



MÁSTERES de la UAM

Facultad de Ciencias
Económicas y
Empresariales /13-14

Máster en
Administración de
Empresas (MBA)



excelencia Campus Internacional
UAM
CSIC+



**Estudio del
mercado espacial
de los colegios pri-
vados y concertados
en la ciudad de
Madrid**

*Adrián Vázquez
Aparicio*



ÍNDICE DE CONTENIDOS

	Pág.
ÍNDICE DE TABLAS	4
ÍNDICE DE FIGURAS	4
AGRADECIMIENTOS	6
1 INTRODUCCIÓN	8
2 OBJETIVOS Y JUSTIFICACIÓN	9
3 REVISIÓN DE LA LITERATURA	11
3.1 Introducción	11
3.2 Colegios públicos y privados: características y diferencias en términos de eficiencia	11
3.2.1 Situación en los países de la OCDE	11
3.2.2 Situación en España	14
3.3 Análisis Exploratorio de Datos Espaciales	19
3.3.1 Análisis exploratorio básico	19
3.3.2 Concepto de autocorrelación	20
3.3.3 Autocorrelación global	22
3.3.4 Autocorrelación local.....	22
3.4 Modelo de regresión lineal	23
3.4.1 Estimación por Mínimos Cuadrados Ordinarios (MCO)	24
3.4.2 Medidas de bondad del ajuste	25
3.4.3 Contrastes de hipótesis del modelo de regresión lineal	27
3.5 Diagnósticos de especificación	28
3.5.1 Multicolinealidad	28

3.5.2 Errores no normalmente distribuidos	29
3.5.3 Heteroscedasticidad	29
3.5.4 Autocorrelación espacial.....	31
3.6 El modelo del retardo espacial.....	33
3.6.1 Definición	33
3.6.2 Estimación por el método de Variables Instrumentales (VI).....	34
3.7 El modelo del error espacial.....	36
3.7.1 Definición	36
3.7.2 Estimación por el Método General de los Momentos (GMM).....	37
4 METODOLOGÍA	38
4.1 Definición de la muestra y las variables.....	38
4.1.1 Selección de la muestra.....	38
4.1.2 Descripción de las variables	40
4.2 Análisis exploratorio espacial de la variable de distancias.....	43
4.3 El modelo empírico	50
4.3.1 Especificación del modelo inicial.....	50
4.3.2 Estimación y contraste del modelo inicial	52
4.3.3 Especificación y estimación de modelos alternativos	54
5 RESULTADOS.....	55
6 CONCLUSIONES	56
7 BIBLIOGRAFÍA.....	58
8 ANEXOS.....	63
8.1 Relación de colegios privados y concertados del municipio de Madrid por distritos	63
8.2 Análisis exploratorio espacial de las variables del modelo.....	70
8.2.1. Construcción de las matrices de vecindad.....	70
8.2.2. Análisis de autocorrelación espacial de las variables	73
8.2.3. Análisis de causalidad entre la variable dependiente y las independientes.....	83

ÍNDICE DE TABLAS

	Pág.
Tabla 3.1.....	14
Tabla 4.1.....	42
Tabla 4.2.....	51
Tabla 4.3.....	53
Tabla 8.1.....	63

ÍNDICE DE FIGURAS

	Pág.
Figura 4.1.....	40
Figura 4.2.....	44
Figura 4.3.....	46
Figura 4.4.....	48
Figura 4.5.....	49
Figura 8.1.....	70
Figura 8.2.....	71
Figura 8.3.....	72
Figura 8.4.....	73
Figura 8.5.....	74
Figura 8.6.....	75
Figura 8.7.....	76

Figura 8.8.....	77
Figura 8.9.....	79
Figura 8.10.....	81
Figura 8.11.....	82
Figura 8.12.....	82
Figura 8.13.....	83
Figura 8.14.....	84
Figura 8.15.....	84
Figura 8.16.....	85
Figura 8.17.....	86
Figura 8.18.....	87
Figura 8.19.....	88
Figura 8.20.....	89
Figura 8.21.....	90
Figura 8.22.....	91
Figura 8.23.....	92
Figura 8.24.....	93
Figura 8.25.....	94
Figura 8.26.....	95
Figura 8.27.....	96
Figura 8.28.....	97
Figura 8.29.....	98
Figura 8.30.....	99
Figura 8.31.....	100
Figura 8.32.....	101
Figura 8.33.....	102

AGRADECIMIENTOS

Ya contemplo muy lejos aquel mes de Septiembre en el que comencé con esta andadura, la cual desconocía en su gran mayoría y, siendo honestos, me asustaba bastante. Cuando comencé el MBA, allá por Enero de 2013, nunca pensé que mi Trabajo de Fin de Master trataría de este tema en concreto. Tenía claro que mi futuro estaba ligado a la rama empresarial del Marketing, pero nunca imaginé realizar un trabajo tan a fondo sobre Geomarketing y Geolocalización. Una vez abordado y empleado tantas y tantas horas, he de decir que he descubierto un campo muy amplio y potente que puede ser muy útil en mi vida profesional y abrirme muchas puertas en el futuro.

Primero me gustaría agradecer a toda la gente que ha pasado por el MBA y he conocido durante este año. Desde profesores como Emilio Ontiveros, por su gran sabiduría y representatividad en el panorama económico español, pasando por Yolanda Bueno, sabiendo decirnos a cada uno nuestros fallos y virtudes y acercándonos a la realidad de los proyectos empresariales en todo lo posible, a Begoña Santos e Isidro de Pablo, por llevarnos a un mundo virtual muy próximo a la realidad profesional de las empresas a través de su simulador, y sin olvidarnos de Hermógenes del Real y sus grandes momentos de hacernos interiorizar las habilidades necesarias para un directivo mediante sus reflexiones enriquecedoras. También agradecer a Cristina Estaban su atención como coordinadora del MBA durante todo el curso y, en especial, durante la búsqueda de prácticas empresariales y posterior tutora de la Memoria de dichas prácticas. A Fernando Casani y Jesús Pomedá, por su paciencia y flexibilidad, siendo conscientes de la dificultad de este proyecto. Se me quedan otros tantos en el tintero a los cuales estoy igualmente agradecido por todo lo que me han aportado en este Máster.

No puedo olvidarme de mis compañeros, en especial de gente como Borja, Pablo, Belén, Confang, Daniel o Álvaro, que he seguido teniendo a mi lado este año, en el cual no nos veíamos a menudo, pero que nos hemos estado apoyando unos a otros para sacar adelante nuestros proyectos.

Mi familia ha sido clave en esta aventura, en especial mis padres, José María y Begoña, ya que Madrid era nuevo para mí, lejos de ellos, pero siempre han sabido estar ahí en los malos momentos, y también en los buenos, para transmitirme toda la fuerza que necesitaba.

He querido dejarla para lo último, ya que sin ella esto habría sido totalmente imposible, de hecho, ni siquiera tendría constancia de todo lo que he aprendido, Coro Chasco. Primero agradecerle sus majestuosas clases de métodos cuantitativos, haciendo que los números y la estadística fuesen más amigables de lo que parecen a simple vista. Gracias a ello un día se le ocurrió ofrecernos un tema para este trabajo, Geomarketing y Geolocalización, algo me transmitió, y aquí estamos ahora, me tiré a la piscina por algo desconocido, pero muy atractivo. Gracias por estar semana tras semana motivándome y despertando en mí un gran interés por esta materia. A parte de todo lo aprendido en la materia tratada, me ha enseñado lo que significa la constancia, sin prisa pero sin pausa, y, sobre todo, la pasión por ponerse delante del ordenador y trabajar hasta conseguir una gran meta. Estaré eternamente agradecido.

1 INTRODUCCIÓN

Con las actuales políticas de privatización del sector educativo en ciertas comunidades autónomas, hemos visto importante estudiar de forma exacta y numérica, y no por posiciones políticas o ideológicas, la accesibilidad espacial a los colegios privados y concertados por parte de todas las clases sociales. Nos hemos centrado en la ciudad de Madrid, ya que muestra un gran porcentaje de este tipo de colegios, que viene incrementándose desde el año 2010 y, por lo tanto, es una nueva línea de negocio muy interesante.

En este estudio, comenzamos analizando la diferenciación entre colegios públicos, privados y concertados, tanto en el ámbito de la OCDE como, más concretamente, de España. Así pues, se analiza, más detalladamente, que los colegios privados y concertados complementan la oferta educativa de los colegios públicos. Los centros concertados son financiados por el Estado, en parte, mientras que los colegios privados carecen de financiación alguna por parte del mismo. (Dronkers, y Robert, 2003; Mancebon y Muñiz ,2008; Bernal ,2005).

Para poder analizar los datos obtenidos utilizaremos el análisis exploratorio de datos espaciales (AEDE), el cual está especialmente capacitado para la descripción y visualización de las distribuciones espaciales. Se trata de un grupo de técnicas capaces de identificar valores espaciales atípicos, esquemas de asociación espacial o agrupamientos ("clusters"), puntos calientes/fríos ("hot spots") u otras formas de heterogeneidad espacial (Anselin, 1999a). Por tanto, el AEDE tiene un carácter descriptivo (estadístico) más que confirmatorio (econométrico), pero, tras la detección de estructuras espaciales en las variables geográficas, hará posible la formulación de hipótesis previas para la modelización econométrica y, en su caso, posible predicción espacial de nuevos datos. (Chasco y Fernández-Avilés, 2009).

También hay que hacer hincapié en el modelo de regresión lineal, para encontrar una relación entre una variable dependiente y un conjunto de variables explicativas: los diagnósticos de especificación, grado en que los supuestos términos de error normal, homoscedásticos y no correlacionados pueden ser violados; así como los modelos de retardo y el error espacial (Anselin, 1992).

Como ya se ha indicado, el objetivo del estudio respondería a la siguiente cuestión:

La accesibilidad espacial a los colegios privados y concertados ¿es más fácil para las clases sociales más acomodadas, en detrimento de los más pobres?

Para dar respuesta a dicha cuestión, hemos aplicado un nuevo enfoque metodológico, utilizando técnicas de geomarketing. Debe indicarse que, hasta donde sabemos, no existen estudios cuantitativos que hayan estudiado este asunto en España (sí los hay en otros países). Por ello, lo que pretendemos es dar una respuesta más científica, que no esté matizada por el debate político e ideológico, que es el que domina en los estudios científicos realizados sobre los colegios en España.

La organización de este Trabajo Fin de Máster es la siguiente: en el capítulo 2, se presentan los objetivos y justificación del trabajo; en el capítulo 3, se realiza una revisión de la literatura sobre los dos aspectos centrales del trabajo: la situación de los colegios y el análisis estadístico-econométrico espacial utilizado en el apartado empírico; el capítulo 4 presenta la metodología concreta empleada para la realización de este estudio y el capítulo 5, los resultados obtenidos; finalmente, los capítulos 6, 7 y 8 presentan las conclusiones, la bibliografía y los Anexos, respectivamente.

2 OBJETIVOS Y JUSTIFICACIÓN

La pregunta principal es importante para dos tipos de agentes: por un lado, los dueños o promotores de colegios privados o concertados, que buscan un determinado perfil de "cliente"; por otro lado, los agentes político-sociales, que pretenden ofertar una educación de calidad para todos (ricos y pobres).

Lo que se espera es que la mejor accesibilidad a los colegios privados y concertados se dé en las clases sociales más "acomodadas" (mayor nivel de renta y mayor nivel educativo).

El artículo en el que se basa nuestra pregunta de investigación no es otro que Bernal (2005), el cual nos cuenta la situación de los colegios tanto públicos como privados y concertados en la ciudad de Zaragoza, la cual, según nos explica el propio autor, es una ciudad que posee las características comunes de la mayoría de las ciudades españolas.

El artículo está más enfocado hacia el punto de vista político o ideológico, pero lo que realmente nos interesa a nosotros, es la diferenciación que hace de los tres tipos de colegios en España,

explicando su funcionamiento, financiación, admisión, composición, clase de alumnos y familias representadas en ellos e, incluso, la forma en que los padres eligen el colegio dependiendo que sea público, privado o concertado. Otro análisis muy interesante en dicho artículo es la descripción que realiza este autor de las clases sociales que predominan en cada uno de los tipos de colegios en nuestro país.

Además del artículo del Bernal, también hemos tenido muy en cuenta los artículos Dronkers y Robert (2003) y Mancebón y Muñiz (2008), los cuales nos cuentan lo mismo que Bernal, pero enfocado en todos los países de la OCDE. El artículo de Mancebón y Muñiz, también nos hace una comparativa de eficiencia muy interesante entre ambos tipos de colegios.

Tal como se ha indicado anteriormente, en este trabajo se pretende dar respuesta a esta pregunta utilizando métodos cuantitativos; en concreto, métodos propios del geomarketing. El geomarketing nace de la confluencia entre marketing y geografía. Ocupa ese lugar común entre geografía y marketing, aunque se trata de un concepto más amplio que engloba otros elementos y ciencias, como la informática, estadística y cartografía. El geomarketing podría ser definido como un sistema integrado por datos, programas informáticos de tratamiento, métodos estadísticos y representaciones gráficas destinado a producir una información útil para la toma de decisiones, a través de instrumentos que combinan la cartografía digital, gráficos y tablas (Latour y Floc'h, 2001).

En nuestro caso, hemos utilizado un sistema geográfico de información para medir las diferencias de accesibilidad y medidas de distancia (euclídea, distancia bloque de Manhattan, etc.) a los distintos colegios para las 2.334 secciones censales de la ciudad de Madrid. Para ello, vamos a especificar y estimar modelos de regresión espacial a colegios privados, concertados y públicos para ajustar la autocorrelación espacial y probar el efecto de la clase social en el área de mercado de los diferentes tipos de escuelas. De todo ello tomaremos constancia de forma teórica en los apartados 3.3 y 3.4, y de forma aplicada a nuestro caso en los apartados 4.2 y 4.3.

Este trabajo busca comprobar empíricamente si las afirmaciones hechas hasta el momento, a partir de análisis cualitativo (observación, "conocimiento", encuestas...), se cumplen o no. Además de ello, también se busca cuantificar qué otras variables (además o en lugar de la renta per cápita) inciden sobre la mejor o peor accesibilidad de la población a los colegios privados y concertados, de cara a poder ofrecer un mejor asesoramiento a todos aquellos promotores interesados en iniciar un negocio de estas características.

Por tanto, a partir de la literatura consultada, podrían formularse las siguientes hipótesis de partida, que se contrastarán a lo largo de este trabajo:

1. La accesibilidad a los colegios privados y concertados no es similar para todas las secciones censales: las secciones en las que predominan las familias de clase alta y media-alta tienen una mejor accesibilidad que las más empobrecidas.
2. Aunque todo el mundo tiene los mismos derechos en España, no todos gozan de las mismas condiciones o posibilidades, ya que según está demostrado los colegios privados son mejores que los colegios públicos en términos de nivel de logros (debido al mayor nivel de disciplina y atención individual).

Una consecuencia de política social sería la identificación de aquellas secciones censales en desventaja económica con objeto de potenciar su desarrollo socioeconómico, como medida crucial para prevenir el fracaso escolar.

3 REVISIÓN DE LA LITERATURA

3.1 INTRODUCCIÓN

Este apartado está dividido en tres partes. La primera de ella (apartado 3.2), se dedica a presentar la situación de los colegios (centros educativos de primaria y secundaria) en la OCDE y en España, analizando las diferencias entre centros públicos y privados. Los apartados 3.3 y 3.4 se dedican a presentar los principios básicos del llamado “análisis exploratorio de datos espaciales” y la regresión espacial, respectivamente, que son las principales herramientas estadístico-econométricas utilizadas en este trabajo.

3.2 COLEGIOS PÚBLICOS Y PRIVADOS: CARACTERÍSTICAS Y DIFERENCIAS EN TÉRMINOS DE EFICIENCIA

3.2.1 Situación en los países de la OCDE

Como consecuencia del conflicto entre la iglesia y el estado en muchas sociedades europeas, los colegios privados de hoy en día pueden tener una diferente relación con el estado. El aspecto fundamental de esta relación es el grado por el cual los colegios privados son financiados por el

gobierno (local, regional o nacional), junto con las cuotas estudiantiles, donaciones, patrocinios y la recaudación de fondos de los padres. En muchas sociedades, los colegios privados tienen un derecho jurídico a la financiación por parte del estado, siempre que cumplan ciertas condiciones. En algunos casos este derecho está recogido en una ley constitucional (Alemania, Holanda), mientras que en otros este derecho es acordado por ley normal (Francia, Hungría). Este derecho de financiación de los colegios privados por el estado también significa una restricción de la autonomía de los colegios privados financiados. Aunque estas restricciones difieren de una sociedad a otra, y varía según el grado de fondos estatales, se puede decir que, en general, esta disminución de la autonomía de estos colegios se ha visto reflejada en diversos aspectos tales como su plan de estudios, forma de examinar, remuneración de profesores y criterios de admisión de estudiantes. Estos colegios privados dependientes del gobierno los podemos encontrar hoy en día en gran número en países como Austria, Bélgica, República Checa, Dinamarca, Finlandia, Francia, Alemania, Hungría, Irlanda, Italia, Holanda, Portugal, España, Suecia, Suiza y Estados Unidos.

Junto a estos colegios privados dependientes del gobierno, existe un número de colegios privados de países de la OCDE, los cuales no obtienen financiación por partes del gobierno (local, regional, nacional). Financieramente, estos colegios dependen completamente de las cuotas de estudiantes, donaciones, patrocinios y la recaudación de fondos de los padres. Sin embargo, su autonomía también puede estar restringida de dos maneras. Por un lado, las autoridades podrían establecer ciertos criterios para los colegios independientes con el fin de asegurar una cierta calidad mínima en la adecuación a la sociedad de la generación venidera. Por otro lado, los colegios privados independientes funcionan dentro de un contexto social y están determinados, en parte, por este. Por ejemplo, los criterios de ingreso a las universidades restringirán la autonomía de los planes de estudios de un colegio privado. Sin embargo, la autonomía del colegio privado será mayor en sus políticas de admisión de estudiantes, especialmente teniendo en cuenta la importancia de las cuotas estudiantiles para la financiación de estos colegios. Estos colegios privados independientes pueden encontrarse actualmente en numerosos países como Austria, Bélgica, Francia, Hungría, Irlanda, Italia, Nueva Zelanda, Polonia, Portugal, España, Suiza, Reino Unido y Estados Unidos.

Dadas las condiciones diferentes para los colegios privados independientes y privados dependientes del gobierno, y sus oportunidades para competir con los colegios públicos, en consecuencia, distinguiremos entre colegios privados independientes, colegios privados dependientes del gobierno y colegios públicos.

También hay que resaltar que esta distinción entre colegios públicos, privados dependientes del gobierno y privados independientes no es común a todas las sociedades, especialmente, debido a que la historia y evolución de estos tipos de colegios puede ser muy diferente en cada una de ellas (iglesia frente a orientación al mercado, conexión fuerte o débil con los diferentes partidos políticos, fuertemente rechazada frente a una amplia aceptación, crecimiento reciente frente a una larga existencia) así como que las disposiciones constitucionales actuales son también diferentes en cada una de estas sociedades (basado en el derecho constitucional frente a acuerdos políticos, un fuerte o débil control de las autoridades estatales). Debido a todo esto, es importante controlar la eficiencia de las diferentes condiciones de enseñanza y aprendizaje, tanto entre los diferentes tipos de colegios como entre distintos países, lo cual haremos a continuación (Dronkers, y Robert, 2003).

Antes de analizar las diferentes causas que pueden provocar esta diferencia de eficiencia entre colegios públicos y privados cabe destacar que tener en consideración lo que narran Mancebon y Muñiz (2008). Hay una creencia generalizada de que los colegios privados son más eficientes que los públicos debido a la competencia en el libre mercado, ya que los colegios privados tienen que reinventarse día tras día para captar a su clientela de alumnos. Si no satisfacen sus necesidades, éstos tienen la oportunidad de irse a otro colegio. Por lo tanto, esto equivaldrá a un uso más eficiente de sus recursos y una mejora en la calidad de la educación para conseguir una ventaja respecto a sus competidores. Estas escuelas están obligadas a todo ello si no quieren desaparecer.

Por el contrario, los colegios públicos son vistos como monopolios con un público cautivo garantizado y cuentan con la ventaja de que el hecho de cambiar de colegio a uno privado para las familias debido a la gratuidad de los públicos. Por lo tanto, solo se dará el cambio si el valor percibido por las familias del colegio privado es muy superior al público y, esta opción, solo es posible para una masa reducida de la población que se lo puede permitir económicamente. Por todo esto, los defensores de la escuela privada, argumentan que en los colegios públicos el logro de la eficiencia y la satisfacción del cliente pasan a un papel secundario.

Después de muchos años de investigación no se tiene nada claro debido a que es un campo muy complicado de analizar. Para poder comprobar los dos lados las familias deben cambiar de un colegio a otro a sus hijos, lo cual es muy costoso tanto económicamente como un trastorno para el alumno causado por la dificultad de adaptación. Otra dificultad es que el usuario que mide la calidad es un niño que posee una capacidad limitada para evaluar la calidad de un

colegio. Por ende, las personas que tienen que evaluarlo son familiares, personas externas a la enseñanza del colegio y que se van a mover por temas religiosos, extra-escolares, proximidad a casa, instalaciones, etc., lo cual tiene una relación nula con la calidad educativa del colegio.

Las causas de estas diferencias de eficiencia entre ambas entidades han sido estudiadas por Dornkers y Robert (2003). En la Tabla 1 se muestran los resultados obtenidos por un test de valores medios de resultados escolares en lectura y matemáticas, diferenciando por tipo de colegio. Los datos indican que ambos resultados son significativamente superiores en los dos tipos de colegios privados en comparación con los colegios públicos, demostrándose así los mayores logros de los estudiantes en el sector privado de la educación. En otras palabras, de acuerdo con el informe PISA, se muestra que los estudiantes de colegios privados independientes tienen un mayor rendimiento.

TABLA 3.1: VALOR MEDIO DE LOS RESULTADOS DE LECTURA Y MATEMÁTICAS

Tipo de Colegio	Privado independiente	Privado dependiente del gobierno	Público	Total
Resultados de lectura	553,5*	529,9*	504,3	509,7
Resultados de matemáticas	546,5*	528,0*	503,8	508,8 (96,2)

* Diferencia significativa de la media comparada con la de los colegios públicos (t-test, varianzas distintas; $p < 0,05$).

3.2.2 Situación en España

- Colegios públicos

Los colegios públicos son caracterizados por la participación democrática en el funcionamiento y dirección de los colegios. Los Consejos Escolares están compuestos por representantes del colegio y de la comunidad, democráticamente elegidos y con poder para tomar decisiones a cualquier nivel. Los Consejos Escolares son responsables de las políticas de admisión y disciplina, aprobando los presupuestos anuales, evaluando los planes de desarrollo de los

colegios, asegurando el mantenimiento y reparación de los edificios del colegio y controlando las actividades administrativas y académicas del personal.

Los colegios tienen muy poca autonomía con respecto a los aspectos financieros y de cualquier otro ámbito de la vida escolar. Son controlados en cada uno de los asuntos fundamentales, así como en los horarios y funcionamiento interno general. El control se ejerce desde el Estado o desde los Consejos de las Comunidades Autónomas. Hasta el momento, no existe competencia entre los colegios públicos. Los padres tienen la libertad de elegir el colegio que desean para sus hijos, generalmente basándose en criterios específicos que, generalmente, limitan la elección, como por ejemplo: proximidad a la escuela, ingresos familiares, número de niños o, incluso, por la ya asistencia de hermanos a dicho colegio. Estos colegios abarcan todo tipo de clases, en cuanto a estudiantes matriculados, pero predominan estudiantes de la clase trabajadora y económicamente en desventaja. La educación impartida es plural.

- Colegios concertados (subvencionados por el estado)

Los colegios concertados, al estar financiados por el estado, deben ser totalmente gratuitos para los alumnos. También permiten la participación democrática, sin embargo, esta participación es de manera diferente a los colegios públicos. Tienen más autonomía en los aspectos financieros y administrativos de la institución. El control estatal se centra, principalmente, en el cumplimiento de los criterios básicos para la admisión de alumnos, horarios y planes de estudio. Por lo tanto, el control del Estado es mínimo y flexible. No hay competencia entre estos colegios, en parte porque no hay suficiente demanda para cada uno de ellos. Esto hace pensar que la cuestión no es la manera de atraer a los estudiantes, sino de cómo seleccionarlos. Hoy en día, el número de estudiantes que se matriculan en este tipo de colegios es decreciente. Por esta razón, los colegios están empezando a desarrollar estrategias de marketing para competir en el mercado.

En principio, los padres pueden elegir colegios concertados de la misma manera que lo hacen con los colegios públicos, es decir, atendiendo a criterios de proximidad, ingreso familiar, número de niños y hermanos que asisten al colegio determinado, aunque algunas veces no se cumplen estos criterios. Esto es debido a que las políticas de admisión no siempre se cumplen al 100%. Los colegios utilizan varios mecanismos de inscripción para evitar cierto tipo de estudiantes (principalmente, inmigrantes y gitanos), así como lo que supone para ciertas familias la compra de uniformes o el coste del comedor y la carga de las actividades extraescolares o las asociaciones de padres. Suelen acudir a estos colegios estudiantes de la clase media de la población. La educación es plural y, en algunos casos, elitista.

- Colegios privados

Los colegios privados no disponen de ningún tipo de subsidio estatal, por lo que participan en un sistema de libre mercado, el cual es bastante "brutal". Estos colegios tienen la libertad para planificar el modelo de participación que prefieran. También tiene total autonomía en aspectos financieros y de gestión. La competencia entre los colegios es muy fuerte, por lo que cada uno gestiona todo tipo de estrategias comerciales y de marketing que van desde anuncios comerciales televisivos hasta descuentos en las propias tiendas que se encuentran en los alrededores de estos colegios. Siguiendo las reglas del mercado, los padres son libres para elegir el colegio que más les guste. Suelen ser colegios donde predominan estudiantes de las clases altas y parte de la clase media. Estos colegios poseen una educación elitista. Bernal (2005)

- Razones por las que las familias españolas escogen un colegio privado o público.

Para analizar esta cuestión atenderemos a los cuatro micro-mercados que hace referencia Bernal (2005):

a) Colegios privados (clase alta y media)

En los colegios privados abundan las familias de clase media y alta con una elevada proporción de amas de casa que trabajan en dichas familias. Este micro-mercado se encuentra, principalmente, en el centro de la ciudad, aunque también se expande hacia otros barrios. Los criterios de los padres para escoger el colegio de sus hijos en este sector ha sido el siguiente:

- El elitismo basado, principalmente, en los resultados académicos en lugar de factores económicos.
- La disciplina y el orden impartidos por el colegio en el sentido de los logros, comportamiento y respeto por la autoridad.
- La participación controlada por la autoridad (un modelo de participación basado en el modelo familiar de tutela)

Esto refleja que este tipo de familias son tradicionales y se organizan en torno a una estructura sólida y patriarcal. Los roles están claramente definidos y jerárquicamente asignados, con el

propósito de garantizar el control de la socialización de los hijos, con lo que la participación se base en el control por autoridad.

b) Colegios públicos (clase media)

Los colegios públicos reúnen, básicamente, alumnos de familias procedentes de los escalafones superior e inferior de la clase media, aunque un gran número de ellos también procede de la clase trabajadora. En este segmento también hayamos un número elevado de amas de casa que trabajan. La mezcla de clases es uno de los factores más interesante. Las características de este micro-mercado:

- Entorno familiar favorable y no conflictivo, no existen personas socialmente aisladas.
- Colegio de prestigio.
- Ideología secular, son padres que prefieren las escuelas privadas, pero no quieren una religiosa.
- Experiencia y expectativas de los padres, suelen elegir el colegio en el que han estudiado ellos anteriormente y, el cual les transmite unas expectativas altas en cuanto al rendimiento de sus hijos.
- Proximidad, la mayoría de las familias viven cerca, aunque, también, un número considerables de las familias viven en otros barrios, pero prefieren estos colegios.
- Disciplina flexible en relación al comportamiento y tolerancia de los estudiantes.
- Las instalaciones que ofrece el colegio.

Este tipo de mercado comparte algunos elementos con el resto de mercados, mientras que en resto de ellos son específicos de los colegios públicos, por ejemplo, un entorno de tolerancia, mayor pluralismo, menos presión académica y la influencia ideológica sobre el contenido impartido. Sin embargo, podríamos decir que estos mercados (colegios de clase media públicos y privados) tienen más aspectos en común que diferencias, principalmente debido a que comparten un fondo social similar.

c) Colegios públicos (clase trabajadora)

Este micro-mercado reúne una gran parte de la clase trabajadora. Las madres de familia raramente trabajan fuera de casa y son, en su gran mayoría, amas de casa. Este mercado es el que acude a lo que podríamos llamar la escuela pública típica, ya que sirve como modelo para las otras escuelas. Sus principales características son las siguientes:

- Diversidad de los estudiantes, espejo del contexto global.
- Disciplina flexible, menor presión y mayor tolerancia.
- Ideología no religiosa.
- Experiencia y expectativas de los padres, suelen elegir el colegio en el que ellos estudiaron y les sugieren altas expectativas de rendimiento de sus hijos.
- Proximidad.
- Las instalaciones que ofrece.

Además, en este micro-mercado, la mayoría de los padres envían a sus hijos al colegio más cercano, sin considerar ninguna otra alternativa y sin hacer ninguna evaluación sobre su elección.

- *Colegios públicos (población más desfavorecida económicamente)*

Estos colegios públicos tienen una alta proporción de familias que pertenecen a los segmentos de la población más desfavorecidos. Dentro de este micro-mercado, los padres, realmente, no piensan donde les gustaría enviar a sus hijos, ellos solo los llevan al colegio más cercano. Su principal criterio de elección es la proximidad. No están especialmente preocupados por los aspectos educativos del colegio. Estas familias se encuadran en el primer nivel de la escala de motivación de Maslow. Su preocupación principal es como luchar en el día a día para salir adelante, y el colegio de sus hijos pasa a cubrir un papel social en el que cuidan de sus niños durante un periodo grande de tiempo, lo que ayuda a evitar que puedan caer en la delincuencia a una edad demasiado temprana.

- Diferencias en la eficiencia de los colegios públicos/privados en España

Una de las principales características del sistema educativo español a niveles no universitarios es su carácter mixto con un conjunto de corriente pública y privada de gran magnitud. Basándonos en el análisis realizado por Mancebón y Muñiz (2008), la mayoría de los colegios analizados son públicos y están localizados en zonas urbanas.

Las comparaciones realizadas sale más favorable la escuela privada que la pública, tanto en términos cuantitativos como cualitativos. Esto se deduce del porcentaje de estudiantes que pasan los exámenes de ingreso a la Universidad Española, el cual es entre un 2 y un 5% superior en los colegios privados que en los públicos, y de la nota media obtenida por los alumnos que aprobaron dichos exámenes, dos o tres décimas superior en los colegios privados.

Además de esto, el número de colegios con calificaciones más altas es mayor en el sector privado que en el público.

Estos valores deben estar relacionados con la calidad de los alumnos en cada tipo de colegio, lo cual nos ayudará a realizar el análisis de la eficiencia. Esta comparación también presenta mejores resultados para los colegios privados. Por lo tanto, estos colegios tienen, en promedio, un mayor porcentaje de alumnos con buen expediente académico, los alumnos dedican más tiempo a estudiar, tienen mayores aspiraciones académicas futuras, un mayor nivel de confianza en sí mismos, y mejor percepción del apoyo familiar en cuanto a lo académico se refiere. Del mismo modo, en los colegios privados hay un mayor porcentaje de estudiantes cuyos padres tiene un trabajo cualificado, han pasado por una educación superior y tienen ingresos mensuales superiores a 1.800 €. También puede verse que las diferencias entre los alumnos de los colegios públicos y privados son mayores si lo relacionamos con los antecedentes que con las características individuales de los estudiantes. Por todo ello, el potencial de los alumnos en la educación privada parece ser mayor que en la educación pública, y esto está en la línea de múltiples estudios que comparan el perfil académico y económico de los alumnos en los colegios públicos y privados.

A pesar de todo esto, en cuanto a eficiencia en la gestión, no se puede concluir una superioridad de los privados frente a los públicos, ya que los mejores resultados académicos de los alumnos en los colegios privados, pueden deberse a que estos tienen un fondo educativo más favorable y, por lo tanto, mayor potencial en términos generales.

3.3 ANÁLISIS EXPLORATORIO DE DATOS ESPACIALES (AEDE)

3.3.1 Análisis exploratorio básico

Dentro del análisis exploratorio básico empezaremos describiendo el histograma de frecuencias. Es una aproximación discreta a la función de densidad de una variable aleatoria y es útil para detectar la asimetría, múltiples modos y otras peculiaridades en la distribución. Se trata de una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. En el eje vertical se representan las frecuencias y en el eje horizontal, los valores de las variables. Este gráfico resulta de especial utilidad en variables de naturaleza continua, pues hace posible una visualización más fácil de su

distribución al agrupar estos valores en categorías o clases diferentes. En nuestro caso, hemos utilizado el programa GeoDa, el cual calcula histogramas de las variables geográficas y permite establecer hasta siete categorías por defecto. Cada una de las barras del histograma tiene un color y es posible realizar una selección en el histograma para ver sobre el mapa las observaciones a las que corresponde.

Como segunda herramienta pasamos a describir el diagrama de dispersión o nube de puntos, el cual, a diferencia del histograma de frecuencias, ya entra dentro del análisis multivariante. El diagrama de dispersión proporciona una buena descripción de la relación o dependencia existente entre dos variables X e Y . Se trata de una representación gráfica de las observaciones de las variables (X, Y) para definir su comportamiento conjunto. La forma que presenta esta nube de puntos refleja el grado de correlación entre las dos variables, que puede ser nula (si los puntos forman un círculo), lineal (si los puntos representan una elipse) o no lineal (si los puntos toman cualquier otra forma). La función más sencilla y útil en la mayoría de los casos, por su simplicidad, es la recta. En este caso, la dependencia entre las variables es medida a través del coeficiente de correlación lineal. El diagrama de dispersión permite determinar la significatividad de este coeficiente, en cuanto que la relación entre ambas variables sea o no lineal, y existan ciertos puntos atípicos que le resten representatividad.

En GeoDa, este diagrama se acompaña de la línea de regresión lineal, ajustada por el método de mínimos cuadrados ordinarios y el coeficiente correspondiente a la pendiente. (Chasco y Fernández, 2009).

3.3.2 Concepto de autocorrelación espacial

Este concepto, también conocido como dependencia espacial, es, junto con la tendencia y los valores atípicos, un elemento caracterizador de las distribuciones espaciales. En concreto, la dependencia espacial consiste en la coincidencia de valores (altos/bajos) de una variable en determinados lugares del espacio geográfico.

El fenómeno de dependencia espacial en una variable puede expresarse matemáticamente como una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que sucede en lugares cercanos (vecinos) al mismo (Anselin, 2001a). Es decir, en una variable se produce el fenómeno de la dependencia o autocorrelación espacial cuando los valores observados en una observación (región, barrio, etc.) dependen de los valores observados en regiones vecinas. En este caso, la distribución de la variable sobre un mapa seguirá una continuidad geográfica.

La dependencia o autocorrelación espacial es un tema que ha suscitado un interés creciente desde los primeros trabajos estadísticos de Cliff y Ord (1973, 1981). En estos libros, los autores estudian este fenómeno para conocer las causas y formas de propagación de epidemias y enfermedades, aunque ya otros autores habían constatado antes la existencia de continuidad geográfica en fenómenos sociales como la intención de voto (Cox, 1969). Además, este efecto espacial ha sido objeto de aplicaciones interesantes en otros campos, como los recursos naturales y medio ambiente (Anselin, 2001a), delincuencia (Baller et al., 2006; Ceccato et al., 2002), actividad de I+D (Anselin et al., 2000; Fisher y Varga, 2003), etc. En el terreno específicamente económico, se ha demostrado como es muy común la existencia de autocorrelación espacial en fenómenos como la producción, paro, renta disponible, procesos de convergencia regional, etc. (algunos ejemplos en este campo serían Rey y Montouri, 1999; Moreno et al., 2000; Rey, 2001, 2004a y b; Chica, 1994); Arbia, 2006; Le Gallo y Chasco, 2008, entre otros muchos).

El efecto de autocorrelación espacial puede ser de signo positivo o negativo, así como nulo.

Se entiende por autocorrelación espacial positiva el fenómeno de asociación entre valores similares de una variable y localizaciones cercanas; es decir, cuando, en el espacio geográfico, los valores altos de una variable están rodeados por valores altos, y viceversa. Éste sería el caso, del llamado efecto contagio o desbordamiento (spillover) que se produce en muchos fenómenos socioeconómicos de renta y desarrollo humano, en general, en los que su presencia en una región es causa de su extensión a regiones vecinas, favoreciendo la concentración del fenómeno en una zona geográfica determinada.

Por el contrario, existe autocorrelación espacial negativa en un espacio cuando los valores altos de una variable se encuentran rodeados por valores bajos de la misma, y viceversa. Se trata de una configuración propia de variables en las que las mayores diferencias de valores se dan en aquellas unidades geográficas que son cercanas entre sí, como sucede con ciertos fenómenos de jerarquías espaciales, con regiones de alta densidad de población rodeadas de regiones despobladas, y viceversa. Según expone Anselin (1988), este efecto espacial es más difícil de detectar en variables de tipo socioeconómico (si se compara con las ciencias naturales, por ejemplo) y está muy condicionado por la escala geográfica en la que se lleven a cabo los análisis.

Por último, se produce ausencia de autocorrelación espacial en una variable geográfica cuando ésta se distribuye de manera aleatoria sobre el espacio.

Sin embargo, no siempre que se produce autocorrelación espacial en un avariable geográfica será consecuencia de la existencia de una verdadera interacción espacial entre regiones, como

los citados fenómenos de desbordamiento y jerarquías espaciales. Algunas veces, las variables espaciales (al igual que las series temporales) pueden estar correlacionadas de forma espuria o falsa. (Chasco y Fernández, 2009).

3.3.3 Autocorrelación espacial global

Para poder definir la autocorrelación global, debemos centrarnos en la técnica de AEDE de dependencia espacial a través de la cual analizamos dicho efecto, el diagrama de dispersión de Moran. Se trata de un diagrama de dispersión que representa en el eje X una variable, previamente estandarizada, y_i , y en el eje Y, se representa la variable espacialmente retardada de dicha variable ($w_{ij}y_j$). Este diagrama permite distinguir los dos signos propios del fenómeno de dependencia espacial: positiva y negativa, mencionados y descritos anteriormente. Las categorías de dependencia espacial positiva se corresponden con los cuadrantes I y III, mientras que el fenómeno de dependencia espacial negativa se encuentra en los cuadrantes II y IV.

En este diagrama de dispersión, en el que se relacionan el valor de la variable observación geográfica y el valor promedio de dicha variable en las observaciones vecinas, la pendiente de la recta de regresión coincide con el valor del estadístico I de Moran de autocorrelación espacial global.

El estadístico I Global de Moran permite conocer el grado de dependencia espacial existente en una variable, es decir, hasta qué punto existe alguna tendencia o concentración de valores de una variables en determinadas zonas de la geografía.(Chasco y Fernández, 2009).

3.3.4 Autocorrelación espacial local

El estadístico I de Moran, al ser una medida de dependencia espacial global en cuanto que se refiere a todo el conjunto geográfico sometido al análisis, no es capaz de detectar la presencia de zonas concretas de inestabilidad (zonas calientes o frías), en las que una variable tenga un comportamiento diferente de la tendencia general. Al producirse este hecho, se dice que la variable no es estacionaria, es decir, que no es posible aceptar una estructura de dependencia espacial similar entre cualquier par de unidades espaciales situadas a igual distancia en un mapa. Muchas veces, la intensidad (y hasta el signo) de la dependencia espacial puede ser significativamente diferente en algunas subzonas del espacio global considerado.

La hipótesis de estacionariedad es la base sobre la que se asienta el análisis estadístico de la autocorrelación espacial, por lo que en el caso de existencia de dependencia espacial local, este

supuesto se incumple y el análisis global no sería válido. El fenómeno de la inestabilidad espacial o dependencia espacial local puede plantearse desde dos puntos de vista:

- En primer lugar, es posible que el estadístico I de Moran no detecte la presencia de dependencia espacial en la distribución de una variable sobre un espacio dado, pero que existan clusters o agrupamientos de unidades en los que dicha variable experimenta una concentración (o dispersión) importante.
- En segundo lugar, también podría ser que el estadístico I de Moran sí que detecte la presencia de dependencia espacial en una variable, pero que no todas las unidades del espacio considerado contribuyan con igual peso en el indicador global. Es decir, dentro del espacio global, pudieran coexistir zonas en las que una variable no esté espacialmente autocorrelacionada con otras zonas en las que el fenómeno de dependencia espacial sea muy importante.

Para responder a todas estas cuestiones planteadas, se ha diseñado el estadístico de dependencia espacial local, estadístico I local de Moran, el cual calcula, para cada unidad i , la existencia de un agrupamiento significativo de valores similares (altos o bajos) alrededor de dicha unidad geográfica del modo siguiente (Anselin, 1995):

$$I_i = \mathcal{Y}_i \sum_{j \in J_i} w_{ij} \mathcal{Y}_j \quad 3.1.$$

siendo $\mathcal{Y}_i, \mathcal{Y}_j$, los valores de la variable Y , en desviaciones a la media, en la unidades espaciales i, j ; w_{ij} son los elementos de la matriz de pesos espaciales W ; J_i es el conjunto de unidades relacionadas con i . (Chasco y Fernández, 2009).

3.4 MODELO DE REGRESIÓN LINEAL

El propósito general del análisis de regresión lineal es encontrar una relación entre una variable dependiente y un conjunto de variables explicativas. Formalmente, esta relación se expresa:

$$y = X\beta + \varepsilon \quad 3.2.$$

donde y es la variables dependiente (en forma vectorial, con N filas), X es una matriz con las observaciones en K variables explicativas (con N filas y K columnas), β es un vector con K

coeficientes de regresión (es decir, de dimensión K por 1), y es un término de error aleatorio (en formato vectorial, con N filas).

Normalmente, este tipo de análisis tiene dos objetivos claros. Por un lado encontrar una buena aptitud entre los valores predichos Xb (suma de valores de las variables explicativas, cada uno multiplicado por su coeficiente de regresión) y los valores observados de la variable dependiente y . Por otro lado, descubrir cuál de las variables explicativas contribuyen significativamente a la relación lineal. (Anselin, 1992)

3.4.1 Estimación por Mínimos Cuadrados Ordinarios (MCO)

El método de mínimos cuadrados ordinarios (MCO) logra ambos objetivos mencionados anteriormente de manera óptima, de acuerdo con una serie de criterios, y se conoce como Mejor Estimador Insesgado Lineal. Las estimaciones de MCO para β se encuentran al minimizar la suma de errores de predicción al cuadrado (por lo tanto, los mínimos cuadrados).

Con el fin de obtener la propiedad insesgada lineal y para ser capaz de hacer inferencias estadísticas acerca de los coeficientes de regresión de población β por medio de sus estimadores b , es necesario asegurarse de las suposiciones sobre la parte aleatoria de la ecuación de regresión (error aleatorio ϵ). Dos de estos supuestos son cruciales para obtener las estimaciones de sesgo y la eficiencia de las estimaciones de MCO:

- El error aleatorio vale 0 (no existe mala especificación sistemática o sesgo en la ecuación de regresión de la población):

$$E[\epsilon_i] = 0 \text{ para todo } i \quad 3.3.$$

- Los términos de error aleatorio no están correlacionados y tienen una varianza constante (homoscedasticidad):

$$E[\epsilon_i \epsilon_j] = 0 \text{ para todo } i + j \quad 3.4.$$

$$E[\epsilon_i^2] = \sigma^2 \text{ para todo } i \quad 3.5.$$

Se necesita un tercer supuesto con el fin de llevar a cabo las pruebas de hipótesis y para evaluar la significación de los coeficientes de regresión:

- El término de error aleatorio sigue una distribución normal:

$$s \sim N(0, \sigma^2) \quad 3.6.$$

En la forma multivariante, es decir, para todos los N conjuntos de observaciones, estas propiedades se expresan como:

- $E[s] = 0$ 3.7.

- $E[ss'] = \sigma^2 I$ 3.8.

- $s \sim N(0, \sigma^2 I)$ 3.9.

Donde s es como antes, un vector N por 1 en términos de error al azar, 0 es un vector N por 1 de ceros, σ es la varianza del error de la población e I es una matriz identidad de dimensión N x N.

Estos supuestos introducen un parámetro adicional a ser estimado (además los coeficientes de regresión), es decir, la varianza del error. (Anselin, 1992)

3.4.2 Medidas de bondad del ajuste

La medida en que los valores predichos coinciden con los valores observados para la variable dependiente se miden por la R^2 . Esta medida de ajuste se base en la descomposición del total de la suma de cuadrados, SST (suma de los cuadrados de la variable dependiente) en una suma explicada de los cuadrados (suma de los cuadrados de los valores predichos), referido a la regresión de la suma de cuadrados, SSR, y una suma residual de la suma de cuadrados (suma de los cuadrados de los residuos), RSS:

$$SST = SSR + RSS \quad 3.10.$$

La R^2 es, entonces, definida como:

$$R^2 = 1 - RSS/SST \quad 3.11.$$

Cuando la regresión incluye un término constante, esta descomposición es equivalente a una descomposición de la varianza total en varianza explicada y residual (cuando todas las sumas de cuadrados se calculan para la media de las desviaciones).

En general, el modelo con mayor R^2 se considera que tiene el mejor ajuste. Sin embargo, la medida de R^2 , solamente, no siempre es un buen indicador de lo bien que se explica la regresión de los valores observados, ya que aumenta con cada variable explicativa adicional que se incluye en la especificación del modelo. Con el fin de proporcionar una mejor guía que compense el "exceso de ajuste" de los datos, un R^2 ajustado (R_a^2) se calcula como:

$$R_a^2 = R^2 - (1 - R^2)(K - 1)/(N - K) \quad 3.12.$$

Este R_a^2 ajustado no necesariamente aumenta cuando se añaden variables adicionales para el modelo. Cuando no es así, significa que la variable adicional no contribuye suficientemente en el ajuste del modelo para justificar la pérdida de grados de libertad debido a su inclusión.

Un conjunto alternativo de medidas de ajuste se basa en la máxima verosimilitud (ML) aproximada a la estimación. Para el modelo de regresión estándar, estos no son realmente necesarios, ya que las estimaciones por MCO son equivalentes a las estimaciones de máxima verosimilitud. Sin embargo, si se desea comparar el ajuste de regresión estándar a la de un modelo de regresión espacial, puede que ya no se confíe en la R^2 y se deben utilizar las medidas basadas en la ML.

La estimación de máxima verosimilitud se base en el concepto de densidad o distribución conjunta en función de los datos observados, y, que se conoce como la función de probabilidad. Esta función de probabilidad se expresa en función de las variables explicativas X y un conjunto de parámetros: β y σ^2 . Las estimaciones de ML para los parámetros son los valores que obtienen la mayor probabilidad o probabilidad conjunta. A fin de aplicar la estimación de ML, es necesario adoptar una forma para la distribución conjunta. Típicamente, esta será la distribución normal. El modelo con la mayor probabilidad registrada es el que logra un mejor ajuste. Como en el caso del R^2 , el logaritmo de verosimilitud siempre aumenta cuando se incluyen variables adicionales en el modelo, y, por lo tanto, puede no ser un buen indicador de lo bien que explica los datos el modelo.

Con el fin de corregir el ajuste por exceso del logaritmo de verosimilitud, han sido propuestos un número de criterios llamados criterios de información (IC). En términos generales, un criterio de información posee la forma siguiente:

$$IC = -2L + f(K, N) \quad 3.13.$$

Donde L es el logaritmo de verosimilitud maximizado y f(K,N) es una función del número de variables (K) y el número de observaciones (N), es decir, de los grados de libertad del modelo. El principio general en el que se basan los IC es que la evaluación del ajuste es penalizado por una función de los grados de libertad. El mejor modelo es el que tienen el valor más bajo para cada criterio de información.

Hay que tener en cuenta que cada factor de corrección corresponde a una función objetivo, y que el mejor modelo según uno, no es necesariamente el mejor según otros. (Anselin, 1992)

3.4.3 Contrastes de hipótesis básicas del modelo de regresión lineal

Cuando el verdadero interés reside en la inferencia estadística, no preocupan tanto los valores particulares de las estimaciones de los coeficientes de regresión b , sino la manera en que se pueden utilizar para sacar conclusiones acerca de los parámetros poblacionales desconocidos. Para llegar a este punto, es necesario hacer una suposición sobre la distribución de los errores aleatorios. También se necesita una estimación de la varianza de los coeficientes de regresión β . La varianza de un coeficiente de regresión individual es el elemento de la diagonal de una matriz correspondiente a la covarianza para todos los coeficientes. Este último se estima como $s^2(X'X)^{-1}$, donde s^2 es el estimador insesgado de la varianza del error, $e'e/N-K$.

Por lo general, interesa saber si el coeficiente de población es diferente de cero, o, en otras palabras, si la variable asociada contribuye a la ecuación de regresión. Formalmente, esta es una hipótesis de prueba para la hipótesis nula de que la regresión del coeficiente poblacional en cuestión (por ejemplo) es cero:

$$H_0: \beta_k = 0 \quad 3.14.$$

Si se rechaza la hipótesis nula, es posible concluir que el coeficiente de población debe ser distinto de cero, dentro de los límites que le ha asignado el nivel de significación elegido (o error Tipo I).

La significación de los coeficientes de regresión individuales puede ser probada por medio del t-test. Bajo el supuesto de términos de error normales, el estadístico:

$$t_b = b_k / SE(b_k) \quad 3.15.$$

Es decir, la estimación de la regresión dividida por su desviación estándar (la raíz cuadrada de su varianza), sigue una distribución t de Student con NK grados de libertad. Si la probabilidad de t-estadístico está por debajo de un nivel crítico escogido, se puede concluir que la hipótesis nula $\beta_k = 0$ es improbable, y, por lo tanto, que el coeficiente de población probablemente sea distinto de cero (o significativo).

Otro tema es la importancia de la especificación de la regresión en su conjunto, o, en otras palabras, si la pendiente de los coeficientes en la regresión de población (es decir, todos excepto el término constante) conjuntamente no sea cero. Esto se traduce en un conjunto $K-1$ hipótesis nulas de la forma dada anteriormente que deben ser satisfechas en su conjunto. Una prueba para este conjunto de hipótesis es la denominada F-test, que es similar al enfoque adoptado en el análisis de la varianza (ANOVA). El estadístico F-test se puede expresar en términos de la

suma residual restringida y no restringida de los cuadrados (RSS, es decir, suma de los residuos al cuadrado):

$$F = \{RSS_C - RSS_U/K - 1\} / \{RSS_U/N - K\} \quad 3.16.$$

La suma residual restringida de los cuadrados (RSS_C) proviene de una regresión en la que se sostiene la hipótesis nula, es decir, una regresión sin coeficientes de pendiente y sólo término constante. El (RSS_C) es, por lo tanto, la suma de los cuadrados de la variable dependiente (en desviaciones desde su media). La suma de cuadrados sin restricciones (RSS_U) es la suma residual de los cuadrados en la regresión. El estadístico F-test se distribuye como una variable aleatoria F con K-1, N-K grados de libertad. Hay que tener en cuenta que sólo en circunstancias excepcionales, este estadístico no será significativo. Típicamente, este será el caso de especificaciones muy pobres del modelo con indica un bajo R^2 o, incluso, un negativo R_a^2 . (Anselin, 1992)

3.5 DIAGNÓSTICOS DE ESPECIFICACIÓN

Obviamente, los supuestos términos de error normales, homoscedásticos y no correlacionados que conducen a la característica AZUL de estimaciones por MCO no están necesariamente satisfechos por modelos reales que utilizan datos reales. Una parte importante del buen desarrollo de la práctica econométrica consiste en comprobar el grado en que estos supuestos pueden ser violados. Cuando se trata de los datos espaciales, se debe prestar especial atención a la posibilidad de que los errores y las variables del modelo muestren la dependencia espacial. (Anselin, 1992)

3.5.1 Multicolinealidad

Un problema que, a menudo, se puede encontrar en el trabajo empírico es la alta correlación entre las observaciones de las variables explicativas incluidas en la especificación de la regresión. En principio, estas variables explicativas deben estar correlacionadas. Si existe una relación lineal exacta entre algunas de las variables (es decir, una correlación igual a 1), la estimación por MCO se descompondrá. Esta situación se conoce como multicolinealidad perfecta. A menudo, sin embargo, el alcance de la multicolinealidad no es perfecto, pero aparece como una relación lineal fuerte (es decir, correlación) entre las variables explicativas. Como consecuencia, la estimación por MCO no se romperá, pero las estimaciones tendrán varianzas

estimadas muy grandes. Por lo tanto, muy pocos coeficientes serán significativos, a pesar de que la regresión en su conjunto pueda parecer lograr un ajuste razonable. Esta combinación de alto R^2 con t estadísticos muy bajos, es, a menudo, un buen indicador de que algo está mal en términos de multicolinealidad. Otro signo revelador es cuando las estimaciones varían considerablemente como resultado de la adición o eliminación de una sola observación.

No existen tests, en sentido estricto, de multicolinealidad. Sin embargo, hay algunos diagnósticos que pueden indicar un problema potencial. Uno de ellos, es el llamado número de condición, popularizado por la obra de Belsley et. Al. (1980). Como regla general, los valores del número de condición mayores de 20 o 30 son considerados como sospechosos. La falta total de multicolinealidad es dada cuando se obtiene un número de condición de 1.(Anselin, 1992)

3.5.2 Errores no normalmente distribuidos

La mayoría de los test de hipótesis y un gran número de diagnósticos de regresión se basan en la hipótesis de una distribución normal del error. Es difícil evaluar el grado en que esto puede ser violado, ya que los mismos errores (es decir, el término de error aleatorio en el modelo de regresión poblacional) no se pueden observar. En cambio, los test para los errores no normales deben calcularse a partir de los residuos de la regresión. Esta es una prueba asintótica, por lo que puede no ser muy fiable para los pequeños conjuntos de datos. El estadístico sigue una distribución χ^2 con 2 grados de libertad. Una baja probabilidad indica rechazo de la hipótesis nula de que el error es normal. Si es el caso, los test para heteroscedasticidad y dependencia espacial deben interpretarse con cautela, ya que se basan en el supuesto de normalidad. En muchos casos, una simple transformación de la variable dependiente, tal como un logaritmo, puede inducir a la normalidad.(Anselin, 1992)

3.5.3 Heteroscedasticidad

La heteroscedasticidad es la situación en la que el error de regresión aleatorio no tiene una variación constante en todas las observaciones (es decir, no es homoscedástico). Como consecuencia, la indicación de la precisión dada que asume una varianza del error constante estimada por MCO será engañosa. Mientras que las estimaciones por MCO son todavía imparciales, ellos no serán más eficientes. Más importante aún, la inferencia basada en los t y f estadísticos usuales será engañosa, y la medida de R^2 de la bondad de ajuste estará

equivocada. En el análisis de datos espaciales, se encuentra con frecuencia este problema, sobre todo cuando se utilizan datos de unidades espaciales irregulares (con diferente zona), cuando hay diferencias regionales sistemáticas en las relaciones del modelo (es decir, los regímenes espaciales), o cuando hay una deriva espacial continua en los parámetros del modelo (es decir, expansión espacial). La presencia de cualquiera de estos efectos espaciales daría un modelo de regresión estándar que ignorara las malas especificaciones. Por lo tanto, una indicación de heteroscedasticidad puede apuntar a la necesidad de una incorporación más explícita de efectos espaciales, en forma de regímenes espaciales o expansión espacial de los parámetros.

Existen muchos test contra la heteroscedasticidad. Todos estos test comienzan con la hipótesis nula de homoscedasticidad:

$$H_0 = E[s_i^2] = \sigma^2 \quad 3.17.$$

La hipótesis alternativa es que el término de error de cada observación tiene una varianza diferente. En muchos casos, esto es demasiado general para ser de gran utilidad. El grado de especificación con la que la hipótesis alternativa de heteroscedasticidad puede expresarse dependerá del conocimiento de los factores que pueden provocarlo. Hay que tener en cuenta que si se está bastante seguro de la existencia de heteroscedasticidad, no se debe estimar un modelo de regresión estándar. En su lugar, se debe incluir, explícitamente, el error heteroscedástico en la especificación del modelo.

Un enfoque común hacia la concreción de la hipótesis alternativa es relacionar la variabilidad en la varianza del error con un número de variables, a través de una forma funcional que incluye algunos parámetros (P parámetros) como en:

$$\sigma_i^2 = \sigma^2 f(\alpha_0 + \sum_p z_{pi} \alpha_p) \quad 3.18.$$

donde σ^2 es un factor de escala simple, f es una función, α_0 y α_p son los parámetros y z_{pi} son las variables P para la observación i. Las formas funcionales comúnmente utilizadas son el uno lineal (denominado heteroscedasticidad aditiva) y la exponencial (llamada heteroscedasticidad multiplicativa).

Las variables z incluidas en la especificación heteroscedástica pueden ser cualquier variable relevante. A menudo, el área de la unidad espacial, o cualquier otra variable que se refiera a su tamaño (población total, ingresos totales) son buenas opciones. Ésta es la situación en la que los coeficientes de regresión no son fijos (como en el modelo estándar), pero varían aleatoriamente en torno a una media, con una determinada varianza. En este caso, las variables z de la

especificación heteroscedástica se convierten en los cuadrados de las variables explicativas del modelo de regresión (es decir, el x_{im}^2), y los coeficientes correspondientes (es decir, α_p) son el coeficiente de la varianza.

Existen tres test contra la heteroscedasticidad. El primero es el test multiplicador de Lagrange desarrollado por Breusch y Pagan (1979), o su versión estandarizada sugerida por Koenker (1981) y Koenker y Bassett (1982). Los cuales dependen del resultado del test de normalidad. Cuando los errores son no normales, el test de Breusch-Pagan (BP) muestra que se ha logrado un pobre poder en pequeñas muestras. Ambos test son asintóticos y consiguen una distribución χ^2 con P grados de libertad (donde P es el número de variables z en la especificación heteroscedástica).

Tanto el test BP como el KB requieren que se especifiquen las variables que se van a utilizar en la especificación heteroscedástica. Se necesita un enfoque diferente para la situación en la que hay muy poca información previa acerca de la forma de la heteroscedasticidad. En tales casos, el test desarrollado por White (1980) es más apropiado, ya que tiene la propiedad de oponerse a cualquier forma no especificada de heteroscedasticidad. Este test también es asintótico y sigue una distribución χ^2 .

Una cuestión a tener en cuenta en situaciones donde tanto la heteroscedasticidad como la dependencia espacial pueden estar presentes es que los test contra la heteroscedasticidad han demostrado ser muy sensibles a la presencia de dependencia espacial. En otras palabras, mientras que los test pueden estar indicando heteroscedasticidad, este puede no ser el problema, pero en su lugar, puede darse dependencia espacial. (Anselin, 1992)

3.5.4 Autocorrelación espacial

La autocorrelación espacial, o más generalmente, dependencia espacial, es la situación en la que la variable dependiente y término de error para cada localización, se correlaciona con las observaciones de la variable dependiente o los valores del término de error en otras localizaciones. El caso general formalmente es:

$$E[y_i y_j] \neq 0 \text{ ó } E[e_i e_j] \neq 0 \quad 3.19.$$

Para localizaciones vecinas i y j. Esta especificación es demasiado general para permitir la estimación potencialmente de N interacciones (N-1) de N observaciones. Por lo tanto, la forma de dependencia espacial se estructura por medio de una matriz de ponderaciones espaciales

(W), lo que reduce el número de parámetros desconocidos a uno, es decir, el coeficiente de asociación espacial en un proceso medio de autorregresión espacial o movimiento espacial.

$$y = \rho Wy + X\beta + \varepsilon \quad 3.20.$$

Donde Wy es una variable dependiente retardada espacialmente y ρ es el coeficiente autorregresivo espacial. La hipótesis nula de no autocorrelación espacial corresponde a:

$$H_0: \rho = 0 \quad 3.21.$$

Si esta forma de autocorrelación espacial es ignorada, las estimaciones por MCO estarán sesgadas y toda inferencia basada en el modelo de regresión estándar (es decir, el modelo sin el término Wy) será incorrecta. En cierto sentido, esto es similar a las consecuencias de la omisión de una variable explicativa significativa en el modelo de regresión.

La segunda forma de autocorrelación espacial en un modelo de regresión se refiere al término de error. También se puede referir a este caso como error espacial o la dependencia espacial como una molestia (ya que solo se refiere a los errores). Formalmente, esta dependencia se expresa por medio de un proceso espacial de los términos de error, ya sea de forma autorregresiva o por media móvil. Dicho proceso autorregresivo se expresa como:

$$y = X\beta + \varepsilon \quad 3.22.$$

$$\varepsilon = \lambda W\varepsilon + \xi \quad 3.23.$$

Con $W\varepsilon$ como término de error retardado espacialmente, λ como el coeficiente autorregresivo y ξ como término de error homoscedástico y no correlacionado). Un proceso de media móvil en el término de error tiene la forma:

$$\varepsilon = \lambda W\xi + \xi \quad 3.24.$$

Donde ahora el retardo espacial se refiere a los errores ξ y no al original ε . En ambos casos, la hipótesis nula tiene la forma siguiente:

$$H_0: \lambda = 0 \quad 3.25.$$

o, el término de error no está correlacionado. Resulta que los test para la presencia de cualquiera de las formas de dependencia espacial del error son los mismos. Las consecuencias de ignorar la dependencia espacial del error son las mismas que para la heteroscedasticidad: el estimador por MCO sigue siendo imparcial, pero ya no es eficiente, ya que no tiene en cuenta la correlación entre los términos de error. Como resultado, la inferencia basada en los estadísticos

T y F será engañosa y las indicaciones de ajuste basadas en R^2 serán incorrectas. (Anselin, 1992)

3.6 EL MODELO DEL RETARDO ESPACIAL

3.6.1 Definición

El modelo del retardo espacial incluye una variable dependiente especialmente retardada, W_y , como una de las variables explicativas:

$$y = \rho W_y + X\beta + \varepsilon \quad 3.26.$$

donde y es un vector de N por 1 observaciones de la variable dependiente, W_y es un vector N por 1 retardos espaciales de la variable dependiente, ρ es el coeficiente autorregresivo espacial, X es una matriz N por K observaciones sobre la variable explicativa con un vector asociado K por 1 de coeficientes de regresión β , y ε es un vector N por 1 de términos de error normalmente distribuidos aleatoriamente, con media 0 y varianza constante σ^2 (homoscedástico).

La presencia del retardo espacial es similar a la inclusión de variables endógenas en la RHS en los sistemas de ecuaciones simultáneas. Por lo tanto, este modelo se refiere a menudo al modelo autorregresivo espacial simultáneo. Si el parámetro autorregresivo es conocido, el modelo se simplificaría a una regresión estándar de las variables dependientes "filtradas" $y - \rho W_y$ en la variable explicativa X , como en:

$$y - \rho W_y = X\beta + \varepsilon \quad 3.27.$$

Sin embargo, típicamente los coeficientes de ρ son desconocidos y se deben estimar conjuntamente con los coeficientes de regresión. La principal consecuencia de la inclusión de W_y en la RHS de la especificación es que MCO ya no alcanza la consistencia. Esto es similar a lo que ocurre en sistemas de ecuaciones simultáneas. En lugar de MCO, la estimación debe basarse en una forma instrumental de variables que se aproximan, o en una maximización explícita de la función de verosimilitud.

Es posible interpretar la especificación del retardo espacial de dos maneras distintas. Por un lado, se tiene en cuenta la inclusión de W_y además de otras variables explicativas como una forma de evaluar el grado de dependencia espacial, mientras se controla el efecto de estas otras variables. Por lo tanto, el interés principal está en el efecto espacial. Alternativamente, la

inclusión de W_y le permite evaluar la importancia de las otras variables (no espaciales), después de que la dependencia espacial está controlada. (Anselin, 1992)

3.6.2 Estimación por el método de Variables Instrumentales (VI)

Como se muestra anteriormente, El modelo del retardo espacial incluye una variable dependiente especialmente retardada, W_y , como una de las variables explicativas:

$$y = \rho W_y + X\beta + \varepsilon \quad 3.28.$$

donde y es un vector de N por 1 observaciones de la variable dependiente, W_y es un vector N por 1 retardos espaciales de la variable dependiente, ρ es el coeficiente autorregresivo espacial, X es una matriz N por K observaciones sobre la variable explicativa con un vector asociado K por 1 de coeficientes de regresión β , y ε es un vector N por 1 de términos de error normalmente distribuidos aleatoriamente, con media 0 y varianza constante σ^2 (homoscedástico). Hay que tener en cuenta que ya no es requerida la normalidad. Para facilitar la notación, se representan el retardo espacial W_y y las variables explicativas "exógenas" X conjuntamente en una matriz Z , de dimensiones N por $K+1$.

La presencia del retardo espacial es similar a la inclusión de variables endógenas en la RHS en los sistemas de ecuaciones simultáneas. El estimador de variables instrumentales (IV) o estimador de mínimos cuadrados de dos etapas (MC2E) explota esta característica mediante la construcción de un instrumento adecuado para el retardo espacial. La estimación resultante es consistente, pero no necesariamente muy eficiente. Se puede utilizar como la base para un procedimiento de arranque.

El principio de la estimación de variables instrumentales se basa en la existencia de un conjunto de instrumentos, Q , que están fuertemente correlacionados con las variables originales, Z , pero asintóticamente correlacionado con el término de error. Una vez que se identifican estos instrumentos, ellos suelen construir un indicador que consiste en sus valores predichos en una regresión sobre los instrumentos y las variables exógenas. Formalmente, este proceso de mínimos cuadrados de dos etapas se obtiene de la estimación:

$$\theta_{IV} = [(Z'Q)(Q'Q)^{-1}(Q'Z)]^{-1}(Z'Q)(Q'Q)^{-1}Q'y \quad 3.29.$$

donde θ es un vector $K+1$ por 1 con la estimación para ρ como primer elemento, seguido por la estimación para β , Q es una matriz de N por P instrumentos (incluyendo las variables exógenas

X), y la otra notación es igual que antes. Se puede demostrar que esta estimación es consistente y asintóticamente eficiente. Sin embargo, sus propiedades en muestras finitas dependen de forma decisiva de la elección de los instrumentos y no siempre son tratables.

Un posible problema con la estimación IV es que dicha estimación para ρ no necesariamente tiene un rango aceptable y se pueden obtener valores superiores a 1 en valor absoluto. Típicamente, esto apunta a posibles problemas con la especificación del modelo.

Elección de los instrumentos

En el marco de ecuaciones simultáneas estándar, los instrumentos son las variables exógenas "excluidas". En el modelo de retardo espacial no existe un equivalente simple de esta noción y una serie de propuestas han sido formuladas (Anselin, 1988, Capítulo 7). Recientemente, Kelejian y Robinson (1992) han demostrado que una serie de variables exógenas espacialmente retardadas, matrices de contigüidad de primer orden y de orden superior son el conjunto adecuado. En la práctica, esta serie se puede trincar y solo las variables explicativas espacialmente retardadas de primer orden pueden ser incluidas. Formalmente, el resultado es una matriz Q que contiene X y WX, donde el término constante y otras variables que podrían causar multicolinealidad perfecta están excluidas de WX.

Medidas de ajuste

La medida tradicional de ajuste R^2 no es aplicable a los resultados de la estimación IV. En lugar de ellos, se pueden utilizar dos pseudo medidas R^2 , lo mismo que en la estimación de ML de modelos espacialmente retardados. Una de ellas es una simple relación de la varianza de los valores predichos sobre la varianza de los valores observados para la variable dependiente. La segunda medida es la correlación al cuadrado entre los valores predichos y observados. Sin embargo, ninguna de estas medidas son comparables con los resultados para la estimación de ML.

Contraste de hipótesis

Toda inferencia estadística para las estimaciones obtenidas con el modelo de variables instrumentales se basa en consideraciones asintóticas, es decir, para tamaños de muestra que

se convierten en infinitamente grandes. Esto puede no ser muy fiable para los pequeños conjuntos de datos.

Dada la naturaleza asintótica de los errores estándar, la significación de los coeficientes del modelo individual se basa en una distribución normal estándar (el resultado asintótico) y no en una distribución t de Student (como es el caso de MCO). El llamado test t asintótico (el cual es en realidad un valor z ya que la distribución limitante es la normal estándar) se obtiene como el cociente de la estimación de su error estándar asintótico. (Anselin, 1992)

3.7 EL MODELO DEL ERROR ESPACIAL

3.7.1 Definición

El modelo de error espacial es un caso especial llamado modelo de error no esférico, es decir, una especificación de la regresión para la que los supuestos homoscedásticos (varianza constante) y errores no correlaciones no son satisfechos. La dependencia espacial en el término de error puede asumir un número diferente de formas. Este modelo es la especificación de regresión estándar con un término de error espacial autorregresivo:

$$Y = X\beta + \varepsilon \quad 3.30.$$

$$\varepsilon = \lambda W\varepsilon + \xi \quad 3.31.$$

donde y es un vector de N por 1 observaciones de la variable dependiente, X es una matriz de N por K observaciones de las variables exploratorias, β es un vector de K por 1 coeficientes de regresión, ε es un vector de N por 1 términos de error, W es un retardo espacial de los errores, λ es el coeficiente autorregresivo y ξ es un error de buen comportamiento, con media 0 y matriz de varianza $\sigma^2 I$.

Como consecuencia de la dependencia espacial, el término de error ya no tiene la usual matriz de varianza diagonal, sino que tienen la siguiente forma:

$$E[\varepsilon\varepsilon'] = \Omega = \sigma^2 [(I - \lambda W)'(I - \lambda W)]^{-1} \quad 3.32.$$

Es fácil demostrar que si el coeficiente autorregresivo se conoce, los coeficientes de regresión podrían estimarse por medio de MCO en un modelo con variables espacialmente filtradas $y - \lambda Wy$ y $X - \lambda WX$:

$$y - \lambda Wy = (X - \lambda WX)\beta + \varepsilon \quad 3.33.$$

donde W_y y W_X son variables dependientes y explicativas espacialmente retardadas, y el término de error ϵ sigue los supuestos del modelo clásico.

Este enfoque se conoce como mínimos cuadrados generalizados (GLS). Sin embargo, el coeficiente λ normalmente no se conoce y debe ser estimado de forma conjunta con los coeficientes de regresión. Para una autocorrelación de serie en el dominio del tiempo, dos etapas llamadas procedimiento Factible y procedimiento de Estimación de Mínimos cuadrados Generalizados (MCGF o EGLS) han sido desarrollados, como el familiar estimador de Cochrane-Orcutt. Sin embargo, debido a la simultaneidad implícita por la naturaleza de dependencia espacial, estos procedimientos no son aplicables en el caso espacial y debe llevarse a cabo una estimación de máxima verosimilitud completa.

Como se ha señalado anteriormente, las consecuencias de ignorar la dependencia espacial del error no son tan graves como ignorar la dependencia espacial del retardo. El problema principal es que las estimaciones por MCO se vuelven ineficaces, pero siguen siendo imparciales. Las estimaciones de máxima verosimilitud son compatibles, y, en la mayoría de los casos (incluyendo el caso espacial) también producirá estimaciones objetivas para los coeficientes de regresión (pero no para el llamado parámetro de molestia). Es todavía una cuestión de debate si la eficiencia asintótica superior del enfoque de máxima verosimilitud se traduce en una mayor eficiencia en muestras infinitas. (Anselin, 1992)

3.7.2 Estimación por el Método General de los Momentos (GMM)

Los principios de la aplicación de la estimación de los métodos generales de los momentos (GMM) del modelo espacial del error se presentaron originalmente en Kelejian y Prucha (1998,1999), y, más recientemente, generalizados en una serie de documentos tales como Kelejian y Prucha (2010), Arraiz et al. (2010) y Drukker et al. (2011), denominados conjuntamente como K-P-D. La estrategia de estimación esbozada por KPD consiste en dos medidas importantes. La primera tiene que ver con la estimación de los coeficientes del modelo utilizando un enfoque de mínimos cuadrados generalizados factibles (Mínimos Cuadrados Ponderados Espacialmente), que son una especie de transformación espacial de Cochrane-Orcutt, con el fin de obtener un parámetro λ autoregresivo espacial consistente (aunque todavía no eficiente). La segunda medida trata de la obtención de una estimación eficiente para λ (Anselin et al., 2012; Chasco, 2013)

4 METODOLOGÍA

4.1 DEFINICIÓN DE LA MUESTRA Y LAS VARIABLES

4.1.1 Selección de la muestra

El escenario de este estudio fue la ciudad de Madrid en España. Madrid está dividida en 21 distritos, 127 barrios y 2.358 secciones censales (*Distritos de Madrid, 2013*). A pesar de que es una dependencia administrativa, existen importantes desigualdades socioeconómicas. Por ejemplo, se puede decir que Madrid es caracterizada por una población cada vez más envejecida, es posible detectar claras diferencias territoriales de edad. Por un lado, el centro de Madrid ("Almendra Central") concentra la población de más avanzada edad, la cual se localiza principalmente en los distritos de Salamanca, Chamberí, Centro, Retiro y Tetuán, los cuales registran la mayor tasa de mortalidad. El porcentaje de personas mayores de 65 años con respecto a la población de estos distritos es del 23% en Salamanca, 24% en Chamberí, 16% en Centro, 23% en Retiro y 20% en Tetuán. Por otro lado, la gente joven principalmente ocupa las zonas periféricas del Este y Sur de la ciudad (Vicálvaro, Villa de Vallecas y Barajas), donde tenemos la mayor tasa de natalidad y los mayores porcentajes de niños por debajo de los 15 años de este municipio. A diferencia de lo anterior, el porcentaje de personas por debajo de los 15 años con respecto a la población de estos distritos es del 19% en Vicálvaro, 18% en Villa de Vallecas y 17 % en Barajas. Sin embargo, hay también ciertos distritos, los cuales están caracterizados por altos porcentajes de población de ambos extremos, mayores de 65 años y menos de 15 años, como pueden ser Usera (18% mayores de 65 años y 15% menores de 15 años), Puente de Vallecas (18% mayores de 65 años y 14% menores de 15 años), Villaverde (18% mayores de 65 años y 16% menores de 15 años) y San Blas (17% mayores de 65 años y 16% menores de 15 años). (*Instituto de Estadística de Madrid, 2013*).

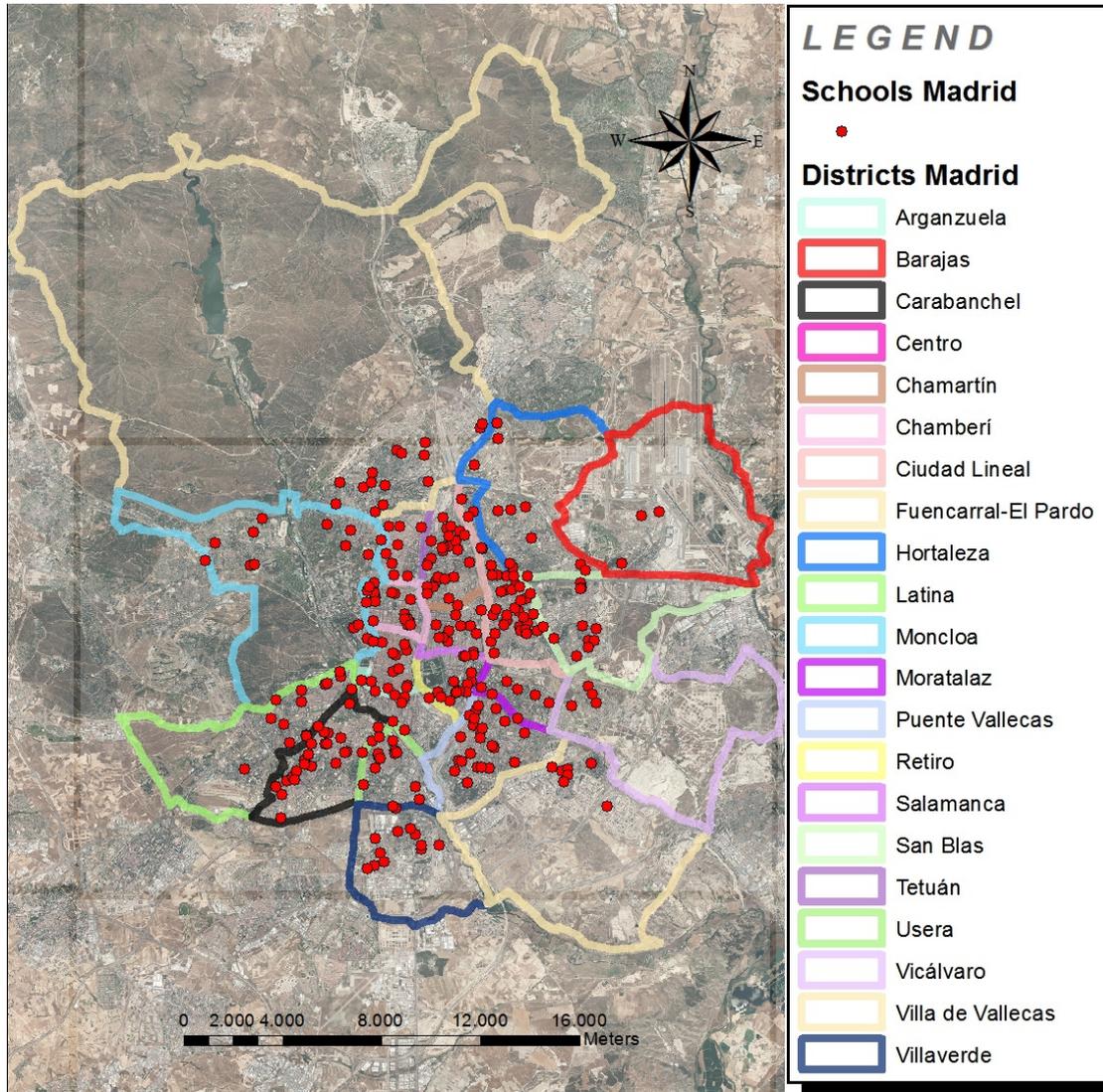
En términos de renta per cápita y niveles de educación, hay también diferencias significativas entre la "Almendra Central" y los distritos periféricos. La población con menos nivel de educación está claramente localizada en los distritos periféricos del Sur (Puente de Vallecas, Vicálvaro, Usera, Villaverde y Carabanchel), los cuales coinciden también con los que menor nivel de renta per cápita poseen en el municipio. La relación entre nivel de educación y renta per cápita en estos distritos es la siguiente, Puente de Vallecas 37% de la población no tiene el graduado en

ESO y solo el 6% tiene una carrera universitaria o superior, la renta per cápita es de 16.970 €; Vicálvaro 26% de la población no tiene el graduado en ESO y el 12% tiene una carrera universitaria o superior, la renta per cápita es de 19.431 €; Usera 35% de la población no tiene el graduado en ESO y solo el 7% tiene una carrera universitaria o superior, la renta per cápita es de 17.002 €; Villaverde 33% de la población no tiene el graduado en ESO y solo el 7% tiene una carrera universitaria o superior, la renta per cápita es de 17.197 €; Carabanchel 29% de la población no tiene el graduado en ESO y el 10% tiene una carrera universitaria o superior, la renta per cápita es de 18.273 €. Inversamente, la mayor concentración de niveles educativos más altos es localizada en la "Almendra Central" (Chamberí, Salamanca, Retiro y Chamartín) como también en los distritos periféricos del Noroeste, Moncloa-Aravaca y Fuencarral-El Pardo, y distritos del Noreste como Hortaleza y Barajas. Al igual que antes, comparando nivel de educación y renta per cápita en estos distritos se observa que, Chamberí 39% de la población posee una carrera universitaria o superior y el 11% no tiene el graduado en ESO, la renta per cápita es de 26.161 €; Salamanca 38% de la población posee una carrera universitaria o superior y solo el 10% no tiene el graduado en ESO, la renta per cápita es de 26.119 €; Retiro 34% de la población posee una carrera universitaria o superior y el 12% no tiene el graduado en ESO, la renta per cápita es de 26.420 €; Chamartín 40% de la población posee una carrera universitaria o superior y solo el 9% no tiene el graduado en ESO, la renta per cápita es de 25.862 €; todos estos distritos en la "Almendra Central". En los distritos periféricos de Noroeste, Moncloa-Aravaca 35% de la población posee una carrera universitaria o superior y el 14% no tiene el graduado en ESO, la renta per cápita es de 24.014 €; Fuencarral-El Pardo 28% de la población posee una carrera universitaria o superior y el 17% no tiene el graduado en ESO, la renta per cápita es de 23.428 €; por último, los distritos del Noreste, Hortaleza 25% de la población posee una carrera universitaria o superior y el 19% no tiene el graduado en ESO, la renta per cápita es de 23.346 €; Barajas 25% de la población posee una carrera universitaria o superior y el 15% no tiene el graduado en ESO, la renta per cápita es de 25.636 €. (*Instituto de Estadística de Madrid, 2010, 2013*).

Finalmente, debería ser recalcada la existencia de clusters (concentraciones) de extranjeros dentro de la ciudad de Madrid. De media, el porcentaje de personas extranjeras con respecto al total de la población por distritos es del 14%. No obstante, el distrito Centro, especialmente, alberga el mayor número de extranjeros, con una diferencia notable en comparación con el resto de distritos (23%). Es seguido de los distritos del Sur, Usera (20%), y Carabanchel (19%) y también por Tetuán (20%), en el extremo noroeste de la "Almendra Central". La menor

proporción de extranjeros se encuentra en Retiro (8%) y los distritos periféricos del Norte y Este de Madrid, Fuencarral-El Pardo (9%), Hortaleza (9%) y Moratalaz (10%). (*Instituto de Estadística de Madrid, 2004, 2013*).

FIGURA 4.1: COLEGIOS PRIVADOS Y CONCERTADOS DISTRIBUIDOS EN LOS DISTINTOS DISTRITOS DE LA CIUDAD DE MADRID



Fuente: elaboración propia

4.1.2 Descripción de las variables

Nosotros utilizamos fundamentalmente datos procedentes del Instituto Nacional de Estadística, como el Censo de Población y Viviendas de 2001 y del Padrón de Habitantes de 2008. Esta fecha viene impuesta por la disponibilidad de la variable de renta familiar disponible por

habitante, que es un de las más importantes de este estudio. Esta variable ha sido elaborada a partir de los datos del precio medio por metro cuadrado de la vivienda en las secciones censales de Madrid, procedentes del portal inmobiliario Idealista.com (base de datos online de bienes inmuebles). Éste es el sitio web de bienes inmuebles más grande del país. Esta base de datos cuenta con más de 36.000 propiedades nuevas y de segunda mano en la ciudad de Madrid, las cuales son ofrecidas tanto por propietarios individuales como por empresas de bienes inmuebles. Cada anuncio de Idealista cuenta con 40 características de la vivienda, incluyendo el precio, m², precio medio por m², distribución, un mapa de la localización de la vivienda, contacto del vendedor, actualizados y con todo tipo de detalles concretos. Está disponible para todos los municipios, sin ningún tipo de restricciones en cualquier tipo de vivienda o rango de precios, aunque las viviendas de segunda mano están más representadas que las de nueva propiedad. Este inconveniente podría sesgar los resultados de nuestro modelo, aunque su impacto es mucho menor en el centro de Madrid donde el espacio para viviendas nuevas es mucho más escaso que en otras áreas de la periferia. Desde Idealista.com se publican, inmediatamente, los precios que piden por las propiedades.

Nuestros registros del precio de la vivienda por metro cuadrado obtenidos de Idealista.com son usados como un indicador indirecto de la renta per cápita. Los datos hacen referencia a Enero de 2008 (*Chasco and Le Gallo, 2012*). Los valores de la variable del precio medio de la vivienda en cada sección censal han servido para distribuir el dato de renta familiar disponible estimado, para el municipio, por el Instituto Lawrence R. Klein de la Universidad Autónoma de Madrid, al que agradecemos nos haya proporcionado esta información.

En cuanto a la información referente a los colegios, se han tomado de la página web del diario El País (que, a su vez, extrae de la Consejería de Educación de la Comunidad de Madrid), en la que se publica la relación de colegios del municipio de Madrid con las calificaciones medias obtenidas por los alumnos de 6º de Primaria:

<http://www.elpais.com/especial/clasificacion-colegios-madrid/>

La ciudad de Madrid tiene 524 colegios (228 públicos y 296 privados), que se han representado en la Figura 4.1. Entre los colegios privados, 256 están subvencionados por el Estado (Colegios concertados). Solo 40 colegios no reciben ninguna financiación por parte del Estado. Por otro lado, la estructura de las clases sociales representadas en esta ciudad, sus indicadores socioeconómicos y la proporción de colegios públicos y privados son típicos de una ciudad española.

Hemos usado un software Sistema de Información Geográfica (GIS) y Google Maps para geocodificar las bases de datos, tanto de las viviendas como de los colegios, a partir de las direcciones postales (a partir de las cuales se han extraído las coordenadas geográficas correspondientes a cada dirección).

TABLA 4.1: N° DE COLEGIOS PRIVADOS POR TIPO Y DISTRITO, EN LA CIUDAD DE MADRID

Distritos	NºColegios Privados	NºColegios Concertados
Arganzuela	1	7
Barajas	1	2
Carabanchel	0	23
Centro	1	10
Chamartín	10	17
Chamberí	3	13
Ciudad Lineal	6	24
Fuencarral-El Pardo	1	15
Hortaleza	4	10
Latina	1	13
Moncloa-Aravaca	4	13
Moratalaz	1	6
Puente de Vallecas	1	22
Retiro	2	11
Salamanca	1	14
San Blas	2	12
Tetuán	0	8
Usera	0	15
Vicálvaro	0	4
Villa de Vallecas	0	7
Villaverde	1	10
<i>Total Colegios</i>	<i>40</i>	<i>256</i>

La accesibilidad espacial de los colegios concertados y privados fue medida como la distancia bloque hasta el supermercado más cercano. La distancia bloque de Manhattan se define de la siguiente manera:

$$d_{ij} = |x_i - x_j| + |y_i - y_j| \quad 4.1.$$

donde d_{ij} es la distancia entre el origen i y el destino j , y x e y son las coordenadas de latitud y longitud del origen i y el destino j . Esta medida de distancia es útil cuando es razonable suponer que los estudiantes viajan a los colegios en una ruta angular, en lugar de en línea recta (una

medida de la distancia Eucladiana) (*Gimpel and, Schuknecht, 2003*). Hemos utilizado un software GIS (GeoDa, Universidad Estatal de Arizona) y SpaceStat 1.93 (terraSeer, Ann Arbor, Mich) para calcular la distancia bloque de Manhattan en base a las coordenadas geográficas de los centroides de las secciones censales y los colegios. Los centroides (centros geográficos) sirven como indicadores de las localizaciones de estas secciones censales. Por lo tanto, la accesibilidad al colegio representa la distancia al colegio privado y concertado más cercano para un residente posicionado en el medio de la sección censal. La mediana de la distancia bloque de Manhattan al colegio más cercano fue 0.40 kilómetros (rango: 0.01 a 10.96 kilómetros). (*Zenk,2004*)

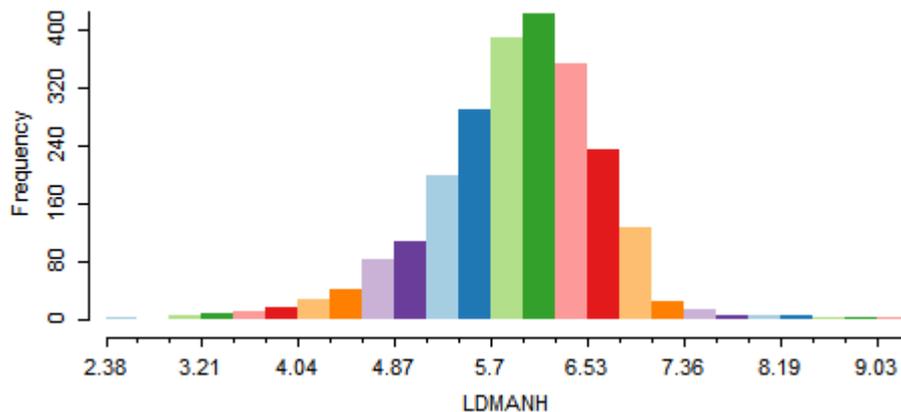
4.2 ANÁLISIS EXPLORATORIO ESPACIAL DE LA VARIABLE DE DISTANCIAS

En este punto, se realiza un análisis exploratorio de la distribución de la variable “distancias al colegio privado y concertado más cercano” en las 2.358 secciones censales existentes en la ciudad de Madrid del año 2008. Hemos obtenido estos datos elaborando la variable con ayuda de un GIS.

El AEDE es considerado como un paso previo al proceso de modelización espacial, dado que permite detectar tendencias y agrupamientos de datos en determinadas zonas geográficas (efectos espaciales), así como las características propias de toda distribución (normalidad, puntos atípicos, etc.). En particular, cuando la variable endógena del modelo presenta alguno de estos rasgos en grado elevado (no normalidad, puntos atípicos, autocorrelación espacial o heterogeneidad espacial) es lógico pensar que los residuos de la regresión puedan adolecer de los mismos. En efecto, si las características de la variable endógena no están suficientemente explicadas por el conjunto de variables exógenas, éstas estarán presentes en los residuos de la regresión, lo que podría alterar las buenas propiedades deseables para la perturbación aleatoria del modelo. Es decir, el AEDE debe constituir la fase previa a toda modelización econométrica espacial (Chasco 2003).

Figura 4.2, se presenta la distribución de las distancias al colegio privado y concertado más cercano. En el gráfico, se representa el histograma de esta variable, que ha sido transformada en logaritmos y desviaciones a la media, por lo que la media es la distancia media de todas las secciones de la ciudad de Madrid de a los colegios privado y concertados más cercanos.

FIGURA 4.2: DISTRIBUCIÓN DE LA DISTANCIA AL COLEGIO PRIVADO Y CONCERTADO MÁS CERCANO DE LAS SECCIONES DE LA CIUDAD DE MADRID, 2008.



Fuente: elaboración propia

En este gráfico se advierte que la distribución de la distancia al colegio privado y concertado más cercano es simétrica, con una gran concentración de secciones con un nivel medio de distancia al colegio privado y concertado más próximo en torno a la media. Destacan dos pequeños grupos de secciones con una distancia medio al colegio privado y concertado más cercano atípicamente alto/bajo.

Todo ello nos indica como la variable sigue una distribución normal.

A continuación, se presentan algunos contrastes que permiten detectar la existencia de los efectos de autocorrelación y heterogeneidad espacial en la variable del precio medio de la vivienda. Para ello, hemos utilizado algunas herramientas propias de la econometría espacial: test I de Moran y diagrama de dispersión de Moran. El efecto de autocorrelación espacial puede ser definido, en palabras de Anselin (2000) como una coincidencia de valores similares en localizaciones cercanas. Este efecto de autocorrelación espacial resulta interesante ya que demuestra la existencia, en una variable (por ejemplo, las distancia al colegio privado y concertado más próximo), de un efecto "contagio" debido a la vecindad geográfica. Esto es lo que suele ser denominado, en la literatura económica, como efectos de desbordamiento ("spillover") o externalidades espaciales. En concreto, se dice que existe autocorrelación espacial positiva en la distancia media al colegio privado y concertado más próximo cuando las distancias mayores o menos tienden a tener lugar de forma agrupada en una zona del espacio geográfico; y esto, con una frecuencia superior a lo que se daría por casualidad. Por otro lado, se dice que existe un efecto de autocorrelación espacial negativa cuando alrededor de las secciones con

distancias medias al colegio privado y concertado mayores se localizan otras secciones vecinas con una distancia media inferior, o viceversa; es decir, cuando se producen agrupamientos espaciales de secciones con una distancia medio al colegio privado y concertado muy diferente. El efecto de autocorrelación espacial puede contrastarse con el test I de Moran (Cliff y Ord, 1973, 1981), que se define como:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{S_0 \sum_{i=1}^n (y_i - \bar{y})^2} \quad 4.2.$$

donde y_i sería, en nuestro caso, el logaritmo de la distancia media al colegio privado y concertado más cercano en las secciones de la ciudad de Madrid; \bar{y} es el valor medio de la variable y ; w_{ij} es uno de los elementos de la matriz W de pesos espaciales; S_0 es un factor de escala igual a la suma de los elementos de la matriz W .

La matriz W es una matriz cuadrada, de orden igual a n (2.358 secciones de la ciudad de Madrid), que expresa las relaciones de vecindad existentes entre las observaciones. Es frecuente definir esta matriz en términos de distancias geográficas (Anselin, 1988; pp. 19-21), ya que de esta forma es más fácil preservar el carácter exógeno que debe tener cuando forma parte de la especificación en los modelos de regresión espacial (Manski, 1993).

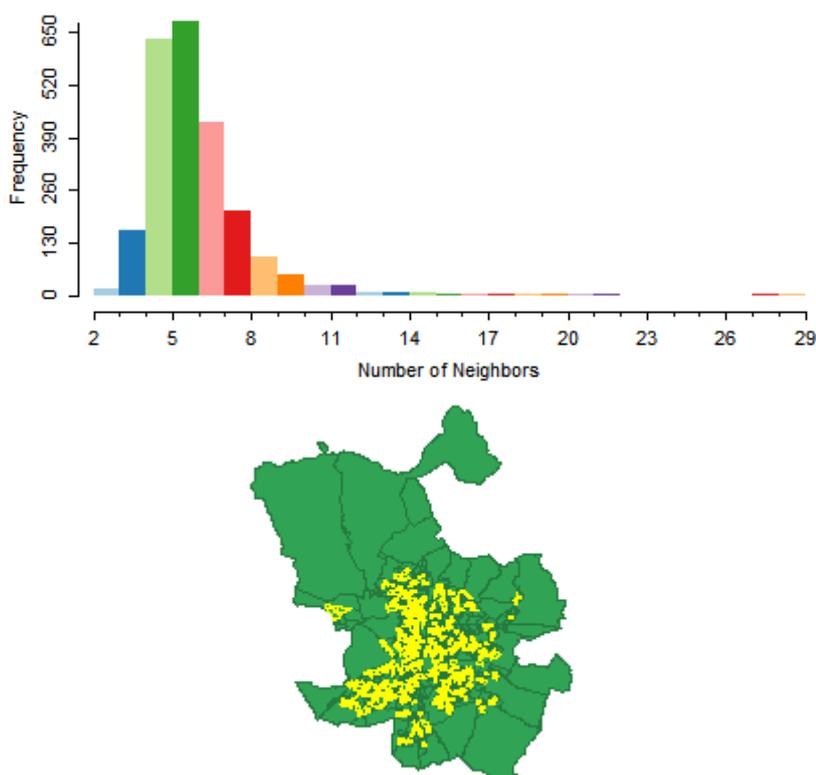
La pre-multiplicación de la matriz W por el vector y_i de la variable distancia media al colegio privado y concertado más cercano da lugar a una nueva variable, Wy , que es la anterior espacialmente retardada (o "retardo espacial"). Por otro lado, es también frecuente someter esta matriz a la llamada "estandarización por filas", que consiste en dividir cada uno de sus elementos por la suma del valor de la fila correspondiente. De este modo, la suma de los elementos de cada fila de la nueva matriz estandarizada por filas, W^* , será siempre la unidad y el valor del retardo espacial W^*y de la distancia al colegio privado y concertado más cercano en la sección i será una media ponderada de la distancia al colegios privado y concertado más cercano en las secciones vecinas a i .

Hemos definido una matriz W para expresar la vecindad. Con ello, deseamos conocer el efecto que la vecindad entre secciones de la ciudad de Madrid tiene sobre la distancia al colegio privado y concertado más cercano, para lo que hemos definido una función inversa y cuadrática de la distancia, tal que $w_{ij} = d_{ij}^{-2}$, siendo d_{ij} la distancia que separa las secciones i, j . Es

decir, para todos y cada una de las secciones de la ciudad de Madrid se ha definido una variable espacialmente retardada de la distancia al colegio privado y concertado más próximo.

Dicha matriz de vecindad utilizada para realizar los siguientes cálculos del test I de Moran está basada en el criterio de vecindad de los 5 vecinos más cercanos. En el cual, las secciones vecinas a una dada serán las 5 secciones más cercanas según la distancia euclídea entre los punto (o centroides, en el caso de unidades poligonales). Para saber el número de vecinos más cercanos, previamente, hemos realizado una matriz binaria de contigüidades por el método de la Torre o Rook (se consideran fronteras a una dada aquellas secciones espaciales que sólo tienen en común algún borde, excluyendo los vértices), dando como resultado el histograma y la selección de vecindad de las secciones con dicho número de vecinos (Figura 4.3).

FIGURA 4.3: HISTOGRAMA DE CONTIGÜIDADES DE LA MATRIZ BINARIA (TORRE) Y SELECCIÓN DEL GRUPO DE VECINDAD CON EL MÁXIMO NÚMERO DE SECCIONES



Fuente: elaboración propia

El cálculo del test I de Moran ofrece resultados muy contundentes para la definición de vecindad. El nivel de autocorrelación espacial positiva de la variable distancia al colegio privado y concertado más cercano es elevado ($I=0.517033$; $p=2.88e-013$), muy superior al valor esperado bajo la hipótesis nula de ausencia de autocorrelación espacial, $E[I]=-0.0036$. Es decir, el

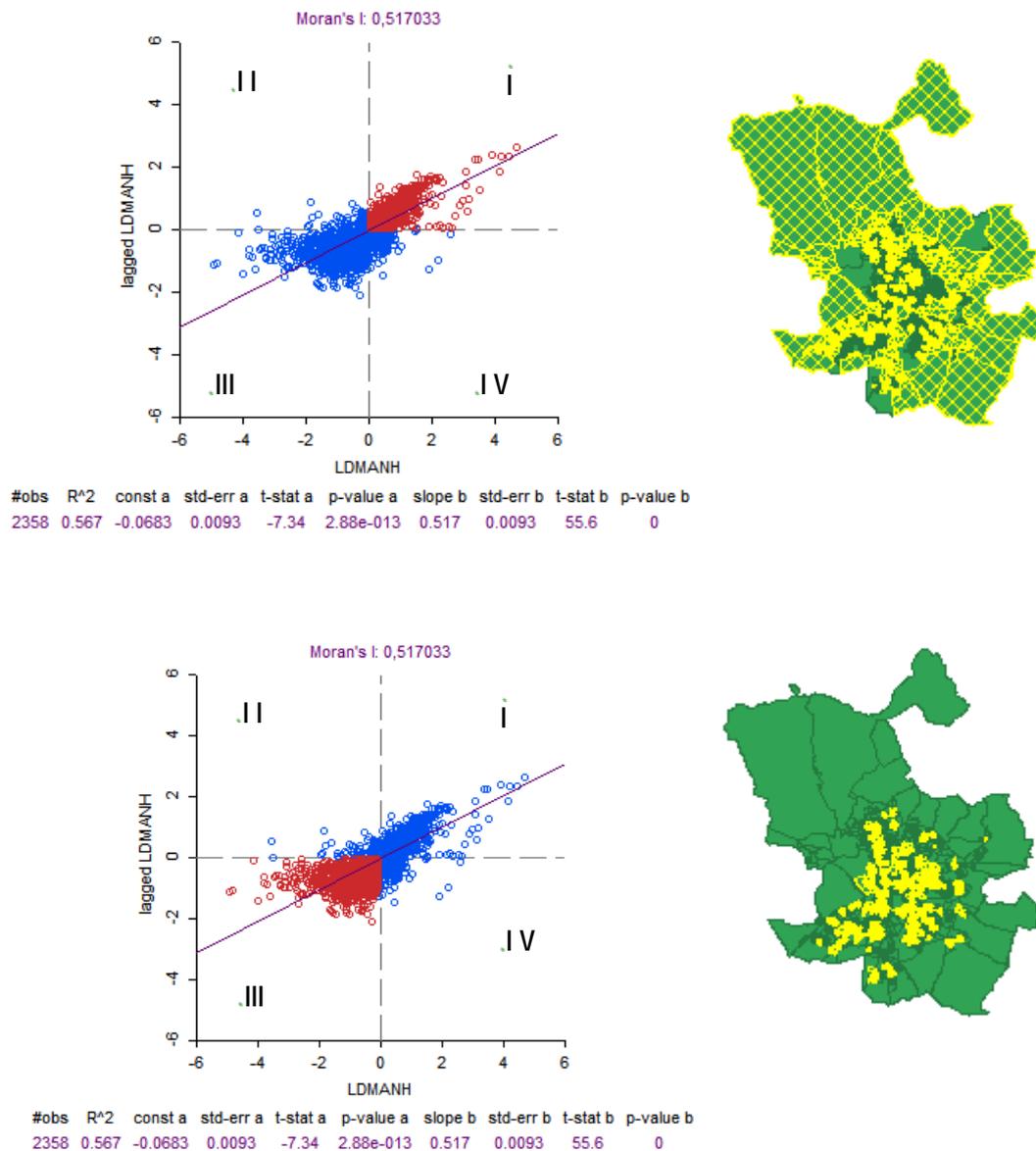
agrupamiento de secciones con niveles altos/bajos de distancia se produce de forma agrupada en el espacio geográfico. La inferencia se basa en el llamado método de permutaciones (999 permutaciones), tal como se expone en (Anselin, 1995). Es decir, parece probado estadísticamente que la distancia media al colegio privado y concertado más cercano aparece agrupado en el espacio geográfico. Es decir, que aquellas secciones de la ciudad de Madrid con una distancia media al colegio privado y concertado más cercano elevada/baja tienden a estar situadas cerca unas de otras con una probabilidad superior de lo que sería considerado como normal. Por esto, los datos de la distancia media al colegio privado y concertado más cercano no pueden ser considerados como observaciones independientes, unas de otras.

En la Figura 4.4, se presenta de forma desagregada las características de este efecto de autocorrelación espacial en la distancia media al colegio privado y concertado más cercano, gracias al diagrama de dispersión de Moran, sugerido por Anselin (1996). En el eje horizontal de este diagrama, se representa la variable estandarizada de la distancia media al colegio privado y concertado más cercano de cada sección de la ciudad de Madrid en 2005 (LDMANH). Así, el valor 0 se corresponde con la distancia media del conjunto de 2.358 secciones de la ciudad de Madrid. Por su parte, en el eje vertical, se representa la variable espacialmente retardada de la distancia media al colegio privado y concertado más cercano (lagged LDMANH). Como ya se ha indicado, el retardo espacial de una sección será la distancia media ponderada de la distancia media al colegio privado y concertado más cercano de las secciones cercanas al mismo.

Los 4 cuadrantes del diagrama de dispersión se corresponden con 4 tipos de asociación espacial local entre una sección y sus vecinas: I ("Alto-Alto"), III ("Bajo-Bajo"), II ("Bajo-Alto") y IV ("Alto-Bajo"). Los cuadrantes I y III se corresponden con asociación espacial positiva, mientras que los II y IV serían los cuadrantes correspondientes a formas de asociación espacial negativa.

En el primer cuadrante del diagrama de dispersión (I) se representan aquellas secciones de la ciudad de Madrid con una distancia al colegio privado y concertado más cercano superior a la media, en 2008, que se encuentran rodeados de secciones de la ciudad de Madrid con una distancia media superior a la media. La selección de las mismas en el diagrama de la izquierda permite su localización en el mapa de la derecha. Se trata de secciones situadas en las zonas periféricas de la ciudad de Madrid de gran dimensión geográfica, aunque también abarca determinadas secciones del centro de la capital.

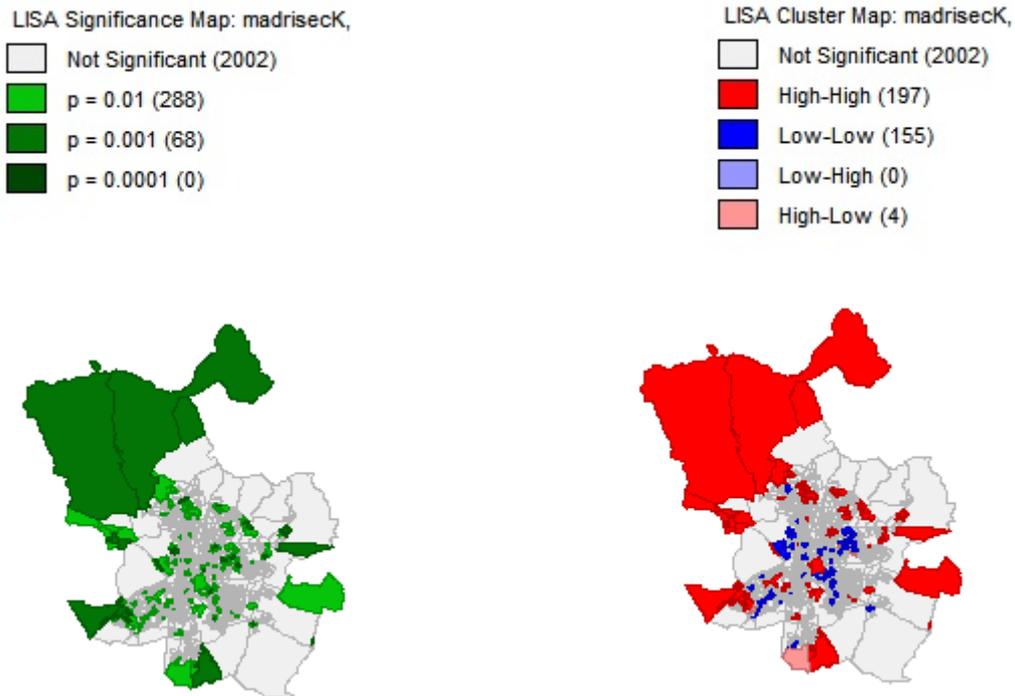
FIGURA 4.4: DIAGRAMAS DE DISPERSIÓN DE MORAN DE LA DISTANCIA AL COLEGIO PRIVADO Y CONCERTADO MÁS CERCANO EN LA CIUDAD DE MADRID, 2005.



Fuente: elaboración propia

Por otro lado, en el tercer cuadrante del diagrama de dispersión (III) se representan aquellas secciones de la ciudad de Madrid con una distancia al colegio privado y concertado más cercano inferior a la media, en 2008, que se encuentran rodeados de secciones de la ciudad de Madrid con una distancia media inferior a la media. Al igual que anteriormente, la selección de las mismas en el diagrama de la izquierda permite su localización en el mapa de la derecha. En este caso nos encontramos con secciones localizadas, en su gran mayoría, en el centro de la ciudad de Madrid, así como algunas secciones del suroeste de la capital.

FIGURA 4.5: CLUSTERS ESPACIALES DE SECCIONES SEGÚN LA DISTANCIA AL COLEGIO PRIVADO Y CONCERTADO MÁS CERCANO



Fuente: elaboración propia

De la representación de los diagramas de dispersión de Moran se deduce que la vecindad es una característica importante que explica las diferencias de distancias al colegio privado y concertado más cercano en Madrid y determina 2 agrupaciones o clusters de secciones de dicha ciudad de Madrid. Para conseguir una mayor exactitud en esta apreciación, nos centraremos en el estadístico I Local de Moran, el cual nos hará ver con más exactitud y certeza esta autocorrelación espacial a nivel local, ya que el estadístico I Global de Moran no es capaz de detectar la presencia de zonas concretas de inestabilidad (zonas calientes o frías), en las que una variable tenga un comportamiento diferente de la tendencia general. Por un lado, estaría el cluster de secciones donde se produce una mayor concentración de distancias más elevadas al colegio privado y concertado más cercano y, por otro, el cluster de secciones donde se produce una mayor concentración de distancias más reducidas.

En la Figura 4.5, siempre siguiendo el método de Anselin de 999 permutaciones y con una significación del 0.01, observamos como el primer grupo o cluster (High-High) se corresponde con secciones de la periferia de la ciudad de Madrid, más concretamente de las zonas norte, noroeste, este y sureste de la capital, muchas de ellas de gran dimensión geográfica, y, también abarca algunas secciones situadas en el centro de la ciudad, pero una pequeña parte. El

segundo grupo o cluster (Low-Low) lo forman secciones situadas, principalmente en el centro de la capital y en el suroeste, las cuales son de una dimensión geográfica mucho menor. La existencia de estos grupos o clusters podría dar lugar a problemas de heteroscedasticidad o inestabilidad paramétrica espacial en el modelo, lo cual será tenido en cuenta más adelante.

4.3 EL MODELO EMPÍRICO

4.3.1 Especificación del modelo inicial

Como ya se ha indicado, este análisis se ha realizado para el territorio urbano de las 2.358 secciones censales de la ciudad de Madrid, cuyo centroide (o punto central) representa la localización de la demanda de educación primaria y secundaria.

Con objeto de contrastar si la accesibilidad de las secciones al colegio privado y/o concertado más cercano depende del nivel de renta medio de los hogares que en ellas residen (*lrta*), se han considerado una serie de variables explicativas de control destacadas por la literatura (ver Tabla 2): a) cercanía al centro urbano y buenas comunicaciones con el resto de la ciudad (*lisol*, *ldens*, *lp6500*, *lcont*, *lnoise*, *lngree*), c) entorno elitista y clase social elevada (*h7180*, *h8190*, *ha90*, *ho70*, *lunem*, *ldirty*, *ldelinc*), b) nivel educativo elevado (*leduc*) y d) densidad de población infantil (*lpo14*).

Por tanto, la variable de accesibilidad podría ser explicada por las 16 variables de la Tabla 4.2.

Sin embargo, un modelo con tantas variables explicativas es muy seguro que tenga problemas de multicolinealidad que dan lugar a elevados valores de la varianza estimada de los coeficientes. De hecho, la estimación de este modelo completo da lugar a un valor en el test del número de multicolinealidad de 1,136.33, muy superior a la cifra óptima de 30-40 propuesta por Anselin (1995).

Para evitar esta situación, hemos contrastado previamente el comportamiento de las variables explicativas para seleccionar aquellas con mayor poder explicativo, clara correlación con la variable de accesibilidad al colegio privado más cercano y menos problemas de multicolinealidad.

TABLA 4.2: VARIABLES QUE INTERVIENEN EN EL MODELO DE ACCESIBILIDAD*

Sigla	Descripción	Fuente estadística
Variable explicada:		
<i>ldist</i>	Distancia (bloque Manhattan) al colegio privado y/o concertado más cercano.	Elaboración propia, 2008
VARIABLES DE RENTA O INGRESO:		
<i>lrta</i>	Renta bruta disponible de los hogares.	Elaboración propia, 2008
Variable de cercanía al centro urbano:		
<i>lisol</i>	Porcentaje de hogares que declaran tener problemas de mala accesibilidad en el entorno de su vivienda.	Censo, 2001
<i>ldens</i>	Densidad de población (en logs.).	Padrón, 2008
<i>lp6500</i>	Porcentaje de población de 65 y más años.	Padrón, 2008
<i>lcont</i>	Porcentaje de hogares que declaran tener problemas de contaminación y malos olores en el entorno de su vivienda.	Censo, 2001
<i>lnoise</i>	Porcentaje de hogares que declaran tener problemas de ruido.	Censo, 2001
<i>lmgree</i>	Porcentaje de hogares que declaran tener problemas de escasez de espacios y zonas verdes en el entorno de su vivienda.	Censo, 2001
Variable de entorno elitista y clase social elevada:		
<i>h7180</i>	Porcentaje de hogares en edificios construidos entre 1971-1980	Censo, 2001
<i>h8190</i>	Porcentaje de hogares en edificios construidos entre 1981-1990	Censo, 2001
<i>ha90</i>	Porcentaje de hogares en edificios construidos después de 1990	Censo, 2001
<i>ho70</i>	Porcentaje de hogares en edificios construidos antes de 1970	Censo, 2001
<i>lunem</i>	Porcentaje de personas en desempleo sobre población activa.	Censo, 2001
<i>ldirty</i>	Porcentaje de hogares que declaran tener problemas de suciedad.	Censo, 2001
<i>ldelinc</i>	Porcentaje de hogares que declaran tener problemas de delincuencia.	Censo, 2001
Variable de nivel educativo:		
<i>ledu</i>	Porcentaje de personas con estudios secundarios y universitarios sobre población de 16 y más años.	Censo, 2001
Variable de densidad de población infantil:		
<i>lp014</i>	Porcentaje de personas de 0 a 14 años.	Padrón, 2008

* Nota: Todas las variables están expresadas en logaritmos.

De este modo, llegamos a un modelo en el que la accesibilidad a los colegios privados y concertados depende de variables de cercanía al centro urbano (*lisol*, *ldens*, *lp6500*, *lmgree*), clase social (*h7180*, *ho70*, *ldelinc*), nivel educativo (*leduc*) e índice de infancia (*lp014*), pero no depende del nivel de renta familiar disponible per cápita (*lrta*):

$$\begin{aligned}
 \ln ldist_t = & \beta_0 + \beta_1 \ln lisol_t + \beta_2 \ln ldens_t + \beta_3 \ln lp6500_t + \beta_4 \ln lmgree_t + \beta_5 \ln h7180_t \\
 & + \beta_6 \ln ho70_t + \beta_7 \ln delinc_t + \beta_8 \ln leduc_t + \beta_9 \ln lp014_t + u_t
 \end{aligned}
 \tag{4.3}$$

4.3.2 Estimación del modelo inicial

Como se puede observar en la Tabla 4.3, todos los coeficientes son bastante significativos y, la gran mayoría negativos, lo que nos indica una tendencia a la convergencia en el conjunto de las variables explicativas del modelo con respecto a la variable a estudiar.

Basándonos en el análisis de la multicolinealidad en el modelo, observamos como el número de condición de multicolinealidad es de 156,07, muy inferior al primer modelo estimado. Este número es superior a la cifra óptima expuesta por Anselin (30-40), pero no es exagerado, por lo tanto, este cierto nivel de multicolinealidad no afecta el estadístico de contraste de significación global, F. La idea es que el estadístico F considera toda la explicación de la variabilidad de la variable endógena, mientras que las variables explicativas comparten una parte de la variabilidad con las demás variables implicadas.

Haciendo referencia a la normalidad de los residuos mediante el test Jarque-Bera comprobamos como existe una probabilidad prácticamente nula, lo que nos mueve a rechazar la hipótesis nula que nos dice que se distribuyen mediante una normal χ^2 con 2 grados de libertad. Esta baja probabilidad indica que los test para heteroscedasticidad y dependencia espacial deben interpretarse con cautela, ya que se basan en el supuesto de normalidad.

Atendiendo a los problemas mencionados anteriormente de heteroscedasticidad, es decir, situación en la que el error de regresión aleatorio no tiene una variación constante en todas las observaciones (no es homoscedástico), hemos realizado tanto el test de Breusch-Pagan como el test de Koenker-Bassett, los cuales discrepan el uno del otro. Mientras que el test de Breusch-Pagan nos muestra un claro rechazo de la hipótesis nula y afirma la presencia de heteroscedasticidad (p-valor=0.01), el test de Koenker-Bassett nos indica todo lo contrario y acepta la hipótesis nula de que el error es homoscedástico (p-valor=0.24). Esto podemos interpretarlo como que algunas variables explicativas del modelo pueden ser heteroscedásticas, pero otras no. Esto podría indicar la existencia de heterogeneidad espacial, lo que nos lleva a ser prudentes en lo que respecta a la eficiencia de los estimadores MCO.

En cuanto a la presencia de autocorrelación espacial se refiere, nos fijaremos en los diferentes test. Empezamos por el test del error I de Moran el cual nos empieza a mostrar esta presencia de autocorrelación espacial siendo muy significativo, aunque no es determinante, ya que este contraste tiene ciertos problemas y, por ese motivo, debemos hacer uso del resto de tests. Pasamos a analizar tanto el test Lagrange Multiplier (estimación) como el test Lagrange Multiplier

(error), ambos muy significativos igualmente, por lo que solo nos queda verificarlo con los tests Robustos, LM (estimación) y LM (error), comprobando que, al igual que los anteriores, también tenemos una significación muy elevada. Aunque el test Robusto LM (error) es, ligeramente, menos significativo. Esto nos lleva a rechazar la hipótesis nula y afirmar la presencia de autocorrelación espacial en el modelo, apuntando hacia un modelo del retardo espacial como especificación más adecuada.

TABLA 4.3: RESULTADOS DE LA ESTIMACIÓN DEL MODELO DE ACCESIBILIDAD A COLEGIOS PRIVADOS Y CONCERTADOS

	Modelo inicial	Modelo del retardo espacial	Modelo del error espacial	Modelo del retardo+error espacial
<i>Método de estimación:</i>	MCO	S2SLS+HAC	GMM+HAC	S2SLS+GMM+HAC
<i>Constante</i>	13.1636514	3.3115	9.5372	3.7931
<i>Variables espacialmente retardadas:</i>				
<i>Wildist</i>	-	0.9555***	-	0.9149***
<i>wε</i>	-	-	0.8399***	0.2666**
<i>Variables de cercanía al centro urbano:</i>				
<i>lisol</i>	0.0295**	-	0.0310**	-
<i>ldens</i>	-0.1882***	-0.1221***	-0.1378***	-0.1262***
<i>lp6500</i>	-0.3695***	-0.0919**	-0.0941**	-0.1121**
<i>lcont</i>	-	0.0641***	0.0664**	0.0574**
<i>Ingree</i>	-0.0446***	-	-	-
<i>Variables de clase social:</i>				
<i>lh7180</i>	-0.0021***	-	-	-
<i>lha90</i>	-	0.0022***	0.0020**	0.0020***
<i>lho70</i>	-0.0014**	-	-	-
<i>ldelinc</i>	-0.2148***	-0.0881***	-0.1905***	-0.0881***
<i>Variable de nivel educativo:</i>				
<i>leduc</i>	-0.4948***	-0.2067***	-0.2449***	-0.2223***
<i>Variable de índice de infancia:</i>				
<i>lp014</i>	-0.4917***	-0.2179***	-0.1822***	-0.2415***
<i>R2</i>	0.1519	-	-	-
<i>Pseudo-R2</i>	-	0.6167	0.1279	0.6156
<i>Multicollinearity #:</i>	156.075	-	-	-
<i>Test Jarque-Bera:</i>	538.248***	-	-	-
<i>Test BP homosced.</i>	20.650**	-	-	-
<i>Test I de Moran</i>	41.784***	-	-	-
<i>Test LM-retardo</i>	1730.669***	-	-	-
<i>Test LM-error</i>	1711.311***	-	-	-
<i>Test robusto LM-retardo</i>	42.477***	-	-	-
<i>Test robusto LM-error</i>	23.119***	-	-	-
<i>Test de Anselin-Kelejian</i>	-	2.1206	-	-

*** Significativo para $p < 0.01$; ** significativo para $p < 0.05$; * significativo para $p < 0.1$; MCO: Mínimos Cuadrados Ordinarios; $w\varepsilon$: término de error autorregresivo espacial; R2: coeficiente de determinación; *Multicollinearity #*: Número condicional de multicolinealidad; BP: test Breusch-Pagan de homoscedasticidad; testsLM: test de autocorrelación espacial del Multiplicador de Lagrange para la matriz de 5 vecinos más cercanos.

De confirmarse este hecho (autocorrelación espacial en los residuos), habría que concluir que los estimadores MCO estarían sesgados y, además, serían inconsistentes. Pese a todo, estos resultados no son totalmente concluyentes dado que la no normalidad de los residuos desvirtúa el comportamiento de los tests basados en el Multiplicador de Lagrange.

En cualquier caso, si ignoramos la posible existencia en el modelo de estos dos problemas, heteroscedasticidad y autocorrelación espacial, obtendremos que, la inferencia basada en los estadísticos T y F podría ser engañosa y que las indicaciones de ajuste basadas en R^2 no serán útiles para valorar la bondad del modelo.

4.3.3 Especificación y estimación de modelos alternativos

El resultado de los tests en el modelo inicial nos lleva a plantearnos la estimación de modelos alternativos de tipo espacial, es decir, modelos que tengan en cuenta la existencia de autocorrelación espacial en los residuos. En concreto, se han estimado tres modelos espaciales alternativos: a) modelo del retardo espacial ("spatial lag model"), b) modelo del error espacial ("spatial error model") y c) modelo del retardo y del error espacial ("spatial lag+error model").

Dado que el término de error no se distribuye normalmente, no podemos estimar estos modelos por el método de máxima-verosimilitud, por lo que optamos por el método de variables instrumentales o mínimos cuadrados en 2 etapas. Además, la existencia de heteroscedasticidad también aconseja el cálculo de una inferencia robusta sobre los estimadores mediante la estimación de una matriz de varianzas y covarianzas de los estimadores de tipo HAC ("Heteroskedasticity and Autocorrelation Consistent Covariance"). Los resultados de esta nueva estimación se encuentran en la Tabla 4.3.

Como puede observarse, el modelo del retardo espacial es capaz de absorber el efecto de autocorrelación espacial en los residuos, ya que el test de Anselin-Kelejian ya no es estadísticamente significativo. Por otro lado, la medida del Pseudo- R^2 se incrementa significativamente ($Pseudo-R^2=0.62$) respecto del valor del coeficiente de determinación en el modelo original ($R^2=0.15$) y el valor del coeficiente autorregresivo espacial está cercano a la unidad ($\rho=0.96$), poniéndose de manifiesto que el efecto de autocorrelación espacial en la variable de distancias constituye casi el 50% de la varianza total de esta variable. La corrección del modelo por el efecto espacial en este modelo altera el valor de algunos estimadores (con

respecto del modelo original), aunque en general, siguen siendo muy significativas las variables de cercanía al centro urbano, clase social, nivel educativo y, sobre todo, el índice de infancia.

Los otros dos modelos ofrecen unos resultados muy parecidos, aunque el valor del Pseudo-R² en el modelo del error espacial es, de nuevo, muy bajo (*Pseudo-R*²=0.13). Por tanto, tal y como se intuía tras los resultados de los tests en el modelo inicial, el modelo finalmente seleccionado para explicar la accesibilidad a los colegios privados es el modelo del retardo espacial, en el que hemos aplicado el método HAC para la inferencia robusta de los estimadores.

5 RESULTADOS

La renta familiar disponible por habitante no parece ser, en sí misma, una variable muy significativa a la hora de explicar la accesibilidad de las familias a los colegios privados y concertados en la ciudad de Madrid.

Pese a todo, hay dos elementos que nos llevarían a matizar estos resultados: 1) que esta variable es una estimación y, por tanto, podría estar sujeta a ciertos errores de medida y 2) que esta variable está muy correlacionada con la variable de nivel educativo, que sí que resulta claramente significativa.

En cualquier caso, las variables que más claramente explican la mayor o menor accesibilidad de los hogares a los colegios privados y concertados son el índice de infancia (en todos los modelos altamente significativa), el nivel educativo, la densidad de población y la ausencia de problemas de delincuencia (como y una variable "proxy" de la clase social).

Estos resultados confirman los estudios anteriormente realizados de que la accesibilidad a estos colegios está muy condicionada por el nivel educativo, la clase social y, en última instancia, también podría decirse que la renta de los hogares. Parece claro que los promotores de este tipo de centros buscan clientes socialmente asentados, que aprecien y potencien la educación de sus hijos (dado que ellos también gozan de un elevado nivel educativo) y, por ese motivo, se encuentran asentados en zonas no conflictivas, aunque con un nivel de densidad de población e índice de infancia que les permita una cuota de mercado suficiente para mantener el negocio.

6 CONCLUSIONES

De los resultados obtenidos, podrían extraerse conclusiones para dos tipos de agentes económicos: los promotores de nuevos colegios privados (y privados concertados), y los encargados de las políticas sociales.

Por un lado, los promotores de nuevos colegios privados y concertados, para que sus centros sean más rentables, deberán elegir localizaciones dentro de la ciudad de Madrid que se encuentren habitadas por clases sociales asentadas, con un alto nivel educativo, que es lo lleva a los progenitores a apreciar y potenciar la educación de sus hijos.

Este tipo de cliente lo localizaremos en zonas con elevado nivel de densidad de población e índice de infancia, que hagan posible una cuota de mercado suficiente para que el colegio sea rentable, aunque siempre en zonas tranquilas y seguras.

Por otro lado, de este trabajo también se extraen interesantes consecuencias de carácter social. En efecto, las políticas sociales deben considerar que existe una desigual accesibilidad a los colegios privados y concertados (que, como ya se ha visto son los más eficientes en términos académicos), que perjudica a las clases sociales más desfavorecidas. Un modo de corregir estos desequilibrios pasa por el fomento de un mayor nivel educativo y cultural en la población. En efecto, nuestro modelo demuestra que el nivel educativo y cultural de los padres, y no tanto su nivel de renta per cápita (aunque sean elementos relacionados), es un factor de primera importancia para el acceso de los hijos a los centros educativos con mayor rendimiento y facilidades al estudio, como son los colegios privados y concertados. Además, este mayor nivel cultural de la sociedad será, a su vez, un revulsivo que llevará a los progenitores a demandar una enseñanza de más calidad, tanto en los colegios privados como en los públicos.

El presente estudio constituye un avance sobre el estado del arte en esta cuestión, al menos en España, en el sentido de que contrasta de forma empírica en una gran ciudad como Madrid, una aseveración que siempre se ha realizado a partir de estudios cualitativos: que *"la accesibilidad espacial a los colegios privados y concertados es más fácil para las clases sociales más acomodadas, en detrimento de los más pobres"*. La principal conclusión de nuestro análisis es que esta aseveración es cierta, pero con matices. Lo que facilita la mejor educación en los alumnos no es el mayor nivel de riqueza de sus familias, sino el nivel educativo y cultural de las mismas.

Finalmente, ha de reseñarse que este estudio tiene algunas limitaciones que deben señalarse. En primer lugar, se trata de un análisis localizado en una ciudad determinada, Madrid, en la que la expansión de colegios privados y concertados está siendo muy importante en los últimos años, pero que no es probablemente extrapolable a todas las ciudades españolas. Resultaría interesante replicar este estudio en otras grandes ciudades pertenecientes a comunidades autónomas con diferente política educativa, como Barcelona, Valencia, Sevilla o Bilbao.

Además, debe indicarse que la variable de renta familiar disponible es una estimación realizada a partir de datos de precios de la vivienda (dado que no se encuentra disponible en los institutos de estadística), con lo que pudiera existir algún sesgo que limitara los resultados. Un análisis posterior podría consistir en investigar la existencia de mejores métodos y variables para la depuración de este importante indicador.

Otra cuestión interesante a cuantificar es la existencia o no de diferencias significativas en las calificaciones escolares (prueba de 6º de Primaria) obtenidas en las pruebas de matemáticas y lengua por alumnos de centros privados, privados-concertados y públicos, así como las variables que afectan a este rendimiento y las posibles diferencias –si existen- entre estos dos tipos de conocimiento.

Por último, un ejercicio interesante de marketing sería la estimación del área de mercado, y su mercado potencial, de un nuevo colegio privado en Madrid.

7 BIBLIOGRAFÍA

Anselin L., (1988). *Spatial Econometrics: Methods and Models*. (Boston, Mass: Kluwer Academic).

Anselin L., (1992). *SpaceStat Tutorial. A Workbook for Using SpaceStat in the Analysis of Spatial Data*. Technical Report S-92-1, National Center for Geographic Information and Analysis. Santa Barbara, CA: University of California.

Anselin L., (1995). "Local indicators of spatial association-LISA". *Geographical Analysis*, 27, pp. 93-115.

Anselin L., (1999). Interactive techniques and exploratory spatial data analysis. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW, eds. *Geographic Information Systems: Principles, Techniques, Management, and Applications*. (New York, NY: John Wiley & Sons;: 253–266).

Anselin L., Varga A. and Acs Z., (2000). "Geographic and sectorial characteristics of academic knowledge externalities". *Papers in Regional Science*, 79 (4), pp. 435-443.

Anselin L., (2001a). " Spatial econometrics". En Baltagi, B. (ed). *A companion to theoretical econometrics*. Oxford: Basil Blackwell, pp. 310-330.

Anselin L., Amaral P.V. and Arribas-Bel D., (2012). "Technical aspects of implementing GMM estimation of the spatial error model in PySAL and GeoDaSpace", Working Paper 2/12 (GeoDa Center, Arizona State University).

Arbia G., (2006). *Spatial econometrics: Statistical foundations and applications to regional convergence*. Berlin: Springer Verlag.

Arraiz I., Drukker D.M., Kelejian H.H. and Prucha I.R., (2010). A spatial Cliff-Ord type model with heteroskedastic innovations: Small and large sample results, *J. Regional Sci.* 50 592-

614.

Bailey T.C. and Gatrell A.C., (1995). *Interactive Spatial Data Analysis*. Essex, (UK: Addison Wesley Longman).

Baller R., Anselin L., Messner S. and Hawkins D., (2006). "Structural covariates of U.S. country homicide rates: incorporating spatial effects". *Criminology*, 39 (3), pp. 561-588.

Belsley et al., (1980). The condition number is the square root of the ratio of the largest to the smallest eigenvalue of the matrix $X'X$, after standardization.

Bernal J.L., (2005). *Journal of Education Policy*. Parental choice, social class and market forces: the consequences of privatization of public services in education. (University of Zaragoza, Spain. 20 (6): 779–792).

Breusch and Pagan, (1979). Lagrange Multiplier test developed.

Ceccato V., Haining R. and Signoreta P., (2002). "Exploring offence statistics in Stockholm city using spatial analysis tools". *Annals of the American Association of Geographers*, 92 (1), pp. 29-51.

Chasco C. and Fernández G., (2009). *Análisis de datos espacio-temporales para la economía y el geomarketing*. La Coruña. pp. 56-94.

Chasco C. and Le Gallo J., (2012). *Economic Geography*. The Impact of Objective and Subjective Measures of Air Quality and Noise on House Prices: A Multilevel Approach for Downtown Madrid. (89(2):127–148).

Chasco C., (2013). *GeoDaSpace: a resource for teaching spatial regression models*. Madrid.

Chica J., (1994). *Teoría de las variables regionalizadas. Aplicación en economía espacial y valoración inmobiliaria*. Granada: Universidad de Granada.

Cox K.R., (1969). "The voting decision in a spatial context". En Board, C., Chorley, R.J., Hagget, P., Stoddart, D.R. (eds). *Progress in Geography*, 1. London: Edward Arnold, pp. 81-117.

Cressie N.A.C., (1993), *Statistics for Spatial Data*. (New York, NY: John Wiley & Sons).

Distritos de Madrid (2013). Available online at: <http://www.distritosdemadrid.com/> (accessed 4 November 2013).

Dronkers J. and Robert P., (2003). *The Effectiveness of Public and Private Schools from a Comparative Perspective*. European University Institute, Florence. Department of Political and Social Sciences. pp. 7-26.

Es Madrid (2004). Observatorio Económico. Reequilibrio Territorial en la Comunidad de Madrid. Estudio Comparativo Situación Social y Económica. Available online at: http://www.esmadrid.com/recursos/doc/es/Negocio/ObservatorioEconomico/1680794832_164200711212.pdf (accessed 3 October 2013).

Fischer M.M. and Varga A., (2003). "Production of knowledge and geographically mediated spillovers from universities: special econometric perspective and evidence from Austria". *Annals of Regional Science*, 37 (2). Pp. 303-322.

Gimpel J.G. and Schuknecht J.E., (2003). Political participation and the accessibility of the ballot box. *PolitGeogr.* (22:471–488).

Haining R. (2003). *Spatial Data Analysis: Theory and Practice*. (Cambridge, Mass: Cambridge University Press).

Instituto de Estadística de Madrid (2013). Padrón de Habitantes, January 1st. Available online at: <http://www.madrid.es/portales/munimadrid/es/Inicio/Ayuntamiento/Estadistica/Areas-de-informacion-estadistica/Demografia-y-poblacion?vgnextfmt=default&vgnnextchannel=4e77a4f094b74210VgnVCM2000000c205a0aRCRD>

Instituto de Estadística de Madrid (2010). Economía. Available online at: <http://www.madrid.es/portales/munimadrid/es/Inicio/Ayuntamiento/Estadistica/Areas-de-informacion-estadistica/Economia?vgnextfmt=default&vgnnextchannel=9f69463ce6b74210VgnVCM2000000c205a0aRCRD>

Kaluzny S.P., Vega S.C., Cardoso T.P. and Shelly A.A., (1998). *S+ SpatialStats User's Manual for Windows and UNIX*. (New York, NY: Springer).

Kelejian H.H. and Robinson D.P., (1992). "Spatial autocorrelation. A new computationally simple test with an application to per capita country police expenditures". *Regional Science and Urban Economics*, 22, 317-331.

Kelejian H.H. and Prucha I.R., (1998). A generalized spatial two-stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances, *J. Real Estate Financ.* 17 99-121.

Kelejjan H.H. and Prucha I.R., (1999). A generalized moments estimator for the autoregressive parameter in a spatial model, *Int. Econ. Rev.* 40 509-533.

Kelejjan H.H. and Prucha I.R., (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *J. Econometrics* 157, 53-67.

Kiefer and Salmon, (1983). It is based on both skewness and kurtosis of the residuals.

Koenker, (1981). Studentized version suggested.

Koenker and Bassett, (1982). Studentized version suggested.

Le Gallo J. and Chasco C., (2008). "Spatial analysis of urban growth in Spain, 1900-2001". *Empirical Economics*, 34 (1), pp. 59-80.

Mancebón M.J. and Muñoz M.A., (2008). Private versus Public High Schools in Spain: Disentangling Managerial and Programme Efficiencies. University of Zaragoza, Zaragoza, Spain; and University of Oviedo, Oviedo, Spain. *The Journal of the Operational Research Society*, Vol. 59, No. 7, pp. 892-901.

Moreno R and Vayá E., (2000). *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial*. Barcelona: Editions Universitat de Barcelona.

Rey S.J. and Montouri B.D., (1999). "U.S. regional income convergence: A spatial econometric perspective". *Regional Studies*, 33, pp. 143-156.

Rey S.J., (2001). "Spatial empirics for economic growth and convergence". *Geographical Analysis*, 33, pp. 195-214.

Rey S.J., (2004a). "Spatial dependence in the evolution of regional income distributions". En A. Getis, J. Mur, H. Zoeller (eds.), *Spatial Econometrics and Spatial Statistics*. Hampshire: Palgrave, pp. 194-213.

Rey S.J., (2004b). "Spatial analysis of regional income inequality". En M. Goodchild, D. Janelle (eds.), *Spatially Integrated Social Science: Examples in Best Practise*. Oxford: Oxford University Press, pp. 280-290.

White, (1980). The White test consists of N times the R^2 in an auxiliary regression of the squared OLS residuals on all cross products between the explanatory variables.

Zenk S.N., (2004). Neighborhood Racial Composition, Neighborhood Poverty, and Food Access in Metropolitan Detroit: Geographic Information Systems and Spatial Analysis [dissertation]. (Ann Arbor, Mich: University of Michigan).

8 ANEXOS

8.1. Relación de colegios privados y concertados del municipio de Madrid por distritos

TABLA 8.1: COLEGIOS PRIVADOS Y CONCERTADOS POR DISTRITOS Y NOTA MEDIA

Distrito	Tipo de Colegio	Colegio	Nota media
Centro	Concertado	Centro de Estudios Castilla	22,32
Centro	Privado	Colegio Privado Internacional SEK Santa Isabel	28,32
Centro	Concertado	Institución Divino Maestro	18,83
Centro	Concertado	La Merced	19,19
Centro	Concertado	La Salle-La Paloma-Fundación Lara	29,18
Centro	Concertado	Madres Mercedarias de D. Juan de Alarcón	21,04
Centro	Concertado	Purísima Concepción	17,04
Centro	Concertado	Real Colegio Santa Isabel-La Asunción	23,62
Centro	Concertado	Sagrado Corazón	26,05
Centro	Concertado	San Alfonso	19,71
Centro	Concertado	Santa Isabel	24,74
Arganzuela	Concertado	Beata María Ana de Jesús	15,29
Arganzuela	Privado	CAU El Mus "Intermezzo"	24,52
Arganzuela	Concertado	Dosparques	25,04
Arganzuela	Concertado	Luz Casanova	27,86
Arganzuela	Concertado	Salesianos Atocha	20,82
Arganzuela	Concertado	San Javier	34,53
Arganzuela	Concertado	San Jose	25,56
Arganzuela	Concertado	San Saturio	26,26
Retiro	Privado	Afuera III A	26,87
Retiro	Concertado	Agustiniano	29,31
Retiro	Concertado	Arcángel	26,81
Retiro	Privado	Los Olmos	25,56
Retiro	Concertado	Montserrat	23,09
Retiro	Concertado	Ntra. Sra. de La Paz	24,55
Retiro	Concertado	Pureza de María	29,78
Retiro	Concertado	Reinado del Corazón de Jesús	21,88
Retiro	Concertado	Retiro	27,58
Retiro	Concertado	Sagrado Corazón de Jesús	26,94
Retiro	Concertado	Santa Elizabeth	29,96
Retiro	Concertado	Santa María del Pilar	27,29
Retiro	Concertado	Virgen de Atocha	28,76
Salamanca	Concertado	Calasancio-Ntra.Sra.de Las Escuelas Pías	24,58
Salamanca	Concertado	Dulce Nombre de Jesús	24,23
Salamanca	Concertado	Fundación Caldeiro	25,62
Salamanca	Concertado	Inmaculada Concepción	28,37
Salamanca	Concertado	Jesús María	26,47
Salamanca	Concertado	Menesiano	27,24

Distrito	Tipo de Colegio	Colegio	Nota media
Salamanca	Concertado	Natividad de Nuestra Señora	27,15
Salamanca	Concertado	Ntra. Sra. de Loreto	25,33
Salamanca	Concertado	Ntra. Sra. del Pilar	29,19
Salamanca	Concertado	Real Colegio Ntra. Sra. de Loreto	19,36
Salamanca	Concertado	Sagrada Familia (Calle Jorge Juan)	28,78
Salamanca	Concertado	Sagrada Familia (Calle Eduardo Aunos)	20,71
Salamanca	Concertado	Santa Ana y San Rafael	27,29
Salamanca	Concertado	Santa Susana	20,94
Salamanca	Privado	Viancor	15,64
Chamartín	Privado	Altair Colegio Internacional	26,10
Chamartín	Concertado	Cardenal Spinola	20,99
Chamartín	Concertado	Claret	27,46
Chamartín	Concertado	Corazón Inmaculado	26,96
Chamartín	Privado	Estilo	25,67
Chamartín	Concertado	Fundación Santamarca	23,98
Chamartín	Concertado	Hispano Alemán	28,00
Chamartín	Privado	Las Naciones	28,15
Chamartín	Privado	Macarena	27,03
Chamartín	Concertado	María Virgen	25,32
Chamartín	Concertado	Nervión	28,25
Chamartín	Concertado	Ntra. Sra. de La Consolación	23,77
Chamartín	Concertado	Ntra. Sra. de La Merced	26,38
Chamartín	Concertado	Ntra. Sra. de La Sabiduría	24,24
Chamartín	Privado	Ntra. Sra. de Las Maravillas (La Salle Maravillas)	27,43
Chamartín	Concertado	Ntra. Sra. del Recuerdo	28,17
Chamartín	Concertado	Paraíso Sagrados Corazones	24,61
Chamartín	Concertado	Sagrado Corazón	27,07
Chamartín	Privado	Saint Anne'sSchool (Británico)	21,98
Chamartín	Concertado	San Agustín	26,14
Chamartín	Concertado	San Francisco de Asís	23,96
Chamartín	Privado	San Patricio	31,15
Chamartín	Concertado	San Ramón y San Antonio	25,13
Chamartín	Concertado	Santa Catalina de Sena	24,67
Chamartín	Privado	Santa Cristina	23,58
Chamartín	Privado	Santa Illa	13,40
Chamartín	Privado	Santa María del Valle Cep	15,31
Tetuán	Concertado	Divino Corazón	21,75
Tetuán	Concertado	Gil-Díaz	14,47
Tetuán	Concertado	María Auxiliadora	22,66
Tetuán	Concertado	Ntra. Sra. del Pilar	24,05
Tetuán	Concertado	Patrocinio de San José	26,77
Tetuán	Concertado	San Antonio	17,18
Tetuán	Concertado	San José	20,12
Tetuán	Concertado	San Juan Bautista	21,41

Distrito	Tipo de Colegio	Colegio	Nota media
Chamberí	Concertado	Amor Misericordioso	25,63
Chamberí	Concertado	Blanca de Castilla	21,40
Chamberí	Privado	CEXTRAN ESP "Suola Materna Italiana de Madrid"	24,52
Chamberí	Concertado	Chamberí Hermanos Maristas	22,54
Chamberí	Concertado	Decroly	23,17
Chamberí	Concertado	Divina Pastora	25,06
Chamberí	Concertado	El Porvenir	24,68
Chamberí	Privado	F.E.M	25,80
Chamberí	Concertado	Jesús Maestro	25,80
Chamberí	Concertado	La Inmaculada-Marillac	25,68
Chamberí	Concertado	La Salle-San Rafael	24,02
Chamberí	Concertado	María Inmaculada	20,23
Chamberí	Concertado	Sagrado Corazón de Jesús	22,71
Chamberí	Concertado	San Diego y San Vicente	30,50
Chamberí	Concertado	Santa María del Yermo	24,61
Chamberí	Privado	ScuolaStatale Italiana Di Madrid – LiceoItaliano	22,15
Fuencarral-El Pardo	Concertado	El Prado	26,28
Fuencarral-El Pardo	Concertado	El Valle III	18,81
Fuencarral-El Pardo	Concertado	Fuentelarreyna	27,33
Fuencarral-El Pardo	Concertado	La Salle-Sagrado Corazón	23,80
Fuencarral-El Pardo	Concertado	Las Tablas Valverde	26,80
Fuencarral-El Pardo	Privado	Montealto	25,50
Fuencarral-El Pardo	Concertado	Ntra. Sra. de La Vega	23,84
Fuencarral-El Pardo	Concertado	Sagrado Corazón	26,39
Fuencarral-El Pardo	Concertado	Sagrado Corazón	28,26
Fuencarral-El Pardo	Concertado	San José de Begoña	20,79
Fuencarral-El Pardo	Concertado	San Rafael Arcangel-Santa Luisa	22,68
Fuencarral-El Pardo	Concertado	Santa Joaquina de Vedruna	23,33
Fuencarral-El Pardo	Concertado	Tres Olivos	23,93
Fuencarral-El Pardo	Concertado	Valdeluz	28,82
Fuencarral-El Pardo	Concertado	Virgen de Mirasierra	26,37
Fuencarral-El Pardo	Concertado	Virgen del Pilar	22,40
Moncloa-Aravaca	Concertado	Bernadette	27,73
Moncloa-Aravaca	Privado	Estudio	23,69
Moncloa-Aravaca	Concertado	Fray Luis De León	27,61
Moncloa-Aravaca	Concertado	La Inmaculada	26,38
Moncloa-Aravaca	Concertado	Liceo Sorolla B	24,52
Moncloa-Aravaca	Concertado	Los Robles	25,64
Moncloa-Aravaca	Concertado	Madres Concepcionistas	25,62
Moncloa-Aravaca	Concertado	María Reina	22,13
Moncloa-Aravaca	Privado	Mater Salvatoris	30,18
Moncloa-Aravaca	Concertado	Ntra. Sra. del Buen Consejo	28,14
Moncloa-Aravaca	Concertado	Ntra. Sra. del Pilar	28,22
Moncloa-Aravaca	Concertado	Sagrado Corazón de Jesús	23,30

Distrito	Tipo de Colegio	Colegio	Nota media
Moncloa-Aravaca	Concertado	Sagrados Corazones	26,33
Moncloa-Aravaca	Privado	Santa María de Los Rosales	20,39
Moncloa-Aravaca	Privado	Santa María del Camino	24,41
Moncloa-Aravaca	Concertado	Santísimo Sacramento	19,83
Moncloa-Aravaca	Concertado	Santo Domingo	21,85
Latina	Concertado	La Salle	23,61
Latina	Concertado	Lourdes	21,15
Latina	Concertado	Luis Feito	19,17
Latina	Concertado	María Cristina	17,00
Latina	Concertado	Nebrija-Rosales	23,88
Latina	Concertado	Ntra. Sra. de Las Escuelas Pías	22,01
Latina	Concertado	Ntra. Sra. del Sagrado Corazón	24,34
Latina	Concertado	San Bernardo	18,19
Latina	Concertado	San Buenaventura	25,99
Latina	Concertado	San José-Lucero	18,61
Latina	Concertado	San Juan García	26,80
Latina	Concertado	San Miguel Arcángel	20,26
Latina	Concertado	Santa Cristina	21,21
Latina	Privado	Santa Gema Galgani	20,77
Carabanchel	Concertado	Amoros	25,94
Carabanchel	Concertado	Cervantes	9,15
Carabanchel	Concertado	Colegio Ártica	19,98
Carabanchel	Concertado	Con-Pe	13,35
Carabanchel	Concertado	Institución Profesional Salesiana	20,25
Carabanchel	Concertado	Liceo Ibérico	14,55
Carabanchel	Concertado	Luz Casanova	21,17
Carabanchel	Concertado	María Inmaculada	25,42
Carabanchel	Concertado	Nazaret-Oporto	25,24
Carabanchel	Concertado	Ntra. Sra. de Las Escuelas Pías	23,27
Carabanchel	Concertado	Ntra. Sra. de Los Dolores	24,31
Carabanchel	Concertado	Ntra. Sra. del Carmen-Nájera	21,00
Carabanchel	Concertado	Ntra. Sra. de La Merced	15,11
Carabanchel	Concertado	San Alberto Magno	19,13
Carabanchel	Concertado	San Gabriel	25,79
Carabanchel	Concertado	San Vicente de Paul	20,27
Carabanchel	Concertado	Santa Beatriz de Silva	19,85
Carabanchel	Concertado	Santa Cruz	27,95
Carabanchel	Concertado	Santa María de Los Apóstoles	22,70
Carabanchel	Concertado	Santa María del Bosque	20,91
Carabanchel	Concertado	Santa María Micaela	20,52
Carabanchel	Concertado	Santa Rita	17,41
Carabanchel	Concertado	Vedruna	21,67
Usera	Concertado	Busmar	8,80
Usera	Concertado	Central	25,28

Distrito	Tipo de Colegio	Colegio	Nota media
Usera	Concertado	La Natividad	14,39
Usera	Concertado	Luz Casanova	27,17
Usera	Concertado	María Reina	14,71
Usera	Concertado	Mater Purissima	21,86
Usera	Concertado	Ntra. Sra. de Fátima	20,85
Usera	Concertado	Ntra. Sra. de La Providencia	24,62
Usera	Concertado	Obra So.N.S. Montserrat-S.Simón y S. Judas	9,79
Usera	Concertado	Rafaela Ybarra	22,00
Usera	Concertado	Sagrado Corazón	20,51
Usera	Concertado	San Fermín	16,46
Usera	Concertado	San Luis-Felca	15,32
Usera	Concertado	San Viator	23,69
Usera	Concertado	Santa María	19,90
Puente de Vallecas	Concertado	Asunción-Vallecas	20,77
Puente de Vallecas	Concertado	Centro Cultural Palomeras	25,11
Puente de Vallecas	Concertado	Centro Cultural Salmantino	14,54
Puente de Vallecas	Concertado	Ciudad de Los Muchachos	17,34
Puente de Vallecas	Concertado	Cumbre-Oxford	19,22
Puente de Vallecas	Concertado	Divina Pastora	22,63
Puente de Vallecas	Concertado	El Ave María	16,38
Puente de Vallecas	Concertado	Gredos San Diego Vallecas	24,17
Puente de Vallecas	Concertado	La Milagrosa	20,16
Puente de Vallecas	Concertado	Liceo Consul	25,20
Puente de Vallecas	Concertado	Los Tilos	18,89
Puente de Vallecas	Concertado	María Inmaculada	20,89
Puente de Vallecas	Privado	Mater Clementisima	23,38
Puente de Vallecas	Concertado	Ntra. Sra. de La Estrella	26,53
Puente de Vallecas	Concertado	Raimundo Lulio	23,74
Puente de Vallecas	Concertado	Sagrada Familia de Urgel	24,02
Puente de Vallecas	Concertado	San José	25,35
Puente de Vallecas	Concertado	San José	23,84
Puente de Vallecas	Concertado	Santa María de Los Pinos	22,83
Puente de Vallecas	Concertado	Santa Rafaela María	17,48
Puente de Vallecas	Concertado	Santo Ángel de La Guarda	21,06
Puente de Vallecas	Concertado	Tajamar	25,78
Puente de Vallecas	Concertado	Ursulinas de Palomeras	19,92
Moratalaz	Concertado	Gredos San Diego-Moratalaz	26,75
Moratalaz	Concertado	Ntra. Sra. de Moratalaz	21,19
Moratalaz	Concertado	Sagrada Familia	29,79
Moratalaz	Concertado	San Martín	19,51
Moratalaz	Privado	Santo Ángel	25,95
Moratalaz	Concertado	Senara	29,22
Moratalaz	Concertado	Siglo XXI	23,75
Ciudad Lineal	Concertado	Ágora	19,30

Distrito	Tipo de Colegio	Colegio	Nota media
Ciudad Lineal	Privado	Arturo Soria	29,72
Ciudad Lineal	Privado	Aula Nueva	24,76
Ciudad Lineal	Concertado	Casa de La Virgen	24,85
Ciudad Lineal	Concertado	Cenit	18,91
Ciudad Lineal	Concertado	Centro Cultural Elfo	25,39
Ciudad Lineal	Privado	Colegio Europeo Almazán	22,17
Ciudad Lineal	Concertado	Corazón De María	21,08
Ciudad Lineal	Concertado	Corpus Christi	22,70
Ciudad Lineal	Concertado	Espíritu Santo	24,59
Ciudad Lineal	Privado	Huérfanos de La Armada	26,14
Ciudad Lineal	Privado	Joyfe	29,41
Ciudad Lineal	Concertado	Liceo Madariaga	14,72
Ciudad Lineal	Concertado	Madre de Dios	21,48
Ciudad Lineal	Concertado	María Reina	24,37
Ciudad Lineal	Concertado	Mater Inmaculata	26,52
Ciudad Lineal	Concertado	Mirasierra	17,48
Ciudad Lineal	Concertado	Montpellier	24,76
Ciudad Lineal	Concertado	Ntra. Sra. de Fátima	19,91
Ciudad Lineal	Concertado	Ntra. Sra. de La Merced	22,77
Ciudad Lineal	Concertado	Ntra. Sra. de Las Victorias	26,76
Ciudad Lineal	Concertado	Ntra. Sra. del Buen Consejo	25,50
Ciudad Lineal	Concertado	Nuevo Equipo	24,96
Ciudad Lineal	Concertado	Obispo Perello	22,42
Ciudad Lineal	Privado	Ramón y Cajal	24,93
Ciudad Lineal	Concertado	San José	20,57
Ciudad Lineal	Concertado	Santa María del Carmen	28,35
Ciudad Lineal	Concertado	Santísimo Sacramento	23,47
Ciudad Lineal	Concertado	Santo Domingo Savio	19,00
Ciudad Lineal	Concertado	Stella Maris	23,40
Hortaleza	Concertado	Asunción Cuestablanca	23,51
Hortaleza	Concertado	Azorín	24,26
Hortaleza	Privado	Bristol	20,11
Hortaleza	Concertado	Colegio Valdefuentes	18,68
Hortaleza	Concertado	Cristo Rey	20,34
Hortaleza	Concertado	El Valle II	20,98
Hortaleza	Concertado	La Inmaculada	27,09
Hortaleza	Privado	Ntra. Sra. Santa María	32,84
Hortaleza	Privado	Parque Conde de Orgaz	26,36
Hortaleza	Concertado	Patrocinio De María	19,29
Hortaleza	Privado	San José del Parque	28,35
Hortaleza	Concertado	Santa Francisca Javier Cabrini	27,15
Hortaleza	Concertado	Santa María de La Hispanidad	22,16
Hortaleza	Concertado	Virgen Milagrosa	17,15
Villaverde	Concertado	Agrupación Escolar Europa	26,00

Distrito	Tipo de Colegio	Colegio	Nota media
Villaverde	Concertado	Comunidad Infantil de Villaverde	25,73
Villaverde	Concertado	Institución Docente Gyrsa	16,45
Villaverde	Privado	Internacional Nuevo Centro	26,66
Villaverde	Concertado	Liceo Oroquieta	18,81
Villaverde	Concertado	Liceo Rosales	20,67
Villaverde	Concertado	Ntra. Sra. de Los Ángeles	22,07
Villaverde	Concertado	Ntra. Sra. del Carmen	25,69
Villaverde	Concertado	Sagrados Corazones	21,50
Villaverde	Concertado	San Pedro	27,88
Villaverde	Concertado	Vedruna	22,79
Villa de Vallecas	Concertado	Gredos San Diego Las Suertes	25,62
Villa de Vallecas	Concertado	Liceo Versalles	20,04
Villa de Vallecas	Concertado	Mater Amabilis	19,12
Villa de Vallecas	Concertado	Nueva Castilla	23,56
Villa de Vallecas	Concertado	Sagrado Corazón	15,37
Villa de Vallecas	Concertado	San Eulogio	14,92
Villa de Vallecas	Concertado	Zazuar	22,83
Vicálvaro	Concertado	El Cid	17,67
Vicálvaro	Concertado	El Valle	27,85
Vicálvaro	Concertado	La Presentación De Ntra. Sra.	22,40
Vicálvaro	Concertado	Minerva	22,26
San Blas-Canillejas	Concertado	Amor de Dios	25,95
San Blas-Canillejas	Privado	Bra-Institución	26,78
San Blas-Canillejas	Concertado	Colegio Internacional J.H. Newman	20,68
San Blas-Canillejas	Concertado	Jesús María-Ntra. Sra. Caridad del Cobre	23,16
San Blas-Canillejas	Concertado	Juan de Valdés	23,65
San Blas-Canillejas	Concertado	Las Rosas	26,83
San Blas-Canillejas	Concertado	López Vicuña	15,68
San Blas-Canillejas	Concertado	Nazaret	19,75
San Blas-Canillejas	Concertado	Ntra. Sra. de Las Nieves	28,93
San Blas-Canillejas	Privado	Ntra. Sra. de Loreto	31,55
San Blas-Canillejas	Concertado	San Vicente	23,95
San Blas-Canillejas	Concertado	Santa Cruz	19,10
San Blas-Canillejas	Concertado	Santo Ángel de La Guarda	27,24
San Blas-Canillejas	Concertado	Timón	22,36
Barajas	Privado	Alameda de Osuna	25,72
Barajas	Concertado	Colegio de Jesús	22,76
Barajas	Concertado	San Pedro Apóstol	20,59

Fuente: Elaboración propia

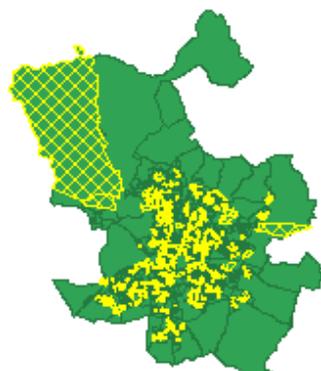
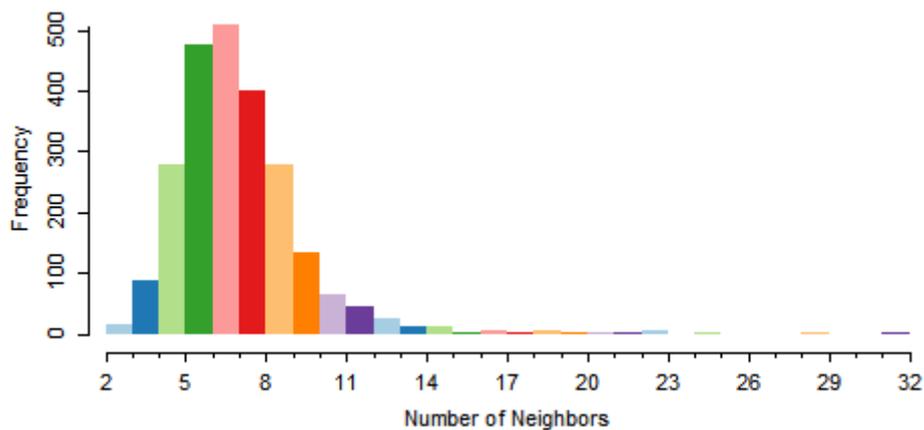
8.2. ANÁLISIS EXPLORATORIO ESPACIAL DE LAS VARIABLES DEL MODELO

8.2.1. Construcción de las matrices de vecindad

La estructura de vecindad espacial puede expresarse matemáticamente a través de una matriz denominada matriz de interacciones, pesos, ponderaciones, distancias o contactos espaciales. Se trata de una matriz cuadrada de orden n (tantos elementos como unidades espaciales tenga el sistema) en la que cada unidad se representa a la vez mediante una fila y una columna. En cada fila, los elementos no nulos de las columnas se corresponden con las unidades vecinas. A continuación se representan las especificaciones más comunes que hemos utilizado en este modelo.

- *Binaria de primer orden:*
 - QUEEN (REINA)

FIGURA 8.1: N° DE VECINOS MÁS CERCANOS POR EL MÉTODO DE LA REINA



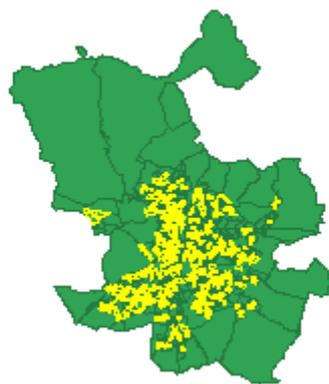
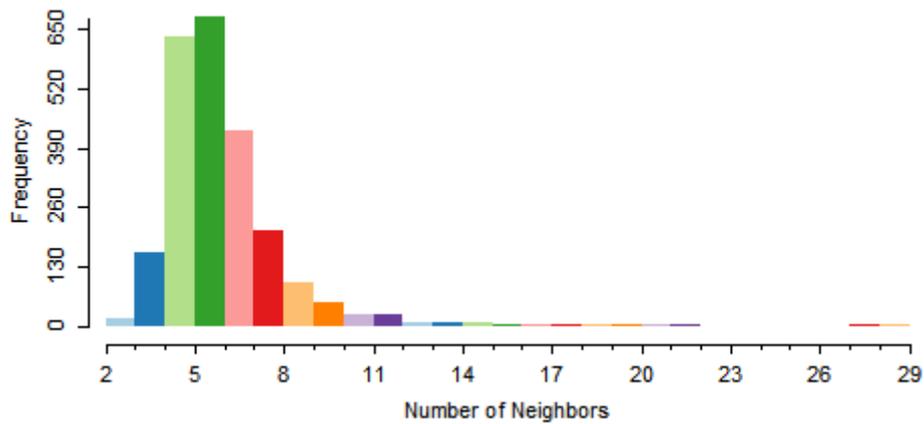
Fuente: Elaboración propia

Con esta matriz podemos observar como 510 secciones tienen 6 secciones vecinas, lo cual lo convierte en el número de secciones más común.

Dichas secciones se corresponden mayormente con la franja central de Madrid, a excepción de alguna sección en el noroeste.

- ROOK (TORRE)

FIGURA 8.2: Nº DE VECINOS MÁS CERCANOS POR EL MÉTODO DE LA TORRE

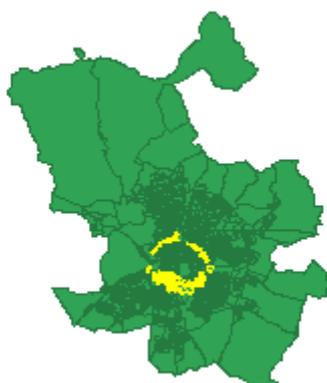
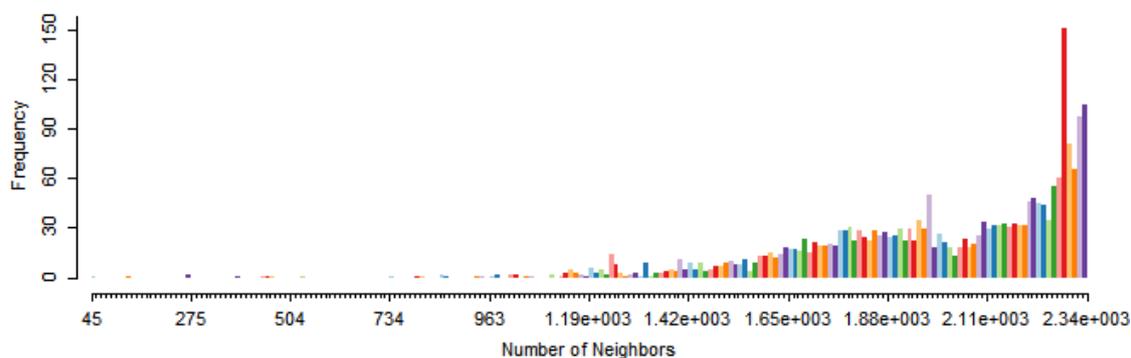


Fuente: Elaboración propia

En este caso podemos observar como el número de secciones vecinas más común son 5 (678 secciones). También vemos como en este caso de vecindad basada solamente en los bordes, excluyendo los vértices, solo nos quedamos con las secciones situadas en la franja central de Madrid.

➤ *Radio de distancias o distancia crítica:*

FIGURA 8.3: Nº DE VECINOS MÁS CERCANOS POR EL MÉTODO DEL RADIOS DE DISTANCIAS O DISTANCIA CRÍTICA



Fuente: Elaboración propia

Con esta matriz, el programa nos ha cogido los vecinos más próximos en un radio de 10,40 Km (coincide con la distancia máxima entre el centroide y el colegio más cercano). El número de secciones vecinas que más se repite dentro de este rango son entre 2283 y 2295. 151 secciones tienen este número de secciones por vecinas en esta matriz y, como podemos observar en el mapa, se corresponden a un círculo en el centro de Madrid.

➤ *Vecinos más cercanos:*

En este caso he realizado la matriz para los grupos de vecindad 5, 6 y 7, ya que son los que tenían el máximo número de secciones.

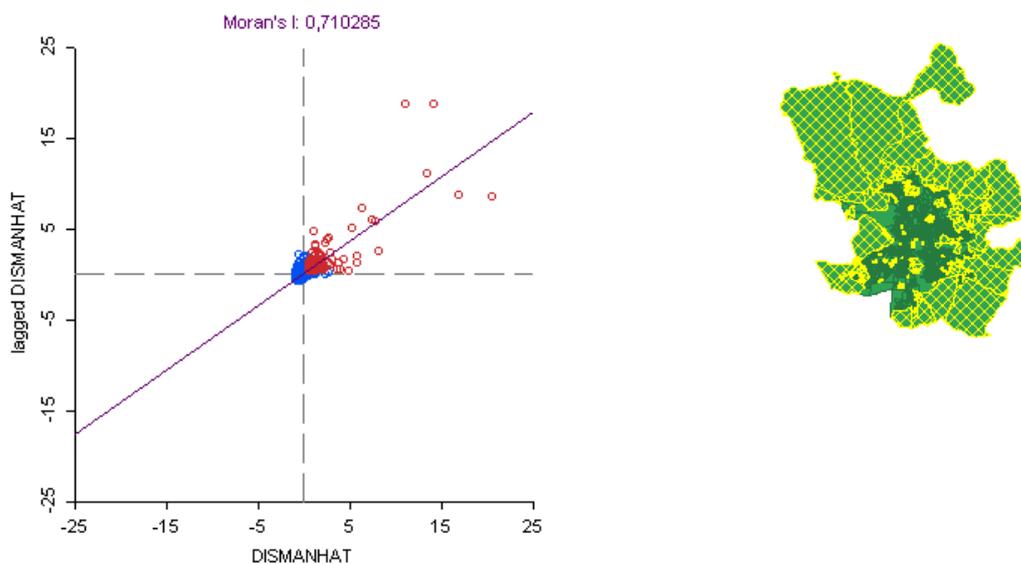
Nos hemos basado en la matriz binaria según el criterio de la Reina (Queen).

8.2.2. Análisis de autocorrelación espacial de las variables

La autocorrelación espacial, o más generalmente, dependencia espacial, es la situación en la que la variable dependiente y término de error para cada localización, se correlaciona con las observaciones de la variable dependiente o los valores del término de error en otras localizaciones.

➤ GLOBAL

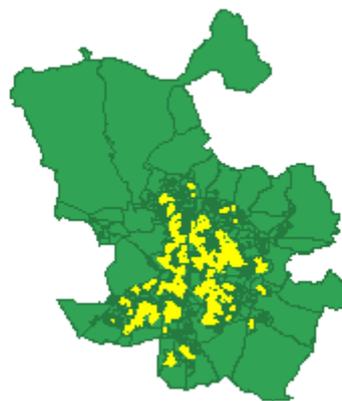
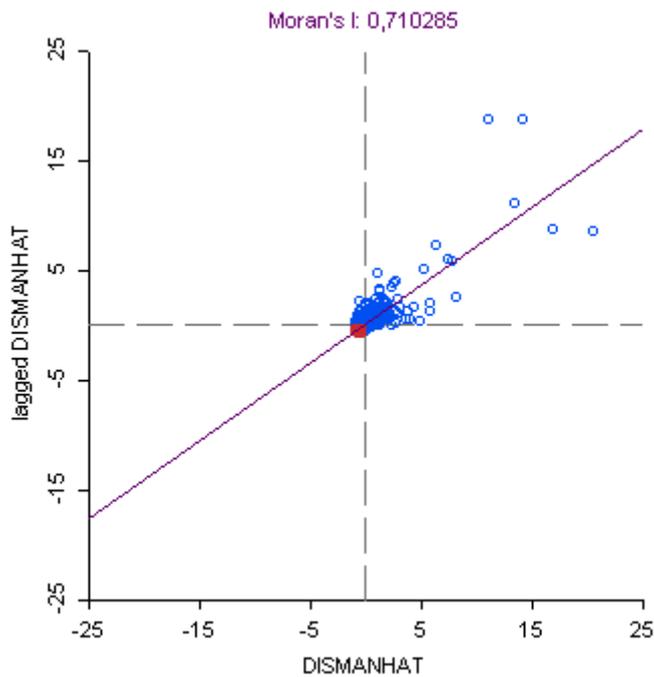
FIGURA 8.4: AUTOCORRELACIÓN ESPACIAL GLOBAL POSITIVA SEGÚN ESTADÍSTICO I DE MORAN CON DISTANCIAS MAYORES



Fuente: Elaboración propia

En este caso podemos ver las secciones que tienen una distancia elevada al colegio más cercano, rodeadas de secciones que tienen también una alta distancia al colegio más cercano. Estas secciones se corresponden, mayormente, con la periferia de la ciudad y secciones de gran dimensión geográfica.

FIGURA 8.5: AUTOCORRELACIÓN ESPACIAL GLOBAL POSITIVA SEGÚN ESTADÍSTICO I DE MORAN CON DISTANCIAS MENORES



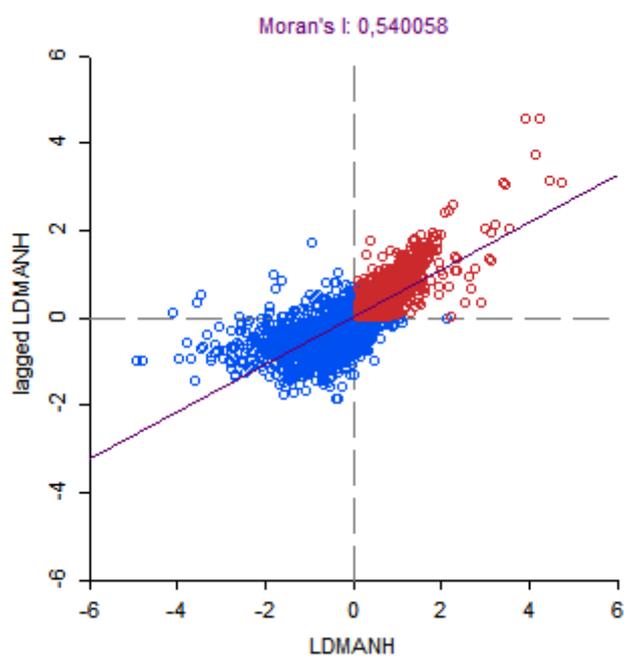
Fuente: Elaboración propia

En el lado opuesto, también con autocorrelación espacial positiva, observamos las secciones que están a una distancia pequeña del colegio más cercano, rodeadas de otras secciones que están también a una distancia reducida del colegio más cercano. En este caso nos referimos a secciones localizadas, en su gran mayoría, en el centro de Madrid.

Las secciones que se encuentra con una autocorrelación espacial negativa son muy reducidas y están situadas en la parte más exterior del centro de Madrid.

Con logaritmos:

FIGURA 8.6: AUTOCORRELACIÓN ESPACIAL GLOBAL POSITIVA SEGÚN ESTADÍSTICO I DE MORAN CON DISTANCIAS MAYORES

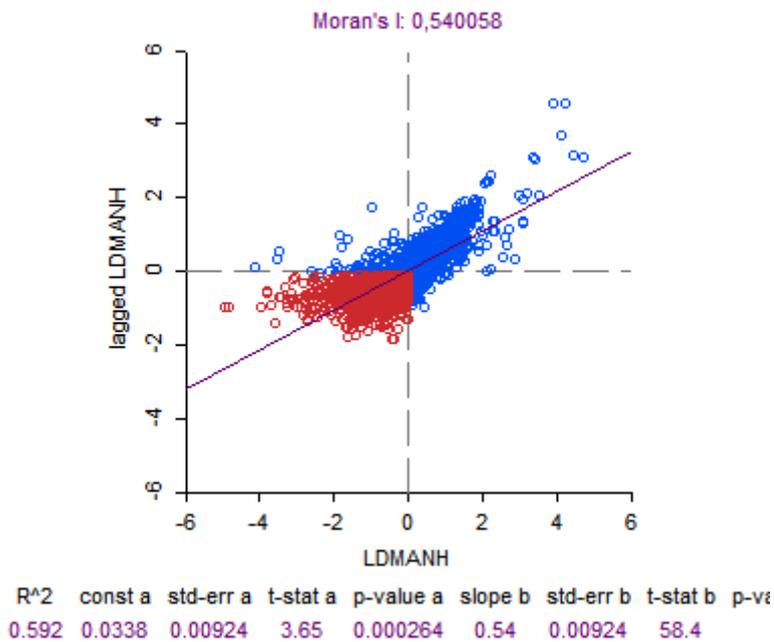


Fuente: Elaboración propia

En este caso podemos ver las secciones que tienen una distancia elevada al colegio más cercano, rodeadas de secciones que tienen también una alta distancia al colegio más cercano.

Se observa como abarca muchas más secciones que las periféricas de la ciudad, llegando ya a muchas del centro de la ciudad de Madrid.

FIGURA 8.7: AUTOCORRELACIÓN ESPACIAL GLOBAL POSITIVA SEGÚN ESTADÍSTICO I DE MORAN CON DISTANCIAS MENORES



Fuente: Elaboración propia

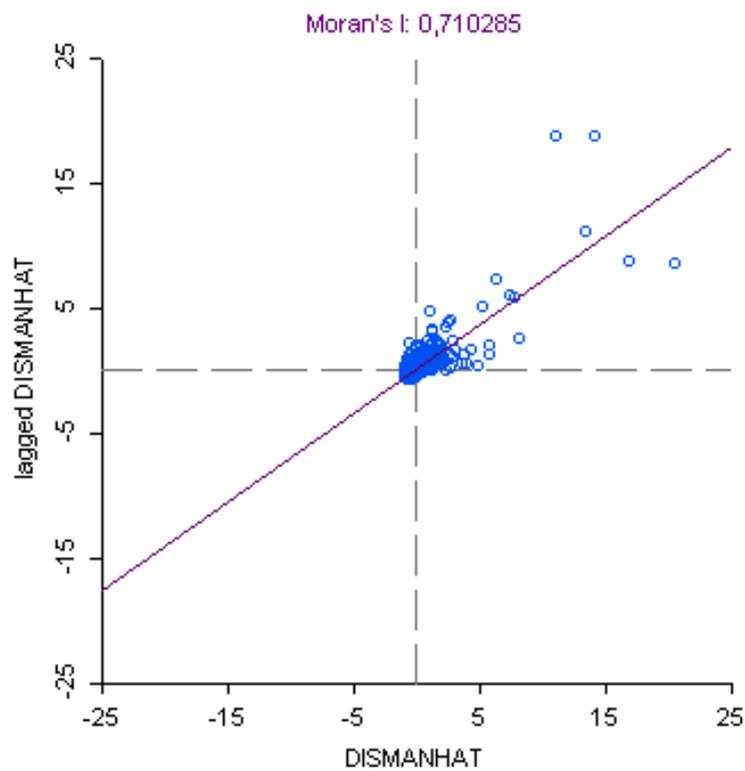
En el lado opuesto, también con autocorrelación espacial positiva, observamos las secciones que están a una distancia pequeña del colegio más cercano, rodeadas de otras secciones que

están también a una distancia reducida del colegio más cercano. En este caso nos referimos a secciones localizadas, en su gran mayoría, en el centro de Madrid. Muy similar al análisis anterior

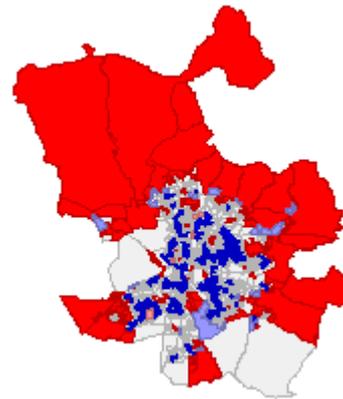
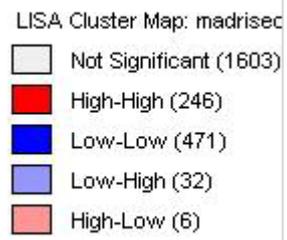
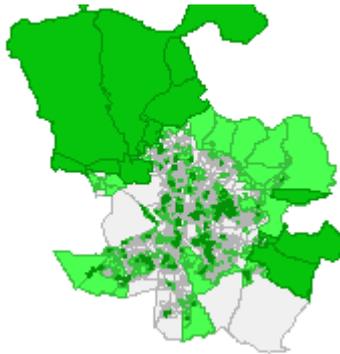
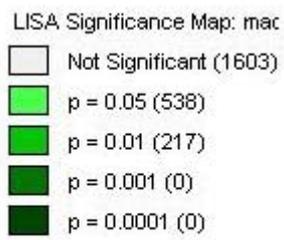
Las secciones que se encuentra con una autocorrelación espacial negativa son más abundantes que anteriormente y se sitúan también en la parte central de la ciudad, ademásde en la parte más exterior del centro de Madrid.

➤ LOCAL

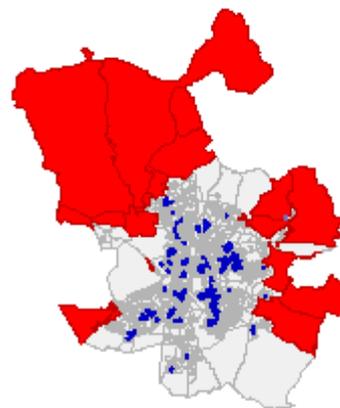
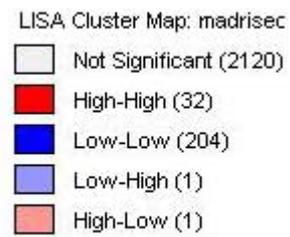
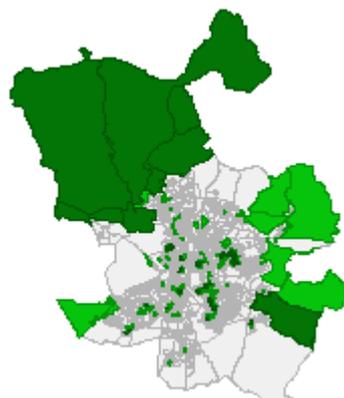
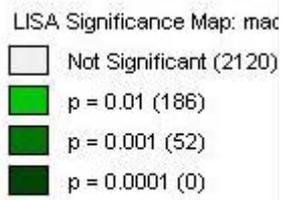
FIGURA 8.8: AUTOCORRELACIÓN ESPACIAL LOCAL POSITIVA SEGÚN ESTADÍSTICO I DE MORAN CON DIFERENTE NIVEL DE PERMUTACIONES Y SIGNIFICACIÓN



99 permutaciones / 0,05



999 permutaciones / 0,01

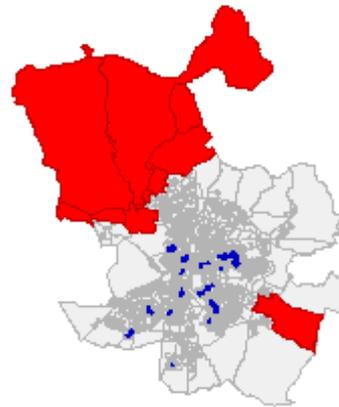


9999 permutaciones / 0,001

LISA Significance Map: mac
Not Significant (2289)
 $p = 0.001$ (50)
 $p = 0.0001$ (19)

LISA Cluster Map: madrisec

Not Significant (2289)
High-High (12)
Low-Low (57)
Low-High (0)
High-Low (0)

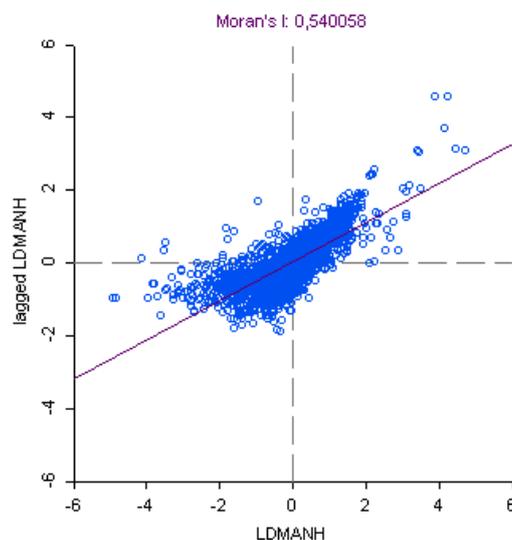


Fuente: Elaboración propia

He ido modificando tanto las permutaciones como los índices de significación y vemos como se posicionan 2 extremos claros, uno en el norte y sur con clusters de secciones con distancias más altas de lo normal con respecto a los colegios más cercanos. Por otro lado, tenemos otro cluster en la parte central de la ciudad, los cuales tienen unas distancias mínimas al colegio más cercano.

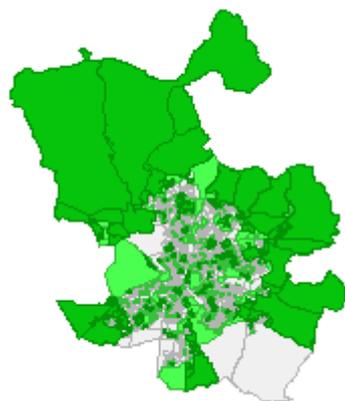
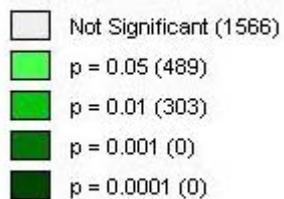
Con logaritmos:

FIGURA 8.9: AUTOCORRELACIÓN ESPACIAL LOCAL POSITIVA SEGÚN ESTADÍSTICO I DE MORAN CON DIFERENTE NIVEL DE PERMUTACIONES Y SIGNIFICACIÓN

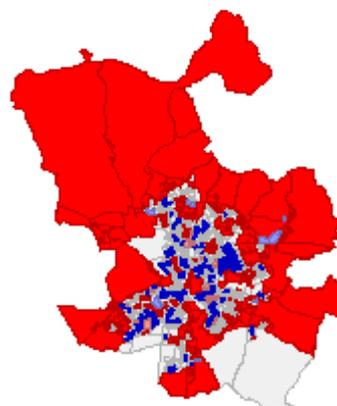
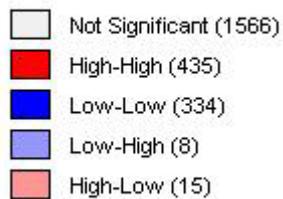


99 permutaciones / 0,05

LISA Significance Map: mac

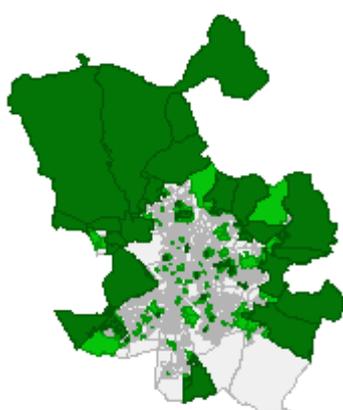
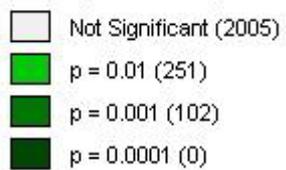


LISA Cluster Map: madrisecc

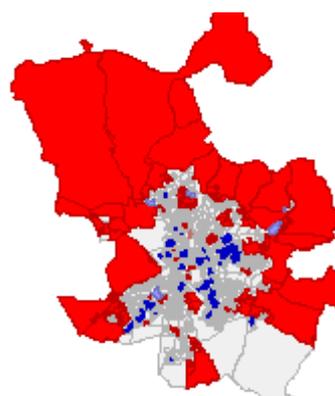
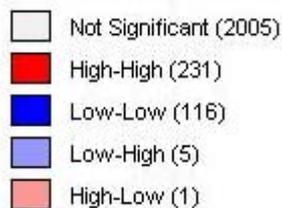


999 permutaciones / 0,01

LISA Significance Map: mac



LISA Cluster Map: madrisecc



9999 permutaciones / 0,001

LISA Significance Map: mac

Not Significant (2253)

$p = 0.001$ (70)

$p = 0.0001$ (35)

LISA Cluster Map: madrisect

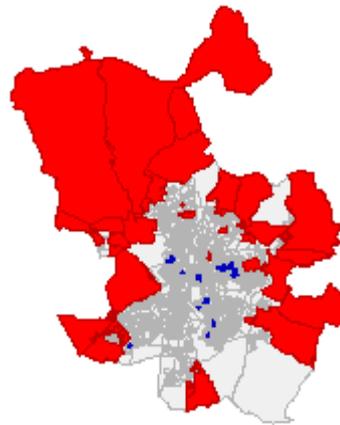
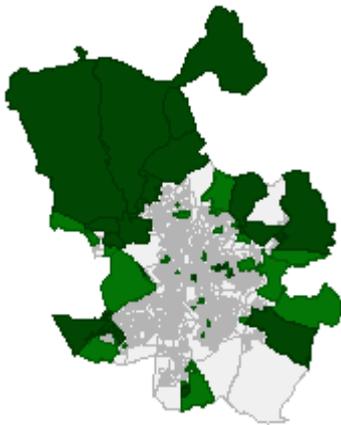
Not Significant (2253)

High-High (80)

Low-Low (25)

Low-High (0)

High-Low (0)



Fuente: Elaboración propia

En este otro análisis con logaritmos, observamos como los clusters de secciones con distancias elevadas al colegio más cercano se extiende de 12 a 80 secciones y abarca también zonas este y oeste de la ciudad. Asimismo, el cluster de secciones situado en el centro de la ciudad se ha reducido de 57 a 25 secciones.

– Valores atípicos

FIGURA 8.10: MAPA DE CAJA DE VALORES ATÍPICOS DE LA VARIABLE DISTANCIAS

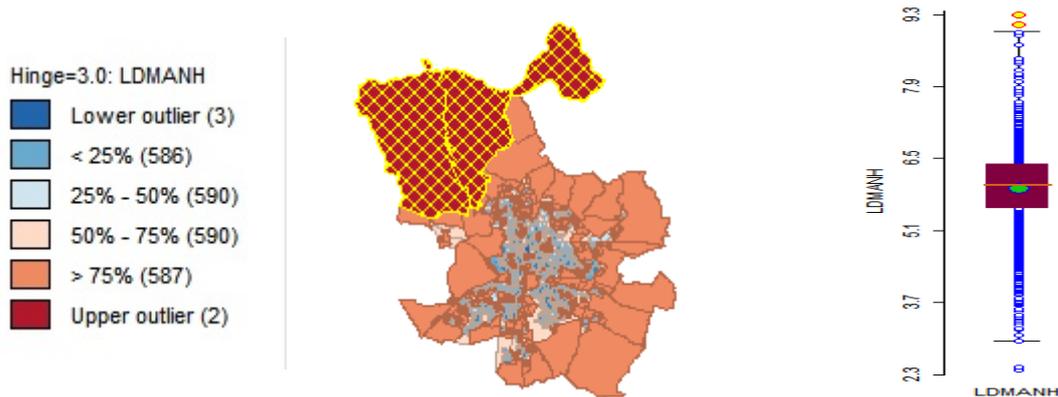


Fuente: Elaboración propia

Podemos observar como existen varios puntos atípicos, pero todos por encima de lo normal. Vemos como casi todas las secciones periféricas de la ciudad de Madrid se encuentran a una distancia atípica muy superior a la media del colegio más cercano.

Con logaritmos:

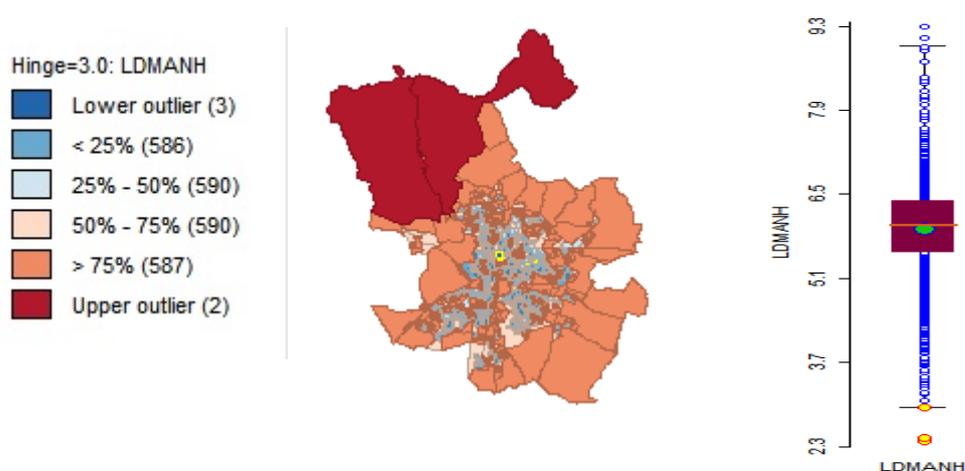
FIGURA 8.11: MAPA DE CAJA DE VALORES ATÍPICOS POR ENCIMA DE LA MEDIA DE LA VARIABLE DISTANCIAS CON LOGARITMOS



Fuente: Elaboración propia

En este caso se ha reducido el número de secciones atípicas que están por encima de la distancia media del colegio más cercano, en solo 2 secciones, las cuales se sitúan en la parte norte de la ciudad de Madrid.

FIGURA 8.12: MAPA DE CAJA DE VALORES ATÍPICOS POR DEBAJO DE LA MEDIA DE LA VARIABLE DISTANCIAS CON LOGARITMOS



Fuente: Elaboración propia

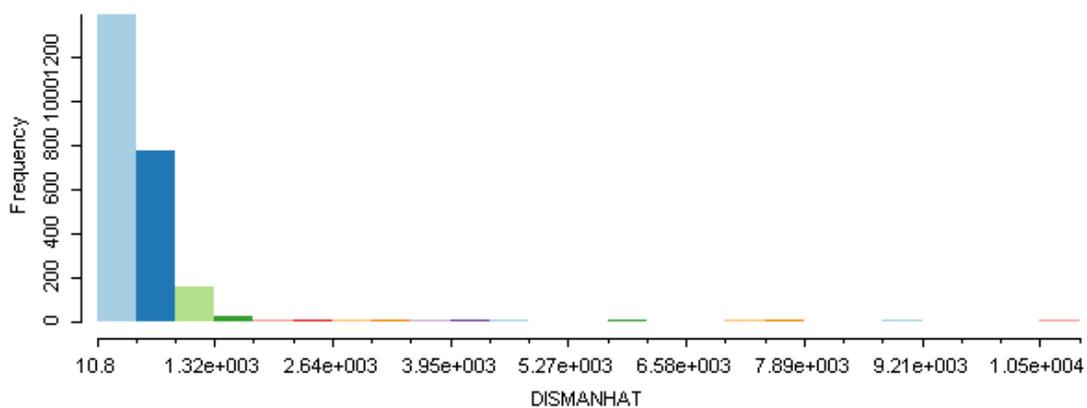
Ahora sí que identificamos valores atípicos en con distancias inferiores a la media respecto al colegio más próximo. Son 3 secciones situadas en el centro de la ciudad de Madrid.

8.2.3. Análisis de causalidad entre la variable dependiente y las independientes

En este apartado del anexo nos centraremos, primero, en el la comprobación de si nuestra variable sigue una distribución normal, como debería de ser según nos dice Anselin, y también la relación de causalidad que existe entre la variable dependiente y cada una de las variables independientes, en qué dirección es dicha relación y con qué intensidad.

➤ *Normalidad/No Normalidad*

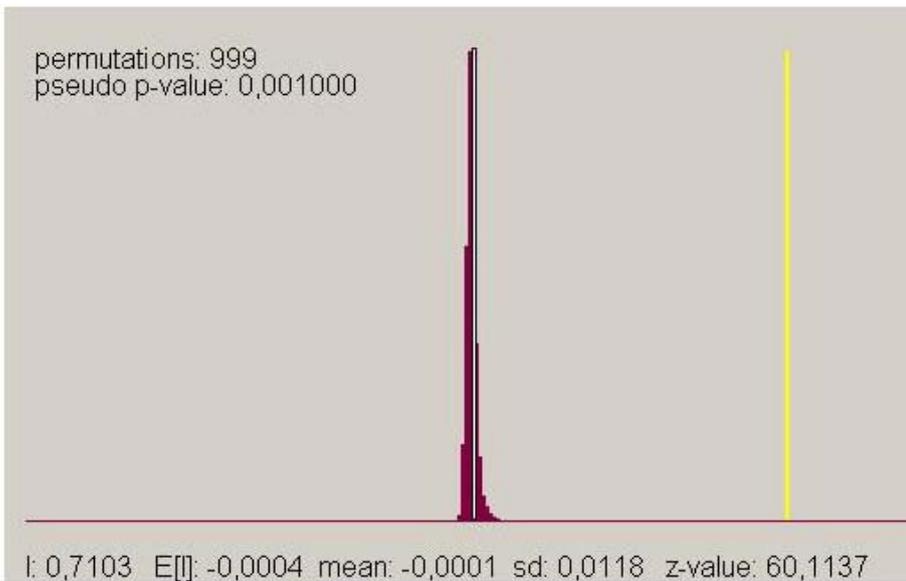
FIGURA 8.13: HISTOGRAMA DE FRECUENCIAS DE LA VARIABLE DISTANCIAS



Fuente: Elaboración propia

Vemos como la variable distancias no sigue una distribución normal. Por ello utilizamos el método de las permutaciones, el cual de forma empírica (con datos simulados) la distribución concreta que debe considerarse como referencia si la variable se distribuye como una normal, el método de permutaciones construye, de forma empírica, la distribución más adecuada. Por lo tanto, si la variable no sigue la normal, el método de permutaciones construirá la distribución más adecuada para el contraste.

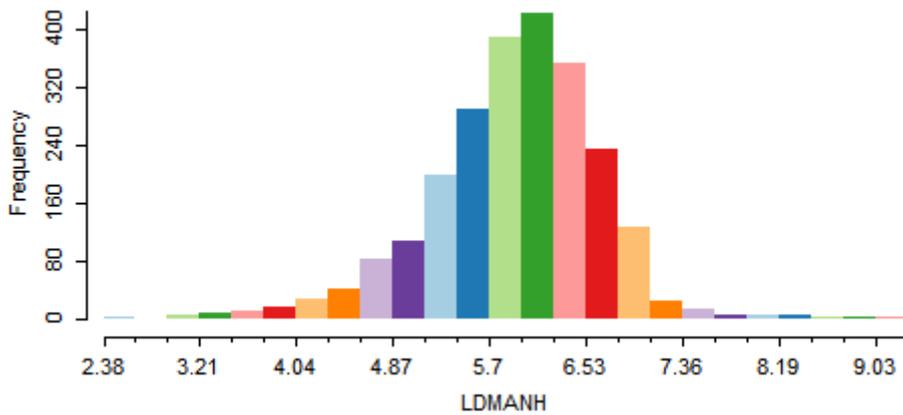
FIGURA 8.14: REPRESENTACIÓN DE LA NORMAL MEDIANTE EL MÉTODO DE LAS PERMUTACIONES



Fuente: Elaboración propia

Con logaritmos:

FIGURA 8.15: HISTOGRAMA DE FRECUENCIAS DE LA VARIABLE DISTANCIAS CON LOGARITMOS

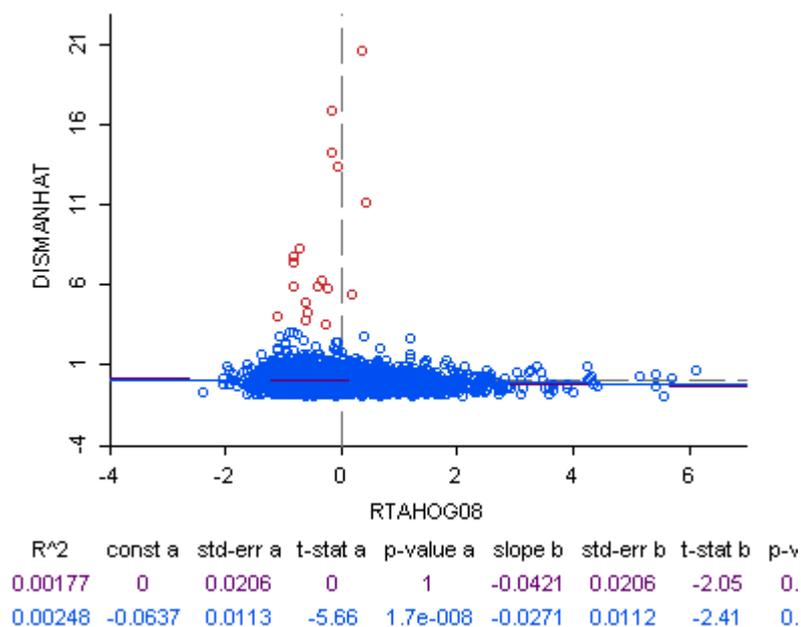


Fuente: Elaboración propia

Al contrario que antes, en esta ocasión, sí que sigue una distribución normal.

- *Relación dos a dos*
 - DISMANHAT vs RTAHOG08

FIGURA 8.16: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y RENTA FAMILIAR



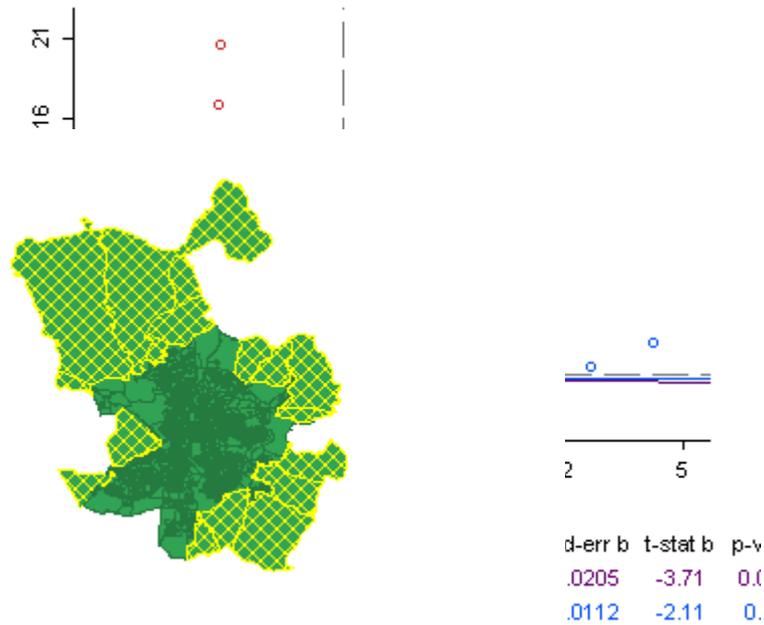
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta la renta familiar disponible por habitante, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– DISMANHAT vs NOISE

FIGURA 8.17: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE RUIDO



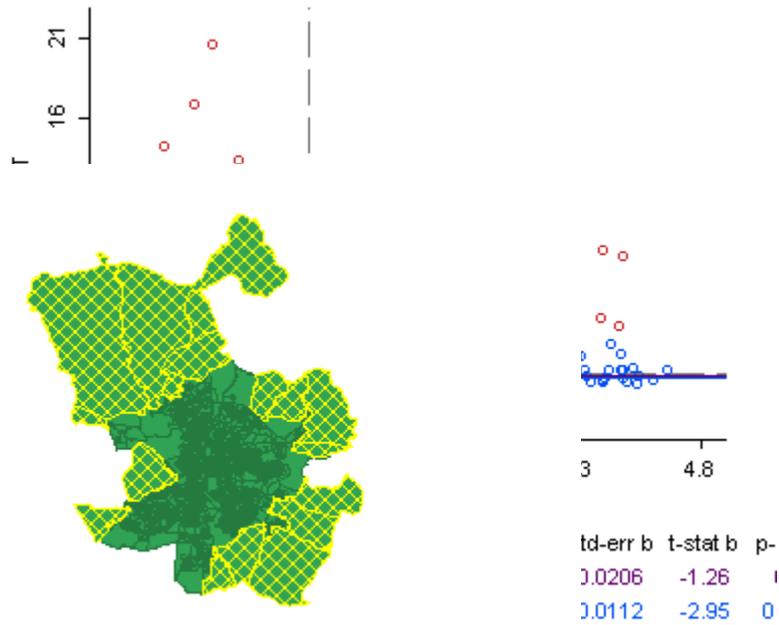
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de ruido en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero, aun así, sigue siendo negativa.

– DISMANHAT vs CONT

FIGURA 8.18: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE CONTAMINACIÓN



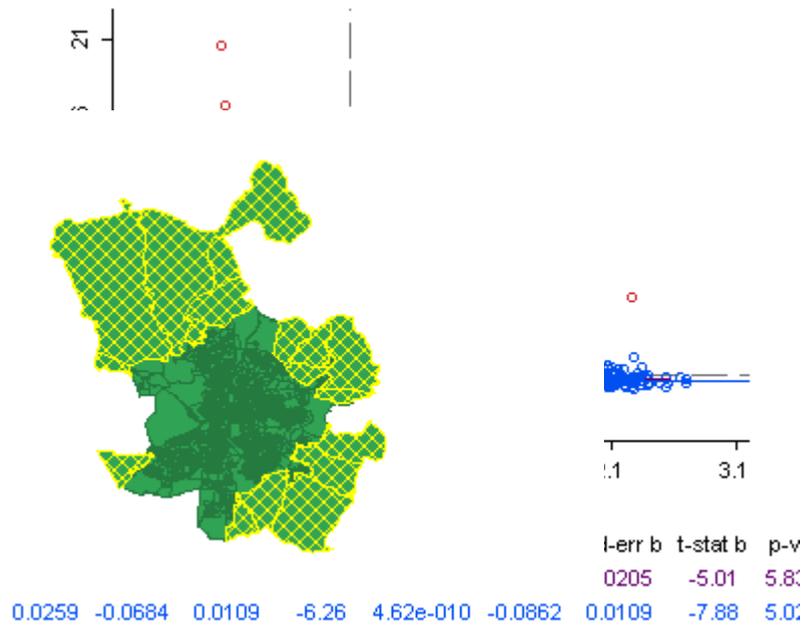
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de contaminación atmosférica en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y, en este caso, se observa como se intensifica ligeramente esta relación indirecta.

– DISMANHAT vs NGREE

FIGURA 8.19: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE AUSENCIA DE ZONAS VERDES



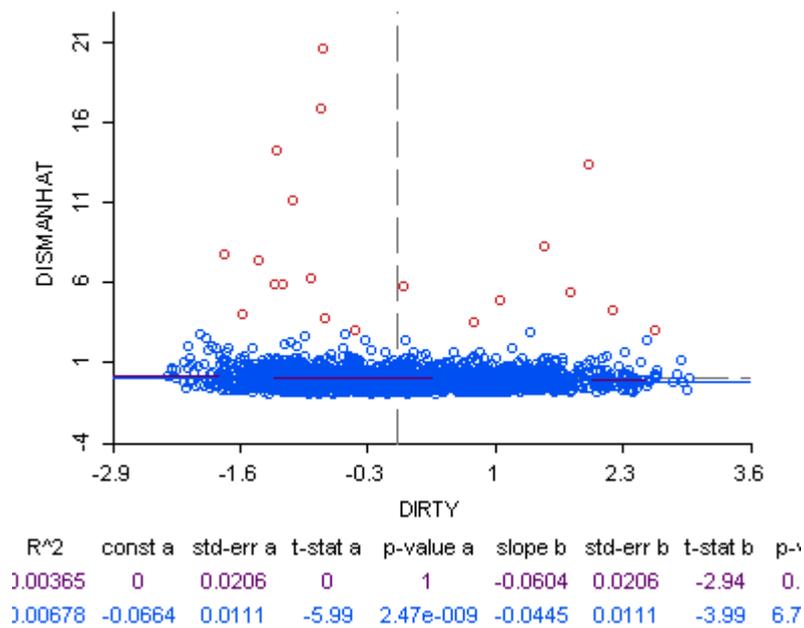
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de ausencia de zonas verdes en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– DISMANHAT vs DIRTY

FIGURA 8.20: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE SUCIEDAD



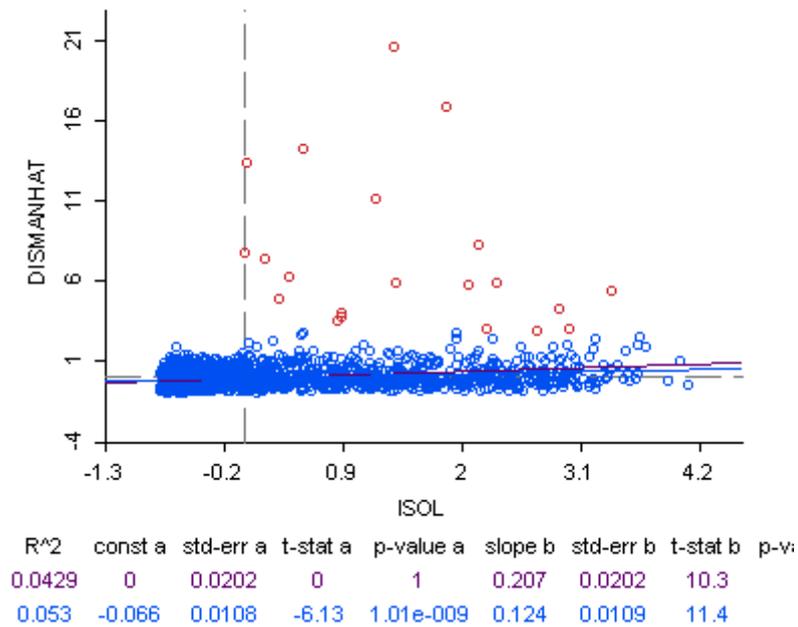
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de suciedad en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– DISMANHAT vs ISOL

FIGURA 8.21: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS POR MALAS COMUNICACIONES



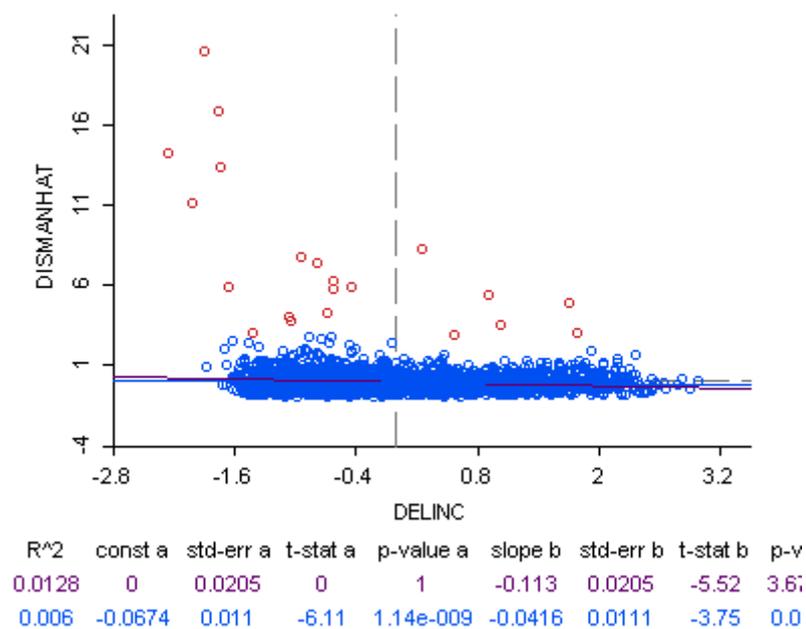
Fuente: Elaboración propia

En este caso, la relación existente es directa o positiva, cuanto más aumenta los hogares que dice tener problemas por malas comunicaciones en su lugar de residencia, aumenta, también, la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo positiva.

– DISMANHAT vs DELINC

FIGURA 8.22: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE DELINCUENCIA



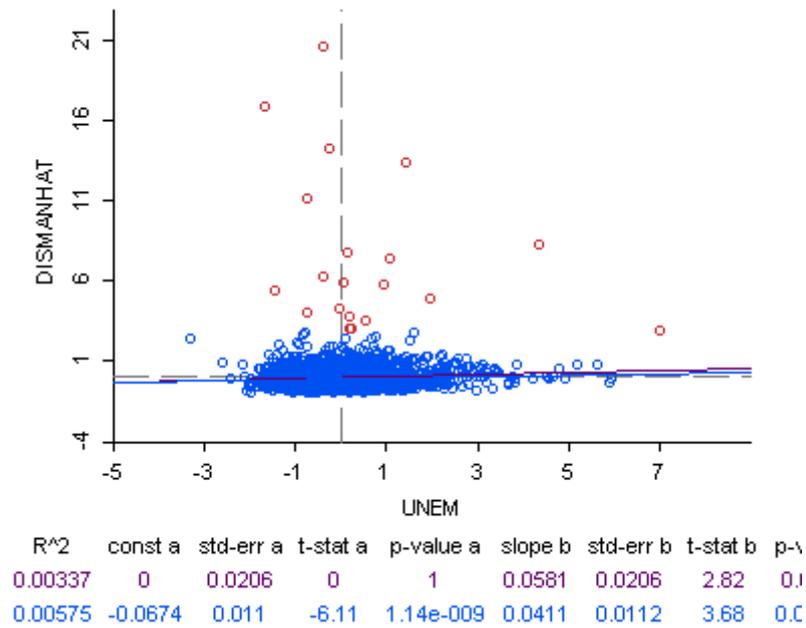
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de delincuencia en su lugar de residencia, disminuye la distancia al colegio más

próximo. He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– DISMANHAT vs UNEM

FIGURA 8.23: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y TASA DE PARO



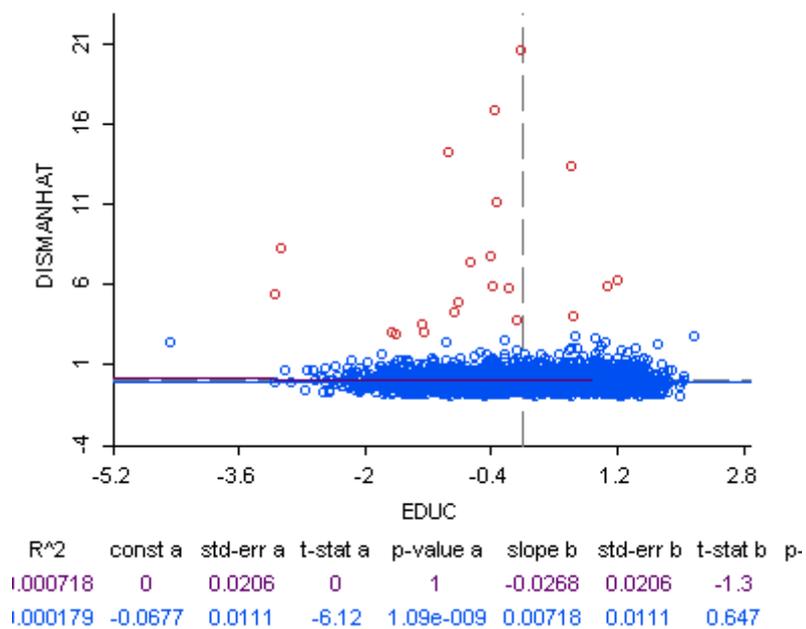
Fuente: Elaboración propia

En este caso, la relación existente es directa o positiva, cuanto más aumenta la tasa de paro, aumenta, también, la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo positiva.

– DISMANHAT vs EDUC

FIGURA 8.24: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y NIVEL EDUCATIVO



Fuente: Elaboración propia

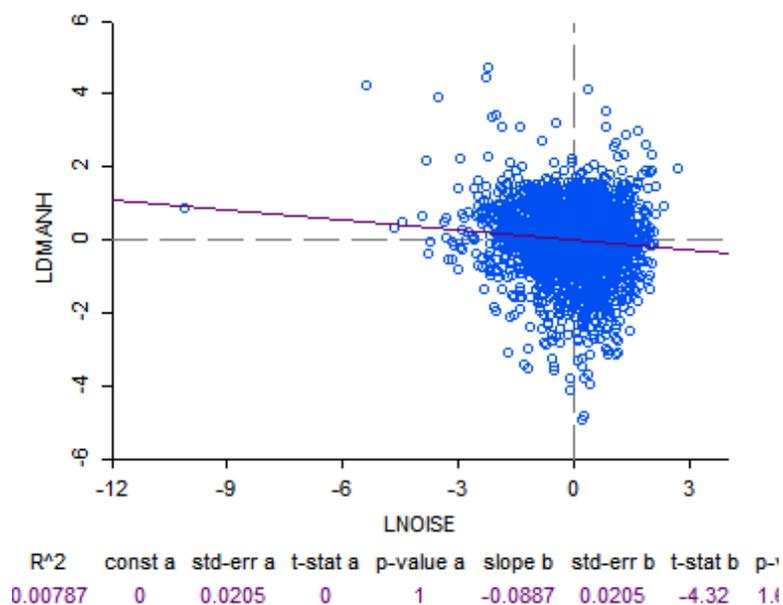
La relación que existe es indirecta o negativa, cuanto más aumenta las de personas con título de secundaria (bachiller y FP) y universitario sobre total población, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

Con logaritmos:

– LDMANH vs LRRTA08

FIGURA 8.25: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y RENTA FAMILIAR CON LOGARITMOS



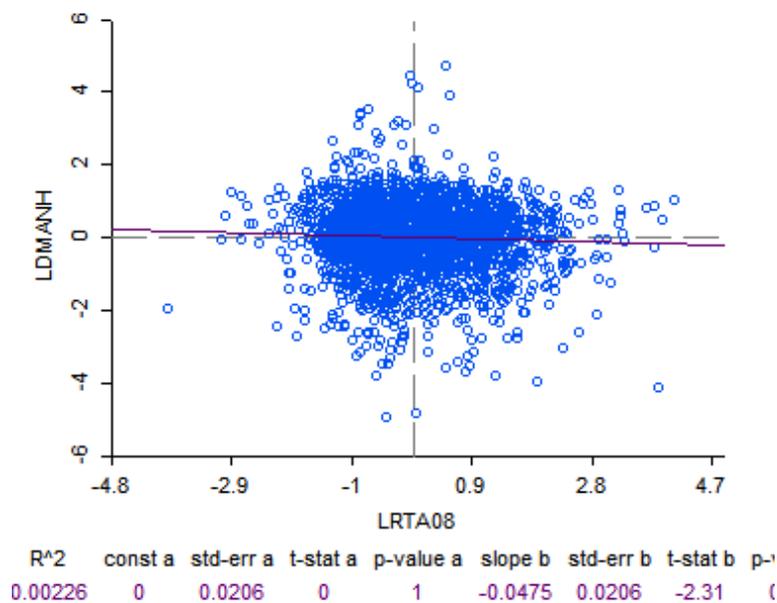
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta la renta familiar disponible por habitante, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– LDMANH vs LNOISE

FIGURA 8.26: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE RUIDO CON LOGARITMOS



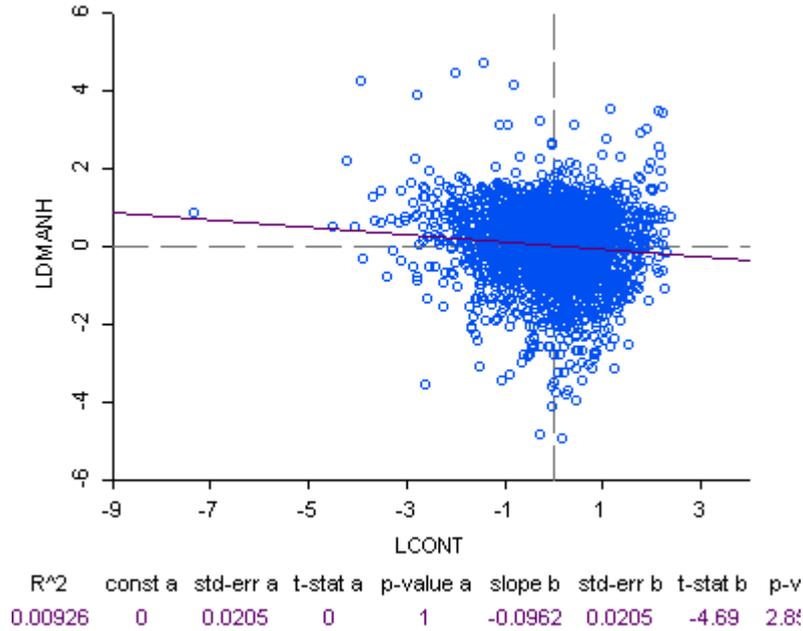
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de ruido en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– LDMANH vs LCONT

FIGURA 8.27: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE CONTAMINACIÓN CON LOGARITMOS



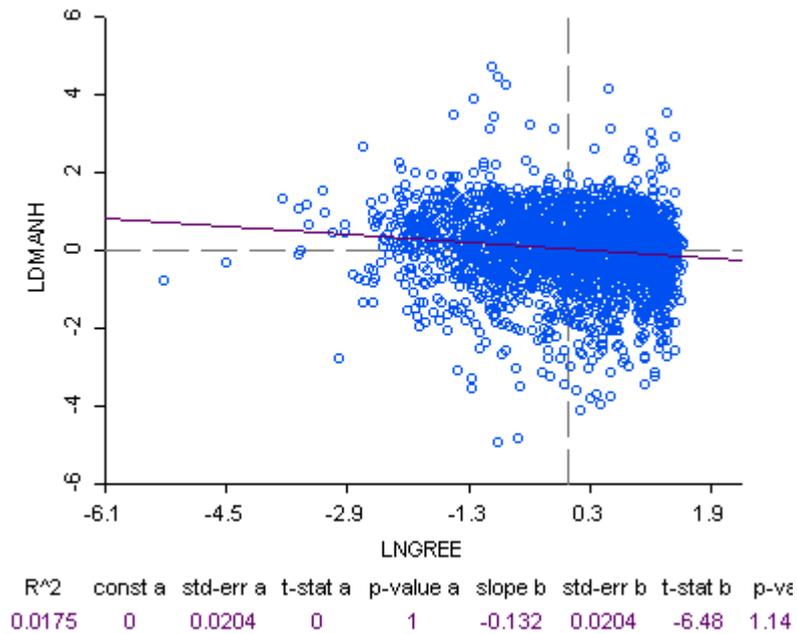
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de contaminación atmosférica en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– LDMANH vs LNGREE

FIGURA 8.28: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE AUSENCIA DE ZONAS VERDES CON LOGARITMOS



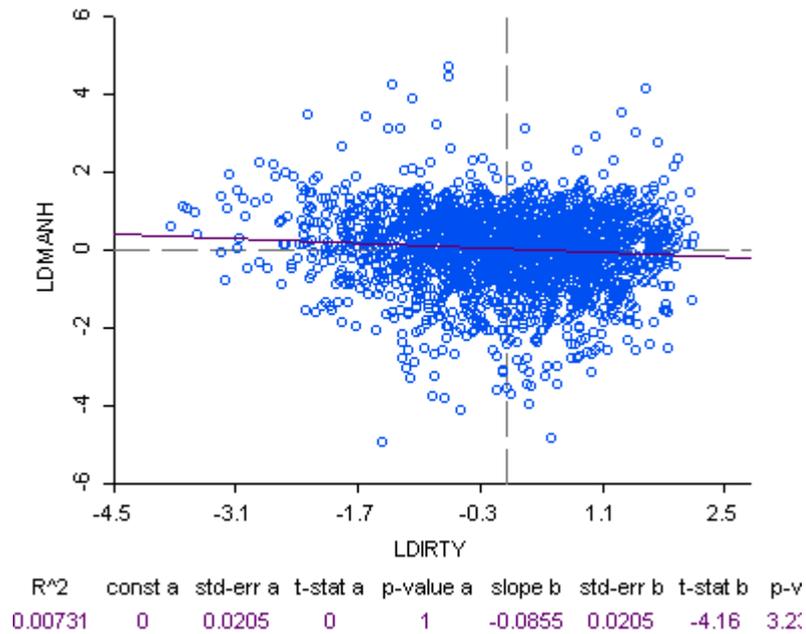
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de ausencia de zonas verdes en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– LDMANH vs LDIRTY

FIGURA 8.29: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE SUCIEDAD CON LOGARITMOS



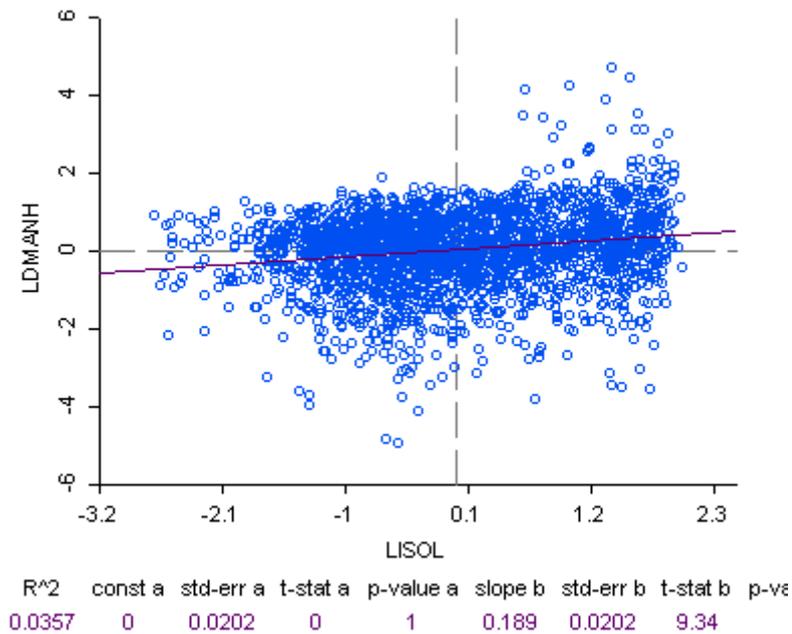
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de suciedad en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– LDMANH vs LISOL

FIGURA 8.30: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS POR MALAS COMUNICACIONES CON LOGARITMOS



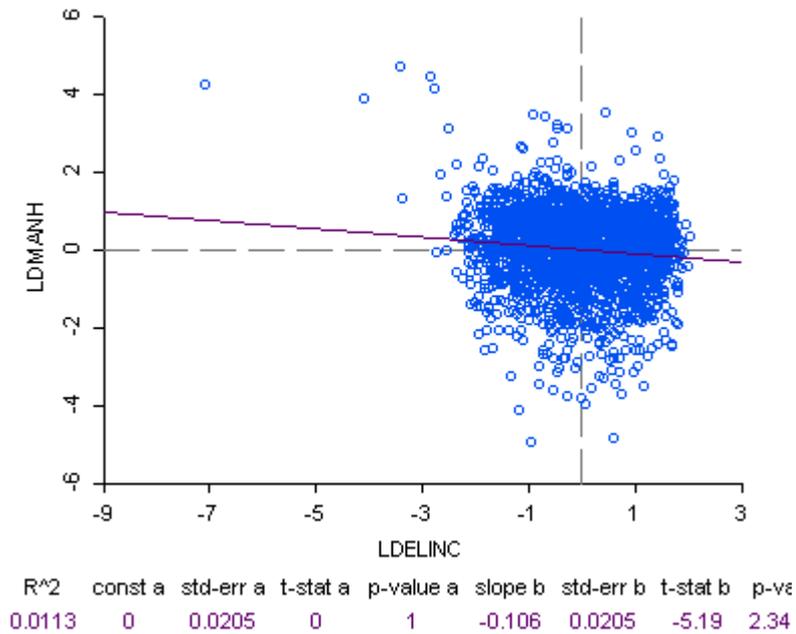
Fuente: Elaboración propia

En este caso, la relación existente es directa o positiva, cuanto más aumenta los hogares que dice tener problemas por malas comunicaciones en su lugar de residencia, aumenta, también, la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo positiva.

– LDMANH vs LDELINC

FIGURA 8.31: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y PROBLEMAS DE DELINCUENCIA



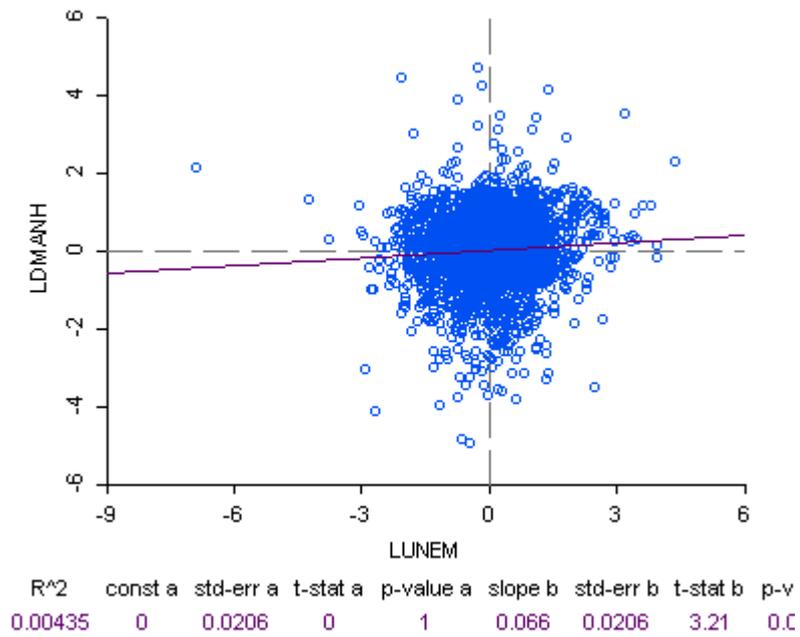
Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta los hogares que dice tener problemas de delincuencia en su lugar de residencia, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

– LDMANH vs LUNEM

FIGURA 8.32: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y TASA DE PARO CON LOGARITMOS



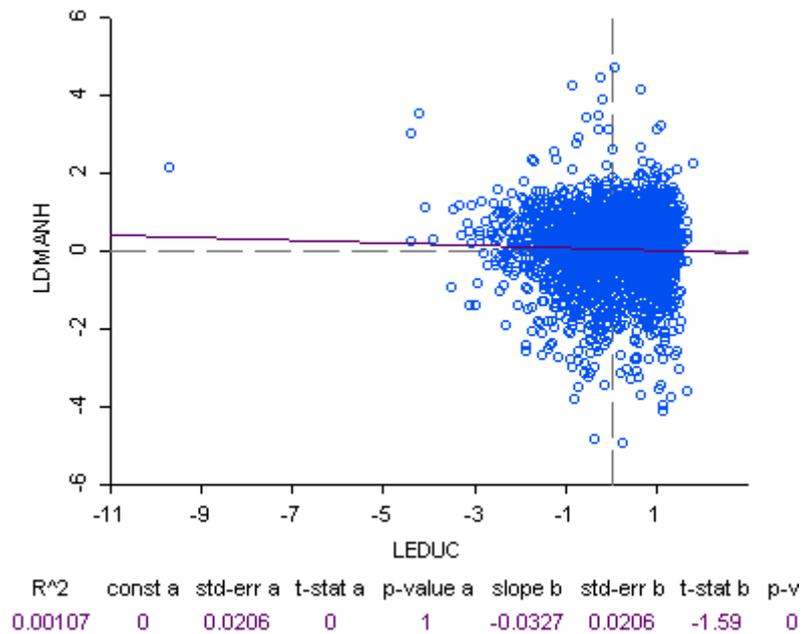
Fuente: Elaboración propia

En este caso, la relación existente es directa o positiva, cuanto más aumenta la tasa de paro, aumenta, también, la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo positiva.

– LDMANH vs LEDUC

FIGURA 8.33: DIAGRAMA DE DISPERSIÓN VARIABLE DISTANCIAS Y NIVEL EDUCATIVO CON LOGARITMOS



Fuente: Elaboración propia

La relación que existe es indirecta o negativa, cuanto más aumenta las de personas con título de secundaria (bachiller y FP) y universitario sobre total población, disminuye la distancia al colegio más próximo.

He excluido los puntos atípicos (indicados en el mapa), y se observa cómo se reduce esta relación, pero aun así sigue siendo negativa.

CONCLUSIONES DEL ANALISIS EXPLORATORIO ESPACIAL

- *Puntos atípicos*

Se reduce considerablemente el número de secciones atípicas analizando las variables logarítmicas que están por encima de la distancia media del colegio más cercano, de 29 secciones abarcando toda la periferia, a solo 2 secciones, las cuales se sitúan en la parte norte de la ciudad de Madrid.

En la variable logarítmica identificamos valores atípicos en con distancias inferiores a la media respecto al colegio más próximo. Son 3 secciones situadas en el centro de la ciudad de Madrid. Cosa que en la variable normal no se detectaban.

- *Problemas de no normalidad*

En este punto veo claramente como la variable normal no sigue una distribución normal en ningún caso y la variable logarítmica sigue una variable normal perfecta. Este aparatado no lo he entendido muy bien en el tema de las permutaciones...

- *Correlación "dos a dos" (bivariante)*

En este apartado no encuentro gran diferencia entre la fuerza de la relación de las variables normales y logarítmicas, ya que es muy similar la pendiente en ambos casos y la relación directa e indirecta coincide en todos los casos. Lo que si veo es que en la variable logarítmica las secciones están más expandidas que en la normal.

- *Autocorrelación espacial global y local*

En el análisis con logaritmos, observamos como los clusters de secciones con distancias elevadas al colegio más cercano se extienden de 12 a 80 secciones y abarca también zonas este y oeste de la ciudad, mientras que el análisis con la variable normal solo se extendía al norte y sur de Madrid.

Así mismo, el cluster de secciones situado en el centro de la ciudad se reducido de 57 a 25 secciones con el análisis de variables logarítmicas.