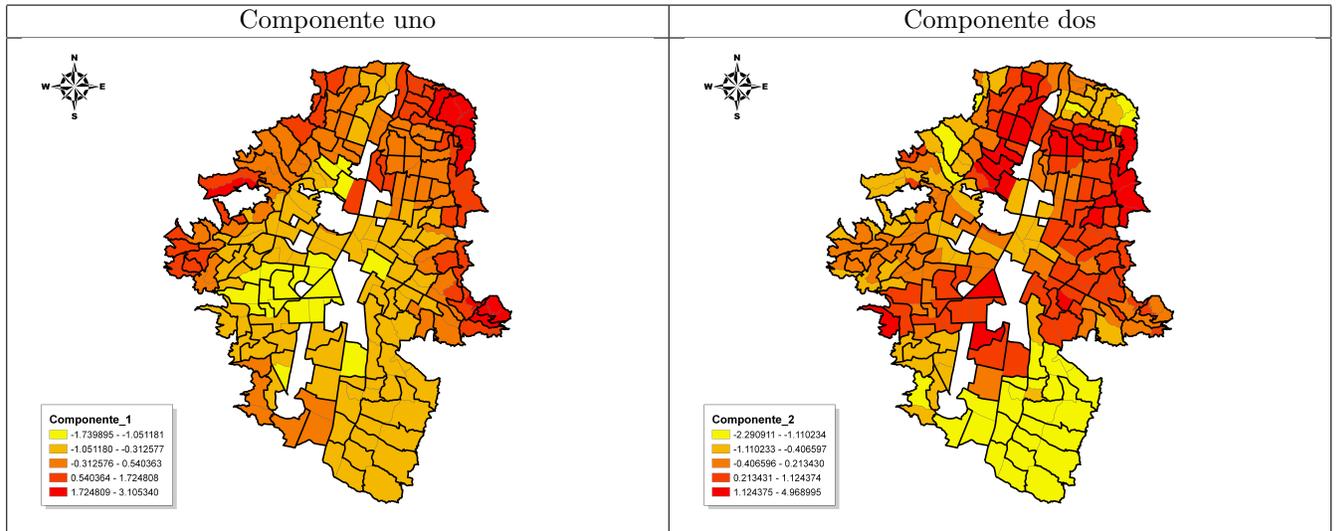


6.3 Max-P-Region

Al organizar la información de la ECV Medellín 2007 por barrios y encontrar que esta unidad geográfica no contaba con muestras representativas, se decidió recurrir a una metodología que permitiera crear unas nuevas unidades geográficas, llamadas regiones analíticas, que maximizaran tanto la homogeneidad entre sus componentes como las diferencias interregionales (Duque et al., 2008; Byfuglien and Nordgard, 1973). En consecuencia se adoptó el modelo *max-p-region* para el cual fue necesario utilizar los componentes calculados en el modelo factorial de componentes principales los cuales sirvieron como variables de agregación. Para fines de este estudio, se buscó que los barrios con tasas similares de las preguntas de la ECV Medellín 2007 compartieran la misma región. Respecto del factor espacial, se eligió una contigüidad de reina de orden uno, es decir, la vecindad para las áreas geográficas se estableció cuando estas compartían un lado o vértice. Por último, antes de correr el algoritmo, se fijó una restricción basada en el número de hogares deseados por región analítica; el umbral fue de 100 hogares por región. En vista de que muchos de los barrios no cumplían con este umbral, el número de regiones analíticas creadas fue inferior al de barrios. A partir de 243 barrios se generaron 131 regiones.

En la figura 6 se aprecian las 131 regiones analíticas creadas por el algoritmo *max-p-region*. Las líneas negras gruesas corresponden a las fronteras de las regiones mientras que las líneas delgadas pertenecen a los límites de los barrios. En otras palabras, el mapa de las regiones analíticas está superpuesto al de los barrios con el objetivo de

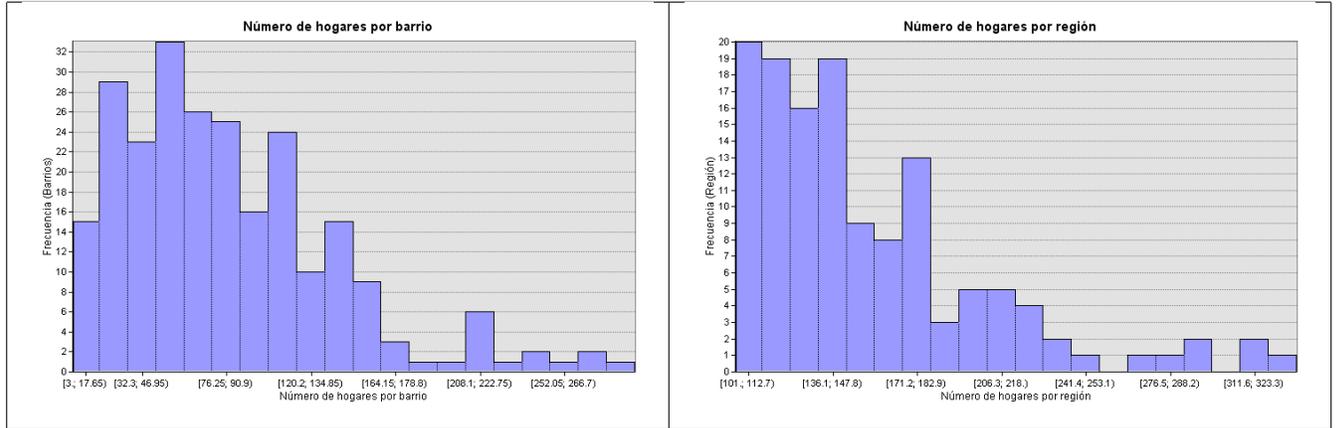
Figura 6: Puntuaciones de los componentes según regiones analíticas



señalar las agrupaciones obtenidas a través del proceso de agregación. A primera vista se destaca el hecho de que el proceso logró el objetivo de reunir en una misma región aquellos barrios con características similares, algo que se evidencia en la homogeneidad de los colores al interior de las regiones. Se confirma la uniformidad intra-regional y la conservación de los patrones espaciales detectados en estos mismos mapas a nivel de barrios. Prueba de ello es que se mantienen los mismos efectos en los mapas de cada uno de los componentes, con un comportamiento más visible en el mapa del componente uno frente al del componente dos debido a que el primero permite detectar *clusters*. En el componente uno las regiones preservan la relación centro-periferia al darse un deterioro gradual de las condiciones de vida a medida que la distancia aumenta respecto del centro.

El mínimo poblacional de 100 hogares por región fue satisfecho y, como lo muestra la figura 7, se atenuaron las marcadas diferencias en las muestras de hogares que presentan los barrios. Mientras estos tienen muestras que oscilan entre 3 y 296 hogares, en las regiones analíticas el rango está entre 101 y 335 hogares. Además, la desviación estándar del número de hogares por región analítica es 49,43, menor a la de hogares por barrio cuyo valor es 56,66. Por último, una media de 156,30 hogares por región superior a la media de 84,26 hogares por barrio mejora las condiciones para la inferencia estadística.

Figura 7: Comparación entre número de hogares por barrio y número de hogares por región analítica



6.4 Índice Slum

Con las regiones analíticas construidas fue posible comenzar a hacer el análisis de pobreza en Medellín basado en las variables *slum* mencionadas en la sección 5.4. Inicialmente se revisó cada una de estas variables por hogares asignando el valor de uno cuando se cumplía el evento de pobreza y registrando el valor de cero en caso contrario. El resultado fue el cálculo del índice *slum* para cada uno de los 21.861 hogares encuestados de acuerdo con la ecuación (6).

La tabla 5 indica que la situación de pobreza de los hogares de Medellín no es preocupante cuando se trata de cobertura en servicios públicos, ya que sólo un porcentaje muy bajo no cuenta con acueducto o alcantarillado. A esto se suma el hecho de que apenas 0,14% de los hogares habita una vivienda con déficit cualitativo dado el material de las paredes. No obstante, el panorama es diferente cuando se analiza la propiedad de las viviendas y la condición de hacinamiento, las cuales tienen porcentajes de dos dígitos pero que, a pesar de ser mayores, se ubican por debajo del 50%. Los resultados también demuestran que ninguno de los hogares presenta la peor situación de pobreza, es decir, ninguno tiene un índice de cinco. Además, apenas dos hogares presentan cuatro eventos de pobreza y 11.204 tienen un índice igual a cero, lo cual refleja que la mayoría de los hogares no está afectada por las características *slum*.

Trasladando el análisis hacia las regiones analíticas, el índice *slum* para cada una de estas se calculó siguiendo la ecuación (7). El comportamiento espacial del índice de las regiones analíticas se analiza a través del estadístico *I de Moran* global que permite establecer la existencia de autocorrelación espacial. Como lo muestra la figura 8, evi-

Tabla 5: Caracterización de los hogares de Medellín por medio de las variables *slum*

Variable <i>slum</i>	Porcentaje de hogares
El hogar no cuenta con servicio de acueducto	0,01
El hogar no cuenta con inodoro conectado a alcantarillado	3,77
Más de dos personas conviven en una misma habitación	16,77
El hogar habita una vivienda cuyo material predominante de las paredes es de mala calidad	0,14
El hogar no es el dueño de la vivienda que habita	35,46
Total de hogares	21.861

dentamente el valor que toma el índice en una determinada región afecta el valor de sus vecinos. Esto se puede afirmar porque la línea azul, cuya pendiente equivale a la *I de Moran* global, se sale de las bandas creadas por las líneas punteadas de color rojo. En el diagrama también puede observarse un patrón en los puntos al coincidir la mayoría de los valores en los cuadrantes superior derecho e inferior izquierdo. La interpretación que se deriva de esto es que regiones con valores altos tienen vecinos con valores altos y regiones con valores bajos tienen vecinos con valores bajos. Este hecho da pie para analizar la situación más a fondo mediante la herramienta de los mapas y confirmar si en efecto existen regiones en Medellín con valores similares que son consideradas como relevantes.

Figura 8: Comportamiento de las regiones analíticas en el diagrama de dispersión de Moran

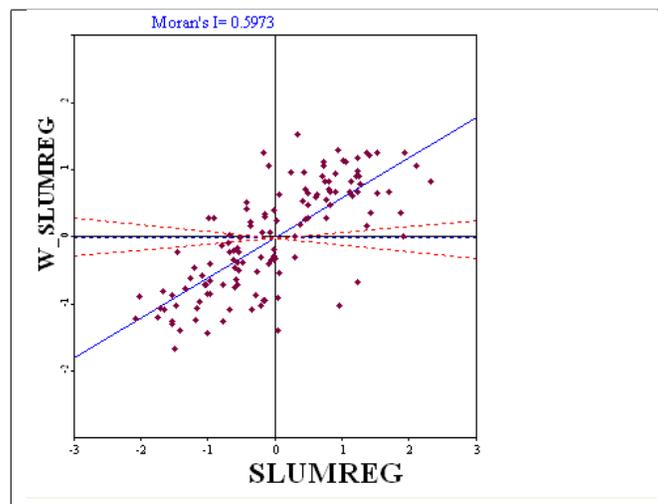


Tabla 6: Situación de las regiones analíticas según variables *slum*

Variable <i>slum</i>	Región	Barrios	Valor
Falta de inodoro conectado a alcantarillado	9	Altamira Córdoba	0,2143
Más de dos personas por habitación	130	Las Estancias	0,4158
Material de las paredes de mala calidad	46	Versalles No. 2 La Cruz Oriente San José La Cima No. 2	0,07547
Residentes no son dueños de la vivienda	24	B. Cerro El Volador La Pilarica Bosques de San Pablo El Diamante	0,5231
Índice <i>slum</i>	46	Versalles No. 2 La Cruz Oriente San José La Cima No. 2	0,9151

6.5 Variables Slum

Para el caso de las variables *slum* individuales, sus valores se computaron sumando la frecuencia de valores iguales a uno por región analítica y luego dividiendo por el número de hogares en la región.

La tabla 6 expone las regiones más afectadas por los diferentes componentes *slum*, así como la región con el mayor índice. Los resultados son muy similares a los obtenidos en el análisis a nivel de hogares, con la variable propiedad de la vivienda siendo la de mayor valor y acueducto la de menor. Con respecto a esta última, llama la atención que todas las regiones presenten un valor igual a cero, por lo cual se ha omitido de la tabla. Esto sugiere, según la información disponible de la ECV Medellín 2007, que hay plena cobertura de servicio de agua. También saltan a la vista los valores en conjunto puesto que, con excepción de propiedad de la vivienda, todos se encuentran por debajo del 50% y esto teniendo en cuenta que se trata de las regiones más afectadas.

En cuanto a la distribución espacial, en la figura 9 una vez más se confirma la tendencia de buenos indicadores en las variables de infraestructura, esto es, alcantarillado, acueducto y material de las paredes. El color amarillo domina estos mapas y sólo se observa una concentración de puntos críticos al nordeste de la ciudad en el caso de alcantarillado. Respecto de las variables propiedad de la vivienda y hacinamiento, existe un

impacto diferente según la zona de la ciudad analizada. Por ejemplo, el hacinamiento empeora cuando la ubicación de una región analítica se aleja del centro y tiende hacia la periferia. La propiedad de la vivienda tiene un comportamiento más disperso y puede afirmarse que, a excepción de la comuna Poblado que presenta algunas regiones en amarillo, es común que los habitantes de Medellín residan en viviendas que no les pertenecen.

En la misma figura puede apreciarse que existe una zona de la ciudad hacia el noroeste que contiene las regiones 9 y 24 las cuales son las más afectadas por la falta de conexión a alcantarillado y por la falta de propiedad de la vivienda, respectivamente. La región 46, ubicada en la parte nordeste de la ciudad, resulta ser la de peores cifras con relación al material de las paredes y al índice *slum*. Sus barrios - Versalles No. 2, La Cruz, Oriente y San José La Cima No. 2 - tienen todos una clasificación de estrato uno según la Alcaldía de Medellín. Algo diferente sucede con la región 24 ya que dos de sus barrios - Barrio Cerro El Volador y El Diamante - están clasificados en estrato dos, mientras que los barrios La Pilarica y Bosques de San Pablo pertenecen al estrato cuatro. Lo anterior demuestra que, si bien los resultados para las regiones son bajos, evidentemente existen zonas de la ciudad donde se concentran algunas características de pobreza.

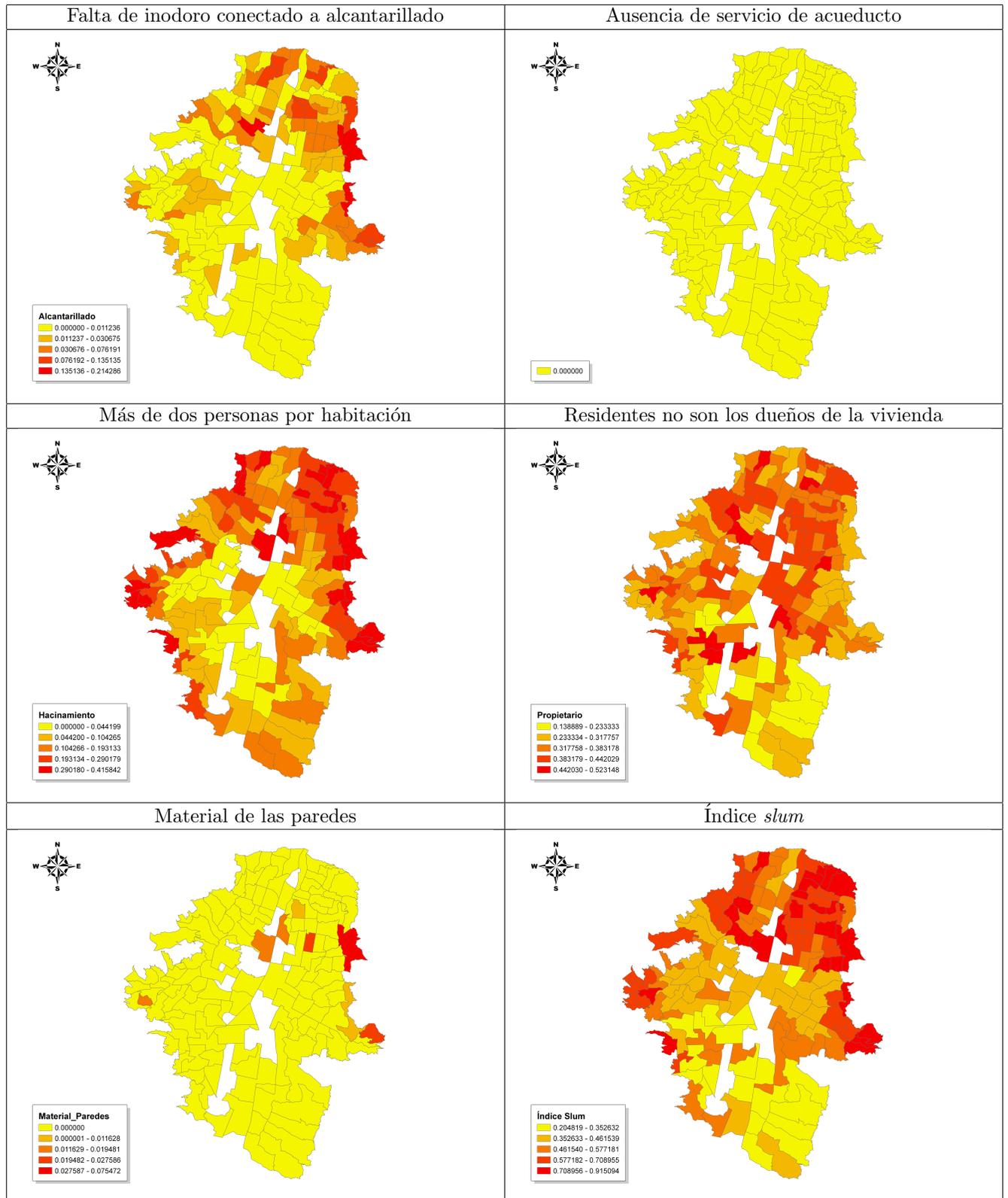
El caso de la región analítica 46 merece un análisis más detallado porque no sólo es la de peores resultados en dos de las variables *slum* sino que entre las 131 regiones creadas ocupa el quinto puesto en la clasificación de las regiones analíticas más afectadas por estas variables.¹¹ Sus barrios pertenecen a la comuna Manrique ubicada en la parte este-nordeste de Medellín. Las condiciones precarias de esta comuna se confirman al analizar los datos más actualizados de la Encuesta de Calidad de Vida realizada por la Alcaldía de Medellín y que corresponden al año 2006 (ECV Medellín 2006). Mientras la ciudad presenta un índice de desarrollo humano (IDH) de 80,35, Manrique está por debajo con 76,44. Esto prueba sus deficiencias en materia de educación, esperanza de vida al nacer e ingreso per cápita que son las medidas para este índice. En cuanto al índice de calidad de vida (ICV), sus números también son desalentadores al presentar un índice de 79,81 en comparación con el nivel de 83,77 de la ciudad. Vale la pena mencionar que entre mayor sea el número de estos índices menor es la vulnerabilidad.¹²

La comuna Manrique tiene una historia de expansión de su territorio a lo largo del siglo pasado especialmente por el auge de la actividad industrial y de la minería. Sin embargo, los barrios que conforman la región 46 - Versalles No. 2, La Cruz, Oriente y San José La Cima No. 2 - se incorporaron a la comuna recientemente, en 1993, y tienen en común que su ocupación se dio como resultado de la agudización de la violencia. He aquí una explicación para las condiciones precarias bajo las cuales viven sus habitantes. En general, la comuna se ha convertido en albergue para las familias

¹¹Ver la tabla 7

¹²A esta información se puede acceder a través de la página de internet de la Alcaldía de Medellín en la dirección: <http://www.medellin.gov.co>

Figura 9: Distribución espacial de los componentes del índice *slum* a nivel de regiones analíticas



desplazadas que se han asentado en las montañas con pendientes altas y caracterizadas por sufrir deslizamientos e inundaciones. Esto quedó registrado en la figura 4 en el mapa de la variable vivienda ubicada en zona de riesgo. Como puede apreciarse, la periferia de la zona nororiental de la ciudad concentra algunos de los barrios más críticos, entre los cuales se encuentran los de la región 46. Con base en las cifras publicadas en el Anuario Estadístico de Medellín 2005, Manrique posee una densidad poblacional de 267 habitantes por hectárea, un grado de hacinamiento comparativamente mayor al nivel de la ciudad. Puede observarse esto en la figura 9 en el mapa de la variable hacinamiento.

Cuando se comparan los mapas de los componentes *slum* con el mapa del índice, queda claro que este último captura los patrones de las variables de manera individual, con zonas críticas en las partes nordeste, noroeste y suroeste de Medellín. En primera instancia se podría afirmar que sólo basándose en el índice se podrían detectar las zonas más vulnerables para cada una de las variables. No obstante, esto se contrastó con la elaboración de la tabla 7 en la cual aparecen las regiones analíticas más afectadas por las diferentes variables del índice calculado.¹³ Por ejemplo, si uno tomara las trece regiones más afectadas por el índice *slum*, obtendría resultados muy próximos a las regiones más golpeadas por todas las características *slum* en conjunto, ya que cada una de estas regiones se ubica entre las diez más afectadas en por lo menos dos de las variables analizadas individualmente. Sin embargo, tomar como referencia únicamente los resultados arrojados por el índice puede conducir a interpretaciones erróneas. Varias regiones cuya clasificación en el índice es mayor al puesto veinte están dentro de las diez primeras posiciones entre las regiones más afectadas por al menos dos de los componentes *slum*. Tal es el caso de las regiones 52 y 92. El criterio del índice también llevaría a la no identificación de la región 9, la cual presenta la peor condición en conexión a alcantarillado pero se ubica en el puesto 67 en el índice. Por estos motivos, es aconsejable analizar el impacto individual en las regiones de cada una de las características *slum* y no solamente basarse en resultados globales.

Tabla 7: Clasificación de las regiones analíticas más afectadas por las variables *slum*

Región	Barrios	Falta alcantarillado	Hacinamiento	Material paredes	Propiedad vivienda	Índice <i>slum</i>
10	La Isla	11	3	9	49	3
130	Las Estancias	7	1	9	60	2
24	Barrio Cerro El Volador La Pilarica	14	56	9	1	9

continúa en la siguiente página

¹³Las regiones que están más arriba en la tabla son las más afectadas y conforme uno se desplaza hacia abajo la vulnerabilidad se reduce. El hecho de que algunas regiones se ubiquen en la misma posición para alguna variable se debe a que presentan el mismo valor. También se hace la observación de que se omitió la variable acueducto porque todas las regiones tienen un valor de cero.

Tabla 7: *continuación*

Región	Barrios	Falta alcantarillado	Hacinamiento	Material paredes	Propiedad vivienda	Índice <i>slum</i>
	Bosques de San Pablo El Diamante					
40	Moravia Caribe Progreso	51	10	5	21	6
46	Versalles No. 2 La Cruz Oriente San José La Cima No. 2	3	4	1	109	1
98	Santander	87	16	9	2	4
41	La Francia	9	8	9	89	10
124	Manrique Oriental	58	42	9	3	13
52	San Isidro	4	54	8	39	22
59	Andalucía Villa Niza	59	33	9	9	17
36	Villa Turbay La Sierra Villa Lilliam	5	11	3	102	7
32	Llanaditas Trece de Noviembre	2	5	7	117	5
47	Veinte de Julio	87	27	9	5	12
126	Versalles No. 1	34	7	9	72	20
3	Popular	83	9	9	38	11
5	Santa Cruz	87	2	9	46	8
108	Kennedy	87	26	9	10	21
62	La Rosa Berlín	10	44	9	73	33
105	Campo Valdés No. 1	16	86	2	16	50
79	Villatina San Antonio	25	36	6	106	51
93	Trinidad	56	110	9	4	58
92	Las Independencias	87	6	4	111	34
69	Barrio Colón Las Palmas	87	87	9	8	61

continúa en la siguiente página

Tabla 7: *continuación*

Región	Barrios	Falta alcantarillado	Hacinamiento	Material paredes	Propiedad vivienda	Índice <i>slum</i>
107	Pedregal	6	78	9	98	69
63	Tejelo	8	106	9	64	74
103	San Bernardo	84	115	9	7	73
9	Altamira Córdoba	1	103	9	115	67
110	Belén Granada Miravalle	85	124	9	6	75

Fuente: Cálculos realizados por el autor

6.6 Clusters

En esta investigación se utilizaron los mapas de *clusters* de *GeoDa* aplicando 999 permutaciones en cada caso para obtener agrupaciones de regiones con la mayor confiabilidad posible. La aplicación del estadístico local *I de Moran* mostró que en Medellín efectivamente existen agrupaciones de regiones estadísticamente significativas en las cuales se aglomera la pobreza y otras, totalmente contrarias, donde se concentran los mejores índices de calidad de vida. Basta con observar en la figura 10 los mapas de las variables hacinamiento e índice *slum* para comprobar el primer caso. El color rojo representa la aglomeración de regiones con valores altos. Se conserva la tendencia de la mayor concentración de pobreza en el nordeste de la ciudad pero también llama la atención la zona este (la parte más saliente del lado derecho del mapa) donde se aprecia la presencia repetida del color rojo en los diferentes mapas. Las regiones 36 y 130 forman parte de este *cluster* de pobreza siendo unas de las regiones más afectadas por las variables *slum* de acuerdo con los resultados expuestos en la tabla 7.

La región 36 comprende los barrios Villa Turbay, La Sierra y Villa Lilliam que están ubicados en la comuna Villa Hermosa. Estos tres barrios presentan la particularidad de ser catalogados en estrato uno en una comuna donde el estrato socioeconómico predominante es el dos. Por su parte, la región 130 está compuesta únicamente por el barrio Las Estancias ubicado también en la comuna Villa Hermosa. De acuerdo con los resultados de la ECV Medellín 2006, Villa Hermosa tiene un IDH por debajo del promedio de Medellín al ubicarse en 77,11 en tanto la ciudad en conjunto presenta un índice de 80,35. Adicionalmente, el ICV de Villa Hermosa es de 79,22 en comparación con 83,77 que presenta la ciudad. A lo anterior se suman los altos niveles de violencia que experimenta la comuna con grupos armados ilegales que ejercen la autoridad por

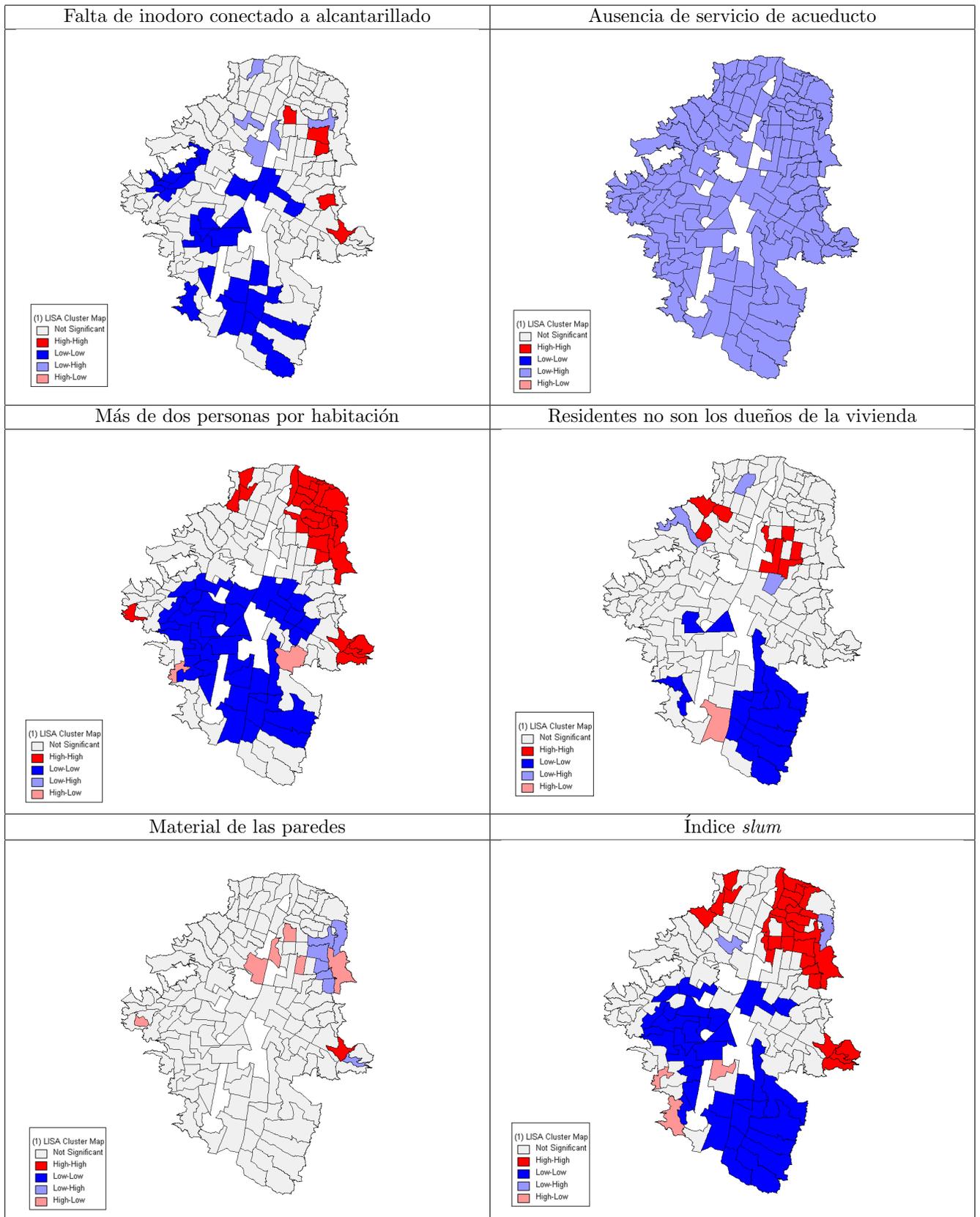
encima de las fuerzas estatales y que han influido negativamente en la deserción escolar y en la generación de una de las tasas más elevadas de desplazamiento intra-urbano.

En cuanto a los demás colores de los mapas de la figura 10, el color azul oscuro resulta de la agrupación de regiones con valores bajos cuya ubicación principal es la zona sur de Medellín que reúne las regiones de la comuna Poblado. Este resultado es muy lógico si se recuerda que a lo largo del estudio de la distribución espacial de las diferentes variables esta zona siempre ha presentado los mejores indicadores. Los colores rosado y azul claro también poseen autocorrelación espacial. Tienen una particularidad por involucrar grupos de áreas con valores disímiles y en la literatura se les conoce como valores atípicos espaciales (*spatial outliers*). Por último, las porciones de los mapas en blanco simplemente representan regiones que no manifiestan un comportamiento de autocorrelación espacial.

7 Conclusiones

La presente investigación ha demostrado que a través del índice *slum* y del AEDE es posible detectar las zonas marginales de la ciudad de Medellín. También ha presentado una alternativa muy importante para el análisis de la pobreza en la ciudad, ya que las regiones analíticas construidas capturan de manera acertada las características de pobreza de los barrios y agrupan aquellos con condiciones similares de tal forma que se logra una homogeneidad intra-regional. Muchas veces las divisiones administrativas son arbitrarias y por esto es más conveniente recurrir a métodos que posibiliten un estudio con otro tipo de delimitación geográfica. Hasta lo que se conoce, este es el primer estudio que incorpora el concepto de regiones analíticas para analizar la pobreza en Medellín, una innovación que podrá seguir siendo implementada en investigaciones futuras en vista de los resultados satisfactorios obtenidos.

Los resultados del índice *slum* comprueban que la pobreza en Medellín es un fenómeno que no se distribuye aleatoriamente en el espacio sino que presenta autocorrelación espacial, donde la pobreza en determinados barrios influye en la pobreza de sus vecinos. Tanto a nivel agregado para el índice como analizando sus componentes individuales pudo confirmarse la existencia de dicha autocorrelación y la concentración de puntos críticos en la ciudad. Debe señalarse que si bien algunas regiones analíticas están en los primeros puestos para más de una de las variables *slum*, la variabilidad es notoria, por lo cual es adecuado analizar el impacto de cada uno de los componentes por separado y no solamente basarse en el índice. El análisis individual es una prueba de que en Medellín la cobertura en servicios públicos es muy buena y que las viviendas presentan buenas condiciones de habitabilidad, por lo menos en lo que concierne al material de las paredes. Como pudo apreciarse en los resultados, las variables no afectan mucho a la población pues casi todas las regiones analíticas estuvieron por debajo del 50% del valor máximo.

Figura 10: Detección de *clusters* espaciales mediante el estadístico local *I de Moran*

No obstante, el hecho de que las zonas más pobres de la ciudad identificadas mediante el índice *slum* coincidieran con las de peores indicadores en otras mediciones de pobreza demuestra la importancia del índice. También vale la pena mencionar los patrones espaciales definidos que presentaron las variables socioeconómicas seleccionadas de la ECV Medellín 2007, con diferencias marcadas entre el norte y el sur de la ciudad y con condiciones de vida más precarias para los hogares ubicados en zonas periféricas.

El AEDE es una herramienta poderosa en el estudio de fenómenos sociales, como la pobreza, porque permite no solamente georreferenciar las áreas más vulnerables sino que sirve como punto de partida para realizar un diagnóstico de las mismas. Su implementación es aconsejable cuando se busca medir la eficacia de las políticas sociales, tal y como quedó registrado en este trabajo en el caso de los mejores índices de educación que presentan las comunas Popular y San Javier tras la puesta en marcha del programa Parques Biblioteca. El hecho de identificar regiones que estén correlacionadas puede llevar a que las autoridades públicas unan sus esfuerzos para enfrentar los problemas sociales de manera conjunta y logren resultados más exitosos. Valiéndose de las regiones analíticas creadas en esta investigación, los hacedores de políticas sociales podrán diseñar sus estrategias no a nivel macro, como las comunas, sino de forma más focalizada destinando los recursos a los barrios con mayores necesidades.

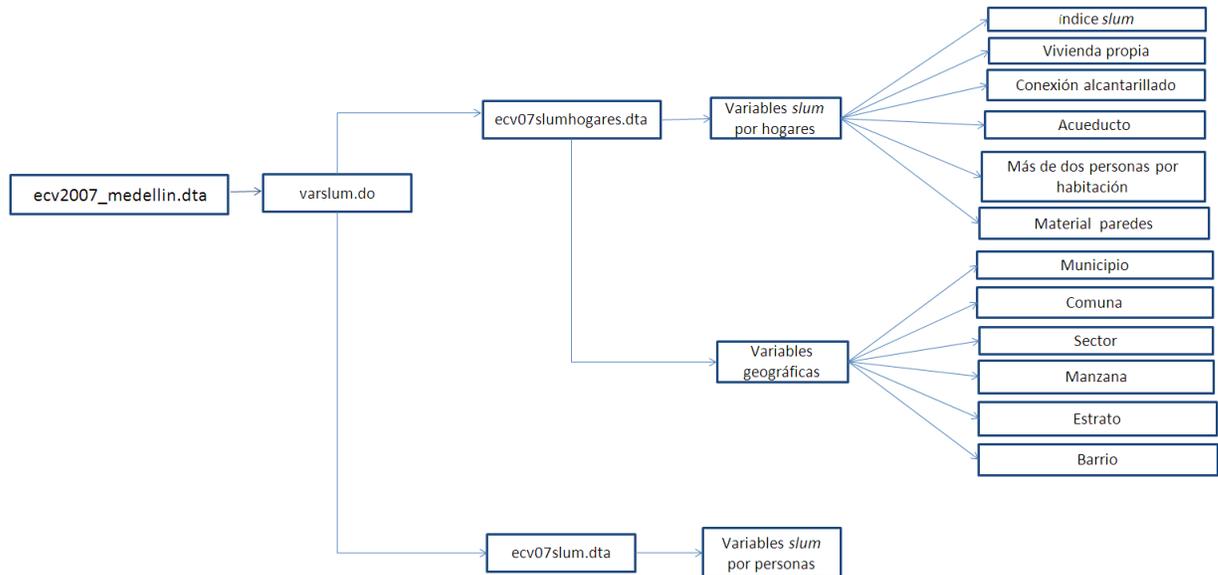
Bibliografía

- Acevedo, I. and Velásquez, E. (2008). Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. *Ecos de Economía*, (27):9–34.
- Agostini, C. A., Brown, P. H., and Góngora, D. P. (2008). Distribución espacial de la pobreza en Chile. *Estudios de Economía*, 35(1):79–110.
- Alcaldía de Medellín (2009a). Encuesta Calidad de Vida 2006. <http://www.medellin.gov.co>.
- Alcaldía de Medellín (2009b). Red de bibliotecas. <http://www.reddebibliotecas.org.co>.
- Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis*, 27:93–115.
- Anselin, L. (2003). GeoDa (tools for exploratory spatial data analysis). Disponible en: <http://geodacenter.asu.edu/software/downloads>.
- Anselin, L., Baller, R. D., Messner, S. F., Deane, G., and Hawkins, D. F. (2001). Structural covariates of U.S. county homicide rates: Incorporating spatial effects. *Criminology*, 39(3):561–590.
- Anselin, L., Kim, Y. W., and Syabri, I. (2004). Web-based analytical tools for the exploration of spatial data. *Journal of Geographical Systems*, 6:197–218.
- Byfuglien, J. and Nordgard, A. (1973). Region-building - a comparison of methods. *Norwegian Journal of Geography*, 27:127–151.
- Carlin, B. and Louis, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman And Hall/CRC, U.S., second edition.
- Departamento Administrativo de Planeación (2007). Estratificación predominante por barrio 2006 suelo urbano municipio de Medellín. <http://www.medellin.gov.co/ReportesSite/index.jsp>.
- Duque, J. C. (2004). *Design of homogeneous territorial units: A methodological proposal and applications*. PhD thesis, Departamento de Econometría Estadística y Económica Española, Universidad de Barcelona.
- Duque, J. C., Anselin, L., and Rey, S. J. (2008). The max-p-region problem.
- Duque, J. C., Ramos, R., and Surinach, J. (2007). Supervised regionalization methods: a survey. *International Regional Science Review*, 30:195–220.
- Fotheringham, S., Brunson, C., and Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage Publications, London.

- Getis, A. and Ord, J. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24:189–206.
- Haddad, M. and Nedovic-Budic, Z. (2003). Reducing intra-urban inequalities: Spatial analysis of public intervention. Research Paper, Iowa State University.
- Hamilton, L. C. (1992). *Regression with Graphics: A Second Course in Applied Statistics*, chapter 8. Wadsworth Publishing Company.
- Knox, G. (1989). *Methodology of Enquiries into Disease Clustering*, chapter Detection of Clusters. Small Area Health Statistics Unit, London.
- Openshaw, S. and Taylor, P. (1981). *Quantitative Geography: A British View*, chapter The modifiable area unit problem, pages 60–69. London: Routledge and Regan Paul.
- Petrucci, A., Salvati, N., and Seghieri, C. (2003). *The application of a spatial regression model to the analysis and mapping of poverty*. Food and Agriculture Organization of the United Nations.
- Pérez, G. J. (2005). Dimensión espacial de la pobreza en Colombia. *Ensayos sobre política económica*, pages 2–51.
- Rosero-Bixby, L. (2004). La fecundidad de reemplazo y más allá en áreas metropolitanas de América Latina. *Notas de Población*, 31(78):35–63.
- Tobler, W. (2004). On the first law of geography: A reply. *Annals of the Association of American Geographers*, 94(2):304–310.
- United Nations Human Settlements Programme (2003). *The challenge of slums: Global report on human settlements 2003*, chapter 1. Earthscan Publications Ltd.
- Vayá, E. and Moreno, R. (2001). *Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial*, chapter 2. Edicions Universitat de Barcelona.
- Weeks, J. R. (2007). *Population: An Introduction to Concepts and Issues*. Ed. 10. Cengage Learning.
- Weeks, J. R., Hill, A., Stow, D., Getis, A., and Fugate, D. (2006). The impact of neighborhood structure on health inequalities in Accra, Ghana. In *Population association of America 2006 annual meeting program*, pages 1–34.
- Weeks, J. R., Hill, A., Stow, D., Getis, A., and Fugate, D. (2007). Can we spot a neighborhood from the air? Defining neighborhood structure in Accra, Ghana. *GeoJournal*, 69(1-2):9–22.

8 Anexos

Figura 11: Flujo de información a nivel de personas, hogares y barrios



NOTA: A partir de la Encuesta de Calidad de Vida 2007 para Medellín se creó el archivo varslum.do. Este contiene las cinco variables que según UN-HABITAT hacen que un lugar se catalogue como un *slum*.

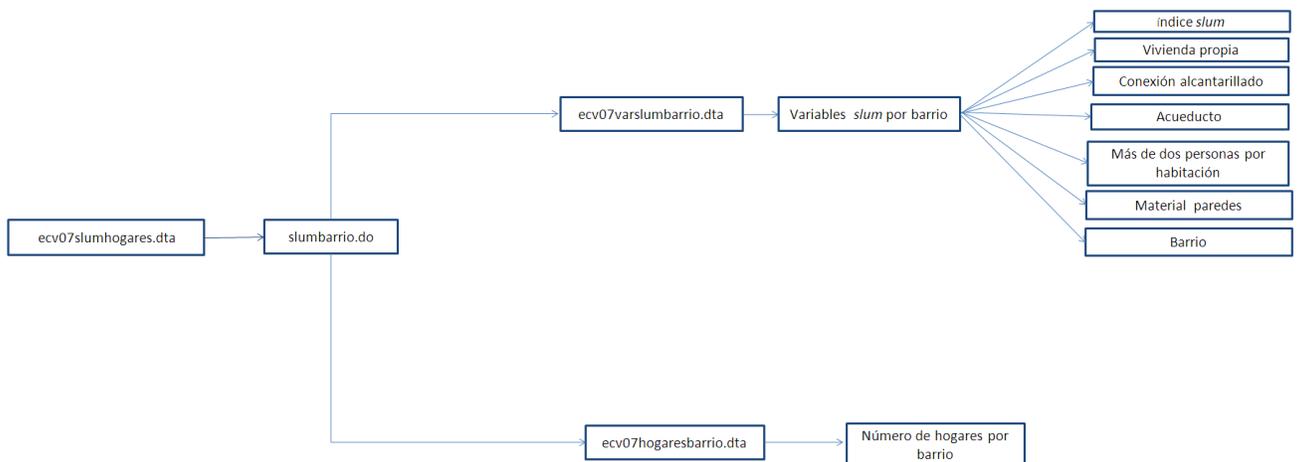


Figura 12: Flujo de información de las variables socioeconómicas de la ECV Medellín 2007

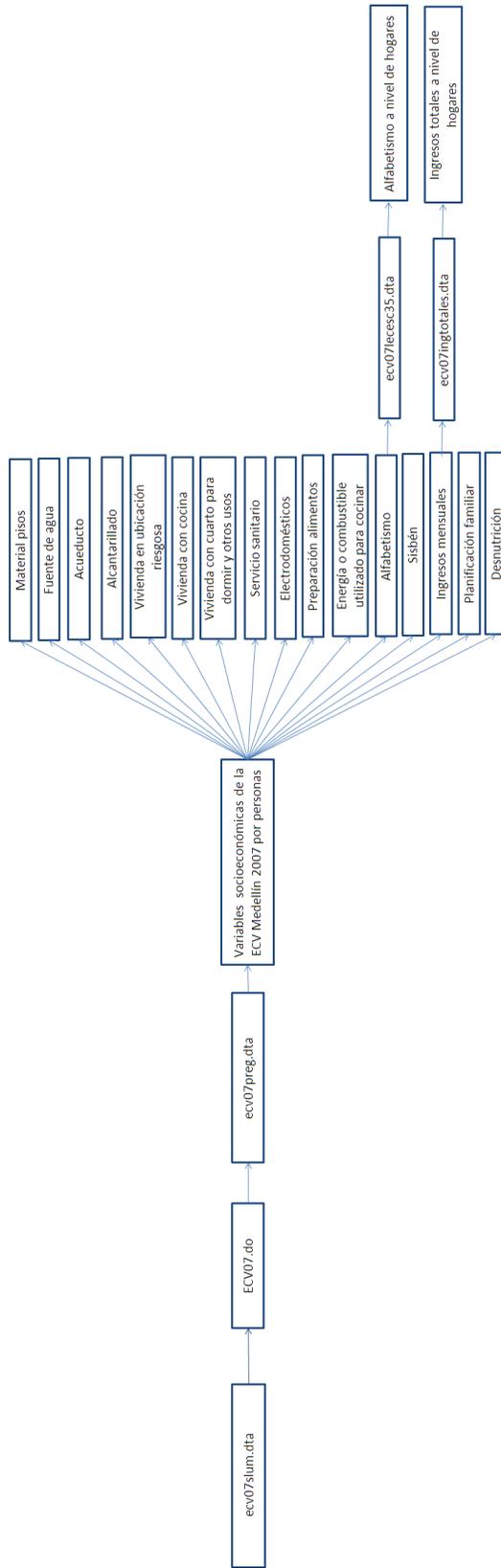


Figura 13: Flujo de información de las variables socioeconómicas de la ECV Medellín 2007

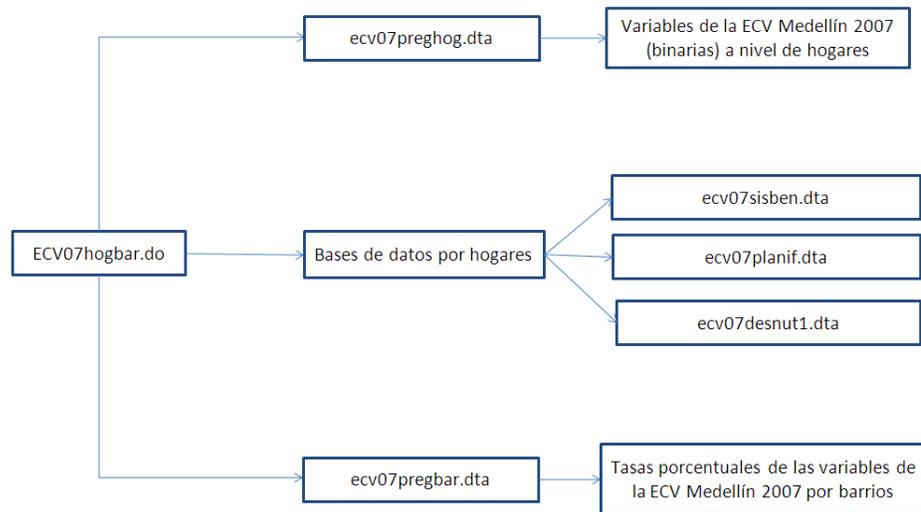


Figura 14: Flujo de información de las regiones analíticas

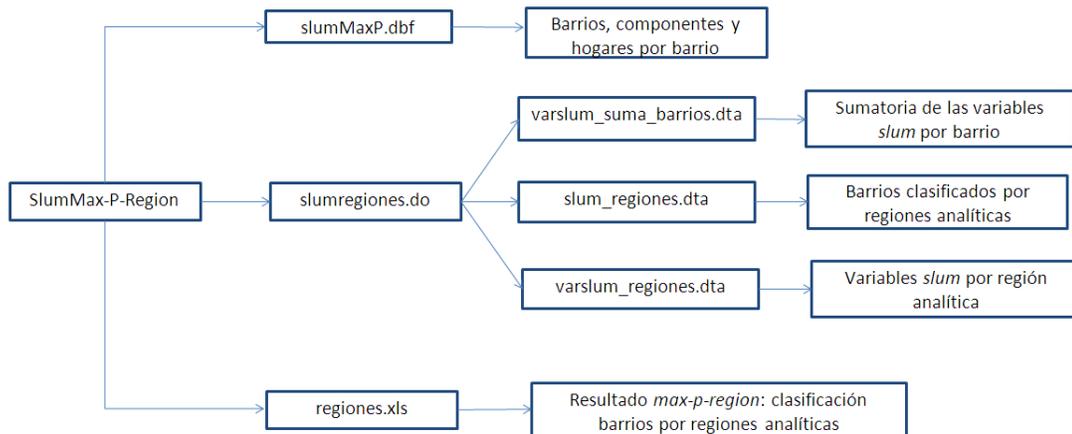


Figura 15: Flujo de información georreferenciada

