

# **Análisis de Autocorrelación Espacial- Precios de ventas de casas de Baltimore, Maryland 1978**



**UNIVERSIDAD DISTRITAL  
FRANCISCO JOSÉ DE CALDAS**

Juan David Gonzalez Hernandez - 20151025060  
Raúl Camilo Martín Bernal - 20151025909

Geoestadística

Ingeniería Catastral y Geodesia

Universidad Distrital Francisco José de Caldas

*Presentado a* Carlos Eduardo Melo Martínez

Bogotá D.C. agosto de 2020

# Tabla de contenidos

<b>1. Introducción</b>	<b>1</b>
<b>2. Justificación</b>	<b>2</b>
<b>3. Objetivos</b>	<b>3</b>
3.1. Objetivo general . . . . .	3
3.2. Objetivos específicos . . . . .	3
<b>4. Estado del Arte</b>	<b>4</b>
<b>5. Marco teórico</b>	<b>10</b>
5.1. Polígonos de Thiessen y Diagramas de Voronói . . . . .	10
5.2. Econometría espacial . . . . .	10
5.2.1. Heterogeneidad Espacial . . . . .	10
5.2.2. Autocorrelación Espacial . . . . .	11
5.3. Análisis exploratorio de datos espaciales (AEDE) . . . . .	11
5.4. Interacciones espaciales . . . . .	12
5.5. Matriz de pesos espaciales . . . . .	12
5.5.1. Estandarización de la matriz W . . . . .	13
5.6. Estadísticos globales de autocorrelación espacial . . . . .	13
5.6.1. I de Moran Global . . . . .	13
5.6.2. C de Geary . . . . .	14
5.6.3. G de Getis-Ord . . . . .	14
5.7. Estadísticos locales de autocorrelación espacial . . . . .	15
5.7.1. Indicador local de asociación espacial (LISA) . . . . .	15
5.7.2. G de Getis y Ord Local . . . . .	16
5.8. Análisis confirmatorio de datos espaciales . . . . .	16
5.8.1. Modelos . . . . .	16
5.8.2. Regresión geográficamente ponderada GWR . . . . .	19
<b>6. Descripción base de datos</b>	<b>20</b>
<b>7. Análisis descriptivo de datos espaciales</b>	<b>22</b>
7.1. Área de estudio . . . . .	22
7.2. Polígonos de Thiessen y Voronói . . . . .	23
7.3. Estadísticas descriptivas . . . . .	23

7.3.1. Variable dependiente . . . . .	23
7.3.2. Variables independientes . . . . .	24
7.4. Matriz W de pesos espaciales . . . . .	27
<b>8. Análisis exploratorio de datos espaciales</b>	<b>30</b>
8.1. Análisis Global . . . . .	31
8.1.1. I de Moran Global variables independientes . . . . .	32
8.1.2. I de Moran Bivariado . . . . .	32
8.1.3. Correlogramas bivariados . . . . .	35
8.2. Análisis Local . . . . .	37
8.2.1. I de Moran Local . . . . .	44
<b>9. Análisis confirmatorio</b>	<b>45</b>
9.1. Modelo Clásico . . . . .	45
9.1.1. Pruebas de los supuestos modelo 2 . . . . .	46
9.1.2. Mapa de residuos . . . . .	47
9.1.3. Multiplicadores de Lagrange . . . . .	47
9.2. Modelo Spatial Lag . . . . .	48
9.3. Modelo Spatial Error . . . . .	48
9.4. Modelo SLX . . . . .	49
9.5. Modelo Spatial Durbin . . . . .	50
9.6. Modelo GLM . . . . .	51
9.7. Modelo SARAR . . . . .	52
9.8. Modelo espacial lineal heterocedastico SARAR . . . . .	53
9.9. Modelo HAC espacial . . . . .	54
<b>10. Análisis de impactos del modelo</b>	<b>56</b>
<b>11. Análisis del modelo</b>	<b>59</b>
<b>12. GWR</b>	<b>60</b>
<b>13. Conclusiones</b>	<b>63</b>
13.1. Anexos . . . . .	66

# Índice de figuras

4.1. Descripción de las muestras utilizadas en Infraestructura pública y precios de vivienda . . . .	5
4.2. Resultado de los Modelos utilizados en Milwaukee . . . . .	8
4.3. Evaluación de los modelos en Milwaukee . . . . .	8
7.1. Ciudad de Baltimore, Maryland,USA . . . . .	22
7.2. Mapa polígonos creados a partir de los puntos . . . . .	23
7.3. Gráficos estadísticos para la variable PRICE . . . . .	23
7.4. Histogramas: Numero de habitaciones, numero de baños y numero de historias . . . . .	24
7.5. Histogramas: Garaje, edad y tamaño de lote . . . . .	25
7.6. Histogramas: Espacio interior,DWELL, patio . . . . .	25
7.7. Histogramas: Chimenea,aire acondicionado,sotano,localización . . . . .	26
7.8. Mapa vecinos por criterios gráficos . . . . .	28
7.9. Correlograma de Moran de la variable dependiente con la matriz de pesos Delaunay . . . . .	29
8.1. Mapa de cuantiles simples para los precios de vivienda . . . . .	30
8.2. I de moran variable dependiente PRICE . . . . .	31
8.3. Dispersograma bivariado de Moran de PRICE con NROOM, DWELL,NBATH . . . . .	33
8.4. Dispersograma bivariado de Moran de PRICE con PATIO,FIREPL,AC . . . . .	34
8.5. Dispersograma bivariado de Moran de PRICE con BMENT,NSTOR,GAR . . . . .	34
8.6. Dispersograma bivariado de Moran de PRICE con AGE,CITCOU,LOTSZ,SQFT . . . . .	34
8.7. Correlogramas bivariado de Moran 9 variables independientes . . . . .	36
8.8. Correlogramas bivariado de Moran 3 variables independientes . . . . .	37
8.9. LISA PRICE . . . . .	38
8.10. LISA NROOM . . . . .	38
8.11. LISA DWELL . . . . .	39
8.12. LISA NBATH . . . . .	39
8.13. LISA PATIO . . . . .	39
8.14. LISA FIREPL . . . . .	40
8.15. LISA AC . . . . .	40
8.16. LISA BMENT . . . . .	41
8.17. LISA NSTOR . . . . .	41
8.18. LISA GAR . . . . .	42
8.19. LISA AGE . . . . .	42
8.20. LISA CITCOU . . . . .	43
8.21. LISA LOTSZ . . . . .	43

8.22. LISA SQFT . . . . .	43
9.1. Mapa de residuos modelo clásico 2 . . . . .	47
9.2. Gráfico determinación Heterocedasticidad . . . . .	55
10.1. Punto 16, seleccionado para realizar el análisis de impactos . . . . .	56
10.2. Impactos . . . . .	57
10.3. Efectos de los impactos . . . . .	57
12.1. R2 local y coeficiente de intercepto . . . . .	60
12.2. Variación de los coeficientes en el espacio para las variables DWELL, NBATH, PATIO . . . . .	60
12.3. Variación de los coeficientes en el espacio para las variables FIREPL, AC, BMENT . . . . .	61
12.4. Variación de los coeficientes en el espacio para las variables GAR, CITCOU, LOTSZ . . . . .	61
12.5. Significancia estadística para las variables DWELL, NBATH, PATIO . . . . .	61
12.6. Significancia estadística para las variables FIREPL, AC, BMENT . . . . .	62
12.7. Significancia estadística para las variables GAR, CITCOU, LOTSZ . . . . .	62
12.8. Predicción de valores, error estándar y R2 local . . . . .	62

# Índice de cuadros

4.1. Variables utilizadas en Infraestructura pública y precios de vivienda . . . . .	5
4.2. Variables utilizadas para el modelado espacialmente de precios de vivienda en Milwaukee . .	8
6.1. Tabla de descripción de las variables . . . . .	20
7.1. Resultados evaluación distintas matrices de pesos . . . . .	28
8.1. Test G de Getis . . . . .	31
8.2. I de moran para las variables independientes . . . . .	32
8.3. Resultados I de Moran Bivariado para todas las variables . . . . .	33
8.4. I de moran variable dependiente PRICE . . . . .	44
9.1. Resultados modelos clásicos . . . . .	45
9.2. Pruebas de supuestos Modelo clásico 2 . . . . .	46
9.3. Multiplicadores de lagrange modelo clásico 2 . . . . .	47
9.4. Modelo spatial lag . . . . .	48
9.5. Supuestos modelo spatial lag . . . . .	48
9.6. Modelo Spatial Error . . . . .	48
9.7. Supuestos modelo spatial error . . . . .	48
9.8. Modelo SLX . . . . .	49
9.9. Supuestos modelo SLX . . . . .	49
9.10. Modelo Durbin espacial . . . . .	50
9.11. Supuestos modelo Durbin espacial . . . . .	50
9.12. Modelo GLM . . . . .	51
9.13. Supuestos modelo GLM . . . . .	51
9.14. Modelo SARAR . . . . .	52
9.15. Supuestos modelo SARAR . . . . .	52
9.16. Modelo SARAR . . . . .	53
9.17. Resultados Normalidad . . . . .	53
9.18. Modelo espacial HAC . . . . .	54
9.19. Resultados Normalidad . . . . .	54
9.20. Resultados multicolinealidad . . . . .	55
10.1. Viviendas mas afectadas, análisis de impactos . . . . .	57
10.2. Tabla de impactos modelo spatial HAC para el precio de la vivienda . . . . .	58

# 1. Introducción

El comportamiento del sector inmobiliario ha sido ampliamente estudiado a lo largo de los años de diferentes maneras y especialmente con la implementación de los modelos de precios hedónicos en el área de precios de la vivienda en donde se consideran como determinantes del precio de la vivienda las características de la misma. Esto hace particularmente importante que los problemas de especificación del modelo estén relacionados con la omisión de variables relevantes o ligadas a una fuerte correlación entre las mismas. Por tal motivo autores como Basu y Thibodeu (1998) discriminaron las características de las viviendas en siete ítems: asociadas al lote, asociadas a las mejoras, la accesibilidad, la proximidad de externalidades, uso de la tierra y época en la que se recolectan los datos. Todo con el fin de evitar sesgos de especificación u omisión de variables relevantes. Esto se desarrolló frecuentemente aplicando la metodología econométrica tradicional y no fue hasta los años 80 donde se comenzaron a ver algunos vestigios de la implementación de análisis espaciales a estos análisis de regresión. A raíz de esto que se quiera analizar para una muestra del mercado inmobiliario de la ciudad de Baltimore del año 1978 como se puede ver afectado el precio de la vivienda por las características de la misma e incluyendo los componentes de asociación espacial que son naturales en las viviendas.

## 2. Justificación

La metodología econométrica tradicional utilizada comúnmente no tiene en cuenta la correlación espacial de los datos, es decir, se omite el espacio como fuente de variabilidad de los datos. Y en segundo lugar, no tiene en cuenta la heterogeneidad espacial de los datos, lo que implica que los valores de los parámetros puedan variar en el espacio Anselin y Bera (1998). Tener presente esto es de vital importancia ya que el mercado inmobiliario por si mismo es de naturaleza espacial y es valido decir esto ya que "las viviendas de mayor y menor precio tienden a estar aglomeradas en lugares específicosz también porque algunas las características de las viviendas pueden ser mayormente relevantes en zonas con un nivel económico alto como lo puede ser un garaje o un patio.



## 3. Objetivos

### 3.1. Objetivo general

Realizar el análisis exploratorio y confirmatorio de los datos asociados al precio de la vivienda en la ciudad de Baltimore para el año 1978, y poder elegir un modelo el cual explique de la mejor manera posible la variable de precio teniendo en cuenta toda la teoría econométrica y espacial para la estimación esperada, y los supuestos que deben cumplir un modelo para inferir de manera correcta en los datos

### 3.2. Objetivos específicos

1. Analizar la base de datos para poder ver el estado de la misma y poder dar suposiciones acertadas de que variables van a ser las mas significativas en la estimación del modelo
2. Comprobar la existencia de correlación espacial entre el precio de la vivienda y sus determinantes mediante pruebas globales y locales.
3. Estimar y comparar los diferentes modelos espaciales que explican el valor de la vivienda y seleccionar el que mejor se ajuste a los datos.
4. Realizar un análisis de impactos en el cual podamos ver como un cambio en un registro de la base de datos puede afectar los demás valores de esta.

## 4. Estado del Arte

### **Infraestructura pública y precios de vivienda: una aplicación de regresión geográficamente ponderada en el contexto de precios hedónicos.**Duque, Ceballos y Agudelo (2011)

Hasta el momento los estudios con la técnica GWR en el área de precios de la vivienda han sido pocos y, en términos generales, se han utilizado en la especificación de las variables intrínsecas de la vivienda, como el área, número de habitaciones, etc. Así como variables dummy con el fin de verificar la relación existente entre el medio que rodea la vivienda y su precio, obteniendo resultados significativos en la mayoría de ellos.

Por ello presentar evidencias de que algunas externalidades como el desarrollo de obras públicas en zonas menos favorecidas de las ciudades como lo es la comuna de San Javier en la ciudad de Medellín puede contribuir a mejorar el bienestar de sus habitantes a través de un aumento de sus patrimonio expresado en términos del aumento del precio de sus viviendas. Además también clarificar los cobros por valorización que realizan las administraciones locales cuando se ejecuta una intervención urbanística o se desarrolla una obra pública.

Para tal fin se utilizaron 98 datos de mercado en los alrededores de la estación del metro de San Javier, localizados en 11 barrios realizados por 4 peritos entre los años 2008 y 2009 esta muestra representa el 0.41 % del total de viviendas de barrios estudiados. Las variables utilizadas y la descripción de estas fueron:

Variable	Definicion
Precio	Avaluo del inmueble en millones de pesos
Tipo	casa o apartamento
Area construida	Area construida de la vivienda en m2
Area de parqueadero	Area medida en m2
Numero de baños	Cantidad de baños de la vivienda
Distancia	Distancia entre la vivienda y la estacion del metro en (Km)
Edad	Edad de la construcción
Edad2	Indica el cuadrado de la edad de construcción
Estrato 2	1 si es estrato 2, 0 de lo contrario
Estrato 3	1 si es estrato 3, 0 de lo contrario
Características	Número de beneficios que tiene una vivienda
Unidad	Indica si el inmueble se localiza en una unidad cerrada
fecha	Fecha en que se realizó el avalúo se cuenta como el numero de meses hacia atrás a partir del ultimo avaluo registrado

Cuadro 4.1: Variables utilizadas en Infraestructura pública y precios de vivienda

	Máximo	Mínimo	Promedio
<b>Valor total</b>	198,420,000	18,988,188	78,612,317
<b>Edad (años)</b>	60	1	21
<b>Construcción (m2)</b>	238	35	104
<b>Estrato</b>	4	2	
<b>No. Baños</b>	4	1	2
<b>Características</b>	6	0	
<b>Distancia (km)</b>	1.845	0.0989	0.778
<b>Parqueadero (m2)</b>	14	0	0.4

Figura 4.1: Descripción de las muestras utilizadas en Infraestructura pública y precios de vivienda

Para el desarrollo se propusieron a utilizar los siguientes modelos:

#### **Estimación por mínimos cuadrados ordinarios OLS:**

El modelo ajustado por OLS explica cerca del 83 % de la variación de los precios de las viviendas además se corrobora la hipótesis de influencia de la estación del metro en los valores de las viviendas. En donde las variable tipo de inmueble no era significativa y las variables estrato2, estrato3, distancia, construcción y baño eran las únicas significativas al 95 %

Para los modelos espaciales se utilizó una matriz W con el método K-vecinos más cercanos igual 4 asumiendo ese como el número de viviendas con características similares en la zona

#### **Modelo autorregresivo espacial de orden uno SAR(1):**

Este modelo presenta un mejor ajuste respecto al modelo OLS. En donde las variables fecha, construcción, parqueadero, distancia, estrato2, estrato3 y baños son significativas al 99 % mientras que la variable características lo es al 90 %

#### **Modelo de error espacial (SEM):**

Los resultados muestran que las variables construcción, parqueadero, distancia, estrato2, estrato3, baños, y el coeficiente lambda que contiene el retardo espacial son significativos al 99 %, fecha y edad lo son al 95 % y edad2 lo es al 90 %

**Modelo GWR:** El modelo GWR presenta el coeficiente de determinación mayor, entre todos los modelos estimados, explica el 92,3 % de variación del precio de las viviendas. El criterio de información de akaike y el ancho de banda luego de seis iteraciones fueron de 3231.515 y 61.026 respectivamente.

Con el fin de analizar diferencias entre la estimación GWR y el método OLS se estimaron los precios de una vivienda tipo localizada en diferentes sectores del área de estudio. De este análisis se concluyó que los precios obtenidos por GWR tienden a variar más en el espacio que los obtenidos por OLS dado que con la regresión geográficamente ponderada los parámetros cambian continuamente en el espacio ajustándose a características como el estrato y la edad, para esto se utilizó la diferencia de precios en dos viviendas a una distancia de 0.9km y 0.91 km de la estación que según GWR se obtuvo una diferencia de precios de \$ 44,940,588 y con OLS una diferencia de \$ 205,478 en cuyo caso los precios obtenidos mediante GWR se acercaban más a la realidad de las viviendas circundantes, es decir a sus respectivos avalúos.

Finalmente del trabajo concluyen la variable distancia entre vivienda y estación resultó ser significativa y muestra una influencia positiva de la estación en los precios de las viviendas localizadas a un radio de 600 metros y que además para determinar la influencia de la estación del metro el modelo GWR se posiciona como el más apropiado gracias a su capacidad de permitir que los coeficientes del modelo hedónico varíen en magnitud y nivel de significancia dependiendo de la localización de la vivienda que se este valorando.

**Modeling spatial dimensions of housing prices in Milwaukee, Wisconsin** Yu, Wei y Wu (2007)

En este estudio se analizaron las dimensiones espaciales del mercado inmobiliario en la ciudad de Milwaukee modelando los diferentes determinantes mediante precios hedónicos, además de los atributos de viviendas convencionales se agregó información de teledetección que corresponde a las fracciones del suelo y la superficie impermeable que representan las condiciones degradadas ambientales de los vecindarios para mejorar el modelo. Se plantearon lograr 3 objetivos haciendo uso de la base de datos “master property” de la ciudad. Primero se examinó la influencia de la autocorrelación espacial en el modelado del mercado inmobiliario haciendo uso de dos tipos de modelos autoregresivos espaciales. Segundo comprender las relaciones de las variables espaciales entre los atributos de la vivienda y los precios de la misma haciendo uso de la regresión geográficamente ponderada (GWR). La tercera parte corresponde la evaluación de los modelos que incorporan información espacial haciendo uso de un conjunto de datos diferente a los usados para la construcción del modelo.

En este estudio se utilizaron datos de la ciudad de Milwaukee en Wisconsin que se encuentra localizada en la costa oeste del lago Michigan en Estados Unidos. Los dos tipos de datos seleccionados para el estudio fueron. Las características y/o atributos estructurales de las viviendas seleccionados de la base de datos del año 2003 “master property” (MPROP) la cual tiene alrededor de 160000 registros de las propiedades dentro del límite de la ciudad, esta base de datos también incluía datos de localización, aunque el estudio solo se quería realizar con los 68728 registros que correspondían a las viviendas unifamiliares ocupadas por sus propietarios los costes computacionales para ese tiempo eran algo restrictivos, finalmente se seleccionaron 2 muestras aleatorias que cubrían la ciudad cada una con 1821 registros una para generar los modelos y otra para evaluarlos. Esta selección tuvo una confianza al 99 % y permitió que las medias de las muestras se encontraran en un error marginal del 3 % de la población. Después de un análisis preliminar trabajaron con las siguientes variables:

El segundo tipo de dato utilizado fue generado a partir de una imagen satelital Landsat ETM+ adquirida el 9 de julio de 2001 que representa las condiciones de deterioro ambiental de los vecindarios, esta se halló utilizando el método de análisis de mezcla espectral normalizado, el impacto de esta variable debe ser negativo.

Los modelos que se construyeron a partir de estos datos fueron OLS, unos modelos espaciales autorregresivos y una regresión geográficamente ponderada GWR. Para los modelos espaciales se hizo uso de una matriz de pesos a partir de una distancia crítica de 2.5km cuyos resultados y evaluación se muestran en la siguientes tablas.

Variable	Significado	impacto esperado
Price	precio de la vivienda	
AirCd	si tiene o no aire acondicionado	positivo
FirePlc	numero de chimeneas	positivo
Nofbath	numero de baños	positivo
HsAge	Edad de la vivienda	negativo
FISize	tamaño del piso	positivo

Cuadro 4.2: Variables utilizadas para el modelado espacialmente de precios de vivienda en Milwaukee

**Table 3.** Modeling results for the ordinary least squares (OLS) and spatial autoregressive models (on modeling samples).

	Estimate <sup>a</sup>	Standard error	t/z value	Pr(>  t / z )
<i>Ordinary least squares</i>				
(Intercept)	8.668	0.243	35.580	0.000
Floor size	0.565	0.036	15.792	0.000
House age	-0.279	0.0212	-13.167	0.000
Fireplace	0.149	0.1186	8.029	0.000
Air conditioner	0.170	0.016	10.877	0.000
Number of bathrooms	0.158	0.032	4.894	0.000
Soil and impervious surface	-0.367	0.066	-5.558	0.003
$F_{6,1814} = 234.1$ , $p$ -value = 0.000				
Log likelihood = -478.937				
<i>Substantive spatial autoregressive model (weight matrix constructed on critical distance = 2.5 km)</i>				
(Intercept)	-3.051	0.138	-22.13	0.000
Floor size	0.554 (98.1)	0.019	28.500	0.000
House age	-0.163 (58.4)	0.012	-13.820	0.000
Fireplace	0.074 (49.7)	0.010	7.412	0.000
Air conditioner	0.069 (40.6)	0.008	8.230	0.000
Number of bathrooms	0.086 (54.4)	0.017	4.968	0.000
Soil and impervious surface	-0.110 (30.0)	0.035	-3.101	0.002
$\rho = 0.980$ , LR test value = 2216.10, $p$ -value = 0.000				
Log likelihood = 629.136				
<i>Nuisance spatial autoregressive model (weight matrix constructed on critical distance = 2.5 km)</i>				
(Intercept)	8.977	1.359	6.605	0.000
Floor size	0.630 (111.5)	0.019	32.584	0.000
House age	-0.255 (91.4)	0.014	-17.706	0.000
Fireplace	0.067 (45.0)	0.010	6.991	0.000
Air conditioner	0.059 (34.7)	0.008	7.155	0.000
Number of bathrooms	0.100 (63.3)	0.016	6.062	0.000
Soil and impervious surface	-0.119 (32.4)	0.035	-3.364	0.001
$\lambda = 0.997$ , LR test value = 2395, $p$ -value = 0.000				
Log likelihood = 718.579				

<sup>a</sup> Numbers in parentheses are the percentages of the corresponding autoregressive model estimated coefficients when compared with those of the OLS.

Figura 4.2: Resultado de los Modelos utilizados en Milwaukee

	OLS	SA(lag)	SA(terr)	GWR
Adjusted $R^2$	0.435	—	—	0.923
<i>Modeling samples</i>				
AIC	973.87	-1240.30	-1419.20	-1869.31
RMSE	0.00738	0.00397	0.00377	0.00272
RE	0.774	0.383	0.358	0.250
<i>Testing samples</i>				
RMSE	0.00734	0.00403	0.0233	0.00332
RE	0.766	0.375	2.932	0.296

Figura 4.3: Evaluación de los modelos en Milwaukee

A partir de estos resultados se concluyó que tanto el modelo OLS como los autoregresivos espaciales indican que todos los atributos de la vivienda y vecindario están significativamente relacionados con el precio de la misma. De manera intuitiva los resultados de los modelos sugieren que las casas más grandes, construidas recientemente, con condiciones ambientales de los vecindarios favorables, que tienen chimeneas, más baños y aires acondicionados tienden a ser aquellas con precios más elevados. Las pruebas para  $\lambda$  y  $\rho$  de los autoregresivos espaciales indican que existe una autocorrelación espacial significativa entre los precios de la vivienda y los términos del error. Una diferencia sustancial entre el modelo OLS y los espaciales es que el modelo OLS tiende a sobreestimar la mayoría de los parámetros cerca del 8,6 % al 70 % y es probable que esta sobreestimación se deba a la existencia de autocorrelación espacial entre las viviendas vecinas que no se tiene en cuenta en este modelo. Cuando esta información espacial se ha incluido explícitamente en los modelos espaciales autorregresivos mediante la matriz W las estimaciones tienden ser más bajas en comparación al modelo OLS y esto representa mejor el valor real.

De los estadísticos AIC se determinó que todos los modelos espaciales autorregresivos y el modelo GWR son mucho mejor que el modelo OLS. Entre los tres modelos espaciales se observó que el modelo GWR se ajustó mejor a los datos. Sin embargo las pruebas pseudo-significativas del modelo GWR en coeficientes locales revelan que no todos los atributos de la vivienda están significativamente relacionados con los valores de la vivienda en todas partes. De hecho, excepto por las variables tamaño de piso y edad, los cuatro atributos estructurales y de vecindario son significativos en áreas específicas.

. Finalmente el documento concluye tres cosas, la primera que cuando se ignora la información espacial al establecer el modelo hedónico de precios, el modelo tiende a sobreestimar los atributos estructurales y que esto se debe a la existencia inamovible de autocorrelación espacial entre las casas vecinas pasando por alto importantes atributos de localización, segundo que al hacer uso del modelo GWR se encontraron que las relaciones entre los valores de vivienda y los atributos no son invariantes sobre el espacio, además que no es necesario que todos los atributos de la vivienda se relacionen de manera significativa en todas las partes de estudio. Tercero en términos de precisión predictiva, en general, los modelos espaciales funcionan mejor. Sin embargo aunque el modelo GWR funciona mejor con las muestras de modelado su precisión de predicción cae más rápido cuando se alimentan las muestras de la prueba uno de los factores que podrían contribuir a tal caída en la precisión de predicción es que usaron una interpolación como medio para la predicción.

## 5. Marco teórico

### 5.1. Polígonos de Thiessen y Diagramas de Voronói

Los polígonos de Thiessen nombrados por el meteorólogo estadounidense Alfred H. Thiessen también estudiados por Georgy Voronoy (también llamados diagramas de Voronói) y por el Matemático Gustav Lejeune Dirichlet (también llamado teselación de Dirichlet) son uno de los métodos de interpolación más simples, basados en la distancia euclidiana, siendo especialmente apropiada cuando los datos son cualitativos. Se crean al unir los puntos entre sí, trazando las mediatrices de los segmentos de unión. Las intersecciones de estas mediatrices determinan una serie de polígonos en un espacio bidimensional alrededor de un conjunto de puntos de control, de manera que el perímetro de los polígonos generados sea equidistante a los puntos vecinos y designando su área de influencia A. (2020).

### 5.2. Econometría espacial

La econometría espacial es una parte de la econometría que estudia los fenómenos económicos espaciales, aunque todas las actividades económicas se desarrollan en un espacio determinado, los fenómenos económicos espaciales se pueden definir como aquellos en los que la variable espacio juega un papel importante en los modelos, un papel en el cual si no se llega a tomar en cuenta para el modelo darían lugar a modelos econométricos con grandes errores. «Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC» (s.f.)

Anselin la define como "La colección de técnicas que tratan las peculiaridades causadas por el espacio en el análisis estadístico de los modelos tradicionales de la ciencia regional Moreno y Vayá (2002).

Al tomar la variable de espacio y con la naturaleza de los datos de corte transversal suelen aparecer los denominados efectos especiales: la heterogeneidad y la dependencia espacial.

#### 5.2.1. Heterogeneidad Espacial

Este efecto aparece cuando se utilizan datos de unidades espaciales muy distintas para explicar un mismo fenómeno, esto hace que aparezcan problemas como la heterocedasticidad e inestabilidad estructural, los



cuales pueden ser resueltos mediante técnicas econométricas. Consiste en la variación de las relaciones en el espacio, hay dos aspectos de heterogeneidad espacial: la inestabilidad estructural y la heterocedasticidad. El primero se refiere a falta de estabilidad en el espacio del comportamiento de la variable en estudio, que lleva a que las variables varíen con respecto a la localización por lo tanto no son homogéneas en toda la muestra. El segundo aspecto, la heterocedasticidad proviene de la omisión de variables u otros errores que llevan a la aparición de errores en la medida Moreno y Vayá (2002).

### **5.2.2. Autocorrelación Espacial**

La dependencia o autocorrelación espacial surge cuando el valor de una variable en un lugar del espacio está relacionado con su valor en otro u otros lugares del espacio, debido a la multidireccionalidad que domina las relaciones entre las unidades espaciales no es posible tratarlo con la econometría estándar.

Aparece como consecuencia de la existencia de relación funcional entre lo que ocurre en dos o mas lugares en el espacio. Esta puede ser positiva o negativa. Cuando se ve concentración de la presencia de un fenómeno en una región que se extiende sobre las regiones que la rodean es un caso de autocorrelación positiva, si por el contrario la aparición de un fenómeno en una región impide o dificulte su aparición en las regiones que la rodean es un caso de autocorrelación negativa. Por ultimo si vemos que la variable se distribuye de manera aleatoria en el espacio no existe autocorrelación espacial Moreno y Vayá (2002).

### **5.3. Análisis exploratorio de datos espaciales (AEDE)**

Para estudios de economía regional urbana y en aquellos casos de estudio a temas relativos a la ubicación y desigualdad regional, es habitual hacer un análisis descriptivo de la distribución en el espacio de las variables objeto de estudio a través de la representación de un mapa de esta variable, complementándose con el calculo de los diversos índices, dichas técnicas presentan importantes limitaciones, como que los resultados son muy subjetivos y dependen de los intervalos seleccionados para visualizar el valor de la variable, además de que datos diferentes en su distribución y sus patrones pueden suministrar valores similares de los índices, por esto se justifica realizar un análisis complementario como el análisis exploratorio de datos espaciales (AEDE), centrado en los efectos espaciales y con un conjunto de técnicas que permiten describir distribuciones espaciales como localizaciones atípicas (outliers espaciales) y asociación espacial (cluster espaciales).

## 5.4. Interacciones espaciales

De un modo amplio la interacción espacial puede ser definida “como las relaciones de movimiento de personas, bienes y objetos y comunicación de ideas e información que tienen lugar en el espacio como un complejo proceso de toma de decisiones”(Stewart Fotheringham). Cuando se habla de autocorrelación espacial, es necesario definir cuáles son los espacios o regiones que tienen una relación espacial, por lo tanto, se establece una matriz de interacciones espaciales. Lo anterior se encuentra estrechamente relacionado con la primera ley de la geografía de Tobler que dice cualquier cosa está relacionada con cualquier otra, pero las cosas cercanas están más relacionadas que las distantes”.

## 5.5. Matriz de pesos espaciales

Es un arreglo matemático que define las posibles interrelaciones entre los elementos que se encuentran en un sistema espacial estableciendo relaciones de interdependencia. Esta matriz permite plasmar las características multidireccionales de los datos espaciales con el fin de determinar la influencia que presentan entre sí. La matriz se supone conocida y tiene que estar definida con anterioridad debido a que entra al proceso de estimación como una constante.

$$W = \begin{pmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{12} & 0 & \dots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1n} & w_{2n} & \dots & 0 \end{pmatrix} \quad (5.1)$$

Es una matriz cuadrada simétrica: la región  $i$  está correlacionada con la región  $j$  por medio de un valor  $w_{ij}$  donde los componentes  $w_{ij}$  de la matriz  $W$  son positivos y finitos,  $w_{ii} = 0$ . Se asume por definición no estocástica, es decir que los valores que la componen no son obtenidos mediante procesos aleatorios sino que sigue parámetros sistemáticos espaciales, razón por la cual es considerada como uno de los factores que causa la dependencia espacial en las mediciones Bohorquez (2009).

### 5.5.1. Estandarización de la matriz $W$

La estandarización de la matriz  $W$  suele hacerse cuando la matriz es binaria, es decir sus componentes  $ij$  son 1 y 0 provenientes de analizar contigüidades de efecto reina o torre

$$\frac{W_{ij}}{\sum_{j \in J} W_{ij}} \quad (5.2)$$

Para el cálculo de los estadísticos globales y locales de autocorrelación espacial, es necesario estandarizar los pesos si la matriz nace a partir de la contigüidad espacial, de manera que cada elemento  $i$  de una fila  $j$  sea dividido por la suma de dicha fila, de tal forma que la suma de cada fila de la matriz estandarizada sea igual a la unidad.

## 5.6. Estadísticos globales de autocorrelación espacial

Para determinar si la configuración espacial de las unidades espaciales se encuentra distribuida aleatoriamente o presenta algún tipo de patrón se realiza el análisis de la autocorrelación espacial global. Los contrastes utilizados para realizar dicho análisis son la  $I$  de Moran,  $C$  de Geary y  $G(d)$  de Getis y Ord Baro-nio, Vianco y Rabanal (2012).

Estas pruebas de asociación espacial tienen las siguientes hipótesis:

$H_0$  : Ausencia de autocorrelación espacial.

$H_a$  : Presencia de autocorrelación espacial.

### 5.6.1. I de Moran Global

El Test I de moran permite medir la autocorrelación espacial con base en las localizaciones y los valores de las variables, evalúa si el patrón presente se encuentra agrupado, disperso o aleatorio en el espacio Yrigoyen (2003). Se expresa de la siguiente manera:

$$I = \frac{N}{S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad i \neq j \quad i, j = 1, 2, \dots, N \quad (5.3)$$

Donde:

$N$ : Cantidad de observaciones o tamaño muestral.

$W_{ij}$ : Elemento de la matriz  $W$  de pesos espaciales.

$S_0$ : Factor de escala igual a la suma de los elementos de la matriz  $W$ .

$x$ : Valor esperado de la variable  $x$ .

### 5.6.2. C de Geary

Se representa de la siguiente forma:

$$C = \frac{(N-1) \sum_{i=1}^N \sum_{j=1}^N W_{ij} (x_i - x_j)^2}{2S_0 \sum_{i=1}^N (x_i - \bar{x})^2} \quad i \neq j \quad i, j = 1, 2, \dots, N \quad (5.4)$$

Donde:

$N$ : Cantidad de observaciones o tamaño muestral.

$W_{ij}$ : Elemento de la matriz  $W$  de pesos espaciales.

$S_0$ : Factor de escala igual a la suma de los elementos de la matriz  $W$ .

$x$ : Valor esperado de la variable  $x$ .

### 5.6.3. G de Getis-Ord

Esta prueba se puede considerar como un complemento a una de las anteriores donde dos pares de regiones  $(i, j)$  serán vecinas cuando se encuentre dentro de una distancia  $d$  específica Yrigoyen (2003). Esta

descrito en la siguiente ecuación:

$$G(d) = \frac{\sum_{i=1}^N \sum_{j=1}^N W_{ij}(d) x_i x_j}{\sum_{i=1}^N \sum_{j=1}^N X_i X_j} \quad i \neq j \quad (5.5)$$

Tiene un estadístico de prueba que se interpreta así: si  $Z(G)$  es significativo y es mayor a cero se rechaza la hipótesis nula y se comprueba la existencia de valores elevados alrededor de  $i$ , si es menor a cero existen valores bajos alrededor de  $i$ , y si el estadístico no es significativo no se rechaza la hipótesis nula y se concluye que existe completa aleatoriedad en la región.

## 5.7. Estadísticos locales de autocorrelación espacial

### 5.7.1. Indicador local de asociación espacial (LISA)

Indicador local de Asociación espacial (LISA) es utilizado para detectar aglomeraciones (clusters) mediante la medición de la autocorrelación espacial donde se localiza cada observación, los mapas LISA ilustran los valores significativos y representan gráficamente los valores atípicos espaciales Serrano y Valcarce (2000). El estadístico local mas utilizado es el  $I$  de Moran es un indicador de dependencia local basado en el índice de Moran, uno de los mas conocidos y extendidos, esta fundamentado en trabajos de Moran (1948), Krishna Iyer (1949) y Geary (1954) requirió algunos ajustes y complementos para su generalización y evolución dando paso a nuevos conceptos, se expresa de la siguiente forma:

$$I_i = \frac{Z_i}{\sum_i Z_i^2 / N} \sum_{j \in J_i} W_{ij} Z_j \quad (5.6)$$

La interpretación después de su estandarización es: cuando  $Z(I_i) > 0$  significativo es porque hay presencia de cluster de valores similares alrededor de la región  $i$ , si esta sigue siendo significativa pero esta vez  $Z(I_i) < 0$  entonces, hay cluster de valores disímiles de la región  $i$ .

### 5.7.2. G de Getis y Ord Local

De igual forma que el indicador global  $G$  este contraste local proporciona una información adicional al indicador local de asociación espacial (LISA) para fortalecer el AEDE, su expresión esta dada por:

$$New - G_i^* = \frac{\sum_{j=1}^N w_{ij} X_j - W_{ij}^* \bar{X}}{S \{[(N S_{1i}^* - W_i^{*2})] / (N - 1)\}^{1/2}} \quad (5.7)$$

$$W_i^* = W_i + w_{ii}, \quad S_{1i}^* = \sum_j w_{ij}^2, \quad S_2 = \sum_j (X_j - \bar{X})^2$$

Si el estadístico es significativo se interpreta de la siguiente forma: cuando  $G_i^* > 0$  muestra hay presencia de valores similares y que además esos son altos alrededor de la región  $i$ ,  $G_i^* < 0$  la región  $i$  estará rodeada de valores similares bajos.

## 5.8. Análisis confirmatorio de datos espaciales

El análisis confirmatorio se ocupa de las diferentes especificaciones posibles para la inclusión del fenómeno espacial en un modelo.

### 5.8.1. Modelos

#### Modelo GLM:

El Modelo Lineal Generalizado, permite codificar todas las situaciones de análisis dentro de un mismo esquema general, este tiene una diferencia con los demás modelos, y es que el componente aleatorio no necesariamente sigue una distribución normal, sino que utiliza cualquier distribución de la familia exponencial.

$$y_i = \beta_j X_{ij} + \epsilon$$

$$\mu_i = E(Y_i) \quad (5.8)$$

$$\eta_i = g(\mu_i)$$

Donde,  $y_i$  es el vector de la variable respuesta,  $\eta_i$  el vector del predictor lineal y el resto de las variables son las ya mencionadas en los modelos anteriores.

Aquí cabe resaltar que, como en GML los residuos pueden seguir cualquier distribución de la familia exponencial, resulta que para la distribución de los errores la homocedasticidad no es imprescindible López-González y Ruiz-Soler (2011).

A partir de esta especificación se generan un conjunto de modelos:

### **Modelo Spatial Durbin:**

Este modelo espacial es uno de los mas robustos y generales, pues los modelos SAR y modelos espaciales de error SMA, son modelos particulares de este, garantiza estimadores consistentes lo que implica estimaciones a los datos espacialmente correlacionados, la forma o ecuación que aquí se maneja es la siguiente:

$$\begin{aligned} y &= \rho W y + X\beta + W X\beta_2 + \epsilon \\ \epsilon &\sim N(0, \sigma^2 I) \end{aligned} \quad (5.9)$$

Donde,  $y$  es el vector de la variable dependiente,  $W$  es una matriz booleana de interacción espacial,  $\rho$  el parámetro que ayuda a determinar la dependencia de la variable dependiente y  $X$  la matriz con  $k$  variables.

### **Modelo Spatial Lag:**

Modelo mixto autorregresivo de regresión espacial o modelo del retardo espacial, incorpora la influencia de las variables omitidas a través de una variable dependiente retardada.

$$\begin{aligned} y &= \rho W y + X\beta + u \\ u &\sim N(0, \sigma^2 I) \end{aligned} \quad (5.10)$$

Siendo  $y$  un vector de dimensiones  $n \times 1$ ,  $W y$  retardo espacial de la variable explicada,  $X$  una matriz con  $k$  variables,  $u$  el término de perturbación de ruido blanco, y  $\rho$  el parámetro espacial autorregresivo Yrigoyen (2003). Si esta forma de autocorrelación espacial fuese ignorada, los estimadores de Mínimos Cuadrados Ordinarios (M.C.O) serían insesgados y además sería incorrecta la inferencia basada en el modelo de regresión clásica, es decir, un modelo sin  $W y$ .

### **Modelo Spatial Error:**

El modelo de regresión con dependencia espacial en la perturbación aleatoria o modelo del error espacial, expone que el efecto de dependencia espacial que tiene la variable endógena es explicado no solo por las

variables independientes sino por otras que se encuentran ausentes, la dependencia espacial residual.

$$\begin{aligned} y &= X\beta + u \\ u &= \lambda W u + \epsilon \end{aligned} \tag{5.11}$$

Donde  $u$ : es la perturbación aleatoria distribuida según un proceso autorregresivo de orden 1,  $\lambda$ : es el parámetro autorregresivo asociado al retardo espacial  $W u$  y  $\epsilon$  es el vector de perturbaciones aleatorias, ruido blanco. Yrigoyen (2003)

### Modelo SAR:

El modelo econométrico espacial autorregresivo de primer orden es una de las formas más simples de expresar la dependencia espacial de los valores que tiene una variable dependiente de la localización de la misma, esta se propone de forma matricial y se deriva del modelo lineal  $y = X\beta + \epsilon$ , el modelo SAR tiene la siguiente forma:

$$\begin{aligned} y &= \rho W y + u \\ u &\sim N(0, \sigma^2 I) \end{aligned} \tag{5.12}$$

Donde  $y$  on los vectores de variables,  $u$  el término de error y  $W$  la matriz de pesos espaciales, lo que indica  $\rho W y$  es una estructura autorregresiva en la variable dependiente que intenta explicar el modelo. Dentro de esta se puede derivar una amplia gama de modelos, los cuales pueden incluir diferentes combinaciones de estructuras autorregresivas y procesos de medidas móviles Pérez Pineda (2006).

### Modelo heterocedastico SARAR:

Estimación de varios pasos generalizada de momentos de la varianza covarianza con un modelo lineal de Cliff y Ord de la forma:

$$\begin{aligned} y &= \lambda W y + X\beta + u \\ u &= \rho W u + \epsilon \end{aligned} \tag{5.13}$$

con

$$\epsilon \sim N(0, \sum_1^2)$$

el modelo permite el rezago espacial en la variable dependiente y las perturbaciones. Los términos en el



proceso de perturbación se asumen de forma desconocida y heterocedastica Piras (2018), Harry H. Kelejian e Ingmar R. Prucha (2010).

## Modelo HAC espacial

Mínimos cuadrados espaciales en dos etapas con errores estándar HAC, estimador no paramétrico heterocedastico y con autocorrelación consistente de la varianza covarianza de un vector de momentos en un contexto espacial, el vector de perturbación se genera de la siguiente manera donde R es una matriz no estocástica Piras (2018).

$$u = RE \quad (5.14)$$

Es necesario primero especificar una clase de función de kernel el cual determina los pesos para las diferentes covarianzas en una función de medición de distancias entre las respectivas unidades Harry H Kelejian e Ingmar R Prucha (2007).

### 5.8.2. Regresión geográficamente ponderada GWR

Tiene como finalidad estimar un vector de parámetros diferente para cada observación, para que este vector dependa de la ubicación espacial de cada observación. Se realiza la estimación debido a la inestabilidad de los parámetros y la ausencia de estacionarias espacial Bohorquez (2009).

$$y_i = X_i\beta(v_i, v_i) + u \quad (5.15)$$

Donde:

$i = 1, 2, 3, \dots, n$  y  $(v_i, v_i)$  coordenadas geográficas. La estimación de  $\beta$  se realiza de la siguiente manera:

$$\hat{\beta}(v_i, v_i) = [X'W(v_i, v_i)X]^{-1}X'W(v_i, v_i)y \quad (5.16)$$

Donde  $W(v_i, v_i) = \text{Diag}(\alpha_{ij})$  y  $(\alpha_{ij})$  es una ponderación determinada en términos de la distancia de la observación espacial  $i$  a la observación espacial  $j$  ( $j = 1, 2, \dots, 3$ ).

## 6. Descripción base de datos

Para este trabajo como insumo principal se utilizó una base de datos con 211 observaciones de viviendas distribuidas aleatoriamente en la ciudad de Baltimore del año 1978, a continuación se presentan las variables a trabajar:

Variable	Descripción
STATION	Variable ID
PRICE	Precio de venta de la casa en \$ 1,000 (MLS)
NROOM	Número de habitaciones
DWELL	1 si unidad separada, 0 de lo contrario
NBATH	Cantidad de baños
PATIO	1 si tiene patio, 0 de lo contrario
FIREPL	1 si tiene chimenea, 0 de lo contrario
AC	1 si tiene aire acondicionado, 0 de lo contrario
BMENT	1 si tiene sótano, 0 de lo contrario
NSTOR	Número de historias
GAR	Número de espacios para automóviles en el garaje (0 = sin garaje)
AGE	Edad de la vivienda en años
CITCOU	1 si la vivienda está en el condado de Baltimore, 0 de lo contrario
LOTSZ	Tamaño de lote en cientos de pies cuadrados
SQFT	Espacio interior en cientos de pies cuadrados
X	Coordenada x en la cuadrícula de Maryland
Y	Coordenada y en la cuadrícula de Maryland

Cuadro 6.1: Tabla de descripción de las variables

Para los diferentes modelos que se van a construir en las secciones siguientes de las variables se espera que entre mas numero de habitaciones y numero de baños el precio de la vivienda sea mayor, de igual forma si tiene patio,sótano, mas espacios para automóviles, chimenea y aire acondicionado el valor de la vivienda va aumentar, también que entre mas grande sea el tamaño del lote y el espacio interior de la vivienda el valor aumente, en cuanto a la edad a medida que esta sea mayor se espera que repercuta negativamente en el valor de la vivienda.

Para realizar un análisis mas completo en cuanto a la matriz de pesos espaciales (W) optamos por generar

unos polígonos de Voronói a partir de las diferentes localizaciones de las viviendas esto con el fin de generar una matriz pesos a partir de la contigüidad espacial haciendo uso del efecto reina y el efecto torre hasta un orden de contigüidad 9 y poder comprarlas con las matrices de pesos a partir de K-vecinos mas cercanos y aquellas obtenidas a partir de criterios gráficos (Triangulación Delaunay, Esfera de influencia, Gráfica de Gabriel y vecinos relativos).

Al realizar la evaluación de los modelos pudimos observar que no se tenía normalidad, y a pesar de intentar con varias técnicas para resolverlo como son las transformaciones de Box-Cox, transformaciones de anamorfosis Gaussiana y cambiando la forma funcional del modelo no se pudo solucionar a lo cual para solucionar se eliminaron datos atípicos en los cual el precio del hogar esta por encima de 145 con el cual se borraron un total de dos registros un 0,94 % de los datos totales, después de eliminar esos datos se pudo solucionar el problema con la normalidad, y al observar las descripciones de variables y los datos vimos como la variable *BMENT* a pesar de que en la definición la tomaban como una variable dummy haciendo referencia a la existencia o no de un sótano tenía valores diferentes a 0 y dado a que la descripción de esta dice que toma valores de 1 y 0 decidimos transformarla tomando todos los valores diferentes a 0 como 1 (si tiene sótano) , por estos motivos en el presente análisis se tiene en cuenta una base de datos que no es exactamente igual a la original.

## 7. Análisis descriptivo de datos espaciales

### 7.1. Área de estudio

Baltimore está en la zona central-norte del estado de Maryland en Estados Unidos, es la ciudad mas poblada del estado y es denominada Ciudad de Baltimore para distinguirla del condado de Baltimore. La ciudad está situada, junto al río Patapsco, cerca de su desembocadura en la Bahía Chesapeake. La ciudad está dividida, oficialmente, en nueve zonas: norte, noroeste, nordeste, oeste, central, este, sur, soroeste y sureste.

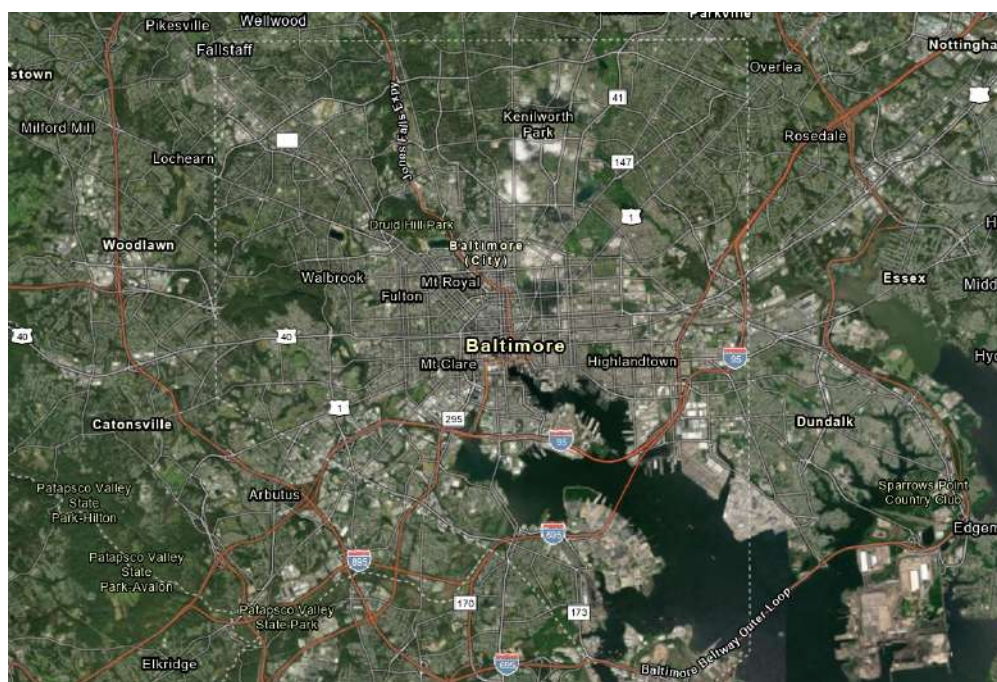


Figura 7.1: Ciudad de Baltimore, Maryland, USA

## 7.2. Polígonos de Thiessen y Voronói

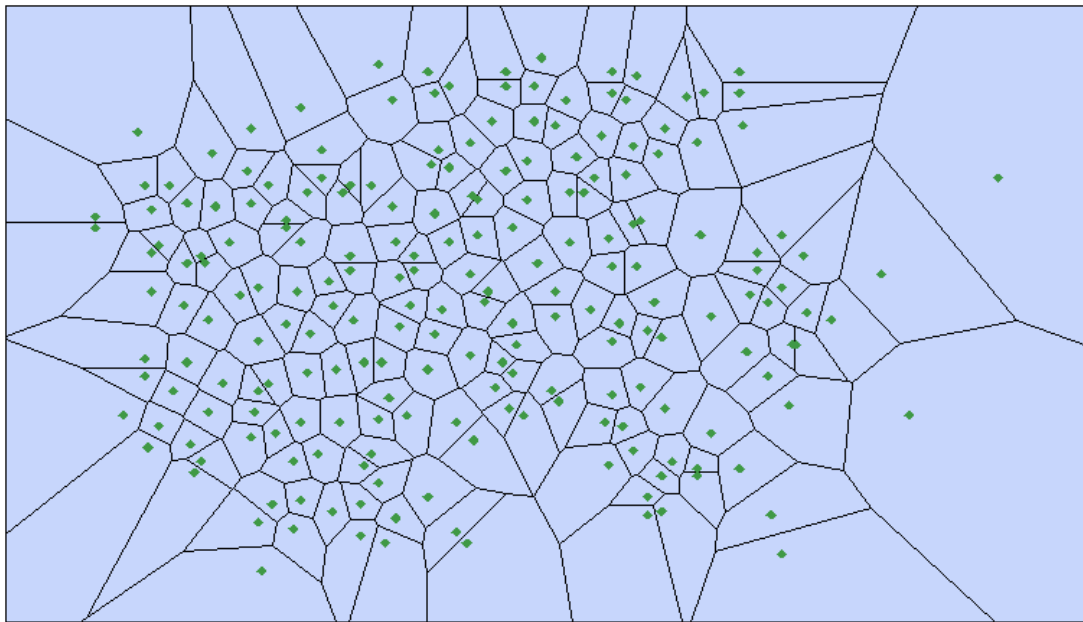


Figura 7.2: Mapa polígonos creados a partir de los puntos

## 7.3. Estadísticas descriptivas

### 7.3.1. Variable dependiente

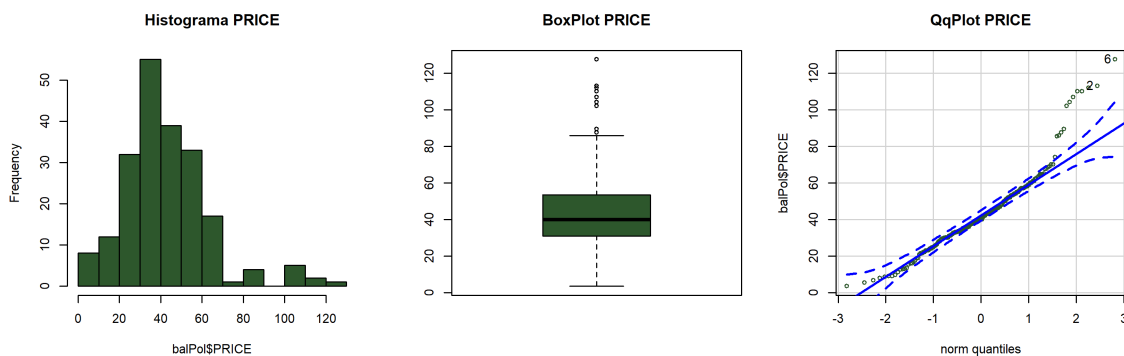


Figura 7.3: Gráficos estadísticos para la variable PRICE

Del histograma de la variable PRICE se puede pensar que la variable sigue una distribución normal con la mayoría de datos en valores de entre 30 mil y 60 mil dolares para la época , para corroborar con el gráfico Q-Q obtenido a través de R concluimos que no sigue una distribución normal y además se observa la exis-

tencia de algunos datos atípicos.

### 7.3.2. Variables independientes

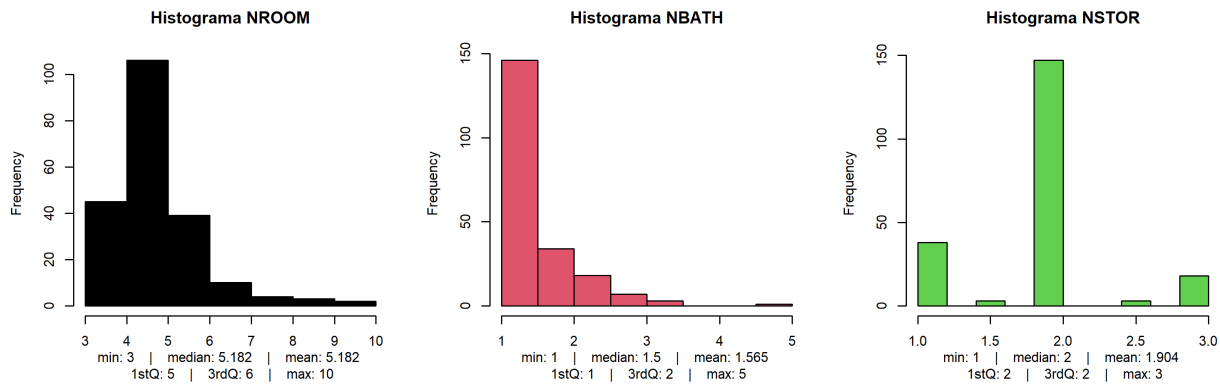


Figura 7.4: Histogramas: Numero de habitaciones, numero de baños y numero de historias

En el histograma de número de habitaciones muestra que la muestra de las viviendas en su mayoría tienen 4 habitaciones y un promedio para la muestra de aproximadamente 5 habitaciones, el número mínimo de habitaciones para la muestra es de 3 y el máximo es de 10.

Se tenía pensado que el número de baños iba a ser proporcional al número de habitaciones, sin embargo en el histograma de número de baños (NBATH) se observa que la mayoría de viviendas tienen entre 1 baño y 1 baño y medio donde 1 baño completo incluye (bañera, inodoro y lavamanos) y medio baño se compone de (inodoro y lavamanos).

Al no tener una definición de la variable NSTOR se asume que representa la cantidad de propietarios que ha tenido la vivienda encontrándose en su gran mayoría con 2 propietarios y con viviendas con un número máximo de 3 propietarios.

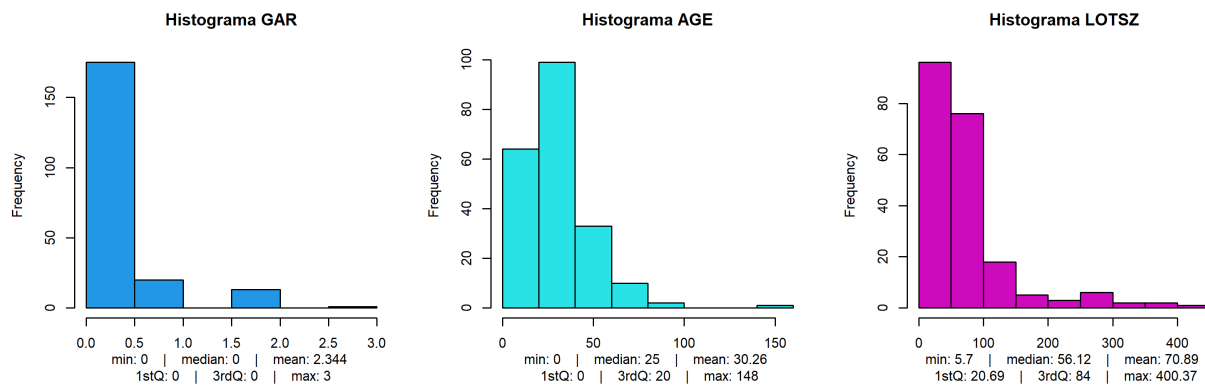


Figura 7.5: Histogramas: Garaje, edad y tamaño de lote

Para la variable GAR número de espacios para vehículos, se observó que la mayoría de viviendas no contaban con un espacio para vehículos, también muy pocas viviendas contaban con 1,2 y hasta 3 garajes como máximo.

Del histograma de la variable AGE (edad) se concluye que la mayoría de viviendas tienen una vetustez de aproximadamente 30 años esto quiere decir que la industria de la construcción se desarrolló fuertemente en la ciudad de Baltimore hacia los años 40, también que existen viviendas con una vetustez de más de 100 años.

La variable independiente tamaño de lote LOTSZ presenta una asimetría hacia la izquierda y se verifica porque su media es mayor a su mediana, lo que indica que el tamaño de lote en su gran mayoría no supera los 10.000 pies cuadrados.

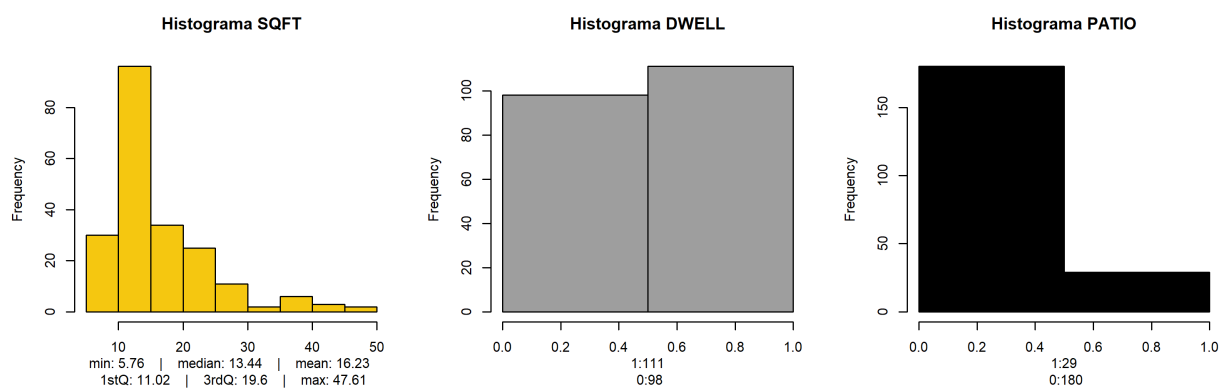


Figura 7.6: Histogramas: Espacio interior,DWELL, patio

Del histograma para el variable espacio interior (SQFT), se puede observar que esta variable no presenta una tendencia específica además que la mayoría de viviendas tienen un espacio interior de aproximadamente

1500 pies cuadrados, la vivienda más pequeña tan solo tiene un espacio interior de 576 pies cuadrados y la más grande 4761 pies cuadrados .

En cuanto a la posesión de un patio en el histograma para la variable PATIO se observa una clara distinción en donde la mayoría de viviendas no cuenta con un patio con un total de 180 al no tener patio se espera que el precio de estas viviendas sea menor de las que si posean, las viviendas que si poseen un patio son un total de 29.

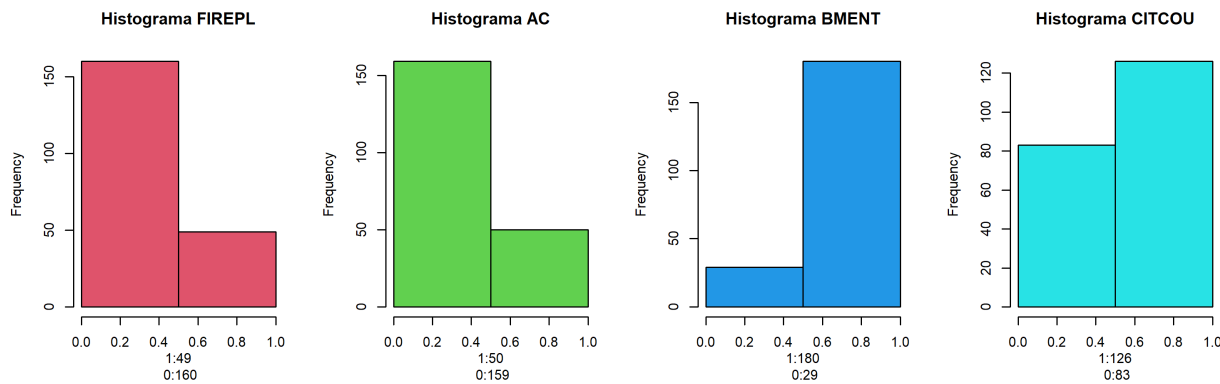


Figura 7.7: Histogramas: Chimenea,aire acondicionado,sotano,localización

Del histograma de la variable FIREPL se concluye que la mayoría de viviendas un total de 160 no cuentan con una chimenea en proporción a las que sí cuentan con esta característica solamente 49.

Al igual que la variable anterior la proporción de viviendas que no cuentan con aire acondicionado es superior con un total de 159 a la que sí tienen esta característica 50.

Con la variable BMENT que se asocia a si la vivienda tiene o no sótano después de haber realizado la conversión de algunos valores de acuerdo al histograma se observa que en su gran mayoría 180 viviendas si poseen un sótano en relación a las que no que son 29.

Para finalizar esta sección de las estadísticas descriptivas en el histograma de la variable CITCOU observamos un número de viviendas de 126 asociadas al valor 1, es decir que pertenecen al condado de Baltimore y 83 que no.



## 7.4. Matriz W de pesos espaciales

Criterio	R2Adj.pos	R2Adj.neg	Pvalue.pos	Pvalue.neg	N.var	R2Adj.select
reina1	0.547573	-0.824658	0.006181	1.000000	8	0.473165
reina2	0.465213	-0.474246	0.006181	1.000000	8	0.459646
reina3	0.353894	-0.374918	0.012325	1.000000	3	0.329427
reina4	0.327964	-0.327964	0.024500	1.000000	3	0.323673
reina5	0.263656	-0.263861	0.169958	1.000000		
reina6	-0.018817	0.021718	1.000000	1.000000		
reina7	-0.329809	0.346989	1.000000	0.018431	8	0.334719
reina8	-0.413956	0.467307	1.000000	0.006181	5	0.282271
reina9	-0.335198	0.384615	1.000000	0.006181	6	0.374866
torre1	0.547573	-0.824658	0.006181	1.000000	9	0.485148
torre2	0.465213	-0.474246	0.006181	1.000000	8	0.459646
torre3	0.353894	-0.374918	0.012325	1.000000	2	0.310358
torre4	0.327964	-0.327964	0.006181	1.000000	3	0.323673
torre5	0.263656	-0.263861	0.122200	1.000000		
torre6	-0.018817	0.021718	1.000000	1.000000		
torre7	-0.329809	0.346989	1.000000	0.018431	8	0.334719
torre8	-0.413956	0.467307	1.000000	0.006181	5	0.282271
torre9	-0.335198	0.384615	1.000000	0.006181	6	0.374866
kvecinos1	0.454351	-0.397577	0.006181	1.000000	17	0.431502
kvecinos2	0.526203	-0.557463	0.006181	1.000000	8	0.440254
kvecinos3	0.511678	-0.596957	0.006181	1.000000	10	0.412976
kvecinos4	0.473696	-0.633369	0.006181	1.000000	14	0.473525
kvecinos5	0.522762	-0.803268	0.006181	1.000000	12	0.428846
kvecinos6	0.540634	-0.938995	0.006181	1.000000	9	0.414369
kvecinos7	0.544673	-1.050988	0.006181	1.000000	10	0.470126
kvecinos8	0.531498	-1.118526	0.006181	1.000000	13	0.451166
kvecinos9	0.536035	-1.128074	0.006181	1.000000	10	0.440463

Criterio	R2Adj.pos	R2Adj.neg	Pvalue.pos	Pvalue.neg	N.var	R2Adj.select
gabriel	0.534737	-0.714985	0.006181	1.000000	6	0.393216
<b>delaunay</b>	<b>0.547004</b>	<b>-0.840518</b>	<b>0.006181</b>	<b>1.000000</b>	<b>9</b>	<b>0.509851</b>
esfera.inf	0.533196	-0.787099	0.006181	1.000000	7	0.429680
vec.relativos	0.516479	-0.502282	0.006181	1.000000	10	0.447336

Cuadro 7.1: Resultados evaluación distintas matrices de pesos

Para seleccionar la matriz de pesos con la cual íbamos a trabajar utilizamos la función `listlw.select` Rdocumentation (2020) la cual selecciona a partir de una lista de matrices de contigüidad la mejor opción, entendiéndose como mejor opción la matriz la cual de como resultado un mayor R cuadrado ajustado para los datos se obtiene como mejor matriz la matriz de criterio gráfico Delaunay (primera de la siguiente figura).

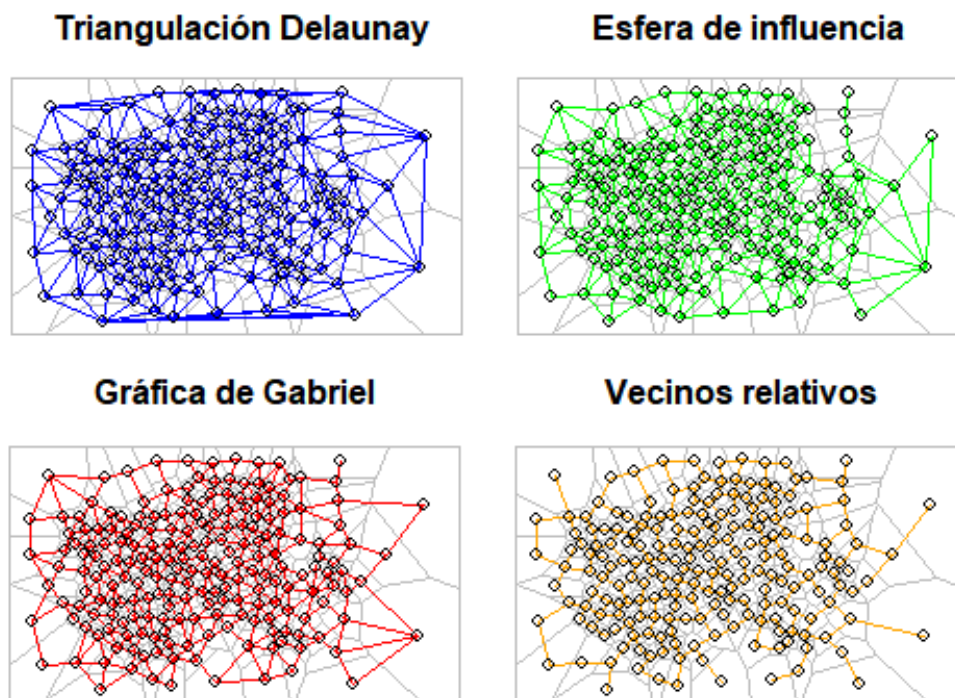


Figura 7.8: Mapa vecinos por criterios gráficos

Con el fin de encontrar estructuras de autocorrelación espacial en la variable precio de la vivienda con la matriz de contigüidad seleccionada, se analiza el correlograma que se muestra a continuación. Este permite examinar patrones de autocorrelación espacial en los datos o modelar residuales, mostrando la correlación entre parejas de observaciones espaciales cuando se aumenta la distancia o el rezago entre ellos. Se encuentra que en todos los órdenes el rezago es significativo, en donde en el primer orden se encuentra el

más significativo y de ahí comienza a disminuir hasta el orden 9.

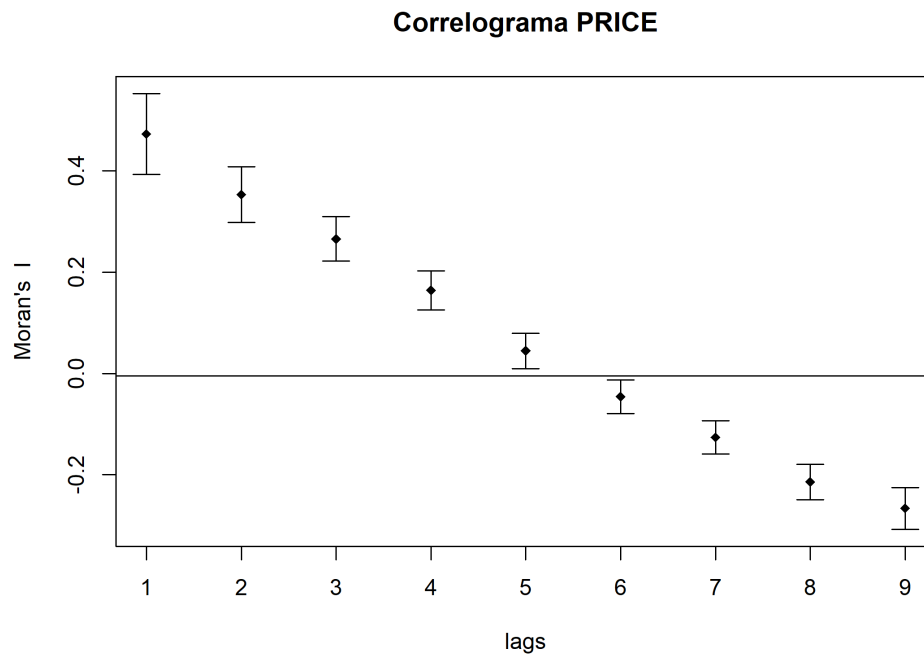


Figura 7.9: Correlograma de Moran de la variable dependiente con la matriz de pesos Delaunay

## 8. Análisis exploratorio de datos espaciales

Para comenzar este análisis se ilustra en los polígonos de Voronoí el comportamiento de los precios de la vivienda a través de un mapa de cuantiles simples.

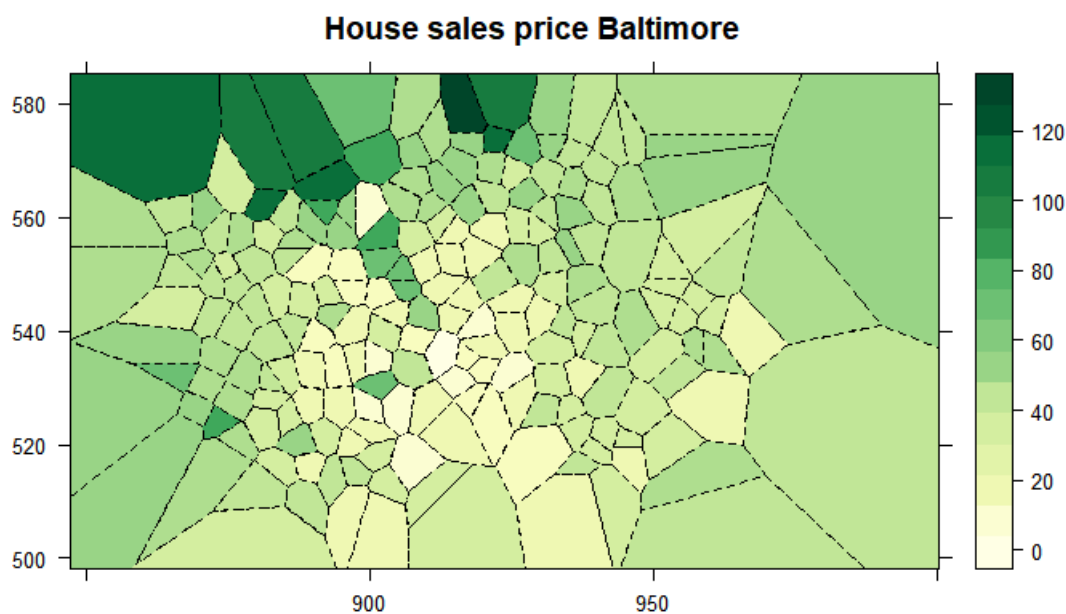


Figura 8.1: Mapa de cuantiles simples para los precios de vivienda

En primera instancia la asociación de colores claros se encuentran concentrados en el centro de la región, es decir que podemos pensar en una concentración de valores bajos para precios de la vivienda, a medida que se comienza a alejar hacia las periferias el valor de la vivienda tiende a aumentar asociándolo con colores mas oscuros, especialmente hacia la región Noroeste.

Para el posterior análisis se realiza con la matriz de peso seleccionada como la mejor en este caso la matriz de pesos Delaunay.

## 8.1. Análisis Global

Una vez observado el comportamiento de los precios en la región se procedió a determinar la presencia de autocorrelación espacial de manera global para la variable precio mediante la prueba del I de moran, se potenció este análisis haciendo uso también de la prueba de Getis y ord, posteriormente para las variables explicativas se hizo la prueba del I de moran a nivel global también para determinar una correlación espacial en estas variables.

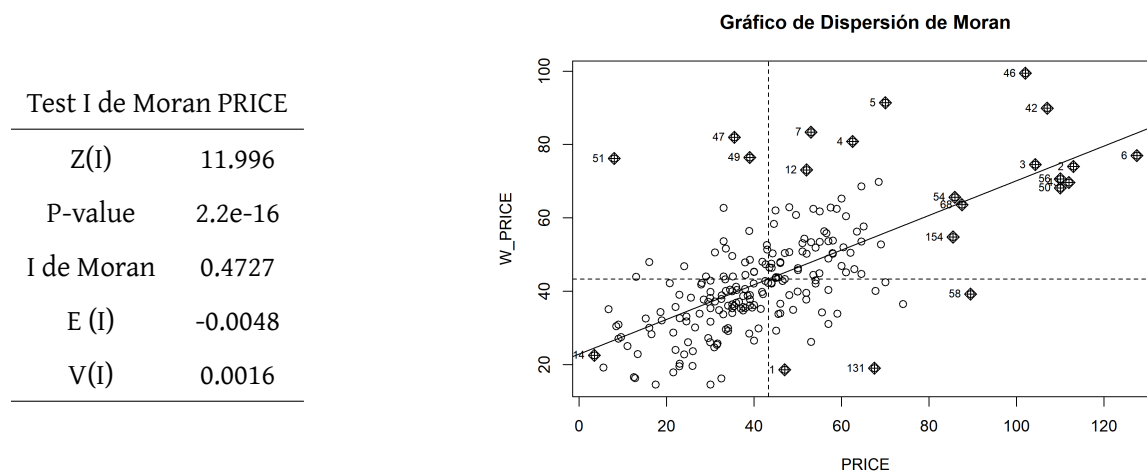


Figura 8.2: I de moran variable dependiente PRICE

Test Getis-Ord global	
Global G	0.00535
p-value	0.0000
E	0.004807
V	0.0000

Cuadro 8.1: Test G de Getis

A partir de los resultados de la prueba del I de moran se concluye que la variable precio presenta una autocorrelación espacial positiva a un nivel de confianza del 95 %. Con la prueba de Getis y Ord concluimos que además de la autocorrelación positiva existen asociaciones de valores similares, donde los precios de vivienda altos están rodeado de viviendas con precios muy similares y que aquellas viviendas cuyos precios son menores también se encuentran rodeadas de viviendas de bajos precios, del dispersograma observamos los valores atípicos. Pensando en las estructuras urbanas el precio de la vivienda puede reflejar la existencia

de barrios mas o menos exclusivos.

### 8.1.1. I de Moran Global variables independientes

VAR	Z	P.value	I.de.Moran	E.I.	V.I.
ROOM	3.053465	0.002262	0.116495	-0.004808	0.001578
DWELL	6.811616	0.000000	0.269344	-0.004808	0.001620
NBATH	5.184488	0.000000	0.201025	-0.004808	0.001576
ATIO	4.510128	0.000006	0.174787	-0.004808	0.001586
FIREPL	3.627006	0.000287	0.140619	-0.004808	0.001608
AC	3.674007	0.000239	0.142531	-0.004808	0.001608
BMENT	3.070780	0.002135	0.117472	-0.004808	0.001586
NSTOR	2.854239	0.004314	0.109373	-0.004808	0.001600
GAR	1.211387	0.225747	0.043077	-0.004808	0.001563
AGE	10.932942	0.000000	0.426376	-0.004808	0.001555
CITCOU	18.513868	0.000000	0.740041	-0.004808	0.001619
LOTSZ	7.335856	0.000000	0.285013	-0.004808	0.001561
SQFT	3.967163	0.000073	0.152919	-0.004808	0.001581

Cuadro 8.2: I de moran para las variables independientes

La anterior tabla revela que la única variable que no es significativa es decir que no presenta una autocorrelacion espacial es la variable que me indica el numero de garajes(GAR). Las demás variables además de presentar una autocorrelacion espacial significativa al 99 % se concluye que también que es positiva, es decir que para estas variables existen concentraciones de valores similares en la ciudad de Baltimore.

### 8.1.2. I de Moran Bivariado

Para el análisis bivariado del I de moran se tuvieron en cuenta las 13 variables que se presume explican el valor de la vivienda en el área seleccionada y cuya tabla de resumen y correspondiente dispersograma se muestra a continuación .

	VAR	Z	P.value	I.de.Moran.Bivariado	E.I.	V.I.
1	NROOM	4.714090	0.000002	0.140264	-0.001919	0.000910
2	DWELL	9.510331	0.000000	0.301561	-0.002636	0.001023
3	NBATH	5.730835	0.000000	0.176165	-0.002328	0.000970
4	PATIO	7.956303	0.000000	0.238228	-0.001929	0.000911
5	FIREPL	7.969101	0.000000	0.248385	-0.002453	0.000991
6	AC	8.411534	0.000000	0.256579	-0.002176	0.000946
7	BMENT	-0.871081	0.383710	-0.025094	-0.000565	0.000793
8	NSTOR	-3.242735	0.001184	-0.090859	0.000644	0.000796
9	GAR	4.193453	0.000027	0.122509	-0.001652	0.000877
10	AGE	-9.810422	0.000000	-0.285271	0.001443	0.000854
11	CITCOU	11.381459	0.000000	0.348083	-0.002182	0.000947
12	LOTSZ	10.895349	0.000000	0.360711	-0.003097	0.001115
13	SQFT	6.153305	0.000000	0.185392	-0.002051	0.000928

Cuadro 8.3: Resultados I de Moran Bivariado para todas las variables

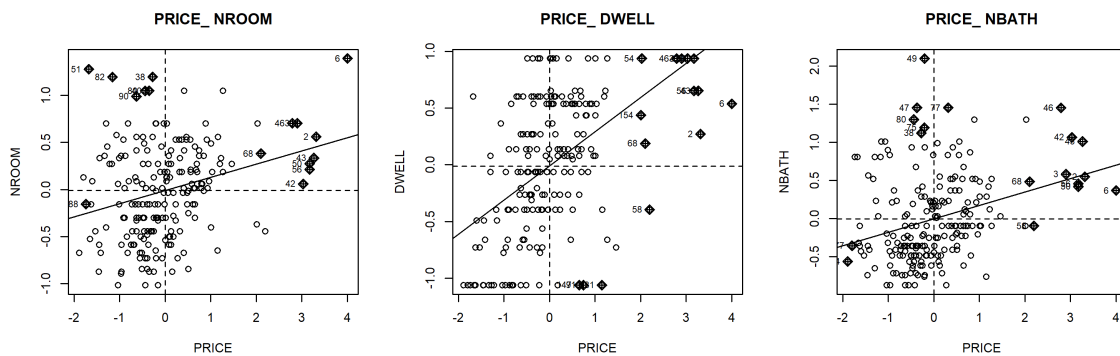


Figura 8.3: Dispersograma bivariado de Moran de PRICE con NROOM, DWELL, NBATH

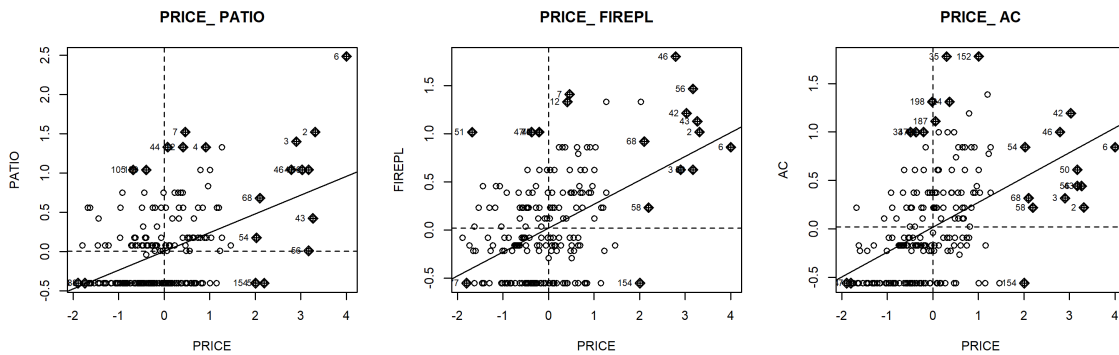


Figura 8.4: Dispersograma bivariado de Moran de PRICE con PATIO,FIREPL,AC

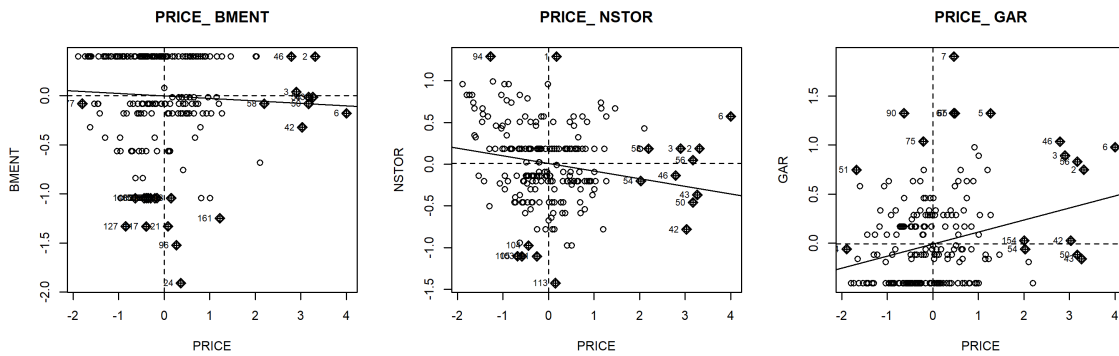


Figura 8.5: Dispersograma bivariado de Moran de PRICE con BMENT,NSTOR,GAR

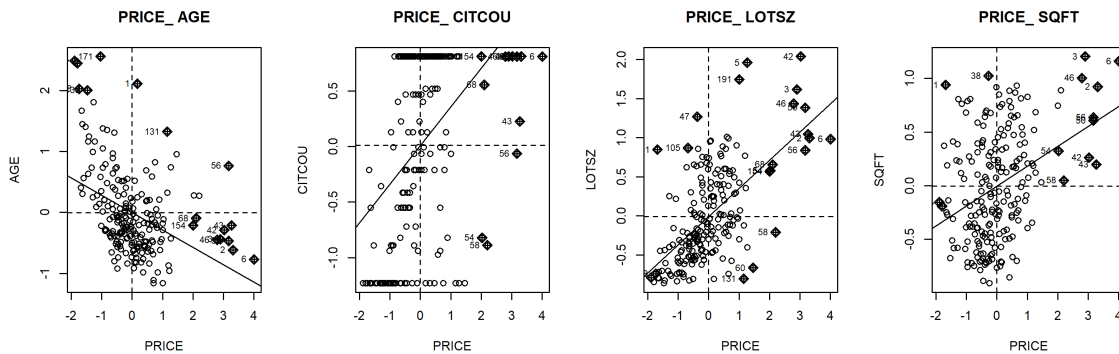


Figura 8.6: Dispersograma bivariado de Moran de PRICE con AGE,CITCOU,LOTSZ,SQFT

Los resultados obtenidos a partir de la aplicación del test bivariado de moran en arrojan que 12 de las 13 variables hay presencia de autocorrelación espacial, únicamente la variable BMENT (existencia de sótano) resulto ser no significativa. Interpretando los demas resultados se tiene:

1.El precio de la vivienda en relación con el numero de habitaciones presenta una autocorrelacion espacial significativa y positiva según el valor (Z) y el p-value y se corrobora con dispersogrma de moran (PRI-



CE\_NROOM), de aquí se puede inferir que las viviendas con precios mas elevados suelen estar rodeadas de viviendas con un alto numero de habitaciones , y aquellas viviendas con precios mas bajos suelen estar rodeadas por viviendas con un numero bajo de habitaciones. Ocurre de manera análoga con las variables (DWELL) y su respectivo dispersograma (PRICE\_DWELL), con el numero de baños (NBATH) y su dispersograma (PRICE\_NBATH), con la variable PATIO y el dispersograma (PRICE\_PATIO), con la existencia de chimenea (FIREPL) y aire acondicionado (AC), con el numero de espacios para vehículos (GAR). En donde el estadístico (Z) es positivo y su P-value es significativo al 1 %. Tambien las variables (CITCOU), tamaño de lote (LOTSZ) y tamaño interior presentan una autocorrelacion espacial positiva y significativa. Según esto lo anterior podemos concluir 2 cosas, la primera es entre mas alto sea el precio de una vivienda, este tiende a estar rodeado por viviendas con las mejores características y segundo las viviendas con precios bajos tienden a estar rodeadas de viviendas con características como un menor numero de habitaciones, de baños, ausencia de espacios para vehículos, chimenea, patio y aire acondicionado, un lote mas pequeño y espacio interior menor.

2. El precio de la vivienda en relación con el numero de propietarios que ha tenido y la edad es diferente que con las demás variables pero concuerda con lo planteado. Primero que todo ambas variables tienen una correlación espacial significativa y negativa, observando los dispersogramas de moran (PRICE\_AGE) y (PRICE\_NSTOR) indican que a bajos precios de la vivienda estas tiende a estar rodeada de viviendas con una edad mayor y que han sido propiedad por un mayor numero de personas, y aquellas viviendas con precios mas elevados tienden a tener viviendas vecinas con pocos años de antigüedad y en donde los propietarios que ha tenido han sido pocos.

### **8.1.3. Correlogramas bivariados**

Para encontrar las estructuras de correlación espacial entre el precio de la vivienda y las 13 variables explicativas con la matriz seleccionada, se analizan los correlogramas que se muestran a continuación el cual nos va a permitir examinar patrones de autocorrelacion espacial en los datos, mostrando la correlación entre parejas de observaciones espaciales cuando se aumenta la distancia o el rezago entre ellos. Se encuentra que en los ordenes 4 y 5 en la variable sótano (BMENT) el rezago es significativo, que las variables (NROOM) es significativa en todos sus ordenes menos en 6 con una mayor significancia en orden 1, la variable (DWELL) es significativa en todos los ordenes del 1 al 9, en cuanto al numero de baños (NBATH) es significativa en todos sus ordenes menos en el 6 con una mayor significancia en orden 1, la variable (PATIO) no es significativa en el orden 5 y 6 de los 9 planteados, en cuanto chimenea (FIREPL) y aire acondicionado (AC) no es significativo en el orden 6 y tiene una mayor significancia en el primer orden, la variable (NSTOR)

deja de ser significativa en los ordenes 5 y 6 al igual que (GAR), la edad (AGE) y el tamaño de lote (LOTSZ) son significativas en todos los ordenes pero mayormente en el primero, finalmente las variables (CITCOU) y (SQFT) en sus respectivos correlogramas se observa que el espacio interior de la vivienda ya no es significativo en el orden 6 y si la vivienda se encuentra o no en el condado de baltimore no es significativa en el orden 5.

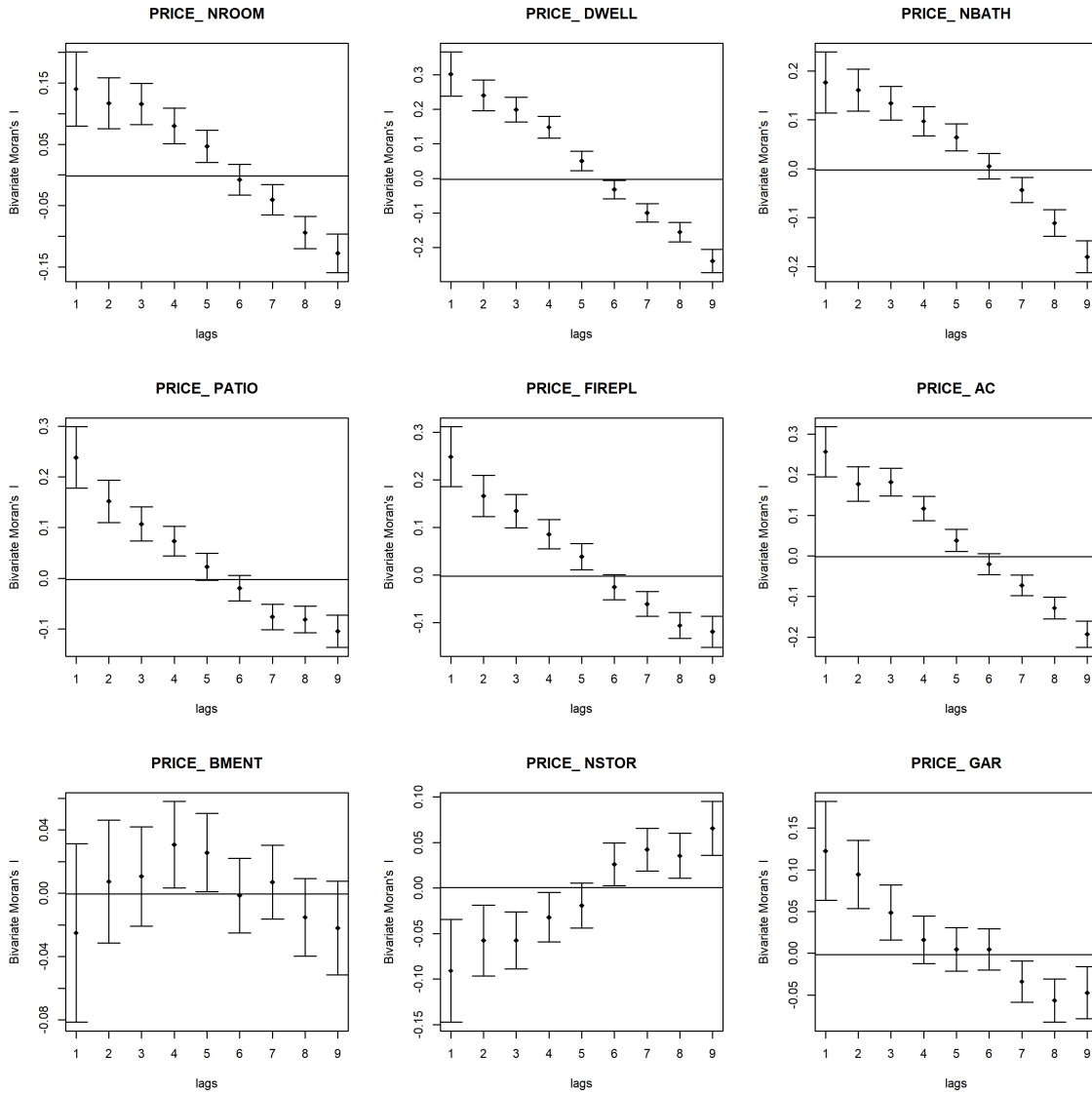


Figura 8.7: Correlogramas bivariado de Moran 9 variables independientes

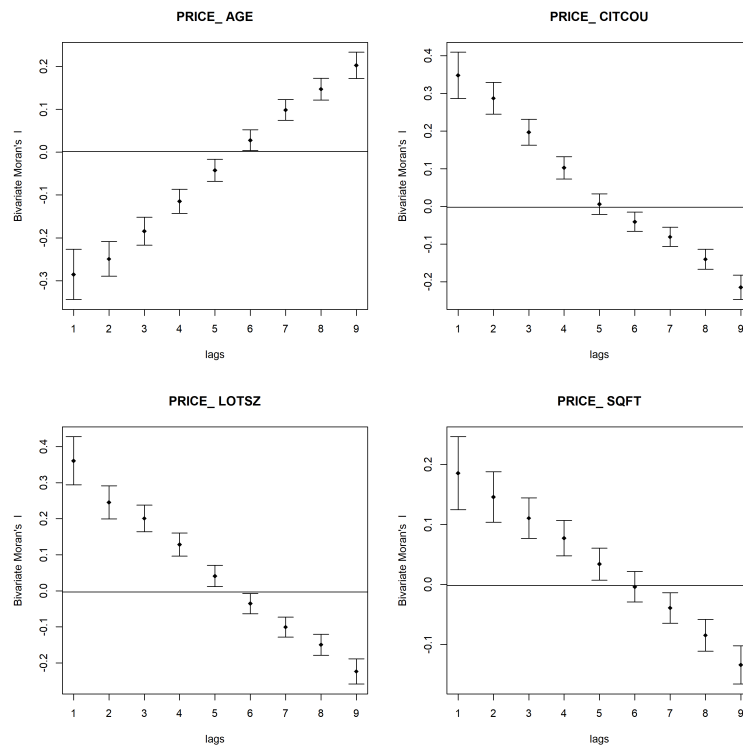


Figura 8.8: Correlogramas bivariado de Moran 3 variables independientes

## 8.2. Análisis Local

EL test de LISA que se presenta a continuación nos va a permitir identificar aquellas zonas donde hay presencia de clusters, al lado izquierdo se tienen los mapas de cluster para las variables y al lado derecho su nivel de significancia en donde el color gris representa la no significancia, para empezar en el test de LISA para la variable precio (PRICE) se observan dos tipos de cluster, el primero de color rojo asociado a valores de precios altos hacia la parte norte de la ciudad, y hacia la zona sur y centro se identifica un cluster de precios bajos para la vivienda.

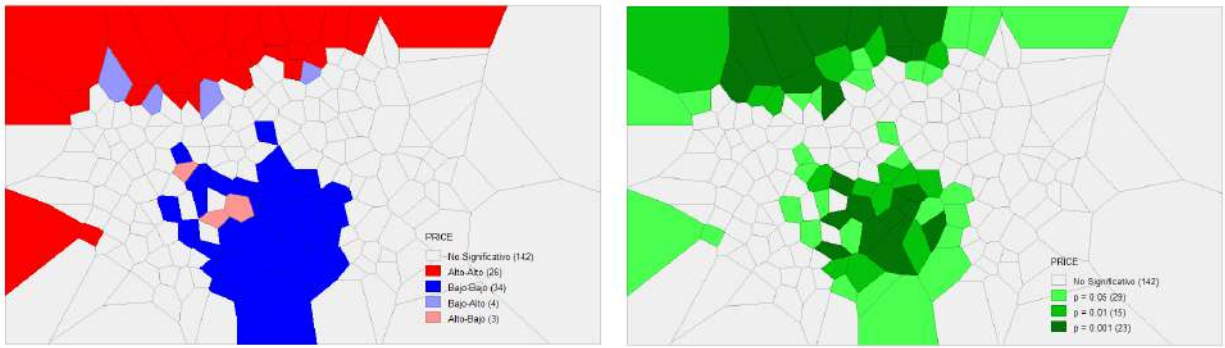


Figura 8.9: LISA PRICE

El test de LISA para las variables independientes y que explican el precio de la vivienda se realizó de la misma manera, siguiendo este orden la variable numero de habitaciones (NROOM) presenta un único cluster de valores bajos (color azul) hacia el sector sur-este del área donde se tomaron las muestras, también presenta de forma aislada algunas viviendas de un bajo numero de habitaciones rodeadas de viviendas de un numero alto de habitaciones.



Figura 8.10: LISA NROOM

La variable DWELL al no tener certeza de un significado en concreto pero determinar que es significativa en los modelos como se muestran en el siguiente capítulo, también se le realizó un test de LISA en donde evidenciamos dos tipos de cluster, el primero hacia el centro de la ciudad con valores bajos rodeados de bajos (azules), otro hacia las periferias de altos rodeados de altos (rojos) y algunos valores aislados y dispersos de valores bajos rodeados de altos (azul opaco), altos rodeados de bajos (rojo opaco).

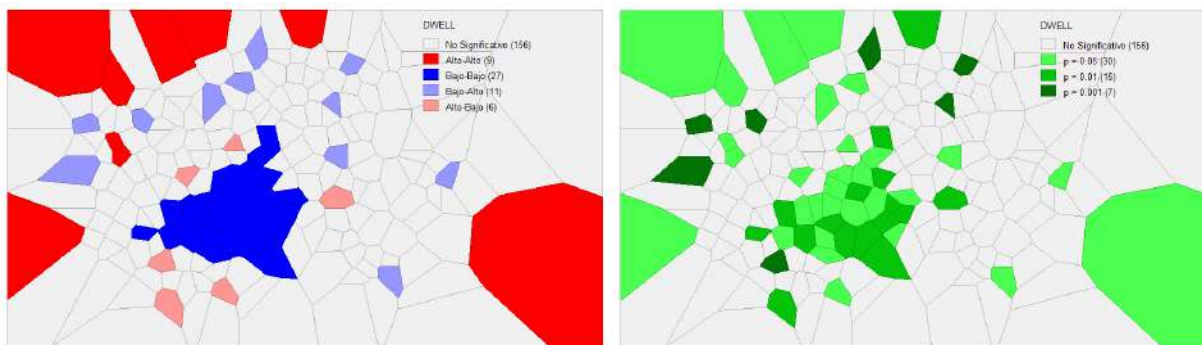


Figura 8.11: LISA DWELL

En cuanto al numero de baños (NBATH) y su test de LISA se observan cluster de bajos hacia el sector sur-este y un cluster de altos hacia zonas norte-este, sin embargo también existen valores aislados y dispersos de valores bajos rodeados de altos (azul opaco), altos rodeados de bajos(rojo opaco).

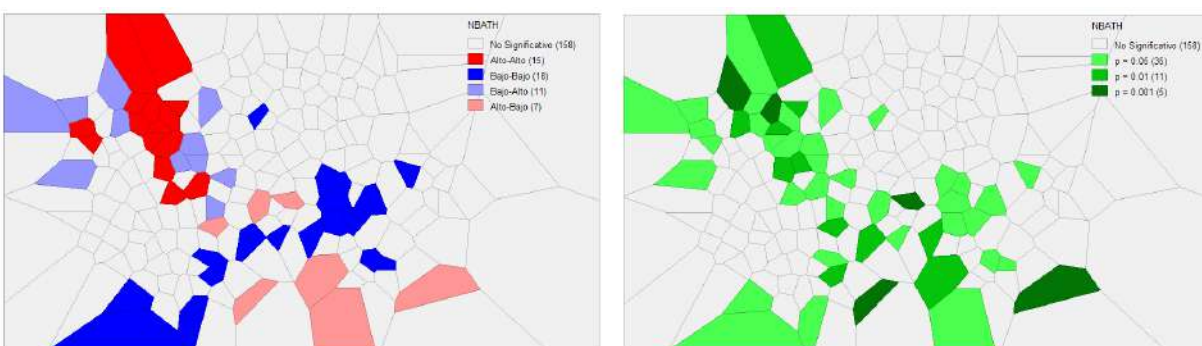


Figura 8.12: LISA NBATH

La variable patio (PATIO) como se observa en su mapa únicamente presenta un cluster de valores altos rodeados de altos hacia la parte norte, que a su vez tiene de forma contigua un cluster de bajos rodeado de altos.

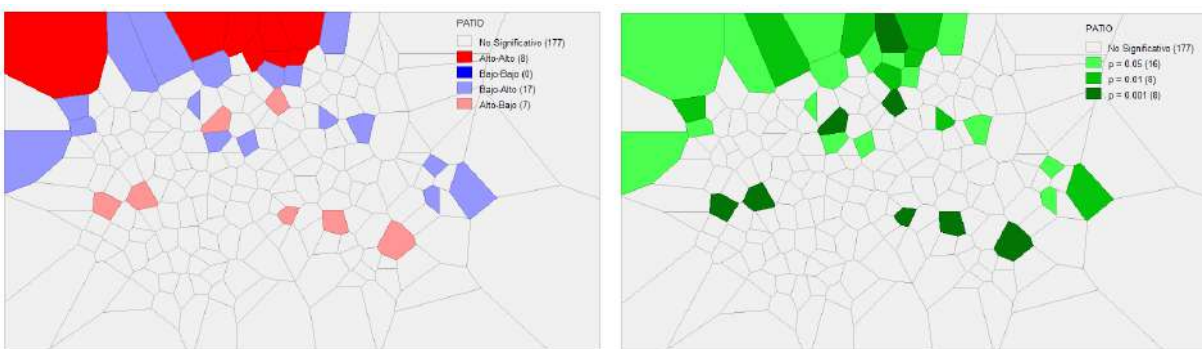


Figura 8.13: LISA PATIO

La variable chimenea (FIREPL) presentan un cluster hacia el norte del área de estudio, el cual según el mapa de cluster se puede concluir que se encuentra rodeado de viviendas sin chimeneas.

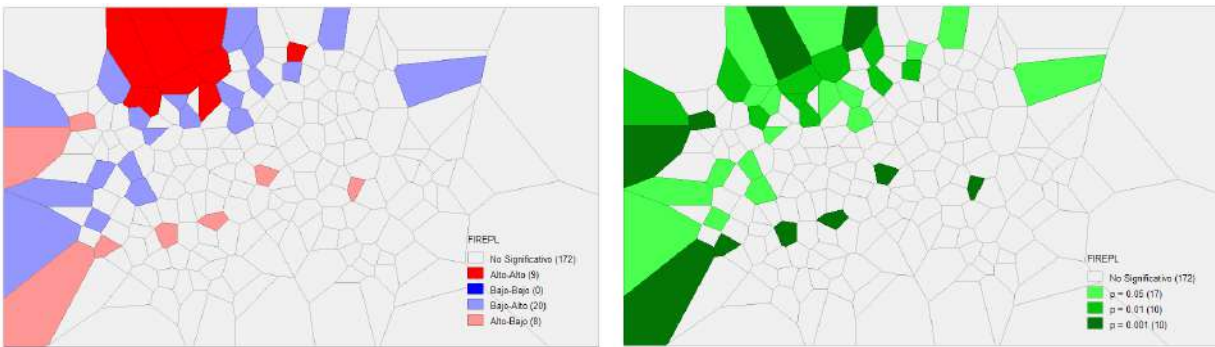


Figura 8.14: LISA FIREPL

El aire acondicionado (AC) a primera vista parece que presenta algunos cluster de altos rodeados de altos, es decir viviendas con aire acondicionado rodeadas de viviendas que también poseen, sin embargo a mas detalle también se encuentran algunos cluster de viviendas sin aire acondicionado rodeadas de viviendas con aire acondicionado.

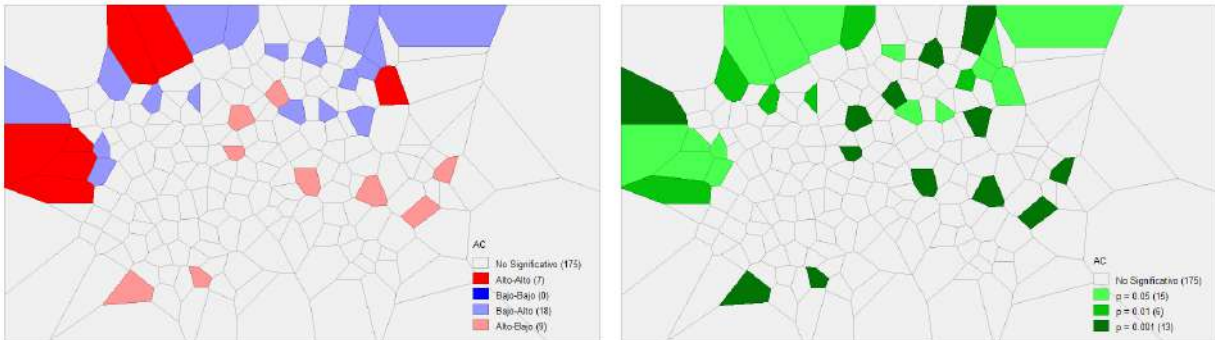


Figura 8.15: LISA AC

Hacia la zona sur de ciudad se puede evidenciar un cluster de bajos para la variable sótano (BMENT) y estos rodeados de altos, sin embargo también existen valores aislados y dispersos de valores bajos rodeados de altos (azul opaco), altos rodeados de bajos(rojo opaco).



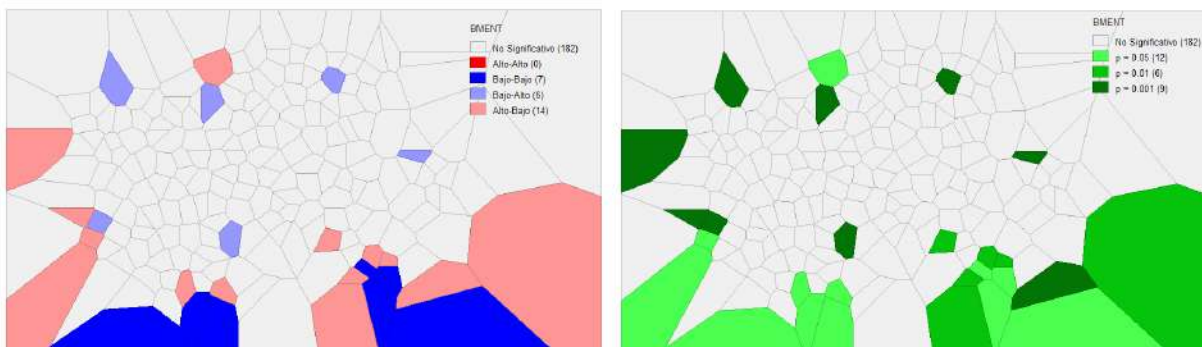


Figura 8.16: LISA BMENT

La variable (NSTOR) presenta un cluster de altos hacia el centro de la región de estudio, en otras palabras una aglomeración de viviendas con que ha tenido un numero alto de propietarios, también un cluster de bajos hacia la región este además existen valores aislados y dispersos de valores altos rodeados de bajos(rojo opaco).

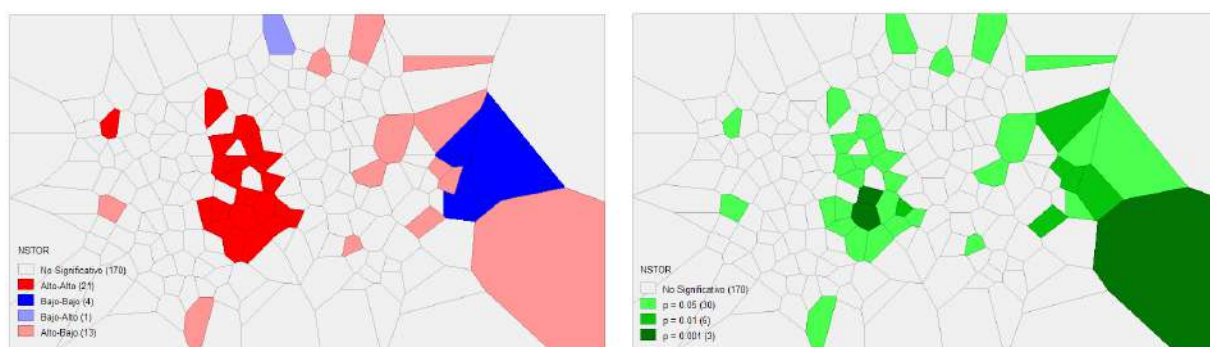


Figura 8.17: LISA NSTOR

La variable (GAR) espacios para vehículos presenta en su gran mayoría valores aislados y dispersos de valores bajos rodeados de altos (azul opaco), altos rodeados de bajos(rojo opaco).

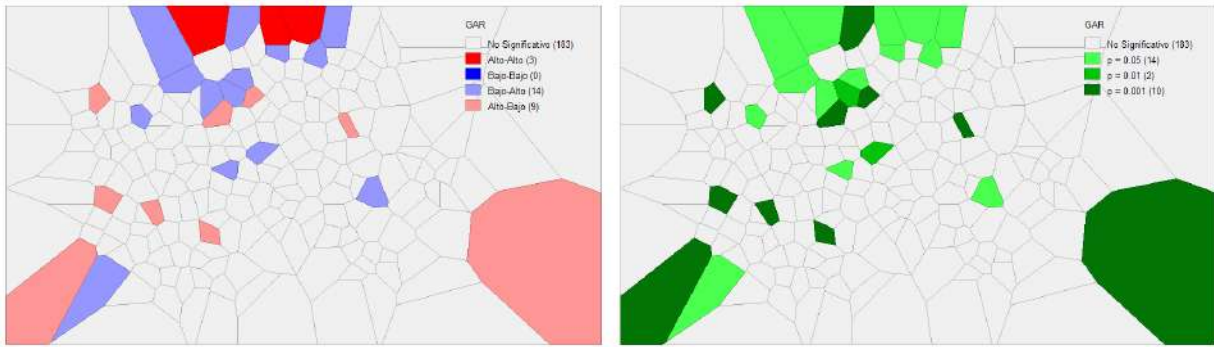


Figura 8.18: LISA GAR

La edad de las viviendas presenta 3 cluster muy notables el primero de altos rodeados de altos hacia el centro y sur del área de estudio, el segundo se presenta como bajos rodeado de bajos hacia el sector oeste y el tercero se presentan hacia el norte-este también como bajos rodeados de bajos, una forma muy convencional en las ciudades modernas en donde hacia el centro de ellas se encuentran las ciudades con una mayor vetustez y hacia las periferias las construidas mas recientemente.

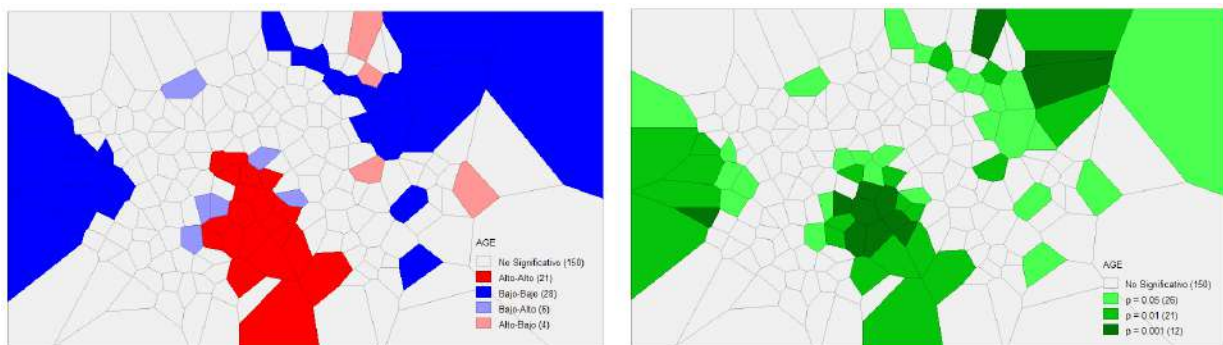


Figura 8.19: LISA AGE

La variable (CITCOU) presenta dos tipos de cluster muy notorios dada la naturaleza de la variable, el primero es un bajo rodeado de bajos al centro del afea de estudio y el otro es un alto rodeado de altos hacia las periferias del area de estudio.



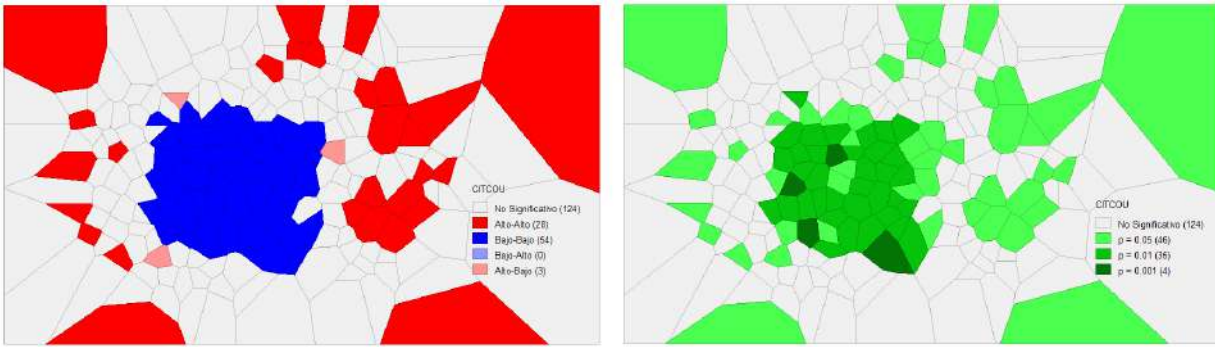


Figura 8.20: LISA CITCOU

El tamaño de lote (LOTSZ) presenta también dos cluster muy notorios, el primero un cluster de altos hacia el norte y hacia el centro de del área de estudio un cluster de bajos, viviendas con un tamaño de lote rodeadas de viviendas con tamaños de lote pequeño.

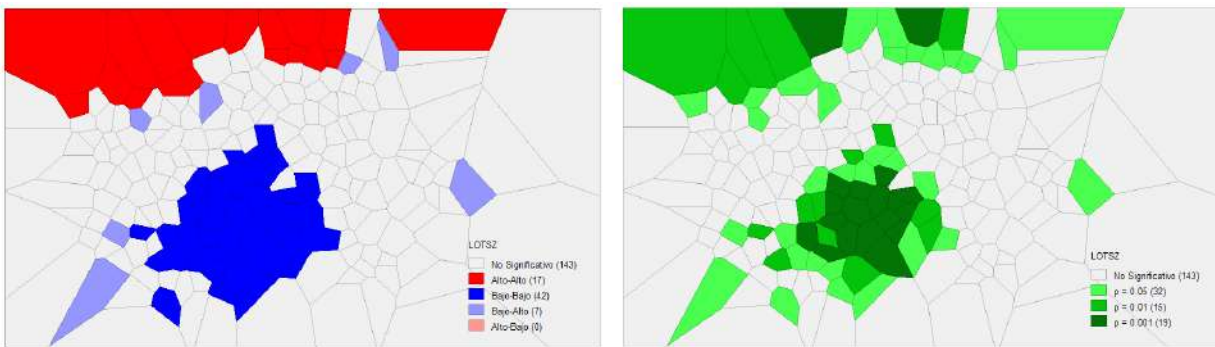


Figura 8.21: LISA LOTSZ

La variable (SQFT) presenta un cluster de bajos hacia el sector este y sur con algunos pequeños clusters dispersos de valores bajos rodeados de altos (azul opaco), altos rodeados de bajos(rojo opaco).

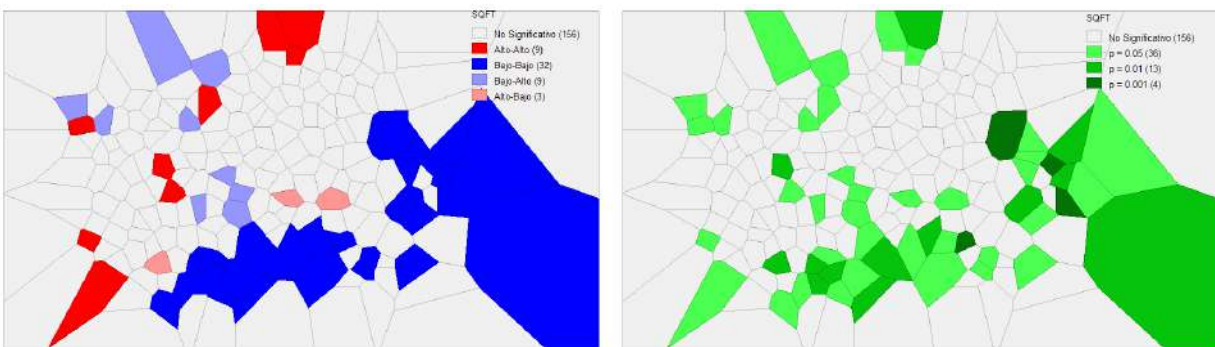


Figura 8.22: LISA SQFT

### 8.2.1. I de Moran Local

ID	Ii	Z.Ii	Pr(z > 0)	Getis
2	4.86	12.20	0.00	3.76
3	4.33	12.61	0.00	4.40
4	1.64	3.76	0.00	4.05
5	2.92	6.68	0.00	5.20
6	6.45	14.73	0.00	3.81
7	0.89	2.24	0.01	4.74
10	0.68	1.72	0.04	2.27
14	1.87	4.28	0.00	-2.27
15	1.87	4.69	0.00	-3.22
16	1.68	4.21	0.00	-3.41
42	6.75	13.75	0.00	4.60
43	4.12	11.18	0.00	3.50
46	7.49	18.79	0.00	6.78
50	3.78	9.47	0.00	3.05
53	0.84	1.71	0.04	2.12
54	2.16	4.95	0.00	2.45
56	4.15	11.27	0.00	3.62
68	2.04	5.94	0.00	2.84
70	0.90	2.26	0.01	-2.43
71	1.20	3.28	0.00	-2.04
72	0.86	1.98	0.02	-3.09
82	0.62	1.70	0.04	-1.45

ID	Ii	Z.Ii	Pr(z > 0)	Getis
86	0.93	2.54	0.01	-2.48
88	0.68	1.71	0.04	-0.99
93	0.77	2.09	0.02	-2.51
94	0.91	2.49	0.01	-1.94
116	0.82	2.23	0.01	-1.71
129	0.72	2.10	0.02	-2.00
130	0.96	1.97	0.02	-1.21
132	1.38	3.16	0.00	-2.21
133	1.87	4.69	0.00	-3.18
135	1.06	2.67	0.00	-2.73
137	1.26	3.17	0.00	-1.94
138	1.34	3.36	0.00	-2.18
153	0.63	1.73	0.04	2.51
154	1.10	2.26	0.01	1.13
171	1.25	3.15	0.00	-3.01
172	0.93	2.14	0.02	-2.55
174	1.09	2.98	0.00	-3.04
176	1.01	2.74	0.00	-1.66
177	2.07	5.19	0.00	-2.89
183	1.52	3.83	0.00	3.15
191	1.22	2.50	0.01	2.44

Cuadro 8.4: I de moran variable dependiente PRICE

Se obtienen 43 registros con I de moran local significativo teniendo en cuenta que existen barrios donde solo existe 1 registro.

## 9. Análisis confirmatorio

### 9.1. Modelo Clásico

	(1)	PRICE (2)	(3)
NROOM	0.469		
DWELL	7.024***	7.570***	10.817***
NBATH	6.738***	7.422***	9.139***
PATIO	7.036***	7.304***	9.565***
FIREPL	10.405***	10.888***	13.739***
AC	9.348***	9.368***	9.251***
BMENT	6.304**	6.126**	
NSTOR	-1.040		
GAR	2.564	2.724*	
AGE	-0.001		
CITCOU	11.532***	11.459***	11.676***
LOTSZ	0.052***	0.053***	
SQFT	0.090		
Constant	4.768	5.210*	9.404***
R <sup>2</sup>	0.732	0.732	0.689
Adjusted R <sup>2</sup>	0.715	0.719	0.681
AIC	1619.954	1612.6	1636.843
BIC	1670.069	1649.365	1663.581
Valor p de significancia *p<0.1; **p<0.05; ***p<0.01			

Cuadro 9.1: Resultados modelos clásicos

En la tabla anterior se muestra en la primera columna el modelo 1 con todas las variables de la base de datos, en la segunda columna el modelo 2 con el mejor modelo según método de stepwise y en la tercera columna el modelo 3 sin ninguna de las variables no significativas.

Al ver los resultados del modelo clásico se observó como el modelo, según indica el p-value es significativo estadísticamente con un nivel de confianza del 99 %, sin embargo tenemos variables las cuales no son significativas como el número de historias, la edad, el número de garajes y el espacio interior no son signi-

ficativos, se decide no tomarlos para el modelo pero conservando el numero de garajes ya que según varios estudios incluidos el documento donde se trabajaron los datos originalmente Dubin (1992) se decidió conservar teniendo una mayor significancia eliminando las demás y ayuda a obtener un mayor  $R^2$ , observamos que al eliminarlas obtenemos unos  $R^2$  mayores, criterios de Akaike y Bayesiano con test de error menor y significancia de casi todas las variables al 99 % sin embargo al eliminar las variables significativas solo al 95 % se obtiene un  $R^2$  Ajustado menor y criterios con test de error mayor por lo tanto se decide seleccionar el modelo 2 de la tabla presentada.

Al realizar el test de Moran del modelo con la matriz de peso seleccionada anteriormente obtenemos un índice de Moran de 0.1329, con p-value de 0.00012 y un  $Z(I)$  de 3.671577 por lo tanto se rechaza la hipótesis de ausencia de autocorrelación espacial por lo tanto los residuos están correlacionados positivamente por lo que podemos decir que cada valor esta rodeado por valores similares.

### 9.1.1. Pruebas de los supuestos modelo 2

Supuesto	Test	Estadístico	p-value
Normalidad	Shapiro-Wilk	0.98693	0.05171
Heterocedasticidad	Breush-Pagan	0.93115	$8.071 \times 10^{-7}$

Cuadro 9.2: Pruebas de supuestos Modelo clásico 2

El supuesto de normalidad evaluado a partir del test de Shapiro-Wilk indica que los residuos si siguen una distribución normal a nivel de significancia de 95 % en el caso de la heterocedasticidad mediante el test de Breusch-Pagan como si es significativo, nos dice que hay heterocedasticidad pero no se corrige debido a que el modelo es el clasico y el para el presente trabajo se necesita modelar la dependencia y autocorrelación espacial y el test de Breush Pagan no es fiable cuando hay autocorrelación espacial en los datos.

### 9.1.2. Mapa de residuos

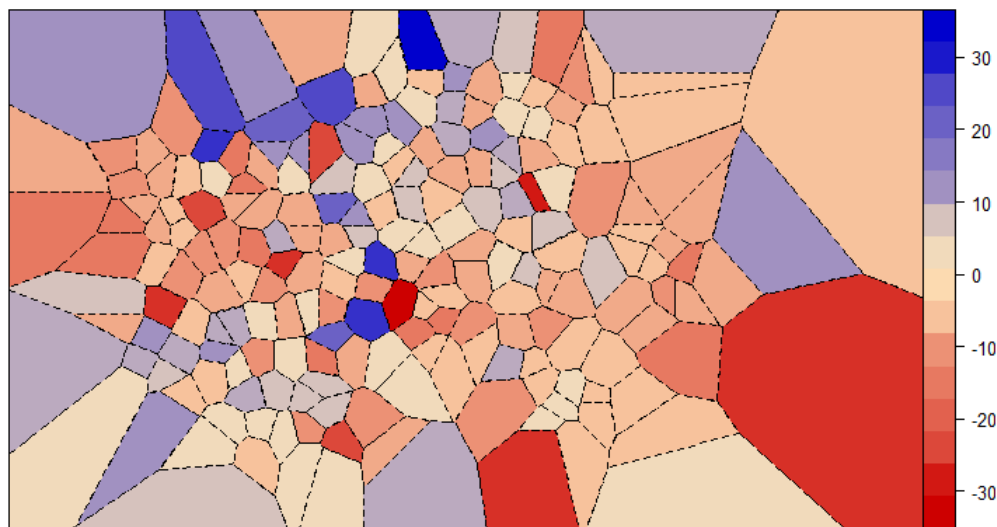


Figura 9.1: Mapa de residuos modelo clásico 2

El mapa de los residuos obtenidos de la estimación del modelo sugieren aunque no de manera muy clara la presencia de esquema de dependencia espacial positiva esta se aclara con los índices sacados en donde podemos decir que si se sigue un esquema de dependencia espacial positiva pero no muy clara

### 9.1.3. Multiplicadores de Lagrange

	statistic.LMerr	parameter.df	p.value
LMerr	10.705429	1.000000	0.001068
LMlag	19.569787	1.000000	0.000010
RLMerr	0.817595	1.000000	0.365884
RLMlag	9.681953	1.000000	0.001861
SARMA	20.387382	2.000000	0.000037

Cuadro 9.3: Multiplicadores de lagrange modelo clásico 2

Se evidencia en la prueba de multiplicadores de Lagrange que los estadísticos tanto el de error como el lag son significativos y al revisar el resultado de los estadísticos robustos se ve que solo es significativo el robusto para el retardo espacial por lo cual según el diagnostico espacial "DIAGNOSTICS FOR SPATIAL DEPENDENCE"decimos que el modelo más adecuado es el de el modelo de retardo espacial.

## 9.2. Modelo Spatial Lag

Modelo spatial lag	
(Intercept)	−1,66
DWELL	7,06 ***
NBATH	7,05 ***
PATIO	5,93 **
FIREPL	9,24 ***
AC	8,23 ***
BMENT	6,09 **
GAR	3,15 *
CITCOU	8,93 ***
LOTSZ	0,04 **
$\rho$	0,25 ***
Nagelkerke pseudo-R-squared	0,75293
Log Likelihood	−786,66
AIC (Spatial model)	1597,33
Significancia:	
*** $p < 0,001$ ; ** $p < 0,01$ ; * $p < 0,05$	

Cuadro 9.4: Modelo spatial lag

El parámetro  $\rho$  es positivo y significativo por lo cual rechazamos la hipótesis nula de que la inclusión del retardo de la variable dependiente no es necesario se rechaza, además de que con este modelo obtenemos un AIC menor

Supuesto	Test	Estadístico	p-value
Normalidad	Shapiro-Wilk	0.98881	0.1019
Heterocedasticidad	Breush-Pagan	39.927	$7.833 \times 10^{-6}$

Cuadro 9.5: Supuestos modelo spatial lag

Podemos observar que hay normalidad pero también hay heterocedasticidad

## 9.3. Modelo Spatial Error

Modelo Spatial Error	
(Intercept)	4,97
DWELL	8,66 ***
NBATH	8,06 ***
PATIO	5,52 *
FIREPL	8,39 ***
AC	8,97 ***
BMENT	6,50 **
GAR	2,37
CITCOU	11,63 ***
LOTSZ	0,04 **
$\lambda$	0,39 ***
Nagelkerke pseudo-R-squared	0,74653
Log Likelihood	−789,34
AIC (Spatial model)	1602,68
Significancia:	
*** $p < 0,001$ ; ** $p < 0,01$ ; * $p < 0,05$	

Cuadro 9.6: Modelo Spatial Error

El parámetro  $\lambda$  es positivo y significativo por lo cual se podría decir que es modelo valido y tendría mejor inferencia que el modelo clásico mas sin embargo vemos que tiene un pseudo  $R^2$  menor y un AIC mayor que en el modelo del rezago espacial.

Supuesto	Test	Estadístico	p-value
Normalidad	Shapiro-Wilk	0.98318	0.01356
Heterocedasticidad	Breush-Pagan	36.209	$3.641 \times 10^{-5}$

Cuadro 9.7: Supuestos modelo spatial error

Podemos observar que no hay normalidad por lo tanto el test de heterocedasticidad no es muy fiable (no se pudo corregir con transformaciones)

## 9.4. Modelo SLX

Modelo SLX	
(Intercept)	17,48 **
DWELL	7,34 ***
NBATH	6,77 ***
PATIO	4,28
FIREPL	9,35 ***
AC	9,17 ***
BMENT	6,31 **
GAR	2,89 *
CITCOU	10,57 **
LOTSZ	0,05 **
lag.DWELL	-5,60
lag.NBATH	-8,09 *
lag.PATIO	7,35
lag.FIREPL	16,78 ***
lag.AC	2,47
lag.BMENT	-4,16
lag.GAR	2,84
lag.CITCOU	-2,88
lag.LOTSZ	0,07
R <sup>2</sup>	0,7316
Adj. R <sup>2</sup>	0,7573
Sigma	10,37
Statistic	37,05
P Value	0,00
DF	19
Log Likelihood	-775,36
AIC	1590,72
BIC	1657,56

Significancia:

\*\*\* $p < 0,001$ ; \*\* $p < 0,01$ ; \* $p < 0,05$

Cuadro 9.8: Modelo SLX

Se observa que salvo el rezago de la variable de chimenea y el numero de baños no son significativas en el modelo pero se obtiene unos valores de  $R^2$  mayores y AIC menores que en el del rezago espacial pudiendo decir que este resultaría ser mejor modelo que los anteriores

Supuesto	Test	Estadístico	p-value
Normalidad	Shapiro-Wilk	0.99584	0.8439
Heterocedasticidad	Breush-Pagan	34.569	0.01071

Cuadro 9.9: Supuestos modelo SLX

Podemos observar que hay normalidad pero también hay heterocedasticidad

## 9.5. Modelo Spatial Durbin

Modelo Durbin espacial	
(Intercept)	15,75 **
DWELL	7,63 ***
NBATH	7,01 ***
PATIO	4,15
FIREPL	8,84 ***
AC	9,08 ***
BMENT	6,36 **
GAR	2,82 *
CITCOU	10,52 ***
LOTSZ	0,04 **
lag.DWELL	-6,60
lag.NBATH	-8,78 **
lag.PATIO	5,65
lag.FIREPL	13,62 **
lag.AC	0,74
lag.BMENT	-5,11 )
lag.GAR	1,73
lag.CITCOU	-4,50
lag.LOTSZ	0,06
$\rho$	0,17
R <sup>2</sup>	21
Log Likelihood	-774,08
AIC (Linear model)	1590,72
AIC (Spatial model)	1590,15

Significancia:

\*\*\* $p < 0,001$ ; \*\* $p < 0,01$ ; \* $p < 0,05$

Cuadro 9.10: Modelo Durbin espacial

Vemos como de las variables con rezago solo son significativas las de numero de baños y de chimenea y vemos como ciertas variables dejan de ser significativas y  $\rho$  no es muy alto ni significativo.

Supuesto	Test	Estadístico	p-value
Normalidad	Shapiro-Wilk	0.99362	0.01258
Heterocedasticidad	Breush-Pagan	0.99362	0.51

Cuadro 9.11: Supuestos modelo Durbin espacial

Podemos observar que no hay normalidad por lo tanto el test de heterocedasticidad no es muy fiable



## 9.6. Modelo GLM

Modelo GLM	
(Intercept)	2,92 ***
W1PRICE	0,01 ***
WDWELL	0,01
WNBATH	-0,27 ***
WPATIO	0,09
WFIREPL	0,19 *
WAC	-0,07
WBMENT	-0,20 *
WGAR	-0,12 *
WCITCOU	0,15 **
WL0TSZ	0,00 ***
X	0,00
Y	0,01 (0,01)
WX	-0,01 *
WY	0,00 (0,01)
Pseudo R <sup>2</sup>	0,4858
AIC	<i>inf</i>
BIC	<i>inf</i>
Significancia:	
*** $p < 0,001$ ; ** $p < 0,01$ ; * $p < 0,05$	

Cuadro 9.12: Modelo GLM

Vemos que casi ninguno de los parámetros espaciales de las coordenadas es significativo solo un poco el rezago de la coordenada x, mientras que vemos que en este caso las variables mas significativas son el tamaño del lote, el numero de baños y el rezago de la variable dependiente de precio.

Supuesto	Test	Estadístico	p-value
Normalidad	Shapiro-Wilk	0.96412	$3.8 \times 10^{-5}$
Heterocedasticidad	Breush-Pagan	57.917	$2.711 \times 10^{-7}$

Cuadro 9.13: Supuestos modelo GLM

Podemos observar que no hay normalidad por lo tanto el test de heterocedasticidad no es muy fiable, igualmente se realizo el modelo con rezagos de ordenes superiores para las variables pero el modelo de mayor Pseudo R<sup>2</sup> que se obtuvo fue con los rezagos de primer orden, de igual manera en ninguno se obtuvo normalidad en los residuos por lo cual se descarta el modelo.

## 9.7. Modelo SARAR

	Model SARAR
(Intercept)	-1,51
DWELL	7,34 ***
NBATH	7,27 ***
PATIO	5,84 **
FIREPL	9,08 ***
AC	8,35 ***
BMENT	6,21 **
GAR	3,08 *
CITCOU	9,21 ***
LOTSZ	0,04 **
$\rho$	0,23 **
$\lambda$	0,06
Pseudo R <sup>2</sup>	0,7530
Log Likelihood	-786,61
AIC (Linear model)	1612,60
AIC (Spatial model)	1599,21

Significancia:

\*\*\* $p < 0,001$ ; \*\* $p < 0,01$ ; \* $p < 0,05$

Cuadro 9.14: Modelo SARAR

Se observa que solo las variables conservan casi la misma significancia que en el modelo clásico y mientras que el rezago de la variable dependiente es significativo el rezago del error no lo es.

Supuesto	Test	Estadístico	p-value
Normalidad	Shapiro-Wilk	0.98811	0.079
Heterocedasticidad	Breush-Pagan	38.971	$1.166 \times 10^{-5}$

Cuadro 9.15: Supuestos modelo SARAR

Podemos observar que hay normalidad pero también hay heterocedasticidad

Considerando que tenemos problemas de heterocedasticidad en todos los modelos se deciden utilizar las opciones de modelos espaciales lineales heterocedasticos, alternativas las cuales tienen en cuenta una heterocedasticidad y autocorrelación consistente HAC Arbia (2014), mediante las funciones de `gstslshet` y `gstslshac` de la librería `sphet`, Piras (2018).

## 9.8. Modelo espacial lineal heterocedastico SARAR

Model heterocedastic SARAR	
(Intercept)	-3,182
DWELL	6,977 ***
NBATH	6,985 ***
PATIO	5,617 *
FIREPL	8,861 ***
AC	7,995 ***
BMENT	6,096 **
GAR	3,242
CITCOU	8,397 ***
LOTSZ	0,039
$\rho$	0,009
$\lambda$	0,305 ***
Wald test	
statistic	5,4104
p-value	0,020017
Significancia:	
*** $p < 0,001$ ; ** $p < 0,01$ ; * $p < 0,05$	

Cuadro 9.16: Modelo SARAR

Los resultados de la estimación muestran que sigue siendo significativa la relación de la variable del precio con las demás, pero al considerar la heterocedasticidad vemos como el tamaño del lote y numero de garajes dejan de ser significativos, el test de Wald rechaza la hipótesis de que los dos parámetros  $\rho$  y  $\lambda$  son cero.

Normalidad, mediante el test de Shapiro Wilk se obtiene un valor W de 0.98 con un valor de p-value de 0.04 de donde se rechaza la hipótesis nula de que los datos son normales con una confiabilidad del 5 por ciento por lo tanto los residuos del modelo no siguen una distribución normal.

Shapiro Wilk	
Estadístico	0.98
p-value	0.04

Cuadro 9.17: Resultados Normalidad

Finalmente estimamos el modelo usando la función para hallar un modelo con autocorrelación y heterocedasticidad espacial constante del modelo SARAR, considerando un kernel Triangular y las distancias entre los puntos de las viviendas mediante la función `stslshac`.

## 9.9. Modelo HAC espacial

Model spatial HAC	
(Intercept)	−3,156
DWELL	6,955 ***
NBATH	6,985 ***
PATIO	5,617 *
FIREPL	8,861 ***
AC	7,995 ***
BMENT	6,096 **
GAR	3,242
CITCOU	8,397 ***
LOTSZ	0,039 *
$W_y$	0,306 ***
Pseudo $R^2$	0,756

Significancia:

\*\*\* $p < 0,001$ ; \*\* $p < 0,01$ ; \* $p < 0,05$

Cuadro 9.18: Modelo espacial HAC

Obtenemos resultados similares con mas significanos en algunas variables pero normalidad en los residuos por esta razon **se decide elegir este como el mejor modelo** a continuación se presenta las pruebas de los supuestos del modelo elegido.

- Normalidad, mediante el test de Shapiro Wilk se obtiene un valor W de 0.98 con un valor de p-value de 0.1834 de donde se acepta la hipotesis nula de que los residuos siguen una distribución normal.

Shapiro Wilk	
Estadístico	0.98
p-value	0.1364

Cuadro 9.19: Resultados Normalidad

- Heterocedasticidad: al graficar los resultados del modelo vs los residuos obtenemos la siguiente gráfica donde se observa homocedasticidad, de igual forma se aclara que la función `stslshac` ya tiene en cuenta la heterocedasticidad en el calculo de los parametros.

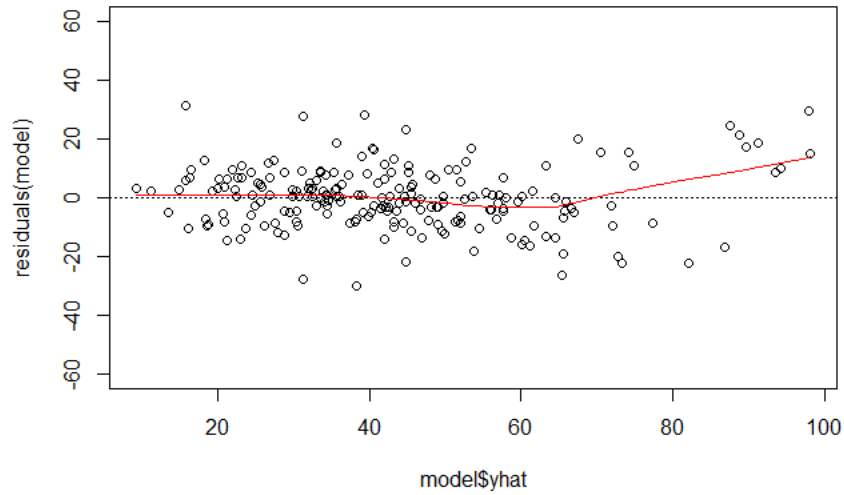


Figura 9.2: Gráfico determinación Heterocedasticidad

- Multicolinealidad: Al evaluar la multicolinealidad obtenemos los siguientes resultados, en donde se calcula el factor de la varianza de la inflación (VIF) de los resultados del modelo obtenemos valores muy bajos por lo que podemos observar que no hay multicolinealidad, solo se da esto cuando estos valores son mayores a 10.

DWELL	NBATH	PATIO	FIREPL	BMENT	AC	GAR	CITCOU	LOTSZ
1.72	1.35	1.161	1.313	1.11	1.24	1.31	1.51	2.03

Cuadro 9.20: Resultados multicolinealidad

## 10. Análisis de impactos del modelo

En el análisis de impactos se decide cambiar el valor de la variable dependiente en una de las regiones cuyo precio esta entre los mas bajos y esta ubicado hacia el centro de los demas, se selecciona el punto 16 y se cambio su precio de 3.5 a 68 (un valor alto pero no de los mas altos con respecto a los demas de la base de datos) en el precio de esta vivienda y analizar los impactos del modelo espacial HAC.

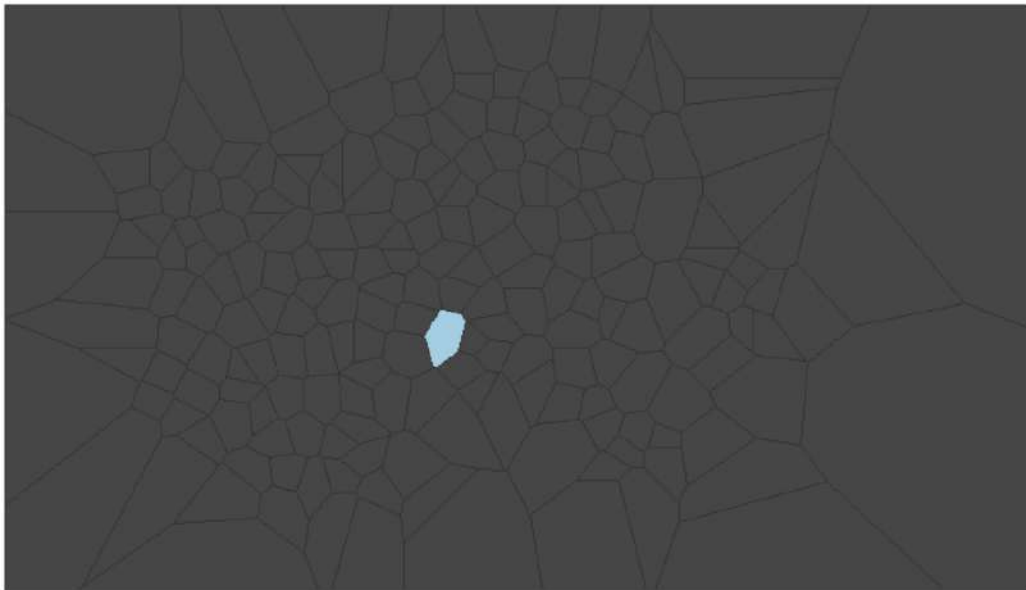


Figura 10.1: Punto 16, seleccionado para realizar el análisis de impactos

Se estudian cuales viviendas son afectadas por el cambio, se utilizo el modelo seleccionado para mirar la diferencia que hubo entre las predicciones realizadas por el modelo de los datos originales y las predicciones realizadas con la base de datos y el valor de el punto 16 para ver como impacta este cambio en las demás, las 10 viviendas mas afectadas son las siguientes.

STATION	Diferencia
85	5.25
94	4.33
89	3.74
14	3.58
16	3.44
177	3.32
48	3.31
131	3.09
72	2.94
1	2.91

Cuadro 10.1: Viviendas mas afectadas, análisis de impactos

La región la cual resulto mas afectada fue la región 85 la cual vario su precio en \$5.250 dolares, se ve que los puntos más afectados positivamente (aumentaron de precio) vienen siendo las mas cercanas a esta aunque unas con unas pocas variaciones negativas, las que se ven afectadas negativamente son las que están mas alejadas, mas hacia la parte externa de la zona de estudio también se ve como las viviendas que no resultaron afectadas son pocas en comparación con el total, en los mapas a continuación se observa a la derecha los cambios de como fueron afectados los precios de la vivienda, positivo, negativo o neutralmente, en el de la derecha se observa mas detalladamente cuanto fue el impacto que tuvieron en el precio los demás puntos de estudio, se observa que los mas cercanos son los mas afectados positivamente y los mas lejanos afectados negativamente, mayormente el impacto fue positivo en las viviendas de la región.

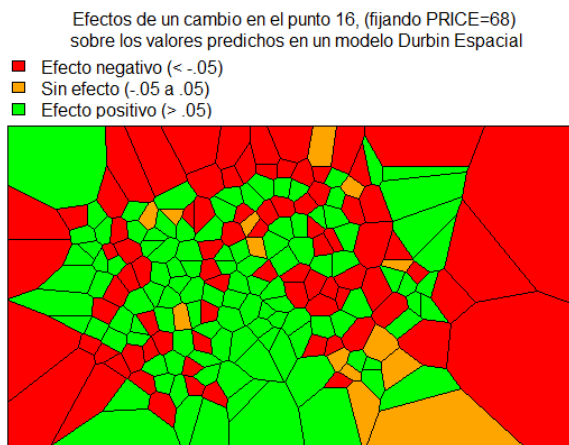


Figura 10.2: Impactos

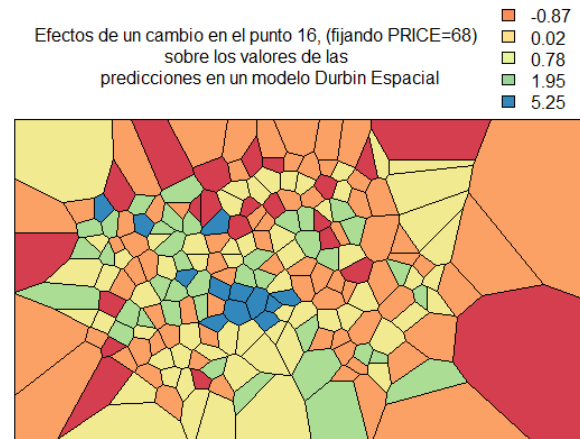


Figura 10.3: Efectos de los impactos

Al analizar la significancia de los impactos vemos como la mayoría de variables son significativas y positivas en los impactos que tiene el modelo de HAC espacial, con coeficientes de los impactos directos e indirectos similares a los obtenidos por el modelo los cuales son explicados a continuación A continuación obtenemos

la tabla de impactos en el cual analizamos el impacto directo que registra la suma de impactos en la región que experimenta un cambio, el impacto indirecto que captura la suma de los impactos debido a los cambios en las otras regiones y el impacto total que es la suma del impacto directo e indirecto, es decir el impacto de una región frente a su cambio y el percibido debido al cambio de otras regiones. Dall’erba y Llamosas-Rosas (2014)

	Directo	Indirecto	Total
DWELL	7.08	2.93	10.02
NBATH	7.09	2.94	10.03
PATIO	5.73	2.37	8.11
FIREPL	9.04	3.75	12.79
BMENT	6.20	2.57	8.76
AC	8.13	3.37	11.50
GAR	3.31	1.37	4.67
CITCOU	8.54	3.54	12.08
LOTSZ	0.04	0.02	0.06

Cuadro 10.2: Tabla de impactos modelo spatial HAC para el precio de la vivienda

En la anterior tabla podemos obtener como resultado que un cambio en todas las variables lleva a un cambio positivo en el precio, se ve como las variables que mas impactan al tener cambios en estas son el tener chimenea o aire acondicionado algo que se puede suponer en una zona lejana del trópico en la cual se tienen estaciones y se tiene que estas condiciones de aclimatación de las viviendas afectan el el precio, también tenemos como el cambiar a que la vivienda se encuentre dentro del condado de Baltimore, es decir fuera de la ciudad de Baltimore una especie de renta de segregación se tiene en cuenta ya que el condado de Baltimore se consideraría como las afueras de la ciudad haciendo un símil de lo que es La Calera o Municipios aledaños de Bogotá en donde las viviendas que se concentran ahí al estar alejadas de la ciudad y tener ciertas diferencias y ventajas con esta se presenta un cambio positivo en el precio de las viviendas, tenemos igualmente que en el posterior análisis de la vivienda vemos como lo que mas afectan vienen siendo las variables dummy, adicional tenemos un impacto positivo en el numero de baños de la casa y finalmente la que se ve tiene menos impacto en el precio es el tamaño del lote.



## 11. Análisis del modelo

El modelo para explicar el precio de la vivienda en la ciudad de Baltimore Maryland resultante teniendo en cuenta la autocorrelación espacial es el siguiente.

$$\begin{aligned} PRICE = & -3,156 + (0,306)WPRICE + (6,955)DWELL + (6,985)NBATH + \\ & (5,617)PATIO + (8,861)FIREPL + (7,995)AC + (6,096)BMENT + (3,242)GAR + \\ & (8,397)CITCOU + (0,039)LOTSZ \end{aligned} \quad (11.1)$$

De donde R es una matriz no aleatoria. Se calculo el HAC espacial con una funcion de kernel tipo triangular necesaria para el calculo de los pesos, los resultados y analisis de los coeficientes de las variables es: Se observa que el tener o no chimenea (*FIREPL*) es lo que más influye en el valor del hogar por lo tanto al tener chimenea costaría \$8.861 dolares más, también vemos que si esta dentro del condado de Baltimore (*CITCOU*) hace que el valor aumente en \$8.397 dolares mas esto se podría suponer dado a que en las ciudades puede ser mayor el costo de la vivienda, igualmente puede aumentar un valor similar si vemos que tiene aire acondicionado (*AC*) en unos \$7.995 dolares lo cual se puede dar dado a que en Baltimore las temperaturas pueden ser en ciertas ocasiones altas, también similarmente las demás variables Dummy como lo son si es una casa independiente (*DWELL*) si tiene patio o no (*PATIO*), o si tiene sotano (*BMENT*) aumenta al rededor de \$6.000 a \$7.000 dolares, se observa en las variables continuas teóricamente por cada baño que tenga (*NBATH*) aumentaría el precio en \$6.985 y por cada espacio para garaje (*GAR*) unos \$3.242 dolares, con respecto al tamaño del interior de la vivienda se tiene que aumenta \$39 dolares por cada 100 pies cuadrados que tenga mas de área.

## 12. GWR

El método GWR se comporta de forma más eficiente porque considera las variaciones espaciales de las relaciones entre las variables y obtenemos parámetros locales que reflejan más exactamente su influencia en el precio de la vivienda en algunos sectores como se muestra en los siguientes mapas .

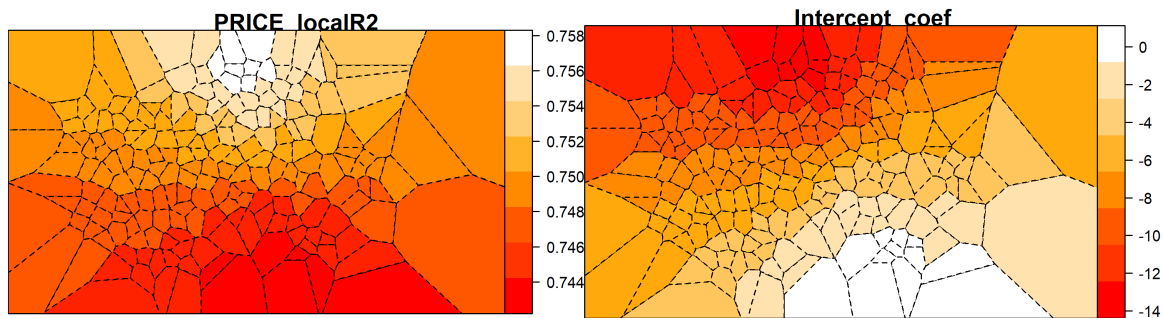


Figura 12.1: R2 local y coeficiente de intercepto

Como se observa en la anterior figura el R2 local varia de 0.744 a 0.758 en donde el mejor ajuste se encuentra hacia el norte de la region de estudio.

El modelo GWR no sólo suele tener un mayor poder explicativo que el OLS, sino que además permite profundizar en el análisis local, ya que se dispone de una ecuación de regresión para cada observación, ajustada localmente. Así es posible conocer dónde el modelo tiene un mayor o menor ajuste (R2) y cómo cambia la relación entre las variables en el espacio (coeficientes de los predictores) y con qué significación estadística (Gutiérrez-Puebla, Garcia-Palomares y Daniel-Cardozo (2012)).

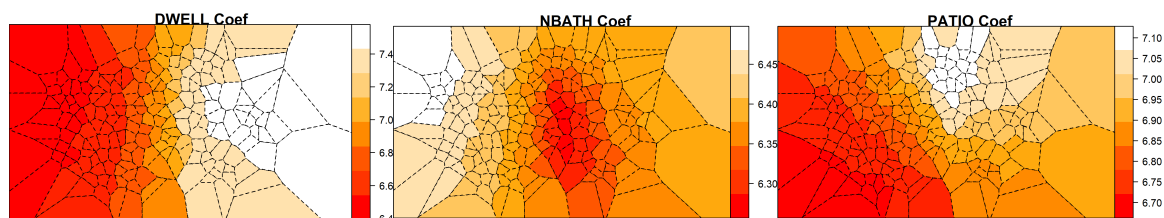


Figura 12.2: Variación de los coeficientes en el espacio para las variables DWELL, NBATH, PATIO

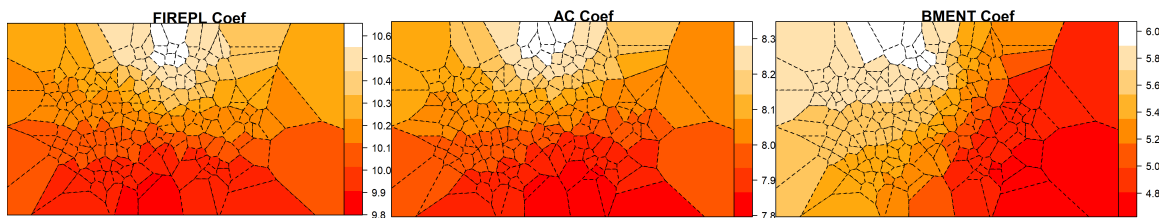


Figura 12.3: Variación de los coeficientes en el espacio para las variables FIREPL, AC, BMENT

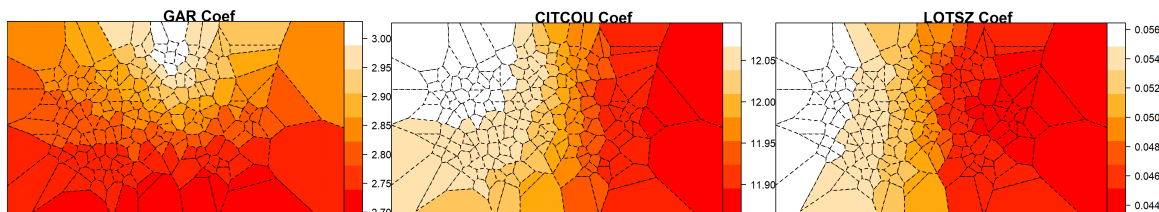


Figura 12.4: Variación de los coeficientes en el espacio para las variables GAR, CITCOU, LOTSZ

Como se ha comentado anteriormente el conocimiento de la variación de cada coeficiente en el espacio (elasticidades) permite extraer conclusiones más realistas, basadas en el análisis local como por ejemplo la variable DWELL que nos indica una clara transición de este a oeste del parámetro del modelo en donde hacia la zona este suele aportar mas (DWELL) en el valor de la vivienda, el coeficiente para el numero de baños es mayor hacia las periferias y tiende a ser menos significativo hacia el centro de la región, la importancia de tener o no un patio (PATIO) se refleja en un coeficiente mayor en la zona norte de la región de estudio, de forma similar poseer chimenea (CHIMENEA), aire acondicionado (AC) y garaje. El coeficiente de la variable sótano (BMENT) tiende a ser mas alto hacia la zona norte y norte-oeste de la región. En cuanto a (CITCOU) y (LOTSZ) los coeficientes locales mas altos se encuentran hacia la región este y reduciéndose a medida que se hace una transición a la región Oeste.

Estimación de GWR t- values, entre mas rojo indica una relación menos significativa

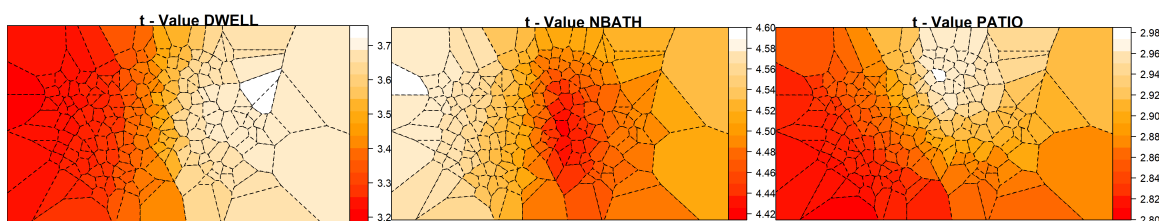


Figura 12.5: Significancia estadística para las variables DWELL, NBATH, PATIO

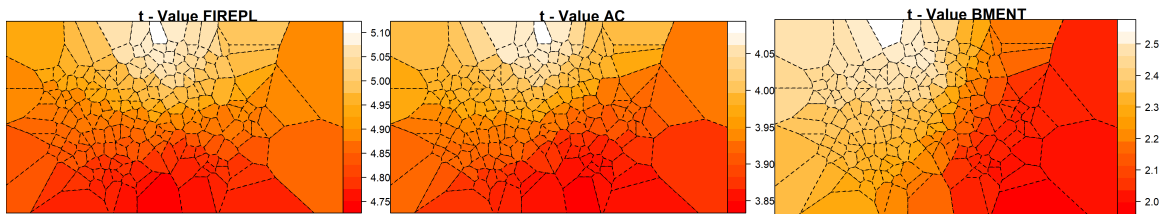


Figura 12.6: Significancia estadística para las variables FIREPL, AC, BMENT

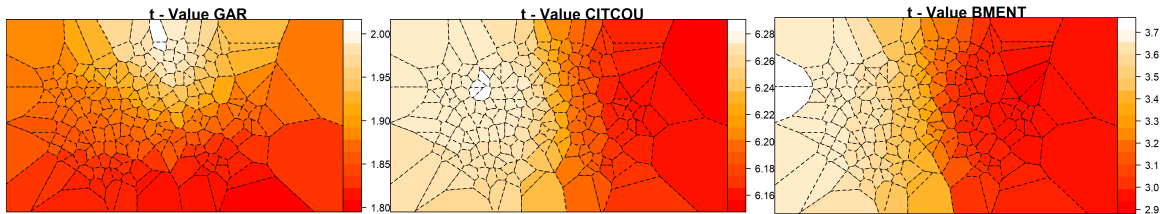


Figura 12.7: Significancia estadística para las variables GAR, CITCOU, LOTSZ

predicted and standard error

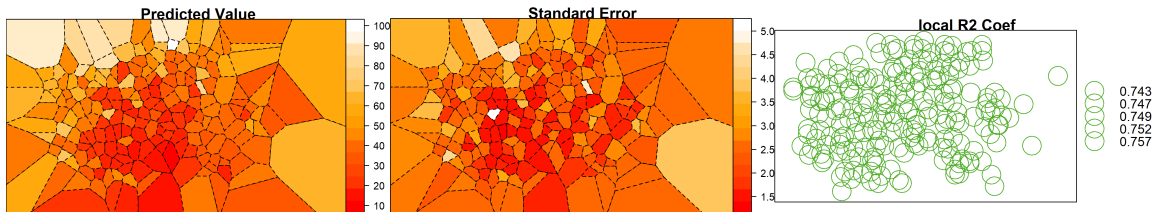


Figura 12.8: Prediccion de valores, error estandar y R2 local

## 13. Conclusiones

- Podemos concluir que las variables dummy juegan un papel muy importante a la hora de estimar el precio de la vivienda en la región de Baltimore, vemos como estas suelen ser las mas significativas y las de mayor coeficiente en el modelo seleccionado
- Las características las cuales dependen de la edad y tamaño de la vivienda no siempre son las mejores a la hora de estimar el valor de un inmueble en este caso, y se vio como otras variables como lo son el numero de baños y/o de garajes continúan pueden repercutir mas en un modelo.
- Podemos decir que las variables mas importantes a la hora de estimar un precio de una vivienda tienen que ver con respecto a como es la población y lugar en el cual se requiera hacer el estudio, en Baltimore por ejemplo se da mayor valor a casas que cuenten con chimenea, esta quizá por motivos sociales de la población, también con respecto a si se encuentra dentro del condado como tal y no a las afueras, esto puede ser dado a motivos de transporte o accesibilidad por motivos laborales, familiares, comerciales, etc, y finalmente el accesos a aire acondicionad, esto puede deberse a las temperaturas en las que se ven en la región.
- Se tiene que tener en cuenta la dependencia espacial que tienen los diferentes fenómenos en los datos en términos de una o mas variables ya que puede ser que las regiones vecinas afecten en medida el valor que toma el fenómeno en cuestión ya que si no se lo tiene no se va a llegar a la mejor estimación posible.
- Se tiene que tener en cuenta también la heterocedasticidad en los modelos para que por si no cumple esta se debe recurrir a otros métodos de la teoría, los cuales tienen en cuenta esta para la estimación del modelo.
- Observamos por el análisis de impactos como al presentar una autocorrelación espacial en los datos vemos como al variar una vivienda en su precio las mas afectadas en su mayoría van a ser las mas cercanas y las mas lejanas positiva y negativamente respectivamente.
- El modelo GWR posee un buen nivel explicativo superior a los demás, también hay que proporcionarle relevancia porque nos permite profundizar el análisis a nivel local y si contáramos con la información

determinar en que barrios de la ciudad son mas relevantes unas características de la vivienda al momento de determinar su precio.

# Referencias

- A., Eric E. Andrade (2020). *Polígonos de Thiessen*. | CivilGeeks.com. URL: <https://civilgeeks.com/2011/09/24/poligonos-de-thiessen/> (visitado 18 de jun. de 2020) (vid. pág. 10).
- Anselin, Luc y Anil K Bera (1998). «Introduction to spatial econometrics». En: *Handbook of applied economic statistics* 237 (vid. pág. 2).
- Arbia, Giuseppe (2014). *A Primer for Spatial Econometrics: With Applications in R*, pág. 246. ISBN: 1137317949. DOI: 10.1057/9781137317940. URL: <https://books.google.com/books?id=P16oBAAQBAJ%7B%5C&%7Dpgis=1> (vid. pág. 52).
- Baronio, Alfredo, Ana Vianco y Christian Rabanal (2012). «Una introducción a la econometría espacial: dependencia y heterogeneidad». En: *Catedra de econometria* (vid. pág. 13).
- Bohorquez, M (2009). «Estadística espacial». En: *Universidad Nacional de Colombia* 76 (vid. págs. 12, 19).
- Dall'erba, Sandy e Irving Llamosas-Rosas (2014). «Spatial Policy for Growth and Equity». En: *Handbook of Regional Science*. Ed. por Manfred M. Fischer y Peter Nijkamp. Springer Berlin Heidelberg: Berlin, Heidelberg, págs. 353-371. ISBN: 978-3-642-23430-9. DOI: 10.1007/978-3-642-23430-9\_21. URL: [https://doi.org/10.1007/978-3-642-23430-9\\_21](https://doi.org/10.1007/978-3-642-23430-9_21) (vid. pág. 58).
- Dubin, Robin A. (1992). «Spatial autocorrelation and neighborhood quality». En: *Regional Science and Urban Economics* 22(3), págs. 433-452. ISSN: 01660462. DOI: 10.1016/0166-0462(92)90038-3 (vid. pág. 46).
- Duque, Juan Carlos, Hermilson Velásquez Ceballos y Jorge Agudelo (2011). «Infraestructura pública y precios de vivienda: una aplicación de regresión geográficamente ponderada en el contexto de precios hedónicos». En: *Ecos de Economía: A Latin American Journal of Applied Economics* 15(33), págs. 99-122 (vid. pág. 4).
- Gutiérrez-Puebla, Javier, JC Garcia-Palomares y O Daniel-Cardozo (2012). «Regresión Geográficamente Ponderada (GWR) y estimación de la demanda de las estaciones del Metro de Madrid». En: *Ponencia presentada en el XV Congreso Nacional de Tecnologías de la Información Geográfica*, págs. 19-21 (vid. pág. 60).
- Kelejian, Harry H. e Ingmar R. Prucha (2010). «Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances». En: *Journal of Econometrics* 157(1), págs. 53-67. ISSN: 03044076. DOI: 10.1016/j.jeconom.2009.10.025. URL: <http://dx.doi.org/10.1016/j.jeconom.2009.10.025> (vid. pág. 19).
- Kelejian, Harry H e Ingmar R Prucha (2007). «HAC estimation in a spatial framework». En: *Journal of Econometrics* 140, págs. 131-154. DOI: 10.1016/j.jeconom.2006.09.005. URL: [www.elsevier.com/locate/jeconom](http://www.elsevier.com/locate/jeconom) (vid. pág. 19).
- López-González, Emelina y Marcos Ruiz-Soler (2011). «Análisis de datos con el Modelo Lineal Generalizado. Una aplicación con R». En: *Revista española de pedagogía*, págs. 59-80 (vid. pág. 17).
- Moreno, Rosina y Esther Vayá (2002). «Econometría espacial: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas». En: *Investigaciones Regionales - Journal of Regional Research* 2002(1), págs. 83-106. ISSN: 1695-7253 (vid. págs. 10, 11).

- «Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC» (s.f.). En: () (vid. pág. 10).
- Pérez Pineda, Jorge A (2006). «Econometría espacial y ciencia regional». En: *Investigación económica* 65(258), págs. 129-160 (vid. pág. 18).
- Piras, Gianfranco (2018). *Estimation of Spatial Autoregressive Models with and without Heteroscedasticity*. R package version 1.7 (vid. págs. 19, 52).
- Rdocumentation (2020). *listw.select function | R Documentation*. URL: <https://www.rdocumentation.org/packages/adespatial/versions/0.2-0/topics/listw.select> (visitado 22 de jun. de 2020) (vid. pág. 28).
- Serrano, Rosina Moreno y Esther Vayá Valcarce (2000). *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial*. Vol. 44. Edicions Universitat Barcelona (vid. pág. 15).
- Yrigoyen, Coro Chasco (2003). *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*. Dirección General de Economía y Planificación (vid. págs. 13, 14, 17, 18).
- Yu, Danlin, Yehua Dennis Wei y Changshan Wu (2007). «Modeling spatial dimensions of housing prices in Milwaukee, WI». En: *Environment and Planning B: Planning and Design* 34(6), págs. 1085-1102 (vid. pág. 6).

### 13.1. Anexos

- Carpeta baltimore con
  - Base de datos original baltim.shp
  - Shapefile filtrado sin valores que causan anomalía baltim\_clear.shp
- Código PesosAEDE.R en lenguaje R en el cual se desarrollo la parte de la selección de matrices de pesos y el AEDE
- Código Modelos.R en lenguaje R en el cual se desarrollo la parte de la selección, evaluación y análisis de los modelos los impactos