



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA DE SISTEMAS – INVESTIGACIÓN DE OPERACIONES

**“DETERMINACIÓN DE CONTRASTES DE REZAGO SOCIAL EN LA ZONA
METROPOLITANA DEL VALLE DE MÉXICO MEDIANTE EL ÍNDICE DE
MORAN.”**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN INGENIERÍA

PRESENTA:

MARTHA PATRICIA OLIVARES TREJO

TESIS DIRIGIDA POR:

M.I. JOSÉ ANTONIO RIVERA COLMENERO
Facultad de Ingeniería

MÉXICO, D.F. AGOSTO 2014



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

Presidente: **Dr. Aceves García Ricardo**

Secretario: **M.I. Fuentes Zenón Arturo**

Vocal: **M.I. Rivera Colmenero José Antonio**

1er. Suplente: **Dra. Elizondo Cortés Mayra**

2do. Suplente: **Dra. Flores De La Mota Idalia**

Lugar o lugares donde se realizó la tesis:

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

CIUDAD UNIVERSITARIA, MÉXICO, D.F.

TUTOR DE TESIS:

M.I. RIVERA COLMENERO JOSÉ ANTONIO

FIRMA



Agradecimientos:

A Dios doy gracias por darme tantas cosas y que hoy me permiten dar un paso más en mi vida.

A mi *alma mater* que me ha cobijado en su seno de la sabiduría y del conocimiento que me permite ser más crítico, científico y mejor persona.

Un agradecimiento especial para el CONACYT por el apoyo que recibí en mi incursión en el posgrado.

A mis profesores, compañeros y amigos, mis sinceras gracias porque sin sus consejos, apoyo y presencia, no hubiera alcanzado este sueño al que todos aspiramos en algún momento de nuestras vidas.

Al Jurado Asignado mi agradecimiento por su apoyo, paciencia y tolerancia.

A mi familia, Coko y Vicente gracias por el apoyo incondicional y por que siempre se mantuvieron pacientes a mi lado a lo largo de todo este tiempo para alcanzar este objetivo.

A mi madre dedico este trabajo, agradeciéndole siempre su paciencia, amor y ejemplo de vida.

ÍNDICE

RESUMEN	8
ABSTRACT	9
INTRODUCCIÓN	10
Capítulo I. Índice de Rezago Social en México	14
1.1 Índice de Rezago Social del CONEVAL	14
1.2 Dimensiones del rezago social	15
1.2.1 Educación	15
1.2.2 Acceso al servicio de salud	16
1.2.3 Servicios básicos en la vivienda	16
1.2.3.1 Activos en el hogar	17
1.2.4 Calidad y espacios en la vivienda	17
1.3 Zonas Metropolitanas de México	18
1.4 Zona Metropolitana del Valle de México	20
1.4.1 Municipios metropolitanos y criterios de delimitación	20
Capítulo 2. Econometría espacial	25
2.1 Efectos espaciales	27
2.1.1 Heterogeneidad espacial	27

2.1.2 Autocorrelación espacial.....	28
2.2 La matriz de pesos espaciales	30
2.2.1 Criterios de contigüidad	32
2.2.2 Otras especificaciones para la matriz de contactos.....	34
Capítulo III. Análisis exploratorio de datos espaciales (AEDE)	37
3.1. Índices de autocorrelación espacial globales y locales.	38
3.2. Índice Global de Moran	40
3.3 Índice Local de Asociación Espacial (LISA)	43
3.4. Técnicas gráficas para la representación de resultados	45
3.4.1 Aproximación geoestadística	45
3.4.2 Aproximación lattice.....	46
3.4.2.1 Scatterplot de Moran.....	46
3.4.2.2 Los mapas LISA.....	50
Capítulo IV. Análisis de Contrastes de Rezago Social en la Zona Metropolitana del Valle de México.....	53
4.1 Análisis municipal.....	53
4.2. Resultados de autocorrelación global	57
4.3. Resultados de autocorrelación local.....	62
Conclusiones.....	66
Bibliografía.	67

Anexo A. Técnico Metodológico	72
A. El Índice de Rezago Social	72
Anexo B. Ejemplo del cálculo de la autocorrelación espacial.....	81
B. Autocorrelación espacial en las delegaciones del Distrito Federal, de lo local a lo global.....	81

RESUMEN

Con base en el índice de Rezago Social (IRS), se realiza un análisis espacial en la Zona Metropolitana del Valle de México (ZMVM), con el propósito de identificar la existencia de patrones de conglomeración geográfica en los municipios con rezago social entre 2005 y 2010.

La metodología utilizada es el Análisis Exploratorio de Datos Espaciales (AEDE) orientado a comprobar como hipótesis que *el Índice de Rezago Social (IRS) NO se distribuye aleatoriamente en el espacio geográfico de la Zona Metropolitana del Valle de México (ZMVM) a nivel municipal*, y que tiende a la concentración de valores altos y bajos formando conglomerados espaciales. Para confirmar esta hipótesis se aplicaron técnicas de análisis geoespacial (dependencia o autocorrelación) como el Índice Global de Moran y el Índice de Asociación Espacial Local (LISA).

Para la identificación y localización del IRS en la ZMVM se utilizaron las bases de datos generadas de los siguientes proyectos: “Principales Resultados por Localidad, 2005” del II Censo de Población y Vivienda (ITER 2005) para el Índice de Rezago Social 2005 y el Censo de Población y Vivienda 2010 para el Índice de Rezago Social 2010, realizado por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL).

Los mapas de rezago social exhibidos en el presente trabajo, se realizaron con las capas disponibles del Instituto Nacional de Estadística y Geografía (INEGI).

Palabras clave: Autocorrelación Espacial, Análisis Exploratorio de Datos Espaciales, I de Moran, Rezago Social.

ABSTRACT

Based on the Social Gap Index (IRS), a spatial analysis is performed in the Metropolitan Area of the Valley of Mexico (ZMVM), in order to identify the existence of patterns of geographic clustering in municipalities with social gap between 2005 and 2010.

The methodology used is Exploratory Spatial Data Analysis (ESDA) aimed at checking the hypothesis that the Social Gap Index (IRS) is randomly distributed in geographic space of the Metropolitan Area of the Valley of Mexico (ZMVM) at the municipal level, and tends to the concentration of high and low values to form spatial clusters. To confirm this hypothesis geospatial analysis techniques (dependency or autocorrelation) as Moran Index and Global Index Local Spatial Association (LISA) were applied.

For the identification and location of the IRS in the ZMVM databases generated from the following projects were used: "Main Results by Locality, 2005" II Population and Housing (ITER 2005) for the Social Gap Index 2005 and census of Population and Housing 2010 for the Social Gap Index 2010, conducted by the National Council for the Evaluation of Social Development Policy (CONEVAL).

Social Gap maps displayed in this paper were performed with the layers available from the National Institute of Statistics and Geography (INEGI).

Keywords: Spatial Autocorrelation, Exploratory Spatial Data Analysis, Moran's I, Social Gap.

INTRODUCCIÓN

En los estudios sobre economía regional, durante los últimos años se han venido desarrollando una serie de herramientas y métodos de análisis espacial, las cuales permiten evaluar la existencia de relaciones socioeconómicas entre sus factores y, a la vez, la relación de estos con su ubicación geográfica. Concretamente, el propósito es establecer la existencia o no de dependencia espacial.

La idea de tomar en cuenta el entorno desde un punto de vista geográfico era comúnmente utilizada en los años sesenta y setenta. En esos años, los científicos sociales prestaron especial atención a la forma en que el espacio geográfico y el contexto social influían en las condiciones de las personas (Massey, 1990; Massey y Eggers, 1990; Bronfenbrenner, 1986).

Recientemente estos enfoques teóricos comúnmente usados en décadas anteriores aparecen con mayor solidez empírica. De este modo “los efectos que el ambiente físico y social tienen sobre el bienestar y el comportamiento de los individuos han adquirido mayor importancia entre los investigadores y tomadores de decisiones” (Sastry, Pebley y Zonta, 2002, p. 2).

Según Weeks et al. (2002), el punto de partida que justifica este enfoque radica en que los seres humanos, como sujetos vivientes sociales, responden a las acciones de los individuos que viven al rededor y sus patrones de vida se articulan de acuerdo con el lugar donde viven: con quién viven, cómo viven y con quién interactúan.

Inicialmente, la configuración del entorno social y las diferencias en los niveles de rezago se construyó a partir de los estilos de vida, niveles de consumo, ingresos y tipo de infraestructura que presentaban los hogares en zonas urbanas, en comparación con las zonas rurales. (Clark, Davis y Johnston, 1974).

Hoy en día, las mediciones se están enfocando en características específicas del ambiente donde los individuos se desarrollan, por ejemplo: el desarrollo

cognoscitivo y psicosocial, la condición de salud, los logros educativos, la dinámica del control social y la conducta socialmente aceptada, las oportunidades de empleo y la posibilidad de los hogares de acumular capital social y capital humano (Brooks–Gunn, Duncan y Aber, 1997; Aber, Gephart, Brooks–Gunn y Connell, 1997; Brooks–Gunn, Duncan, Kato y Sealander, 1993).

Los investigadores en geografía humana y segregación social, defienden el análisis por niveles, individuo-hogar-entorno, para explicar los resultados de la pobreza y la exclusión. Las mediciones en estos niveles múltiples han generado diferentes aproximaciones del estatus socioeconómico, la estratificación social, el nivel educativo, el acceso a bienes y servicios básicos; pero además se han validado de nuevas técnicas para aproximar la exposición de los individuos al peligro, el crimen y la violencia, el grado de seguridad ciudadana, la distancia física y el aislamiento, la formación de las redes sociales y el control social. (Wilson, 1996; Sampson, Raudenbush y Earls, 1997; Gephart, 1997; Massey, 1998).

Los sistemas de información geográfica unen este nuevo compendio de mediciones con la información del lugar de residencia de los individuos, de modo que en la actualidad es común encontrar mapas de carencia básica o de rezago social.

En materia de software, la aplicación más desarrollada en el campo de la econometría espacial es *GeoDa*® Center for Geospatial Analysis and Computation¹; programa libre desarrollado por el Laboratorio de Análisis Espacial de la Universidad de Illinois, que realiza análisis exploratorio de datos espaciales, utiliza Sistemas de Información Geográfica y los mapas de este tipo de programas como *input*. El programa permite efectuar todas las etapas de un análisis empírico

¹ El software *GeoDa* se puede descargar directamente desde la Web del Centro para el Análisis Geoespacial y Computacional, <https://geodacenter.asu.edu/>

de datos espaciales; representación cartográfica, estudio analítico y gráfico de la autocorrelación espacial, análisis exploratorio multivariante y regresión espacial.

Cabe destacar el esfuerzo del profesor Roger Bivand de la *Norwegian School of Economics and Business* que ha implementado un grupo importante de funciones en el lenguaje de programación R, que abarca la construcción de diversas formas de matrices espaciales, autocorrelación espacial y un grupo muy extenso de contrastes y métodos de estimación de modelos econométricos espaciales. Otro recurso interesante, y también de libre acceso lo constituye el programa de estadística espacial *CrimeStat*®, aunque enfocado a la criminología, permite análisis geográficos, epidemiológicos, botánicos o geológicos. Cubre la mayoría de tópicos de la estadística espacial más descriptiva y es un programa pensado para su utilización conjunta con *ArcView*®, *ArcGis*® o *MapInfo*®.

A pesar de esta aparente mayor difusión de la econometría espacial, la distancia que la separa, en términos de su conocimiento y difusión, de la econometría clásica todavía es notable (Anselin y Griffith, 1988; Anselin y Hudak, 1992). Este hecho es incuestionable a nivel teórico dado que, en la actualidad, la mayoría de los textos de econometría y estadística no hacen referencia al problema de los efectos espaciales y, en caso de hacerlo, se limitan a un breve apunte.

El objetivo de esta investigación es proponer el análisis espacial y específicamente el uso de los sistemas de información geográfica (SIG) como herramienta para la detección de conglomerados geográficos para la ubicación de zonas específicas de rezago social y otros fenómenos de exclusión social. El método, además de ubicar geográficamente zonas de rezago social, permite asociar el espacio físico con la conducta de los individuos, sus condiciones y estilos de vida.

Cabe señalar que el presente trabajo sigue las propuestas y herramientas desarrolladas por Luc Anselin (1995, 2004) en GeoDa², quien continúa en la

² <https://geodacenter.asu.edu/>

construcción, mejora y difusión de algoritmos para el análisis geográfico, como el Índice Global y el Índice Local de Asociación Espacial (LISA).

La tesis se estructura en cuatro capítulos. En el primer capítulo se describe el Índice de Rezago Social (IRS) del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) a nivel municipal en la Zona Metropolitana del Valle de México (ZMVM).

La metodología del Análisis Exploratorio de Datos Espaciales (AEDE), así como las principales técnicas de autocorrelación espacial: Índice Global de Moran e Índice de Asociación Espacial Local (LISA) se presenta en el segundo capítulo.

El tercer capítulo, expone la existencia de la econometría espacial para abordar los efectos espaciales; la heterogeneidad y la autocorrelación espacial, analizando a su vez la matriz de pesos espaciales como instrumento básico para reflejar las interdependencias en el espacio.

En el capítulo 4 se presentan y comentan los resultados de las estimaciones realizadas a nivel global y local mediante el I de Moran para la Zona Metropolitana del Valle de México (ZMVM).

Finalmente, se presentan las principales conclusiones del trabajo de tesis.

Capítulo I. Índice de Rezago Social en México

El rezago social en México es fenómeno no menos importante que la pobreza, su objetivo es establecer criterios para la determinación de las Zonas de Atención Prioritaria, tal como lo dispone la Ley General de Desarrollo Social (LGDS).

En 2006, el Consejo Nacional para la Evaluación de la Política de Desarrollo Social (CONEVAL) dio a conocer el Índice de Rezago Social. Para darle un seguimiento periódico, se realiza un informe de su comportamiento cada cinco años a nivel estatal, municipal y por localidad.

Hoy día, el CONEVAL cuenta con las bases de los índices de rezago social 2000, 2005 y 2010.

1.1 Índice de Rezago Social del CONEVAL

El Índice de Rezago Social, es una medida ponderada que resume cuatro dimensiones de carencias sociales: educación, salud, servicios básicos y espacios en la vivienda en un solo índice, que tiene como finalidad ordenar a las unidades de observación según sus carencias sociales.

La estimación de este Índice tiene como fuente de información: la base de datos “Principales Resultados por Localidad, 2005” del II Censo de Población y Vivienda (ITER 2005) para el índice de Rezago Social 2005 y el Censo de Población y Vivienda 2010 para el índice de Rezago Social 2010.

La técnica utilizada para la elaboración de este índice es la de componentes principales, que permite resumir en un indicador agregado las diferentes dimensiones del fenómeno en estudio. El rezago social se calculó a tres niveles de agregación geográfica: estatal, municipal y localidad.

Además, esta técnica es la misma que utiliza el Consejo Nacional de Población (CONAPO) para la construcción del índice de marginación y que ha sido ampliamente utilizado.

Los resultados de la estimación del índice de rezago social se presentan en cinco estratos. Se utiliza la estratificación con base en la metodología de *Dalenius & Hodges*, dado que permite que dentro de cada estrato las unidades sean lo más homogéneas posible y entre los estratos lo más distintas posible. Los cinco estratos en que se distribuye el índice son: muy bajo, bajo, medio, alto y muy alto rezago social.

1.2 Dimensiones del rezago social

Como se mencionó anteriormente, el CONEVAL considera cuatro dimensiones de carencia social: educación, salud, servicios básicos y espacios en la vivienda. A continuación se abordan dichas dimensiones y los indicadores socioeconómicos empleados para su medición.

1.2.1 Educación

Esta dimensión se integra por cuatro indicadores. La educación en México es un derecho consagrado por la constitución vigente, que en su artículo 3° declara que la educación impartida por el Estado debe ser gratuita, laica y obligatoria para todos los habitantes del país. La educación es uno de los pilares más importantes de toda cultura, da la capacidad de leer y escribir un recado; el no poder realizar estas acciones limita las posibilidades de comunicación de la gente, literalmente al ámbito oral, y prácticamente trunca toda posibilidad de adquirir conocimientos tanto en el sistema educativo ortodoxo, como de manera autodidacta.

En la educación se transmiten los valores, principios, actitudes, gustos, sentimientos que le dan “personalidad” a cada pueblo. Sus indicadores socioeconómicos son:

-
- a) Porcentaje de población de 15 años o más analfabeta;
 - b) Porcentaje de la población de 6 a 14 años que no asiste a la escuela;
 - c) Porcentaje de los hogares con población de 15 a 29 años, con al menos un habitante con menos de 9 años de educación aprobados;³
 - d) Porcentaje de la población de 15 años o más con educación básica incompleta.⁴

1.2.2 Acceso al servicio de salud.

- a) Porcentaje de la población sin derechohabiencia a servicios de salud.

1.2.3 Servicios básicos en la vivienda

La vivienda es el único espacio físico constante durante las etapas de la vida de los individuos, desde la infancia hasta la edad adulta en plenitud, por tanto, es determinante para el desarrollo de capacidades, habilidades, madurez emocional y conocimientos de toda persona. En consecuencia, explorar las condiciones de las viviendas resulta esencial al tratar el rezago social.

Esta dimensión se integra por cuatro indicadores. Sus indicadores socioeconómicos son:

- Porcentaje de las viviendas particulares habitadas que no disponen de excusado o sanitario;
- Porcentaje de las viviendas particulares habitadas que no disponen de agua entubada de la red pública;

³ Este indicador de rezago educativo se empleó para los niveles estatal y municipal.

⁴ Este indicador se empleó como indicador de rezago educativo en los niveles de localidad.

-
- Porcentaje de las viviendas particulares habitadas que no disponen de drenaje;
 - Porcentaje de las viviendas particulares habitadas que no disponen de energía eléctrica.

1.2.3.1 Activos en el hogar

La disponibilidad de refrigerador y lavadora se encuentran condicionadas por el ingreso del que se dispone en las viviendas, ya sea por trabajo o transferencias monetarias o en especie; aún más, el no tener refrigerador limita las posibilidades de contar con alimentos perecederos frescos e incrementa los riesgos de salud asociados con la ingesta de alimentos con algún grado de descomposición y con una dieta deficiente.

En virtud de lo anterior, se considera los siguientes dos indicadores socioeconómicos:

- e) Porcentaje de las viviendas particulares habitadas que no disponen de lavadora;
- f) Porcentaje de las viviendas particulares habitadas que no disponen de refrigerador.

1.2.4 Calidad y espacios en la vivienda

Los dos indicadores socioeconómicos considerados en la dimensión de espacio de vivienda son:

- a) Porcentaje de las viviendas particulares habitadas con piso de tierra.
- b) Promedio de ocupantes por cuarto.

1.3 Zonas Metropolitanas de México

En el año 2004, la Secretaría de Desarrollo Social (Sedesol), el Consejo Nacional de Población (CONAPO) y el Instituto Nacional de Estadística, Geografía e Informática (INEGI)⁵ publicaron el libro “Delimitación de las zonas metropolitanas de México”, resultado de un trabajo conjunto en torno a la definición de las zonas metropolitanas del país y su delimitación territorial, mediante el establecimiento y aplicación de criterios operativos para identificar las unidades que las integran, a partir de la cartografía geoestadística y los datos del XII Censo General de Población y Vivienda 2000.

Dada su importancia, en 2007 se actualizó la delimitación de las zonas metropolitanas a partir de los resultados del III Censo de Población y Vivienda 2005, generándose una nueva publicación.

Desde el ejercicio 2005, han ocurrido cambios estructurales que justifican un actualización de la delimitación, como son el crecimiento demográfico de la mayoría de las zonas metropolitanas, la construcción de grandes desarrollos habitacionales en la periferia que han inducido la expansión urbana en municipios incorporados al universo metropolitano, el levantamiento del Censo de Población y Vivienda 2010, la aparición de temas emergentes como el crecimiento verde y el impacto del cambio climático, el consenso sobre el papel creciente de las zonas metropolitanas como motores del desarrollo económico y social de los países, y la consolidación del *Fondo Metropolitano* como mecanismo de financiamiento.

La última delimitación de las zonas metropolitanas se realizó con base en la cartografía geoestadística, de los resultados definitivos del Censo de Población y Vivienda 2010, y en los programas publicados de ordenación y desarrollo urbano de zonas conurbadas y zonas metropolitanas correspondientes.

⁵ Actualmente Instituto Nacional de Estadística y Geografía.

En México, los resultados de la delimitación de las zonas metropolitanas de México 2010, dan cuenta de la existencia de 59 zonas metropolitanas integradas por 367 municipios, de los cuales 263 son municipios centrales, 43 municipios exteriores definidos con base en criterios estadísticos y geográficos, y, finalmente, 61 municipios exteriores definidos con base en criterios de planeación y política urbana. Estas 59 zonas metropolitanas representan 56.8 por ciento de la población nacional, con 63.8 millones de habitantes. Conviene resaltar que la población urbana del país es eminentemente metropolitana, tal y como se vislumbraba hace algunos años.

El fenómeno metropolitano se ha intensificado y dispersado en las últimas décadas, tal y como se observa en los resultados obtenidos de estas investigaciones, las cuales reconocen un número reciente de zonas metropolitanas y de municipios que las conforman, aun cuando los resultados no son estrictamente comparables en el tiempo.

Cuadro 3. Indicadores del proceso de metropolización, 1960 -2010

Indicador	1960	1980	1990	2000	2005	2010
Zonas metropolitanas	12	26	37	55	56	59
Delegaciones y municipio metropolitanos	64	131	155	309	345	367
Entidades federativas	14	20	26	29	29	29
Población total (millones)	9.0	26.1	31.5	51.5	57.9	63.8
Porcentaje de la población nacional	25.6	39.1	38.8	52.8	56.0	56.8

Nota: Los datos no son estrictamente comparables.

Fuente: Para 1960, Unikel, et al. (1978), El desarrollo urbano de México, México, El Colegio de México.

Para 1980, Negrete y Salazar (1986), "Zonas metropolitanas en México", Estudios Demográficos y Urbanos, vol. I, núm. 1.

Para 1990, Sobrino (1993), Gobierno y administración metropolitana y regional, México, INAP.

Para 2000, S EDESOL, C ONAPO e I NEGI (2004), Delimitación de las zonas metropolitanas de México, México.

Para 2005, S EDESOL, C ONAPO e I NEGI (2007), Delimitación de las zonas metropolitanas de México, México.

Para 2010, resultados de la presente publicación.

1.4 Zona Metropolitana del Valle de México.

Se define como zona metropolitana al conjunto de dos o más municipios donde se localiza una ciudad con 50 mil o más habitantes, cuya área urbana, funciones y actividades rebasan el límite del municipio que originalmente la contenía, incorporando como parte de sí misma o de su área de influencia directa a municipios vecinos, predominantemente urbanos, con los que mantiene un alto grado de integración socioeconómica. También se incluyen a aquellos municipios que por sus características particulares son relevantes para la planeación y política urbanas de las zonas metropolitanas en cuestión.

Adicionalmente, se define como zonas metropolitanas a todos aquellos municipios que contienen una ciudad de un millón o más habitantes, así como aquellos con ciudades de 250 mil o más habitantes que comparten procesos de conurbación con ciudades de Estados Unidos de América.

1.4.1 Municipios metropolitanos y criterios de delimitación

Para la delimitación de las zonas metropolitanas se definieron a su vez tres grupos de municipios metropolitanos, con sus respectivos criterios:

- a) *Municipios centrales:* corresponden a los municipios donde se localiza la ciudad central que da origen a la zona metropolitana;
- b) *Municipios exteriores definidos con base en criterios estadísticos y geográficos:* son municipios contiguos a los anteriores, cuyas localidades geoestadísticas urbanas no están conurbadas a la ciudad central, pero que manifiestan un carácter predominantemente urbano, al tiempo que mantienen un alto grado de integración funcional con los municipios centrales de la zona metropolitana;
- c) *Municipios exteriores definidos con base en criterios de planeación y política urbana:* son municipios que se encuentran reconocidos por los

gobiernos, federal y locales, como parte de una zona metropolitana, a través de una serie de instrumentos que regulan su desarrollo urbano y la ordenación de su territorio, además de que presentan un carácter predominantemente urbano y mantienen un alto grado de integración funcional con los municipios centrales.

La **Zona Metropolitana del Valle de México** o por sus siglas ZMVM,⁶ es el área metropolitana constituida por las 16 delegaciones del Distrito Federal, 59 municipios del estado de México y 1 municipio del estado de Hidalgo. En las tablas 1, 2 y 3 se detalla las delegaciones y municipios que conforman la ZMVM.

De acuerdo con los resultados del censo elaborado por el INEGI en el año 2010 esta zona contaba con una población de poco más de 20 millones de habitantes, tan solo en el Distrito Federal son 8,851,080 habitantes.⁷

Tabla 1. Distrito Federal, 16 delegaciones.

Población Total: 8,851,080⁸	
Delegación	Población
Álvaro Obregón	727,034
Azcapotzalco	414,711
Benito Juárez	385,439
Coyoacán	620,416
Cuajimalpa de Morelos	186,391
Cuauhtémoc	531,831
Gustavo A. Madero	1,185,772
Iztacalco	384,326
Iztapalapa	1,815,786
La Magdalena Contreras	239,086

⁶ Consejo Nacional de Población. Delimitación de las zonas metropolitanas de México 2010.

⁷ INEGI. México en cifras 2010.

⁸ INEGI (2010). Instituto Nacional de Estadística y Geografía, [Censo de población y vivienda 2010 - Estados Unidos Mexicanos Resultados Definitivos por Entidad y Municipio](#) (HTML).

Población Total: 8,851,080⁸	
Delegación	Población
Miguel Hidalgo	372,889
Milpa Alta	130,582
Tláhuac	360,265
Tlalpan	650,567
Venustiano Carranza	430,978
Xochimilco	415,007

Tabla 2. Estado de Hidalgo, 1 municipio.

Población Total: 97,461⁹	
Municipio	Población
Tizayuca	97,461

Tabla 3. Estado de México, 59 Municipios.

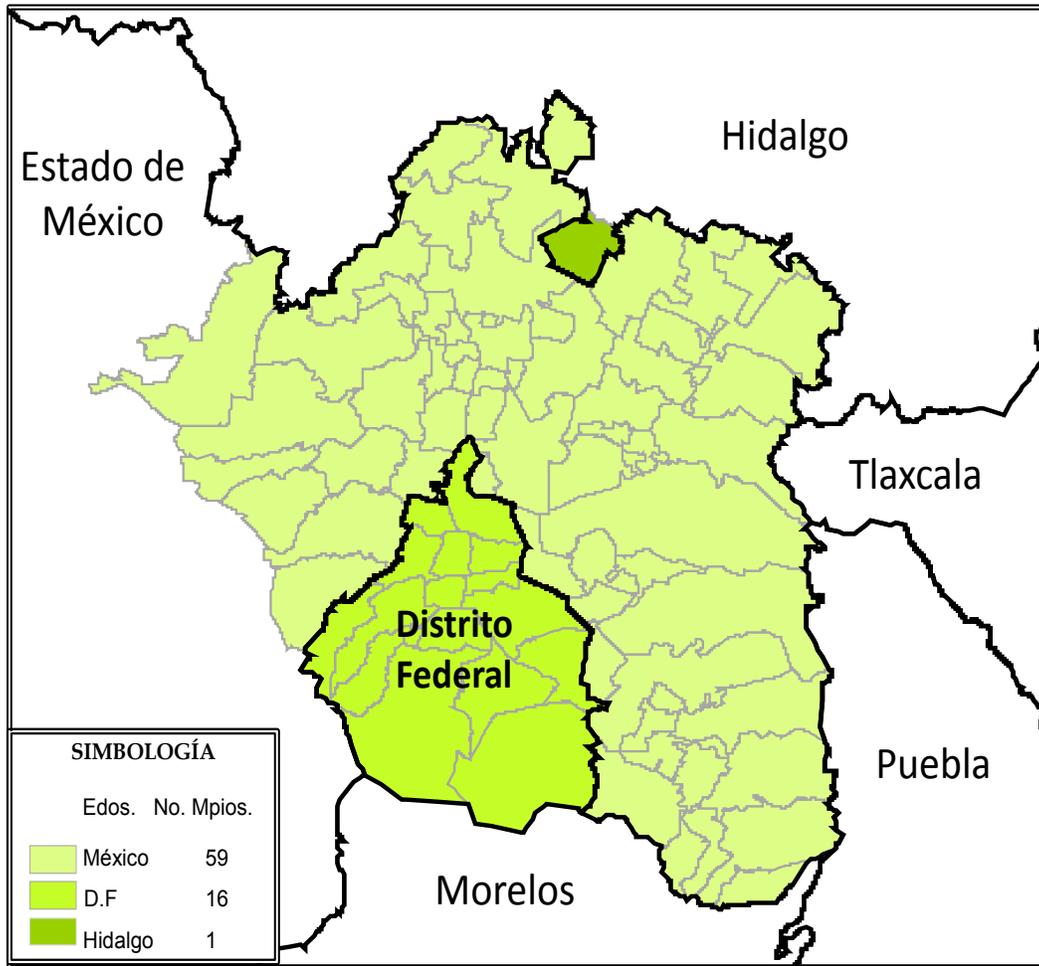
Población Total: 11.158.085¹⁰			
Municipio	Población	Municipio	Población
Acolman	136,558	Naucalpan de Juárez	833,779
Amecameca	48,421	Nextlalpan	34,374
Apaxco	27,521	Nezahualcóyotl	1,110,565
Atenco	56,243	Nicolás Romero	366,602
Atizapán de Zaragoza	489,937	Nopaltepec	8,895
Atlautla	27,663	Otumba	34,232
Axapusco	25,559	Ozumba	27,207
Ayapango	8,864	Papalotla	4,147
Chalco	310,130	San Martín de las Pirámides	24,851
Chiautla	26,191	Tecámac	364,579
Chicoloapan	175,053	Temamatla	11,206
Chiconcuac	22,819	Temascalapa	35,987
Chimalhuacán	614,453	Tenango del Aire	10,578
Coacalco de Berriozábal	278,064	Teoloyucan	63,115

⁹ INEGI (2010). Instituto Nacional de Estadística y Geografía, [Censo de población y vivienda 2010 - Estados Unidos Mexicanos Resultados Definitivos por Entidad y Municipio](#) (HTML).

¹⁰ INEGI (2010). Instituto Nacional de Estadística y Geografía, [Censo de población y vivienda 2010 - Estados Unidos Mexicanos Resultados Definitivos por Entidad y Municipio](#) (HTML).

Población Total: 11.158.085¹⁰			
Municipio	Población	Municipio	Población
Cocotitlán	12,142	Teotihuacán	53,010
Coyotepec	39,030	Tepetlaoxtoc	27,944
Cuautitlán	140,059	Tepetlixpa	18,327
Cuautitlán Izcalli	511,675	Tepotzotlán	88,559
Ecatepec de Morelos	1,656,107	Tequixquiac	33,907
Ecatzingo	9,369	Texcoco	235,151
Huehuetoca	100,023	Tezoyuca	35,199
Hueypoxtla	39,864	Tlalmanalco	46,130
Huixquilucan	242,167	Tlalnepantla de Baz	664,225
Isidro Fabela	10,308	Tonanitla	10,216
Ixtapaluca	467,361	Tultepec	91,808
Jaltenco	26,328	Tultitlán	524,074
Jilotzingo	17,970	Valle de Chalco Solidaridad	357,645
Juchitepec	23,497	Villa del Carbón	44,881
La Paz	253,845	Zumpango	159,647
Melchor Ocampo	50,240		

Mapa 1. Zona Metropolitana del Valle de México.



Fuente: Elaboración Propia

Capítulo 2. Econometría espacial

La econometría espacial es la parte de la econometría que se dedica al estudio de los *fenómenos económicos espaciales*. Si bien todas las actividades económicas se desarrollan en un espacio determinado, los fenómenos económicos espaciales pueden definirse como aquellos en los que la variable espacial, entendida en sentido amplio y definida por alguna matriz de contactos, juega un rol tan importante que su exclusión podría dar lugar a modelos econométricos con severos errores de especificación.

El término econometría espacial fue acuñado por Jean Paelink al comienzo de la década de los setenta, para abordar la autocorrelación espacial en el término de perturbación de una regresión.

Anselin (1988), es probablemente la referencia más citada en los trabajos de econometría espacial, que la define como “la colección de técnicas que lidian con las peculiaridades causadas por el espacio en el análisis estadístico de los modelos de la ciencia regional.” Once años más tarde, Anselin (1999) extiende la definición diciendo que “la econometría espacial es una rama de la econometría que se preocupa del tratamiento adecuado de la interacción espacial (autocorrelación espacial) y la estructura espacial (heterogeneidad espacial) en modelos de regresión con datos de corte transversal.”

Para más detalle, Paelinck y Klaassen (1979) mencionan cinco *principios fundamentales* que rigen al análisis espacial:

1. *Interdependencia*: la dependencia recíproca entre las diferentes unidades de análisis debe ser debidamente incorporada.
2. *Asimetría*: refleja la idea de concentración y desconcentración gradual en diferentes áreas.
3. *Alotopía*: la causa de un fenómeno espacial debe buscarse en otro lugar.
4. *No linealidad*

-
5. *Topología*: Inclusión de variables de distancias entre dos localizaciones, coordenadas, densidades y otras.

De acuerdo con Corrado y Fingleton (2011), a pesar de que varios economistas se resisten a la econometría espacial, por considerar que las variables espaciales se adicionan al análisis sólo por mostrar significatividad estadística sin ninguna justificación teórica, los estudios espaciales dentro de la corriente principal de la economía han ido incorporando variables del tipo espacial. En particular, la econometría espacial ha proporcionado herramientas valiosas para el estudio de las externalidades en forma de spillovers (efectos indirectos) espaciales.

Existe una gran cantidad de ramas de la economía que han incorporado al análisis la econometría espacial: la economía urbana, la economía regional y la macroeconomía entre otras. En este sentido, la teoría de los centros de desarrollo constituye uno de los principales fundamentos para el análisis espacial de los fenómenos económicos. En este sentido, los aportes teóricos más significativos han provenido desde la Teoría del Desarrollo Económico y de la denominada Teoría de la Organización Espacial (Posada, 1978).

Habitualmente, en el estudio econométrico de fenómenos económicos espaciales suelen surgir, dos problemas principales, que son inherentes a la propia naturaleza de la estructura de datos a analizar: la heterogeneidad espacial y la autocorrelación espacial –también llamada dependencia espacial por algunos autores (Moreno y Vayá, 2000).

2.1 Efectos espaciales

Los efectos espaciales son la razón esencial para la existencia de un campo de estudio espacial, y se pueden agrupar en torno a dos grandes bloques: heterogeneidad espacial y dependencia espacial.

Los problemas causados por la heterogeneidad espacial y por la dependencia espacial, y sus efectos en la validez de los métodos estadísticos convencionales no es nuevo, se remonta al año 1914 con un trabajo de Student, donde por vez primera se hace referencia a este tipo de problemas. Sin embargo, el desarrollo o el nacimiento de un campo aparte no se produce hasta la década de los cincuenta con los trabajos de Moran (1948) y Geary (1954), donde se presentan los primeros índices formales para medir la autocorrelación o dependencia espacial.

2.1.1 Heterogeneidad espacial

La heterogeneidad espacial consiste en la variación de las relaciones en el espacio. En general, se puede decir que existen dos aspectos distintos de heterogeneidad espacial: *la inestabilidad estructural* y *la heteroscedasticidad*.

La inestabilidad estructural se refiere a la falta de estabilidad en el espacio del comportamiento de la variable bajo estudio, su forma funcional y los parámetros de una regresión puedan variar según la localización, siendo no homogéneas en toda la muestra. El segundo caso, *la heteroscedasticidad*, proviene de la omisión de variables u otras formas de errores de especificación que llevan a la aparición de errores de medida.

El tratamiento de la heterogeneidad espacial puede realizarse por medio de las técnicas econométricas estándar, entre otras, la consideración explícita de los parámetros variantes, coeficientes aleatorios (Hildreth y Houck, 1968) y switching regressions (Quandt, 1958) o las técnicas de filtraje adaptativo espacial (Foster y Gorr, 1983, 1984, 1986). Sin embargo, en el caso específico de la heterogeneidad espacial, en donde ésta viene causada por cuestiones inherentes a la localización

de las observaciones, el conocimiento teórico de la estructura espacial de los datos puede conducirnos a procedimientos más complejos pero más eficientes como la expansión espacial de parámetros (Casetti, 1972) o las regresiones ponderadas geográficamente (Fotheringham, 1998). A pesar de que dicho efecto puede tratarse en la mayoría de los casos mediante las técnicas econométricas tradicionales, el presente trabajo se centra en la problemática entorno a la autocorrelación espacial.

2.1.2 Autocorrelación espacial

La dependencia o autocorrelación espacial (AE) aparece como consecuencia de la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que ocurre en otro lugar (Cliff y Ord, 1973; Paelink y Klaassen, 1979; Anselin 1988). Es decir, el valor que toma una variable en una región no viene explicado únicamente por condicionantes internos sino también por el valor de esa misma variable en regiones vecinas, incumpléndose por tanto el supuesto de independencia entre las observaciones muestrales.

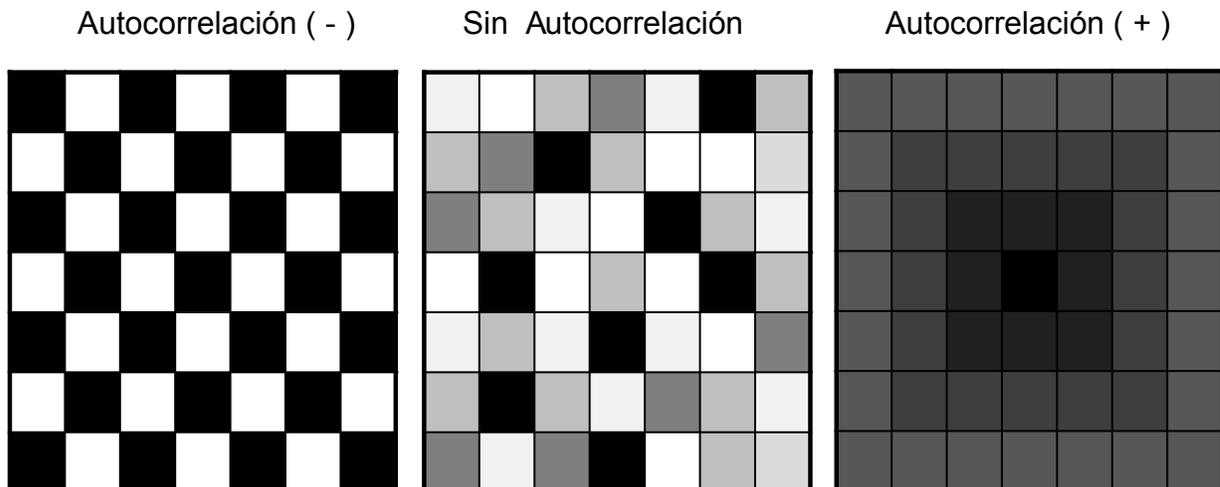
La autocorrelación espacial (AE) de una misma variable da lugar a una de estas tres posibilidades:

- *Autocorrelación espacial positiva*: la presencia de un fenómeno determinado en una región lleva a que se extienda ese mismo fenómeno hacia el resto de regiones que lo rodean, favoreciendo la concentración o agrupamiento del mismo, en las unidades espaciales. Por ejemplo, la variable crecimiento de la productividad muestra un esquema de dependencia espacial positiva, las regiones vecinas se verían contagiadas de una buena evolución de la productividad en dicha región. En este caso, se dice que existe un efecto de “contagio”, que favorece la concentración de un determinado fenómeno en una zona geográfica concreta.

- *Autocorrelación espacial negativa:* La presencia de un fenómeno en una región impide o dificulta su aparición en las regiones vecinas a ella (como ocurre en la disposición de las casillas blancas y negras en un tablero de ajedrez), es decir, cuando unidades geográficas cercanas sean netamente más disímiles entre ellas que entre regiones alejadas en el espacio. En este caso, se produce un efecto de “absorción” de un determinado espacio geográfico, que hace que el fenómeno estudiado se manifieste con mayor intensidad en dicho espacio y con mucha menor intensidad en los espacios geográficos próximos.
- *Sin autocorrelación:* no ocurre ninguna de las dos situaciones anteriores. Por lo tanto, los valores de las unidades espaciales vecinas presentan valores producidos en forma aleatoria.

En la figura 1. Se visualiza estos tres eventos.

Figura 1. Tipos de autocorrelación espacial



Fuente: Elaboración Propia

De esta manera el concepto de autocorrelación tiene sustento con *la primera ley de la geografía de Tobler (1970)* en la que se afirma: “*que en el análisis geográfico todo está relacionado con todo, pero las cosas cercanas están más relacionadas entre sí que las cosas lejanas*”.

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

A pesar de su creciente importancia en el marco del análisis exploratorio de datos espaciales, su uso es reducido en trabajos geográficos que continúan recurriendo a los coeficientes tradicionales de correlación y a la estadística descriptiva.

Cabe señalar, que las dos principales causas que pueden inducir a la aparición de AE son: la existencia de errores de medida y la propia interacción espacial de las unidades. En términos económicos, los efectos desbordamiento (spillovers) y de jerarquías espaciales.

2.2 La matriz de pesos espaciales

En el análisis de series temporales es usual utilizar un operador de rezago o retardos, para capturar la influencia de las observaciones pasadas en la dinámica de la serie y el valor de la realización contemporánea. Otra razón por la que resulta relevante tiene que ver con la posibilidad de realizar pronósticos. De esta manera, la influencia del retardo temporal es unidireccional: las realizaciones pasadas explican el presente, y harán lo propio con las futuras, conforme a una estructura dinámica. Ahora bien, en el marco del análisis espacial se establecen relaciones multidireccionales: una región puede no sólo estar afectada por otra región contigua a ella sino por otras muchas que la rodean, al igual que ella puede influir sobre aquéllas. Este hecho imposibilita la utilización del operador de retardo L presente en el contexto temporal, el cual recoge únicamente una relación unidireccional. La solución al problema de la multidireccionalidad en el contexto

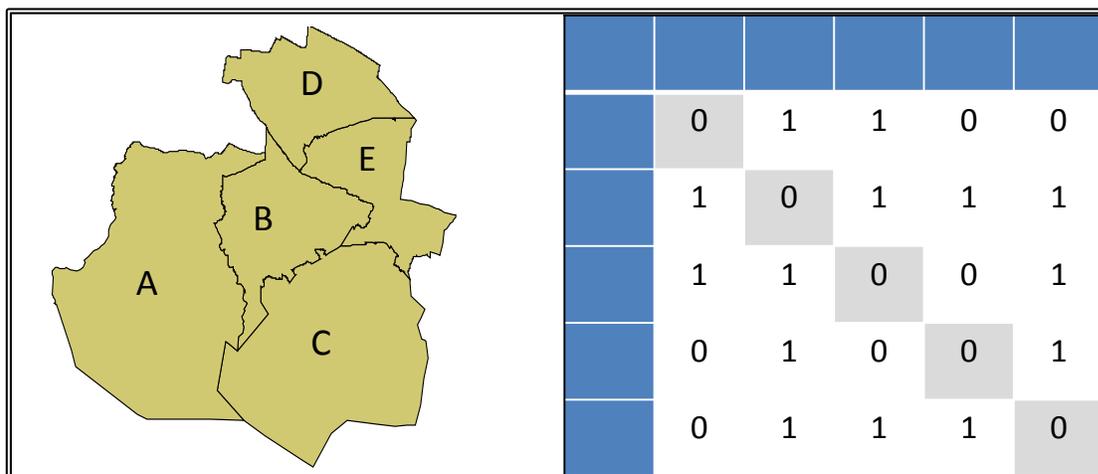
espacial pasa por la definición de la denominada matriz de pesos espaciales (también denominada matriz de retardos o de contactos o matriz de proximidad espacial) simbolizada con W , una matriz cuadrada de $N \times N$ (siendo N el número de unidades espaciales) no estocástica, compuesta por ceros y unos, cuyos elementos W_{ij} reflejan la intensidad de la interdependencia existente entre cada par de regiones i y j (Moreno y Vayá, 2000)

Con respecto a cómo definir los mencionados pesos, cabe destacar que no existe una definición unánimemente aceptada, si bien se ha de cumplir que dichos pesos sean no negativos y finitos (Anselin, 1980) A pesar de ello, de forma habitual se recurre al concepto de contigüidad física de primer orden, utilizado inicialmente por Moran (1948) y Geary (1954), donde:

$$W(i,j) \begin{cases} 1 & \text{si las regiones } i \text{ y } j \text{ son físicamente contiguas} \\ 0 & \text{si las regiones } i \text{ y } j \text{ NO son físicamente contiguas} \end{cases}$$

Se asume por definición que $W_{ii} = 0$ ¹¹. La Figura 2, ilustra un ejemplo hipotético.

Figura 2. Ejemplo de una matriz de pesos espaciales.



Fuente: Elaboración propia.

¹¹ Es posible obtener matrices de contigüidad de órdenes superiores. Se dice que dos regiones i y j serán contiguas de segundo orden si ambas están separadas por una tercera región h que es contigua de primer orden a ambas. La misma idea es extensible para órdenes superiores.

Cabe destacar que, de forma habitual, se recurre a la estandarización de la matriz W por filas. Esto es, se divide cada elemento w_{ij} , por la suma total de la fila a la que pertenece.

$$w_{ij}^* = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}} \quad (1)$$

De esta forma, se verificará que la suma de los pesos asociados a cada área es igual a 1, es decir; $\sum_{j=1}^n w_{ij}^* = 1$

La nueva matriz W^* se conoce con el nombre de matriz de ponderaciones espaciales estandarizada por filas (Anselin, 1995). Por otra parte, de acuerdo con Kapoor, et. al. (2007) la matriz W debe ser delimitada de manera uniforme y en valor absoluto, lo que implica la existencia de una constante $c < \infty$ tal que:

$$\overbrace{1 \leq i \leq N \sum_{j=1}^N |W_{ij}|}^{\text{máximo}} \leq c \quad \text{y} \quad \overbrace{1 \leq j \leq N \sum_{i=1}^N |W_{ij}|}^{\text{máximo}} \leq c$$

para producir los resultados asintóticos se exige una estimación consistente.

La matriz de pesos espaciales desempeña un papel fundamental en la incorporación de las relaciones espaciales de las variables al modelo. En el contexto de una única ecuación, por ejemplo, es posible generar una matriz de rezago espacial de la variable endógena WY , a partir de multiplicar la matriz W por un vector de variables endógenas Y , de orden $N \times 1$. Otra posibilidad consiste en incorporar otras variables espacialmente rezagadas, desde una matriz de variables X de orden $N \times k$, reflejado en las columnas de la matriz WX .

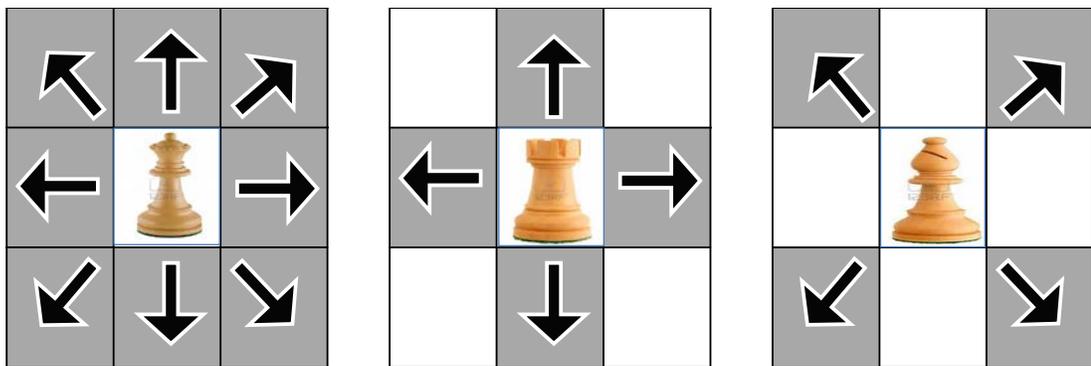
2.2.1 Criterios de contigüidad

Considerando un espacio representado en una cuadrícula regular, la contigüidad (o vecindad) espacial entre dos unidades espaciales puede definirse de la siguiente forma:

- a) Contigüidad tipo Reina (*Queen*): es aquella que considera vecinas a las unidades espaciales que comparten un borde o un vértice del cuadrado. También se conoce con el nombre de criterio de la reina.
- b) Contigüidad tipo Torre (*Rook*): una unidad espacial será vecina de otra si comparte con ésta un lado o borde del cuadrado. Este criterio se conoce con el nombre de torre.
- c) Contigüidad tipo Alfil (*Bishop*): dos unidades espaciales serán vecinas si comparten un vértice del cuadrado. Este criterio se conoce también con el nombre de criterio del alfil.

Estos nombres corresponden al movimiento que realizan la reina, la torre y el alfil en un tablero de ajedrez, Figura 3. El método Torre, por su simplicidad es el más utilizado.

Figura 3. Criterios de contigüidad de pesos espaciales. Retardo espacial de primer orden

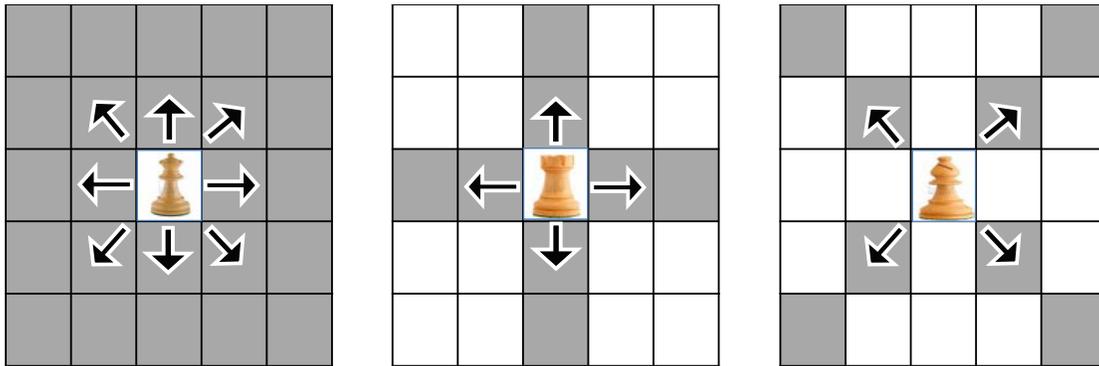


- a) Criterio de Reina (Queen) de primer orden
- b) Criterio de Torre (Rook) de primer orden
- c) Criterio de Alfil (Bishop) de primer orden

Fuente: Elaboración propia.

Los criterios presentados reflejan situaciones de contigüidad de primer orden. No obstante, en algunos problemas puede resultar particularmente útil el abordaje a partir de contigüidades de segundo orden.

Figura 4. Criterios de contigüidad de pesos espaciales. Retardo espacial de segundo orden



- a) Criterio de Reina (Queen) de segundo orden b) Criterio de Torre (Rook) de segundo orden c) Criterio de Alfil (Bishop) de segundo orden

Fuente: Elaboración propia.

En general, no existen situaciones que requieran la adopción de un criterio de vecindad que vaya más allá del segundo orden. En todo caso, sería necesario explorar otras hipótesis de vecindad que respondan a otras formulaciones sobre la base de especificaciones económicas o geográficas alternativas, como pueden ser: la distancia entre dos unidades, el nivel de intercambio comercial entre diferentes regiones o países, etc. En todos los casos, lo que resultará importante aquí, será escoger variables que sean operativas. En la próxima sección se presentan algunas de las formulaciones alternativas más difundidas.

2.2.2 Otras especificaciones para la matriz de contactos

Alternativamente, siguiendo el criterio de la proximidad física, diversos autores han propuesto definiciones de W basadas en la utilización de la distancia entre regiones. A continuación, se presentan algunos procedimientos que constituyen las alternativas más difundidas.

a) Cliff y Ord (1981) construyeron una matriz sobre la base de la distancia que separa a dos regiones i, j , (d_{ij}) , ponderada por el tamaño de la frontera que presenta en común (β_{ij}) . En este caso la formulación da lugar a una matriz asimétrica. De esta forma los elementos de W , se reducen a:

$$w_{ij} = (d_{ij})^{-a} (\beta_{ij})^b \quad (2)$$

Donde a y b son parámetros a estimar. Anselin (1980), comenta que dichos parámetros suelen ser dados *a priori* y no estimados conjuntamente).

b) Bodson y Peeters (1975) han sugerido que la matriz W considere un criterio de accesibilidad general. Esto es, que tenga en cuenta y combine los distintos conductos de comunicación entre regiones. En consecuencia, esta formulación permite redefinir el concepto de vecindad, y a que para tenerlos se deberá tener buenos conductos de comunicación entre las regiones. La formulación analítica de esta propuesta se expresa como:

$$w_{ij} = \sum_{n=1}^N k_n \left\{ \frac{a}{[1 + b * \exp(-c_j d_{ij})]} \right\} \quad (3)$$

Donde:

k_n : la importancia relativa de la vía de comunicación n

N : el número de vías de comunicación

d_{ij} : la distancia entre las regiones i y j .

a, b y c_j : son parámetros a estimar.

c) Case *et al.* (1993) han propuesto que la matriz W se sustente en distancias económicas. En este caso, se asumirá que una región es vecina de otra

cuando es a distancias, en términos de la variable económica operativa, sea pequeña (lo que deberá ser definido por el investigador). A menudo, desde esta perspectiva, la distancia económica es entendida como la reducción de los costos de transacción asociados a la interrelación económica de regiones lejanas entre sí (no contiguas). Esto es presumiblemente provocado por mejoras que se asocian a condiciones económicas estructurales similares y costos de información más bajos.

- d) Anselin (1980) propone la utilización de una *matriz inversa de distancias al cuadrado*, de manera que claramente la intensidad de la interdependencia entre dos regiones disminuye con la distancia que separa sus respectivos centros.

Capítulo III. Análisis exploratorio de datos espaciales (AEDE)

El Análisis Exploratorio de Datos Espaciales AEDE (o ESDA por sus siglas en inglés *Exploratory Spatial Data Analysis*) es comúnmente considerado un subconjunto del Análisis Exploratorio de Datos AED (o EDA por sus siglas en inglés *Exploratory Data Analysis*) que se ocupa de las características distintivas de datos geográficos, con especial énfasis en los problemas de autocorrelación espacial y heterogeneidad espacial (Anselin, 1998; Haining, 1990).

El AEDE puede definirse como un “una colección de técnicas para describir y visualizar distribuciones espaciales, identificar localizaciones espaciales atípicas o outliers espaciales, descubrir patrones de asociación espacial, clusters o puntos calientes, y sugerir regímenes espaciales u otras formas heterogeneidad espacial” (Anselin, 1998; p. 4). De esta forma, al igual que el AED, el AEDE procura detectar en los datos posibles patrones y sugerir hipótesis sobre posibles relaciones entre las variables involucradas en el análisis. Dentro de la colección de técnicas incluidas en el AEDE se encuentran:

- Visualización de distribuciones espaciales;
- Visualización de asociación espacial;
- Indicadores locales de asociación espacial (*LISA Local Indicator of Spatial Association*);
- Indicadores multivariados de asociación espacial.

El desarrollo de nuevos paquetes informáticos y el mejoramiento de otros que ya existían, han permitido a dichas técnicas alcanzar su potencial como procedimiento para proponer posibles conexiones entre las variables.

En el presente estudio, se trabajará con el **software libre OpenGeoDa** desarrollado por Luc Anselin en la Arizona State University.

3.1. Índices de autocorrelación espacial globales y locales.

Para verificar la existencia o ausencia de autocorrelación espacial, se pueden utilizar diferentes índices, siendo el de Moran el más conocida y utilizada en la práctica. Todos estos índices tratan de contrastar la hipótesis nula de ausencia de autocorrelación espacial (esto es, la isotropía espacial) frente a la hipótesis alternativa de existencia de autocorrelación espacial (positiva o negativa).

Los índices de autocorrelación espacial pueden tener un carácter global o local.

Entre dichos estadísticos globales, se encuentran el de Moran (Moran, 1948) y el $G(d)$ (Getis y Ord, 1992), los estadísticos se resumen en un indicador único, presentándose sus expresiones en el cuadro 1.

El índice de Moran (1948), a pesar de su antigüedad no fue muy utilizado debido a que el cálculo manual era sumamente laborioso, por lo que fue necesario contar con programas informáticos para poder realizarlo. Es por ello, que en la década de los noventa fue rescatado del olvido y comenzó a utilizarse en trabajos de investigación de manera consistente.

Dado que los tests de autocorrelación espacial global analizan todas las observaciones de la muestra de forma conjunta, no resultan sensibles a situaciones donde predomine una importante inestabilidad en la distribución espacial de la variable objeto de estudio (procesos no estacionarios espacialmente), es decir, no contemplan la posibilidad de que el esquema de dependencia detectado a nivel global (ausencia de autocorrelación espacial) pueda no mantenerse en todas las unidades del espacio analizado. Dicha limitación es fácilmente superable por medio del cálculo de los denominados contrastes locales de autocorrelación espacial entre los que se encuentran el estadístico local de Moran, I_i , (Anselin, 1995) y los tests New- G_i y New- $G_i^*(d)$ (Ord y Getis, 1995). En el cuadro 2 se presentan sus expresiones. En este caso, se

obtendrá un valor de dichos estadísticos para cada observación de la muestra, realizando así un análisis de la situación de cada unidad espacial por separado.

Finalmente, es necesario resaltar la complejidad de la información suministrada por estos índices, tanto globales como locales (Moreno y Vayá, 2000).

Cuadro 1. Estadísticos de asociación espacial global

<i>I</i> de Moran	<i>G(d)</i> de Getis y Ord
$I = \frac{n}{S_0} * \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{para } i \neq j$ <p>Dónde: x_i: valor de la variable x en la región i \bar{x}: media muestral de la variable x w_{ij}: pesos de la matriz W n: tamaño muestral</p> $S_0: \sum_{i=1}^n \sum_{j=1}^n W_{ij}$	$G(d) = \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij}(d) x_i x_j}{\sum_{i=1}^n x_i x_j} \quad \text{para } i \neq j$ <p>Donde dos pares de regiones i y j son vecinas si se encuentran dentro de una distancia d determinada.</p>
<p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> ✓ $Z(I) > 0$ y signif.: autocorrelación positiva ✓ $Z(I) < 0$ y signif.: autocorrelación negativa 	<p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> ✓ $Z(G) > 0$ y signif.: concentración de valores elevados ✓ $Z(G) < 0$ y signif.: concentración de valores bajos

Cuadro 2. Estadísticos de asociación espacial local

Local de Moran I_i	<i>G(d)</i> de Getis y Ord
$I_i = \frac{Z_i}{\sum_{i=1}^n Z_i^2 / n} \sum_{j \in J_i} W_{ij} Z_j$ <p>z_i: Valor correspondiente a la región i de la variable normalizada J_i: Conjunto de regiones vecinas a i</p>	$New - G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - W_i^* \bar{x}}{s\{[nS_{1i}^* - W_i^{*2}] / (n - 1)\}^{1/2}}$ <p>$W_i^* = W_i + w_{ij}$ $S_{1i}^* = \sum_j w_{ij}^2$ $S^2 = \frac{1}{n - 1} \sum_j (x_j - \bar{x})^2$</p>
<p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> ✓ $Z(I_i) > 0$ y signif.: cluster de valores similares alrededor de i ✓ $Z(I_i) < 0$ y signif.: cluster de valores disímiles alrededor de i 	<p>Interpretación:</p> <ul style="list-style-type: none"> ✓ $New - G_i^* > 0$ (y signif.: cluster de valores similares elevados alrededor de i) ✓ $New - G_i^* < 0$ (y signif.: cluster de valores similares bajos alrededor de i)

3.2. Índice Global de Moran

El Índice Global de Moran es un estadístico que mide la presencia o ausencia de autocorrelación espacial de una variable. Esta medida es similar a la estructura del coeficiente de correlación de Pearson, sus valores fluctúan entre 1 y -1, en donde 1 indica autocorrelación positiva perfecta, -1 expresa autocorrelación negativa perfecta y el valor 0 indica la existencia de patrones completamente aleatorios en la distribución espacial de la variable.

El índice se calcula con base en una matriz de contigüidades o distancias (W_{ij}), predefiniendo los valores vecinos. En la construcción de la matriz de pesos espaciales, nos aseguramos que todas las entidades tengan al menos un vecino (primer orden), para lo cual se utiliza al menos uno de los parámetros de vecindad anteriormente mencionados, que considera cualquier intersección con un polígono como vecino.

El índice Global de Moran viene dado por la siguiente expresión¹²:

$$I = \left(\frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \right) \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{para } i \neq j \quad (4)$$

Donde:

n : número de unidades geográficas en el mapa (municipios)

W_{ij} : matriz de distancias que define si las áreas geográficas i y j son contiguas o no.

La matriz de vecindad espacial se normaliza dividiendo cada elemento de la fila por la sumatoria de los elementos de la fila, situación en la que la suma de los elementos de las filas es igual a 1. En este caso, la expresión se simplifica y queda de la siguiente manera:

¹² El cálculo manual del algoritmo es laborioso, por lo que es necesario usar la estructura formal en el marco de un SIG con capacidad geoestadística como GeoDa.

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{para } i \neq j \quad (5)$$

La regla de decisión para validar la hipótesis de investigación¹³ partir del Índice de Moran es la siguiente:

- ✓ $H_0: I = 0 \rightarrow H_0$: Coeficiente Global de Moran (I) es igual a 0, por lo que no existe autocorrelación espacial, es decir, se observa una **distribución aleatoria** de los valores del IRS en el espacio municipal.
- ✓ $H_1: I \neq 0 \rightarrow H_1$: Coeficiente Global de Moran (I) no es igual a 0, por lo que existe autocorrelación espacial, entonces es posible afirmar que los valores del IRS **NO se distribuyen aleatoriamente** en el espacio geográfico municipal.

Como se puede observar, la expresión (5) es similar al coeficiente de correlación, la gran distinción es la inclusión del término W_{ij} , que cumple un propósito de ponderar de acuerdo con las distancias entre las observaciones. Sin embargo, el estadístico de Moran no está centrado en torno a cero, sino que la esperanza de dicho estadístico es:

$$E(I) = \frac{-1}{(n-1)} \quad (6)$$

Por tanto, el valor esperado es negativo y depende del número de observaciones. Además, la media tiende a cero al aumentar el número de observaciones. Si el valor del estadístico es superior a $\frac{-1}{(n-1)}$, nos encontraremos en presencia de autocorrelación positiva y si, por el contrario, es inferior, estaremos en presencia de autocorrelación negativa. Este índice de Moran se sujeta a una prueba de

¹³ La comprobación de la hipótesis comprende un nivel de significancia del 95% por lo que se espera un p-value <0.05

significancia estadística de v valores Z , es decir, bajo el supuesto de una distribución normal (Cliff y Ord, 1981).

Por último, si el valor da cercano a $\frac{-1}{(n-1)}$, indica que no hay autocorrelación espacial y por lo tanto significa un patrón espacial totalmente aleatorio.

Este valor $Z(l)$ es asintóticamente normal, de manera que si toma valores positivos y estadísticamente significativos (mayores que 1.96 a un nivel de significación del 5%), se podrá concluir que los datos presentan autocorrelación espacial positiva. Si, por el contrario, el valor es estandarizado es negativo y estadísticamente significativo (menor que -1.96 , a un 5% de nivel de significación) se puede afirmar que los datos presentan autocorrelación espacial negativa. Finalmente, si el valor es estandarizado e incluye dentro del intervalo $[-1.96 - 1.96]$ se podrá admitir la hipótesis nula de aleatoriedad espacial.

Lo anteriormente explicado es aplicable sólo para una variable.

Si se desea obtener un estadístico para estudiar la autocorrelación entre dos variables p y q , se recurre a la expresión:

$$I^{pq} = \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i^p - \bar{x}) (x_j^q - \bar{x})}{\sum_{i=1}^n (x_i^p - \bar{x}) (x_j^q - \bar{x})} \quad \text{para } i \neq j \quad (7)$$

Donde:

x_i^p : observación de la variable p en la unidad espacial i

x_j^q : observación de la variable q en la unidad espacial j

Para Anselin, esta expresión permite descubrir los valores en los cuales una variable, presente en una localización dada, muestra una asociación sistemática con otra variable observada en las unidades espaciales más próximas.

En general, los índices globales presentan una fuerte limitación: no son capaces de considerar situaciones de aglomeración (*clusters*) en un área determinada, para la que cabría esperar valores más bajos o más altos si existiese una distribución homogénea (Moreno y Vayá, 2000). Esto implica que un determinado esquema de autocorrelación espacial, detectado mediante contrastes globales, pueda no mantenerse para toda la muestra.

3.3 Índice Local de Asociación Espacial (LISA)

Para la detección de aglomeraciones (*clusters*) resulta útil el análisis de *autocorrelación espacial local*. Este procedimiento permite medir la autocorrelación espacial para la ubicación de cada observación y está basado en el Indicador Local de Moran.

De acuerdo con Anselin (1995), este índice cumple con dos requisitos fundamentales para ser considerado como estadístico local:

a) por un lado, proporciona una cuantificación del grado de agrupamiento significativo de valores similares alrededor de una observación.

b) La suma de estos índices para todas las observaciones es proporcional a un índice global de asociación espacial, por lo que resulta útil para medir la contribución de cada observación al valor del contraste global.

Este análisis es también particularmente útil para detectar la posible presencia de localizaciones que muestren valores altos respecto a la media de otros emplazamientos colindantes, y para las variables de interés, o viceversa.

La forma local estandarizada de *I* de Moran para la observación *i* es la siguiente:

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^n W_{ij} (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2 / n} \quad \text{para } i \neq j \quad (8)$$

Es decir, se obtiene un índice para cada unidad espacial, el cual mide el grado de autocorrelación espacial de cierta variable de la unidad espacial i respecto al valor promedio de la misma en la vecindad establecida por la matriz W . De igual forma, se puede notar que:

$$I = \sum_{i=1}^n I_i \quad (9)$$

Esta fórmula se destaca sobre otras similares porque **la suma de las unidades espaciales locales es igual al valor global**.

De la ecuación (9) se puede deducir que la versión local del Índice de Moran permite estimar el grado de contribución de cada unidad espacial al Índice de Moran global.

Por último, se define como variable **LAG** a la siguiente expresión contenida en el Índice de Moran local:

$$LAG_I = \sum_{j=1}^n W_{ij} Z_j \quad (10)$$

Donde:

$$Z_j = (x_j - \bar{x})$$

Esta variable representa un promedio ponderado por las distancias de la variable Z (asociada al vector de atributos X), en torno a la unidad espacial i . Es decir, representa un valor promedio de la vecindad de la unidad espacial i . La denominación **LAG** que emplea Anselin (1988) se debe a la analogía de esta variable con las series de tiempo, en que se requiere un operador **LAG** para trasladar las variables por uno o más periodos de tiempo. En el caso espacial, las variables se trasladan a distintas localizaciones y el operador **LAG**, en este caso, sería la matriz W .

3.4. Técnicas gráficas para la representación de resultados

Además de los índices globales y locales anteriormente presentados, el análisis exploratorio de datos espaciales cuenta con diversos gráficos que permiten visualizar de forma clara y directa la presencia o ausencia de autocorrelación espacial.

De acuerdo con Anselin (1998), existen dos perspectivas para llevar a cabo un AEDE a partir de técnicas gráficas. Por un lado, la aproximación geoestadística y por el otro, la aproximación *lattice*. La principal diferencia entre estas dos técnicas se encuentra en la forma en la que la proximidad espacial es formalizada.

3.4.1 Aproximación geoestadística

En este método, el supuesto consiste en que el proceso espacial continuo conduce al uso de la distancia métrica, como medio para organizar las observaciones.

Al respecto, Anselin (1998, p. 5) sostiene: “Dado que la asociación espacial se asume como una función suave de la distancia, una medida formal de la (no)similitud entre dos observaciones, es como la diferencia al cuadrado, es comparada con la distancia que los separa. Un mayor grado de autocorrelación espacial implica pequeñas diferencias en distancias cortas y diferencias crecientes en distancias mayores. La función formal que operacionaliza esta noción es el variograma¹⁴. Las técnicas del AEDE desde la geoestadística radican en las formas en las que el variograma puede ser visualizado, resumido y probado para la presencia de no estacionalidad local u otra conducta atípica. Un importante aspecto de esta visualización es que las entidades en un variograma pertenecen a pares de observaciones (separadas por una distancia dada), y no a localizaciones individuales.” Las técnicas gráficas de esta perspectiva se concentran en:

¹⁴ Para detalles técnicos revise Cressie (1993)

-
- La función de distribución acumulativa;
 - El *Box Plot* del variograma;
 - La nube del variograma multivariado.

3.4.2 Aproximación lattice

En este método las observaciones tienen carácter discreto. En consecuencia, la formalización de las similitudes espaciales se desarrolla de una manera diferente a la empleada por el anterior abordaje.

“El concepto clave a aquí es la noción de vecino espacial, que conduce a la construcción de matrices de pesos espaciales y variables espaciales rezagadas” (Anselin, 1998; p. 6).

De esta forma, las herramientas del AEDE en la perspectiva lattice intentan reflejar la asociación entre las variables y sus rezagos espaciales, para diferentes definiciones de valores similares y para diferentes pesos espaciales.

Los instrumentos más útiles son:

- El *box map*;
- El histograma regional;
- El *scatterplot* de Moran;
- Los mapas LISA;
- El *scatterplot* multivariante.

De entre todos los gráficos, el más utilizado es, sin duda, el *scatterplot* de Moran y su asociado *Scatter map*.

3.4.2.1 Scatterplot de Moran

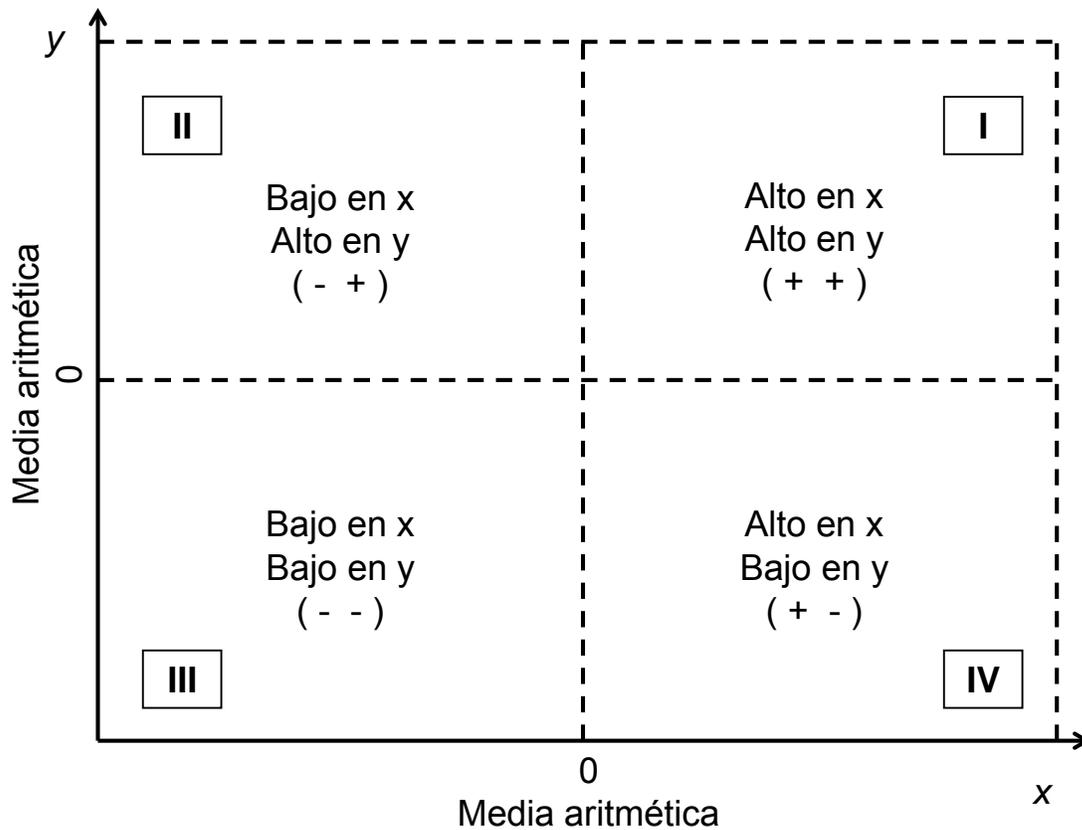
En el año de 1993 Anselin presenta el *Scatterplot* de Moran, gráfico de dispersión que permite observar el comportamiento de cada unidad espacial. Esta novedad

fue uno de los primeros pasos hacia el análisis local, es decir la desagregación del valor global de AE.

El *scatterplot* de Moran univariante, se construye a partir de un sistema de dos ejes cartesianos, de forma que en el eje de abscisas se representan los valores normalizados de la variable objeto de análisis, mientras que en el eje de ordenadas se representan los valores del retardo espacial de la variable estandarizada. Se entiende por retardo espacial de la unidad espacial i el valor medio de la variable estandarizada en todas las unidades espaciales vecinas de la misma. De esta forma, se divide en cuatro cuadrantes (ver figura 5).

En el cuadrante I podemos identificar las observaciones espaciales con valores superiores a la media que, a su vez, también cuentan con vecinos con valores altos (situación *alto-alto*, también conocidos como *hot spots* en el *scatterplot* de Moran). La situación inversa se registra en el cuadrante III (situación *bajo-bajo*, también denominados *cold spots*). Ambos cuadrantes permiten detectar los *clusters* o agrupamientos de unidades espaciales con valores similares a los de sus vecinos. En contraparte, los *outliers* espaciales responden a contextos mixtos, en otras palabras, unidades espaciales con valores bajos (inferiores a la media) con vecinos que registran valores altos (situación *bajo-alto*) en el cuadrante II. El escenario opuesto (situación *alto-bajo*) se encuentra en el cuadrante IV. La importancia de los *outliers* radica en que no responden al principio de AE ya que el valor de las unidades espaciales que poseen a estas características se diferencia de los valores de sus vecinos. Además, por su carácter atípico sirven, inicialmente, para detectar errores en el ingreso de la información, ya sea de forma accidental o deliberada.

Figura 5. Moran Scatterplot



De esta manera, en función de la posición de las observaciones en los cuatro cuadrantes del gráfico, se podrá determinar el patrón de dependencia espacial existente. Si la nube de puntos está dispersa en los cuatro cuadrantes es indicio de ausencia de correlación espacial. Si la nube de puntos se sitúa alrededor de la diagonal que va desde el primer cuadrante (extremo superior derecho del gráfico) al tercer cuadrante (extremo inferior izquierdo), será un claro indicio de la existencia de autocorrelación espacial positiva. Por el contrario, cuando los valores representados se sitúen al rededor de la diagonal que cruza desde el segundo cuadrante (extremo inferior derecho) hasta el cuarto cuadrante (extremo superior izquierdo), se podrá hablar de autocorrelación espacial negativa.

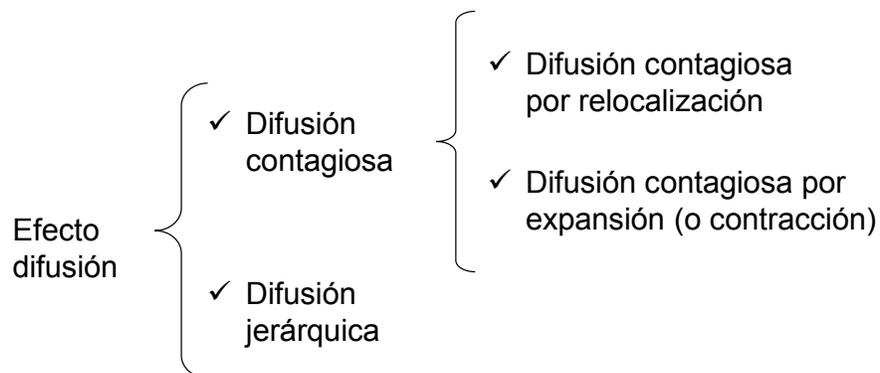
Finalmente, el gráfico ajustado a una recta de regresión a la nube de puntos representados, de manera que la pendiente de dicha recta de regresión coincide con el estadístico I global de Moran, anteriormente presentado.

También se puede utilizar un diagrama de dispersión de Moran bivalente. En este gráfico, se representa en el eje de ordenadas el retardo espacial de la variable que se está analizando, y en el eje de abscisas los valores estandarizados de la variable explicativa. Una vez representada la nube de puntos, la pendiente de la recta de regresión ajustada (el valor de la I de Moran) cuantifica la relación lineal que existe entre la variable explicativa y la explicada en las unidades espaciales vecinas a una unidad dada.

Cabe mencionar, que el *Scatter map* es la traslación de la información del *Scatterplot* un mapa del territorio, donde cada uno de los cuadrantes es resaltado con un color diferente.

Este tipo de análisis permite observar los diferentes tipos de dinámica que pudieran presentar las observaciones durante dos o más periodos a comparar. Por ejemplo, es posible que durante este lapso de tiempo las observaciones cambien su ubicación en el plano cartesiano a cualquier otro cuadrante, lo que se conoce como *efecto difusión*. Estos efectos pueden clasificarse de dos tipos, difusión espacial contagiosa y difusión espacial jerárquica (ver Figura 6).

Figura 6. Clasificación de los tipos de difusión.



La difusión contagiosa, se presenta, si un municipio llega a afectar a un municipio vecino, de modo que este último que era pobre (rico) pasa ahora a clasificarse como rico (pobre). Esta difusión se puede llevar a cabo por causas de relocalización o de expansión (o contracción). En el caso de una relocalización espacial de la pobreza, el municipio pobre puede dejar de serlo pero no sin antes causar efectos negativos en sus vecinos. La expansión o contracción, por su parte, implica que la situación de pobreza o riqueza cambia para el municipio local pero no para sus vecinos, es decir, que se produce un contagio que va de los vecinos hacia el municipio local.

3.4.2.2 Los mapas LISA.

Para detectar autocorrelaciones espaciales a nivel local, se utiliza el gráfico denominado mapa LISA¹⁵. Este mapa utiliza los índices de Moran Locales calculados para todas las unidades espaciales analizadas. La principal utilidad de este gráfico es que permite localizar geográficamente los agrupamientos espaciales (que se producen cuando una unidad espacial que registra un valor alto (bajo) de la variable se encuentra rodeada de unidades espaciales que también registran un valor alto (bajo) de dicha variable, esto es, alto-alto o bajo-bajo) y los atípicos espaciales (que son aquellos que surgen cuando una unidad espacial con un elevado valor de la variable analizada se encuentra rodeada de unidades espaciales en las que la variable registra valores pequeños, o viceversa, es decir, alto-bajo o bajo-alto).

El mapa LISA tiene una versión bi variante, que permite visualizar para cada unidad espacial considerada, la relación existente entre los valores estandarizados de la variable objeto de interés y el retardo espacial de otra variable diferente. De esta manera, es posible también identificar tanto agrupamientos como atípicos espaciales teniendo en cuenta simultáneamente dos variables.

¹⁵ LISA es el acrónimo de “*Local Indicators of Spatial Association*”.

Además de los gráficos anteriormente mencionados, la visualización de la autocorrelación espacial se mejora con otros gráficos adicionales. Los más interesantes, en este sentido, son los siguientes:

- a) *Gráfico de coordenadas paralelas*: se trata de un gráfico compuesto por dos o tres líneas paralelas (dependiendo de que se quiera analizar la relación entre dos o entre tres variables), de forma que en cada línea se representa el valor observado de cada variable. A continuación, para cada unidad espacial estudiada se unen mediante líneas los valores de las dos o tres variables objeto de análisis. De esta forma, es posible detectar la presencia tanto de clusters espaciales (cuando las líneas de varias unidades espaciales sean muy similares) como de atípicos espaciales (cuando se aprecien diferencias importantes entre las líneas de diferentes unidades espaciales).

- b) *Cartograma*: se trata de un mapa en el que el valor de la variable para cada unidad espacial se representa mediante un círculo, cuyo tamaño es proporcional al citado valor. Además, los círculos poseen diferentes colores dependiendo de que el valor de la variable sea anormalmente alto o bajo. Así, el color por defecto del círculo es el verde, pero su color será el negro si la variable toma un valor negativo, el blanco si la variable toma un valor nulo en esa unidad espacial, el rojo si el valor es anormalmente elevado o el azul si el valor es anormalmente bajo.

- c) *Proceso conocido como randomization o de aleatorización*: los datos de las unidades espaciales se intercambian (permutan) al azar obteniéndose diferentes valores de autocorrelación que se comparan con el generado en la distribución real. Posteriormente, se especifica el nivel de significancia que indica la probabilidad de rechazar la hipótesis nula si esta es verdadera. Por lo tanto, es la probabilidad que se está dispuesto a cometer al aceptar la hipótesis alternativa. Se suele elegir de acuerdo a la

importancia del problema y generalmente es del 5 % (0.05) y 1 % (0.01) (Buzai y Baxendale, 2006).

Asimismo, **el *p-valor*** es el resultado que nos brinda el *test* de hipótesis. Si el nivel de significancia es superior al *p-valor*, se rechaza la hipótesis nula y se acepta la alternativa. Por el contrario, si se comprueba la hipótesis nula se puede decir que la configuración espacial se produce de forma aleatoria.

Para el análisis local, al igual que en el caso del global, una vez aplicado el procedimiento estadístico, a continuación hay que determinar su distribución. Como este desarrollo es todavía bastante novedoso no se ha podido establecer su distribución exacta. Es por ello, que en la práctica, también se recomienda recurrir a procedimientos como el de aleatorización y permutación (Anselin, 1995). El valor esperado de *I* de Moran para una observación individual es:

$$E(I) = - \frac{W}{(n - 1)} \quad (5)$$

Donde ***W*** es la suma de los elementos de la fila.

De esta forma, muchas de las técnicas de análisis exploratorio de datos espaciales, aprovecha la capacidad de visualización y manipulación de los datos que proporcionan los Sistemas de Información Geográfica (SIG). Por ejemplo, el SIG ArcView (ESRI, 1995) se organiza entorno a varias vistas vinculadas de los datos (un mapa, una tabla y varios tipos de gráficos), lo cual permite un cierto grado de interacción dinámica en el sentido de que una selección determinada es una de las vistas se refleja de forma inmediata en el resto. Asimismo, se desarrollaron en los años noventa varias vinculaciones de ArcView con otros módulos de software, convirtiéndolo en una herramienta adecuada para el análisis exploratorio de datos espaciales. Así, por ejemplo, se ha vinculado el software GeoDa para el análisis de datos espaciales creado por Anselin.

Capítulo IV. Análisis de Contrastes de Rezago Social en la Zona Metropolitana del Valle de México.

El estudio del índice de rezago social, describe la situación de un área geográfica, sea por entidad, municipio o localidad, sin tomar en cuenta la localización espacial; es decir tratando las unidades territoriales como unidades aisladas sin ningún tipo de conexión con las áreas vecinas.

La metodología propuesta en este trabajo, busca hacer aportes en dos vías: en primer lugar, realizar un análisis municipal; es decir, analizando la localización espacial del índice de rezago social en cada uno de los municipios de la ZMVM. En segundo lugar, determinar si el índice de rezago social de un municipio está influenciado por los índices de las unidades vecinas, formando lo que se denomina un cluster locacional.

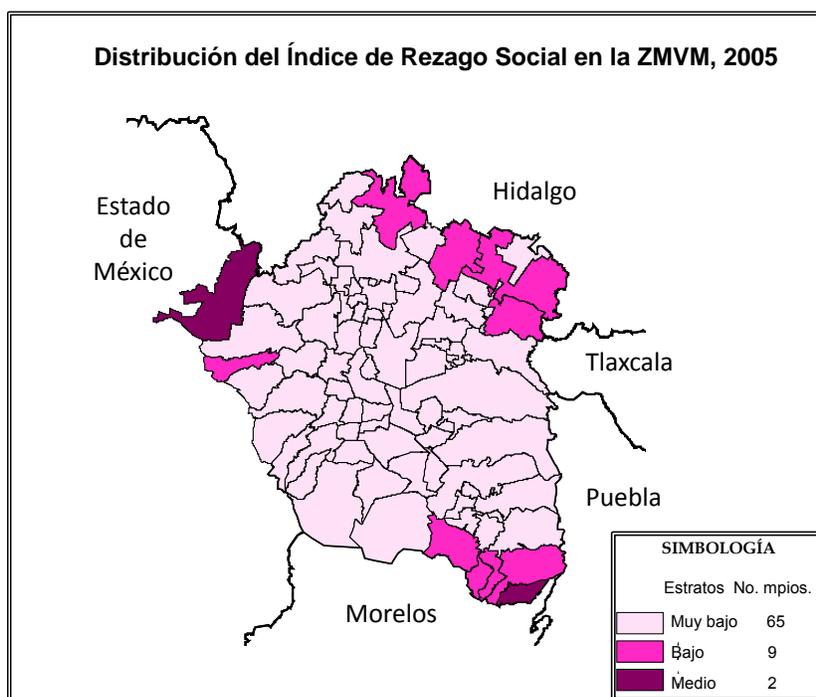
4.1 Análisis municipal

En el capítulo anterior, se mencionó que con base en la metodología de *Dalenius & Hodges*, se identificaron cinco estratos en que se distribuye el índice de rezago social: muy bajo, bajo, medio, alto y muy alto.

Para la ZMVM se identifican tres estratos: muy bajo, bajo y medio; esto ocurre para cada año analizado: 2005 y 2010. Los mapas 2 y 3 muestran la distribución de los estratos.

Cabe mencionar, que los colores claros en los mapas muestran los municipios con un grado de rezago social muy bajo, mientras que los municipios con colores más oscuros muestran municipios con un grado medio de rezago social. De este modo es posible observar una distribución del rezago social, que claramente se encuentra concentrada en algunas zonas geográficas determinadas.

Mapa 2. Índice de Rezago Social en la ZMVM, 2005.

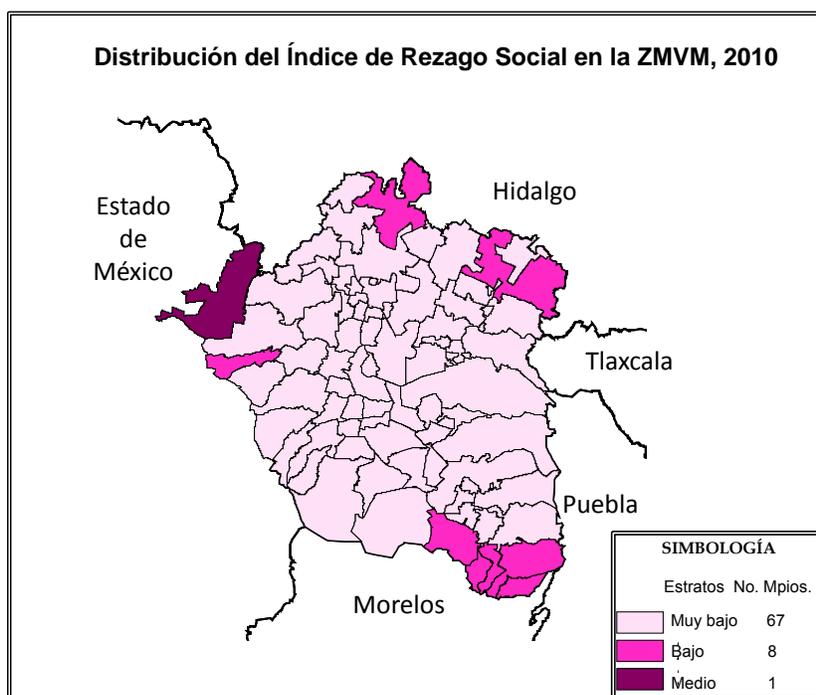


Fuente: elaboración propia

Para el 2005, los tres estratos del índice de rezago social en la ZMVM se integran por los siguientes municipios y delegaciones:

- Medio: Ecatzingo y Villa del carbón, municipios correspondientes al estado de México;
- Bajo: Atlautla, A xapusco, Hueypoxtla, I sidro F abela, J uchitepec, Otumba, Ozumba, Temascalapa y Tepetlixpa, municipios correspondientes al estado de México;
- Muy B ajo: dieciséis delegaciones del D istricto F ederal, c uarenta y oc ho municipios del Estado de México y un municipio del estado de Hidalgo.

Mapa 3. Índice de Rezago Social en la ZMVM, 2010.



Fuente: elaboración propia

En el 2010, los tres estratos identificados se conforman por los siguientes municipios y delegaciones de la ZMVM.

- Medio: Villa del carbón;
- Bajo: Atlautla, Axapusco, Ecatzingo, Hueypoxtla, Isidro Fabela, Juchitepec, Ozumba y Tepetlixpa;
- Muy Bajo: dieciséis delegaciones del distrito federal, cincuenta municipios del estado de México y un municipio del estado de Hidalgo.

En resumen, los municipios y delegaciones que se mantienen en su estrato del periodo 2005 al 2010, son los siguientes:

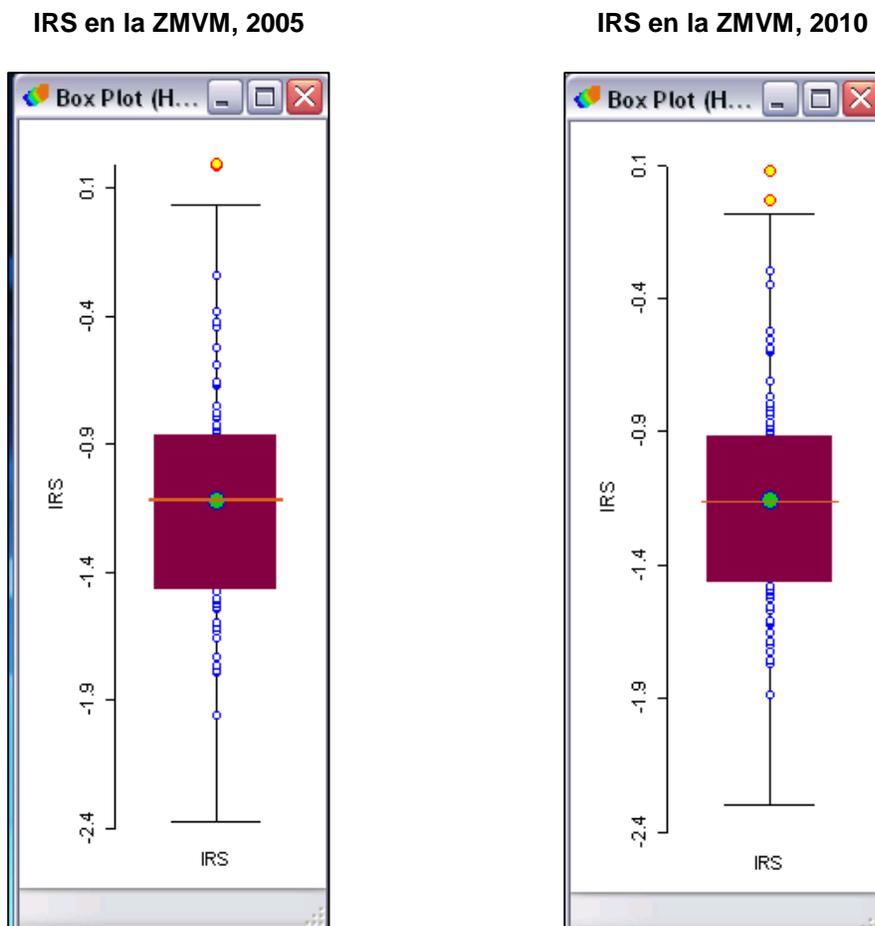
- Medio: Villa del carbón;
- Bajo: Atlautla, Axapusco, Hueypoxtla, Isidro Fabela, Juchitepec, Ozumba y Tepetlixpa;

- Muy B ajo: dieciséis del egaciones del di strito federal, cuarenta y oc ho municipios del estado de México y un municipio del estado de Hidalgo.

Este primer resultado parece indicar que la distribución espacial del IRS, no se comporta en f orma al eatoria, tal hec ho da ev idencia de algún t ipo d e dependencia entre los municipios.

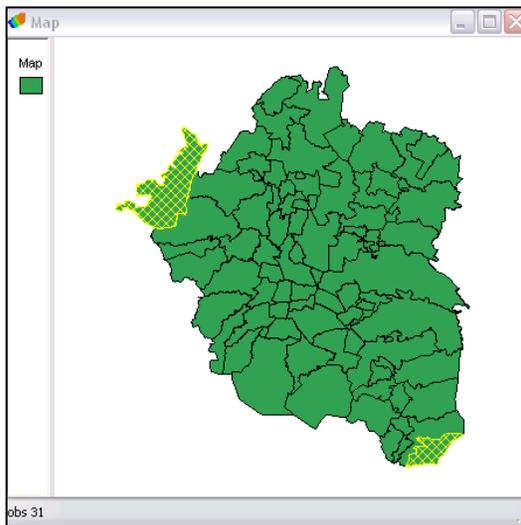
En el gráfico 1 y mapa 4, se identifican los valores outliers para cada periodo analizado en la ZMVM. En ambos casos, los municipios outliers resultan ser: Ecatzingo y Villa del carbón; cabe señalar que para el año 2010, el municipio de Ecatzingo se ubica más cerca del límite del tercer cuartil, es decir su IRS poco a poco se aleja de ser un municipio outliers.

Gráfico 1. Box plot del Índice de Rezago Social, 2005 y 2010.

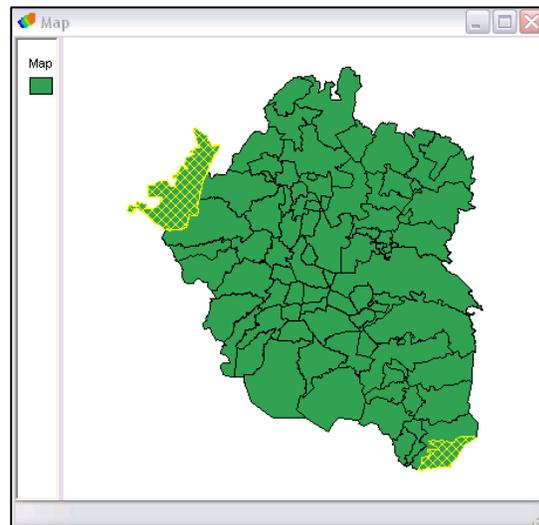


Mapa 4. Municipios outliers, 2005 y 2010

ZMVM, 2005



ZMVM, 2010



4.2. Resultados de autocorrelación global

Para validar formalmente la posible evidencia de algún tipo de dependencia entre los municipios de la ZMVM, para el 2005 y 2010, la regla de decisión a partir del Índice de Moran queda de la siguiente manera:

- ✓ $H_0: I = 0 \rightarrow H_0$: **No** existe autocorrelación espacial, es decir, se observa una **distribución aleatoria** de los valores del RS en el espacio municipal.
- ✓ $H_1: I \neq 0 \rightarrow H_1$: **Existe** autocorrelación espacial, entonces es posible afirmar que los valores del RS **NO se distribuyen aleatoriamente** en el espacio geográfico municipal.

En la tabla 4, se observan los resultados del Índice de Moran para los años 2005 y 2010.

Tabla 4. Índice de Moran

Año	Índice de Moran (Primer orden)	Índice de Moran (Segundo orden)
2005	0.5467	0.3021
2010	0.5313	0.2730

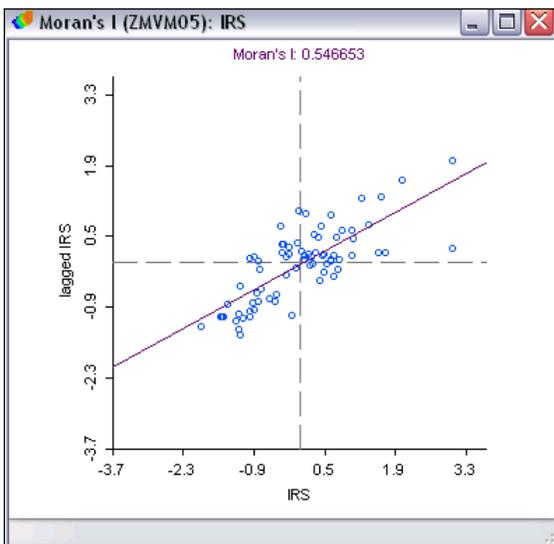
El p- value del índice de Moran es 0.05

Para su cálculo se definieron dos tipos de matriz de contigüidad, de primer y segundo orden, utilizando el criterio tipo torre (Rook).

En los siguientes gráficos scatterplot del índice de Moran, se muestra la distribución de los municipios según la relación espacial que los agrupa. Ya sea formando un cluster RS en el cuadrante I y III (municipios con valores similares de RS) o un outlier RS en los cuadrantes II y IV (municipios vecinos con valores disimilares de RS).

Gráfico 2. Diagramas de dispersión de Moran de primer orden, 2005 y 2010.

Scatterplot 2005



Scatterplot 2010

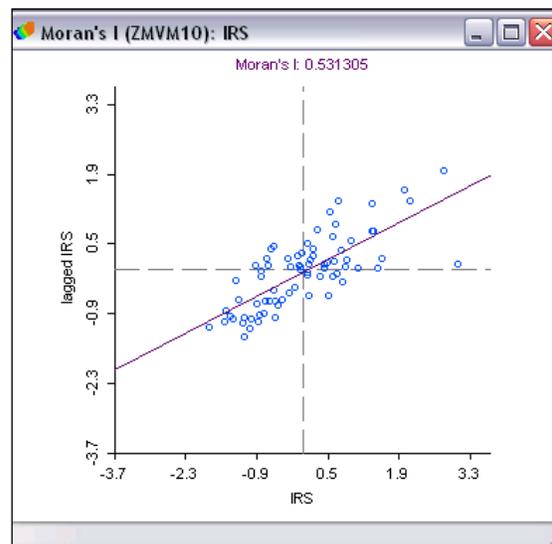
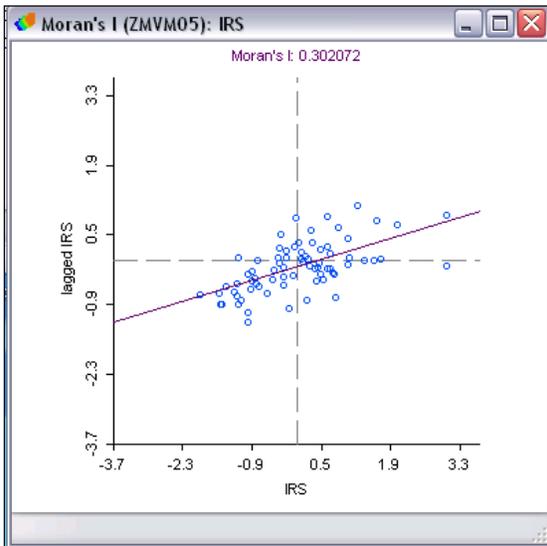
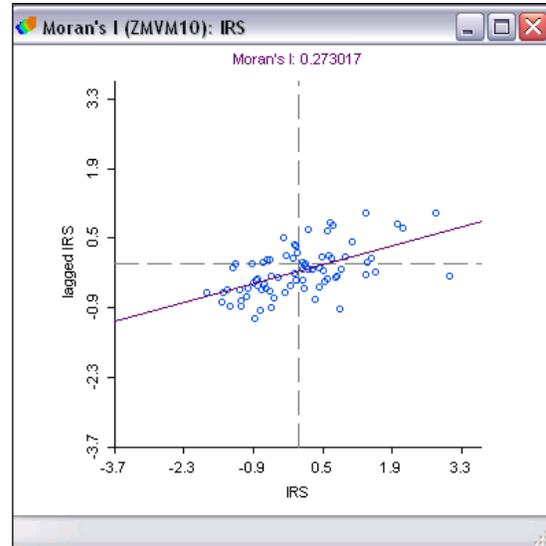


Gráfico 3. Diagramas de dispersión de Moran de segundo orden, 2005 y 2010.

Scatterplot 2005



Scatterplot 2010



Los valores del coeficiente de Moran (tabla 4) sugieren validar la hipótesis H_1 que considera la existencia de autocorrelación espacial en el RS. En tanto, todos los coeficientes cumplen la regla de decisión que indica que I es distinto a cero ($I \neq 0$), lo que permite rechazar la hipótesis nula (H_0) que sostenía que el RS es resultado de un proceso espacialmente aleatorio. Por ello, se validó que el RS es un fenómeno que posee una estructura geográfica que tiende a concentrar (cluster) en sus valores altos – altos (Hot-Spots) y bajos – bajos (Cold-Spots) y, por ende, no es resultado de un proceso al azar.

Por otro lado, las categorías de asociación mixta también, llamados outliers espaciales, presentes en los cuadrantes II y IV, representan los municipios y delegaciones con valores bajos/altos de RS que están rodeados por municipios y delegaciones con valores diferentes. A nivel teórico, se confirma que el RS en los municipios y delegaciones de la ZMVM posee relaciones espaciales específicas como la autocorrelación espacial. En este sentido se señala que:

1. En los municipios y delegaciones de la Z MVM, el RS no se distribuye aleatoriamente en el espacio, más bien presenta una geografía que tiende a la concentración y autocorrelación espacial, es decir; cada municipio o delegación depende no solamente de sus condiciones propias sino además, de cada situación de su vecino más cercano.
2. Desde el 2005 a 2010 existe autocorrelación espacial del RS en los municipios y delegaciones de la Z MVM. La que se expresa en la regionalización geográfica del mismo.
3. El espacio geográfico municipal y delegacional es una variable explicativa del RS.

A continuación, se calcula el valor esperado de I de Moran para cada periodo, mediante el uso de la fórmula (6), con la intención de fortalecer los resultados anteriores.

Tabla 5. Valor esperado de Índice de Moran

Año	Valor esperado
2005, 2010	$E[I] = -\frac{1}{75} = -0.013$

Para ambos periodos el valor de los índices, son superiores al valor esperado $E[I] = -0.013$, por lo que nos encontramos en presencia de autocorrelación positiva.

Otra alternativa que nos ayuda a ratificar la dependencia, es a través de la prueba de randomization, donde se generan 499 permutaciones de nuevos valores de I de Moran. En la parte inferior izquierda del gráfico se registran el valor observado $I: 0.5467$ y el esperado: -0.0133 . Ambos resultados coinciden con los cálculos

realizados con anterioridad. Por lo que nos permite concluir que existe una autocorrelación positiva.

El (*p-value*) es de 0.0020 siendo menor a 0.05^{16} , se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

Gráfico 4. Randomization de primer orden 2005.

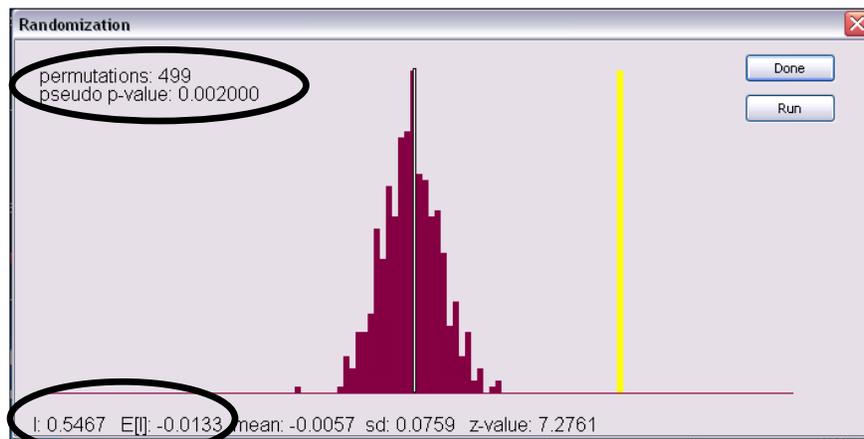
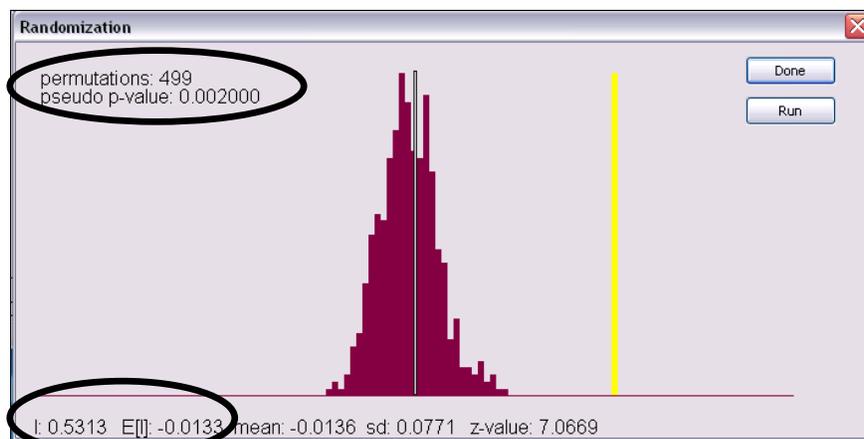


Gráfico 5. Randomization de primer orden 2010.



¹⁶ La comprobación de la hipótesis comprende un nivel de significancia del 95% por lo que se espera un *p-value* < 0.05

Los resultados anteriores validan que el Rezago Social es un fenómeno que posee una estructura geográfica que tiende a concentrarse en sus valores altos – altos y bajos – bajos y por ende, no es un resultado de un proceso al azar o aleatorio.

4.3. Resultados de autocorrelación local

El siguiente paso es, realizar el análisis de dependencia espacial a nivel local (LISA), para identificar y localizar geográficamente los municipios que tienden a agruparse en clusters espaciales de alto o bajo Rezago Social.

El análisis LISA se realizó con el software GeoDa que descompone el índice de Moran, estableciendo en cuanto contribuye cada unidad espacial (municipio y delegaciones) a la formación del valor general y, a su vez, identifica el nivel de significancia de cada conglomerado. Los mapas LISA representan las áreas geográficas que forman cluster con mayor significancia estadística en la concentración de los valores altos y bajos del RS. Con GeoDa se generan dos productos cartográficos: 1) Mapa con los polígonos municipales y delegacionales que conforman un cluster espacial y 2) Mapa con las áreas de significancia (p-value) para los cluster identificados.

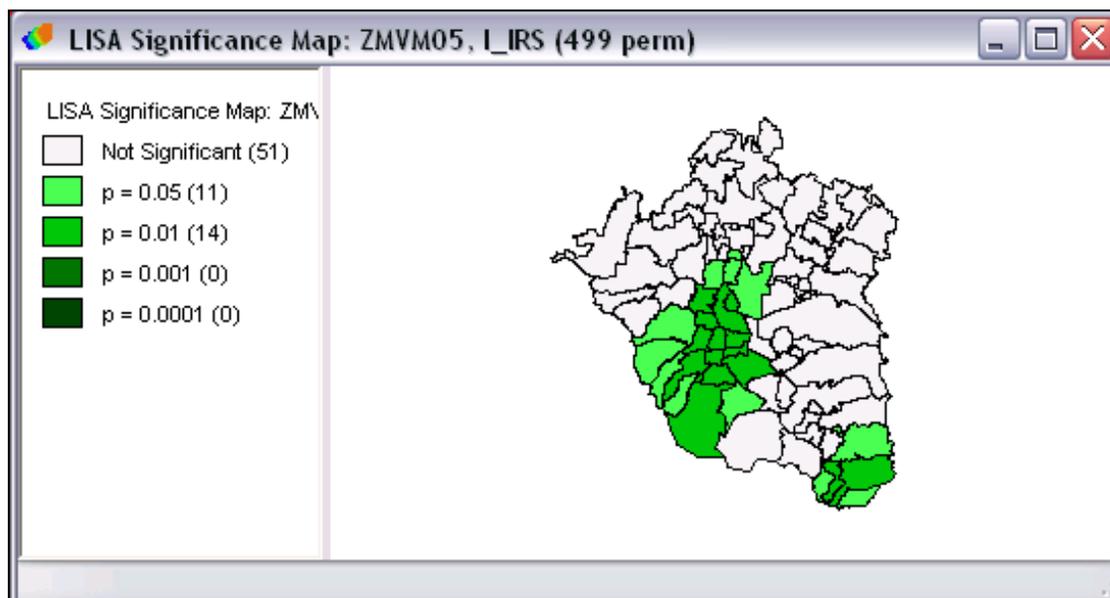
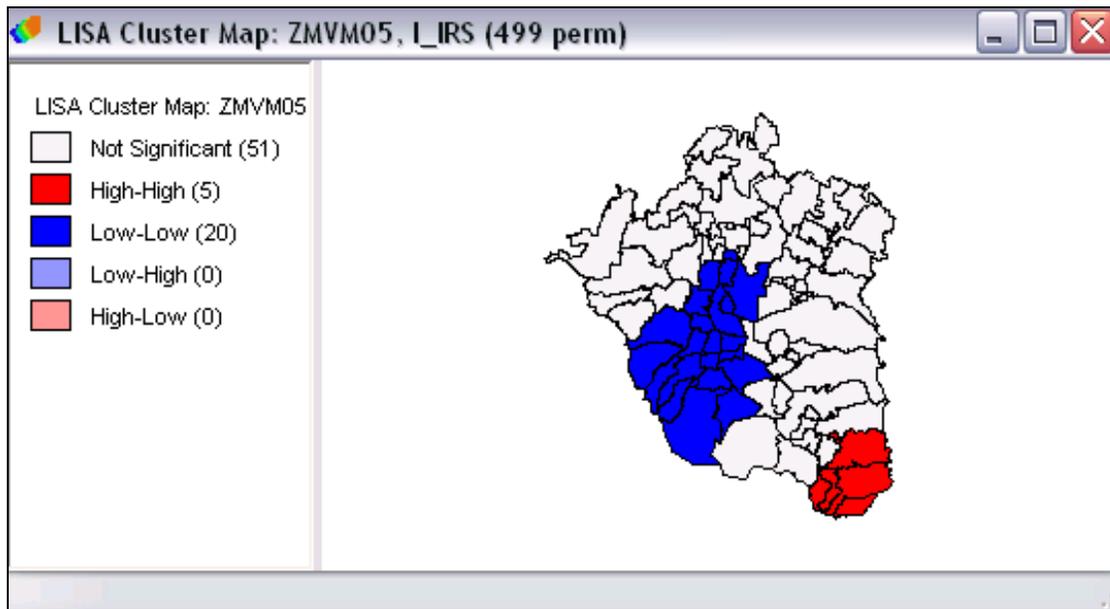
En el mapa 5, se observa una geografía de RS marcada por la concentración de los municipios y delegaciones, formando clusters espaciales de alto (rojo) y bajo (azul) RS. El cluster de RS (hot-spots) que presenta autocorrelación y significancia estadística se localiza en los siguientes municipios:

- Atlautla y Ozumba con una nivel de significancia del 1%
- Amecameca, Ecatzingo, y Tepetliltla con un nivel de significancia del 5%

Cabe mencionar, que el municipio de Ecatzingo se ubica con un grado de rezago social medio, mientras que los demás municipios se encuentran en un grado de rezago “muy bajo” y “bajo”, este hecho es un claro ejemplo de que cada municipio no solamente depende de sus condiciones propias sino además, de cada situación de su vecino más cercano.

Por otra parte, se identifica que los niveles de menor RS (azul) se localizan en su totalidad en el distrito federal y en seis municipios del estado de México (Cold Spots).

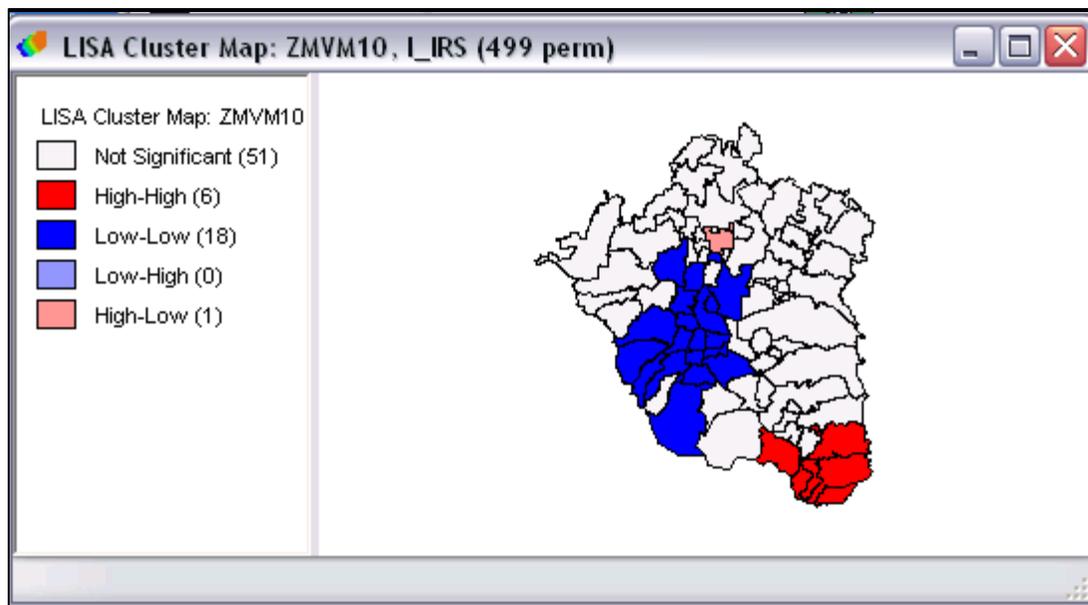
Mapa 5. Cluster y significancia (LISA) del Índice de Rezago Social en la ZMVM, 2005

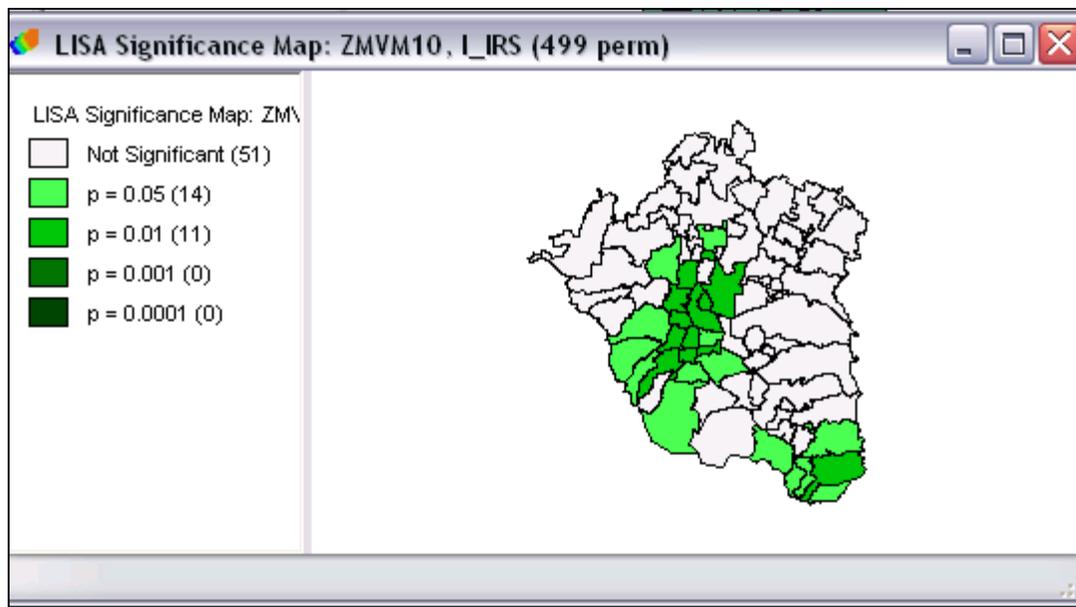


En el mapa 6, se observa una geografía de RS diferente del 2005, note que el cluster alto (rojo) adhiere el municipio de Juchitepec que tiene un grado de RS bajo.

Por otra parte, la delegación La Magdalena Contreras y el municipio de Coacalco de Berriozabal, dejan de pertenecer al cluster de menor RS (azul, Cold Spots).. Además, el municipio de Nextlalpan, cae a la categoría de asociación mixta, llamada outliers espacial, y a que para el 2010 los municipios que lo rodean cambiaron de valores en su índice de Rezago Social.

Mapa 6. Cluster y significancia (LISA) del Índice de Rezago Social en la ZMVM, 2010





Los mapas muestran la clasificación de los municipios de acuerdo con la significancia del tipo de asociación espacial encontrada. Los municipios sombreados corresponden a aquellos en los que existen evidencias de dependencia espacial significativa. Los resultados indican la existencia de dos clusters de municipios con distintos niveles de RS: el primero (Alto – Alto) indica que son municipios con alto nivel de RS (muy bajos) y están rodeados de municipios con altos niveles de RS (muy bajos)

Conclusiones.

En este trabajo de investigación, por medio del análisis geoespacial, se analiza e identifica la existencia de autocorrelación espacial en los valores del Índice de Rezago Social en el espacio geográfico municipal de la ZMVM en ambos periodos analizados. Es posible afirmar que la ubicación geográfica juega un papel fundamental en la determinación del rezago social. Es decir, el rezago social de un municipio depende en forma importante del nivel de rezago social de los municipios vecinos. Cabe señalar que en los dos periodos analizados no solamente resultó significativa la dependencia espacial de primer orden sino también la de segundo orden, es decir, que para cada municipio no sólo son importantes sus vecinos inmediatos, sino además los vecinos de sus vecinos, lo que puede llamarse *efecto vecindario*.

Un segundo resultado, es la identificación de dos tipos de clusters (alto - alto) y (bajo - bajo). La determinación de contrastes de rezago social en la ZMVM, muestra que es un fenómeno regionalizado y concentrado en ciertas zonas geográficas. Con esto se concluye que el espacio es una variable fundamental en los estudios sociales y el análisis geoespacial es una metodología pertinente para su estudio. Este no muy nuevo enfoque permite aportar mayor conocimiento y difusión de los fenómenos complejos multicausales como el rezago social, así como sus implicaciones para la toma de decisiones sobre programas de intervención focalizada a la priorización de los recursos.

Bibliografía.

- ✓ ABER, J.L., M.A. Gephart, J. Brooks-Gunn y J.P. Connell (1997), "Development in Context. Implications for Studying Neighborhood Effects", en J. Brooks-Gunn, G.J. Duncan, and J .L. A ber (editores) *Neighborhood Poverty*, V ol. I , N ueva Y ork. Russell Sage Foundation.
- ✓ ÁLVAREZ, María Eugenia, Martínez, Horacio, *El desafío de la pobreza*, Fundación Social - Siglo del hombre editores – Confederación Colombiana de ONG (CCONG), Bogotá, 2001.
- ✓ ANSELIN, Luc, (1980), "Estimation methods for spatial autoregressive structures". Ithaca N Y: Cornell University, Regional Science Dissertation and Monograph Series #8.
- ✓ ANSELIN, L uc, (1988), *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, The Netherlands.
- ✓ ANSELIN, Luc (1995), "Local Indicators of Spatial Association-LISA", *Geographical Analysis*, vol. 27, núm. 2, abril, pp. 93-115, Ohio State University Press, EUA.
- ✓ ANSELIN, L uc, (1996), "Local I ndicators o f S patial A nalysis – LISA", en *Geographical Analysis*, vol. 27, 199 5, p p. 93-115. ANSELIN, Luc , " The Moran Scatterplot as a n ESDA Tool to Asses local I nstability in Spatial Asociation", e n *Spatial Analytical Perspectives on GIS in Enviromental and Socio- Economic Sciences*, London.
- ✓ ANSELIN, Luc . (2003). *GeoDa 0.9 User's Guide*. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign, IL.
- ✓ ANSELIN, L uc y B era, A .K. (1998), " Spatial depe ndence i n l inear r egression models w ith an i ntroduction t o s patial econometrics" *Handbook o f A pplied Economic Statistics* (eds.) Aman Ullah and D.E.A Giles. New York, Marcel Dekker, Inc.

-
- ✓ ANSELIN, Luc y Griffith, D.A. (1988), "Do spatial effects really matter in regression analysis?", Papers Regional Science Association, No. 65, pp. 11-34
 - ✓ ANSELIN, Luc y Hudak, S. (1992) "Spatial econometrics in practice. A review of software options", Regional Science and Urban Economics, No. 22, pp 509-536.
 - ✓ ANSELIN, Luc, Ibnu Syabri and Youngihn Kho (2004), GeoDa: An Introduction to Spatial Data Analysis, Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, E U A. Disponible en <http://geodacenter.asu.edu/pdf/geodaGA.pdf>
 - ✓ BODSON, P. y Peeters, D. (1975), "Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation. An application to a Belgian labour-demand function", Environment and Planning A, No. 7, pp. 455-472.
 - ✓ BRONFENBRENNER, U. (1986), "Ecology of the Family as context for Human Development", en Developmental Psychology, Vol. 22, No. 6; pp 723 – 742.
 - ✓ BROOKS–Gunn, J., G.J Duncan, P. Kato y N. Sealand (1993), "Do Neighborhoods Influence Child and Adolescent Behavior?", en American Journal of Sociology, No. 99; pp.353 - 395.
 - ✓ BROOKS–Gunn, J., G.J Duncan, y L. Aber (1997), Neighborhood Poverty", Vol. 1, Nueva York. Russell Sage Foundation.
 - ✓ BUZAI, G. y Baxendale, C. (2006), "Análisis socioespacial con sistemas de información geográfica. Lugar Editorial, Buenos Aires, Argentina.
 - ✓ CASSETTI, E. (1972), "Generating models by the expansion method: applications to geographical research" Geographical Analysis, No. 4, pp. 81 -91.
 - ✓ CLARK. D,W.K.D. Davis y R. J. Johnston (1974), "The Application of Factor Analysis in Human Geography" en The Statistician, Vol. 23, Nos. 3-4, pp. 259-281.
 - ✓ CLIFF, A. y Ord, J. (1973), "Spatial Autocorrelation". London, Pion.
 - ✓ CLIFF, A. y Ord, J. (1981), "Spatial Process. Models and Applications". London, Pion.

-
- ✓ Comisión Económica para América Latina y el Caribe (CEPAL) y el Centro Latinoamericano y Caribeño de Demografía (CELADE) – División de la Población de la CEPAL, “Pobreza Infantil en América Latina y el Caribe 2010”.
 - ✓ CONEVAL, *Datos en línea*, en <http://www.coneval.gob.mx/Medicion>. México D.F.
 - ✓ CORRADO, L. y Fingleton, B. (2011): “Where is the Economics in the Spatial Econometrics”, Discussion Papers in Economics, N° 11-01, pp. 1-34.
 - ✓ CRESSIE, N. (1993), “Statistics for Spatial Data” (Revised Version). New York. Wiley.
 - ✓ FOSTER, S. Y Gorr, W. (1983), “Adaptive filtering approaches to spatial modeling”, Modeling and Simulation, No. 15, pp. 29 – 34.
 - ✓ FOSTER, S. Y Gorr, W. (1986), “An adaptive filter for estimating spatially-varying parameters: Application to modeling police hours spent in response to calls for service” Management Science, No. 32, pp. 878 – 889.
 - ✓ FOSTER, S. Y Gorr, W. (1983), “Adaptive filtering approaches to spatial modeling”, Modeling and Simulation, No. 14, pp. 745 – 750.
 - ✓ FOTHERINGHAM, A.S., Charlton, M. y Brunsdon, C. (1998), “Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis”, Environment and Planning A, No. 30, pp.1905 - 1927
 - ✓ GEARY, R. (1954), “The contiguity ratio and statistical mapping”, The Incorporated Statistician, No. 5, pp. 115-145
 - ✓ GETIS, A. y Ord, J. (1992), “The analysis of spatial association by use of distance statistics” Geographical Analysis, No. 24, pp. 189 -206.
 - ✓ GEPHART, M.A. (1997),” Neighborhoods and Communities as Contexts for Development”, en J. Brooks-Gunn, G. Duncan, y J.L. Aber (editores) Neighborhood Poverty, Vol. I, Nueva York. Russell Sage Foundation.
 - ✓ HAINING, R. (1978), “ Estimating spatial interaction models”, Environment and Planning A, No. 10, pp. 305 – 320.
 - ✓ HILDRETH, C. y Houck, J. (1968), “ Some estimators for a linear model with random coefficients”, Journal of the American Statistical Association, No. 63, pp. 584-595.
 - ✓ KAPOOR, M.; Kelejian, H. H. y Prucha, I. R. (2007), “Panel Data Models with Spatially Correlated Error Components”, en Journal of Econometrics, Vol. 140 (1), pp. 97-130.

-
- ✓ LIZAMA C., Guillermo, Geografía electoral del abstencionismo en los municipios de México (1994 – 2009), Universidad Nacional Autónoma de México. Doctorado en Geografía.
 - ✓ MASSEY, D .S. (1990), “American Apartheid, Segregation and Making of the Underclass”, en American Journal of Sociology, Vol. 96, No 2; pp.329 - 357.
 - ✓ MASSEY, D.S. y M.L. Eggers. (1990) “The Ecology of Inequality. Minorities and the Contemporary Sociology of Poverty 1970 – 1980”, en American Journal of Sociology, Vol. 95, No 5; pp.1153 - 1158.
 - ✓ MASSEY, D .S. y N.A. Denton (1993) American Apartheid. Segregation and the Making of the Underclass. Cambridge. Harvard University Press.
 - ✓ MORAN, Patrick (1948), “ The interpretation of statistical maps”, Journal of the Royal Statistical Society”, vol. 10, 1948, Londres pp. 243-251. Disponible en <http://www.jstor.org/pss/2983777>
 - ✓ MORENO, Rosina; VAYÁ, Esther, *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial*, Universidad de Barcelona, Barcelona, 2000.
 - ✓ NÚÑEZ, Jairo; SÁNCHEZ, Fabio, “ Estimaciones trimestrales de la línea de pobreza y sus relaciones con el desempeño macroeconómico colombiano 1977-1997”, *Archivos de macroeconomía*, nú m. 110, Departamento Nacional de Planeación (DNP), mayo de 1999.
 - ✓ ORD, J.K. y Getis, A. (1995), “Local spatial autocorrelations statistics: distributional issues and an application”, *Geographical Analysis*, No. 27, pp. 286 - 306.
 - ✓ PAELINCK, J.H.P y Klaassen, L. H. (1979), “Spatial Econometrics”. Farnborough, Saxon House.
 - ✓ POSADA, L . J . (1978), “Fundamentos económicos-espaciales de la teoría de Centros de Desarrollo”, *Agricultura y Sociedad*, N° 6, Ministerio de Agricultura y Comercio de España. Pp. 137-180.
 - ✓ QUANDT, R . (1958), “ The estimation of the parameters of a linear regression system obeying two separate regimes”, *Journal of the American Statistical Association*, 53:873-880.
 - ✓ SAMPSON, R.J., S. W. Raudenbush y F. Earls (1997), Neighborhood and Violent Crime. A Multilevel Study of Collective Efficacy” en *Science*, Vol. 227, No.15; pp 918-924

-
- ✓ SASTRY, N., A. Pebley y M. Zonta (2002), "Neighborhood Definitions and the Spatial Dimension of Daily Life in the Angeles".
 - ✓ TOBLER, Waldo (1970), "A Computer Movie Simulating Urban Growth in the Detroit Region", *Economic Geography*, vol. 46, núm. 2, junio, Clark University, EUA, pp. 234-240.
 - ✓ WEEKS, J. et al. (2002), "Spatial Patterns as Predictors of Fertility Change in Rural Egypt".
 - ✓ WILSON, W.J. (1987), *The truly disadvantaged. The inner city, the underclass and public policy*. Chicago. University of Chicago Press.
 - ✓ WILSON, W.J. (1996), *When work disappears. The world of the new urban poor*. Nueva York. Vintage.

Anexo A.Técnico Metodológico.

A. El Índice de Rezago Social

En el documento *Los Mapas de Pobreza en México*, junto con su anexo estadístico y geográfico, se da a conocer los resultados de las estimaciones realizadas por el CONEVAL para la medición multidimensional de la pobreza a nivel estatal y municipal. Los mapas muestran dos medidas de carencias: la pobreza por ingresos y el índice de rezago social. Estas mediciones se construyeron a partir de los indicadores que marca la Ley General de Desarrollo Social y que deben ser considerados en la definición, identificación y medición de la pobreza.

El índice de rezago social se calculó empleando la técnica estadística de componentes principales, la cual permitió combinar información de indicadores de carencias en un índice que sintetiza numéricamente diferentes dimensiones de la pobreza. El índice se calculó para niveles estatal, municipal y localidad.

La construcción del índice de rezago social tiene como intención cumplir con tres criterios básicos:

En primer lugar, se consideró la oportunidad de la información, incorporando la información disponible acorde a los indicadores de pobreza y a los niveles de desagregación que marca la Ley y de la información disponible.

En segundo lugar, se seleccionó una base de datos cuya estructura permitiera obtener indicadores en los niveles de agregación de localidades, municipal, estatal y nacional.

Sobre la base de estos dos primeros criterios se decidió utilizar la base de datos "Principales Resultados por Localidad, 2005" del II Censo de Población y Vivienda (ITER 2005).

En tercer lugar, se optó por la técnica estadística de componentes principales ya que permite resumir en un indicador agregado las diferentes dimensiones de l fenómeno en estudio.⁶ El índice resultante permite ordenar las unidades de observación (localidad, municipio, estado) según sus carencias sociales. Además, esta técnica es la misma que utiliza CONAPO para la construcción del índice de marginación y que ha sido ampliamente utilizado en la definición

Indicadores

Para la construcción del índice de rezago social se consideraron los siguientes indicadores.

Educativos

- Porcentaje de población de 15 años o más analfabeta.

$$I_{analf} = \frac{p_{15maan}}{p_{15ymas}} * 100$$

donde:

p_15maan: Población de 15 años y más analfabeta

p_15ymas: Población de 15 años y más

- Porcentaje de la población de 6 a 14 años que no asiste a la escuela.

$$I_{asistesc} = \frac{p_{6a14noa}}{p_{6a14an}} * 100$$

donde:

p6a14noa: Población de 6 a 14 años que no asiste a la escuela

p_6a14_an: Población de 6 a 14 años

- Porcentaje de los hogares con población de 15 a 29 años, con algún habitante con menos de 9 años de educación aprobados.¹⁷

$$I_{reeduc} = \frac{hog_pob_15_29sin9}{hog_pob_15_29} * 100$$

donde:

hog_pob_15_29sin9: Hogares con población de 15 a 29 años, con algún poblador con menos de 9 años aprobados

hog_pob_15_29: Hogares con población de 15 a 29 años

- Porcentaje de la población de 15 años o más con educación básica incompleta.¹⁸

$$I_{edbasinc} = \frac{p15ymase + p15ym_ebin}{p15ymase + p15ym_ebin + p15ym_ebc + p15ymapb} * 100$$

donde:

p15ymase: Población de 15 años y más sin escolaridad

p15ym_ebin: Población de 15 años y más con educación básica incompleta

p15ym_ebc: Población de 15 años y más con educación básica completa

¹⁷ Este indicador de rezago educativo se empleó para los niveles estatal y municipal.

¹⁸ Este indicador se empleó como indicador de rezago educativo en los niveles de localidad.

p15ymapb: Población de 15 años y más con educación posbásica

Acceso a servicio de salud.

- Porcentaje de la población sin derechohabiencia a servicios de salud.

$$I_{sdersalud} = \frac{p_sinder}{p_total} * 100$$

donde:

p_sinder: Población sin derechohabiencia a servicios de salud

p_total: Población total

Calidad y espacios en la vivienda

Los dos indicadores socioeconómicos considerados en la dimensión de espacio de vivienda son:

- Porcentaje de las viviendas particulares habitadas con piso de tierra.

$$I_{ptierra} = \frac{vph_con_pt}{vivparha} * 100$$

donde:

vph_con_pt: Viviendas particulares habitadas con piso de tierra

vivparha: Viviendas particulares habitadas

- Promedio de ocupantes por cuarto.¹⁹

$$I_{hacin} = pro_c_vp$$

¹⁹ En la aplicación de la metodología de componentes principales, se ocupó el algoritmo natural del indicador de hacinamiento.

donde:

pro_c_vp: Promedio de ocupantes por cuarto en viviendas particulares habitadas.

Servicios básicos en la vivienda

- Porcentaje de las viviendas particulares habitadas que no disponen de excusado o sanitario.

$$I_{nosan} = 1 - \frac{vph_excsa}{vivparha} * 100$$

donde:

vph_excsa: Viviendas particulares habitadas que no disponen de excusado o sanitario.

- Porcentaje de las viviendas particulares habitadas que no disponen de agua entubada de la red pública.

$$I_{noagua} = \frac{vph_noag}{vivparha} * 100$$

vph_noag: Viviendas particulares habitadas que no disponen de agua entubada de la red pública

- Porcentaje de las viviendas particulares habitadas que no disponen de drenaje.

$$I_{nodren} = \frac{vph_nodren}{vivparha} * 100$$

donde:

vph_nodren: Viviendas particulares habitadas que no disponen de drenaje

- Porcentaje de las viviendas particulares habitadas que no disponen de energía eléctrica.

$$I_{noenel} = \frac{vph_enel}{vivparha} * 100$$

donde:

vph_enel: Viviendas particulares habitadas que disponen de energía eléctrica.

Activos en el hogar

- Porcentaje de las viviendas particulares habitadas que no disponen de lavadora.

$$I_{notlav} = 1 - \frac{vph_lava}{vivparha} * 100$$

donde:

vph_lava: Viviendas particulares habitadas que disponen de lavadora.

- Porcentaje de las viviendas particulares habitadas que no disponen de refrigerador.

$$I_{noref} = 1 - \frac{vph_refr}{vivparha} * 100$$

donde:

vph_refr: Viviendas particulares habitadas que disponen de refrigerador.

Resultados.

En el Cuadro 1 se presentan los resultados de la aplicación de la técnica de componentes principales, de acuerdo a la varianza explicada por cada componente.

Cuadro 1. Proporción de la varianza explicada por cada componente

Componente	Nivel de Agregación		
	Localidad	Municipal	Estatad
1	0.4005	0.5487	0.6760
2	0.0880	0.1015	0.0983
3	0.0796	0.0847	0.0640
4	0.0721	0.0669	0.0500
5	0.0656	0.0474	0.0434
6	0.0611	0.0422	0.0276
7	0.0532	0.0321	0.0214
8	0.0466	0.0228	0.0065
9	0.0447	0.0185	0.0054
10	0.0382	0.0161	0.0034
11	0.0342	0.0140	0.0026
12	0.0162	0.0050	0.0013

Fuente: Estimaciones del CONEVAL con base en el *ITER 2005*.

El Índice de Rezago Social se construye como una suma ponderada de los diferentes indicadores. Para su construcción se utilizan como ponderadores los coeficientes de la primera componente. Éstos se muestran en el cuadro 2 para los diferentes niveles de desagregación. El índice es estandarizado de tal forma que su media sea cero y su varianza unitaria.

En virtud de que el Índice de Rezago Social, por la forma en la cual se construye, cumple con el propósito de ordenar las diferentes unidades de observación (localidades, municipios y estados), estas se estratificaron en cinco categorías, de tal forma que dentro de cada categoría las unidades fueran lo más homogéneas

posibles y entre los estratos lo más distintos posibles. La técnica de estratificación de Dalenius & Hodges cumple con este propósito.

De esta forma se crearon cinco estratos que son: muy bajo, bajo, medio, alto y muy alto rezago social, y que dan la idea de grupos de localidades, municipios y estados que van de aquellos que muestran un menor a una mayor carencia en los indicadores que conforman el índice, respectivamente.

Cuadro 2. Coeficientes de las variables utilizados en la construcción del puntaje

Variable	Nivel de Agregación		
	Estatad	Municipal	Localidad
% Población de 15 años y más analfabeta	0.3401	0.3387	0.2900
% Población de 6 a 14 años que no asiste a la escuela	0.2644	0.1903	0.1808
% Hogares con población de 15 a 29 años, con algún habitante con menos de 9 años de educación aprobados ¹	0.3036	0.3202	0.2853
% Población sin derechohabiencia a servicios de salud	0.2992	0.2427	0.1951
Promedio de ocupantes por cuarto ²	0.3170	0.3259	0.2782
% Viviendas particulares habitadas con piso de tierra	0.3222	0.3389	0.3371
% Viviendas particulares habitadas que no disponen de excusado o sanitario	0.2117	0.1448	0.2729
% Viviendas particulares habitadas que no disponen de agua entubada de la red pública	0.3002	0.2481	0.2117
% Viviendas particulares habitadas que no disponen de drenaje	0.2858	0.3031	0.3066

Variable	Nivel de Agregación		
	Estatad	Municipal	Localidad
% V viviendas particulares habitadas que no disponen de energía eléctrica	0.1269	0.2303	0.3089
% V viviendas particulares habitadas que no disponen de refrigerador	0.3097	0.3486	0.3765
% V viviendas particulares habitadas que no disponen de lavadora	0.3161	0.3429	0.3498

^{/1} Para el caso de las estimaciones a nivel localidad, se utilizó el % de hogares con población de 15 años y más que no completó la educación básica.

^{/2} Para la construcción del índice de rezago social se utilizó el logaritmo natural del promedio de ocupantes por cuarto.

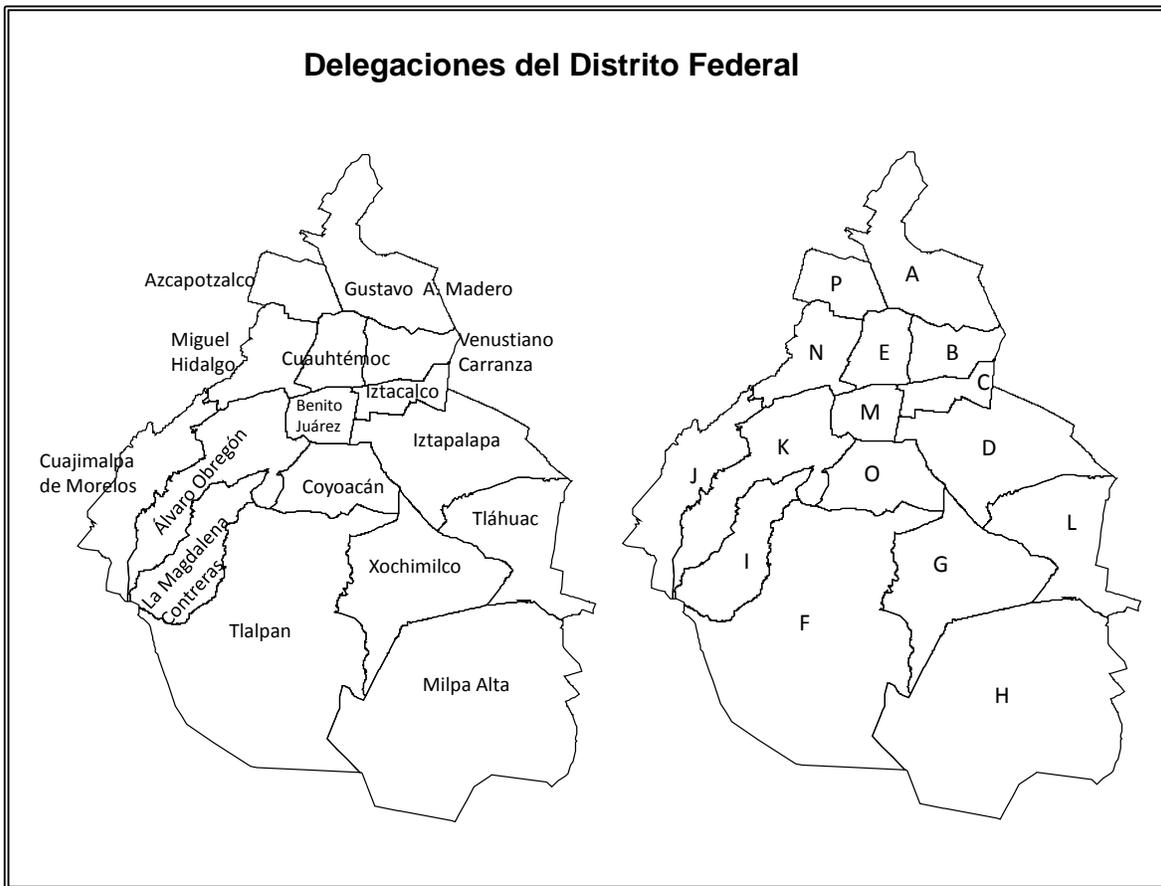
Fuente: Estimaciones del CONEVAL con base en el ITER 2005.

Anexo B. Ejemplo del cálculo de la autocorrelación espacial

B. Autocorrelación espacial en las delegaciones del Distrito Federal, de lo local a lo global.

Se ejemplificará el funcionamiento del procedimiento de AE que lleva a cabo el programa GeoDa internamente, se utiliza el Índice de rezago social 2005 y se aplicó a las 16 delegaciones del Distrito Federal.

Mapa B.1 Asignación de las delegaciones del Distrito Federal por polígono.



Fuente: elaboración propia

Índice de Rezago Social 2005.

Delegación	IRS
A Gustavo A. Madero	-1.634
B Venustiano Carranza	-1.598
C Iztacalco	-1.659
D Iztapalapa	-1.468
E Cuauhtémoc	-1.621
F Tlalpan	-1.520
G Xochimilco	-1.324
H Milpa Alta	-1.000
I La Magdalena Contreras	-1.534
J Cuajimalpa de Morelos	-1.542
K Álvaro Obregón	-1.631
L Tláhuac	-1.463
M Benito Juárez	-1.960
N Miguel Hidalgo	-1.782
O Coyoacán	-1.792
P Azcapotzalco	-1.765

Fuente: elaboración propia

El primer paso es establecer la matriz de pesos a partir de la cantidad de vecinos que tiene cada delegación. Se utilizó la relación de contigüidad tipo torre (*Rook*). Por ejemplo, la delegación A tiene tres vecinos a los que les corresponde un 1 en la matriz y 0 al resto.

Matriz de pesos

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Total
A	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	3
B	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	3
C	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	4
D	0	0	1	0	0	0	1	0	0	0	0	1	1	0	1	0	5
E	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	1	6
F	0	0	0	0	0	0	1	1	1	0	1	0	0	0	1	0	5
G	0	0	0	1	0	1	0	1	0	0	0	1	0	0	1	0	5
H	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	3
I	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Total
J	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	2
K	0	0	0	0	0	1	0	0	1	1	0	0	1	1	1	0	6
L	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	3
M	0	0	1	1	1	0	0	0	0	0	1	0	0	1	1	0	6
N	0	0	0	0	1	0	0	0	0	1	1	0	1	0	0	1	5
O	0	0	0	1	0	1	1	0	0	0	1	0	1	0	0	0	5
P	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	3
Suma																	68

Fuente: elaboración propia

A continuación se estandarizan los pesos para que cada fila suma 1. Este procedimiento no es necesario pero facilita la interpretación de las estadísticas (Anselin, 1995: 95).

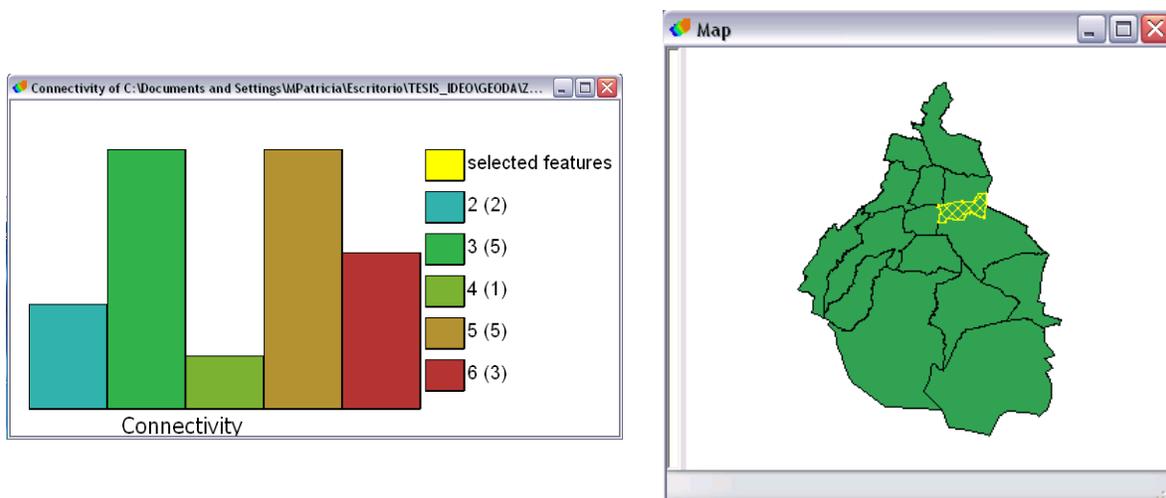
Matriz de pesos estandarizada

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Total
A	0	0.33	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0.33	1
B	0.33	0	0.33	0	0.33	0	0	0	0	0	0	0	0	0	0	0	1
C	0	0.25	0	0.25	0.25	0	0	0	0	0	0	0	0.25	0	0	0	1
D	0	0	0.20	0	0	0	0.20	0	0	0	0	0.20	0.20	0	0.20	0	1
E	0.17	0.17	0.17	0	0	0	0	0	0	0	0	0	0.17	0.17	0	0.17	1
F	0	0	0	0	0	0	0.20	0.20	0.20	0	0.20	0	0	0	0.20	0	1
G	0	0	0	0.20	0	0.20	0	0.20	0	0	0	0.20	0	0	0.20	0	1
H	0	0	0	0	0	0.33	0.33	0	0	0	0	0.33	0	0	0	0	1
I	0	0	0	0	0	0.50	0	0	0	0	0.50	0	0	0	0	0	1
J	0	0	0	0	0	0	0	0	0	0	0.50	0	0	0.50	0	0	1
K	0	0	0	0	0	0.17	0	0	0.17	0.17	0	0	0.17	0.17	0.17	0	1
L	0	0	0	0.33	0	0	0.33	0.33	0	0	0	0	0	0	0	0	1
M	0	0	0.17	0.17	0.17	0	0	0	0	0	0.17	0	0	0.17	0.17	0	1
N	0	0	0	0	0.20	0	0	0	0	0.20	0.20	0	0.20	0	0	0.20	1
O	0	0	0	0.20	0	0.20	0.20	0	0	0	0.20	0	0.20	0	0	0	1
P	0.33	0	0	0	0.33	0	0	0	0	0	0	0	0	0.33	0	0	1
Suma																	16

Fuente: elaboración propia

Para confirmar la cantidad de vecinos que tiene cada delegación, se hace uso del histograma de conectividad²⁰, el cual tiene todas las propiedades de un histograma estándar y puede ser vinculado a los otros puntos de vista en un proyecto.

Por ejemplo, en la siguiente figura, la única delegación que tiene cuatro vecinos es Iztacalco (polígono C), esto se muestra al seleccionar la barra con mayor número de vecinos contiguos - 4(1) -, e l efecto se confirma con los resultados generados de la matriz de pesos.



En base a la matriz estandarizada se puede aplicar la formula (8) para conocer el valor local de I de Moran para cada unidad espacial (delegación) y conocer como contribuye al valor global:

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^n W_{ij} (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2 / n} \quad \text{para } i \neq j$$

²⁰ Herramienta del software GeoDa.

$$I_A = \left(\frac{-1.634 - (-1.581)}{.7127} \right) * [(0.333 x - 0.017) + (0.333 x - 0.040) + (0.333 x - 0.184)] = 0.0060$$

$$I_B = \left(\frac{-1.598 - (-1.581)}{.7127} \right) * [(0.333 x - 0.053) + (0.333 x - 0.078) + (0.333 x - 0.040)] = -0.0014$$

$$I_C = \left(\frac{-1.659 - (-1.581)}{.7127} \right) * \left[\frac{(0.25 x - 0.017) + (0.25 x 0.113) +}{(0.25 x - 0.040) + (0.25 x - 0.379)} \right] = 0.0088$$

$$I_D = \left(\frac{-1.468 - (-1.581)}{.7127} \right) * \left[\frac{(0.20 x - 0.078) + (0.20 x 0.257) + (0.20 x 0.118) +}{(0.20 x - 0.379) + (0.20 x - 0.211)} \right] = -0.0093$$

$$I_E = \left(\frac{-1.621 - (-1.581)}{.7127} \right) * \left[\frac{[(0.17 x - 0.053) + (0.17 x - 0.017) + (0.17 x - 0.078) +]}{(0.17 x - 0.379) + (0.17 x - 0.2011) + (0.17 x - 0.184)} \right] = 0.0085$$

$$I_F = \left(\frac{-1.520 - (-1.581)}{.7127} \right) * \left[\frac{(0.20 x 0.257) + (0.20 x 0.580) + (0.20 x 0.047) +}{(0.20 x - 0.050) + (0.20 x - 0.211)} \right] = 0.0106$$

$$I_G = \left(\frac{-1.324 - (-1.581)}{.7127} \right) * \left[\frac{(0.20 x 0.113) + (0.20 x 0.061) + (0.20 x 0.580) +}{(0.20 x 0.118) + (0.20 x - 0.211)} \right] = 0.0476$$

$$I_H = \left(\frac{-1.000(-1.581)}{.7127} \right) * [(0.333 x 0.061) + (0.333 x 0.257) + (0.333 x 0.118)] = 0.1181$$

$$I_I = \left(\frac{-1.534(-1.581)}{.7127} \right) * [(0.50 x 0.61) + (0.50 x - 0.050)] = 0.0004$$

$$I_J = \left(\frac{-1.542 - (-1.581)}{.7127} \right) * [(0.50 x 0.047) + (0.50 x - 0.201)] = -0.0068$$

$$I_K = \left(\frac{-1.631 - (-1.581)}{.7127} \right) * \left[\frac{(0.17 x 0.061) + (0.17 x 0.047) + (0.17 x 0.039) +}{(0.17 x - 0.379) + (0.17 x - 0.201) + (0.17 x - 0.211)} \right] = 0.0075$$

$$I_L = \left(\frac{-1.463 - (-1.581)}{.7127} \right) * [(0.333 x 0.113) + (0.333 x 0.275) + (0.333 x 0.580)] = 0.0523$$

$$I_M = \left(\frac{-1.960 - (-1.581)}{.7127} \right) * \left[\frac{(0.17 x - 0.078) + (0.17 x 0.113) + (0.17 x - 0.040) +}{(0.17 x - 0.050) + (0.17 x - 0.201) + (0.17 x - 0.211)} \right] = 0.0414$$

$$I_N = \left(\frac{-1.782 - (-1.581)}{.7127} \right) * \left[\frac{(0.20 x - 0.040) + (0.20 x - 0.039) + (0.20 x - 0.050) +}{(0.20 x - 0.379) + (0.20 x - 0.184)} \right] = 0.0347$$

$$I_o = \left(\frac{-1.792 - (-1.581)}{.7127} \right) * \left[\frac{(0.20 \times 0.113) + (0.20 \times 0.061) + (0.20 \times 0.257)}{(0.20 \times -0.050) + (0.20 \times -0.379)} \right] = 0.0185$$

$$I_p = \left(\frac{-1.765 - (-1.581)}{.7127} \right) * [(0.333 \times -0.053) + (0.333 \times -0.046) + (0.333 \times -0.201)] = 0.0253$$

La suma de los valores locales conforman el valor global de *I* de Moran:

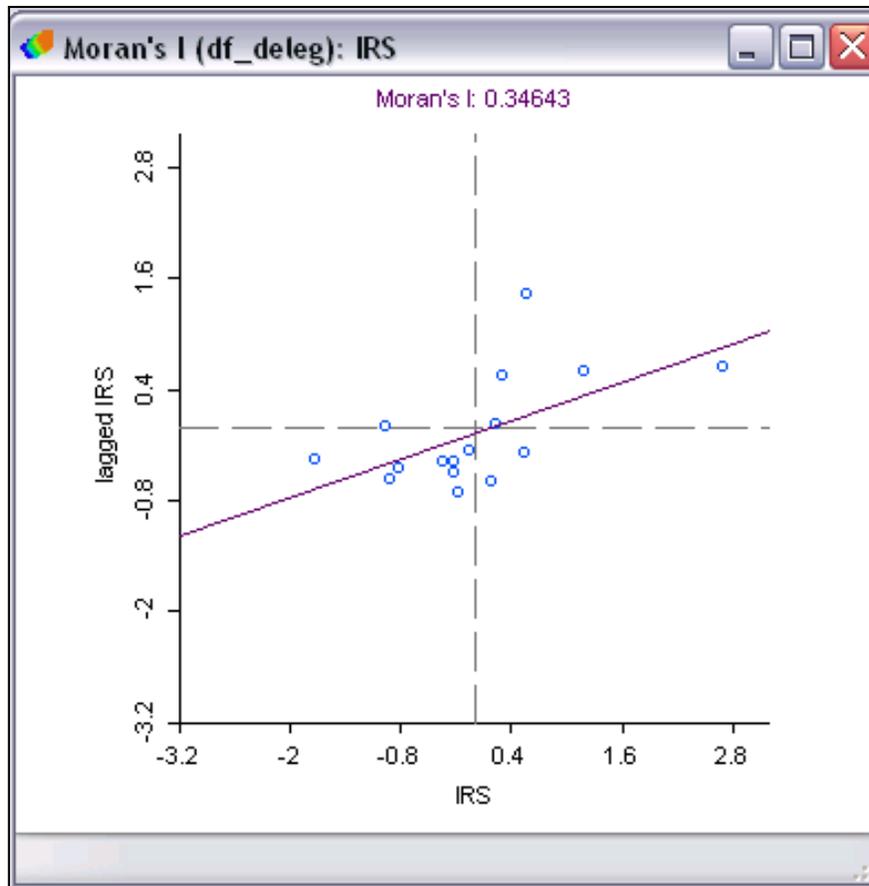
$$\sum_{i=1}^n I_i = \sum_{i=1}^n Z_i \sum_{j=1}^n W_{ij} Z_j$$

Es decir,

$$I = I_A + I_B + I_C + I_D + I_E + I_F + I_G + I_H + I_I + I_J + I_K + I_L + I_M + I_N + I_O + I_P$$

$$I = 0.346$$

El resultado se puede visualizar en la parte superior del *scatterplot*, además, permite ubicar en los cuadrantes los valores estandarizados de la variable en cuestión para cada unidad espacial.



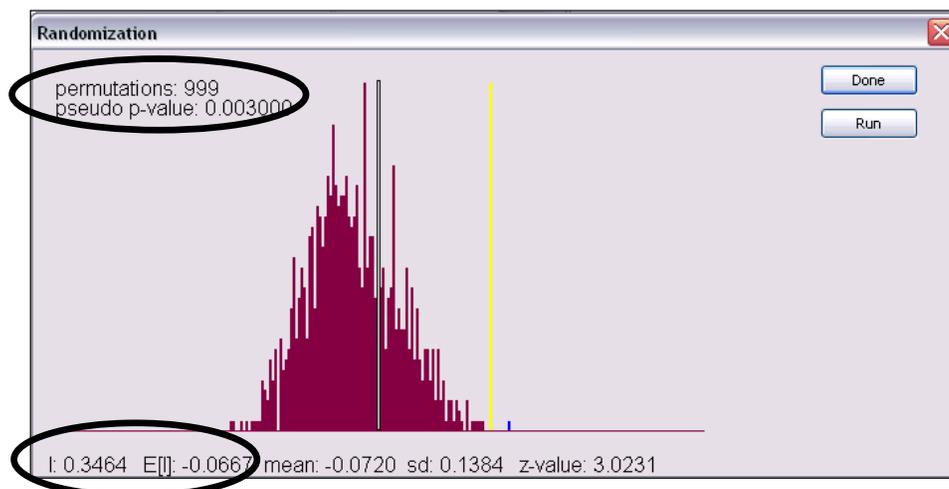
Fuente: elaboración propia con base al programa GeoDa

Con la formula (6), se calcula el valor esperado de I de Moran:

$$E[I] = -\frac{1}{15} = -0.0667$$

La otra alternativa para probar si la configuración es aleatoria, es por medio de las permutaciones ya explicadas y que utiliza el GeoDa. Esta prueba se basa en la idea de que si hay n observaciones sobre una región en particular, $n!$ permutaciones de los datos son posibles². El valor de I puede obtenerse por cualquiera de las permutaciones. Luego se puede comparar el valor observado de I respecto a la distribución de permutaciones para determinar si es significativo (Lloyd, 2007).

En la siguiente figura se puede observar el valor global de la I de Moran y su valor esperado, $E[I]$ calculado por medio del procedimiento de aleatorización. A través de las permutaciones se generan 999 nuevos valores de I de Moran que componen una distribución de la cual se destaca la línea que determina el promedio de la misma. En la parte inferior izquierda del gráfico se registran el valor observado $I: 0.3464$ y el esperado: -0.0667 . Ambos resultados coinciden con los cálculos realizados con anterioridad.

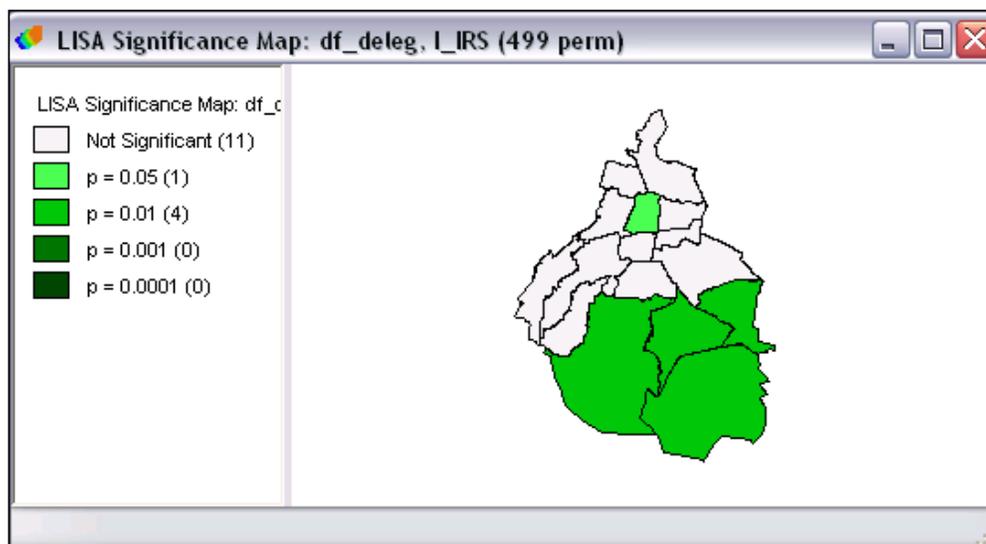


Fuente: elaboración propia con base al programa GeoDa

En este caso el (p -value) es de 0.003 y por ser menor a 0.05 se rechaza la hipótesis nula y se acepta la alternativa. De todas maneras, los valores son cercanos de manera que hay que tomarlos con cautela, recordando nuevamente la finalidad explicativa que tiene el presente trabajo. Como los resultados no son azarosos continuamos explorando la información de las unidades espaciales por medios de los mapas de significancia y *cluster*.

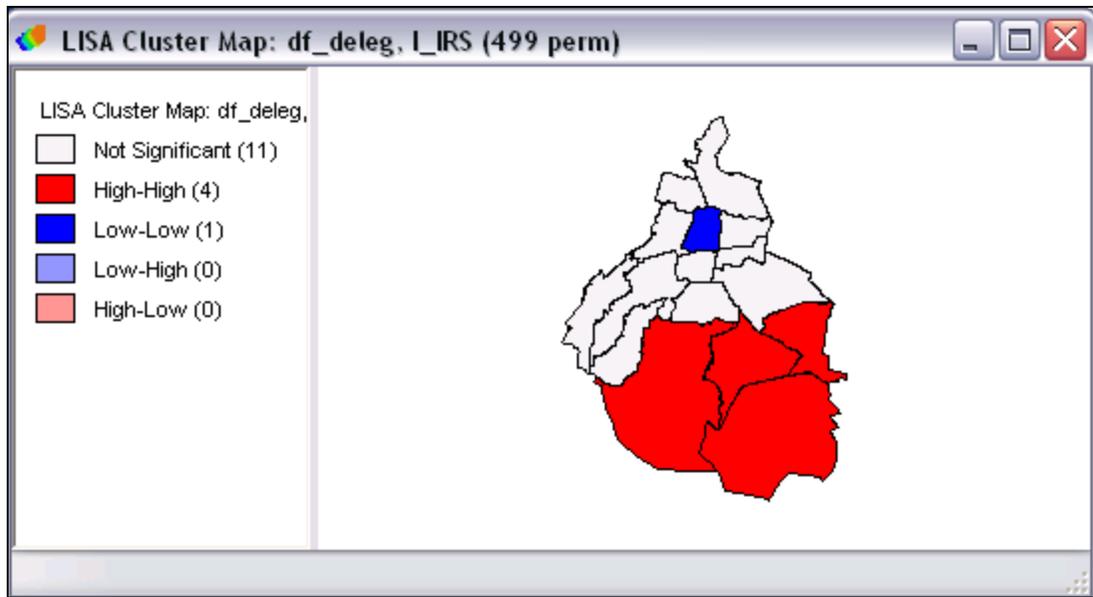
En la siguiente figura se muestra la probabilidad de que las relaciones de contigüidad se produzcan de manera aleatoria y se destaca una delegación significativa; E (Cuauhtémoc) con un valor que denota una probabilidad de error de 0.05 (5 en 100). Con un valor que denota una probabilidad de error de 0.01 (1 en 100) se encuentran cuatro delegaciones; F (Tlalpan), G (Xochimilco), L

(Tláhuac) y H (Milpa Alta). En el resto predomina el color blanco que denota la falta de significancia.



Fuente: elaboración propia con base al programa GeoDa

En la siguiente figura, se hace presente dos categorías de las cinco posibles. Predomina el color blanco que significa que los valores no son significativos para las delegaciones que lo poseen. Se registran dos *clusters*, cuatro delegaciones pertenecen al cluster (alto - alto) F, G, L y H (Tlalpan, Xochimilco, Tláhuac y Milpa Alta) respectivamente y un segundo cluster de (bajo - bajo) E (Cuauhtémoc) respectivamente que ya aparecen en el mapa anterior.



Fuente: elaboración propia con base al programa GeoDa.