

Sistemas de Información Geográfica y Econometría Espacial en la Tasación de Inmuebles Urbanos. Proyecto Piloto en la Ciudad de Cartagena.

Fernando A. López Hernández

Departamento de Métodos Cuantitativos e Informáticos

Universidad Politécnica de Cartagena

e-mail: fernando.lopez@upct.es

Resumen

La tasación de un bien inmobiliario está siempre sujeta a un elevado número de factores que son difíciles de controlar de forma simultánea por el tasador. Con el fin de abordar este problema, el objetivo básico que se persigue en esta comunicación, es proporcionar al tasador de una herramienta informática que, mediante la conjunción de técnicas de gráficas (Sistemas de Información Geográfica) y estadísticas (Econometría Espacial), permitan asignar un valor de referencia a la tasación de bienes inmobiliarios. La utilización simultánea de estas técnicas se aplica a una muestra de viviendas tasadas en la ciudad de Cartagena.

Palabras clave: Sistemas de Información Geográfica, Econometría Espacial, Tasación.

1. Introducción.

Uno de los problemas más importantes a los que se enfrenta las empresas del sector dedicado a la tasación de bienes inmobiliarios es el que se refiere a la valoración de un inmueble. Este problema es especialmente complejo debido principalmente a la gran cantidad de factores que inciden, de una u otra forma, en la determinación del precio final en el que se realiza una transacción. Además, el mercado inmobiliario tiene características económicas muy particulares que lo distinguen y diferencian de otros mercados, en él inciden no sólo los factores propios del mercado, sino también las políticas económicas y sociales que adoptan los gobiernos, produciendo fluctuaciones de notable variación.

Por otra parte, los compradores y vendedores de inmuebles, en general pueden no estar informados sobre los valores y tendencias del mercado. Cada bien inmueble es único, y está ligado a una única ubicación, que lo distingue de los demás que lo rodean

y la transferencia de los mismos requiere conocimientos legales y técnicos que no siempre posee el comprador. Los inmuebles no se adquieren con frecuencia y casi siempre representan la mayor inversión individual que una persona hace, por lo que, una vez adquirido, tienden a tener una larga vida física y económica.

Esta situación está experimentando en los últimos años fuertes cambios, debido en primer lugar a los altos precios de los bienes inmuebles de tal forma que a la hora de asignar un valor a una vivienda cada vez se exigen mayor precisión ya que de ello puede depender decisiones relevantes. Por otra parte el elevado número de transacciones ha obligado a profundizar en los sistemas de valoración a la vez que se ha creado una demanda de formación por parte de los consumidores que no permiten que las necesidades de tasación sigan siendo atendidas por profesionales para tal propósito, pues se encuentran en juego grandes cantidades de dinero como para que se tome a la ligera una actividad tan importante. Esta demanda en formación tenido la correspondiente respuesta mediante la oferta de un amplio abanico de cursos de postgrado y master en valoración inmobiliaria.

El profesional de la tasación, no sólo debe conocer procedimientos matemáticos, sino que ha de estar al tanto de las teorías económicas y financieras, de los derechos que se ejercen sobre la propiedad. Así la estadística y las técnicas de computación (más asequible cada día con ordenadores personales y herramientas como las hojas de cálculo) pueden ser instrumentos de inestimable valor para sus análisis.

Las técnicas de valoración de inmuebles están siendo objeto de un profundo análisis y revisión en los últimos años por parte de los profesionales tanto los que trabajan en los mercados hipotecarios como los que prestan servicio en el catastro.

La determinación del precio de compra-venta al que finalmente se transfiera una propiedad es fruto en la mayoría de situaciones de la valoración subjetiva que realiza el vendedor, que siendo un buen conocedor del mercado y en base a su experiencia, determina de forma más o menos acertada el precio de venta de un bien urbano.

En la mayor parte de los casos, el precio de compra venta de un inmueble es el resultado de la convergencia de dos factores, por un lado, el **valor** del mismo, entendiendo por ello un **valor objetivo**, fruto de las características físicas del mismo y de su localización espacial en el contexto urbano en el que está inmerso, por otro lado, el **precio** del inmueble, como una **medida subjetiva**, fruto de apreciaciones individuales y sujeta a la factores externos de difícil determinación (utilidad, necesidad de inversión, necesidad de venta, etc).

Con el fin de reducir la subjetividad de la valoración inmobiliaria en la determinación del precio de los bienes urbanos y en particular del precio de la vivienda, en el presente artículo vamos a presentar la forma en que pueden utilizarse los métodos estadísticos e informáticos para ayudar a fijar el precio de una vivienda.

Así, el objetivo básico que se persigue en este artículo, es presentar algunos indicios de cómo es posible proporcionar a los tasadores una herramienta informática que, mediante la conjunción de técnicas gráficas (Sistemas de Información Geográfica) y Estadísticas (Econometría Espacial) permitan asignar un valor de referencia a la tasación de bienes inmobiliarios.

Los sistemas de información geográfica están cada vez mas implantados en la Sociedad de la Información y han alcanzado niveles de desarrollo que permiten una fácil aplicación a diversos campos. La herramienta gráfica que se apunta en este artículo ayudará al tasador a la visualización de toda la información de forma simultánea en la pantalla de su ordenador. Poder observar de forma global en un plano georeferenciado la localización del inmueble y las últimas valoraciones realizadas sobre otras viviendas en la misma zona será útil a la hora de tomar una decisión mas correcta.

En segundo lugar, la metodología estadística asociada a esta herramienta asignaría un valor de referencia al inmueble en base a la información suministrada. El valor de referencia asignado, no sólo tendría en cuenta como es habitual las características físicas del inmueble (superficie, antigüedad, estado de conservación, etc) y a su localización (Ciudad, Barrio, Código Postal,...) sino que también mediante técnicas de econometría espacial, modificaría la estimación atendiendo al valor de Tasación asociado a las seis viviendas más próximas. Este valor de referencia es un valor objetivo, carente de apreciaciones individuales y resultado de un modelo econométrico.

Como objetivos secundarios derivados de este objetivo central se plantean también las siguientes metas:

Permitir al Tasador visualizar la localización exacta de la vivienda y su entorno en un plano georeferenciado permitiendo variar la escala de observación.

Realizar un esfuerzo por cuantificar, especificar y analizar de forma clara y sistemática las variables que realmente determinan el valor de tasación de un inmueble.

Detección de tasaciones atípicas espacialmente. Valores muy elevados o muy bajos respecto a su entorno.

Cuantificar valores medios de tasaciones anteriores por zonas (Códigos Postales) al igual que valores medios de las características de las viviendas de cada zona.

Atendiendo a lo anterior, el presente trabajo, presenta la siguiente estructura: En el segundo punto se introducen la regresión lineal simple y múltiple como vía para la valoración inmobiliaria y se pasan a describir los factores de macro y micro valoración como factores determinantes del precio de una vivienda. En el tercer apartado se incluyen los factores de macro-localización y en el cuarto los factores de micro-valoración realizando una descripción teórica de la forma en que debe presentarse el modelo y el proceso de estimación de los parámetros. El quinto apartado se dedica a presentar un caso práctico haciendo uso de la hoja de cálculo EXCEL. Finalmente se presenta la bibliografía y un anexo que incorpora la programación mediante Matlab del problema solucionado en el quinto apartado.

1. Metodología: Tasación por Regresión.

Existen diversos métodos y técnicas en la valoración de bienes inmuebles, y dentro de las técnicas de valoración de tipo estadístico destaca una de las alternativas que está tomando mas fuerza debido a que introduce una mayor objetividad a la hora de asignar un valor a un bien raíz, es aquel que está basado en el empleo de técnicas estadísticas y econométricas, conocido como Tasación por Regresión.

Las técnicas estadísticas evolucionan y a cada vez más presentan modelos que pueden adaptarse a la compleja realidad del mercado. Centraremos esta comunicación en plantear una alternativa dentro de la tasación por regresión basada en la introducción de una doble componente espacial con el fin de determinar el precio de la vivienda, no sólo en base a las características propias del inmueble, como habitualmente se desarrolla en el modelo clásico de regresión, sino en base a uno de los factores que con mayor peso incide en el valor de un inmueble: su localización espacial.

Habitualmente las técnicas de regresión que se aplican para la valoración de bienes inmuebles puede clasificarse como:

Regresión Simple: $Y = a + bX + e$.

En este caso, el modelo estadístico que determina el precio de una vivienda está en función de una única variable explicativa, que debe ser el factor clave. Este tipo de modelos es muy simple pero puede tener su aplicación en aquellas situaciones en las que se intente valorar un conjunto de inmuebles de características muy similares, con una única diferencia básica. Este sería el caso por ejemplo, de un conjunto de viviendas

dentro de una misma promoción, cuya única diferencia sea la superficie de la parcela en la que se encuentra la vivienda.

Regresión Múltiple: $Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n + e$

En la mayoría de las situaciones no es suficiente con una única variable para explicar el comportamiento aleatorio del precio de un inmueble, ya que también contribuyen otros factores objetivos, que diferencian el valor de la vivienda (factores a tener en cuenta pueden ser: calidad de la construcción, antigüedad de la edificación, anchura de la calle en la que se encuentre, altura en caso de edificios, si tiene o no ascensor, el número de baños, etc)

En este caso, el planteamiento matemático del modelo de regresión múltiple es semejante al trazado anteriormente. Ahora los coeficientes b_i representan el peso de cada variable explicativa X_i sobre la variable dependiente que en nuestro caso se corresponde con el precio de compra venta de la vivienda. Así, el coeficiente b_i indicará la variación que sufre el precio de la vivienda al incrementar en una unidad el factor X_i

El espacio como elemento determinante del valor de un inmueble:

Los modelos estadísticos que acabamos de presentar, son los que históricamente se han utilizado en la valoración de inmuebles. Todos estos modelos aún gozando de buenas propiedades estadísticas no son capaces de captar un aspecto básico en la determinación del valor de una vivienda: *la localización*. Bajo el supuesto de idénticas características físicas de dos inmuebles, el valor de cada bien vendrá determinado por su localización dentro del núcleo urbano en el que se encuentre el inmueble.

Aquí es importante que diferenciamos entre dos factores espaciales que influyen en la determinación del valor final de un inmueble: Factores de macro-localización y factores de micro-localización.

Los **factores de macro-localización**, se refieren a los propios del barrio en el que se encuentra la vivienda. Cada barrio o distrito en una ciudad tiene sus propias particulares que inciden en el precio de la vivienda: servicios, estatus social, etc.

Los **factores de micro-localización**, hacen referencia al efecto de mimetismo que se produce en la valoración de los bienes de carácter urbano por el cual, el valor de un inmueble viene en buena parte determinado por el precio de venta de los inmuebles que se encuentran en su entorno o aquellos que le son más próximos.

En este sentido los modelos de regresión clásicos captan con dificultad estos dos tipos de factores, y nuestro objetivo es presentar la utilidad de los modelos espaciales de regresión propios de la econometría espacial en la valoración de inmuebles, mediante la inclusión de factores de macro y micro localización.

3. Técnicas de regresión espaciales: Factores de macro-localización:

Con el fin de incluir los factores de macro-localización, se introducirán en el modelo de regresión como variables explicativas del valor de la vivienda, las coordenadas de longitud y latitud de cada unidad analizada (esto se puede hacer bien mediante un sistema GPS, o simplemente superponiendo una cuadrícula sobre el plano urbano de la ciudad y asignando coordenadas a cada vivienda). Este tipo de modelos no cambia sustancialmente con respecto al modelo de regresión clásico y recoge la tendencia del proceso analizado.

Mediante este tipo de modelos se intentan recoger la variación en el precio de un inmueble en base a su localización global sobre la superficie analizada, siendo común encontrarse con esquemas de tipo lineal que marquen diferencias Norte-Sur o Este-Oeste o bien esquemas cuadráticos para modelar estructuras de Centro-Periferia.

En este caso, el modelo de regresión vendrá representado, en el caso más general, por una función polinómica de un orden determinado, de tal forma que la tendencia del proceso que se analiza puede expresarse matricialmente como,

$$Y = A \theta + e \quad (1)$$

donde A es la matriz de localización de las n zonas y θ es el vector de parámetros.

$$A = \begin{bmatrix} 1 & a_{11} & a_{12} & a_{11}^2 & a_{12}^2 & a_{11}a_{12} & \cdot & \cdot & \cdot & \cdot & \cdot & a_{11}^p a_{12}^q \\ 1 & a_{21} & a_{22} & a_{21}^2 & a_{22}^2 & a_{21}a_{22} & \cdot & \cdot & \cdot & \cdot & \cdot & a_{21}^p a_{22}^q \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & a_{n1} & a_{n2} & a_{n1}^2 & a_{n2}^2 & a_{n1}a_{n2} & \cdot & \cdot & \cdot & \cdot & \cdot & a_{n1}^p a_{n2}^q \end{bmatrix} \quad (2)$$

donde (a_{i1}, a_{i2}) definen la localización de la zona i (por ejemplo, coordenadas de latitud y longitud), y el vector de coeficientes

$$\theta^T = (\theta_{00} \theta_{10} \theta_{01} \theta_{20} \theta_{02} \theta_{11} \dots \theta_{pq})$$

Esta expresión matricial puede también formularse como:

$$Y_i = \sum_{s=0}^p \sum_{r=0}^q \theta_{rs} a_{i1}^r a_{i2}^s + e_i \quad (3)$$

Expresando la idea de que el valor del inmueble coincide con el valor que toma el polinomio en las coordenadas que tiene asignada dicha zona. Este modelo de regresión es conocido como el modelo de superficie tendencia de orden k y la primera aplicación de este modelo se debe a Student (1.914)

La suma $p+q$ representa el orden ($k=p+q$) de la superficie tendencia. Así, una superficie de orden cero ($p+q=0$) se referirá a un plano horizontal, lo que equivale a suponer un valor constante para todas las unidades analizadas; una superficie lineal o de primer orden ($p+q=1$) se corresponde con una superficie plana inclinada indicando una tendencia Norte-Sur o Este-Oeste. De tal forma que al aumentar el orden se generará una superficie cada vez más complicada (cuadrática indicando un modelo Centro periferia, cúbica, etc.)

4. Técnicas de regresión espaciales: Factores de micro-localización:

La introducción de factores de micro-localización dentro del modelo de regresión basándose en las técnicas de econometría espacial requiere un tratamiento más complejo, del presentado hasta ahora que a continuación se detalla:

La matriz de contactos para definir estructura de vecindades

En primer lugar es necesario la definición del concepto de vecindad. El concepto de micro-localización antes definido, hace referencia a la influencia que ejercen aquellas unidades que se encuentran próximas o en el entorno de cada unidad analizada. Por tanto, es necesario formalizar en cada caso que se entiende por “vecino de” cada unidad “i” o, dicho de forma distinta, cuales son las unidades que interaccionan con una dada. La formalización de esta estructura de vecindades conducirá a la determinación de una maya que será introducida en el modelo de regresión haciendo uso de la conocida como matriz W de vecindades o matriz de contactos.

Así la solución a la introducción de una estructura de vecindades pas pues por la definición de la denominada matriz de pesos espaciales, habitualmente denominada como W:

$$W = \begin{bmatrix} 0 & w_{12} & w_{13} & \cdot & w_{1n} \\ w_{21} & 0 & w_{23} & \cdot & w_{2n} \\ w_{31} & w_{32} & 0 & \cdot & w_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ w_{n1} & w_{n2} & w_{n3} & \cdot & 0 \end{bmatrix} \quad (4)$$

Así $W = \{w_{ij}\}$ es una matriz cuadrada no estocástica donde por w_{ij} se cuantifica el efecto de la zona j sobre la zona i . La determinación de los elementos de la matriz W es uno de los problemas que pueden generar más controversia en el análisis de este tipo de modelos, debido a la amplia variedad de escenarios que pueden plantearse modificando los coeficientes que determinan esta matriz y las repercusiones que estas modificaciones tienen sobre el modelo de regresión.

Es también una práctica común en aquellas situaciones para las que sea posible, relativizar las ponderaciones una vez definidas, dividiendo cada elemento de la matriz W por la suma de la fila correspondiente.

$$w'_{ij} = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}} \quad (5)$$

Así tendremos que $\sum_j w_{ij} = 1$, de tal forma que todas las ponderaciones que afectan a la zona i sumen uno (es decir, la matriz W será una matriz estocástica). Esta estandarización de la matriz permite, de forma semejante a lo que se hace en procesos temporales, definir retardos espaciales. Si con la matriz estandarizada (a la que se seguirá llamando W) realizamos el producto WY el i -ésimo elemento del vector columna que se obtiene se corresponde con el valor medio ponderado de los vecinos de la zona i -ésima.

Modelos de regresión espaciales.

En términos generales, los modelos de regresión espacial suelen clasificarse atendiendo a la forma en que se introducen los efectos espaciales en el modelo de regresión. Las alternativas que se plantean tienen básicamente dos líneas diferentes que representan a su vez diferentes concepciones sobre la naturaleza del proceso y no son excluyentes. La primera alternativa introduce retardos espaciales de las variables endógenas y/o exógenas como variables explicativas del modelo, mientras que la segunda alternativa introduce un esquema de dependencia espacial en el término de perturbación. De esta manera se diferenciarán entre:

- Modelos con términos autorregresivos.
- Modelos con errores autocorrelacionados.

Centraremos este trabajo en el primer grupo de modelos, los modelos con términos autorregresivos. Estos modelos incluyen como variables explicativas retardos espaciales de los factores que intervienen en la modelización incluyendo así los términos WY y/o WX como variables exógenas del modelo de regresión, intentando explicar una variable en función de la información que concurre en el entorno (W) de cada localización espacial.

La autocorrelación espacial que puede ser estructurada mediante esta clase de modelos se conoce como autocorrelación espacial *sustantiva* y aparece en los residuos de un modelo de regresión clásico debido a la omisión de alguno de los retardos espaciales que debían haber sido incluidos en el modelo.

En el segundo de los casos, la dependencia espacial no causada por la omisión de retardos en las variables del modelo, se plantea la dependencia espacial residual (también conocida como autocorrelación *nuisance*). Que puede ser causada por la omisión de variables no consideradas en el modelo o por la falta de correspondencia entre la unidad de observación y el ámbito espacial del fenómeno.

Sea cual sea la forma en la que se presente la dependencia espacial, la presencia de autocorrelación lleva consigo, al igual que en los procesos temporales, una serie de consecuencias indeseables en el proceso de inferencia.

Así, en el caso de los modelos con términos autorregresivos, las consecuencias de aplicar MCO como método de inferencia, si bien las estimaciones son insesgadas, estas serán ineficientes lo que conduce a que las estimaciones por mínimos cuadrados de β se encuentren sobrestimadas en comparación con las obtenidas por mínimos cuadrados generalizados (supuesto conocida la estructura de covarianzas del proceso). También se producirá una sobre valoración del coeficiente de determinación, y valores erróneos de t y F . (Anselin 1.988). En el caso de que en el modelo de regresión se presenten retardos de la variable independientes como variables dependientes, las consecuencias son aún más graves ya que la estimación mínimo cuadrática presentará, estimadores sesgados e inconsistentes, independientemente de que la perturbación esté o no correlacionada espacialmente.

En el caso de la aplicación de estas técnicas estadísticas a la tasación inmobiliaria, el producto de esta matriz W por el vector Y que recoge en nuestro caso los precios de las viviendas, que escribiremos como WY será también un vector cuya i -ésima componente hará referencia al precio de las viviendas del entorno. De esta manera, la inclusión de esta nueva variable como un factor endógeno más del modelo recogerá la influencia que ejerce sobre el precio de un inmueble el precio de sus vecinos. Esta variable rectificará por tanto los valores previstos en función del precio de venta de las viviendas del entorno.

EL PRECIO DE LA VIVIENDA EN LA CIUDAD DE CARTAGENA.

El objetivo de este trabajo, es intentar construir un modelo de valoración de viviendas no en función de factores temporales (evolución de la bolsa, tipos de interés,...etc) sino en función de las características físicas de las viviendas en un momento determinado del tiempo. Vamos a intentar determinar cuales son los factores que la población en mayor medida va a valorar a la hora de adquirir un bien para toda la vida o lo que es lo mismo de que va a depender el precio de la vivienda en un momento determinado del tiempo, haciendo especial hincapié en la localización del bien ya que suponemos a priori que es un factor fundamental a la hora de su adquisición.

A continuación y antes de realizar el análisis exploratorio de datos vamos a presentar, la muestra y los softwares utilizados tanto para localizar como para analizar la información disponible gracias a TABIMED (Sociedad Tasadora de la Caja de Ahorros del Mediterráneo).

Una vez realizado el análisis exploratorio de datos a través de SPSS y Excel por barrios, distritos y códigos postales vamos a realizar el análisis exploratorio de datos espaciales a través de Geoda, no sin olvidarnos de realizar un análisis de la regresión lineal a través de SPSS primero para finalizar con un análisis a través de Geoda.

5.1. Los Datos: La muestra y sus características

Para realizar este estudio se ha utilizado el contenido de 344 tasaciones, correspondientes a 344 viviendas de la ciudad de Cartagena, proporcionados por TABIMED (Sociedad Tasadora de la Caja de Ahorros del Mediterráneo) y realizadas en el primer semestre del 2004.

Inicialmente contábamos con el contenido de 438 tasaciones pertenecientes a viviendas de la ciudad de Cartagena y sus alrededores (La Vaguada, Isla Plana, Isla Menores, Los

Belones, Canteras, Molinos Marfagones, Playa Honda, Los Nietos, Mar de Cristal, La Palma, El Albuñón, Cabo de Palos y La Manga), éstas últimas han sido eliminadas ya que su estudio no era de nuestro interés.

En concreto, tenemos información de distintas características para cada una de las viviendas:

- _Superficie Útil.
- _Superficie Construida.
- _Superficie Construida con Comunes.
- _Fecha de la Tasación.
- _Barrio donde se ubica la vivienda.
- _Calle.
- _Número.
- _Código Postal.
- _Valor global de la Tasación.

A partir de estos datos hemos creado dos nuevas variables:

- _Valor de la tasación por metro cuadrado (Valor Global de Tasación/Superficie Útil).
- _Valor de la tasación por metro cuadrado (Valor Global de Tasación/Superficie Construida con Comunes).

Y dos últimas variables correspondientes a la ubicación de cada una de las viviendas en la cartografía digital de Cartagena a través de Autocad:

- _Coordenadas de latitud y longitud para cada vivienda en el plano de Cartagena.

5.2.) El software: localización y análisis de la información.

- i. Autocad.
- ii. GeoDa.
- iii. ArcView.
- iv. SPSS.
- v. Excel.

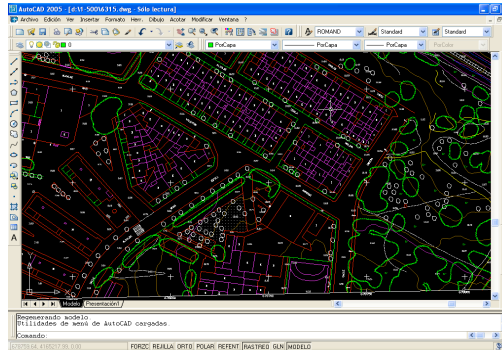
Los tres primeros son Sistemas de Información Geográfica, que van a ser utilizados para localizar las viviendas en el plano.

Se conocen como sistemas de información geográfica, SIG, o GIS en la denominación inglesa (Geographic Information Systems), a un conjunto de programas y aplicaciones informáticas que permiten la gestión de datos organizados en base de datos, referenciados espacialmente y que pueden ser visualizados mediante mapas.

Los dos últimos, SPSS (Base de Datos) y Excel (Hoja de Cálculo) va a ser utilizados para analizar la información disponible.

i) Autocad.

En concreto, para situar las viviendas en el plano de Cartagena, a través de su cartografía digital hemos utilizado uno de los programas de diseño gráfico más utilizado en el campo de la investigación como es el **Autocad**, también he de señalar la utilización de forma complementaria con Autocad del sistema de localización desarrollado por Cartomur (Producciones Cartográficas de la Región de Murcia) a través de su página web www.cartomur.com.



ii) GeoDa.

Se trata de un nuevo programa informático, para el análisis exploratorio de datos espaciales desarrollado por el Profesor Luc Anselin de la Universidad de Illinois, y tiene la ventaja de ser, hasta el momento, un producto gratuito y de libre acceso en Internet (<http://sal.agecon.uiuc.edu/csidd/geoda.html>).

En términos generales, las diferentes funciones de GeoDa podrían ser clasificadas en 6 categorías: tratamiento de datos geográficos, transformación de datos, representación gráfica en mapas, gráficos estadísticos, dependencia espacial y regresión espacial.

iii) ArcView.

ArcView es la herramienta SIG más extendida en todo el mundo dadas sus avanzadas capacidades de visualización, consulta y análisis de información geográfica, además de las numerosas herramientas de integración de datos desde todo tipo de fuentes.

Por sí solo, ArcView permite la explotación de toda la información tanto en sistemas monousuario como en sistemas departamentales, pero es al integrarse en la [arquitectura ArcGIS](#) donde se consigue una solución global en el manejo de información geográfica y escalable según las necesidades del usuario.

iv) SPSS y v) Excel.

SPSS y la Hoja de cálculo de Excel son dos de las herramientas más utilizadas en la actualidad para el análisis exploratorio de datos, en este trabajo van a ser utilizadas para llevar a cabo el mismo destacando su simplicidad y claridad a la hora de estudiar y analizar las relaciones existentes entre los datos.

Capítulo 6

ANÁLISIS EXPLORATORIO DE DATOS ESPACIALES.

6.1.) INTRODUCCIÓN

El análisis exploratorio de datos espaciales, es una disciplina más dentro del general Análisis Exploratorio de Datos, que ha sido diseñada para el tratamiento específico de datos espaciales o geográficos. Puede definirse como el conjunto de técnicas que permiten describir y visualizar las distribuciones espaciales, identificar localizaciones atípicas (“outliers”), descubrir esquemas de asociación espacial (“clusters”) y sugerir estructuras espaciales u otras formas de heterogeneidad espacial (Anselin 1999).

El AEDE combina el análisis estadístico con el análisis gráfico, dando lugar a lo que podría denominarse “visualización científica” (Haining *et al.*, 2000), ya que a los contrastes estadísticos sobre los efectos espaciales de dependencia y heterogeneidad,

une un amplio marco de gráficos o “vistas” múltiples y dinámicas sobre la información geográfica (Unwin, 2000).

En este apartado vamos a desarrollar un conjunto de técnicas de especificación, estimación, contraste y predicción necesarias para el tratamiento de datos espaciales, en modelos de regresión de corte transversal y datos panel.

La naturaleza de los datos geográficos da lugar a los llamados “efectos espaciales” de dependencia o autocorrelación espacial y heterogeneidad espacial, por ello deben de ser considerados de forma explícita y adecuadamente tratados en todo ejercicio de modelización y predicción de datos geográficos.

Para llevar a cabo el análisis espacial de los datos vamos a utilizar, el paquete informático GeoDa Versión 0.9.5-i (Beta), se trata de un software para el análisis de datos espaciales, que va a ser utilizado para localizar y estudiar la relaciones existentes

entre las viviendas en el plano de la ciudad de Cartagena. Geoda, está concebido como un producto autosuficiente que no requiere de un sistema específico de GIS (Sistema de Información Geográfica) y funciona en cualquiera de los sistemas operativos de Windows y Macintosh.

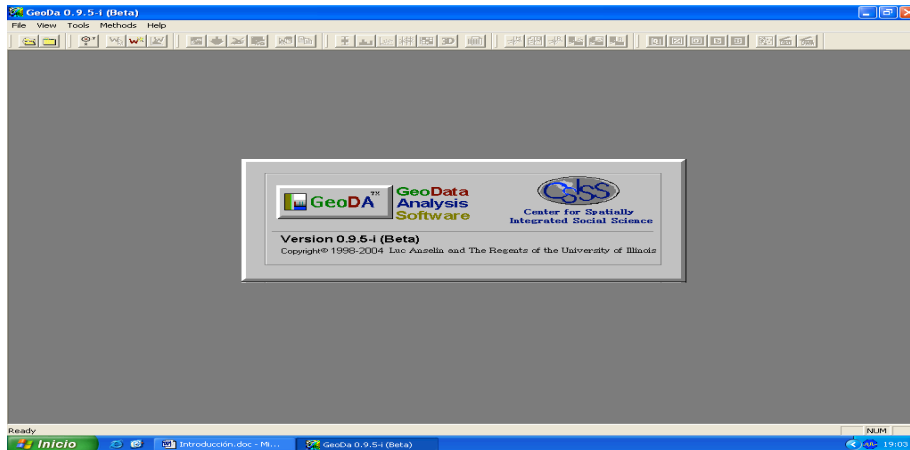
Se conocen como sistemas de información geográfica, SIG, o GIS en la denominación inglesa (Geographic Information Systems), a un conjunto de programas y aplicaciones informáticas que permiten la gestión de datos organizados en base de datos, referenciados espacialmente y que pueden ser visualizados mediante mapas.

Se trata de un nuevo programa informático (Figura 1), para el análisis exploratorio de datos espaciales desarrollado por el Profesor Luc Anselin de la Universidad de Illinois, y tiene la ventaja de ser, hasta el momento, un producto “gratis”, es decir, de libre acceso en Internet.

(<http://sal.agecon.uiuc.edu/csidd/geoda.html>), en esta página web, hay diversos materiales de apoyo así como una película del programa “Quicktime” en la que se hace una demostración de las características principales de este programa. La versión más reciente del programa, 0.9.5-i5, data de Agosto de 2004 y está siendo muy bien recibida por todo tipo de usuarios, especialmente los dedicados al mundo académico y de la investigación, en general (en septiembre de 2004, GeoDa tenía registrados a 3.500 usuarios, que se incrementan en una tasa de 150 nuevos cada mes).

En términos generales, las funciones de GeoDa podrían ser clasificadas en 6 categorías: tratamiento de datos geográficos, transformación de datos, representación gráfica en mapas, gráficos estadísticos, dependencia espacial y regresión espacial.

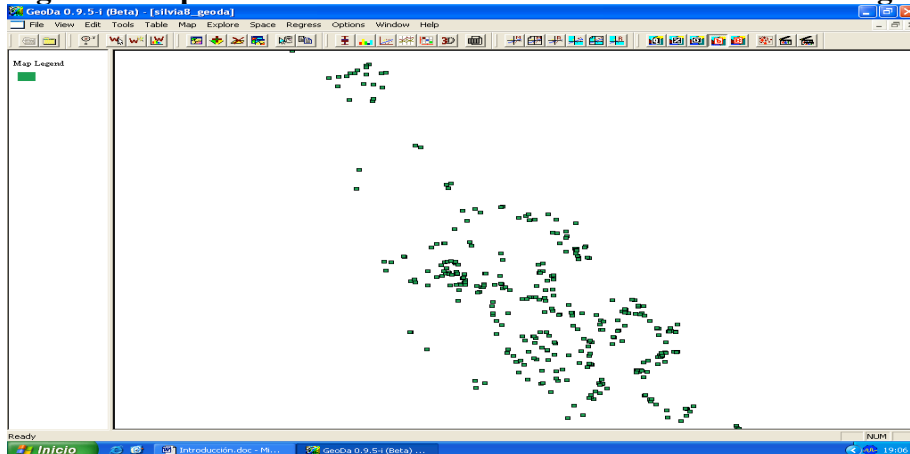
Figura 1: GeoData Análisis Software



Fuente: Elaboración propia a partir de GeoDa

GeoDa, nos va a permitir representar las viviendas de Cartagena (Figura 2), de la siguiente manera:

Figura 2: Representación de las viviendas de la ciudad de Cartagena



Fuente: Elaboración propia a partir de GeoDa

Para el cometido del presente trabajo, es decir, para analizar el precio de las viviendas en la ciudad de Cartagena, disponemos de 295 viviendas con sus respectivas características físicas y de localización (superficie útil, valor de tasación, barrio, código postal,...tc), como así estudiamos en el análisis descriptivo de los datos, además de una serie de características que nos van a medir la calidad de las viviendas (Entorno Comercial, Número de Viviendas en el edificio, Antigüedad de la Construcción, Renta media de los habitantes de la zona, Orientación de la vivienda y Calidad de la Construcción).

Cada punto en el plano representa una vivienda, localizada en la ciudad de Cartagena.

En el apartado siguiente, se presentan los principales métodos de AEDE con ayuda de GeoDa, pero antes de llegar a la presentación de dichos métodos vamos a introducir y definir un elemento importantísimo de cara al AEDE y es la llamada “matriz de pesos”

Para analizar las relaciones existentes entre las viviendas en la ciudad de Cartagena, vamos a definir la matriz de pesos (W_{ij}) de distintos órdenes, para cada una de las variables.

$$W = \begin{bmatrix} 0 & w_{12} & \cdot & w_{1N} \\ w_{21} & 0 & \cdot & w_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ w_{N1} & w_{N2} & \cdot & 0 \end{bmatrix}$$

Al ser la dependencia espacial multidireccional, no podemos utilizar el operador de retardos L , $L^p X = x_{t-p}$, presente en el contexto temporal, el cual recoge únicamente una relación unidireccional (Pasado>Presente o Presente>Futuro), sino que es necesario la utilización de una matriz exógena que recoja a través de cada uno de sus elementos

denominados pesos (w_{ij}) la intensidad de la dependencia que existe entre cada par de viviendas i y j , dicha matriz es conocida como la matriz de pesos.

Pero, ¿cómo se construye dicha matriz?

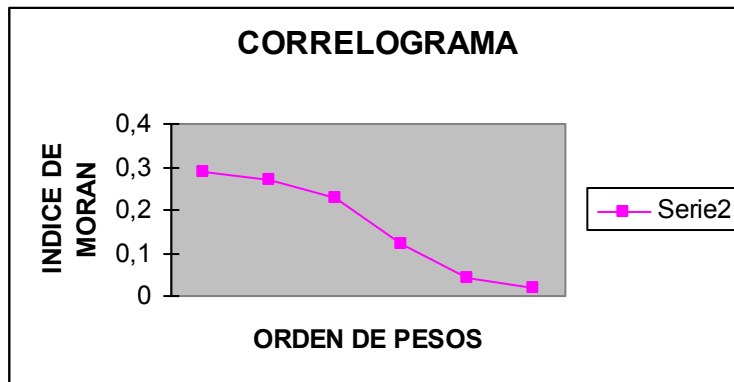
Según el criterio del vecino más próximo:

Suponiendo una muestra de n casos, se elige un subconjunto de dicha muestra entonces; $W_{ij} = 1$ si el caso j que se elija está dentro de ese subconjunto de observaciones vecinas a i

$W_{ij} = 0$ en caso contrario

La matriz de pesos puede tener distintos órdenes, dependiendo del criterio de agrupación que elijamos, como vamos a mostrar, a continuación lo que sí debemos de tener claro es que el Índice de Moran (pendiente de la recta de regresión), que nos va a medir el grado de autocorrelación espacial, va a disminuir (Figura 3) conforme incrementemos el orden de dicha la matriz.

Figura 3: Representación gráfica del valor del Índice de Moran conforme aumenta el orden de la matriz de pesos.



Fuente: elaboración propia a partir de Excel

Para ver el contenido de la matriz de pesos vamos a abrir desde el block de notas (Figura 4) de la matriz de pesos de orden 6 (W_6), recogiendo para cada vivienda (w_{ij}) sus 6 vecinas más próximas.

Así tenemos;

Figura 4: Block de notas de la matriz de pesos de orden 6

Fuente: elaboración propia a partir del block de notas

Cuya interpretación, por ejemplo para la vivienda número 1 sería; Las 6 viviendas vecinas a la vivienda 1 son las números, 36, 106, 74, 80, 198 y 37 y la distancia correspondiente de cada una de ellas a dicha vivienda es la columna de al lado (80, 199, 233, 234, 239 y 256 respectivamente).

Efectivamente estas viviendas pertenecen todas al barrio de San Antón, lo que indica que son todas vecinas de la primera vivienda perteneciendo todas ellas al mismo barrio.

6.2.) METODOS GRAFICOS DEL AEDE: VISUALIZACION DE LA PRESENCIA DE EFECTOS ESPACIALES.

Aunque existen gran cantidad de métodos gráficos para el Análisis Exploratorio de Datos Espaciales (AEDE), existen pocos estudios que valoren la utilidad y efectividad de los mismos. Un buen método de AEDE es aquel capaz de analizar y representar dos características fundamentales en toda distribución espacial: tendencia y puntos atípicos. La tendencia de una distribución tiene un sentido global, es decir, referido a todo el mapa (y no a una parte del mismo), mientras que los atípicos espaciales (“spatial outliers”), suelen producirse en determinadas zonas del mapa, normalmente en forma de clusters de valores muy altos/bajos en comparación con los valores de su entorno.

TABLA 19: MÉTODOS DE ANALISIS EXPLORATORIO EN GEODA.

	Tendencia espacial	Atípicos espaciales
AED general	Histograma de frecuencias	Diagrama de caja
	Diagrama de dispersión	
	Gráfico de coordenadas paralelas	
	Gráfico de dispersión en 3D	
AEDE reticular	Mapas temáticos	Mapa de caja
	Mapa dinámico	Mapa de percentiles
	Gráficos condicionales	Cartograma
	Diagrama de dispersión de Moran	Gráficos LISA
	Diagrama de dispersión de	Gráficos LISA multivariantes

Fuente: Coro Chasco (2004)

En la primera columna de la Tabla 16 se exponen los métodos más destacados, según que sean métodos generales del AED o métodos explícitos del AEDE reticular.

En ambos casos, se presentarán no sólo técnicas univariantes, sino también multivariantes.

En el presente trabajo se va a trabajar con algunos de estos métodos gráficos, para demostrar la funcionalidad del programa Geoda (para algunas de las variables estudiadas). El primer grupo de ellos tratan de estudiar, la tendencia espacial a nivel global mientras que el segundo grupo va a estudiar la dependencia a nivel local.

6.2.1.) Métodos de representación de la tendencia espacial global.

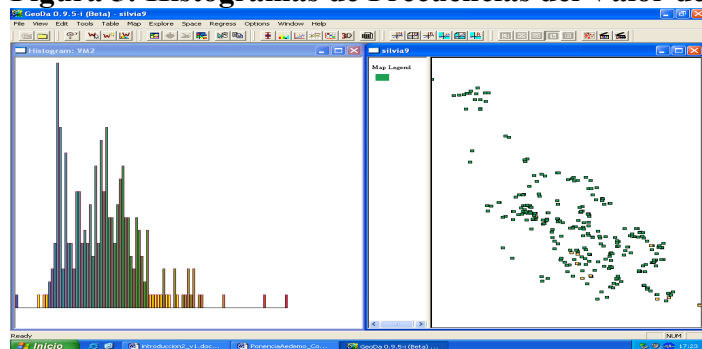
Son herramientas de representación cartográfica de sentido global, para las que lo importante no es el mapa en sí mismo, sino la representación gráfica de los estadísticos básicos.

A) Histogramas de frecuencias.

El histograma de frecuencias, es un gráfico estadístico del análisis general exploratorio de datos, que va a representar las frecuencias de la variable elegida, para distintas clasificaciones, aunque el número por defecto es 7. 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.

Así el histograma de frecuencias para la variable valor por metro cuadrado (VM2), sería;

Figura 5: Histogramas de Frecuencias del Valor de Tasación por metro cuadrado

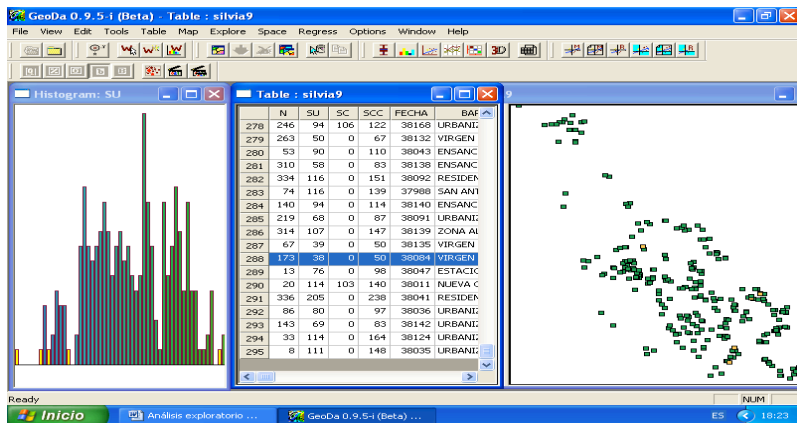


Fuente: Elaboración propia a partir de GeoDa

En el gráfico (Figura 5) observamos que las viviendas con mayor valor de tasación por metro cuadrado (con menor longitud de barra) son las que pertenecen a la zona Centro de Cartagena o a zonas residenciales como Ciudad Jardín o el Polígono Residencial de Santa Ana.

Para la Superficie Útil (SU) el histograma de frecuencias sería:

Figura 6: Histograma de frecuencias para la Superficie Útil



Fuente: Elaboración propia a partir de GeoDa

Al igual que para el Valor de Tasación, el histograma de frecuencias de la superficie útil (Figura 6), nos muestra las viviendas agrupadas por categorías. Dicha agrupación nos permite visualizar las viviendas con menor superficie útil (correspondientes a zonas céntricas de Cartagena) y las de mayor superficie útil (zonas residenciales a las afueras de la ciudad).

B) Mapas temáticos.

Los mapas temáticos, en general, consisten en la representación cartográfica de una variable geográfica. Esta representación puede llevarse a cabo mediante símbolos y colores que pongan de manifiesto el valor de una variable en cada una de las unidades geográficas consideradas (países, regiones, etc.). Puede utilizarse un color/símbolo diferente para cada valor o para cada intervalo de valores de la variable.

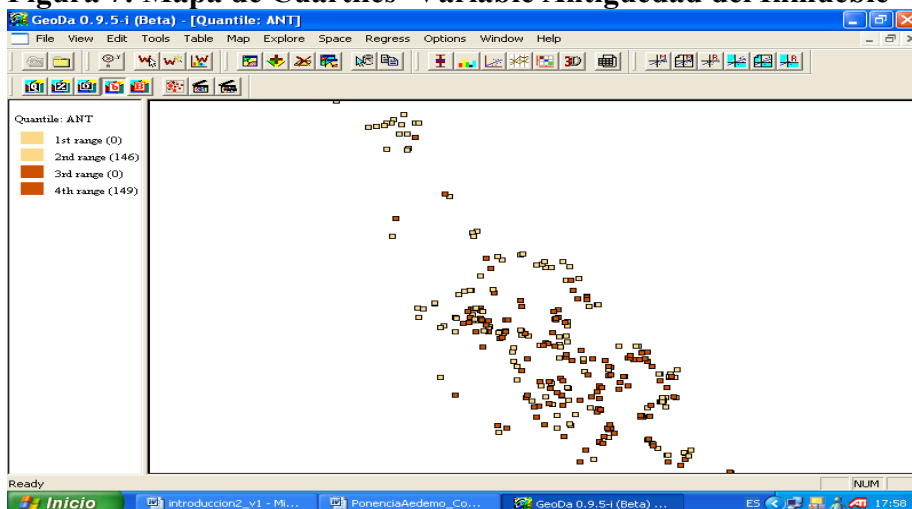
Los mapas temáticos más importantes para la representación de la tendencia espacial de una variable son el mapa de cuantiles y el mapa de la desviación típica.

B.1) Mapa de cuantiles.

En este tipo de mapas, los datos se dividen y agrupan en una serie de categorías (cuantiles) con igual número de observaciones. Este tipo de mapa será un mapa de cuantiles si la distribución se divide en 4 grupos con igual número de unidades territoriales.

Por ejemplo para la variable Antigüedad de la construcción (ANT) el mapa de cuantiles sería;

Figura 7: Mapa de Cuantiles- Variable Antigüedad del Inmueble



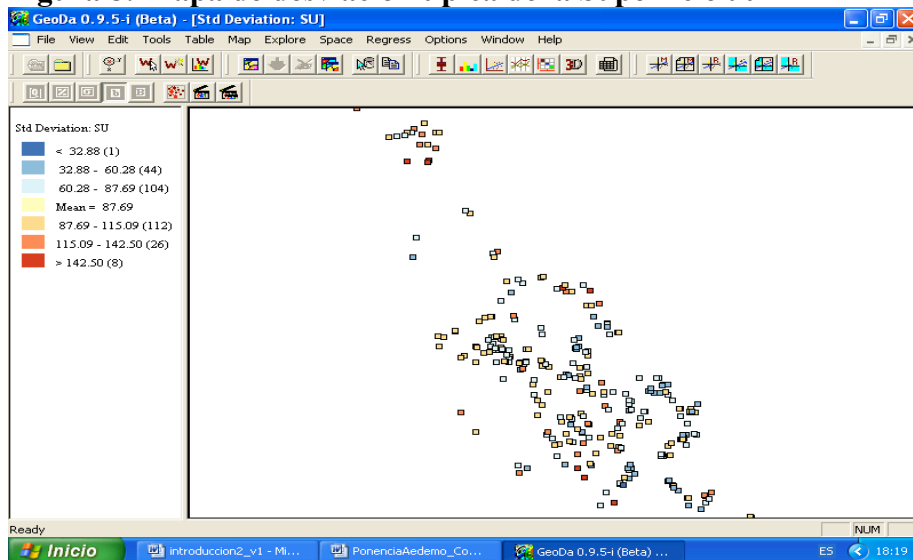
Fuente: Elaboración propia a partir de GeoDa

De esta forma y a través de la representación gráfica del mapa de cuartiles (Figura 7) observamos, que las viviendas con mayor antigüedad se encuentran en la zona centro de Cartagena, mientras que las de menor antigüedad se encuentran en las zonas periféricas de la ciudad de Cartagena.

B.2) Mapa de la desviación típica.

Este mapa agrupa las observaciones según que sus valores caigan dentro de un rango, entendido éste como un número determinado de unidades de la desviación típica a partir de la media. Se trata de un mapa temático en el que las categorías en las que se divide la variable se corresponden con múltiplos de las unidades de la desviación típica. Así, tenemos para la Superficie Útil, el siguiente mapa de desviación típica:

Figura 8: Mapa de desviación típica de la Superficie útil



Fuente: Elaboración propia a partir de GeoDa

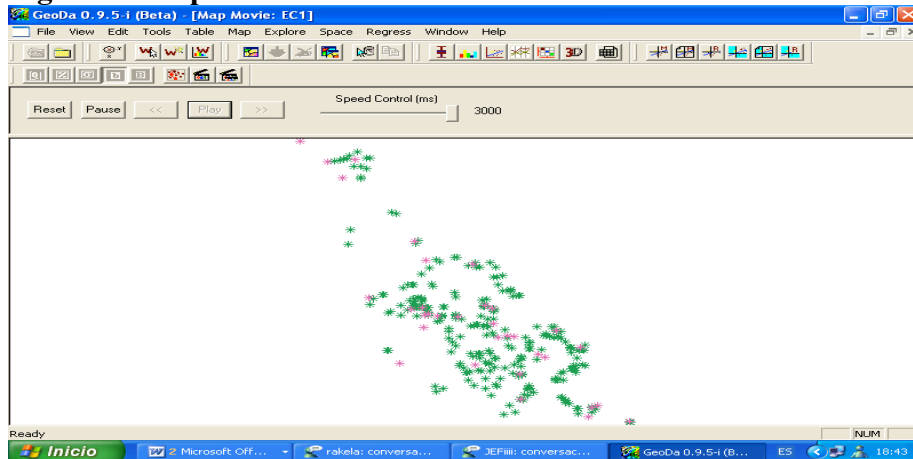
GeoDa divide la variable en 7 intervalos, de forma que el intermedio (nº 4) coincide con el valor de la media y los 6 restantes vienen determinados por 1, 2 y 3 unidades de la desviación típica. Por ejemplo, el intervalo 5º (87,69 a 115,09) está constituido por las 26 viviendas, cuya superficie útil se encuentra entre la media (87,69) y una unidad de la desviación típica (115,09-87,69) como así observamos en la figura número 8.

B.3) Mapa dinámico.

El mapa dinámico (“map movie”) es una forma de animación del mapa que consiste en la iluminación, sobre el mismo, de las distintas unidades geográficas según su orden respecto del valor que adoptan en una variable (siempre de menor a mayor). GeoDa permite la visualización de este mapa de forma simple (las unidades se iluminan una vez, cuando les corresponde, y se apagan a continuación) o acumulativa (las unidades permanecen iluminadas hasta que el mapa entero queda iluminado).

Para la variable entorno comercial (EC1) el mapa dinámico acumulativo sería:

Figura 9: Map Movie EC1



Fuente: Elaboración propia a partir de Geoda.

De esta forma, observamos que las viviendas del centro de Cartagena, y las de La Alameda de San Antón junto con las de Ciudad Jardín son las últimas que se van a iluminar, al ser las zonas con mayor entorno comercial de Cartagena (Figura 9).

6.2.2.) Análisis exploratorio de la dependencia espacial global

La dependencia o autocorrelación espacial consiste en la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que sucede en lugares cercanos o vecinos al mismo. Es decir, una variable estará espacialmente autocorrelacionada cuando los valores observados en un punto o región dependan de los valores observados en regiones vecinas, de forma que se produzca una cierta continuidad geográfica en la distribución de esta variable, por ejemplo, sobre un mapa. Los diagramas de dispersión de Moran (univariante/bivariante), nos van a permitir estudiar este tipo de asociación espacial.

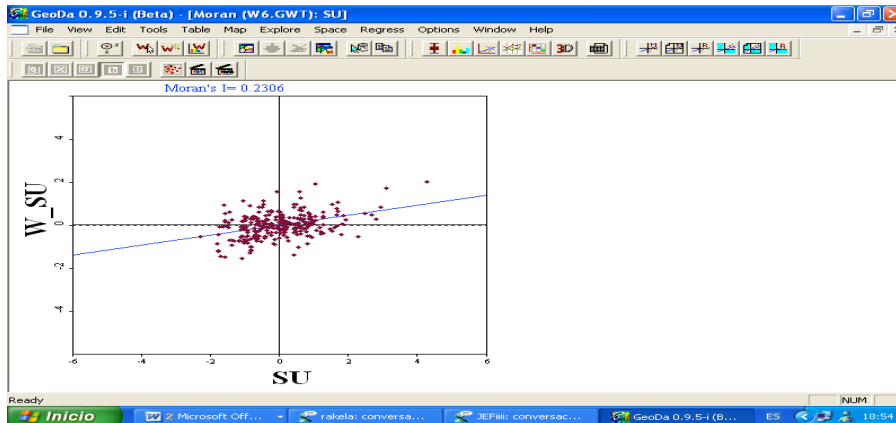
A) Univariante Moran

Se trata de un diagrama de dispersión que representa en el eje X la variable previamente estandarizada y en el eje Y se representa el retardo espacial de dicha variable estandarizada. Se entiende por retardo espacial el promedio ponderado de los valores que adopta una variable en el subconjunto de observaciones vecinas a una dada.

En este tipo de diagrama de dispersión, en el que se relacionan, para cada observación, el valor de la variable en la misma y el valor promedio en sus correspondientes observaciones vecinas, la pendiente de la recta de regresión es el valor del denominado estadístico I de Moran de autocorrelación espacial global¹. Así, cuanto mayor sea el valor de este estadístico, es decir, el ángulo que forme la recta de regresión con el eje de abscisas, más fuerte será el grado de autocorrelación espacial en la

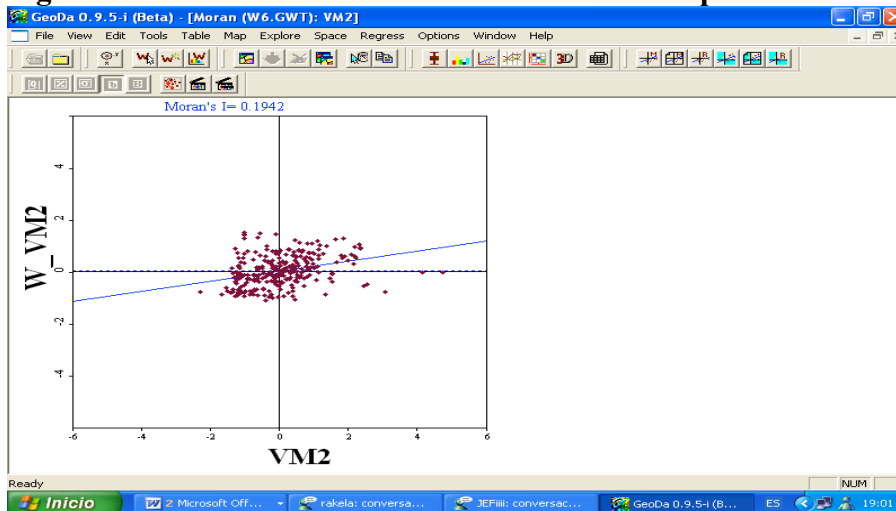
variable, y viceversa. Opción disponible en Geoda 0.9.5-i (Beta) en Space > Univariate Moran.

Figura 10: Índice de Moran de la Superficie Útil



Fuente: Elaboración propia a partir de Beoda

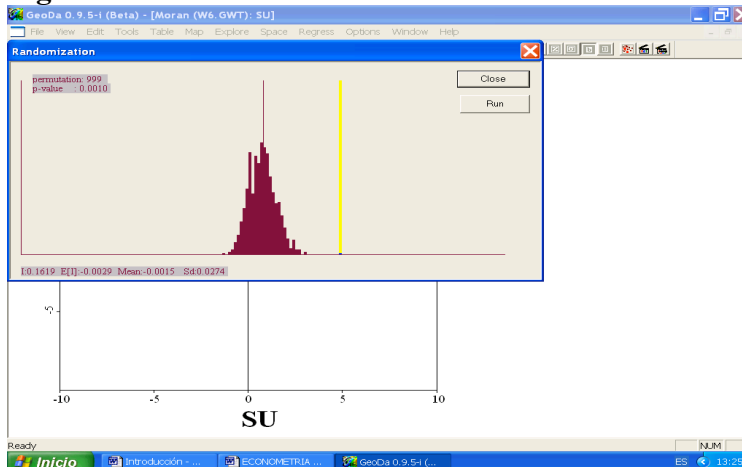
Figura 11: Índice de Moran del Valor de Tasación por metro cuadrado



Fuente: Elaboración propia a partir de GeoDa

En el estudio que llevamos a cabo la variable superficie útil tiene un mayor grado de dependencia espacial que el valor de tasación por metro cuadrado, ya que el Índice de Moran de la Superficie Útil (0,23) es superior al valor del Índice de Moran del valor de tasación por metro cuadrado (0,19).

Figura 12

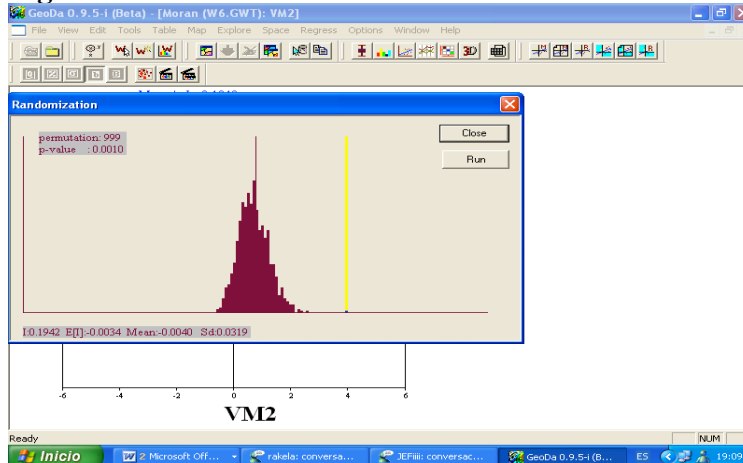


Fuente: Elaboración propia a partir de GeoDa

Con un índice de Moran igual a 0.23, la variable superficie útil presenta una estructura de dependencia espacial significativa ya que tiene un p-value inferior a 0.05 (0.0010) a 999 permutaciones (Figura 12).

Para el Valor de Tasación por metro cuadrado tenemos también un p-value inferior a 0.05 (0.0010) a 999 permutaciones por lo que la variable VM2 también presenta una estructura de dependencia espacial significativa (Figura 13):

Figura 13



Fuente: Elaboración propia a partir de GeoDa

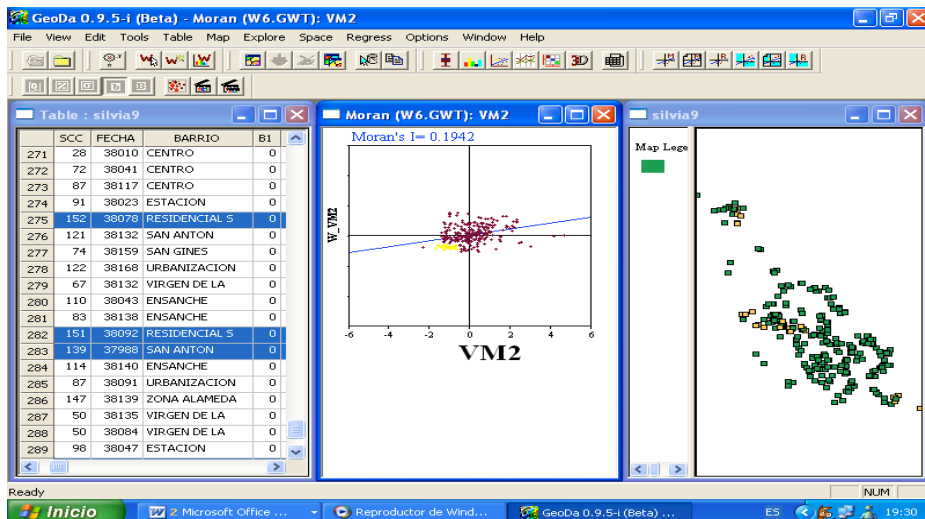
La interpretación del test I de Moran permite conocer el grado en que este estadístico resume la estructura global de asociación o dependencia lineal existente que, en el caso de la superficie útil, para el total de la distribución, sería del 23% y en el caso del Valor de Tasación por metro cuadrado el 19% .

Este diagrama de dispersión suele dividir el tipo de asociación espacial en cuatro categorías: dos para autocorrelación espacial positiva (valores altos de una variable rodeados de valores altos o valores bajos rodeados de valores bajos) y dos para autocorrelación espacial negativa (valores altos rodeados por valores bajos, y viceversa). Las categorías de asociación espacial positiva se corresponden con los cuadrantes I y III, mientras que los cuadrantes II y III recogen las categorías de asociación espacial negativa.

Así para el valor de tasación por metro cuadrado;

En los cuadrantes I y III se presentan aquellos barrios que, con un valor de tasación por metro cuadrado superior/inferior a la media municipal (valor estandarizado positivo/negativo de la variable) se encuentran rodeados de barrios que también disponen de valores de tasación por metros cuadrado superiores/inferiores a la media municipal (retardo espacial positivo/negativo), respectivamente. Así en el cuadrante, número I, tenemos viviendas situadas en la zona de La Alameda, Ciudad Jardín y el Centro de Cartagena y en el cuadrante número III viviendas que pertenecen a la zona de Santa Lucía, Urbanización Mediterráneo, Residencial Santa Ana y San Antón (Figura 14).

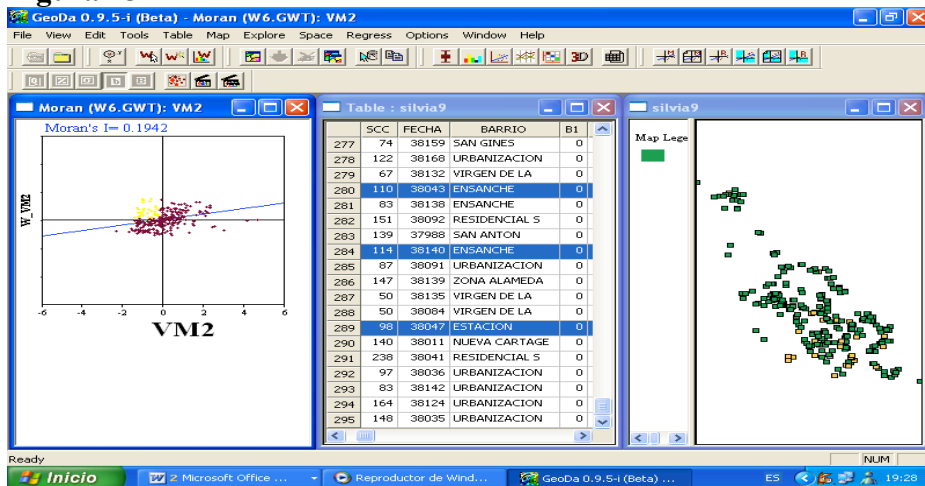
Figura 14



Fuente: Elaboración propia a partir de GeoDa

Por el contrario, las categorías de asociación negativa vienen dadas por los cuadrantes II y IV de este diagrama, en los que se representan los viviendas con valores bajos/altos de valor de tasación por metro cuadrado rodeadas por viviendas con valores altos/bajos de dicha variable (Ensanche, Estación, Virgen de la Caridad y Barrio Peral), respectivamente (Figura 15).

Figura 15



Fuente: Elaboración propia a partir de GeoDa

Antes de terminar, debemos de señalar que el índice de Moran proporciona información tanto de la asociación espacial global (pendiente de la recta de regresión) como de la asociación espacial local. De este modo, los valores en el diagrama de dispersión de Moran situados a más de dos unidades del origen (valor 0) pueden considerarse como puntos atípicos de la distribución.

A) Bivariante de Moran.

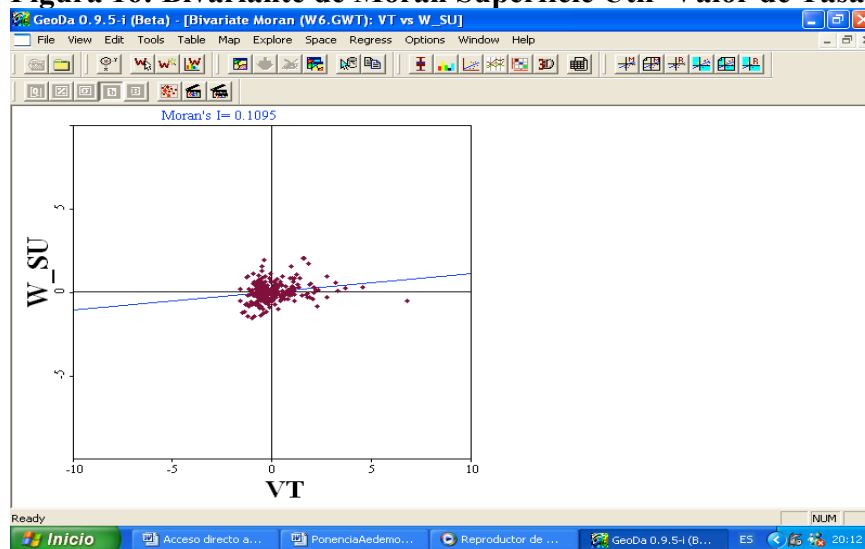
Se trata de un diagrama de dispersión de Moran en el que se representa en el eje vertical Y el retardo espacial de la variable que se desea explicar y en el eje X la variable explicativa. Ambas variables deben estar previamente estandarizadas, de tal forma que la media de ambas sea cero y la desviación típica sea uno. Es decir a través de este gráfico vamos a ver la relación que existe, en cada punto de la muestra, entre los valores de una variable y el valor medio de otra variable en el entorno de dicho punto.

La pendiente de la línea de regresión muestra el grado de relación lineal existente entre la variable del eje horizontal y los valores de la variable del eje vertical en los puntos vecinos a uno dado.

Este tipo de gráficos es muy útil ya que al cambiar el valor de una determinada variable provoca un cambio en el resto de variables de sus vecinas.

Permite estudiar como se encuentran distribuidas dos variables en el modelo que estamos planteando, pudiendo estudiar y señalar los casos atípicos. Para la Superficie Útil-Valor de Tasación el gráfico sería;

Figura 16: Bivariante de Moran Superficie Útil- Valor de Tasación.



Fuente: Elaboración propia a partir de GeoDa

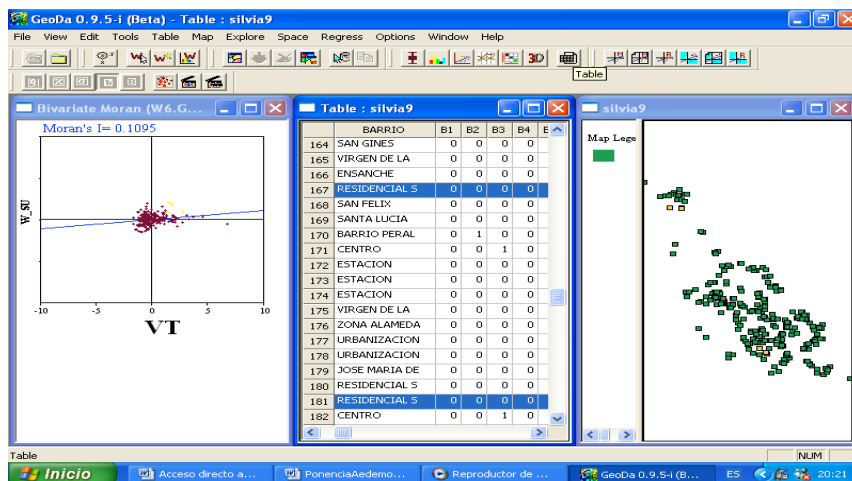
En el cuadrante superior de la derecha (Figura 17), tendríamos aquellas viviendas que tienen un alto valor de tasación y una alta superficie útil (si las señalamos, aparecen marcadas tanto en la tabla como en el mapa) y son:

Caso nº 167 (Polígono Residencial de Santa Ana; Superficie útil>173 Valor de Tasación>223709€).

Caso nº 181 (Polígono Residencial de Santa Ana; Superficie Útil> 205, Valor de Tasación>212185€)

Caso nº 291 (Polígono Residencial de Santa Ana; Superficie Útil>205, Valor de Tasación>209545€)

Figura 17



Fuente: elaboración propia a partir de GeoDa

Y así sucesivamente, estos casos son llamados “normales” ya que si una vivienda tiene un gran valor de tasación, lo “normal” es que tenga una superficie útil elevada. Estas viviendas suelen estar situadas en zonas residenciales a las afueras de la ciudad, o suelen ser viviendas antiguas situadas en la zona centro.

En el cuadrante superior de la izquierda tendríamos la primera clase de viviendas “raras”, ya que tienen un alto valor de tasación y un valor de la superficie útil pequeño y son:

Caso nº 31 (Centro; Superficie Útil>59, Valor de Tasación>113.128€)

Caso nº 94 (Centro; Superficie Útil>44, Valor de Tasación>117.653€)

Caso nº 272 (Centro; Superficie Útil>50, Valor de Tasación>97.653€)

Y así con todas las marcadas

En el cuadrante inferior de la izquierda, tenemos aquellas viviendas que son consideradas normales ya que poseen un valor de tasación bajo con un superficie útil, también pequeña. La mayoría de estas viviendas corresponden a barrios periféricos de clase obrera (Barrio Peral, Urbanización Mediterráneo y barriada Virgen de la Caridad).

En el cuadrante inferior de la derecha tenemos el segundo grupo de viviendas denominadas “raras”, ya que poseen un valor de tasación bajo con una superficie útil elevada (Este tipo de viviendas corresponden a zonas no muy céntricas de la ciudad, soliendo ser viviendas antiguas como algunas de las que se encuentran en la zona del Ensanche).

El Índice de Moran, alcanza el valor de 0.1095, siendo significativo (p-value: $0.010 < 0.05$) a 999 permutaciones.

6.2.3.) Métodos de representación de los atípicos espaciales.

Los métodos de representación de los atípicos espaciales, analizan la existencia de concentraciones de observaciones cuyo valor se encuentra a cierta distancia de la tendencia general (mediana). Este hecho suele producirse en determinadas zonas del mapa general, en forma de clusters de valores muy altos/bajos en comparación con los del entorno. Vamos a estudiar los gráficos de caja, el mapa de percentiles y los gráficos LISA como métodos de representación de los atípicos espaciales

A) Gráficos de caja (diagrama y mapas de caja).

Los gráficos de caja constituyen un método de representación a nivel univariante. El diagrama de caja parte del cálculo de los cuartiles y la media de una variable, así como de la obtención de las llamadas cotas superiores e inferiores, que se obtienen, a su vez, como el producto de los valores del tercer (primer) cuartil por 1,5 veces el recorrido

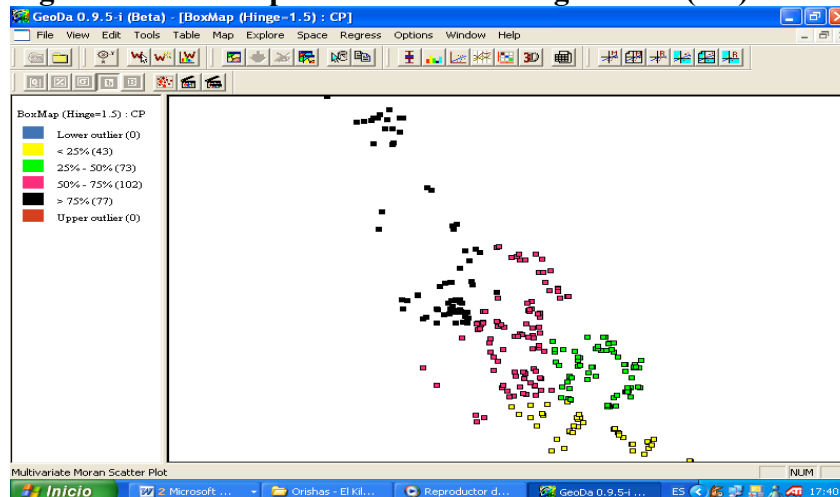
intercuartílico. De esta forma, se consideran como valores atípicos aquéllos situados por encima (o por debajo) de dichas cotas. Un criterio un poco más estricto que el anterior consistiría en multiplicar por tres el recorrido intercuartílico para la fijación de las cotas. El mapa de caja (box map), es una extensión del mapa cuartil junto con los outliers superiores e inferiores, definidos estos últimos como las observaciones fuera de los límites en un box plot (diagrama de caja). Comparando box maps para diferentes

variables se puede conseguir una idea inicial de asociaciones multivariantes en el sentido de que es probable que mapas con cuartiles (y outliers) coincidentes correspondan a variables correlacionadas. Este tipo de análisis puede constituir uno de los puntos de partida en la selección de las variables que deben incluirse en un modelo cuando no existen fundamentos teóricos fuertes a priori que nos oriente en la especificación del mismo.

Vamos a construir el box map (Map> Box Map) para las principales variables de nuestro estudio y así poder analizar los resultados:

Los box map y box plot para la superficie útil, la superficie construida con comunes, el código postal, el valor de tasación y el valor de tasación por metro cuadrado son los siguientes:

Figura 18: Box Map de la variable Código Postal (CP)

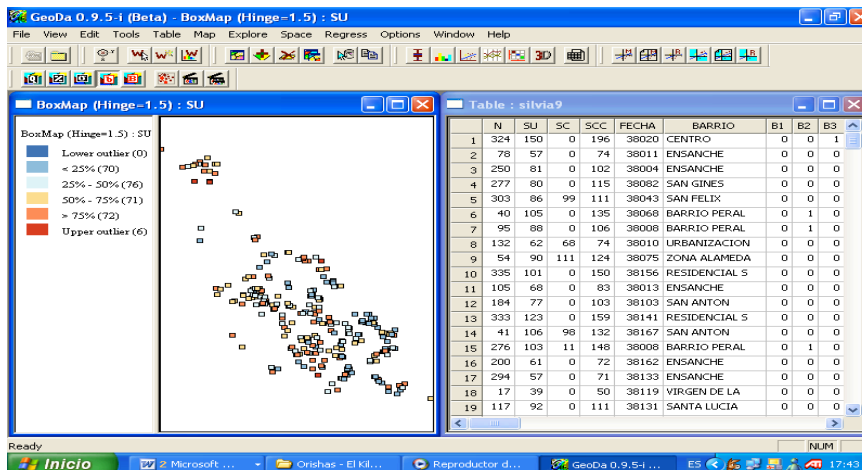


Fuente: elaboración propia a partir de Geoda

El box map de la variable código postal, identifica por colores zonas por el mismo color que corresponden al mismo código postal. De esta manera se puede comprobar que una gran parte de las viviendas se sitúan en la zona centro de Cartagena

En el box map de la superficie útil aparecen 6 outliers superiores vamos a ver a que viviendas corresponden, seleccionándolos en el mapa, aparecen marcados en la tabla:

Figura 19: Box Map de la variable Superficie Útil (SU)



Fuente: elaboración propia a partir de Geoda

Corresponden a las viviendas:

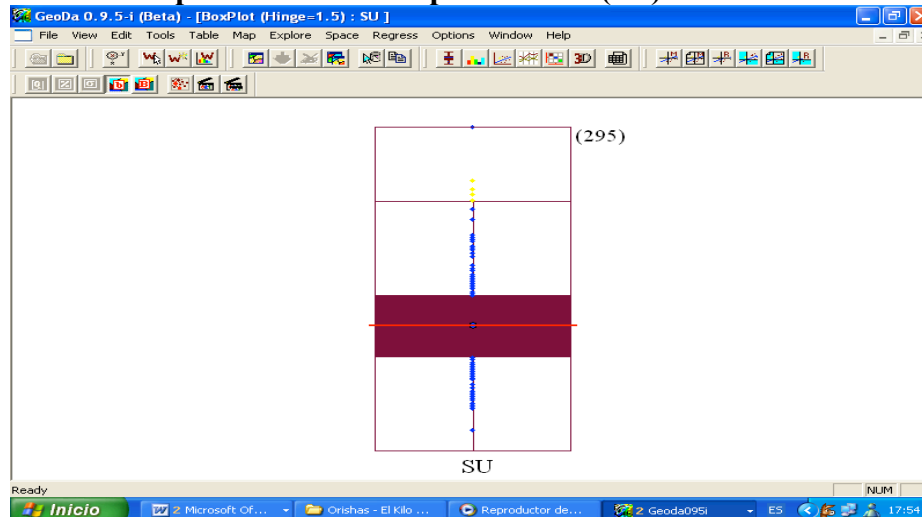
- 336, 337, 342 del Polígono Residencial de Santa Ana.
- 84, de San Félix.
- 175 de Ciudad Jardín.
- 316 del Centro de Cartagena.

Estos resultados son lógicos ya que tres de estas zonas (Ciudad Jardín, Polígono Residencial de Santa Ana y San Félix) son zonas residenciales situadas las dos últimas a las afueras de la ciudad de Cartagena.

En la zona Centro y San Antón, es lógico que encontremos viviendas grandes ya que se tratan de construcciones antiguas con mayor superficie que las viviendas de última construcción. Vamos a construir el box plot de la Superficie Útil, para comprobar que los 6 outliers del box map coinciden con los del box plot.

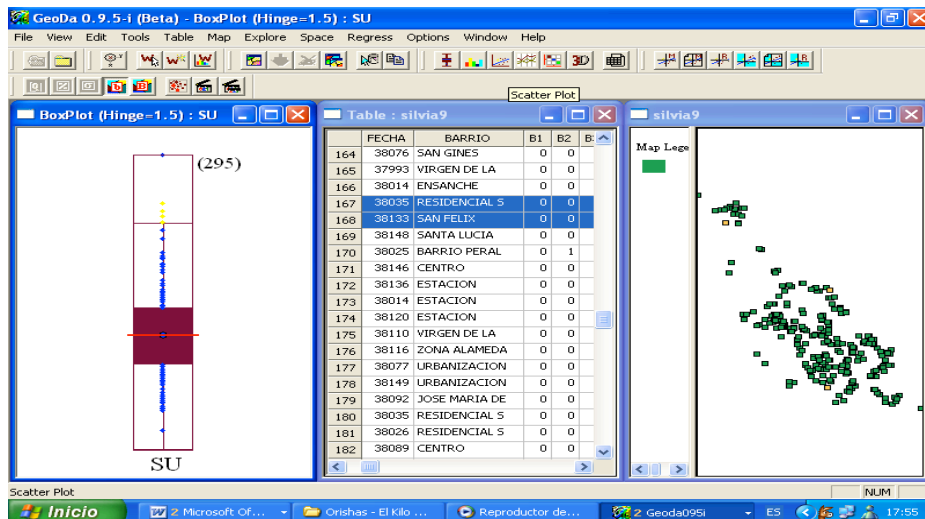
El box plot (Explore> Box plot) de la Superficie Útil sería:

Figura 20: Box Map de la variable Superficie Útil (SU)



Fuente: elaboración propia a partir de Geoda

Figura 21



Fuente: elaboración propia a partir de GeoDa

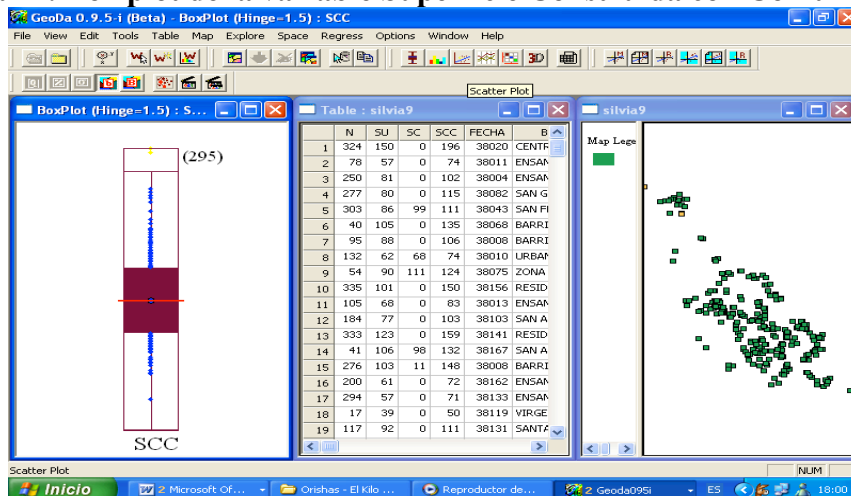
Efectivamente los 8 outliers superiores del box map coinciden con los el box plot de la superficie útil.

Para la superficie construida con comunes tenemos 3 outliers superiores que corresponden a las viviendas:

- 325 y 336 del Polígono Residencial de Santa Ana
- 175 de Ciudad Jardín

El box plot sería:

Figura 22: Box plot de la variable Superficie Construida con Comunes (SCC)

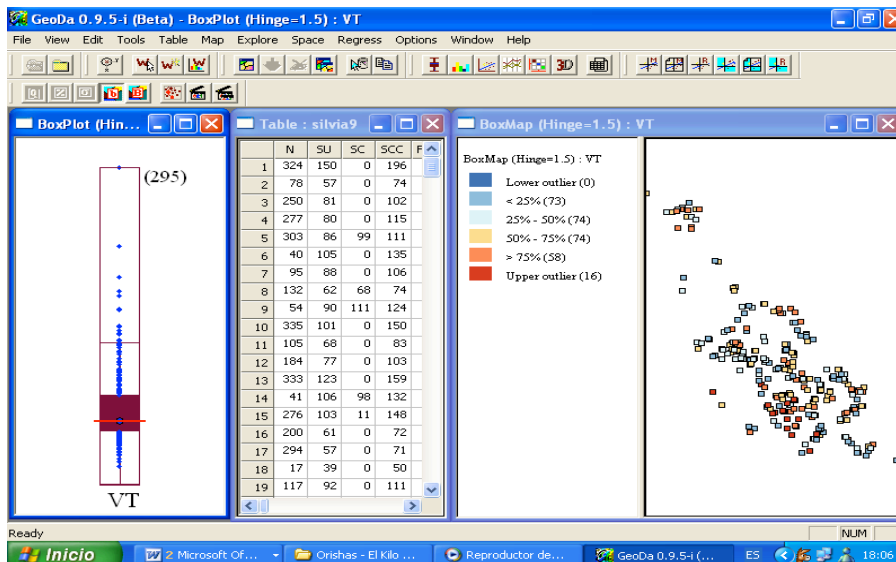


Fuente: elaboración propia a partir de GeoDa

De nuevo los outliers del box plot coinciden con los del box map, y muchos de ellos (viviendas 336 y, 175) con los de superficie útil ya que el concepto es parecido.

El box plot y el box map para el valor de tasación:

Figura 23: Box plot y Box map de la variable Valor de Tasación (VT)



Fuente: elaboración propia a partir de Geoda

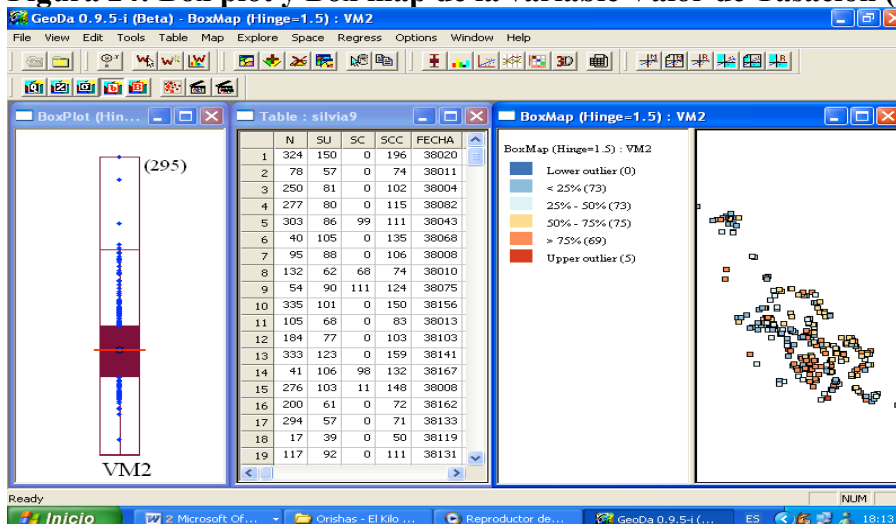
En este caso tenemos 16 outliers superiores que corresponden con las viviendas:

- 324, 316, 300, 307 y 309 en el Centro de Cartagena.
- 175 y 306 en Ciudad Jardín.
- 319, 321, 318, 314 y 317 en la Alameda de San Antón.
- 315 y 312 en el barrio del Ensanche.
- 47 en San Antón.
- 342 en el Polígono Residencial de Santa Ana.

Lo cual es lógico ya que se trata de zonas residenciales y exclusivas de la ciudad de Cartagena de alto valor.

Muchas de estas viviendas coincidirán con el siguiente caso que es el estudio del box plot y box map del valor de tasación por metro cuadrado:

Figura 24: Box plot y Box map de la variable Valor de Tasación (VT)



Fuente: elaboración propia a partir de Geoda

Existen 5 outliers superiores que corresponden a las viviendas:

- 324 y 322 en la zona Centro de Cartagena.
- 321 en la Alameda de San Antón

- 344 en el Polígono Residencial de Santa Ana
- 320 en el Residencial Puerta de Hierro

Igual que anteriormente las viviendas más caras por metro cuadrado en la ciudad de Cartagena se sitúan en la zona Centro y Alameda de San Antón lo que es lógico no sólo por el efecto Corte Inglés en el caso de la Alameda, sino porque se trata de zonas muy próximas altamente demandadas y con poca oferta de viviendas (poca oferta, alta demanda lo que da lugar a la subida del precio por metro cuadrado)

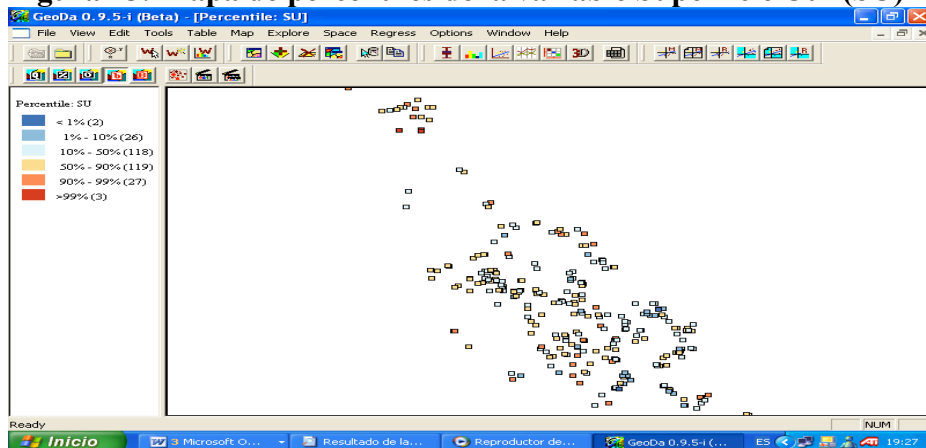
El Polígono Residencial de Santa Ana, Ciudad Jardín y el Residencial Puerta de Hierro son zonas exclusivas, residenciales a las afueras del núcleo urbano.

C) Mapas de percentiles

El mapa de percentiles es un caso particular del mapa de cuantiles, es decir, un mapa temático. Pero, a diferencia del mapa de cuantiles, este mapa permite la detección de puntos atípicos en la distribución espacial. Las categorías se diseñan de modo que queden acentuados los valores extremos de la variable. El programa GeoDa crea, en concreto, se crean 6 categorías correspondientes a los siguientes percentiles²: <1, [1,10), [10,50), [50,90), [90,99), >99.

Vamos a construir el mapa de percentiles de la variable Superficie útil (SU):

Figura 25: Mapa de percentiles de la variable Superficie Útil (SU)



Fuente: elaboración propia a partir de GeoDa

Como puede observarse, hay solamente 3 casas con valores atípicamente altos en cuanto a superficie útil, situadas las tres en el Polígono residencial de Santa Ana, mientras que sólo dos casas tienen un valor atípicamente bajo y se encuentran situadas en la zona Centro y en la barriada Virgen de la Caridad.

6.2.4.) Análisis exploratorio de la dependencia espacial local.

El estadístico I de Moran y el gráfico de dispersión de Moran son métodos que permiten analizar el fenómeno de dependencia o autocorrelación espacial desde un punto de vista global. Es decir, no son capaces de detectar la presencia de grupos o asociaciones de datos que se desvían de la tendencia general de una variable. Es cierto que el diagrama de dispersión de Moran es capaz de detectar ciertos aspectos locales de una distribución señalando aquellos puntos especialmente alejados de la media (superiores al valor 2). Pero no puede determinar sobre la existencia de especiales “concentraciones” o “ausencia de concentración” de valores altos/bajos de dicha variable. Para ello, se han

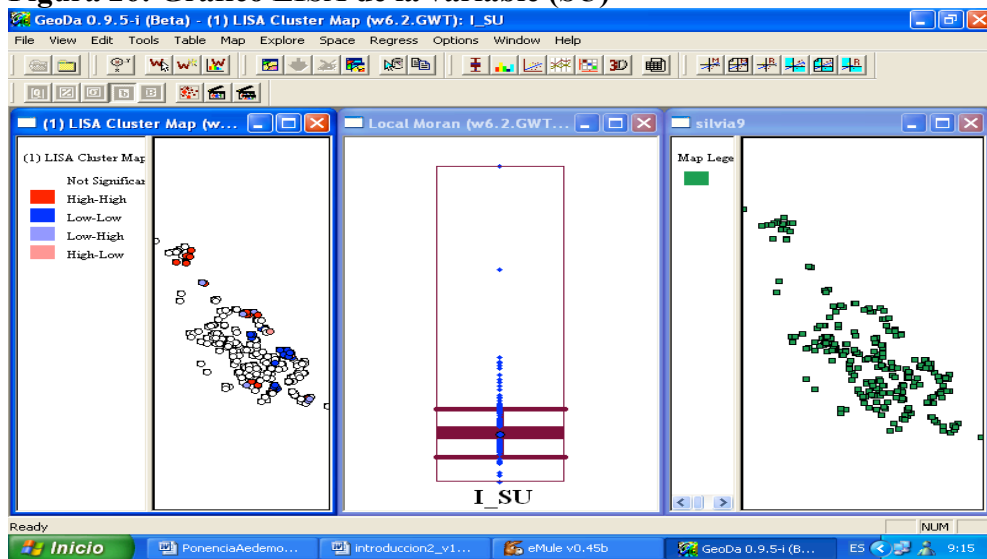
² El paréntesis cuadrado indica que ese extremo se incluye en el intervalo y el curvo indica que no se incluye.

diseñado los gráficos LISA de dependencia espacial local que ofrecen una indicación del grado de concentración de valores similares (altos o bajos) en torno a cada unidad geográfica.

Mapa LISA.

Las letras LISA significan “Local Indicator of Spatial Association”. El mapa LISA es un mapa en el que se representan las localizaciones con valores significativos en determinados indicadores estadísticos de asociación espacial local. En concreto, los gráficos LISA incluidos en GeoDa se basan en el estadístico I de Moran de asociación local. Se trata de un estadístico que, a diferencia del estadístico I de Moran anteriormente expuesto, no se calcula de forma global para todas las observaciones del mapa, sino que adquiere un valor diferente para todas y cada una de ellas. Efectivamente, este estadístico mide el grado de concentración de valores altos/bajos de una variable en el entorno geográfico de cada una de las observaciones de la muestra. Para cada valor del estadístico es posible realizar una inferencia para evaluar el nivel de significatividad estadística de rechazo de la hipótesis nula de ausencia de similitud o disimilitud de valores en una localización geográfica. De este modo, se pone de manifiesto la presencia de puntos calientes (“hot spots”) o atípicos espaciales, cuya mayor o menor intensidad dependerá de la significatividad asociada de los citados estadísticos.

Figura 26: Gráfico LISA de la variable (SU)



Fuente: elaboración propia a partir de GeoDa

Hemos representado el mapa LISA de agrupamientos espaciales (“spatial clusters”) de la superficie útil (Figura 26). Como puede observarse, a través de una gradación de colores, este mapa representa las zonas calientes de concentraciones de viviendas con alto valor (color rojo) y bajo valor (azul) en la superficie útil.

En color rojo tenemos viviendas situadas en zonas residenciales como el Polígono de Santa Ana, San Félix o Ciudad Jardín mientras que en color azul tenemos viviendas situadas en zonas de clase obrera (Santa Lucía o Virgen de la Caridad) o céntricas de la ciudad (Alameda de San Antón).

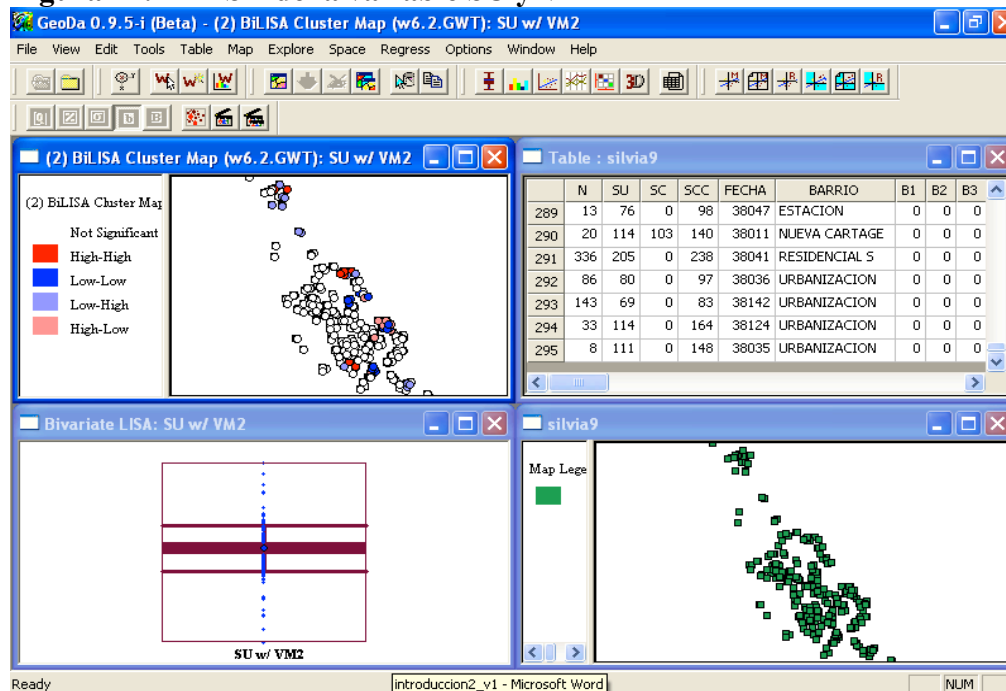
Gráficos LISA bivariantes.

Se trata de la versión multivariante de los gráficos LISA: el mapa y diagrama de dispersión LISA bivariantes. En el caso univariante, el estadístico LISA de asociación

espacial comparaba los valores de una variable en cada unidad geográfica con los valores de su correspondiente retardo espacial (variable con los valores medios de las unidades vecinas a una dada). De este modo, es posible determinar la existencia o no de concentraciones significativas de una variable en torno a una unidad geográfica. En el caso bivalente, el estadístico LISA tiene en cuenta para cada unidad geográfica, los valores de una variable y el retardo espacial de otra variable diferente. Por ejemplo, para

cada vivienda se calcula un estadístico LISA. Este estadístico mide el grado de asociación por ejemplo entre la superficie útil en unas viviendas y el valor de tasación por metro cuadrado medio de sus vecinas.

Figura 27: BiLISA de la variable SU y VM2



Fuente: elaboración propia a partir de GeoDa

De este modo el mapa LISA bivalente destaca, bastantes casos de asociación local positiva; alta concentración de valores altos (viviendas situadas en zonas céntricas o residenciales) y bajos (viviendas situadas en zonas periféricas de clase obrera).

- Anselin, L. (1.988). *Spatial econometrics: Methods and Models* Kluwer Academic Publishers
- Anselin, L. (1.995). "Local Indicators of Spatial Associations-LISA." *Geographical Analysis*, vol.27:93-115
- Cliff, A. D. y Ord, J. K. (1.981). *Spatial processes: Models and Applications* Pion Limited, London.
- Cressie, N. (1.993). *Statistics for spatial data* John Wiley, New York.
- Geary, R.C. (1.954) "The contiguity ratio and statistical mapping" *The Incorporated Statistician* 5,115-145
- Getis, A. y Ord J.K. (1.992) "The analysis of spatial association by use of distance statistics" *Geographical Analysis*, 24: 189-199.
- Griffith D.A. (1.988) *Advanced Spatial Statistics* Kluwer
- Haining R. (1.990). *Spatial data analysis in the social and environmental sciences* Cambridge University Press
- Knox E.G. (1964). "The detection of space-time interactions" *Applied Statistics* 13, 25-29
- LeSage, J.P. (1.997) "Bayesian Estimation of Spatial Autoregressive Models", *International Regional Science Review*, 1997 Volume 20, number 1&2, pp. 113-129.
- LeSage, J.P. (1.999). *Spatial Econometrics*. Department of Economics. University of Toledo.
- LeSage, J.P. (1.999). *The Theory and Practice of Spatial Econometrics*. Department of Economics. University of Toledo.
- López, Palacios (2.000) Distintos Modelos de Dependencia Espacial. Análisis de Autocorrelación. *XIV Reunión Asepelt España, Oviedo 22-24, Junio 2.000*.
- López, Palacios y Ruiz (2.001) Distintos modelos explicativos del desempleo en términos de localización. Una aplicación a las provincias españolas. *XXVII Reunión de Estudios Regionales, Madrid 28-30, Noviembre 2.001*.
- Moran, P.A.P. (1.948) "The interpretation of statistical maps" *Journal of the Royal Statistical Society B*, v.10, 243-251.
- Ord, J.K. y Getis, A (1.995) "Local spatial autocorrelation statistics: distributional issues and an application" *Geographical Analysis*, 27: 286-296.
- Ripley, B.D. (1.981). *Spatial Statistics* Wiley, New York.
- Vayá Valcarce E. y Moreno Serrano R. (2.000). La Utilidad de la Econometría Espacial en el Ámbito de la Ciencia Regional. *Documento de Trabajo 2000-13*.