



**UNIVERSIDAD SAN FRANCISCO DE QUITO**

**Colegio de Posgrados**

**Análisis Geo-estadístico de la delincuencia  
para Quito Urbano usando GeoDa**

**Eddy César de Jesús Coba Rubio**

**Carlos Eugenio Balseca Carrera**

**Carlos Jiménez Mosquera, PhD., Director de Tesis**

**Tesis de grado presentada como requisito para la obtención  
del título de Magister en Matemáticas Aplicadas**

**Quito, abril de 2015**

**Universidad San Francisco de Quito**

**Colegio de Posgrados**

**HOJA DE APROBACIÓN DE TESIS**

**Análisis Geo-estadístico de la delincuencia para Quito Urbano usando  
GeoDa**

**Eddy César de Jesús Coba Rubio**

**Carlos Eugenio Balseca Carrera**

Carlos Jiménez Mosquera, Ph.D.  
Director de Tesis

---

Eduardo Alba, Ph.D.  
Miembro del Comité de Tesis

---

Carlos Jiménez Mosquera, Ph.D.  
Director de la Maestría en Matemáticas Aplicadas y  
Miembro del Comité de Tesis

---

Santiago Gangotena, Ph.D.  
Decano del Colegio de Ciencias e Ingeniería

---

Víctor Viteri Breedy, Ph.D.  
Decano del Colegio de Posgrados

---

Quito, abril de 2015

## ©Derechos de Autor

Por medio del presente documento certificamos que hemos leído la Política de Propiedad Intelectual de la Universidad San Francisco de Quito y estamos de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo de investigación quedan sujetos a lo dispuesto en la Política.

Así mismo, autorizamos a la USFQ para que realice la digitalización y publicación de este trabajo de investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma: \_\_\_\_\_

Nombre: Eddy César de Jesús Coba Rubio

C. I.: 1706387238

Firma: \_\_\_\_\_

Nombre: Carlos Eugenio Balseca Carrera

C. I.: 0501522940

Quito, abril de 2015

## **DEDICATORIA**

A Dios Padre de todos, a nuestras respectivas familias que con paciencia y dedicación nos han apoyado siempre.

## **AGRADECIMIENTO**

De manera especial deseamos dejar constancia de nuestro imperecedero agradecimiento a nuestro Maestro y Director de Tesis, Carlos Jiménez Mosquera, PhD.

A las siguientes personas que desde los diferentes estamentos del Gobierno, previa la firma de los respectivos Acuerdos de Confidencialidad hicieron posible la consecución de los datos estadísticos y mapas geográficos necesarios para la realización del presente trabajo:

Ing. Norman Dávila (INEC)

Ing. Viviana Guerra (ex CEASI)

Ing. Oswaldo Padilla (Universidad de las Fuerzas Armadas)

## RESUMEN

En la actualidad el interés por análisis de datos espaciales es cada vez mayor debido en gran parte a recientes desarrollos y contribuciones de varios investigadores interesados en esta temática, sobre todo en lo que tiene que ver con aplicaciones a ciudades. En este trabajo se presenta una implementación extensa de base de datos de Quito urbano tanto espaciales como transversales usando GoeDa. Esta implementación permitirá presentar el comportamiento de ciertas variables de interés cuyo análisis depende principalmente de los diferentes tamaños que poseen sus poblaciones, para ello se emplearán herramientas tales como las Estadísticas de Morán, que permiten representar gráficamente valores significativos de los índices locales y/o globales de Morán a través del ajuste empírico de Bayes. Buscándose principalmente, detectar si existe similitud de los valores de las variables entre zonas censales (autocorrelación espacial AE). Se propone un análisis de estos índices basado en los valores de las variables observadas en las diferentes zonas censales urbanas del Distrito Metropolitano de Quito. Siendo por tanto posible aplicarlo para cualquier tipo de variable socioeconómica, cuando sus observaciones correspondan a datos numéricos. Se analiza cada zona censal de tal manera que aquellas con reducido número de población no afecten los resultados si se comparan con otras zonas censales con poblaciones mayores. Además este estudio propuesto permitirá a través del tiempo detectar cómo evolucionan las variables analizadas.

## ABSTRACT

Today is of great interest to spatial data analysis, especially when it has to do with applications to cities. In this work an extensive database implementation urban Quito using both spatial and temporal GoeDa presented. This implementation will allow to present the behaviour of certain variables of interest whose analysis depends mainly on the different sizes that have their populations by tools such as statistics Moran, that can represent graphically values significant local and/or global indices will be used Moran through empirical Bayes smoothing. Searching mainly detect whether there is similarity of the values of the variables between parishes (spatial autocorrelation). An analysis of these indices based on the values of the observed variables in the different urban parishes in the city of Quito is proposed. Being thus possible to apply for any socio-economic variable, when their observations correspond to numerical data. Each parish so that those with limited number of population do not affect the results when compared with other parishes with larger populations is analysed. Furthermore, this proposed study will allow detecting how they evolve through the variables analysed time.

# Índice general

<b>1. INTRODUCCIÓN AL PROBLEMA</b>	<b>15</b>
1.1. Antecedentes . . . . .	16
1.2. El Problema . . . . .	17
1.3. Hipótesis . . . . .	20
1.4. Contexto y marco teórico . . . . .	20
1.4.1. El propósito del estudio . . . . .	20
1.4.2. El significado del estudio . . . . .	21
1.5. Presunciones de los autores del estudio . . . . .	21
1.6. Supuestos del estudio . . . . .	21
<b>2. ANÁLISIS ESPACIAL</b>	<b>22</b>
2.1. Dependencia Espacial . . . . .	23
2.2. El Proceso Espacial Autorregresivo . . . . .	25
2.3. El Proceso Espacial Autorregresivo de Generación de Datos . . . . .	29
2.4. Modelos de Regresión Espacial Autorregresivos . . . . .	32
2.5. Estimación por Máxima Verosimilitud . . . . .	34
2.5.1. Estimación de modelos SAR y SDM . . . . .	34
2.5.2. Estimación del modelo SEM . . . . .	39
<b>3. TRATAMIENTO A VALORES DE DATOS EMPLEA-</b>	
<b>DOS</b>	<b>42</b>
3.1. Descripción y tratamiento de valores de datos de las variables empleadas . . . . .	42
3.2. Análisis Exploratorio de Datos Espaciales (ESDA) . . . . .	46
3.2.1. Quantil Map . . . . .	46
3.2.2. Box Plot . . . . .	47

	10
3.2.3. Scatter Plot . . . . .	53
3.2.4. Linking and Brushing . . . . .	55
3.3. Exploración de Datos Multivariados . . . . .	55
3.3.1. 3-D Scatter plot . . . . .	55
3.4. Detección y Medición de Autocorrelación Espacial . . . . .	57
3.5. Índice de Morán en forma matricial . . . . .	58
3.6. Indicador Local de Autocorrelación Espacial (LISA) . . . . .	59
3.6.1. Ejemplo práctico para el cálculo del Índice Local de Autocorrelación Es- pacial (LISA) . . . . .	61
3.7. Análisis del Índice de Moran . . . . .	64
3.8. Análisis de Autocorrelación Espacial para tasas . . . . .	65
3.9. Índice de Moran Bivariado . . . . .	70
3.10. Regresión Espacial . . . . .	72
3.11. Modelo Matricial Propuesto . . . . .	83
3.12. Diagnóstico de la Regresión . . . . .	84
3.12.1. Modelo de Regresión de Superficie de Tendencia (Trend Surface Regression) . . . . .	84
3.12.2. Mapas de residuos y valores pronosticados . . . . .	101
3.12.3. Mapas de residuos . . . . .	101
3.12.4. Máxima Verosimilitud en el modelo de rezagos espaciales . . . . .	103
<b>4. MODELOS ESPACIALES PARA PANEL DE DATOS</b>	<b>107</b>
4.1. Estructuras de datos . . . . .	108
4.2. Clases y métodos para modelos de panel espacial . . . . .	109
4.3. Modelos Espaciales de Panel de Datos . . . . .	111
4.4. Implementación ML . . . . .	114
4.4.1. Modelo de Efectos Aleatorios . . . . .	114
4.4.2. Modelo de Efectos Fijos . . . . .	116
4.5. Pruebas . . . . .	119
4.5.1. Pruebas LM . . . . .	120
4.5.2. Prueba espacial de Hausman . . . . .	122
4.6. Análisis de Modelos Espaciales para Datos de Panel . . . . .	123
4.6.1. Modelo Pooling (Modelo1) . . . . .	123
4.6.2. Modelo Pooling (Modelo2) . . . . .	123
4.6.3. Modelo Pooling (Modelo3) . . . . .	124

4.6.4. Modelo Within (Modelo4) . . . . .	125
4.6.5. Modelo Within (Modelo5) . . . . .	125
4.6.6. Modelo Within (Modelo6) . . . . .	126
4.6.7. Modelo Random (Modelo7) . . . . .	126
4.6.8. Modelo Random (Modelo8) . . . . .	127
4.6.9. Modelo Random (Modelo9) . . . . .	127
4.6.10. Modelo Pooling (Modelo plm) . . . . .	128
4.6.11. Modelo Within (Modelo plm) . . . . .	129
4.6.12. Modelo Random (Modelo plm) . . . . .	130
4.6.13. Pruebas Sobre los Modelos . . . . .	130
<b>5. CONCLUSIONES</b>	<b>138</b>
<b>6. REFERENCIAS BIBLIOGRÁFICAS</b>	<b>143</b>
<b>7. ANEXOS</b>	<b>151</b>
7.1. Definición de términos . . . . .	151

# Índice de figuras

1.1. División territorial del D.M.Q. en 448 Zonas Censales Urbanas . . . . .	19
2.1. Moran Scatter plot relación entre observaciones . . . . .	28
3.1. Zonas censales con mayor tasa de robos año 2013 . . . . .	48
3.2. Zonas censales con viviendas que presentan hacimaniento . . . . .	49
3.3. Zonas censales con población no analfabeta . . . . .	50
3.4. Zonas censales con mayor presencia femenina . . . . .	51
3.5. Zonas censales de viviendas con personas que han migrado . . . . .	52
3.6. Gráfico de dispersión de mujeres vs robos . . . . .	54
3.7. Gráfico de dispersión de viviendas con migrantes vs robos . . . . .	54
3.8. Exploración de datos multivariados: Migrantes–Mujeres–Robos . . . . .	56
3.9. Proyección de datos sobre planos:Migrantes–Mujeres–Robos . . . . .	56
3.10. Ejemplo de Matriz de contiguidad . . . . .	61
3.11. Ejemplo de Matriz estandarizada por filas . . . . .	62
3.12. Ejemplo de cálculo de $Z_i$ . . . . .	62
3.13. Ejemplo de cálculo del Índice Local de Autocorrelación Espacial (LISA) . . . . .	63
3.14. Gráfico del Índice de Moran global univariado para robos y sus rezagos espaciales .	64
3.15. Moran I estandarizado con (EB rate) . . . . .	66
3.16. Gráficos de scatter plot, mapa de clústeres y mapa de significancia para Moran I con (EB rate) . . . . .	66
3.17. Gráficos scatter plot, mapa de clústeres, mapa de significancia con 9999 permuta- ciones para Moran I con (EB rate) . . . . .	67
3.18. Índice de Moran con 9999 permutaciones . . . . .	68
3.19. Gráfico de dispersión del Índice de Morán con mapas de significancia y de clústeres	69
3.20. Gráfico de dispersión del Índice de Moran Bivariado con mapas de significancia y de clústeres . . . . .	70

3.21. Índice de Moran Bivariado para hogares con personas que han migrado y robos con 9999 permutaciones . . . . .	71
3.22. Índice de Moran Bivariado entre lprob13 y lpnanlf . . . . .	72
3.23. Índice de Moran Bivariado entre lprob13 y lpnuj . . . . .	73
3.24. Índice de Moran Bivariado entre lprob13 y lpemig . . . . .	73
3.25. Índice de Moran Bivariado entre lprob13 y Inviv . . . . .	74
3.26. Zonas censales con sus centroides . . . . .	86
3.27. Polígonos de Thiessen correspondientes a Quito urbano . . . . .	87
3.28. Quadratic trend surface predicted value map contiguidad de orden 2 . . . . .	101
3.29. Mapa de Residuos, Superficie de Tendencia Cuadrática (Quadratic Trend Surface) contiguidad de orden 2 . . . . .	102
3.30. Quadratic trend surface residual /fitted value plot (contiguidad de orden 2) . . . . .	103
3.31. Moran scatter plot para Quadratic Trend Surface Residual (contiguidad de orden 2)	103

# Índice de cuadros

3.1. Variables y descripción del tratamiento dado a sus valores . . . . .	45
3.2. Resultados de la estimación por MCO . . . . .	76
3.3. Resultados de la estimación del Modelo espacial considerando siete variables ini- ciales . . . . .	79
3.4. Resultados de la estimación del Modelo considerando variables iniciales con sus rezagos . . . . .	80
3.5. Resultados de la estimación del modelo con rezagos espaciales . . . . .	82
3.6. Linear Trend Surface Primer Semestre . . . . .	89
3.7. Linear Trend Surface Segundo Semestre . . . . .	91
3.8. Linear Trend Surface Tercer Semestre . . . . .	93
3.9. Quadratic Trend Surface Primer Semestre . . . . .	95
3.10. Quadratic Trend Surface Segundo Semestre . . . . .	97
3.11. Quadratic Trend Surface Tercer Semestre . . . . .	99
3.12. Resultados del modelo de rezagos espaciales-Estimación de Máxima Verosimilitud	106

# Capítulo 1

## INTRODUCCIÓN AL PROBLEMA

En este trabajo se propone realizar un análisis geoestadístico de “*delitos*” orientado específicamente a “*los robos en todas sus modalidades*”, que actualmente afectan al Distrito Metropolitano de Quito en su parte urbana, aplicando para ello la metodología denominada “*Análisis Exploratorio de Datos Espaciales*”. Entendiéndose por *datos espaciales*, aquellos valores de la variable que tienen una referencia geográfica asociada. Las referencias geográficas corresponden a la georeferenciación de las distintas zonas censales que conforman a esta ciudad. Se trabaja con zonas censales porque corresponden a las divisiones territoriales más pequeñas para la cual fue posible obtener información y que permiten contar con un número relativamente grande de ellas (448), a las que más adelante se las denominará unidades geográficas.

Esta metodología permite describir cómo se encuentra distribuida geográficamente una variable de interés, determinar cuán intensa es su presencia en cada una de las unidades geográficas, permite además visualizar la presencia de patrones, de datos atípicos, etc., en mapas que previamente son generados para la ciudad, país o región donde se realice el estudio. Los diversos gráficos estadísticos y los valores o índices obtenidos del respectivo análisis espacial, describen las principales características de los datos y contribuyen a establecer la presencia de autocorrelaciones espaciales, la formación de patrones o agrupamientos, entre otros. Y finalmente permite, por medio de la realización de varias regresiones espaciales, obtener un modelo matemático que busca analizar si los robos en una determinada zona censal, están siendo influenciados o dependen del promedio de robos en las zonas censales vecinas y de ciertos indicadores socioeconómicos, tales como pobreza, hacinamiento, migración, falta de servicios básicos, acceso a educación, etc., debiéndose hacer notar que con ello, no se pretende determinar causalidad.

Generalmente la mayoría de modelos estadísticos simples, parten de variables aleatorias indepen-

dientes e idénticamente distribuidas (*iid*), donde el supuesto de independencia permite a la teoría estadística ser más entendible. Sin embargo al trabajar con datos de tipo espacial, los modelos que mejor se ajustan a la realidad, son aquellos que pueden involucrar dependencia (DE)<sup>1</sup> o autocorrelación espacial (AE),<sup>2</sup> entre sus variables. Los datos que se emplearán, son precisamente de este tipo, en los que se estima que la dependencia espacial, entre valores de la variable, va en las todas direcciones y en general se vuelve más intensa en las localizaciones que se hallen más cercanas (vecinas entre sí). Tobler, 1970.

## 1.1. Antecedentes

Como una de las políticas del actual gobierno se tiene, previa la firma de los respectivos compromisos de confidencialidad, la posibilidad de acceder a bases de datos de todo el país, que están disponibles principalmente, en el Ministerio del Interior, en el ex ex CEASI<sup>3</sup> (Datos y Estadísticas relacionadas con problemáticas sociales), en el INEC<sup>4</sup> (Censos de Población y Vivienda del país), Instituto de la Ciudad (Complejidad de la ciudad), Observatorio Metropolitano de Seguridad Ciudadana (OMSC) y en la Policía Zona 9 del Distrito Metropolitano de Quito (Institución encargada de brindar seguridad pública donde se reporta y registra los delitos). A estas Instituciones del Estado, se acudió para conseguir información relacionada a los robos en todas sus modalidades, cometidos en esta ciudad y a varios indicadores socioeconómicos, pertenecientes a las diferentes zonas censales del Distrito Metropolitano de Quito, variables necesarias para la consecución de este estudio.

Se empleará métodos de Análisis Exploratorio de Datos Espaciales, siglas en inglés (ESDA), Sistemas de Información Geográfica (SIG) y los programas de distribución libre, “GeoDa” y “R”, los cuales para poder hacer uso de sus potencialidades y que proporcionen los resultados buscados, tanto visuales como numéricos, requieren de información previamente extraída, procesada y compatibilizada<sup>5</sup>. Por lo que como punto de partida, se extraen de las extensas bases de datos disponibles, únicamente, la información pertinente a la parte Urbana del Distrito

<sup>1</sup>DE: dependencia espacial se produce cuando el valor de una variable en una unidad geográfica es parcialmente función del valor de la misma variable en unidades vecinas (Flint, Harrower & Edsall, 2000, p. 4)

<sup>2</sup>AE: autocorrelación espacial refleja el grado en que objetos o actividades en una unidad geográfica (zona censal) son similares a otros objetos o actividades en unidades geográficas próximas (Goodchild, 1987).

La diferencia entre AE y DE está fundamentalmente en el uso de las palabras, AE hace simultáneamente referencia a un fenómeno y técnica estadística y la segunda a una explicación teórica (Villalta, 2005)

<sup>3</sup>ex CEASI: Centro Ecuatoriano de Análisis de Seguridad Integral

<sup>4</sup>INEC: Instituto Ecuatoriano de Estadísticas y Censos

<sup>5</sup>que corresponda a una misma unidad geográfica “zona censal”

Metropolitano de Quito y que como referencia para la realización de este estudio, se considera a los “Delitos cometidos durante el año 2013” y el primer semestre del año 2014, específicamente, los diferentes tipos de “Robos”, información obtenida en el ex CEASI, y a los que posteriormente los relacionaremos con ciertos indicadores socioeconómicos, extraídos a su vez de la base de datos del Censo de Población y Vivienda correspondiente al año 2010 (último disponible).

La presencia de autocorrelación espacial y de heterogeneidad espacial (HE)<sup>6</sup>, hace necesaria la aplicación de diversas técnicas de suavizado, tendientes a reducir los efectos de la heteroscedasticidad en el modelo de regresión, la misma que puede deberse, a la variación espacial de la variable en cuestión y/o a la varianza de las perturbaciones. GeoDa, implementa este procedimiento de ajuste, basándose en el principio de suavizado Empírico de Bayes.

Posteriormente se presentará a través del empleo del programa estadístico R los resultados que se pueden obtener si a los datos disponibles se los trata como datos dentro de la metodología de “*Datos de Panel Espaciales*” cuyo principal propósito será realizar un estudio espacio-temporal de la variable “*robos*” pero esta vez analizados en tres períodos diferentes en el tiempo (primer semestre del año 2013, segundo semestre 2013 y primer semestre del año 2014).

## 1.2. El Problema

En la actualidad, sobre todo en las grandes ciudades o lugares con asentamientos humanos considerables, es importante poder realizar análisis de datos espaciales, que permitan explicar de alguna manera los diversos fenómenos sociales que ocurren en ellos, así como sus relaciones.

En el Distrito Metropolitano de Quito se han realizado algunos estudios asociados a este tipo de problemáticas, sin embargo, el uso de modelos matemáticos y de herramientas provenientes de software específico para el análisis espacial de datos, aún no ha logrado desarrollarse debidamente, de tal manera que su implementación acorde a las exigencias tecnológicas actuales se vuelve imprescindible.

Este trabajo pretende fomentar su implementación, mostrar su versatilidad y hasta cierto punto la confiabilidad de lograr, diversas formas de explorar y analizar datos espaciales, tanto de manera descriptiva, matemática como visual. Pudiéndose apreciar a través de mapas choropleths<sup>7</sup>, el

<sup>6</sup>HE: Variación de las relaciones entre las variables en el espacio (LeSage, 1999.)

<sup>7</sup>Mapa temático, con áreas sombreadas en proporción a la medida de la variable estadística analizada.

comportamiento de las variables de interés, y su incidencia en lugares específicos de la ciudad. Los mapas se convierten entonces en herramientas muy útiles, que facilitan el análisis visual, de lo que está presente o lo que está ocurriendo en las diferentes zonas de la ciudad, más aún si recordamos que, “una imagen vale más que mil palabras”. Y los valores numéricos obtenidos permiten corroborar lo que se puede percibir visualmente en los mapas.

La falta de uniformidad en el manejo de tamaños y formas de las unidades geográficas<sup>8</sup> para los datos (en bruto o tasas netas), entre las diferentes instituciones gubernamentales, fue uno de los principales problemas que se debió solventar, pues los mapas que generalmente se disponen para el Distrito Metropolitano de Quito (DMQ) están divididos a nivel de parroquias como unidad geográfica más pequeña; sin embargo los datos correspondientes al menor tamaño y forma, disponibles en el INEC, están a nivel de zonas censales, sectores censales y hasta manzanas; en cambio en el ex CEASI, la información que se disponía para la unidad geográfica más pequeña, está a nivel de parroquias, la información proveniente de la Policía Zona 9 del DMQ, se encuentra en cambio a nivel de circuitos y subcircuitos dentro de Distritos (denominación que emplea la Policía para unidades geográficas correspondiente a un grupo de barrios que en ocasiones pertenecen a diferentes parroquias). Por lo que el estudio a nivel de parroquias se volvía impracticable; para compatibilizar toda la información, se considera a las zonas censales, como las unidades geográficas más pequeñas, para las que se puede desagregar la información disponible además presentan la ventaja de contar con un mayor número de estas unidades.

Como resultado adicional de este trabajo, se dispondrá, de una nueva y amplia base de datos espaciales, generadas al unir y relacionar los datos existentes en las Instituciones del Estado anteriormente mencionadas, y que consta de “*tablas*”, formadas por filas generalmente denominadas “*registros*” y columnas denominadas “*campos o atributos*”. Además se contará con las representaciones espaciales (mapas geográficos) que se almacenan como entidades vectoriales<sup>9</sup> y/o ráster,<sup>10</sup> compatibilizados y actualizados, correspondientes a 448 zonas censales del Distrito Metropolitano de Quito en su parte urbana, que se almacenan y administran en columnas de atributos (variables de estudio), junto con sus registros (nombre de la parroquia a la que pertenece cada zona censal). Las tablas así obtenidas, constan de tantas filas como zonas censales se tienen y a cada fila le corresponde el mismo número de columnas. Figura (1.1)

<sup>8</sup>Valor del polígono, para especificar que dicho campo contiene las coordenadas y la geometría que lo define.

<sup>9</sup>Entidades vectoriales, son puntos, líneas o polígonos que representan objetos físicos

<sup>10</sup>Ráster, representan entidades geográficas dividiendo el espacio en celdas discretas cuadradas o rectangulares dispuestas en una cuadrícula, cada celda tiene un valor que se utiliza para representar alguna característica de dicha ubicación.

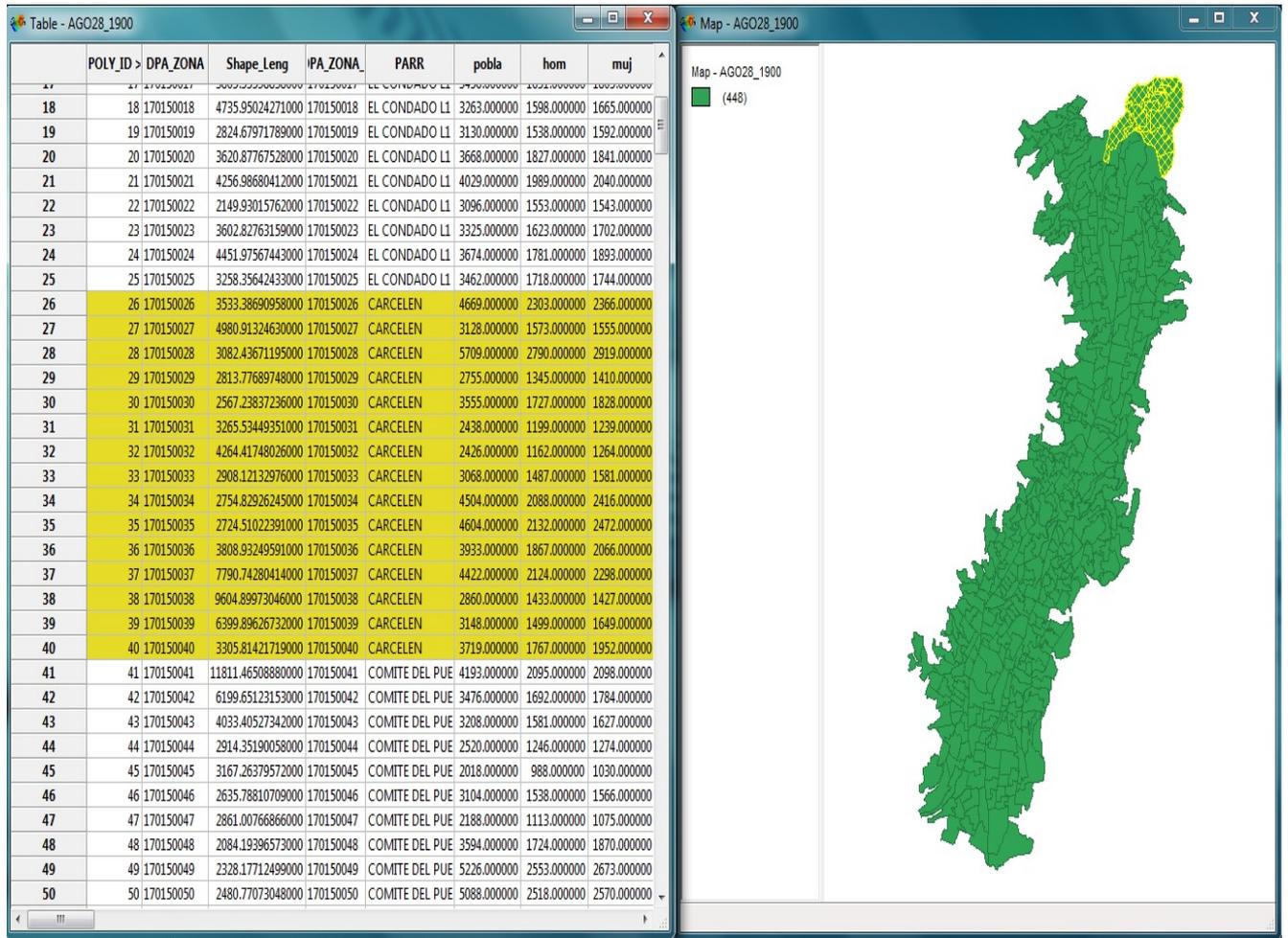


Figura 1.1: Zonas censales urbanas del D.M.Q. (448)

### **1.3. Hipótesis**

Como hipótesis para este análisis se plantea: “los robos en las diferentes zonas censales del Distrito Metropolitano de Quito, no están influenciados por la cantidad de robos en las zonas censales vecinas y por tanto la distribución de los datos obedece a una distribución espacialmente aleatoria”.

Para verificar la validez de la afirmación anterior se planea realizar el análisis geo-estadístico de estos datos espaciales a través del empleo de herramientas y aplicaciones presentes en los software de distribución gratuita “GeoDa” y “R” y de la implementación de modelos matemáticos que pretendan explicar, la dependencia espacial de la tasa de robos (si ésta existe) entre zonas censales, su dependencia de ciertos indicadores socioeconómicos, considerados como relevantes para este estudio, así como tratar de visualizar hacia que sectores de la ciudad se encuentran tendencias de crecimiento o disminución de esta problemática social.

### **1.4. Contexto y marco teórico**

La delincuencia que afecta al Distrito Metropolitano de Quito, parece obedecer a ciertos factores principalmente de índole social, psicológico y económico, este hecho no afecta, exclusivamente a la ciudad de Quito, sino a la mayoría de pueblos y ciudades sobre todo en Latinoamérica. Sin embargo esta problemática puede ser analizada desde varios enfoques el: psicológico, social, económico, político, administrativo, entre otros. Para este trabajo el enfoque que se plantea, supone adoptar una visión territorial (ubicación geográfica) de los delitos además del análisis geoestadístico y matemático de este problema.

#### **1.4.1. El propósito del estudio**

Este análisis pretende describir la distribución geográfica de los delitos (robos) en el Distrito Metropolitano de Quito en su parte urbana. Se trata de encontrar o poner de manifiesto la presencia de patrones, verificar la presencia de dependencia espacial, de autocorrelaciones espaciales entre variables analizadas, la formación de agrupamientos y la presencia de valores atípicos, pues se cree que los robos son influenciados por sí mismos y por ciertas variables exógenas que pueden crear patrones identificables visual y estadísticamente.

### **1.4.2. El significado del estudio**

Este estudio es de gran importancia pues permitirá describir el comportamiento de la variable, robos, en las diferentes zonas censales y verificar si esta variable depende de otras de tipo socio-económico. Así mismo servirá para que las instituciones encargadas de la seguridad pública, en vista de los resultados obtenidos, tomen acciones tendientes a controlar o disminuir la tasa de robos en las diferentes zonas de esta ciudad.

### **1.5. Presunciones de los autores del estudio**

Se presume que los datos espaciales obtenidos con respecto al número de robos perpetrados en una determinada zona censal dentro del sector urbano del Distrito Metropolitano de Quito, son influenciados por los robos presentes en las zonas censales vecinas. Además se cree que la tasa de robos a su vez están siendo influenciada en mayor o menor medida por ciertos indicadores socioeconómicos presentes en cada una de estas zonas censales.

### **1.6. Supuestos del estudio**

Se pretende visualizar en qué sectores de Quito se han concentrado la mayoría de los casos de robos. Se supone que la mayoría de ellos pueden estarse ejecutando en los sectores con menores seguridades, en donde se encuentre menos personal encargado de la seguridad pública, en los sectores menos atendidos por los gobiernos locales o en los sectores donde se concentra la actividad económica y social (diversiones) de la ciudad.

A continuación se presenta la revisión de la literatura que sustenta este estudio. Seguidamente la explicación de la metodología de investigación aplicada, el análisis de datos obtenidos y finalmente las principales conclusiones obtenidas.

# Capítulo 2

## ANÁLISIS ESPACIAL

Las técnicas incluidas en el set de herramientas de GeoDa, están dirigidas entre otras a la exploración de valores atípicos o a la formación de agrupamientos en mapas que representan rangos o proporciones, tales como por ejemplo rangos de homicidios, rangos de incidencia de cáncer, rangos de mortalidad, etc.

Tres amplias clases de procedimientos son considerados:

- Mapas de outliers,
- Procedimientos de suavizado, y
- Análisis de Autocorrelación.

En el primero se implementan mapas de la variable expresada en percentiles, mapas de caja de exceso de rango.

Un segundo conjunto de métodos engloba procedimientos de suavizado, con el fin de obtener estimaciones más precisas por ejemplo del riesgo subyacente que se produce por mapas que utilizan la tasa cruda correspondiente a cada localización.

Por último, el análisis de autocorrelación, se refiere al análisis de cierta variable correspondiente a una determinada zona geográfica con relación a la influencia que puede ejercer sobre esta misma variable pero de sus vecindades y/o la influencia que pueden ejercer sus vecinos sobre la variable antes mencionada.

## 2.1. Dependencia Espacial

Se comienza con algunas definiciones que son importantes, para entender el término de dependencia espacial, se tienen muestras de datos espaciales, que están compuestas por observaciones de alguna variable de interés que está asociada con puntos, regiones o zonas geográficas de un determinado país, provincia, cantón, parroquia, zona censal, etc.

Si se considera una variable representada por un vector transversal<sup>1</sup> en el tiempo, que representa observaciones obtenidas que incluyen una referencia geográfica asociada (puntos o regiones en el espacio). Las variables pueden ser por ejemplo, número de robos en cada unidad geográfica escogida, o pueden ser número de casos de personas contagiadas con determinado virus, que de pronto se ha convertido en endémico en cierta región y requerimos conocer cómo se está expandiendo en determinada dirección, esto es, hacia qué regiones se está expandiendo dicho virus, otro ejemplo puede ser los precios de venta de casas en determinadas unidades geográficas.

Georreferenciar información geográfica, significa proporcionar un código, que lo asocia con cada unidad geográfica, a esto también se lo denomina geo-codificación, esto permite posteriormente asociar cada una de las unidades geográficas con coordenadas de latitud y longitud (centroides de cada unidad geográfica). Que a su vez permite relacionar las distintas variables que se desea analizar con sus respectivas ubicaciones dentro de las unidades geográficas que previamente deben ser geo-referenciadas. Es importante señalar que en modelos de regresión espacial, cada observación corresponde a una unidad geográfica en particular.

Para explicar el procedimiento que se llevará adelante, la notación en forma de vectores y/o matrices se dejará de lado por el momento y en general se tratará como términos de observaciones individuales.

El proceso de generación de datos (DGP) para una determinada muestra transversal en el tiempo no considerada espacialmente, compuesta de  $n$  observaciones independientes  $y_i$ , con  $i = 1, 2, \dots, n$  que están relacionadas linealmente a una matriz  $X$  de variables explicativas, toma la forma de (2.1) donde se ha omitido término del intercepto, el cual puede ser incluido en la matriz  $X$ .

$$y_i = X_i\beta + \varepsilon_i \quad (2.1)$$

---

<sup>1</sup>valor de la variable tomada en un tiempo  $t$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1, 2, \dots, n \quad (2.2)$$

Donde:

$N(0, \sigma^2)$  denota una distribución normal univariada con media 0 y varianza  $\sigma^2$ .

$X_i$  vector de  $1 \times k$  de variables explicativas.

$\beta$  vector de  $k \times 1$  de parámetros asociados

En este modelo, cada observación tiene media de  $X_i\beta$  y una componente aleatoria  $\varepsilon_i$ . Una implicación de este modelo es que para situaciones donde las observaciones  $i$  representan unidades geográficas, o puntos en el espacio es que los valores observados en cada unidad geográfica (o puntos en el espacio), son *independientes* de observaciones hechas en otras localizaciones o regiones geográficas, este hecho simplifica el modelo. Esto último implica que  $E(\varepsilon_i\varepsilon_j) = E(\varepsilon_i)E(\varepsilon_j) = 0$ , en el contexto espacial esta simplificación parece forzada, pues en sentido espacial, existen muchos factores que deberían ser agregados en el modelo,(2.1) en contraste, el término *dependencia espacial*, refleja la situación donde los valores observados en una unidad geográfica (la observación  $i$ ), *depende* de los valores de las observaciones de sus *vecinos* (localizaciones geográficas vecinas).

Supongamos que son vecinos las observaciones  $i = 1$  y  $j = 2$  por ejemplo con fronteras comunes entre una y otra región, entonces, el proceso de generación de datos tomará la forma mostrada en (2.3).

$$y_i = \alpha_i y_j + X_i \beta + \varepsilon_i \quad (2.3)$$

$$y_j = \alpha_j y_i + X_j \beta + \varepsilon_j$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1$$

$$\varepsilon_j \sim N(0, \sigma^2) \quad j = 2$$

La fórmula (2.3) indica un proceso simultáneo de generación de datos, donde el valor tomado por  $y_i$  depende del valor de  $y_j$  y viceversa.

La fórmula (2.3) es importante ya que refleja dependencia, pues para obtener el valor de la variable dependiente  $y_i$  se ve que ésta depende del valor de la variable  $y_j$  (en este caso como variable explicativa) de la región vecina  $j$ . En forma similar, se puede decir que para obtener el valor de la variable dependiente  $y_j$  se ve que ésta depende del valor de la variable  $y_i$  (en este caso como variable explicativa) de la región vecina  $i$ .

A veces la dependencia espacial puede deberse a variables omitidas en el modelo. También puede verse dependencia espacial reflejada en los residuos (en el modelo de regresión no espacial) y se ha trabajado mucho en este sentido tratando de justificar lo que llamaremos *una variable dependiente rezagada* (López-Bazo et al., 2004; Ertur and Koch, 2007).

Así, *una variable dependiente rezagada* es un vector, incluido como una variable explicativa construida usando un promedio de valores de la variable dependiente en las regiones vecinas. Estos modelos tratan de exteriorizar mediante un modelo, la interdependencia que existe entre regiones, lo que se reduce a incluir en el modelo de regresión un rezago espacial de la variable dependiente, entonces, una vez añadida en el modelo esta variable dependiente espacialmente rezagada, se la puede utilizar para realizar estimaciones e inferencias, pero se debe analizar la *significancia* del parámetro asociado con la variable dependiente rezagada.

## 2.2. El Proceso Espacial Autorregresivo

En (2.3), se pueden generar más observaciones como se muestra en (2.4):

$$y_i = \alpha_{i,j}y_j + \alpha_{i,k}y_k + X_i\beta + \varepsilon_i \quad (2.4)$$

$$y_j = \alpha_{j,i}y_i + \alpha_{j,k}y_k + X_j\beta + \varepsilon_j$$

$$y_k = \alpha_{k,i}y_i + \alpha_{k,j}y_j + X_k\beta + \varepsilon_k$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1$$

$$\varepsilon_j \sim N(0, \sigma^2) \quad j = 2$$

$$\varepsilon_k \sim N(0, \sigma^2) \quad k = 3$$

Pero por este camino se tendría muchos más parámetros que observaciones, existiendo por tanto un problema de sobre-parametrización, ante esto LeSage, propone imponer una estructura determinada sobre las relaciones de dependencia espacial. Ord, (1975) propone una forma de parametrización para las relaciones de dependencia, esta estructura desemboca en un proceso de generación de datos conocido como *Proceso Espacial Autoregresivo* aplicado a las relaciones de dependencia entre las observaciones sobre la variable  $y$ , que puede expresarse como (2.5):

$$y_i = \rho \sum_{j=1}^n W_{ij}y_j + \varepsilon_i \quad (2.5)$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1, \dots, n$$

Donde no consta el intercepto, asumiendo que el vector de observaciones de la variable  $y$  está en la forma de desviaciones desde la media. El término  $\sum_{j=1}^n W_{ij}y_j$  es llamado *un rezago espacial* ya que representa una combinación lineal de los valores de la variable  $Y$ , construída desde las observaciones/unidad geográfica vecinas a la observación  $i$ . Esto es llevado a efecto gracias a los elementos  $W_{ij}$  pertenecientes a la *Matriz de pesos espaciales*  $W$ , así  $\sum_{j=1}^n W_{ij}y_j$  resulta en un escalar que representa una combinación lineal de los valores tomados por observaciones vecinas.

Se define *vecino de primer orden de la región  $i$*  como aquellos vecinos directos cuyas fronteras las comparten con la región  $i$ . Se dice *vecinos de segundo orden  $k$*  si la región  $k$  es vecina de la vecina directa de la región  $i$  (es decir, vecinos de los vecinos de primer orden); y así sucesivamente.

Es importante anotar que la región  $i$  es vecina de segundo orden para sí misma, esto de acuerdo a la definición anteriormente dada de vecino de segundo orden, pues es vecina de su vecino. Y si las relaciones de vecindad son simétricas, cada región siempre será un vecino de segundo orden de sí misma. Por naturaleza, las relaciones de contiguidad son simétricas, pero como indica LeSage, existen otras definiciones de relaciones de vecindad que pueden resultar no ser simétricas.

Se puede escribir una versión en forma matricial de este proceso autorregresivo como (2.6):

$$y = \rho W y + \varepsilon_i \quad (2.6)$$

$$\varepsilon_i \sim N(0, \sigma^2 I_n)$$

Donde:

$W$  es la matriz  $n \times n$  de pesos espaciales usando relaciones de contiguidad de primer orden que resulta de un proceso a describirse a continuación, partiendo de una matriz inicial  $C$ , en ella se asocian las filas de la matriz con sus vecinos: si es vecino del tipo de primer orden se pone 1 y si no lo es, se pone 0, también se pone 0 si es vecino de sí mismo, es por esto que la diagonal de esta matriz está compuesta de ceros.

Con el propósito de formar un rezago espacial o combinaciones lineales de valores de las observaciones de la variable estudiada en las vecindades, se procede a normalizar la matriz  $C$  de tal manera que la suma de los elementos de las filas resulta la unidad. Esta es la matriz  $W$  nombrada *row - stochastic* dicho término se refiere a una matriz no negativa normalizada utilizando la suma de los elementos de cada fila de la matriz  $C$ , por ejemplo si en la fila  $i$  hay tres

unos, su suma es 3 y dividiremos a cada *uno* en dicha fila para tres, dando de esta manera la suma de cada fila de esta matriz normalizada *la unidad*.

La matriz  $W$  puede ser multiplicada por un vector  $y$  de dimensión  $n \times 1$  de valores tomados de cada región para producir un *vector espacial rezagado* de la variable dependiente, este vector mencionado toma la forma:  $Wy$ . Este producto matricial  $Wy$  cuyo resultado es un vector  $n \times 1$  que representa el vector de rezagos espaciales para cada observación  $i$ ,  $i = 1, \dots, n$ . Esta matriz  $W$ , en la cual se trata a los pesos de cada vecindad en forma igualitaria, da como resultado un vector de rezagos con la característica que viene a ser un simple promedio de los valores de las observaciones de las vecindades que cumplan con el criterio de vecindad de primer orden para cada región.

El parámetro  $\rho$  en (2.6) que es un escalar, describe la intensidad de la dependencia espacial en la muestra de observaciones. El uso de un único parámetro para reflejar un nivel promedio de dependencia sobre todas las relaciones de dependencia que surgen de las observaciones  $i = 1, \dots, n$ , parecería que no es suficiente para reflejar la real estructura autorregresiva.

Como se ha mencionado, existen otras alternativas para definir y ponderar las observaciones de las vecindades.

Resulta interesante examinar gráficamente mediante el llamado *Moran scatter plot* la relación entre las observaciones en el vector  $y$  (en la forma de desviaciones de la media) versus los valores promedio de las observaciones de las vecindades. En virtud de esta transformación a desviaciones de las medias, se tienen en dicho gráfico cuatro cuadrantes:

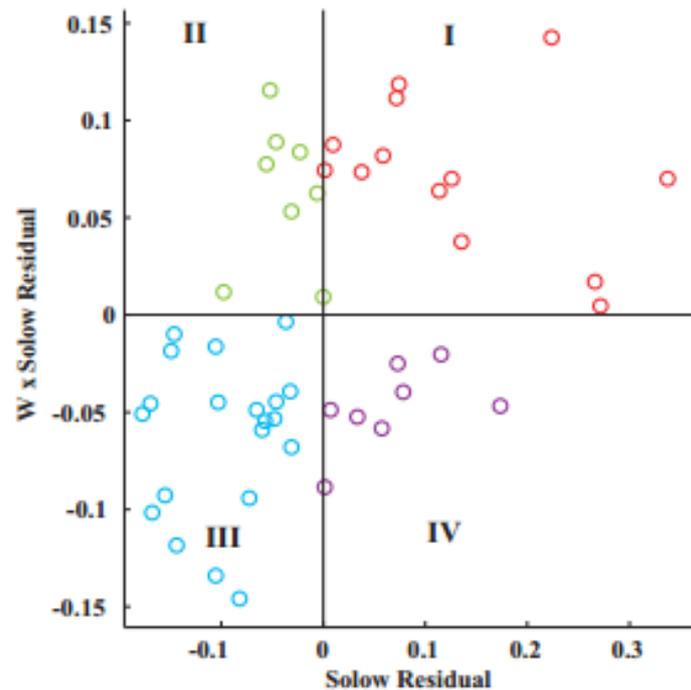


Figura 2.1: **Moran Scatter plot relación entre observaciones del vector  $y$  vs valores promedio de observaciones vecinas** (Fuente: Introduction to Spatial Econometric LeSage, J)

Cuadrante I: (puntos en color rojo) en nuestro caso serían las zonas censales que tienen un valor de la variable analizada (residuos) está por encima de la media, e igualmente, donde el promedio de la variable analizada de las vecindades  $Wy$  es también mayor que la media.

Cuadrante II: (puntos en color verde) serían las zonas censales que exhiben un valor de la variable que está bajo la media, pero que el promedio de dicha variable de sus vecindades  $Wy$  está sobre la media.

Cuadrante III: (puntos en color azul) corresponderían a las zonas censales que presentan un valor que está por debajo de la media y también el promedio de la variable en sus vecindades  $Wy$  está por debajo de la media.

Cuadrante IV: (puntos en color violeta) zonas censales que presentan un valor de la variable sobre la media, y el promedio del valor de la variable en sus vecindades  $Wy$  está por debajo de su media.

Del análisis del (Moran Scatter plot), si se observara una asociación positiva entre la variable dependiente  $y$  ubicada en el eje horizontal y las observaciones espacialmente rezagadas  $Wy$  en

el eje vertical, esto sugeriría que el parámetro espacial  $\rho$  es mayor que cero.

La presencia de poquísimos puntos de colores verde y violeta en el Moran scatter plot es una indicación de una *fuerte asociación positiva*. Por otra parte, la presencia de gran cantidad de puntos en los cuadrantes II y IV con pocos puntos en los cuadrantes I y III puede sugerir una dependencia espacial negativa y por ello el parámetro  $\rho$  para este caso es menor que cero ( $\rho < 0$ ).

Los puntos en el scatter plot pueden ser visualizados en el mapa usando el mismo código de colores, en nuestro caso en el mapa de Quito urbano, dividido en unidades territoriales denominadas zonas censales. En este mapa podemos si existe, visualizar los llamados clustering (agrupamientos) de zonas censales, por ejemplo zonas censales con observaciones de color rojo rodeadas de zonas censales de color rojo, ó zonas censales de color azul rodeadas de zonas censales de color azul lo que indicaría en ambos casos una *dependencia espacial positiva*. Esto indicará, previo un análisis mediante test y pruebas de significancia, que dicha asociación o agrupamiento no es obra de un proceso aleatorio (es decir, no es obra de la casualidad).

Mientras que zonas censales de color verde rodeadas de color violeta o viceversa, indicarían la presencia de una *dependencia espacial negativa*.

Un comentario realizado por LeSage sobre el parámetro  $\rho$ : El interpretar a este parámetro dentro de este proceso espacial autorregresivo como el conocido coeficiente de correlación entre el vector  $y$  y el vector espacial rezagado  $Wy$  no es enteramente preciso, pues entre otras cosas el rango del coeficiente de correlación está entre  $[-1, 1]$  mientras que  $\rho$  no puede ser igual a uno.

### 2.3. El Proceso Espacial Autorregresivo de Generación de Datos

El proceso espacial autorregresivo es mostrado en (2.7), y su proceso de generación de datos es mostrado en (2.8). LeSage introduce un vector constante de *unos* denominado  $\iota_n$ , y un parámetro asociado  $\alpha$  para acomodar situaciones donde el vector  $y$  no tiene una media de valor cero.

$$y = \alpha \iota_n + \rho Wy + \varepsilon \quad (2.7)$$

$$(I_n - \rho W)y = \alpha \iota_n + \varepsilon$$

$$y = (I_n - \rho W)^{-1} \iota_n \alpha + (I_n - \rho W)^{-1} \varepsilon \quad (2.8)$$

$$\varepsilon_i \sim N(0, \sigma^2 I_n)$$

Donde:

$y$  vector  $n \times 1$  contiene la variable dependiente.

$\rho$  es un parámetro escalar

$W$  matriz  $n \times n$  de pesos espaciales

Se asume que  $\varepsilon$  sigue una distribución normal multivariada, con media cero y una matriz constante de varianza-covarianza  $\sigma^2 I_n$  cuya diagonal está compuesta de valores escalares.

El modelo representado por (2.7) como un indicativo de que el valor esperado de cada observación  $y_i$  dependerá del valor medio  $\alpha$  más una combinación lineal de los valores tomados de las observaciones de las vecindades escalado por el parámetro  $\rho$ . El proceso de generación de datos en (2.8) expresa la naturaleza simultánea de un proceso espacial autorregresivo.

Se puede representar  $(I_n - \rho W)^{-1}$  usando la serie infinita convergente (si por el momento asumimos que  $|\rho| < 1$ ):

$$(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots \quad (2.9)$$

Sustituyendo esto en (2.8), obtenemos para el proceso autorregresivo de generación de datos para el vector  $y$ :

$$y = (I_n - \rho W)^{-1} \iota_n \alpha + (I_n - \rho W)^{-1} \varepsilon$$

$$y = \alpha \iota_n + \rho W \iota_n \alpha + \rho^2 W^2 \iota_n \alpha + \dots + \varepsilon + \rho W \varepsilon + \rho^2 W^2 \varepsilon + \dots \quad (2.10)$$

Además, la expresión (2.9) se simplifica si ( $|\rho| < 1$ ), entonces  $\iota_n \alpha + \rho W \iota_n \alpha + \rho^2 W^2 \iota_n \alpha + \dots$  converge a  $(1 - \rho)^{-1} \iota_n \alpha$  ya que  $\alpha$  es un escalar,  $|\rho| < 1$ , y  $W$  es estocástica por filas. Por definición  $W \iota_n = \iota_n$  y por tanto  $W(W \iota_n)$  es también igual a  $W \iota_n = \iota_n$  y  $W^q \iota_n = \iota_n$  para  $q > 0$ , teniendo  $W^0 = I_n$ , (2.9) se simplifica a:

$$y = \frac{1}{1 - \rho} \iota_n \alpha + \varepsilon + \rho W \varepsilon + \rho^2 W^2 \varepsilon + \dots \quad (2.11)$$

En la expresión (2.11) se tiene potencias de la matriz de pesos:  $W^2$ ,  $W^3$ ,  $W^4$ , ..., Se asume que las filas de la matriz de pesos  $W$  son construidas para representar vecindades de primer orden. La matriz  $W^2$  reflejará vecindades contiguas de segundo orden (Aquellos que son vecinos de los

vecinos de primer orden), y ya que el vecino del vecino a una observación  $i$  incluye la observación  $i$  en sí misma, pues cumple con la definición de vecindad de segundo orden, entonces la diagonal de  $W^2$  está compuesta por elementos positivos cuando cada observación tiene al menos un vecino. Se tiene que rezagos espaciales de órdenes mayores pueden acarrear una relación de conectividad para una observación  $i$  tal que  $W^2\varepsilon$  extraerá observaciones del vector  $\varepsilon$  que apunte de regreso a la misma observación  $i$ . Esto presenta un marcado contraste con la relación descrita en (2.1) donde las asunciones Gauss-Markov descartan dependencia de  $\varepsilon_i$  sobre otras observaciones  $j$ , asumiendo cero covarianza entre las observaciones  $i$  y  $j$  en el proceso de generación de datos.

En LeSage, se muestra en un ejemplo que dado que  $|\rho| < 1$ , el proceso de generación de datos asigna menor influencia a términos de perturbaciones asociadas a vecindades de orden superior, con un decaimiento tipo geométrico cuando el orden de vecindad crece. También señala que una fuerte dependencia espacial reflejada en altos valores de  $\rho$  asigna una mayor influencia de los vecinos de mayores órdenes.

La dependencia de cada observación  $y_i$  por perturbaciones asociadas con observaciones de las vecindades como vecindades de altos órdenes, sugiere una estructura de media y varianza-covarianza para las observaciones en el vector  $y$  que dependen en una manera complicada de otras observaciones (implica modelos más complejos).

El valor esperado de la variable  $y$  de la expresión (2.11), asumiendo que la matriz de pesos espaciales es exógena o fija en repetidos muestreos, entonces:

$$E[y] = \frac{1}{1-\rho}\alpha\iota_n + E[\varepsilon] + \rho W E[\varepsilon] + \rho^2 W^2 E[\varepsilon] + \dots \quad (2.12)$$

$$E[y] = \frac{1}{1-\rho}\alpha\iota_n$$

El objetivo de modelos espaciales dice LeSage, es analizar la variación en los flujos entre regiones que ocurren en el tiempo (tomando una foto transversal en el tiempo de flujos observados entre regiones origen y regiones destino que toman lugar sobre un período de tiempo finito, pero medido en un simple punto del tiempo).

Esta situación de dependencia simultánea no ocurre en el análisis de las Series de tiempo, y esto nos dice nuevamente que modelar un proceso espacial autorregresivo es muy distinto del proceso autorregresivo de las series de tiempo.

## 2.4. Modelos de Regresión Espacial Autorregresivos

La estructura espacial autorregresiva puede ser combinada con un modelo de regresión para producir una extensión espacial de un modelo de regresión estándar, la relación (2.13) nos muestra este modelo mixto y su correspondiente proceso de generación de datos. LeSage se refiere a este modelo como *Modelo Espacial Autorregresivo* nominado con las siglas *SAR*.

$$y = \rho W y + X\beta + \varepsilon \quad (2.13)$$

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1} \varepsilon \quad (2.14)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Los parámetros a ser estimados son  $\beta$ ,  $\sigma$ ,  $\rho$ , hay que notar que si  $\rho$  toma el valor de cero indica que no existe dependencia espacial en el vector de observaciones y tomado en un instante transversal en el tiempo, de esto se tiene que el modelo de regresión por mínimos cuadrados resulta en un caso especial del modelo *SAR*.

LeSage realiza una prueba para evaluar el impacto espacial al duplicar la densidad de población de una región específica y realiza una predicción del impacto en la variable dependiente en las regiones vecinas a la región donde se duplicó la población, para dicha predicción usa el siguiente modelo basado en la variable explicativa X:

$$\hat{y} = (I_n - \hat{\rho}W)^{-1} X\hat{\beta} \quad (2.15)$$

Donde  $\hat{\rho}$ ,  $\hat{\beta}$  son estimadores de máxima verosimilitud.

De esta experiencia se tiene que el cambio en la densidad de población ocasionó un impacto indirecto o efectos secundarios (llamado en inglés *Spillovers*) en el valor de la variable dependiente en las otras regiones y la intensidad de los efectos secundarios resultaron mayores en las regiones más próximas y se fueron atenuando en regiones más alejadas. El impacto total ocasionado en todas las regiones por el cambio realizado, es la suma de los efectos directos e indirectos experimentados por dicho cambio.

La razón de cambio de la variable dependiente con respecto a la o a las variables explicativas puede expresarse como  $\frac{\partial y_i}{\partial X_{i2}}$ , donde  $X_{i2}$  se refiere a la  $i$ -ésima observación de la segunda variable explicativa en el modelo. Así, la derivada parcial cruzada representa los

*efectos indirectos asociados con este cambio.* Para representar estos efectos, el proceso de generación de datos para el modelo SAR puede ser escrito como (2.16), donde el subíndice  $r$  denota la variable explicativa  $r$

$$y = \sum_{r=1}^k S_r(W)X_r + (I_n - \rho W)^{-1}\varepsilon \quad (2.16)$$

$$E(y) = \sum_{r=1}^k S_r(W)X_r \quad (2.17)$$

Donde  $S_r(W) = (I_n - \rho W)^{-1}\beta_r$  actúa como una matriz multiplicadora que aplica las relaciones de vecindad de órdenes superiores a  $X_r$ . Derivando parcialmente (2.17):

$$\frac{\partial E(y_i)}{\partial E(X_{jr})} = S_r(W)_{ij} \quad (2.18)$$

Donde  $i, j$  representan distintas observaciones;  $S_r(W)_{ij}$  representa el elemento  $ij$  de la matriz  $S_r(W)$ .

La expresión (2.18) indica una ecuación muy diferente a la tradicional interpretación del coeficiente de una regresión estándar:  $\hat{\beta}_r = \frac{\partial y}{\partial X_r}$ .

También puede ser de interés medir el impacto sobre  $y_i$  de los cambios en la observación  $j$  de la variable  $r$ , expresado como:  $\frac{\partial y_i}{\partial X_{jr}}$ , con  $i \neq j$ . Así mismo, debería ser claro que los impactos derivados de un cambio en las variables explicativas influenciará en mayor proporción en los vecinos de menor orden que en los vecinos de mayor orden, se esperaría que dicho impacto decline en magnitud conforme nos movemos de menor a mayor orden de vecindad.

En conclusión, debido a la transformación de  $X_r$  ejecutada por la matriz  $S_r(W)$  de dimensión  $n \times n$ , entonces cualquier cambio a una variable explicativa en una región determinada, puede afectar a la variable dependiente en todas las regiones a través de la matriz inversa. Más aún, si se hace  $\rho = 0$  en el modelo, lo que estamos haciendo es tener un modelo de regresión no espacial donde las estimaciones  $\hat{y}^{(1)} = X\hat{\beta}_0$  y  $\hat{y}^{(2)} = \tilde{X}\hat{\beta}_0$ , y la diferencia  $(\tilde{X} - X)\hat{\beta}_0$  donde los parámetros estimados  $\hat{\beta}_0$  serían aquellos de una regresión por mínimos cuadrados, y ya que este modelo SAR con  $\rho = 0$  ignora el rezago espacial inherente de la variable dependiente, las estimaciones por mínimos cuadrados serán sesgadas e inconsistentes.

## 2.5. Estimación por Máxima Verosimilitud

La mayoría de estimaciones de modelos espaciales a través de modelos por mínimos cuadrados ordinarios, como se comprobará más adelante, puede conducir a estimaciones inconsistentes de los parámetros de la regresión cuando los modelos contienen variables dependientes rezagadas, obteniéndose estimaciones inconsistentes que pueden ser tanto para los parámetros espaciales como para la estimación de los errores estándar. Por esta razón el método de máxima verosimilitud es un método *consistente* para modelos espaciales (Lee 2004), por lo que este estudio se enfoca también en la teoría de estimación por máxima verosimilitud del modelo de regresión espacial.

### 2.5.1. Estimación de modelos SAR y SDM

El modelo espacial de Durbin (SDM) es mostrado en (2.19), junto con su proceso de generación de datos asociado (2.20),

$$y = \rho W y + \alpha \iota_n + X\beta + WX\theta + \varepsilon \quad (2.19)$$

$$y = (I_n - \rho W)^{-1}(\alpha \iota_n + X\beta + WX\theta + \varepsilon) \quad (2.20)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Donde

$0$  vector de ceros de  $n \times 1$

$\iota_n$  vector de unos de  $n \times 1$  asociado con el término constante del parámetro  $\alpha$ .

Este modelo puede ser escrito como un modelo espacial autorregresivo (SAR) definido por:  $Z = [\iota_n \ X \ WX]$  y  $\delta = [\alpha \ \beta \ \theta]^T$ , que llevan a (2.21). Esto significa que la función de verosimilitud para los modelos SAR y SDM pueden ser escritos en la misma forma donde:  $Z = [\iota_n \ X]$  para el modelo SAR y  $Z = [\iota_n \ X \ WX]$  para el modelo SDM.

$$y = \rho W y + Z\delta + \varepsilon \quad (2.21)$$

$$y = (I_n - \rho W)^{-1}Z\delta + (I_n - \rho W)^{-1}\varepsilon \quad (2.22)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Del modelo de la ecuación (2.21), si el verdadero valor del parámetro  $\rho$  fuese conocido digamos  $\rho^*$ , se puede ordenar la ecuación (2.21) como se muestra en la ecuación (2.23)

$$y - \rho^* W y = Z \delta + \varepsilon \quad (2.23)$$

Esto sugiere un estimado para  $\delta$  de:  $\hat{\delta} = (Z^T Z)^{-1} Z^T (I_n - \rho^* W) y$ . En este caso también puede encontrarse una estimación para el parámetro del ruido de la varianza  $\hat{\sigma}^2 = n^{-1} e(\rho^*)^T e(\rho^*)$ , donde  $e(\rho^*) = y - \rho^* W y - Z \hat{\delta}$ .

Esto motiva la total concentración en el (log) de verosimilitud con respecto a los parámetros  $\beta$ ,  $\sigma^2$  y reducir la máxima verosimilitud a un problema de optimización univariado en el parámetro  $\rho$ .

Maximizar la función log-verosimilitud completa para el caso de un modelo SAR implica obtener las primeras derivadas con respecto a los parámetros  $\beta$ , y  $\sigma^2$  con  $\rho$  igual a cero, y resolviendo simultáneamente estas condiciones de primer orden para todos los parámetros.

Por otro lado las estimaciones de máxima verosimilitud equivalentes pueden ser encontradas utilizando la función log-verosimilitud concentrada con respecto a los parámetros  $\beta$  y  $\sigma^2$ . Esto implica la sustitución de soluciones de forma cerrada de las condiciones de primer orden para los parámetros  $\beta$  y  $\sigma^2$  para producir una log-verosimilitud que se dice está concentrada con respecto a estos parámetros. Estas expresiones se han etiquetado como  $\hat{\beta}(\rho)$ ,  $\hat{\sigma}^2(\rho)$ , haciéndose notar que ellas dependen de los datos de la muestra más el parámetro desconocido  $\rho$ . En el caso de modelos SAR, esto produce una log-verosimilitud concentrada que depende sólo del único simple parámetro escalar  $\rho$ . Optimizando la función log-verosimilitud concentrada con respecto a  $\rho$ , el encontrar la máxima verosimilitud de  $\hat{\rho}$  estimada permite utilizar esta estimación en las expresiones cerradas para  $\hat{\beta}(\hat{\rho})$  y  $\hat{\sigma}^2(\hat{\rho})$  para producir estimaciones de máxima verosimilitud para estos parámetros.

Trabajar con log-verosimilitud concentrada produce exáctamente la misma estimación de máxima verosimilitud de  $\hat{\beta}$ ,  $\hat{\sigma}$  y  $\hat{\rho}$  como surgiría de la maximización de log-verosimilitud completa (Davidson and Mackinnon, 1993, p. 267-269). La motivación para optimizar la log-verosimilitud concentrada es que esto simplifica problemas de optimización multivariante reduciéndolos a problemas de optimización univariante. Otra ventaja de utilizar log-verosimilitud concentrada es que ajustes simples en las salidas de problemas de optimización, pueden ser usados para producir cálculos eficientes de la matriz de varianza-covarianza que se utiliza para la inferencia

con respecto a los parámetros. Estas inferencias son idénticas a las que se obtendrían resolviendo problemas engorrosos de optimización que impliquen log-verosimilitud completa.

La función log-verosimilitud para modelos SDM y SAR toman la forma de (2.24) (Anselin, 1988, p. 63), donde  $\omega$  es el vector  $n \times 1$  de valores propios de la matriz  $W$ .

$$\ln L = -(n/2)\ln(\pi\sigma^2) + \ln|I_n - \rho W| - \frac{e^T e}{2\sigma^2} \quad (2.24)$$

$$e = y - \rho W y - Z\delta$$

$$\rho \in (\min(\omega)^{-1}, \max(\omega)^{-1})$$

Si  $\omega$  contiene sólo valores propios reales, se garantiza una matriz de varianza-covarianza definida positiva por la condición:  $\rho \in (\min(\omega)^{-1}, \max(\omega)^{-1})$  esto se muestra en Ord (1975). La matriz  $W$  siempre puede ser construida para obtener un valor propio máximo de 1. Por ejemplo, la ampliación de la matriz de pesos por su valor propio máximo según lo observado por Barry y Pace (1999); Kelijian y Prucha (2007). En este caso el intervalo de  $\rho$  llega a ser  $(\min(\omega)^{-1}, 1)$  y un subconjunto de esto empleado ampliamente en la práctica es  $\rho \in [0, 1)$ . Los valores admisibles pueden llegar a ser más complicados para matrices de pesos  $W$  no simétricas ya que estas pueden tener valores propios complejos.

Como se ha señalado la log-verosimilitud puede ser concentrada con respecto al coeficiente  $\delta$  del vector y el parámetro  $\sigma^2$  ruido de la varianza. Pace y Barry (1997) sugieren un enfoque conveniente para la concentración de los parámetros  $\delta$  y  $\sigma^2$ , como se muestra en (2.25). el término  $\kappa$  es una constante que no depende del parámetro  $\rho$ , y  $|I_n - \rho W|$  es el determinante de esta matriz  $n \times n$ . Se usa la notación  $e(\rho)$  para indicar que este vector depende de los valores tomados por el parámetro  $\rho$ , al igual que el valor escalar de la función log-verosimilitud concentrada  $\ln L(\rho)$

$$\ln L(\rho) = \kappa + \ln|I_n - \rho W| - (n/2)\ln(S(\rho)) \quad (2.25)$$

$$S(\rho) = e(\rho)^T e(\rho) = e_0^T e_0 - 2\rho e_0^T e_d + \rho^2 e_d^T e_d$$

$$e(\rho) = e_0 - \rho e_d$$

$$e_0 = y - Z\delta_0$$

$$e_d = Wy - Z\delta_d$$

$$\delta_0 = (Z^T Z)^{-1} Z^T y$$

$$\delta_d = (Z^T Z)^{-1} Z^T W y$$

Para simplificar la optimización del logaritmo de verosimilitud con respecto al parámetro escalar  $\rho$  Pace y Barry proponen evaluar el logaritmo de verosimilitud usando un vector  $q \times 1$  de valores de  $\rho$  en el intervalo  $[\rho_{min}, \rho_{max}]$ , etiquetados como  $\rho_1, \rho_2, \dots, \rho_q$  en (2.26).

$$\begin{pmatrix} \ln L(\rho_1) \\ \ln L(\rho_2) \\ \vdots \\ \ln L(\rho_q) \end{pmatrix} = \kappa + \begin{pmatrix} \ln(|I_n - \rho_1 W|) \\ \ln(|I_n - \rho_2 W|) \\ \vdots \\ \ln(|I_n - \rho_q W|) \end{pmatrix} - (n/2) \begin{pmatrix} \ln(S(\rho_1)) \\ \ln(S(\rho_2)) \\ \vdots \\ \ln(S(\rho_q)) \end{pmatrix} \quad (2.26)$$

Para este análisis se asume que estos valores están disponibles durante la optimización de log-verosimilitud. Dada una rejilla suficientemente fina de valores de  $q$  para log-verosimilitud, la interpolación puede suministrar puntos intermedios con la precisión deseada (que se desprende del suavizado de la función de log-verosimilitud). Notar que los momentos escalares  $e_0^T e_0$ ,  $e_d^T e_0$ ,  $e_d^T e_d$  y los vectores  $\delta_0$ ,  $\delta_d$  de  $(k \times 1)$  se deben calcular antes de la optimización y así obtener un valor para  $\rho$ , por lo que el cálculo de  $S(\rho)$  simplemente requiere tres números de ponderación. Dado el valor óptimo de  $\rho$ , este llega a ser la estimación de máxima verosimilitud de  $\rho$  denotado como  $\hat{\rho}$ . Por tanto se requieren pocos cálculos para llegar al vector de valores concentrados de log-verosimilitud.

Dado  $\hat{\rho}$  estimado por máxima verosimilitud, (2.27), (2.28), y (2.29) muestran los estimados por máxima verosimilitud del coeficiente  $\hat{\delta}$ , del parámetro del ruido de la varianza  $\hat{\sigma}^2$ , y la matriz asociada de varianza-covarianza para las perturbaciones.

$$\hat{\delta} = \delta_0 - \hat{\rho} \delta_d \quad (2.27)$$

$$\hat{\sigma}^2 = n^{-1} S(\hat{\rho}) \quad (2.28)$$

$$\hat{\Omega} = \hat{\sigma}^2 [(I_n - \hat{\rho} W)^T (I_n - \hat{\rho} W)]^{-1} \quad (2.29)$$

La función verosimilitud combina un término de suma de los errores cuadrados transformado con

el logaritmo del determinante término que actúa como una función de penalización que impide que la estimación por máxima verosimilitud de  $\rho$  sea igual a una estimación basada solamente en la minimización (transformada) de la suma de los errores cuadrados  $S(\rho)$ . El enfoque de vectorización proporciona la ventaja adicional de asegurar un óptimo global en lugar de un óptimo local.

La estimación de máxima verosimilitud podría proceder usando variedad de técnicas de optimización univariadas. Estas podrían incluir el enfoque vectorizado que se acaba de mencionar basado en una fina malla de valores de  $\rho$  (gran  $q$ ), métodos de búsqueda no derivativos tales como el simplex o el esquema de búsqueda de bisección de Nelder-Mead, o mediante la aplicación de una técnica de optimización basada en derivadas. Algunas formas del método de Newton con derivadas numéricas tiene la ventaja de proveer el óptimo tales como la segunda derivada del log-verosimilitud concentrada en el óptimo  $\hat{\rho}$ . Este estimado numérico de la segunda derivada en conjunción con otra información puede ser muy útil en producir un estimado numérico de la matriz de varianza-covarianza para los parámetros.

Como se menciona arriba, una barrera aparente para implementar estos modelos para  $n$  grande, es la matriz  $W$  de  $n \times n$ . Si  $W$  contiene todos los elementos diferentes de cero, ello puede requerir una inmensa cantidad de memoria para almacenar la matriz por ejemplo para casos como los censos de Estados Unidos donde  $n > 60000$ . Afortunadamente,  $W$  es generalmente escasa lo que significa que tiene una gran proporción de ceros, por ejemplo si uno se basa en regiones contiguas o algún número  $m$  de regiones vecinas más cercanas para formar  $W$ , la matriz espacial de pesos contendrá solamente  $m \times n$  no-ceros como opuesto a una matriz densa con  $n^2$  elementos no-ceros. La proporción de no-ceros llega a ser  $m/n$  que disminuye con  $n$ . Matrices de contiguidad (de pesos) tienen un promedio de 6 vecinos por fila (para conjuntos de puntos espacialmente aleatorios sobre un plano). Como un ejemplo, usando los 3111 condados representando 48 estados más el distrito de Columbia, hay 9678321 elementos en la matriz  $W$  de  $3111 \times 3111$ , pero solamente  $3111 \times 6 = 18666$  serían diferentes de cero, es decir, 0.1929 por ciento de las entradas. Adicionalmente, calcular productos matriz-vector tales como  $Wy$  y  $WX$  toma mucho menos tiempo para matrices no densas. En ambos casos, matrices no densas requieren  $n$  operaciones lineales ( $O(n)$ ) mientras que una matriz densa requiere  $n$  operaciones cuadráticas ( $O(n^2)$ ). Técnicas con matrices no densas acelera grandemente el cálculo del logaritmo del determinante y otras cantidades de interés.

Para resumir, un número de técnicas facilitan el cálculo de estimaciones por máxima verosimilitud

para los modelos SDM y SAR. Estas técnicas incluyen la log-verosimilitud concentrada, cálculo previo del log-determinantes, así como también momentos tales como  $e_0^T e_d$  y usando una matriz no densa  $W$ . Tomando juntas, estas técnicas se reducen grandemente los cálculos así como también la cantidad de memoria requerida para resolver problemas con muestras grandes de datos.

### 2.5.2. Estimación del modelo SEM

El modelo SEM en que se toma en cuenta la dependencia espacial de los residuos se muestra en la ecuación (2.30), con el DGP (Proceso de Generación de Datos) para este modelo mostrado en (2.31), donde se define  $X$  como la matriz  $n \times k$  de variables independientes que puede o no incluir un término constante, y  $\beta$  el vector  $k \times 1$  asociado de parámetros.

$$y = X\beta + u \quad (2.30)$$

$$u = \lambda W u + \varepsilon$$

$$y = X\beta + (I_n - \lambda W)^{-1} \varepsilon \quad (2.31)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

La función log-verosimilitud completa viene dada por la ecuación (2.32).

$$\ln L = -(n/2) \ln(\pi \sigma^2) + \ln |I_n - \lambda W| - \frac{e^T e}{2\sigma^2} \quad (2.32)$$

$$e = (I_n - \lambda W)(y - X\beta)$$

Para un  $\lambda$  dado, la optimización de la función log-verosimilitud mostrado en (Ord, 1975; Anselin, 1988)  $\beta(\lambda) = (X(\lambda)^T X(\lambda))^{-1} X(\lambda)^T y(\lambda)$ , donde  $X(\lambda) = (X - \lambda W X)$ ,  $y(\lambda) = (y - \lambda W y)$ , y  $\sigma^2(\lambda) = e(\lambda)^T e(\lambda) / n$  donde  $e(\lambda) = y(\lambda) - X(\lambda) \beta(\lambda)$ . Por lo tanto se puede concentrar log-verosimilitud con respecto a  $\beta$  y a  $\sigma^2$  para llegar a log-verosimilitud concentrada como una función de  $\lambda$  como se muestra en (2.33).

$$\ln L(\lambda) = \kappa + \ln |I_n - \lambda W| - (n/2) \ln(S(\lambda)) \quad (2.33)$$

$$S(\lambda) = e(\lambda)^T e(\lambda) \quad (2.34)$$

A diferencia de los casos SAR o SDM,  $S(\lambda)$  no es un simple parámetro espacial cuadrático. Como se muestra en (2.34), evaluando la función log-verosimilitud concentrada para cualquier valor dado

de  $\lambda$  requiere manipulación de matrices  $n \times 1$  y  $n \times k$  para cada  $\lambda$  escogido. Esto llega a ser tedioso para grandes bases de datos. Sin embargo variables que requieren  $O(n)$  cálculos, pueden ser precalculadas, para que el cálculo de  $S(\lambda)$  durante la optimización sólo requiera trabajar con matrices de momentos de dimensión  $k \times k$  o menores. Estas matrices de momentos involucran las variables dependientes e independientes como una función de  $\lambda$ .

$$A_{XX}(\lambda) = X^T X - \lambda X^T W X - \lambda X^T W^T X + \lambda^2 X^T W^T W X$$

$$A_{Xy}(\lambda) = X^T y - \lambda X^T W y - \lambda X^T W^T y + \lambda^2 X^T W^T W y$$

$$A_{yy}(\lambda) = y^T y - \lambda y^T W y - \lambda y^T W^T y + \lambda^2 y^T W^T W y$$

$$\beta(\lambda) = A_{XX}(\lambda)^{-1} A_{Xy}(\lambda)$$

$$S(\lambda) = A_{yy}(\lambda) - \beta(\lambda)^T A_{XX}(\lambda) \beta(\lambda)$$

Con estos momentos y una malla de log-determinantes previamente calculados (acoplados con una rutina de interpolación) se realiza una actualización del log-verosimilitud en (2.33) para un nuevo valor de  $\lambda$ , de esta manera este proceso es rápido. Aplicando una técnica de optimización univariada tal como el método de Newton, ecuación (2.33) para encontrar  $\hat{\lambda}$  y sustituyendo esto en  $\sigma^2(\lambda)$ ,  $\beta(\lambda)$  y  $\Omega(\lambda)$ , permite las estimaciones de máxima verosimilitud, ecuaciones (2.35) a (2.37).

$$\hat{\beta} = \beta(\hat{\lambda}) \quad (2.35)$$

$$\hat{\sigma}^2 = n^{-1} S(\hat{\lambda}) \quad (2.36)$$

$$\hat{\Omega} = \hat{\sigma}^2 [(I_n - \hat{\lambda} W)^T (I_n - \hat{\lambda} W)]^{-1} \quad (2.37)$$

Como se indicó anteriormente, aplicando el método de Newton con derivación numérica para encontrar el óptimo, se produce un estimado numérico de la segunda derivada de la función log-verosimilitud concentrada en un óptimo  $\hat{\lambda}$ . Este estimado numérico de la segunda derivada, puede ser usado en conjunción con otra información para producir una matriz de varianza-covarianza estimada.

Notar que el modelo SDM anida al modelo SEM como un caso especial. Para ver esto, considérese una forma alternativa del modelo SEM en (2.38). Para evitar problemas de colinealidad para filas estocástica  $W$ , se asume que la matriz  $X$  no contiene un término constante y especifica esto separadamente. Esto es necesario para evitar crear un vector columna  $W \iota_n = \iota_n$  en  $WX$  que puede duplicar el término del intercepto.

$$y = \alpha_n t_n + X\beta + (I_n - \lambda W)^{-1} \varepsilon$$

$$(I_n - \lambda W)y = \alpha_n (I_n - \lambda W)t_n + (I_n - \lambda W)X\beta + \varepsilon$$

$$y = \lambda W y + \alpha (I_n - \lambda W)t_n + X\beta + WX(-\beta\lambda) + \varepsilon \quad (2.38)$$

El modelo en (2.38) representa un modelo SDM donde el parámetro sobre los rezagos espaciales de las variables independientes ( $WX$ ) ha sido restringido e igualado a  $-\beta\lambda$ . Estimando el modelo más general SDM  $y = \lambda W y + X\beta + WX\theta + \varepsilon$  y realizando un test para la restricción  $\theta = -\beta\lambda$  puede llevar al rechazo del modelo SEM relativo a SDM.

# Capítulo 3

## TRATAMIENTO A VALORES DE DATOS EMPLEADOS

Para el tratamiento que se ha dado a los valores disponibles de las variables empleadas, se cita como ejemplo el tratamiento dado a la variable de interés (variable dependiente “*lprob13*”), esta variable corresponde al logaritmo natural de la tasa formada por la cantidad de robos en cada zona censal del Distrito Metropolitano de Quito, en el año 2013 y por cada mil habitantes de dichas zonas. El empleo de tasas así como del logaritmo, se justifica debido a la naturaleza exploratoria que posee este estudio. De manera similar se procede con la mayoría de valores para los datos correspondientes tanto a los tipos de robos, cantidad de población, áreas geográficas; como con algunas de las variables socioeconómicas tomadas para este estudio.

Con los datos así obtenidos se pretende, detectar la presencia de dependencia espacial entre sus valores, encontrar patrones espaciales de concentración y difusión, que serán evidenciados en los respectivos mapas de Quito cuya división territorial interna está a nivel de zonas censales, utilizando para ello el software GeoDa.

### **3.1. Descripción y tratamiento de valores de datos de las variables empleadas**

Las tablas que se han elaborado, contienen datos que provienen de la tabulación, unión, compatibilización, y tratamiento de los valores de las variables tanto socioeconómicas y del número de habitantes de cada zona censal extraídos del último Censo; como de la cantidad de robos registrados en el ex CEASI valores que son concensuados con la Fiscalía, Ministerio del Interior y Policía del DMQ, para cada zona censal durante el año 2013 y el primer semestre del año 2014, los mismos que posteriormente se los vinculará a su respectivo mapa de la ciudad cuyas divisiones territoriales internas también están a nivel de zonas censales.

Se ha establecido de acuerdo a la investigación preliminar realizada que las zonas censales cuentan como mínimo con poblaciones de alrededor de 1500 habitantes, y como para datos de poblaciones, generalmente se habla del valor de la variable en tasas por mil, por diez mil, etc., entonces en primera instancia a algunos valores de las variables, se las transformó, a tasas por mil, con el fin de reducir posibles efectos adversos asociados a la diferencia de tamaño de las unidades geográficas (zonas censales). Adicionalmente, se calcula el logaritmo natural de dichas tasas, tendiente esto a reducir la influencia de valores extremos en el modelo que se obtendrá como consecuencia de todo este análisis.

Se presenta a continuación un resumen de la tabla generada con las variables que se han considerado, el significado de ellas y la forma como fueron tratadas.

VARIABLE	DESCRIPCIÓN (para cada Zona Censal)	TRATAMIENTO
DPA_ZONA	División Político Administrativa de zonas censales	
PARR	Parroquia urbana de Quito	
pobla	Número de personas	
pob5m	Número de personas mayores a 5 años	
hom	Número de hombres	
muj	Número de mujeres	
esprm	Escolaridad promedio de las personas que habitan cada zona censal	Para cada zona censal, se considera los años de escolaridad de las personas que viven en dicha zona y se calcula su promedio
nanlf	Número de personas que NO son analfabetas	
anlf	Número de personas analfabetas	
nviv	Número de viviendas	
nagua	Número de viviendas que no tienen acceso a red pública de agua	
nalcan	Número de viviendas que no tienen acceso a red de alcantarillado público	
emig	Número de viviendas en las cuales existe al menos una persona migrante	
hcin	Número de viviendas que presentan hacinamiento	Para cada zona censal, se consideran a aquellas viviendas en las cuales existen 4 o más personas por dormitorio
nhog	Número de hogares	
hpobr	Número de hogares pobres	Pobreza calculada mediante criterio de necesidades básicas insatisfechas
hpext	Número de hogares en pobreza extrema	Pobreza calculada mediante criterio de necesidades básicas insatisfechas
rob13	Número de robos (año 2013)	
rob13m	Número de robos a mujeres (año 2013)	
rob13h	Número de robos a hombres (año 2013)	
area	Área en metros cuadrados	Área de cada polígono
phom	Tasa de hombres por 1000 habitantes	$phom * 1000 / pobla$
pmuj	Tasa de mujeres por 1000 habitantes	$pmuj * 1000 / pobla$
pnanlf	Tasa de NO analfabetos por 1000 habitantes mayores de 5 años	$pnanlf * 1000 / pob5m$
panlf	Tasa de analfabetos por 1000 habitantes mayores de 5 años	$panlf * 1000 / pob5m$
pnagua	Tasa de viviendas que no tienen acceso a agua por red pública por 1000 viviendas	$pnagua * 1000 / nviv$

VARIABLE	DESCRIPCIÓN (para cada Zona Censal)	TRATAMIENTO
pnalcan	Tasa de viviendas que no tienen acceso a red de alcantarillado público por 1000 viviendas	$pnalcan*1000/nviv$
pemig	Tasa de viviendas en las cuales existe al menos una persona migrante por 1000 viviendas	$pemig*1000/nviv$
phcin	Tasa de viviendas que presentan hacinamiento por 1000 viviendas	$phcin*1000/nviv$
phpobr	Tasa de hogares pobres por 1000 hogares	$phpobr*1000/nhog$
phpext	Tasa de hogares en pobreza extrema por 1000 hogares	$phpext*1000/nhog$
prob13	Tasa de robos (año 2013) por 1000 habitantes	$prob13*1000/pob5m$
prob13m	Tasa de robos a mujeres (año 2013) por 1000 habitantes	$prob13m*1000/muj$
prob13h	Tasa de robos a hombres (año 2013) por 1000 habitantes	$prob13h*1000/hom$
pob_ar	Tasa de habitantes por área	$pob*1000000 /área$
viv_ar	Tasa de viviendas por área	$nviv*1000000 /área$
hom_ar	Tasa de hombres por área	$hom*1000000 /área$
muj_ar	Tasa de mujeres por área	$muj*1000000 /área$
lpobla	Logaritmo natural de número de habitantes	$\ln*pobla$
lnviv	Logaritmo natural de número de viviendas	$\ln*nviv$
larea	Logaritmo natural del área	$\ln*area$
lpob_ar	Logaritmo natural de la tasa de habitantes por área	$\ln*pob_ar$
lviv_ar	Logaritmo natural de la tasa de viviendas por área	$\ln*viv_ar$
lhom_ar	Logaritmo natural de la tasa de hombres por área	$\ln*hom_ar$
lpmuj	Logaritmo natural de tasa de mujeres por 1000 habitantes	$\ln(pmuj+0.5)$
lpnanlf	Logaritmo natural de tasa de NO analfabetos por 1000 habitantes mayores de 5 años	$\ln(pnanlf+0.5)$
lpnanlf	Logaritmo natural de tasa de analfabetos por 1000 habitantes mayores de 5 años	$\ln(pnanlf+0.5)$
lpnagua	Logaritmo natural de tasa de viviendas que no tienen acceso a agua por red pública por 1000 viviendas	$\ln(pnagua+0.5)$
lpnalcan	Logaritmo natural de tasa de viviendas que no tienen acceso a red de alcantarillado público por 1000 viviendas	$\ln(pnalcan+0.5)$
lpemig	Logaritmo natural de tasa de viviendas en las cuales existe al menos una persona emigrante por 1000 viviendas	$\ln(pemig+0.5)$
lphcin	Logaritmo natural de tasa de viviendas que presentan hacinamiento por 1000 viviendas	$\ln(phcin+0.5)$
lphpobr	Logaritmo natural de tasa de hogares pobres por 1000 hogares	$\ln(phpobr+0.5)$
lphpext	Logaritmo natural de tasa de hogares en pobreza extrema por 1000 hogares	$\ln(phpext+0.5)$
lprob13	Logaritmo natural de tasa de robos (año 2013) por 1000 habitantes	$\ln(prob13+0.5)$
lprob13m	Logaritmo natural de tasa de robos a mujeres (año 2013) por 1000 habitantes	$\ln(prob13m+0.5)$

Cuadro 3.1: Variables y descripción del tratamiento dado a sus valores

## 3.2. Análisis Exploratorio de Datos Espaciales (ESDA)

El análisis exploratorio de datos espaciales, consiste en una serie de técnicas que permiten visualizar y estimar la presencia de autocorrelación espacial, el programa GeoDa, nos permite aplicar varias de estas técnicas de análisis por medio de herramientas tales como:

*Quantil Map*

*Box Plot*

*Scatter plot*

*3 – D Scatter plot,*

entre otras, que son aplicaciones útiles, para explorar y analizar cualquier variable de interés.

El objetivo principal de este programa es proporcionar dichos análisis, comenzando con mapeos y geo-visualizaciones que posteriormente permiten pasar al análisis de autocorrelación espacial y terminar con la respectiva regresión espacial.

### 3.2.1. Quantil Map

Como ejemplo de esta aplicación se presenta el *Quantil Map*, correspondiente al análisis del número de robos reportados en cada zona censal del Distrito Metropolitano de Quito para su parte Urbana, durante el año 2013; figura (3.1). Para ello GeoDa ordena internamente los datos y acorde a los requerimientos, los divide en el número de grupos que se desee, procurando contener igual número de observaciones en cada uno de ellos. Para nuestro caso, se selecciona 10 grupos, que permite el análisis por deciles. Como dato informativo, si se hubiera seleccionado, cuatro grupos, implícitamente se está solicitando al programa, los datos correspondientes a los cuartiles, que generalmente son los más usados en el análisis descriptivo de ellos.

Analizando los valores reportados, se tiene para el último decil: [3.17:5.79] (45), que indica la existencia de 45 zonas censales en las que ha ocurrido la mayor cantidad de robos, esto es de 24 a 327,<sup>1</sup> robos por cada 1000 habitantes mayores de 5 años.

En el mapa ver figura (3.1), se puede apreciar que las zonas censales asociadas al último decil, se resaltan en color rojo intenso, y conforme decrece la intensidad de estos colores, hasta llegar al

---

<sup>1</sup> $e^{3,17} \approx 24$  y  $e^{5,79} \approx 327$

color blanco que a su vez representa al primer decil, y cuyos valores  $[-2.07:-0.472](45)$  indican que se tienen 45 zonas censales en las que se ha reportado, desde 0.126 hasta 0.624,<sup>2</sup> robos. Es preciso aclarar que estos valores, en realidad corresponden a zonas censales en donde no se reportaron robos durante el año 2013, los valores presentados menores a 1, se deben al hecho de que inicialmente, a la cantidad de robos ocurridos en el año 2013 (rob13), se los transformó a tasa por cada mil habitantes mayores de 5 años, pasando entonces a ser una tasa de los robos (prob13), luego a ésta tasa y con la finalidad de obtener su logaritmo y tratando de evitar problemas en zonas donde no se reportaron robos pues el logaritmo de cero no está definido, se incrementó a dicha tasa 0.5, obteniéndose entonces (prob13+0.5), y al obtener su logaritmo pasa a ser (lprob13) que representa finalmente el logaritmo natural de la tasa de robos reportada en cada zona censal durante el año 2013, este proceso se describe en la tabla presentada anteriormente. Cuadro (3.1).

### 3.2.2. Box Plot

El diagrama de caja o “*box plot*” ubicado junto al “*Quantil Map*”, figura (3.1), para este ejemplo proporciona información relacionada a los datos ubicados en el primer cuartil ( $0,6168 \approx (2 \text{ robos})$ ), segundo cuartil o mediana ( $1,411 \approx (4 \text{ robos})$ ), media ( $1,391 \approx (4 \text{ robos})$ ), tercer cuartil ( $2,312 \approx (10 \text{ robos})$ ), el rango intercuartílico ( $1,695 \approx (5 \text{ robos})$ ), valor mínimo ( $-2,072 \approx (0 \text{ robos})$ ), valor máximo ( $5,79 \approx (327 \text{ robos})$ ), y desviación estándar ( $1,493 \approx (4,5 \text{ robos})$ ), este tipo de gráficos en estadística descriptiva permite detectar posibles valores atípicos, que generalmente, se encuentran sobre y bajo los bigotes. GeoDa permite seleccionar en el gráfico aquellos datos que se encuentran fuera de los bigotes y los relaciona dinámicamente, con las zonas censales a los que ellos representan, marcándose de un color diferente (generalmente amarillo) tanto en el mapa como en la tabla de datos que las contienen. Los datos atípicos así obtenidos no corresponden a los datos atípicos (outliers) correspondientes al análisis espacial de los mismos.

---

<sup>2</sup> $e^{-2,07} \approx (0,126)$     y     $e^{-0,472} \approx (0,624)$

## Número de robos por zona censal en Quito Urbano (Año 2013)

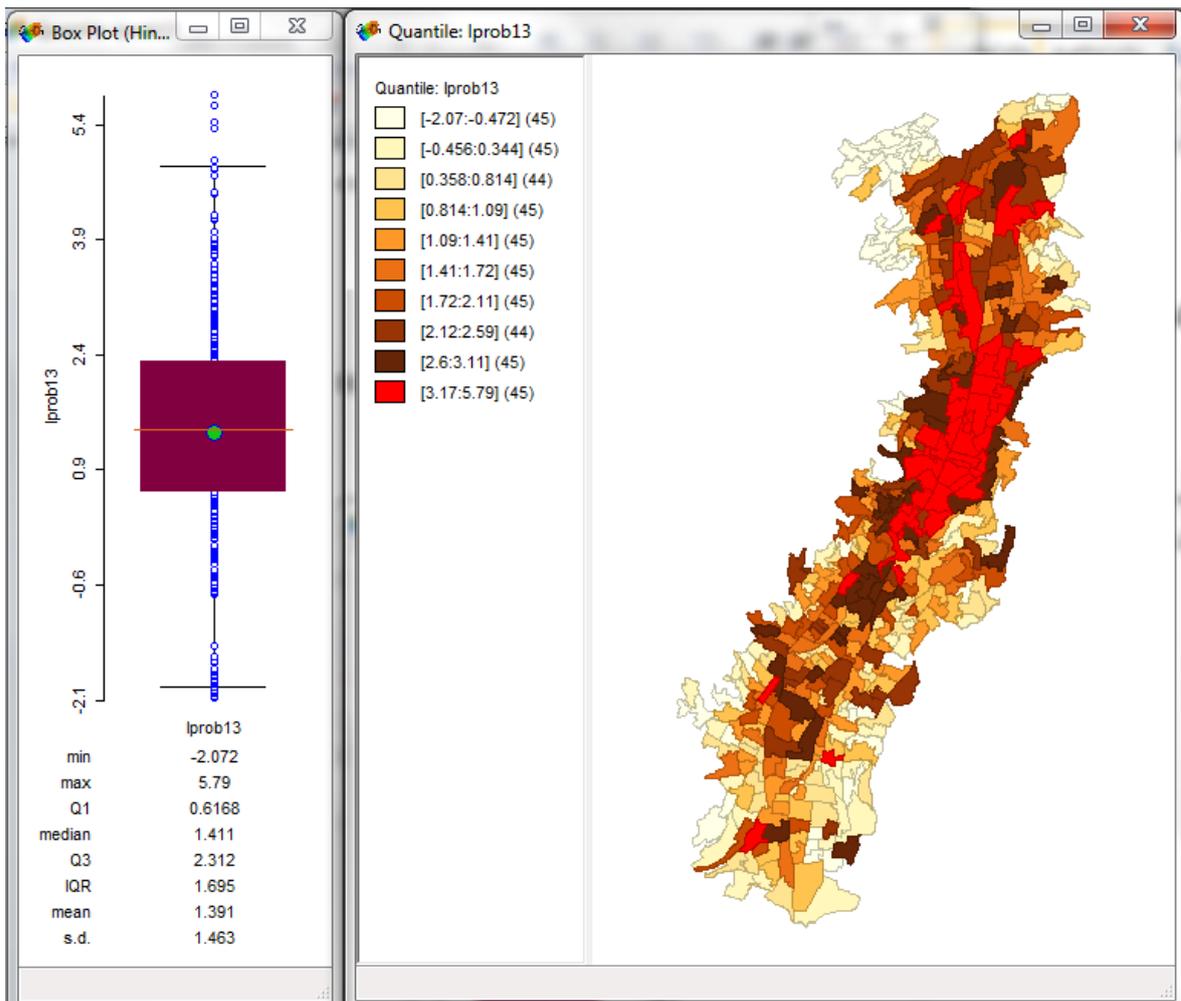


Figura 3.1: Box plot y Quanti Map para (*lprob13 en deciles*)

De forma análoga al ejemplo anterior se obtiene el Cuantil Map y su respectivo diagrama de caja, para la variable (*lphcin*) logaritmo natural de tasa de viviendas que presentan hacinamiento por cada 1000 viviendas, figura (3.2). Donde los valores del último decil [4.95;5.29](45) corresponden a 45 zonas censales que contendrán entre  $e^{4,95} \approx (142)$  y  $e^{5,29} \approx (199)$  viviendas, que presentan hacinamiento por cada 1000 viviendas.

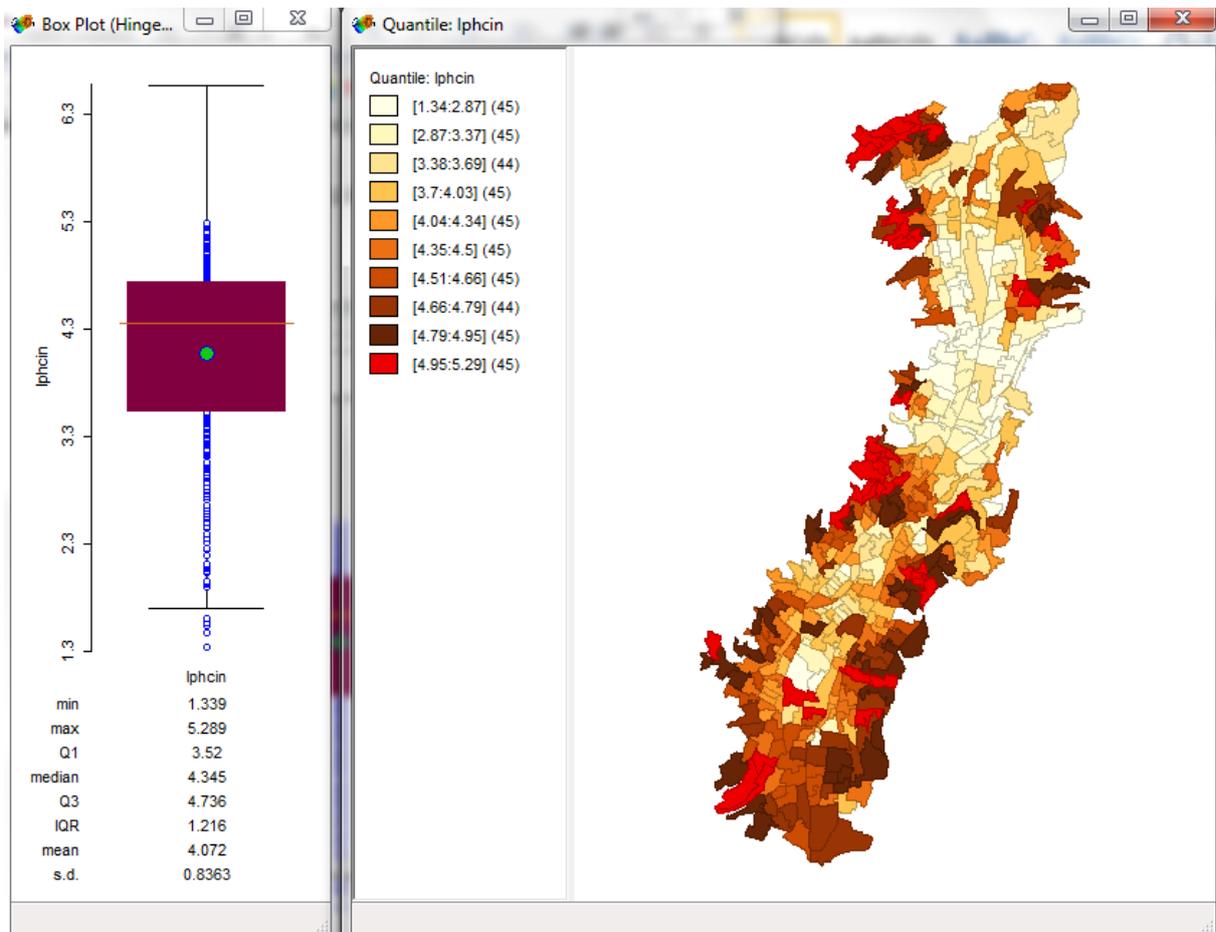


Figura 3.2: Box plot y Quanti Map para Viviendas que presentan hacinamiento (*lphcin en deciles*)

En el siguiente gráfico se pretende poner de manifiesto las zonas censales con mayor presencia de personas que han tenido acceso a la educación o a cierto nivel de estudio y compararlo con las zonas en donde se registra la mayor tasa de robos, pudiéndose notar que de alguna manera si estos dos mapas se superponen, los colores más intensos, y que corresponde a los niveles más altos de estas dos variables, en gran parte de ellas coinciden. Sin querer decir con ello que a mayor educación mayor tasa de robos, pero si nos permite pensar que en estos sectores con mejor estatus de vida, que generalmente está asociado al nivel académico de sus habitantes, pueden ser atractivos para la delincuencia.

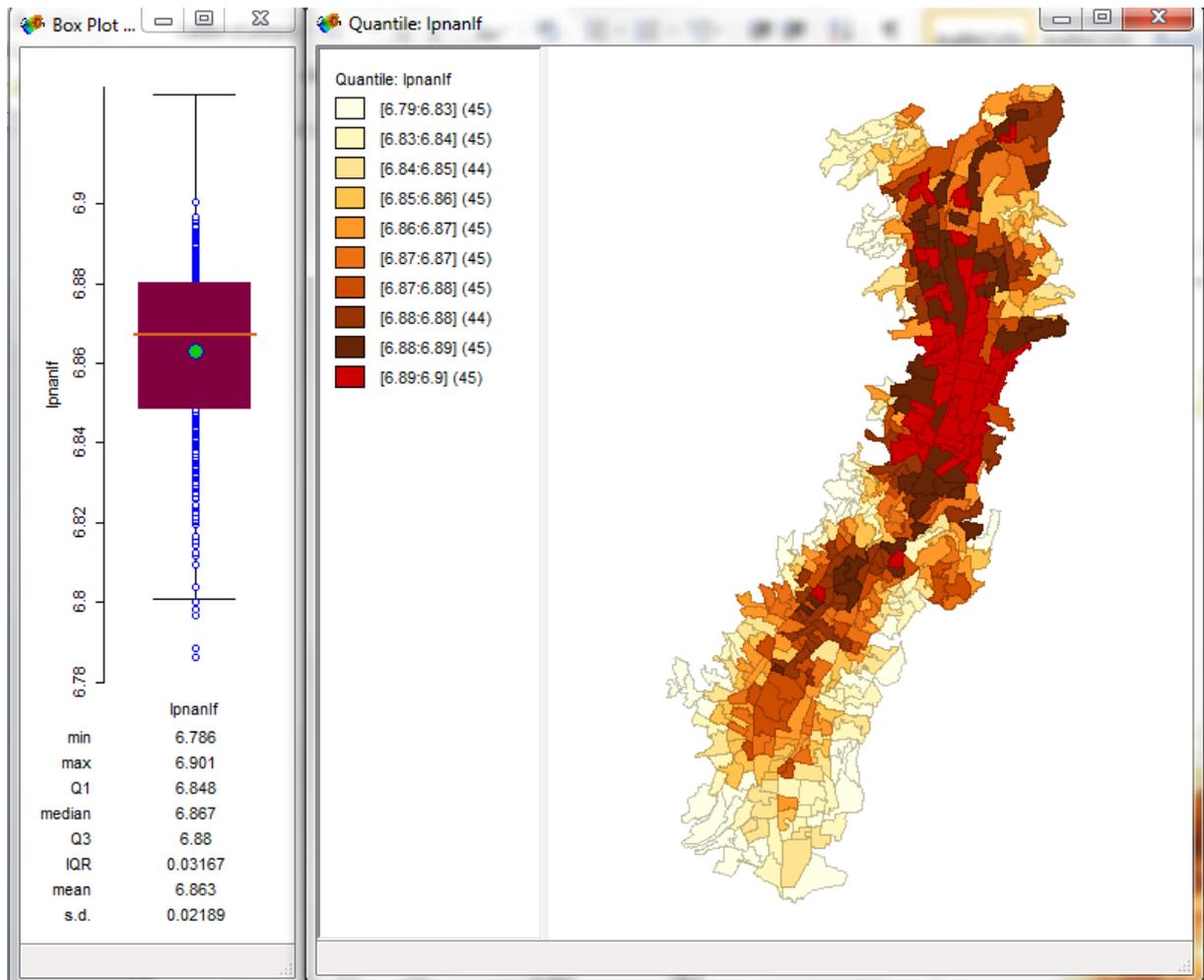


Figura 3.3: Box plot y Quanti Map para Número de no Alfabetos por zona censal en Quito Urbano (*lpananf en deciles*)

En el siguiente mapa y como exploración se trató de descubrir, si las zonas con mayor presencia femenina, puede ser un blanco para los delincuentes, de forma análoga a lo anterior, se puede visualizar que las zonas con más alta tasa de robos coincide con zonas donde la presencia femenina predomina.

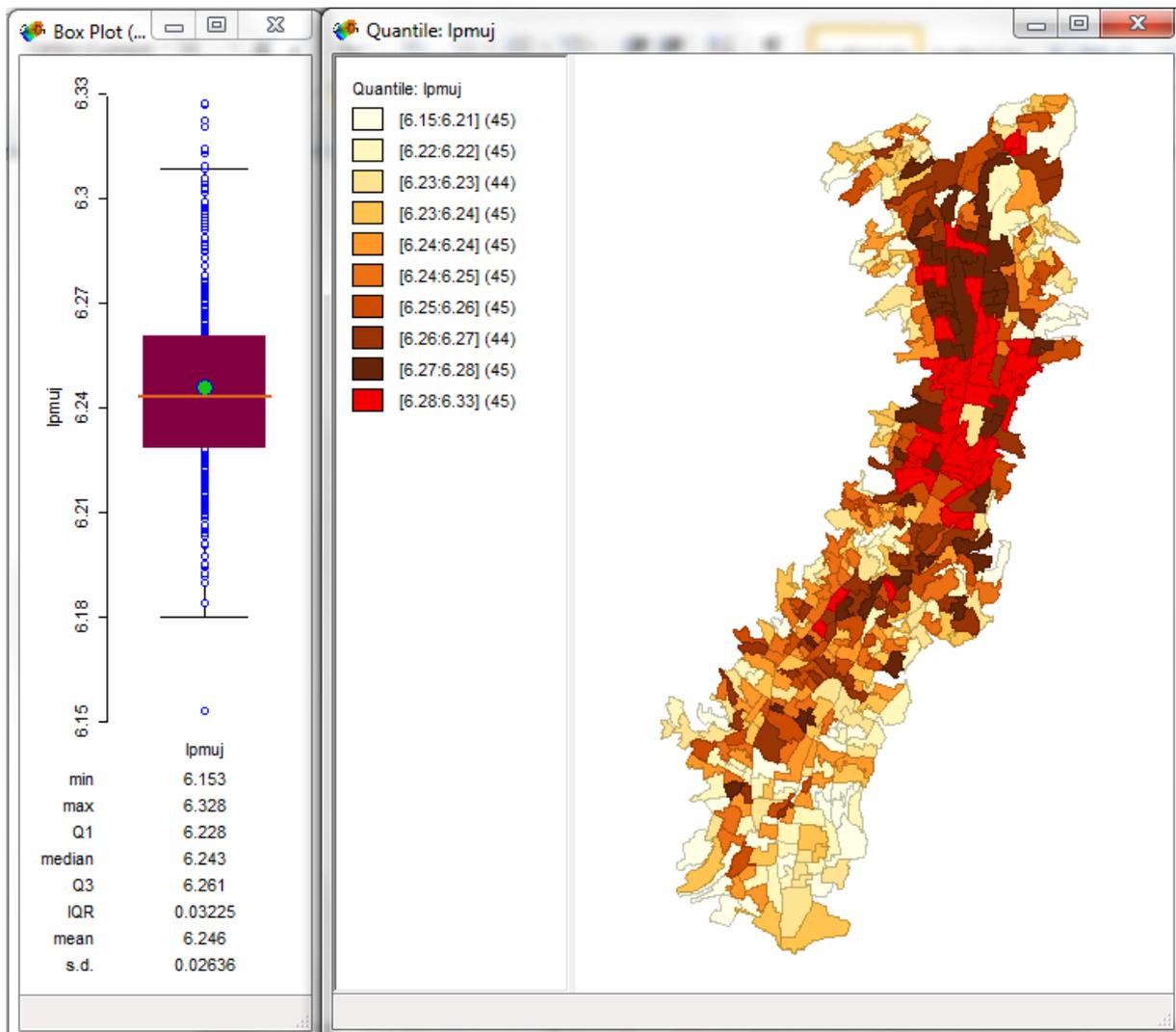


Figura 3.4: Box plot y Quanti Map para Número de mujeres por zona censal en Quito Urbano (*lpmuj en deciles*)

Finalmente se buscó determinar si las zonas en donde los hogares tienen personas que han migrado serían afectados por los delincuentes, para este caso en cambio se puede ver que no se comportan de la misma manera como en los casos anteriores por lo que se cree que no estarían relacionadas de manera relevante.

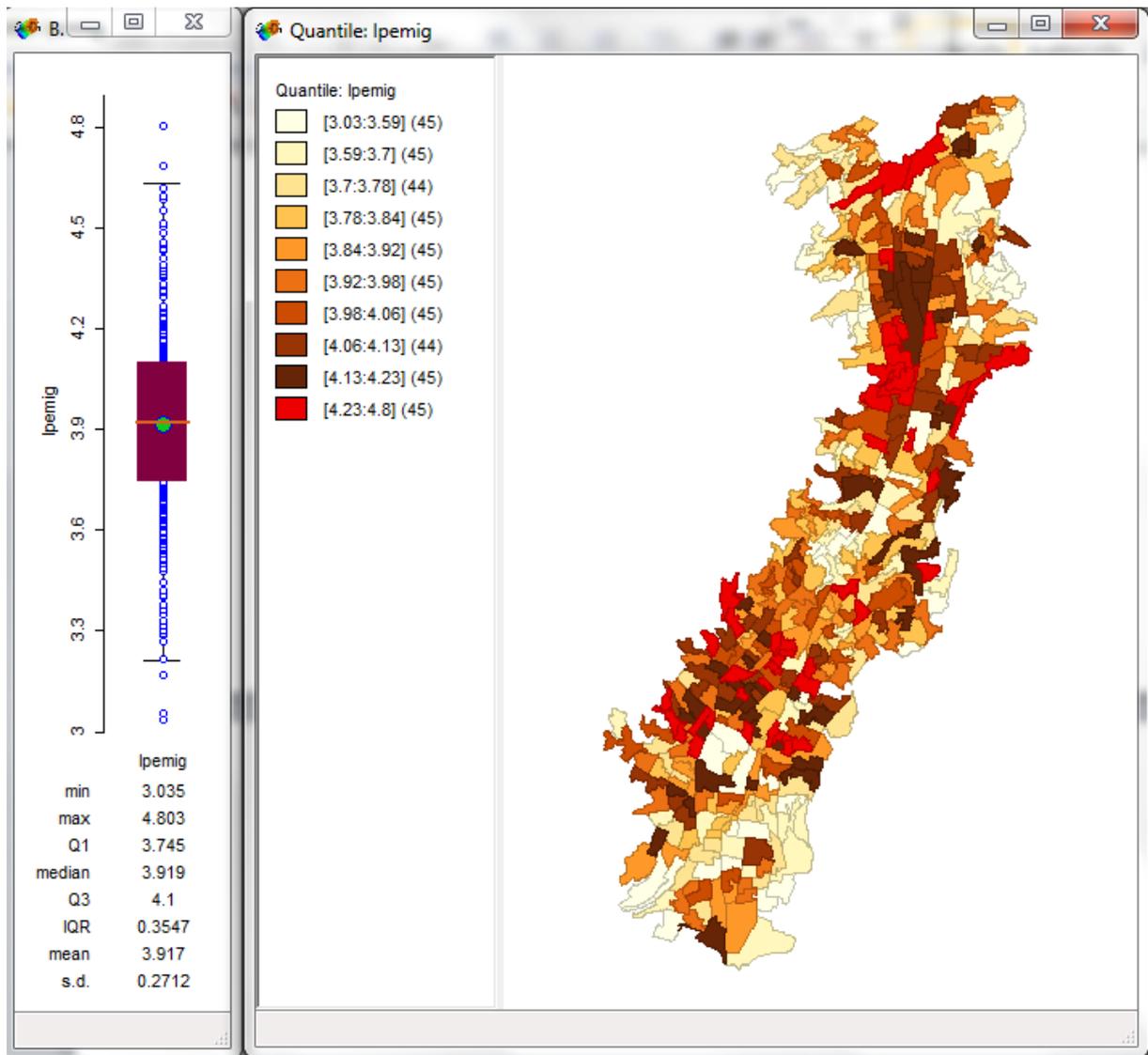


Figura 3.5: Box plot y Quanti Map para Número de Viviendas con personas que han migrado de Quito Urbano (*Ipemig en deciles*)

### 3.2.3. Scatter Plot

Esta aplicación genera un gráfico de dispersión bivariado, para su exploración inicialmente consideramos las variables, logaritmo de la tasa de robos durante el año 2013 ( $\ln(\text{prob13})$ ) ubicada en el eje  $Y$  y el logaritmo de la tasa de mujeres ( $\ln(\text{pmuj})$ ) en el eje  $X$ . En el gráfico resultante se visualiza la dispersión de los datos y la recta correspondiente al ajuste lineal por mínimos cuadrados, que refleja una dependencia positiva de los mismos, y en la parte inferior se presentan, los resultados correspondientes al, número de observaciones (448), coeficiente de determinación ( $R^2 = 0,241$ ), constante (intercepto) (-169), error estándar de la constante (14,3), estadístico  $t$  para la constante (-11,8),  $p$ -valor para la constante ( $3,99e^{-28}$ ), pendiente (27,2), error estándar para la pendiente (2,29), estadístico  $t$  para la pendiente (11,9), y  $p$ -valor para la pendiente (0), datos que corresponden a los estadísticos asociados a las dos variables analizadas. Figura 3.6.

De forma análoga se procede para el análisis de las variables: viviendas con migrantes vs. tasa de robos durante el año 2013. Figura 3.7.

Este procedimiento lo repetimos, relacionando siempre la tasa de robos del año 2013, con algunas de las variables socioeconómicas extraídas de las bases de datos originales, para tratar de visualizar el grado de dependencia que puede haber entre ellas. De las que finalmente tomamos las que presentaban mayores valores de dependencia

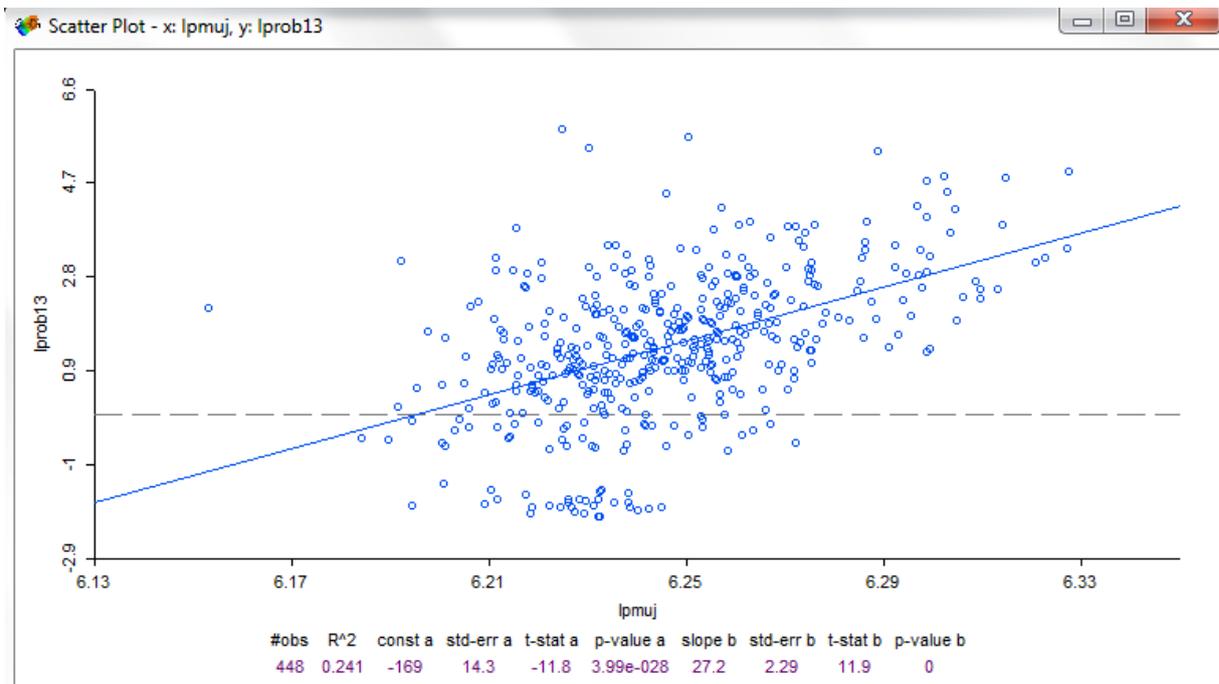


Figura 3.6: Scatter Plot.- Mujeres – Robos (*lpmuj – lprob13*)

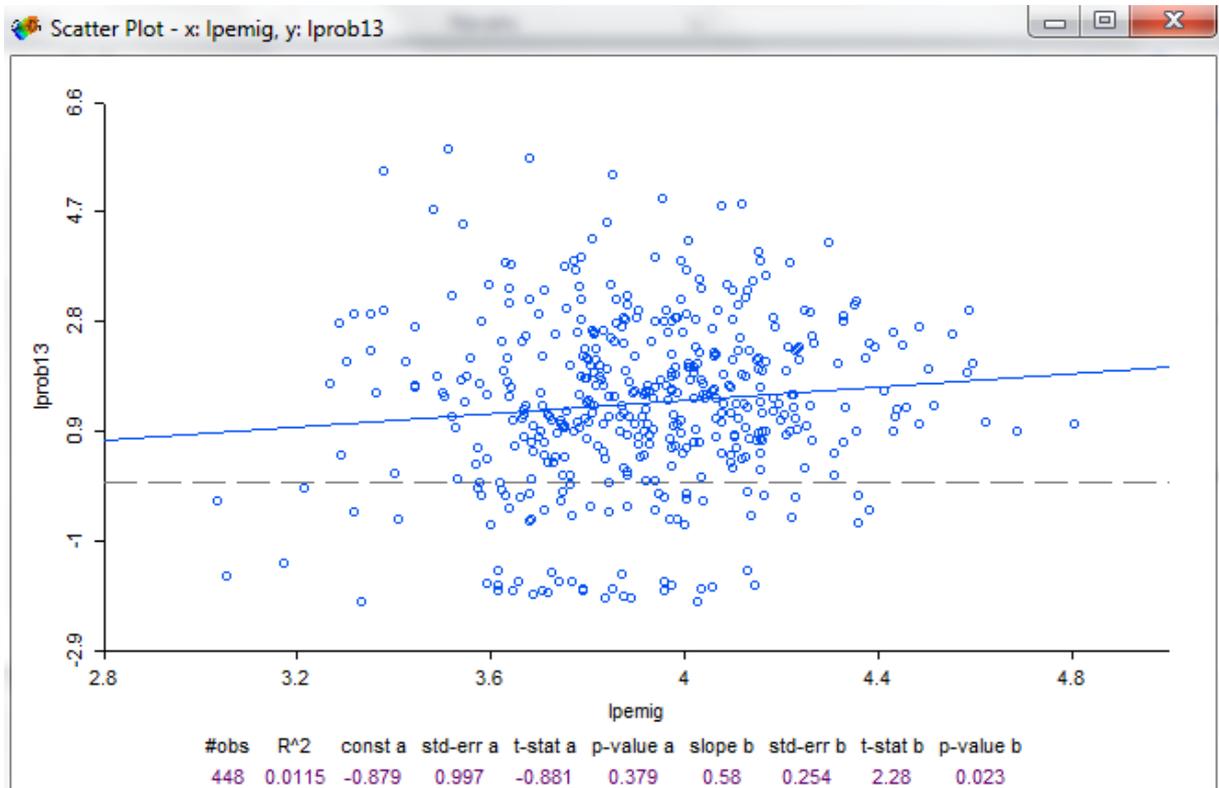


Figura 3.7: Scatter Plot.- Viviendas con migrantes – Robos (*lpemig – lprob13*)

### 3.2.4. Linking and Brushing

Como una de las principales herramientas que dispone GeoDa, esta la posibilidad de ejecutar tareas denominadas, "*Linking and Brushing*", que consiste en seleccionar marcando ya sea en el mapa, en la tabla de datos o en los diferentes gráficos estadísticos, una porción de datos, para eliminar su efecto sobre los demás, para eliminar el efecto de datos atípicos si ese fuera el interés, o simplemente para realizar análisis de una porción de la totalidad de ellos. Esto es posible seleccionándolos, en una de las tres opciones antes mencionadas, y GeoDa recalcula automáticamente entregando los nuevos valores de los diferentes estadísticos, tanto para los datos seleccionados, como para los datos sobrantes. Es importante hacer notar que una vez seleccionados, en cualquiera de las tres opciones (mapa, tabla o gráfico estadístico), dicha selección, se vincula dinámicamente con las restantes, apareciendo en todas ellas marcadas con un distintivo color amarillo.

## 3.3. Exploración de Datos Multivariados

### 3.3.1. 3-D Scatter plot

Es una técnica considerada para explorar asociaciones multivariadas y que consiste en un diagrama de dispersión en tres dimensiones. Como ejemplo presentamos el gráfico correspondiente a las variables asociadas a la tasa de hogares con personas que han salido del país (lpemig), la tasa de mujeres (lpmuj) y la tasa de robos en Quito Urbano (lprob13). Figura 3.8.

En la figura 3.9, se puede observar la distribución de las proyecciones de los datos sobre cada uno de los planos existentes.

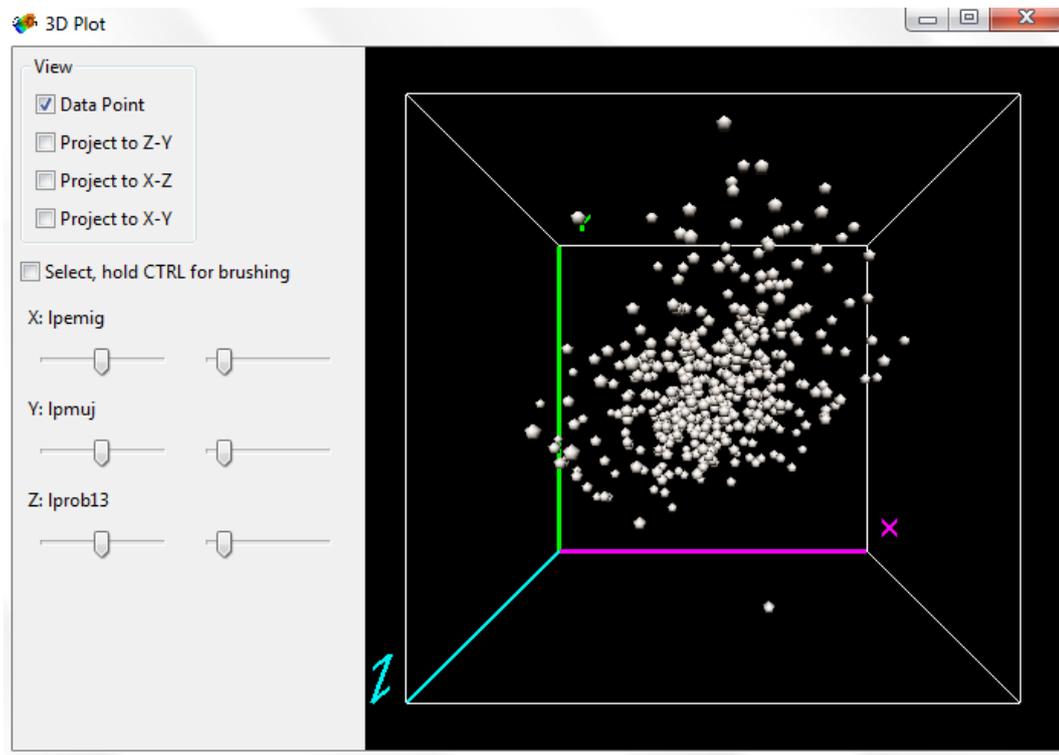


Figura 3.8: 3D Scatter Plot.- Migrantes – Mujeres – Robos (*lpemig-lpmuj-lprob13*)

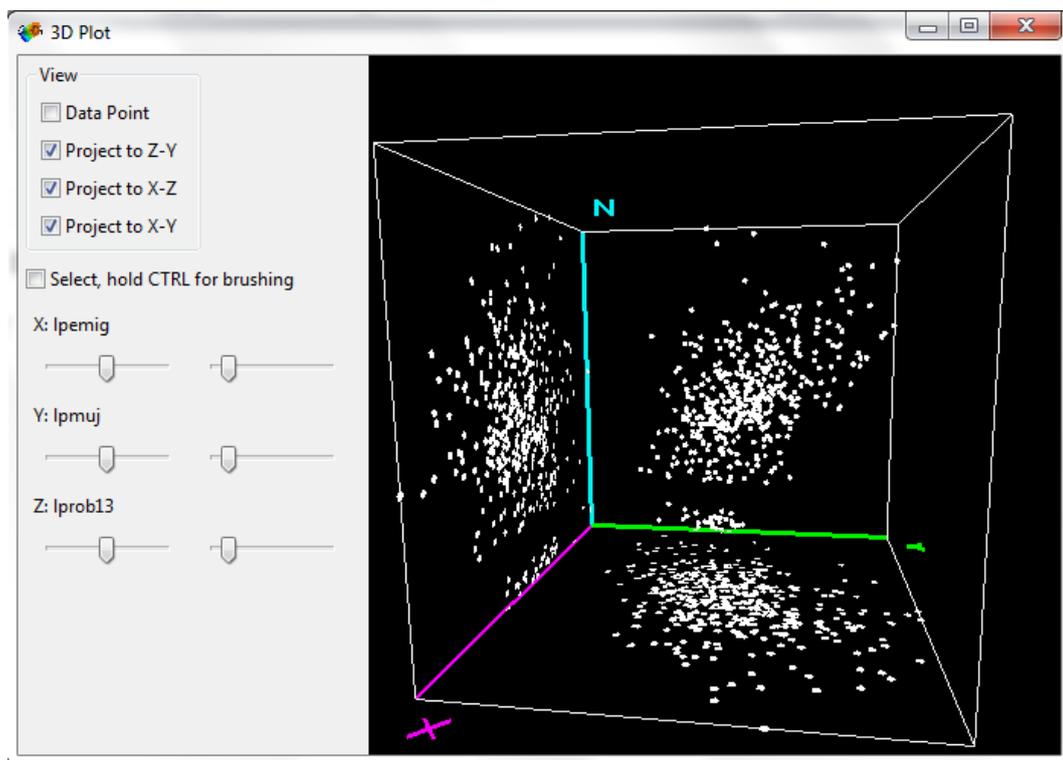


Figura 3.9: 3D Scatter Plot.- Migrantes – Mujeres – Robos (*proyección de los datos sobre los planos*)

### 3.4. Detección y Medición de Autocorrelación Espacial

En problemas con datos de tipo espacial, generalmente no es posible aislar una observación de otra, pues normalmente existe autocorrelación espacial. Entonces el análisis de autocorrelación espacial, se emplea para verificar si se cumple la hipótesis de que una variable (por ejemplo los robos) tiene o no, una distribución aleatoria, es decir puede ser que exista una agrupación considerable de valores similares, altos o bajos entre zonas censales vecinas (autocorrelación positiva), o por el contrario se tenga valores no similares entre zonas vecinas (autocorrelación espacial negativa). El término "*significativo*", en la mayoría de la literatura estadística corresponde a un umbral de porcentajes generalmente escogido de manera empírica.

El estadístico de prueba que sirve para medir la autocorrelación espacial, es el índice Global  $I$  de Moran (Moran, 1950). Este índice proporciona un valor como medida resumen de la intensidad de la correlación entre los valores de las zonas censales consideradas. Está definido por:

$$I = \frac{m \sum_{i=1}^m \sum_{j=1}^m W_{ij} (p_i - \bar{p})(p_j - \bar{p})}{S_o \sum_{j=1}^m (p_i - \bar{p})^2} \quad i \neq j \quad (3.1)$$

donde:

$m$  es el número de zonas censales (448),

$$S_o = \sum_{i=1}^m \sum_{j=1}^m W_{ij}$$

$x_i$  población en riesgo en la zona censal  $i$ ,

$n_i$  número de robos en la zona censal  $i$ ,

$p_i = \frac{n_i}{x_i}$  tasa observada de robos,

$\bar{p} = \sum_i \frac{p_i}{m}$  es el promedio de las tasas observadas  $p_i$  en todas las zonas censales,

$W_{ij}$  son los pesos de la matriz de ponderación o contiguidad espacial,

$W_{ij} = 1$  si  $i \neq j$ , zonas censales vecinas, y

$W_{ij} = 0$  para cualquier otro caso.

A decir de algunos autores, es parecido al coeficiente de Pearson, pues su valor varía entre +1 y -1. Donde +1 indica una autocorrelación "*positiva*" perfecta, y -1 corresponde a una autocorrelación "*negativa*" perfecta, valores de 0 o cercanos a él, indicará ausencia de autocorrelación.

El índice de Moran se basa en que la asociación entre los valores del conjunto de datos está determinada por la matriz de contiguidad (matriz de pesos espaciales), que toma valores de 1 para

zonas censales adyacentes (vecinos de primer orden) y 0 para los demás casos. Los vecinos se pueden definir de acuerdo a los movimientos de la Torre en el juego de Ajedrez en donde se requiere que todo un borde (frontera) sea común entre zonas censales vecinas o de acuerdo a los movimientos de la Reina, en donde se requiere que cualquier parte del borde (frontera) o vértices de las zonas censales vecinas se toquen, es decir serán vecinos todos aquellos que compartan fronteras o vértices comunes.

Una vez que se tiene disponible la cartografía digital correspondiente a las zonas censales del Distrito Metropolitano de Quito, filtrada exclusivamente para esta unidad geográfica y para la parte Urbana de la ciudad, empleando las herramientas del programa GeoDa, se construye el ponderador correspondiente a la matriz de pesos espaciales, que se almacena como un nuevo archivo con extensión *\*.gal* para la variable de interés y se determina cuál será el criterio de vecindad a emplear, para este estudio se eligió el criterio de la Reina. El programa crea automáticamente dicho archivo y un identificador para cada una de las zonas censales.

Como ejemplo se presenta más adelante la figura (3.14), denominado scatterplot de Moran I, en el que se tiene la tasa normalizada de robos durante el año 2013 para las 448 zonas censales, en el eje de las abscisas y el denominado rezago espacial de esta misma tasa, definido como el producto del vector de observaciones de  $x$  y la matriz de pesos espaciales  $W$ , en el eje de las ordenadas.

### 3.5. Índice de Morán en forma matricial

La distribución de este estadístico estandarizado  $Z(I)$  se estima que es normal cuando el tamaño muestral es suficientemente grande, por tanto este coeficiente I de Moran, se ajustará a una prueba de significancia estadística de valores “ $Z$ ”, es decir se partirá del supuesto de una distribución normal (Cliff & Ord, 1981).

$$I_{lk} = \frac{Z_k^T W Z_l}{Z_k^T Z_l} = \frac{Z_k^T W Z_l}{n} \quad (3.2)$$

Donde:

$I_{lk}$  : índice de Morán generalizado para las variables  $Z_k$  y  $Z_l$

$Z_k$  y  $Z_l$  : variables estandarizadas con media cero y varianza uno

$W$  : matriz estandarizada de pesos espaciales

$n$  : número de observaciones.

### 3.6. Indicador Local de Autocorrelación Espacial (LISA)

Como la asociación significativa entre los valores de variable puede no darse en toda la ciudad, sino en determinadas zonas, se recurre entonces a encontrar el indicador local de autocorrelación espacial LISA,<sup>3</sup> cuyo propósito es obtener el estadístico para cada zona censal, que proporcione información sobre la relevancia de valores similares que se tengan alrededor de ella. El estadístico para autocorrelación espacial local  $I_i$  de Moran (Anselin, 1995) está definido por:

$$I_i = \frac{(x_i - \bar{x})}{S_i^2} \sum_{j=1}^n W_{ij}(x_j - \bar{x}) \quad j \neq i \quad (3.3)$$

donde:

$$S_i^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n - 1} - (\bar{x})^2 \quad j \neq i \quad (3.4)$$

$n$  número de zonas censales (448),

$x_i$  tasa observada de robos en la zona censal  $i$ ,

$\bar{x}$  es el promedio de las tasas observadas  $x_i$  considerando todas las zonas censales,

$W_{ij}$  son los pesos de la matriz de ponderación o contigüidad espacial,

De forma simplificada el indicador local  $I_i$  de Moran para cada zona censal, puede ser calculado con la siguiente expresión :

$$I_i = z_i \sum_{j=1} W_{ij} z_j \quad (3.5)$$

donde:

$z_i$  es la variable original  $x_i$  en su forma estandarizada

$$z_i = \frac{x_i - \bar{x}}{SD_x}$$

$W_{ij}$  es la matriz de contigüidad, y el

$\sum_j$  es a través de cada fila  $i$  de la matriz de pesos espaciales.

Al igual que el global  $I$  de Moran, este indicador también varía entre  $-1$  y  $+1$ , y representa el grado de correlación entre los valores de la variable (robos) de una zona censal con los valores de la variable en las zonas vecinas.

El indicador local  $I_i$  de Moran, ofrece a través de sus mapas de significancia y del scatterplot,

<sup>3</sup>LISA Local Indicator of spatial autocorrelation

figura (3.14), información sobre la presencia de patrones o clústeres y de valores atípicos (outliers). Pudíendose identificar zonas censales con presencia alta de robos rodeadas de zonas censales también con presencia alta de robos denominadas conglomerados calientes o (*hot spots*), que corresponden al primer cuadrante del scatterplot de Moran (High - High); o zonas censales con presencia baja de robos rodeadas de zonas censales también con presencia baja de robos conglomerados fríos (*cold spots*), tercer cuadrante del scatterplot de Moran (Low - Low); o bien zonas censales con presencia baja de robos rodeadas por zonas censales con presencia alta de robos, segundo cuadrante del scatterplot de Moran (Low - High) o viceversa (High - Low) que corresponderá al cuarto cuadrante. En todos estos casos considerados, estaremos refiriéndonos a zonas clústeres. Por último se puede también detectar zonas censales sin autocorrelación significativa a los que se les considerará como outliers.

En general el índice de autocorrelación espacial  $I$  global de Moran, presenta un valor promedio o resumen de la intensidad de la autocorrelación espacial entre las unidades geográficas consideradas, pero carece de la indicación de cuales unidades geográficas forman parte de un determinado cluster y donde se localizan, es decir no reporta el patrón de esta relación. En cambio el  $I_i$  local de Moran que se deriva del índice global de Moran, tiene la particularidad de hacer posible la identificación de patrones espaciales entre unidades geográficas consideradas.

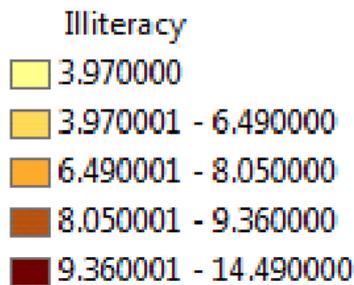
En todos estos casos el valor-p, reportado debe ser lo suficientemente pequeño para que el cluster o el valor atípico se considere estadísticamente significativo. Es así que los valores-p, obtenidos en los mapas de significancia que se presentan en las figuras (3.14, y 3.18), confirman este hecho.

### 3.6.1. Ejemplo práctico para el cálculo del Índice Local de Autocorrelación Espacial (LISA)

A continuación se presenta un ejemplo práctico tomado de la página de internet “<http://www.utdallas.edu/briggs/henan/11SAlocal.ppt>”, muy didáctico que se adapta a nuestras consideraciones y permite visualizar claramente el procedimiento que se sigue hasta la obtención de los índices de Moran tanto local como global, y que el programa GeoDa lo realiza automáticamente.

Contiguity Matrix		1	2	3	4	5	6	7	Sum	Neighbors	Illiteracy
Code		Anhui	Zhejiang	Jiangxi	Jiangsu	Henan	Hubei	Shanghai			
Anhui	1	0	1	1	1	1	1	0	5	6 5 4 3 2	14.49
Zhejiang	2	1	0	1	1	0	0	1	4	7 4 3 1	9.36
Jiangxi	3	1	1	0	0	0	1	0	3	6 2 1	6.49
Jiangsu	4	1	1	0	0	0	0	1	3	7 2 1	8.05
Henan	5	1	0	0	0	0	1	0	2	6 1	7.36
Hubei	6	1	0	1	0	1	0	0	3	1 3 5	7.69
Shanghai	7	0	1	0	1	0	0	0	2	2 4	3.97

Each row in the contiguity matrix describes the neighborhood for that location.



<http://www.utdallas.edu/briggs/henan/11SAlocal.ppt>

Figura 3.10: Ejemplo de Matriz de contigüidad

## MATRIZ DE VECINDAD Y MATRIZ DE PESOS ESPACIALES ESTANDARIZADOS POR FILA

Contiguity Matrix		1	2	3	4	5	6	7	
Code		Anhui	Zhejiang	Jiangxi	Jiangsu	Henan	Hubei	Shanghai	Sum
Anhui	1	0	1	1	1	1	1	0	5
Zhejiang	2	1	0	1	1	0	0	1	4
Jiangxi	3	1	1	0	0	0	1	0	3
Jiangsu	4	1	1	0	0	0	0	1	3
Henan	5	1	0	0	0	0	1	0	2
Hubei	6	1	0	1	0	1	0	0	3
Shanghai	7	0	1	0	1	0	0	0	2

Row Standardized Spatial Weights Matrix		Anhui	Zhejiang	Jiangxi	Jiangsu	Henan	Hubei	Shanghai	Sum
Code		Anhui	Zhejiang	Jiangxi	Jiangsu	Henan	Hubei	Shanghai	Sum
Anhui	1	0.00	0.20	0.20	0.20	0.20	0.20	0.00	1
Zhejiang	2	0.25	0.00	0.25	0.25	0.00	0.00	0.25	1
Jiangxi	3	0.33	0.33	0.00	0.00	0.00	0.33	0.00	1
Jiangsu	4	0.33	0.33	0.00	0.00	0.00	0.00	0.33	1
Henan	5	0.50	0.00	0.00	0.00	0.00	0.50	0.00	1
Hubei	6	0.33	0.00	0.33	0.00	0.33	0.00	0.00	1
Shanghai	7	0.00	0.50	0.00	0.50	0.00	0.00	0.00	1

Diagram illustrating the row standardization process: A circle around the value 3 in the Jiangsu row of the Contiguity Matrix (row 4, column 4) has an arrow pointing to a circle containing 1/3. Another arrow points from this 1/3 circle to the value 0.33 in the Jiangsu row of the Row Standardized Spatial Weights Matrix (row 4, column 4). A '10' is written at the bottom right of the diagram.

<http://www.utdallas.edu/briggs/henan/11SAlocal.ppt>

Figura 3.11: Ejemplo de Matriz estandarizada por filas

### ESTANDARIZADO (CALCULO DE $Z_i$ )

Deviations from Mean and z scores.				
	X	X-Xmean	$(X-\text{Mean})^2$	$z \leftarrow Z_i = \frac{x_i - \bar{x}}{SD_x}$
Anhui	14.49	6.29	39.55	2.101
Zhejiang	9.36	1.16	1.34	0.387
Jiangxi	6.49	(-1.71)	2.93	(0.572)
Jiangsu	8.05	(0.15)	0.02	(0.051)
Henan	7.36	(0.84)	0.71	(0.281)
Hubei	7.69	(0.51)	0.26	(0.171)
Shanghai	3.97	(-4.23)	17.90	(-1.414)

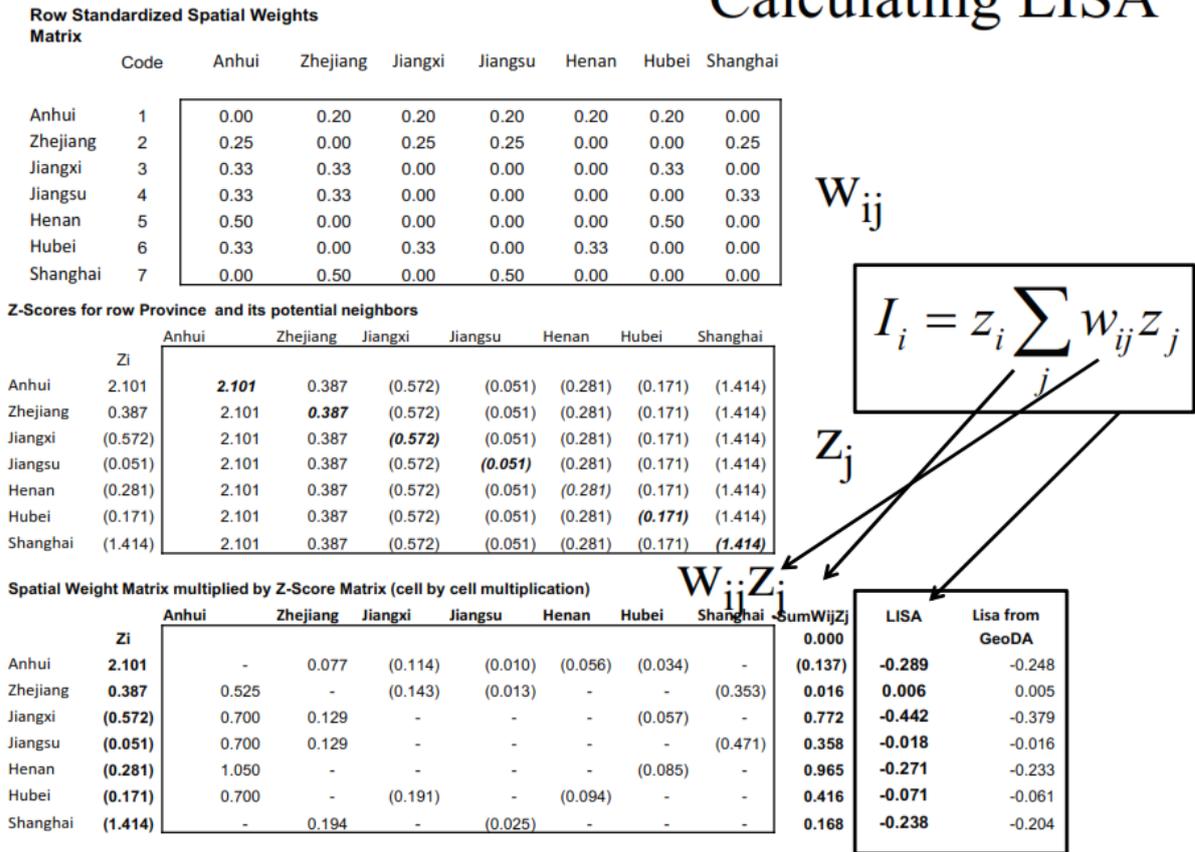
  

Mean and Standard Deviation			
Sum	57.41	0.00	62.71
Mean	57.41 / 7 =		8.20
Variance	62.71 / 7 =		8.96
SD	$\sqrt{8.96}$	=	2.99

<http://www.utdallas.edu/briggs/henan/11SAlocal.ppt>

Figura 3.12: Ejemplo de cálculo de  $Z_i$

# Calculating LISA



<http://www.utdallas.edu/briggs/henan/11SAlocal.ppt>

Figura 3.13: Ejemplo de cálculo del Índice Local de Autocorrelación Espacial (LISA)

### 3.7. Análisis del Índice de Moran

El índice global  $I$  de Moran calculado para nuestro estudio corresponde al valor de  $(0,647753)$ , que a su vez representa la pendiente de la regresión lineal de la figura (3.14), y de acuerdo con este valor y la recta obtenida, se tendrá una situación de autocorrelación positiva, nos permitiría concluir entonces, que la tasa de robos en los sectores urbanos del Distrito Metropolitano de Quito, muestran patrones de dependencia espacial o autocorrelación. Según Messner et al., estos patrones se generan debido a la trasmisión contagiosa, de episodios de violencia (delitos) que se esparsen en un determinado sector o en sus alrededores. Acorde con este principio, se puede apreciar en el mapa de significancia respectivo, la formación de clústeres de zonas censales con alta presencia de tasa de robos rodeadas por otras zonas censales con también alta tasa de robos, lo cual de paso contradice la hipótesis nula, que para este análisis se plantea inicialmente, en la que se asume que los datos están distribuidos aleatoriamente y no presentan ningún tipo de patrón identificable, es decir se tiene presente la formación de patrones geográficos, "*NOALEATORIOS*".

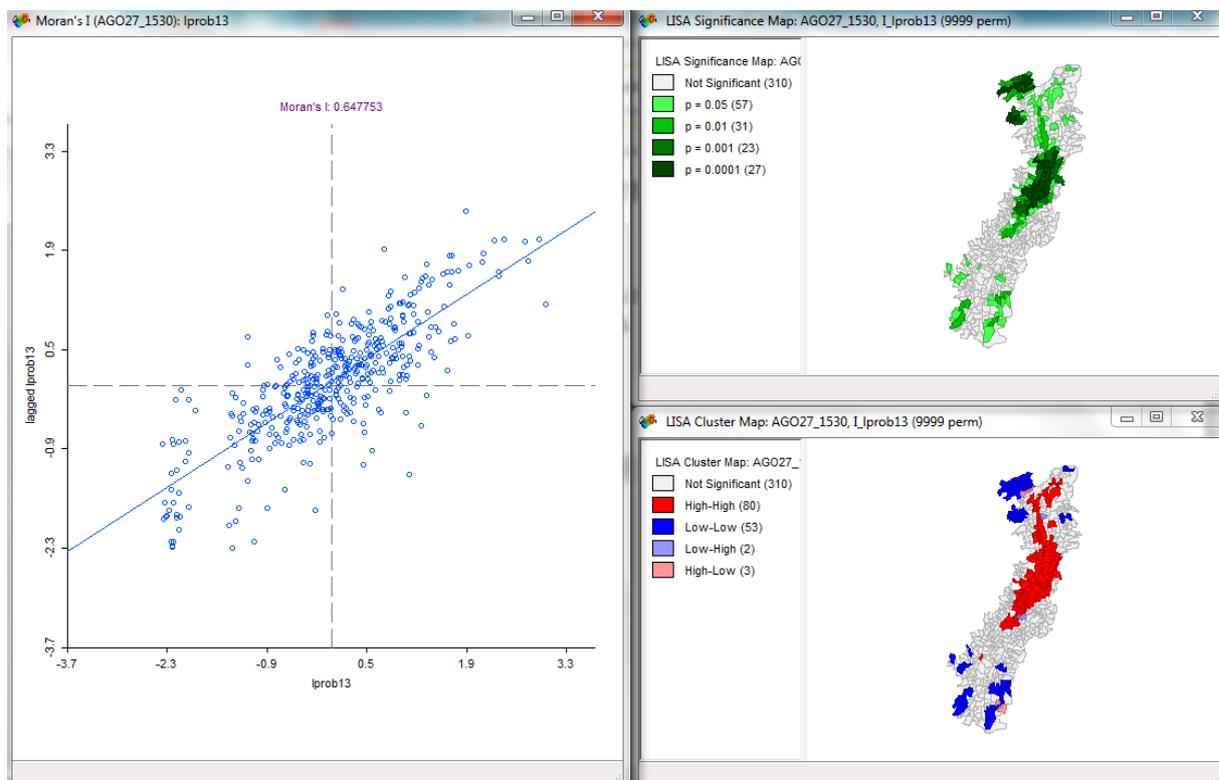


Figura 3.14: ÍNDICE DE MORAN GLOBAL UNIVARIADO.- ROBOS - REZAGO ESPACIAL DE ROBOS CON (9999) PERMUTACIONES

### 3.8. Análisis de Autocorrelación Espacial para tasas

Al calcular del modo anterior, los índices de autocorrelación global y local, se puede violar el principio de estacionalidad del proceso espacial, por existir inestabilidad en la varianza. Esta inestabilidad puede deberse a las diferencias entre los valores de la población total de cada zona censal. Para intentar corregir esto se pueden emplear técnicas de suavizado. GeoDa permite implementar un procedimiento de ajuste por medio de la transformación de variables basada en el principio de estandarizado empírico de Bayes (Assuncao y Reis, 1999). Se hace notar que la tasa estandarizada de Empirical Bayes (EB) en este análisis no es lo mismo que el suavizado empírico de Bayes.

En los gráficos siguientes se ilustra el análisis de autocorrelación espacial con ajuste de la inestabilidad de la varianza para tasas (rates), Assuncao and Reis (1999).

A continuación se presenta el análisis realizado para datos dados en forma de tasas, que consta fundamentalmente de gráficos como el scatter plot de Moran I estandarizado, en el que, en lugar de analizar las variables independientemente; se las analiza como "*tasas*" que pueden estar formadas por los valores de una variable asociada a algún factor de transformación o estar formada por la razón entre dos variables que se deseen considerar en el eje  $X$  y la matriz de pesos espaciales asociada a esta tasa en el eje  $Y$ . GeoDa dispone dentro de sus aplicaciones para suavizado el ajuste Empírico de Bayes tanto local como global, ajuste que para este ejemplo propuesto, toma en cuenta la inestabilidad de la varianza de la tasa estandarizada de  $(lprob13 / lpemig)$ , formada en este caso por la razón entre los valores de la variable logaritmo de la tasa de robos en Quito durante el año 2013 ( $lprob13$ ) y el logaritmo de la tasa de hogares que tienen familiares que han emigrado ( $lpemig$ ).

En la figura (3.15), se tiene el scatter plot del índice de Moran estandarizado correspondiente al ajuste empírico de Bayes para la tasa  $(lprob13 / lpemig)$  y  $W(lprob13 / lpemig)$ ; en la figura (3.16) se presenta además del Moran I estandarizado el mapa de clústeres y el mapa de significancia para las mismas variables anteriores. Finalmente en la figura (3.17) se tiene los resultados cuando se realizan 9999 permutaciones.

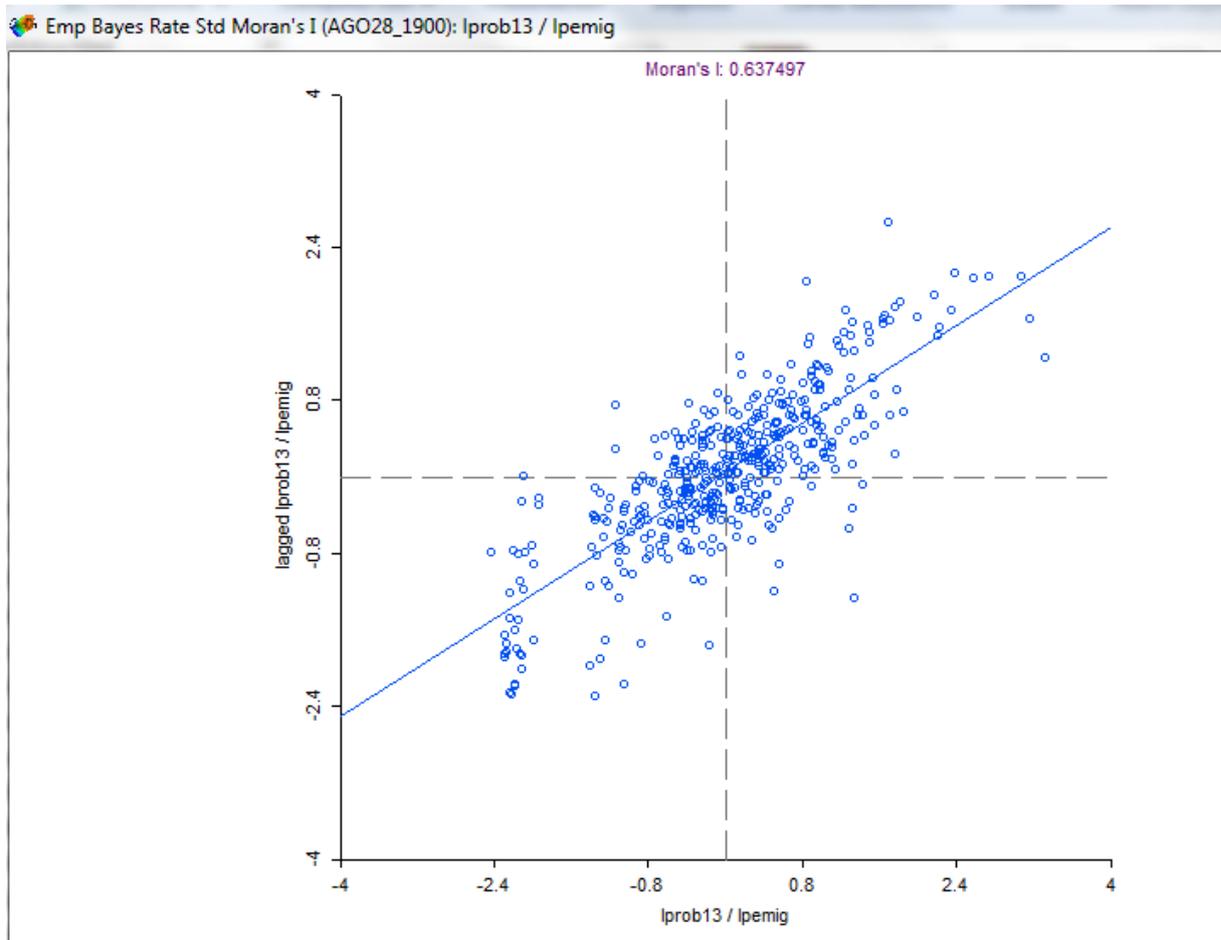


Figura 3.15: MORAN I CON AJUSTE (EB) PARA TASAS

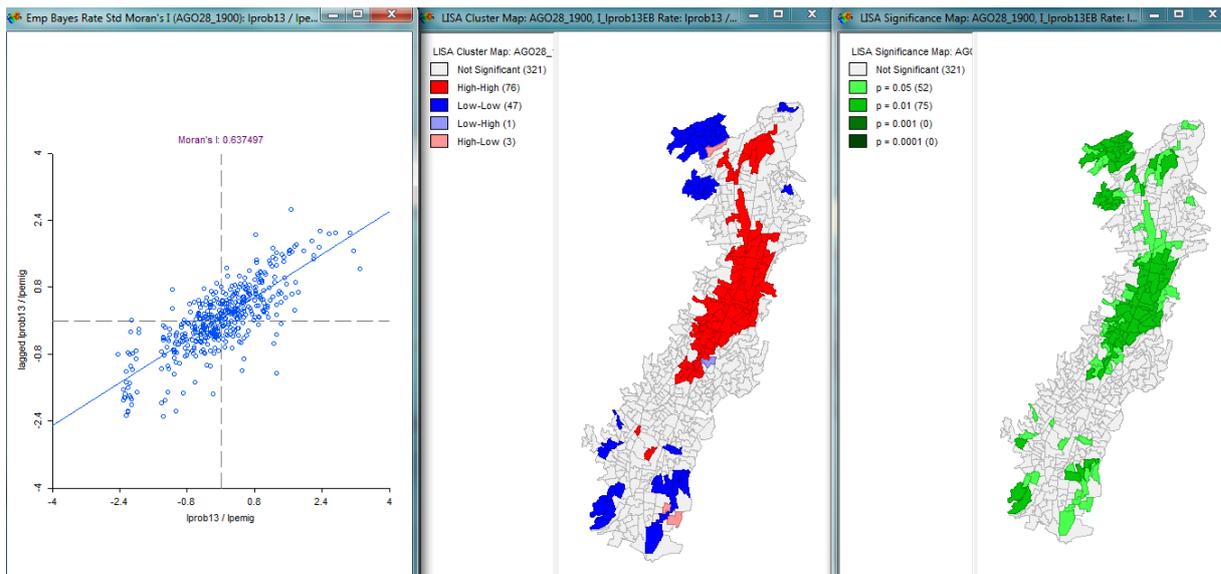


Figura 3.16: USO DEL AJUSTE EMPÍRICO DE BAYES

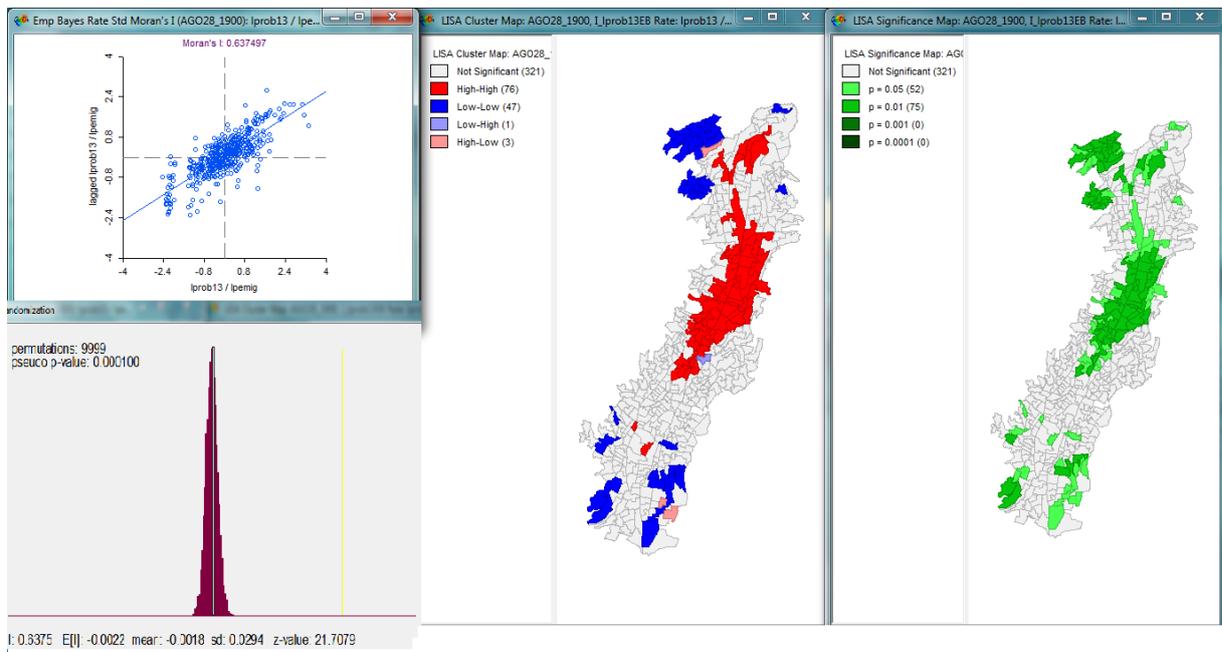


Figura 3.17: EMPLEO DEL AJUSTE EMPÍRICO DE BAYES CON 9999 PERMUTACIONES

Como se puede establecer de la revisión de las dos formas propuestas para el manejo de los datos y que se han presentado, los valores para el índice de Moran obtenidos son relativamente altos y contrastando además con la información presente en los mapas de clústeres y de significancia donde efectivamente se puede visualizar la formación de agrupamientos perfectamente identificables lo cual nos permite concluir definitivamente que existe la presencia de autocorrelación espacial entre las variables utilizadas en este estudio, y rechazar la hipótesis nula que trata de sostener la idea de que los datos empleados para este trabajo obedecen a una distribución aleatoria.

La inferencia del índice de Moran está basada en las permutaciones que se pueden realizar con los valores de la cantidad de robos (dato conocido), que como una manera de explicación práctica, podemos mencionar, que si se pudiera de alguna manera extraer los valores correspondientes a los robos en cada zona censal y los lanzáramos aleatoriamente sobre las zonas censales, estos valores que permanecen constantes, caerían en diferentes zonas censales. Si este proceso se lo puede realizar un gran número de veces, se lograría finalmente obtener una distribución normal de los datos, acorde al teorema de límite central. Por tanto el índice de Moran obtenido inicialmente es contrastado con los diferentes valores que se obtendrían al realizar varias permutaciones (9999). Figura (3.18).

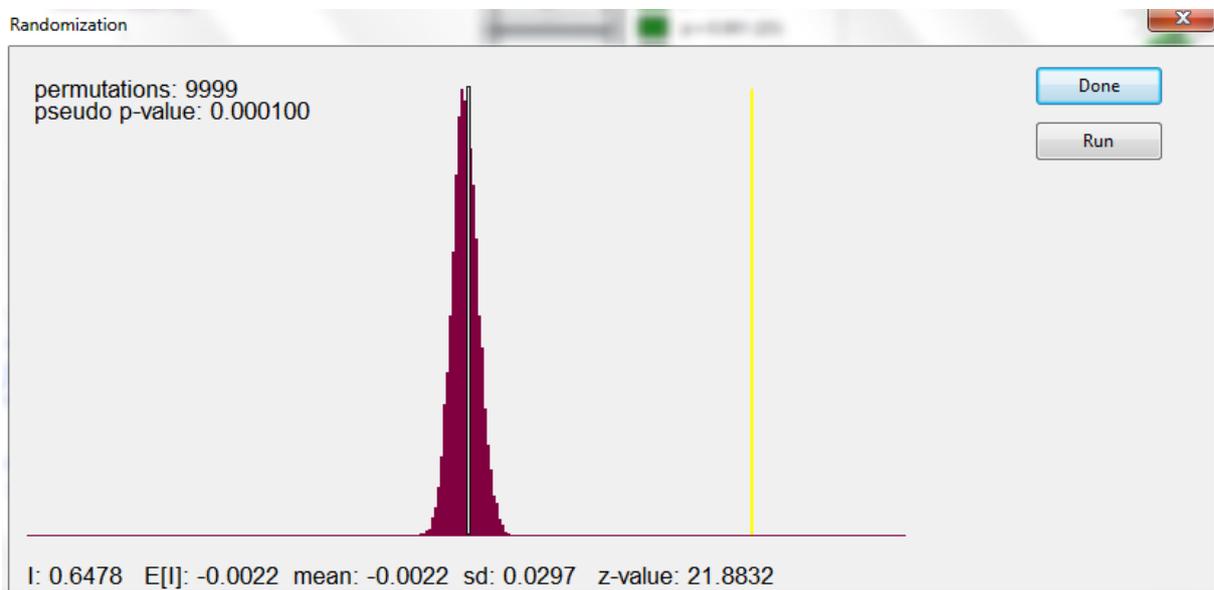


Figura 3.18: Índice de Moran con 9999 permutaciones

En el scatterplot de Moran I, de los robos y sus rezagos espaciales, figura (3.19) , en el cuadrante I, se tiene autocorrelación positiva (High - High, color rojo), que representa zonas censales con alta tasa de robos rodeadas por zonas censales cuyo promedio de tasa de robos también es alta (*hot Spots*).

En el cuadrante III, en cambio se tiene autocorrelación positiva (Low - Low, color azul), que representa zonas censales con baja tasa de robos rodeadas por zonas censales cuyo promedio de tasa de robos también es baja (*cold spots*).

En este mismo gráfico se muestra la funcionalidad de la aplicación denominada *linking and brushing*, para ello se seleccionó únicamente en el gráfico de dispersión, aquellas zonas censales que aparecen ubicadas en el I cuadrante del scatterplot de Moran (puntos en color rojo), estas mismas zonas censales se ven automáticamente resaltadas en color amarillo, tanto en los mapas de significancia y de clústeres, como en la tabla de datos. Adicionalmente en el scatterplot aparecen dos nuevas líneas de regresión, una de color rojo correspondiente a las zonas censales seleccionadas y otra de color lila correspondiente a las zonas sobrantes, manteniéndose inalterable la recta de color azul correspondiente al índice de Moran previamente calculado.

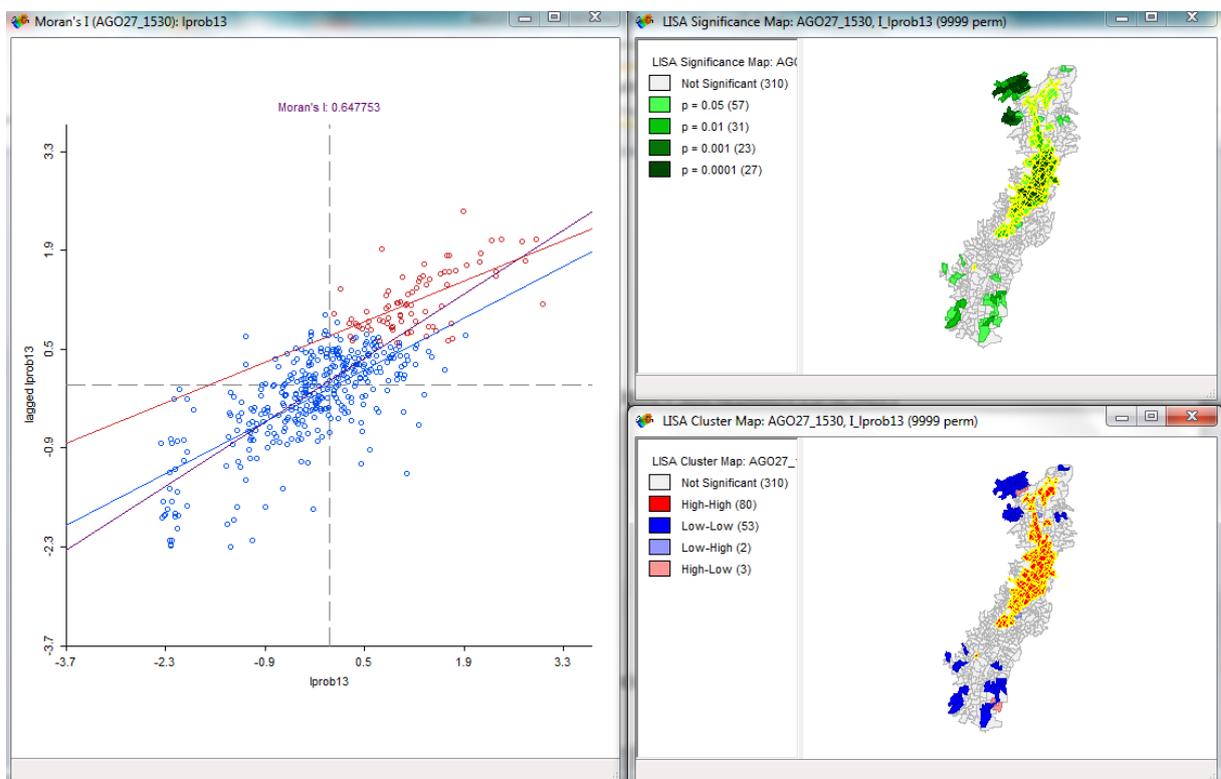


Figura 3.19: Gráfico de dispersión del Índice de Morán con mapas de significancia y de clústeres

### 3.9. Índice de Moran Bivariado

Los índices univariados de Moran  $I$  establecen la correlación entre una variable de estudio  $v$  y el rezago de esta misma variable  $v$  pero siendo esta última de las zonas vecinas. Para obtener el índice de Moran bivariado, que permite establecer una correlación entre el rezago de la variable  $v$  en el eje  $Y$  y una variable diferente en las zonas vecinas en el eje  $X$ , como ejemplo mostramos el scatterplot y los mapas de significancia y de clústeres, para la tasa de robos (rezagada) en el eje  $Y$ , vs. la tasa de hogares con extrema pobreza en el eje  $X$ . Obteniéndose para este caso un índice de Moran bivariado de  $-0,486952$ , cuyo signo nos permite establecer que existe una autocorrelación negativa, es decir se tendrá una zona censal con una tasa baja de hogares con extrema pobreza, rodeada de zonas censales cuyo promedio de tasa de robos es alta (rezago).

#### ÍNDICE DE MORAN LOCAL BIVARIADO ENTRE: HOGARES CON POBREZA EXTREMA Y ROBOS

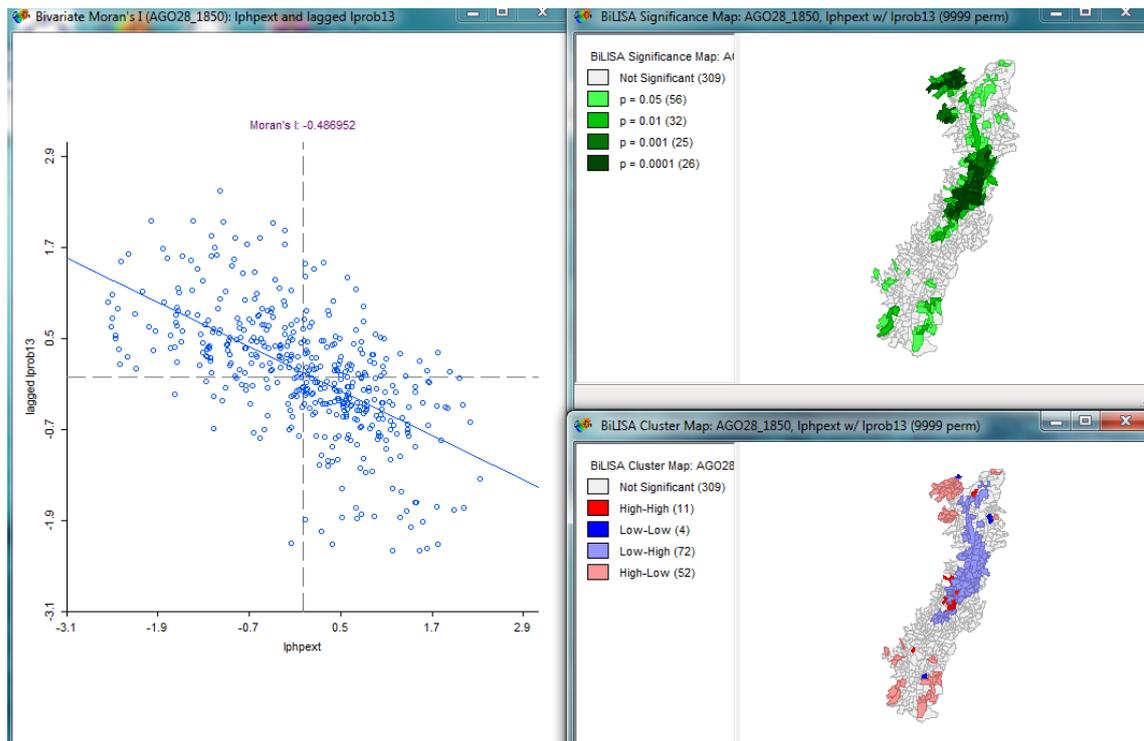


Figura 3.20: Gráfico de dispersión del Índice de Moran Bivariado con mapas de significancia y de clústeres

## ÍNDICE DE MORAN LOCAL BIVARIADO ENTRE: HOGARES CON PERSONAS QUE HAN EMIGRADO Y ROBOS

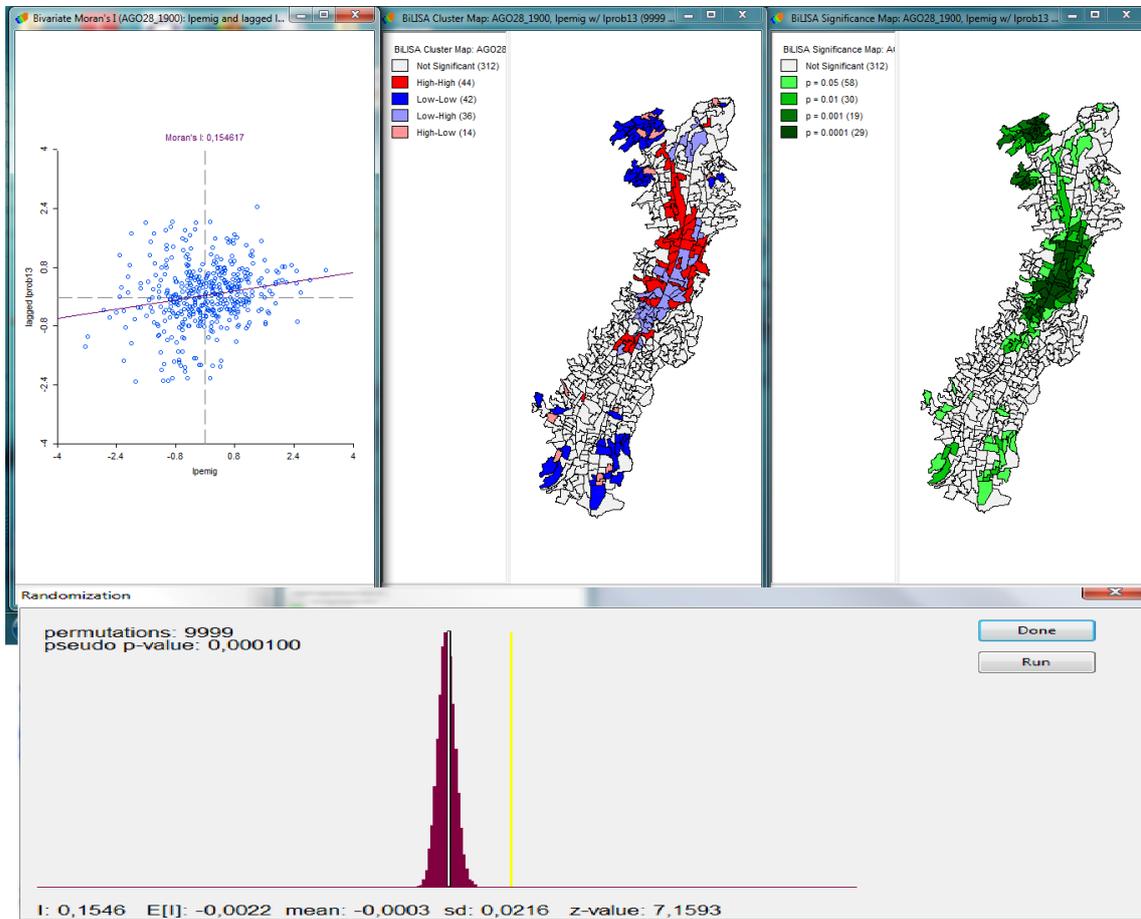


Figura 3.21: Índice de Moran Bivariado para hogares con personas que han migrado y robos con 9999 permutaciones

### 3.10. Regresión Espacial

Para establecer la presencia de dependencia espacial entre las variables de interés tasa de robos, sus rezagos espaciales y las variables consideradas como exógenas, se presenta a continuación una parte del análisis exploratorio espacial realizado entre estas variables para verificar su relevancia o no. Para ello se han ejecutado las aplicaciones de tipo espacial disponible en GeoDa, considerado entonces como posibles variables exógenas aquellas que luego del análisis exploratorio se estiman como más representativas ver figuras: (3.22), (3.23), (3.24) y (3.25). Sin embargo, esto también es posible hacerlo directamente generando un modelo de regresión espacial, que permita verificar si existe dependencia espacial, como consecuencia de la presencia de variables sistemáticas (endógenas y/o exógenas) correlacionadas espacialmente, o como consecuencia de la presencia de dependencia espacial en las perturbaciones. (Moreno y Vayá, 2004).

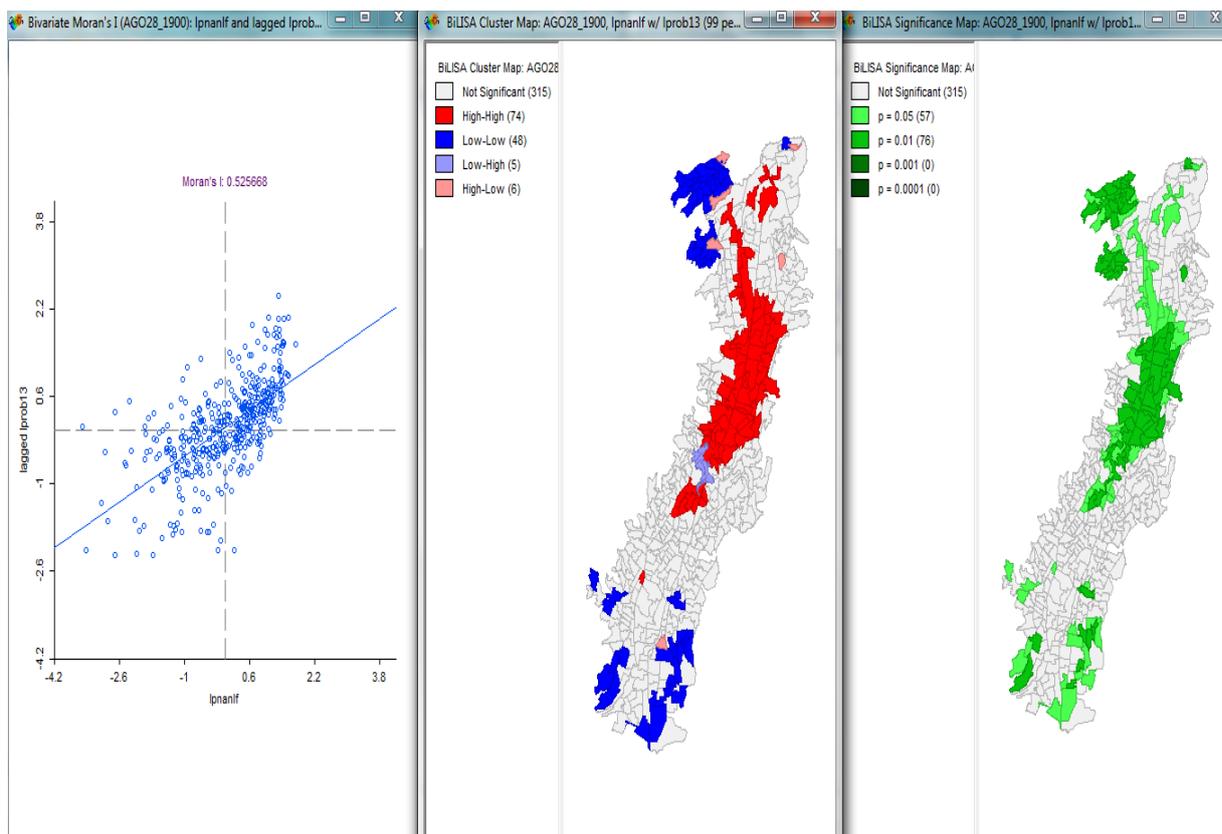


Figura 3.22: Índice de Moran Bivariado entre *lprob13* y *lpanlf*

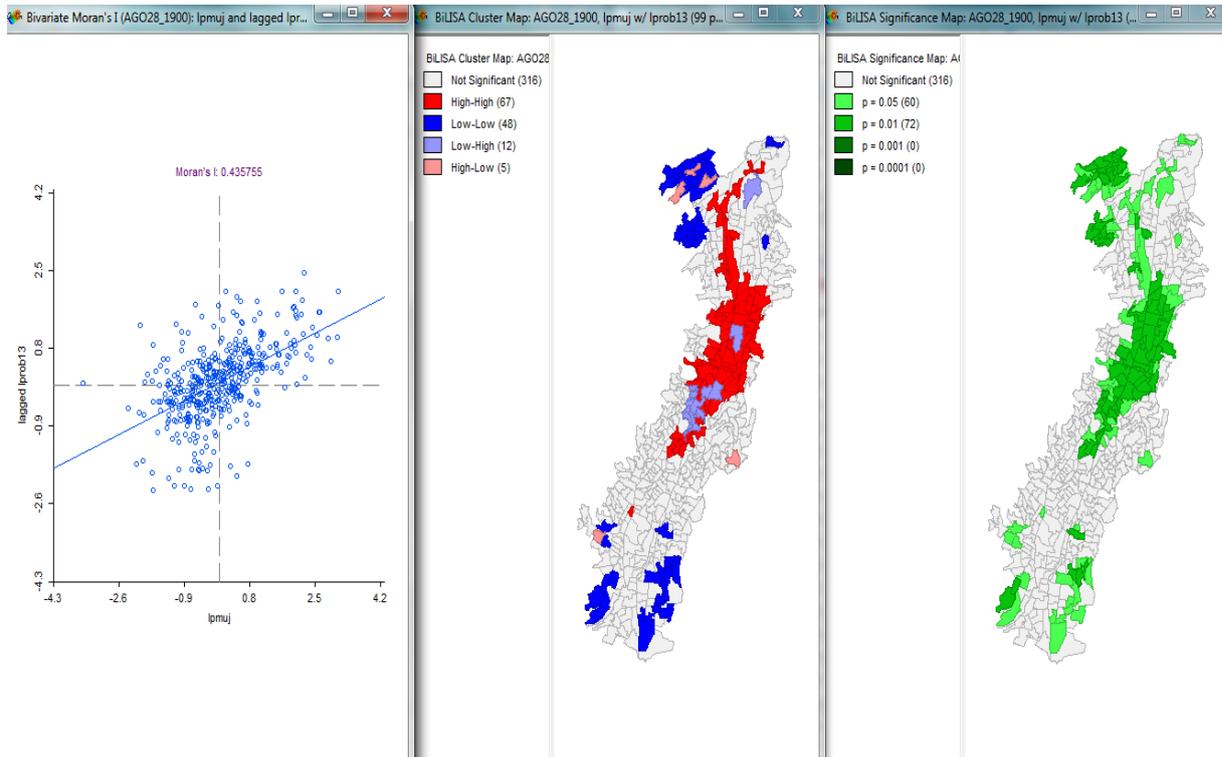


Figura 3.23: Índice de Moran Bivariado entre  $lprob13$  y  $lpmuj$

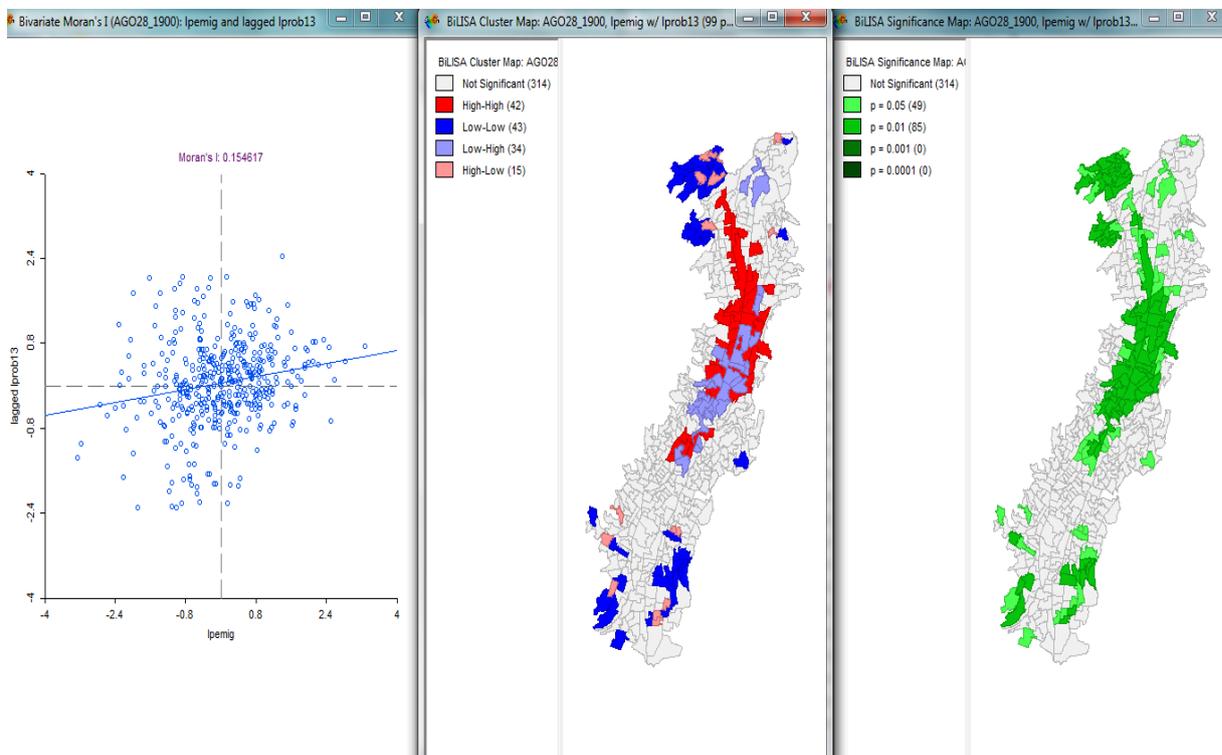


Figura 3.24: Índice de Moran Bivariado entre  $lprob13$  y  $lpmig$

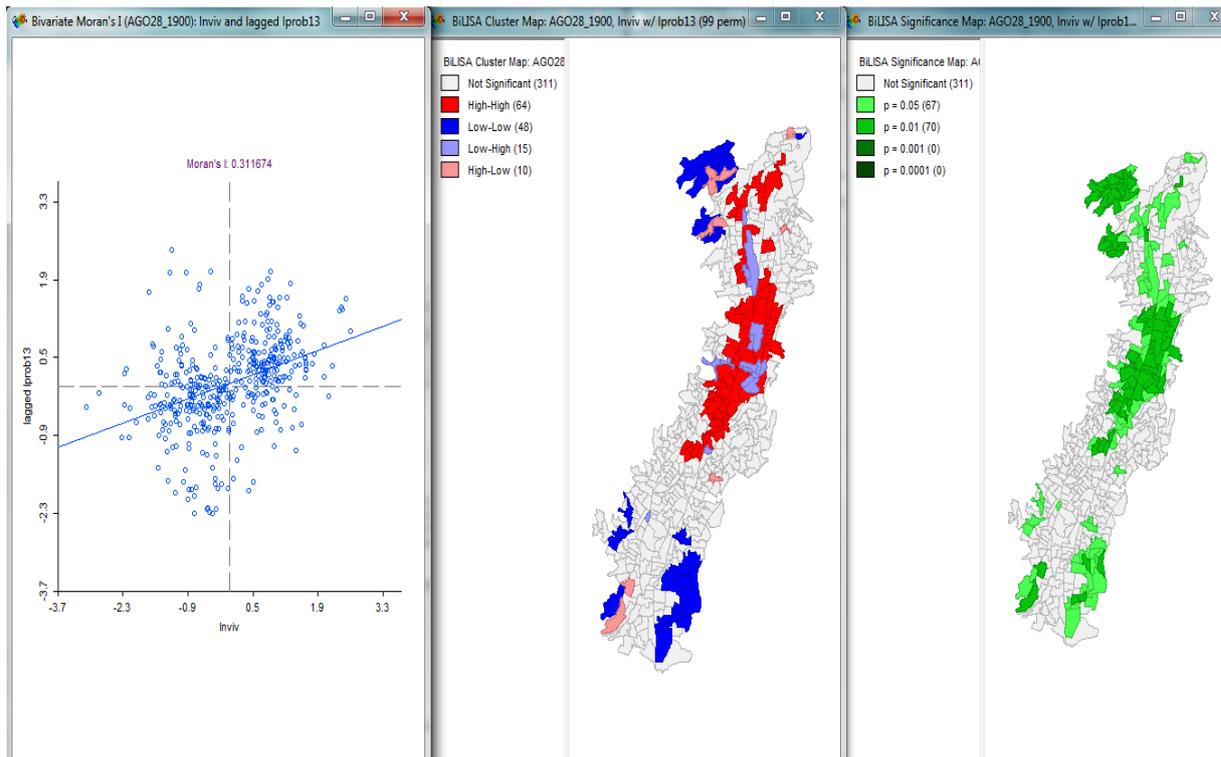


Figura 3.25: Índice de Moran Bivariado entre lprob13 y Inviv

Previo a obtener regresiones de tipo espacial se realiza una regresión de mínimos cuadrados ordinarios en la que se incorporan las variables consideradas como explicativas y la variable de interés, robos durante el año 2013 para la parte Urbana del Distrito Metropolitano de Quito, ver cuadro (3.2) se realiza ésto, con la finalidad de obtener su diagnóstico que generalmente presenta un resumen de características del modelo: los coeficientes estimados asociados, error estándar, estadístico  $t$  y su probabilidad, las medidas de ajuste  $R^2$  y  $R^2$  ajustado, la suma de los residuos al cuadrado, el estadístico  $F$  sobre la hipótesis nula que todos los coeficientes de la regresión son en conjunto cero con su respectiva probabilidad, la log-verosimilitud, el criterio de información de Akaike, el criterio de Schwarz, estas tres medidas mencionadas recientemente se basan en el supuesto de normalidad multivariada y la correspondiente función de verosimilitud para el modelo de regresión estándar. Además se presenta parte de la matriz de covarianza para los coeficientes estimados de la regresión, así como los valores pronosticados y los residuos para cada observación. Los nombres de las variables consideradas se encuentran en la parte superior de las columnas de la matriz de varianzas (esta matriz es simétrica) y finalmente las condiciones de multicolinealidad y las pruebas estadísticas para normalidad y heteroscedasticidad.

## Regresión por Mínimos Cuadrados Ordinarios

### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

```

Data set           :  datos_panel_robos
Dependent Variable :      lprob13  Number of Observations:  448
Mean dependent var :      1.3911  Number of Variables   :    8
S.D. dependent var :      1.46128  Degrees of Freedom    :  440
R-squared          :      0.422935  F-statistic           :      46.0684
Adjusted R-squared :      0.413754  Prob(F-statistic)    :          0
Sum squared residual:      552.04  Log likelihood       :     -682.462
Sigma-square       :      1.25464  Akaike info criterion :     1380.92
S.E. of regression :      1.12011  Schwarz criterion    :     1413.76
Sigma-square ML    :      1.23223
S.E of regression ML:      1.11006
  
```

---

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-298.88	38.08652	-7.847395	0.00000
larea	0.7395264	0.116295	6.359058	0.00000
lnviv	-0.2826789	0.2256379	-1.252799	0.21094
lpemig	-0.518562	0.2132663	-2.431523	0.01543
lphcin	0.3326592	0.1345923	2.471605	0.01383
lpmuj	13.89597	3.121491	4.45171	0.00001
lpnalcan	-0.2659568	0.06219326	-4.276296	0.00002
lpnanlf	30.208	4.948788	6.104121	0.00000

---

### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 2592.942596

#### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	6.7054	0.03499

#### DIAGNOSTICS FOR HETEROSKEDASTICITY

##### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	7	37.5898	0.00000
Koenker-Bassett test	7	29.0365	0.00014

#### SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	35	64.6738	0.00166

#### DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : datos\_panel\_robos.gal

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.4216	14.6895	0.00000
Lagrange Multiplier (lag)	1	230.1572	0.00000
Robust LM (lag)	1	32.7403	0.00000
Lagrange Multiplier (error)	1	201.4944	0.00000
Robust LM (error)	1	4.0776	0.04346
Lagrange Multiplier (SARMA)	2	234.2348	0.00000

```

COEFFICIENTS VARIANCE MATRIX
  CONSTANT      larea      lnviv      lpemig      lphcin
1450.583185    -0.623935     1.909175     0.594563    -3.656635
-0.623935      0.013525     -0.008520     0.002756     0.005648
  1.909175     -0.008520     0.050912     0.001705    -0.003529
  0.594563      0.002756     0.001705     0.045483     0.001572
-3.656635      0.005648     -0.003529     0.001572     0.018115
-53.830525     0.037480     -0.046519    -0.083303     0.195861
-0.690877     -0.003658     0.002551     0.000528    -0.002012
-161.109455    0.036692     -0.272144    -0.044782     0.336704
  lpmuj      lpnalcan      lpnanlf
-53.830525    -0.690877    -161.109455
  0.037480    -0.003658     0.036692
-0.046519     0.002551    -0.272144
-0.083303     0.000528    -0.044782
  0.195861    -0.002012     0.336704
  9.743704    -0.002318    -1.114066
-0.002318     0.003868     0.106537
-1.114066     0.106537     24.490503
OBS          lprob13          PREDICTED          RESIDUAL
  1          -0.63056          -0.07161          -0.55895
  2          -1.93442           0.13434          -2.06875
  3          -1.85348          -0.16913          -1.68435
  4           1.64990           1.42982           0.22008
  5           1.53824           2.27111          -0.73286
  .
  .
440          0.16158          -0.62761           0.78919
441          0.85530           1.12339          -0.26810
442          1.86291           2.94801          -1.08510
443          4.74039           3.13541           1.60499
444          0.74655           0.76969          -0.02314
445          2.56410           2.46710           0.09700
446          3.84702           2.53736           1.30966
447          2.44305           2.18083           0.26222
448          1.01591           1.78311          -0.76721
===== END OF REPORT =====

```

Cuadro 3.2: Resultados de la estimación por MCO

Siguiendo la metodología sugerida en (Moreno y Vayá, 2004), se incorporan la tasa de robos como variable endógena (dependiente), y como variables exógenas esta misma variable pero rezagada junto a las variables anteriormente consideradas en el modelo MCO, para obtener las primeras estimaciones de modelos con rezagos espaciales y/o con errores espaciales, que a su vez se logran por la aplicación de métodos de Máxima Verosimilitud; cuyas técnicas implementadas en GeoDa están basadas en los algoritmos de Smirnov y Anselin (2001). Siendo necesario hacer notar que para el caso espacial solamente funciona cuando toma en cuenta la matriz de pesos espaciales

que previamente debe ser generada con pesos que reflejen un arreglo espacial simétrico, tales como vecindades de contiguidad tipo: “*Rook*” o “*Queen*”, pero no para vecindades del tipo “*K – nearest*”. Los resultados para la primera regresión espacial, se pueden observar en el cuadro (3.3)

Cuando la variable endógena del modelo de regresión espacial esté correlacionada espacialmente, se sugiere emplear el siguiente modelo (Moreno y Vayá, 2004).

$$y = \rho W y + X \beta + u \quad (3.6)$$

donde:

$$u \sim N(0, \sigma^2 I)$$

$y$  vector  $n \times 1$

$\rho$  parámetro autorregresivo que recoge la intensidad de las interdependencias entre las observaciones muestrales.

$W_y$  rezago espacial de la variable  $y$

$X$  matriz de  $k$  variables exógenas

$\beta$  vector de  $k \times 1$  parámetros

$u$  término de perturbaciones, ruido blanco

$n$  número de observaciones

De forma similar, si la autocorrelación espacial está únicamente presente en el término de las perturbaciones se tendrá: (Moreno y Vayá, 2004).

$$y = X \beta + \varepsilon \quad (3.7)$$

donde:

$$\varepsilon = \lambda W \varepsilon + u$$

$$u \sim N(0, \sigma^2 I)$$

$u$  término de ruido blanco

$\lambda$  parámetro autorregresivo que recoge la intensidad de las interdependencias.

### Regresión espacial inicial

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

```

Data set           : datos_panel_robos
Spatial Weight     : datos_panel_robos.gal
Dependent Variable : lprob13   Number of Observations: 448
Mean dependent var : 1.3911   Number of Variables   : 9
S.D. dependent var : 1.46128  Degrees of Freedom    : 439
Lag coeff. (Rho)  : 0.659358

R-squared          : 0.664056  Log likelihood         : -584.982
Sq. Correlation    : -          Akaike info criterion  : 1187.96
Sigma-square       : 0.717356  Schwarz criterion     : 1224.91
S.E of regression  : 0.846969
  
```

Variable	Coefficient	Std.Error	z-value	Probability
W_lprob13	0.6593575	0.03869924	17.038	0.00000
CONSTANT	-154.8185	30.04471	-5.152936	0.00000
larea	0.5586771	0.08962683	6.233369	0.00000
lnviv	-0.4514215	0.1706425	-2.645423	0.00816
lpemig	-0.4031252	0.1620855	-2.487114	0.01288
lphcin	0.2667066	0.1018008	2.619887	0.00880
lpmuj	7.95624	2.379821	3.343209	0.00083
lpnalcan	-0.1637934	0.04704947	-3.481301	0.00050
lpnanlf	14.93074	3.863307	3.864756	0.00011

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	7	37.7816	0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : datos\_panel\_robos.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	194.9591	0.00000

```

COEFFICIENTS VARIANCE MATRIX
  CONSTANT      larea      lnviv      lpemig      lphcin
902.684453    -0.505061      1.117195      0.479696     -2.070030
 -0.505061      0.008033     -0.004923      0.001293      0.003188
  1.117195     -0.004923      0.029119      0.001024     -0.002011
  0.479696      0.001293      0.001024      0.026272      0.000938
 -2.070030      0.003188     -0.002011      0.000938      0.010363
-33.381446      0.026698     -0.027507     -0.052593      0.111251
 -0.382706     -0.002116      0.001463      0.000325     -0.001147
-100.338657      0.037618     -0.158474     -0.041281      0.190193
  0.331310     -0.000670      0.000116      0.000632      0.000094
  lpmuj      lpnalcan      lpnanlf      W_lprob13
-33.381446     -0.382706    -100.338657      0.331310
  0.026698     -0.002116      0.037618     -0.000670
 -0.027507      0.001463     -0.158474      0.000116
 -0.052593      0.000325     -0.041281      0.000632
  0.111251     -0.001147      0.190193      0.000094
  5.663550     -0.001763     -0.344966     -0.011767
 -0.001763      0.002214      0.059533      0.000056
 -0.344966      0.059533     14.925143     -0.037167
 -0.011767      0.000056     -0.037167      0.001498
OBS      lprob13      PREDICTED      RESIDUAL      PRED ERROR
  1      -0.63056      -0.44020      0.78613      -0.19036
  2      -1.9344      -0.13782     -1.11833     -1.79660
  3      -1.8535      0.26591      -1.69465     -2.11939
  4       1.6499      1.37769      0.60559      0.27221
  5       1.5382      1.95097     -0.10845     -0.41273
  .
  .
  .
441       0.8553      0.73649     -0.26681      0.11881
442       1.8629      2.78118     -0.76044     -0.91827
443       4.7404      3.80544      0.40644      0.93495
444       0.74655      0.78649      0.12298     -0.03995
445       2.5641      2.54295     -0.01759      0.02115
446       3.847       2.56421      1.37380      1.28281
447       2.4431      2.16007      0.37902      0.28299
448       1.0159      1.99947     -0.75314     -0.98357
===== END OF REPORT =====

```

Cuadro 3.3: Resultados de la estimación del Modelo espacial considerando siete variables iniciales

A continuación se adicionan al modelo anterior los rezagos de todas las variables exógenas para verificar como cambia el modelo, ver cuadro (3.4), seguidamente se procede a eliminar variables que de acuerdo a los resultados reportados en cada regresión no presentan mayor significancia estadística, hasta lograr finalmente un modelo que se ajusta de mejor manera a este estudio cuadro (3.5).

### Regresión espacial considerando los rezagos de las variables iniciales

```

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set           : datos_panel_robos
Spatial Weight     : AGO28_1850.gal
Dependent Variable : lprob13   Number of Observations: 448
Mean dependent var : 1.3911   Number of Variables   : 16
S.D. dependent var : 1.46128  Degrees of Freedom    : 432
Lag coeff. (Rho)  : 0.644446
R-squared          : 0.683544  Log likelihood         : -570.297
Sq. Correlation    : -          Akaike info criterion  : 1172.59
Sigma-square       : 0.675743  Schwarz criterion     : 1238.27
S.E of regression  : 0.822036

Variable           Coefficient      Std.Error      z-value      Probability
W_lprob13          0.644446        0.04358603    14.78561     0.00000
CONSTANT           -89.41878       50.03978      -1.786954    0.07394
LA_NANLF           -12.84243       7.525772      -1.70646     0.08792
LA_NALCAN          -0.2876075     0.09126513    -3.15134     0.00163
LA_MUJ             1.958464        4.41485       0.4436081    0.65733
LA_HCIN            0.1822836       0.2105712     0.8656622    0.38668
LA_EMIG            0.03697046     0.2978935     0.1241063    0.90123
LA_NVIV            0.7302894       0.3412896     2.139794     0.03237
LA_AREA           -0.2784151     0.1688943     -1.648458    0.09926
larea              0.7010607       0.09723402    7.210036     0.00000
lnviv              -0.706415       0.1853523     -3.811202    0.00014
lpemig             -0.4401885     0.1737759     -2.533082    0.01131
lphcin             0.155239        0.1136744     1.365646     0.17205
lpmuj              8.490906        2.413321      3.51835      0.00043
lpnalcan           -0.09011765     0.05017771    -1.79597     0.07250
lpnanlf            15.77582        5.113857      3.084915     0.00204

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                DF        VALUE        PROB
Breusch-Pagan test  14        56.5213     0.00000
DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : AGO28_1850.gal
TEST                DF        VALUE        PROB
Likelihood Ratio Test  1        152.4586    0.00000
===== END OF REPORT =====

```

Cuadro 3.4: Resultados de la estimación del Modelo considerando variables iniciales con sus rezagos

### Regresión espacial con variables más significativas

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : SHAPE\_lag\_GEODAABR9\_1650  
 Spatial Weight : datos\_panel\_robos.gal  
 Dependent Variable : lprob13 Number of Observations: 448  
 Mean dependent var : 1.3911 Number of Variables : 6  
 S.D. dependent var : 1.46128 Degrees of Freedom : 442  
 Lag coeff. (Rho) : 0.615149

R-squared : 0.654668 Log likelihood : -587.465  
 Sq. Correlation : - Akaike info criterion : 1186.93  
 Sigma-square : 0.737403 Schwarz criterion : 1211.56  
 S.E of regression : 0.858722

Variable	Coefficient	Std.Error	z-value	Probability
W_lprob13	0.6151488	0.04175786	14.73133	0.00000
CONSTANT	-48.3278	12.09947	-3.99421	0.00006
LA_NALCAN	-0.3336017	0.0534234	-6.244487	0.00000
lpemig	-0.3917289	0.1626375	-2.408601	0.01601
lpmuj	7.389413	1.951388	3.786748	0.00015
larea	0.3898426	0.07375699	5.2855	0.00000

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	4	34.6940	0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : datos\_panel\_robos.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	153.8032	0.00000

## COEFFICIENTS VARIANCE MATRIX

CONSTANT	LA_NALCAN	lpemig	lpmuj	larea
146.397055	-0.178755	0.367503	-23.497825	-0.066214
-0.178755	0.002854	0.001030	0.028601	-0.000945
0.367503	0.001030	0.026451	-0.079029	0.001528
-23.497825	0.028601	-0.079029	3.807914	-0.000989
-0.066214	-0.000945	0.001528	-0.000989	0.005440
0.149698	0.001028	0.000383	-0.023557	-0.000694

W\_lprob13

0.149698  
0.001028  
0.000383  
-0.023557  
-0.000694  
0.001744

OBS	lprob13	PREDICTED	RESIDUAL	PRED ERROR
1	-0.63056	-0.38834	0.85350	-0.24222
2	-1.9344	0.12841	-1.34003	-2.06283
3	-1.8535	0.57663	-2.10597	-2.43012
4	1.6499	1.17628	0.74111	0.47362
5	1.5382	1.52879	0.27138	0.00945
6	2.3565	2.20729	0.19846	0.14919
7	0.092359	0.70107	0.38621	-0.60871
8	2.0545	0.94524	1.91245	1.10922
9	-1.605	0.96691	-2.07067	-2.57194
10	-2.0528	-0.15760	-0.97815	-1.89524
.				
.				
.				
200	1.7287	0.84075	0.10456	0.88792
201	0.6145	0.54906	-0.04157	0.06544
202	0.075648	0.76107	-0.60264	-0.68542
203	0.76157	0.92677	0.15576	-0.16520
204	-0.55809	0.80500	-1.61272	-1.36310
205	1.6947	0.79827	0.84086	0.89641
.				
.				
.				
440	0.16158	0.61013	-0.49717	-0.44856
441	0.8553	1.19449	-0.55844	-0.33920
442	1.8629	3.01171	-0.83055	-1.14880
443	4.7404	3.71798	0.45474	1.02242
444	0.74655	0.71675	0.22641	0.02980
445	2.5641	3.21466	-0.39967	-0.65056
446	3.847	2.91386	1.12044	0.93315
447	2.4431	2.20861	0.25646	0.23444
448	1.0159	2.23608	-0.98534	-1.22017

===== END OF REPORT =====

Cuadro 3.5: Resultados de la estimación del modelo con rezagos espaciales

Para modelos con rezagos espaciales, se debe hacer una distinción entre residuo y error pronosticado, este último es la diferencia entre el valor observado y el valor pronosticado que sólo utiliza variables exógenas, en lugar de tratar el retardo espacial  $Wy$ . El reporte de resultados consiste de los coeficientes estimados y sus errores asintóticos, pruebas  $t$ , medidas de ajuste log-verosimilitud, criterio de información de Akaike (AIC), criterio de Schwarz (SC), prueba Breusch Pagan para heteroscedasticidad, y prueba Razón de Verosimilitud sobre los parámetros del rezago espacial.

### 3.11. Modelo Matricial Propuesto

Considerando la última regresión presentada, se propone el siguiente modelo:

$$lprob13 = \alpha + \rho Wlprob13 + X\beta + \gamma Wlpnalcan + \varepsilon \quad (3.8)$$

Donde:

$lprob13$  : vector de dimensión  $448 \times 1$  que corresponde a la variable dependiente: logaritmo natural de la tasa de robos por mil habitantes ( $lprob13$ ).

$\alpha$  : Vector de interceptos de dimensión  $448 \times 1$ .

$\rho$  : coeficiente asociado al rezago espacial de la variable dependiente

$W$  : matriz de pesos espaciales (estandarizada por filas) de dimensión  $448 \times 448$

$Wlprob13$  : rezago espacial de la variable dependiente ( $lprob13$ )

$Wlpnalcan$  : rezago espacial de la variable independiente ( $lpnalcan$ )

$X$  : matriz de dimensión  $448 \times 3$  que contiene a las siguientes variables: logaritmo de área ( $larea$ ), logaritmo de la tasa de viviendas con presencia de emigración ( $lpemig$ ) y logaritmo de la tasa de mujeres ( $lp muj$ ).<sup>4</sup>

$\beta$  : vector de dimensión  $3 \times 1$  que contiene los parámetros asociados a las variables independientes.

$\varepsilon$  : vector de errores de dimensión  $448 \times 1$ . Puede tener estructura espacial.

Si  $\varepsilon$  tuviese estructura espacial se tendría:

$$\varepsilon = \lambda W\varepsilon + u \quad (3.9)$$

$\lambda$  : coeficiente asociado al rezago espacial de los errores ( $\lambda$  es análogo al  $\rho$  pero para los errores)

$W$  : matriz de pesos espaciales estandarizada de dimensión  $448 \times 448$

$u$  : vector de errores aleatorios  $u \sim N(0, \Omega)$

---

<sup>4</sup>El cálculo de las tasas de estas variables se las puede encontrar en el cuadro (3.1)

$\Omega$  : matriz diagonal de varianza – covarianza.

### 3.12. Diagnóstico de la Regresión

Como complemento al análisis anterior que originalmente fué realizado para el período completo correspondiente al año 2013, ahora se presenta un análisis a través de regresiones realizadas para tres periodos semestrales denominados: 1S (ltot131) correspondiente al primer semestre del año 2013, 2S (ltot132) al segundo semestre del año 2013 y 3S (ltot141) correspondiente al primer semestre del año 2014 de forma individualizada, empleando únicamente la variable dependiente del modelo propuesto “*Robos*”. Por medio GeoDa se generan nuevas variables como resultantes de los análisis hechos en cada regresión tales como los valores pronosticados de las variables así como los valores pronosticados de los residuos y que son adicionados en la tabla general de variables, posteriormente se aprovecha de dichas nuevas variables para realizar predicciones, y análisis de residuos como resultado del pronóstico obtenido de las regresiones antes mencionadas, en especial se buscará si existe a través del tiempo (tres semestres), patrones de autocorrelación espacial ó valores altos de variables (Outliers), y si dichos patrones se mantienen o varían en forma notoria a través del tiempo.

Se propone además en este estudio analizar en tiempo y espacio utilizando la funcionalidad del GEODA si coinciden los mapas de tendencia de cada semestre versus el mapa presentado en el análisis de un año.

#### 3.12.1. Modelo de Regresión de Superficie de Tendencia (Trend Surface Regression)

A continuación se presentan alternativas de modelos de regresión espacial de tendencia de crecimiento o disminución de la variable de interés *Robos*, para cada uno de los tres semestres para los cuales se dispone información (ltot131, ltot132 y ltot141). Para realizar estas regresiones fue preciso considerar las coordenadas  $X$  y  $Y$  de los centroides de cada una de las zonas censales, como los puntos donde se concentran la cantidad de robos reportados para cada zona censal.

Se analiza dos tipos de regresiones, la de superficie de tendencia lineal (con los valores de las coordenadas  $X$  y  $Y$  como las variables explicativas) y la de superficie de tendencia cuadrática (con  $X$ ,  $Y$ ,  $X^2$ ,  $Y^2$ ,  $XY$  como las variables explicativas), con la finalidad de verificar cual es

el modelo que mejor se ajuste a lo que se trata de investigar (comportamiento de los robos en las diferentes zonas censales).

La idea de esto es poder proyectar en cada semestre hacia donde existe la tendencia en que se espera se incremente o disminuya (de acuerdo a los coeficientes de la regresión) la variable en estudio, y al realizar un análisis comparativo de las tres regresiones en conjunto y ayudándonos de los polígonos de Thiessen ver si existe algún patrón consistente en alguna dirección específica.

Para llevar a cabo lo propuesto a través de GeoDa se generan tanto los centroides ( $X$  y  $Y$ ) Fig (3.26), las variables cuadráticas ( $X^2$ ,  $Y^2$ ) y el producto de estas variables ( $XY$ ) que son necesarias para la ejecución de las regresiones, como también los polígonos denominados de Thiessen, Fig (3.27).

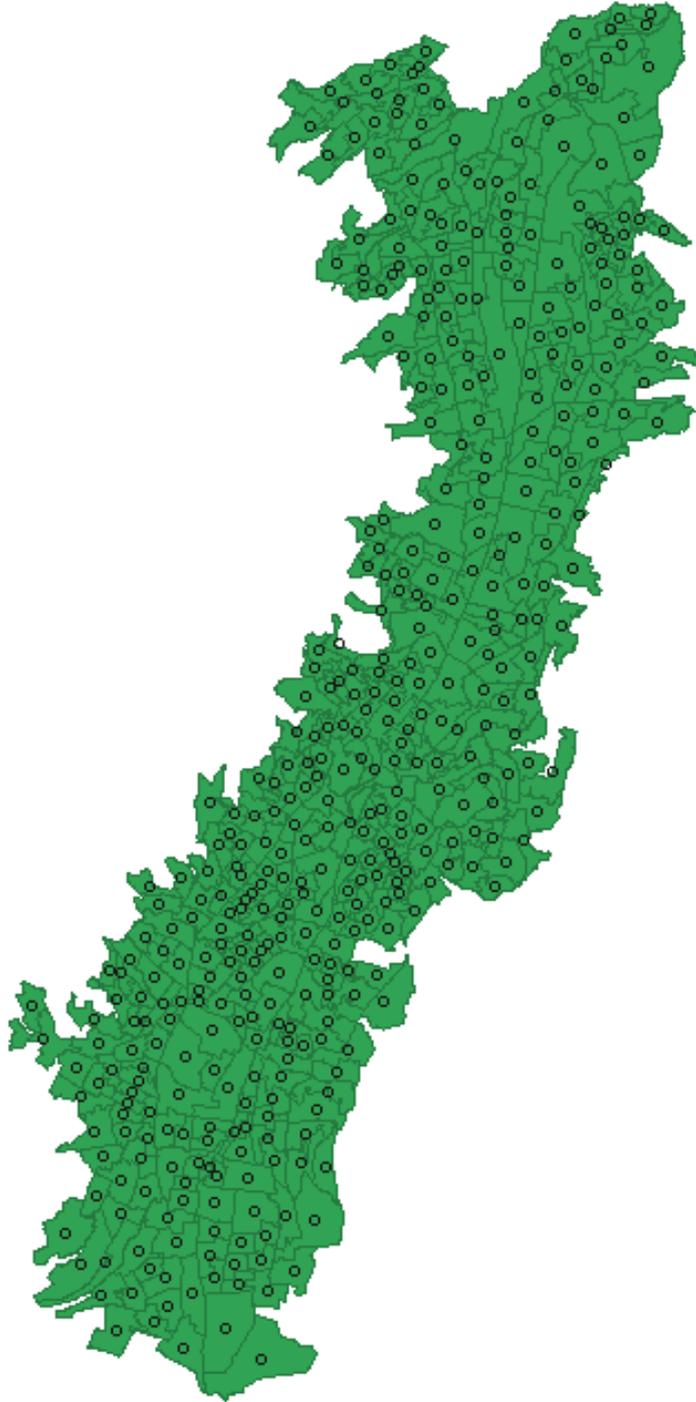


Figura 3.26: Zonas censales con sus centroides



Figura 3.27: Polígonos de Thiessen correspondientes a Quito urbano

En los cuadros siguientes (3.6), (3.7) y (3.8), se presenta los resultados de las regresiones de superficie de tendencia lineal para los tres semestres, pudiéndose determinar consistencia en los valores y los signos de los coeficientes obtenidos producto de las regresiones, pues se puede evidenciar una tendencia de crecimiento porcentual (signo positivo) hacia el norte (longitud o eje  $X$ )<sup>5</sup>. Para el coeficiente correspondiente a la variable  $Y$  se verifica en cambio la presencia del signo negativo en los resultados de las tres regresiones, esto sugeriría una reducción en porcentaje del valor de la variable analizada de oeste a este, se hace notar que los valores obtenidos para esta variable son tan pequeños (en el orden de  $10^{-5}$ ) que se puede decir que la variación de tendencia en porcentaje es prácticamente nula en esta dirección.

Para estas regresiones es necesario generar y utilizar una matriz de pesos espaciales de segundo orden de contiguidad que debe incluir además la contiguidad de primer orden.

<sup>5</sup>En sistemas geográficos de tipo esférico las líneas horizontales latitud (este-oeste) corresponden al eje  $Y$  y las líneas verticales longitud (norte-sur) corresponden al eje  $X$ .

## Linear Trend Surface

### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

```

Data set           : ROBTHIESENABR8_1105
Dependent Variable :   ltot131   Number of Observations:  448
Mean dependent var :   0.915253   Number of Variables   :    3
S.D. dependent var :   1.42223    Degrees of Freedom    :  445

R-squared          :   0.106433    F-statistic           :    26.5019
Adjusted R-squared :   0.102417    Prob(F-statistic)     : 1.33604e-011
Sum squared residual:   809.741    Log likelihood        :   -768.275
Sigma-square       :   1.81964    Akaike info criterion :   1542.55
S.E. of regression :   1.34894    Schwarz criterion     :   1554.86
Sigma-square ML    :   1.80746
S.E of regression ML:   1.34442
  
```

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	396.8148	105.9602	3.744942	0.00020
XCNTRD	0.0002126701	2.992365e-005	7.107089	0.00000
YCNTRD	-5.623991e-005	1.233874e-005	-4.557994	0.00001

### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 4428.904356

#### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	0.4786	0.78717

#### DIAGNOSTICS FOR HETEROSKEDASTICITY

##### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	12.9483	0.00154
Koenker-Bassett test	2	14.0145	0.00091

##### SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	5	21.6265	0.00062

#### DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : ROBTHIESENABR8\_1105.gal

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.5591	20.5266	0.00000
Lagrange Multiplier (lag)	1	399.7776	0.00000
Robust LM (lag)	1	0.5427	0.46130
Lagrange Multiplier (error)	1	399.3129	0.00000
Robust LM (error)	1	0.0781	0.77996
Lagrange Multiplier (SARMA)	2	399.8556	0.00000

## COEFFICIENTS VARIANCE MATRIX

CONSTANT	XCNTRD	YCNTRD
11227.563158	0.002179	-0.001295
0.002179	0.000000	-0.000000
-0.001295	-0.000000	0.000000

OBS	ltot131	PREDICTED	RESIDUAL
1	-1.72917	0.18916	-1.91834
2	-1.93442	0.26712	-2.20153
3	-1.85348	0.39233	-2.24581
4	1.11582	0.47601	0.63981
5	1.11538	0.45092	0.66446
.			
.			
.			
200	1.34301	1.87511	-0.53210
201	-0.17395	1.84576	-2.01971
202	0.07565	1.77287	-1.69722
203	0.45141	1.79689	-1.34548
204	-0.55809	1.69873	-2.25682
205	0.97255	1.59412	-0.62157
.			
.			
.			
440	0.16158	0.50139	-0.33981
441	0.30875	1.12514	-0.81639
442	1.39629	1.28456	0.11173
443	4.29196	1.40653	2.88543
444	-0.20896	0.85502	-1.06399
445	2.05328	1.58252	0.47076
446	3.45784	0.92114	2.53670
447	1.75938	0.95177	0.80762
448	0.55328	0.95412	-0.40084

===== END OF REPORT =====

**Cuadro 3.6: Linear Trend Surface Primer Semestre**

## SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : ROBTHIESENABR8\_1105  
 Dependent Variable : ltot132 Number of Observations: 448  
 Mean dependent var : 0.540274 Number of Variables : 3  
 S.D. dependent var : 1.36595 Degrees of Freedom : 445

R-squared : 0.077111 F-statistic : 18.5908  
 Adjusted R-squared : 0.072964 Prob(F-statistic) : 1.76083e-008  
 Sum squared residual: 771.432 Log likelihood : -757.418  
 Sigma-square : 1.73356 Akaike info criterion : 1520.84  
 S.E. of regression : 1.31665 Schwarz criterion : 1533.15  
 Sigma-square ML : 1.72195  
 S.E of regression ML: 1.31223

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	264.6029	103.4234	2.558445	0.01084
XCNTRD	0.000168745	2.920724e-005	5.777505	0.00000
YCENRD	-3.960521e-005	1.204333e-005	-3.288559	0.00109

## REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 4428.904356

## TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	0.5823	0.74742

## DIAGNOSTICS FOR HETEROSKEDASTICITY

## RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	13.5535	0.00114
Koenker-Bassett test	2	14.8629	0.00059

## SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	5	23.7970	0.00024

## DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : ROBTHIESENABR8\_1105.gal

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.5286	19.4175	0.00000
Lagrange Multiplier (lag)	1	356.8740	0.00000
Robust LM (lag)	1	0.1661	0.68359
Lagrange Multiplier (error)	1	356.8431	0.00000
Robust LM (error)	1	0.1352	0.71305
Lagrange Multiplier (SARMA)	2	357.0092	0.00000

## COEFFICIENTS VARIANCE MATRIX

CONSTANT	XCNTRD	YCNTRD
10696.389450	0.002076	-0.001234
0.002076	0.000000	-0.000000
-0.001234	-0.000000	0.000000

OBS	ltot132	PREDICTED	RESIDUAL
1	-0.63056	0.01571	-0.64627
2	-1.93442	0.08043	-2.01485
3	-1.85348	0.18206	-2.03555
4	0.84755	0.24953	0.59803
5	0.56884	0.23308	0.33576
.			
.			
.			
200	0.70702	1.29001	-0.58299
201	0.16252	1.26349	-1.10097
202	-1.87026	1.20023	-3.07049
203	-0.33704	1.22207	-1.55912
204	-1.65670	1.14694	-2.80365
205	1.08378	1.06490	0.01887
.			
.			
.			
440	-1.44786	0.14005	-1.58791
441	0.10808	0.68934	-0.58126
442	0.92087	0.87966	0.04121
443	3.73113	0.94087	2.79026
444	0.37882	0.41861	-0.03979
445	1.67680	1.09834	0.57846
446	2.72424	0.59250	2.13174
447	1.75938	0.62850	1.13089
448	0.11797	0.62823	-0.51027

===== END OF REPORT =====

Cuadro 3.7: Linear Trend Surface Segundo Semestre

## SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : ROBTHIESENABR8\_1105  
 Dependent Variable : ltot141 Number of Observations: 448  
 Mean dependent var : 0.935715 Number of Variables : 3  
 S.D. dependent var : 1.20949 Degrees of Freedom : 445

R-squared : 0.096785 F-statistic : 23.8422  
 Adjusted R-squared : 0.092725 Prob(F-statistic) : 1.45731e-010  
 Sum squared residual: 591.939 Log likelihood : -698.093  
 Sigma-square : 1.3302 Akaike info criterion : 1402.19  
 S.E. of regression : 1.15334 Schwarz criterion : 1414.5  
 Sigma-square ML : 1.32129  
 S.E of regression ML: 1.14947

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	59.04487	90.59581	0.6517395	0.51491
XCNTRD	0.0001390795	2.558468e-005	5.436047	0.00000
YCNTRD	-1.66509e-005	1.054961e-005	-1.578343	0.11520

## REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 4428.904356

## TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	8.5468	0.01393

## DIAGNOSTICS FOR HETEROSKEDASTICITY

## RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	18.7384	0.00009
Koenker-Bassett test	2	14.7860	0.00062

## SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	5	25.0976	0.00013

## DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : ROBTHIESENABR8\_1105.gal

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.5064	18.6124	0.00000
Lagrange Multiplier (lag)	1	325.5570	0.00000
Robust LM (lag)	1	1.1610	0.28127
Lagrange Multiplier (error)	1	327.5078	0.00000
Robust LM (error)	1	3.1118	0.07773
Lagrange Multiplier (SARMA)	2	328.6688	0.00000

## COEFFICIENTS VARIANCE MATRIX

CONSTANT	XCNTRD	YCNTRD
8207.600073	0.001593	-0.000947
0.001593	0.000000	-0.000000
-0.000947	-0.000000	0.000000

OBS	ltot141	PREDICTED	RESIDUAL
1	-1.72917	0.66766	-2.39684
2	-1.93442	0.73013	-2.66454
3	0.09243	0.82117	-0.72874
4	1.22704	0.88017	0.34687
5	1.30644	0.87766	0.42878
.			
.			
.			
200	0.70702	1.51582	-0.80880
201	-0.17395	1.48366	-1.65762
202	-0.77165	1.41424	-2.18589
203	-0.33704	1.44111	-1.77815
204	-0.55809	1.38798	-1.94607
205	0.84739	1.32343	-0.47604
.			
.			
.			
440	-0.34925	0.37694	-0.72619
441	0.85530	1.00292	-0.14763
442	1.34072	1.36314	-0.02242
443	4.14510	1.30025	2.84486
444	0.12751	0.60006	-0.47255
445	2.46526	1.48687	0.97839
446	2.81201	1.13027	1.68174
447	1.89997	1.19722	0.70274
448	1.01591	1.19020	-0.17430

===== END OF REPORT =====

Cuadro 3.8: Linear Trend Surface Tercer Semestre

### Quadratic Trend Surface

#### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

```

Data set           : ROBTHIESENABR8_1105
Dependent Variable :   ltot131  Number of Observations:  448
Mean dependent var :   0.915253  Number of Variables   :    6
S.D. dependent var :   1.42223   Degrees of Freedom    :  442

R-squared          :   0.344648   F-statistic           :    46.4894
Adjusted R-squared :   0.337235   Prob(F-statistic)    : 1.40961e-038
Sum squared residual:   593.872   Log likelihood       :   -698.824
Sigma-square       :    1.3436   Akaike info criterion:   1409.65
S.E. of regression :    1.15914   Schwarz criterion    :   1434.28
Sigma-square ML    :    1.32561
S.E of regression ML:    1.15135

```

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-1477511	120754.9	-12.23562	0.00000
XCNTRD	-0.4746029	0.05587586	-8.493881	0.00000
YCNTRD	0.3331427	0.02720527	12.24552	0.00000
X2	-7.948163e-008	1.034809e-008	-7.680802	0.00000
Y2	-1.903057e-008	1.567106e-009	-12.14376	0.00000
XY	5.995937e-008	6.98568e-009	8.583183	0.00000

#### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 14851036.790657

#### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	2.7613	0.25142

#### DIAGNOSTICS FOR HETEROSKEDASTICITY

##### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	8.5844	0.12683
Koenker-Bassett test	5	7.5482	0.18296

##### SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	20	N/A	N/A

#### DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : ROBTHIESENABR8\_1105.gal

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.4087	15.4808	0.00000
Lagrange Multiplier (lag)	1	217.0218	0.00000
Robust LM (lag)	1	4.4492	0.03492
Lagrange Multiplier (error)	1	213.3155	0.00000
Robust LM (error)	1	0.7428	0.38876
Lagrange Multiplier (SARMA)	2	217.7646	0.00000

## COEFFICIENTS VARIANCE MATRIX

CONSTANT	XCNTRD	YCNTRD	X2	Y2	
14581739517.150635	4337.203390	-3260.435744	0.000542	0.000184	0.000184
4337.203390	0.003122	-0.001112	0.000000	0.000000	
-3260.435744	-0.001112	0.000740	-0.000000	-0.000000	
0.000542	0.000000	-0.000000	0.000000	0.000000	
0.000184	0.000000	-0.000000	0.000000	0.000000	
-0.000519	-0.000000	0.000000	-0.000000	-0.000000	

XY

-0.000519  
-0.000000  
0.000000  
-0.000000  
-0.000000  
0.000000

OBS	ltot131	PREDICTED	RESIDUAL
1	-1.72917	-0.64330	-1.08588
2	-1.93442	-0.46383	-1.47059
3	-1.85348	-0.12051	-1.73297
4	1.11582	0.10010	1.01572
5	1.11538	-0.12086	1.23624
.			
.			
.			
445	2.05328	1.61142	0.44186
446	3.45784	1.38905	2.06879
447	1.75938	1.08723	0.67215
448	0.55328	1.15931	-0.60603

===== END OF REPORT =====

Cuadro 3.9: Quadratic Trend Surface Primer Semestre

## SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : ROBTHIESENABR8\_1105  
 Dependent Variable : ltot132 Number of Observations: 448  
 Mean dependent var : 0.540274 Number of Variables : 6  
 S.D. dependent var : 1.36595 Degrees of Freedom : 442

R-squared : 0.272893 F-statistic : 33.1777  
 Adjusted R-squared : 0.264668 Prob(F-statistic) : 9.36746e-029  
 Sum squared residual: 607.78 Log likelihood : -704.009  
 Sigma-square : 1.37507 Akaike info criterion : 1420.02  
 S.E. of regression : 1.17263 Schwarz criterion : 1444.65  
 Sigma-square ML : 1.35665  
 S.E of regression ML: 1.16475

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-1293929	122160.7	-10.59202	0.00000
XCNTRD	-0.435577	0.05652636	-7.705733	0.00000
YCNTRD	0.2932935	0.02752199	10.6567	0.00000
X2	-7.043767e-008	1.046856e-008	-6.728497	0.00000
Y2	-1.68257e-008	1.58535e-009	-10.61324	0.00000
XY	5.463634e-008	7.067006e-009	7.731186	0.00000

## REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 14851036.790657

## TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	2.1127	0.34773

## DIAGNOSTICS FOR HETEROSKEDASTICITY

## RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	11.6147	0.04047
Koenker-Bassett test	5	11.0137	0.05111

## SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	20	N/A	N/A

## DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : ROBTHIESENABR8\_1105.gal

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.4130	15.6383	0.00000
Lagrange Multiplier (lag)	1	219.3520	0.00000
Robust LM (lag)	1	2.2716	0.13176
Lagrange Multiplier (error)	1	217.8156	0.00000
Robust LM (error)	1	0.7352	0.39119
Lagrange Multiplier (SARMA)	2	220.0872	0.00000

## COEFFICIENTS VARIANCE MATRIX

CONSTANT	XCNTRD	YCNTRD	X2	Y2	
14923230504.009998	4438.776721	-3336.792163	0.000555	0.000188	
4438.776721	0.003195	-0.001138	0.000000	0.000000	
-3336.792163	-0.001138	0.000757	-0.000000	-0.000000	
0.000555	0.000000	-0.000000	0.000000	0.000000	
0.000188	0.000000	-0.000000	0.000000	0.000000	
-0.000531	-0.000000	0.000000	-0.000000	-0.000000	

## XY

-0.000531
-0.000000
0.000000
-0.000000
-0.000000
0.000000

OBS	ltot132	PREDICTED	RESIDUAL
1	-0.63056	-0.77014	0.13958
2	-1.93442	-0.60686	-1.32756
3	-1.85348	-0.29897	-1.55451
4	0.84755	-0.10155	0.94911
5	0.56884	-0.28915	0.85799
6	1.57739	-0.05532	1.63271
7	-0.49543	-0.60657	0.11114
8	1.66242	-0.54384	2.20626
9	-1.60503	-1.02453	-0.58050
10	-2.05284	-1.39344	-0.65940
.			
.			
.			
440	-1.44786	-0.66254	-0.78532
441	0.10808	1.02923	-0.92115
442	0.92087	1.22492	-0.30405
443	3.73113	1.23466	2.49647
444	0.37882	-0.78868	1.16750
445	1.67680	1.13445	0.54235
446	2.72424	1.01036	1.71388
447	1.75938	0.77538	0.98401
448	0.11797	0.83286	-0.71489

===== END OF REPORT =====

Cuadro 3.10: Quadratic Trend Surface Segundo Semestre

## SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : ROBTHIESENABR8\_1105  
 Dependent Variable : ltot141 Number of Observations: 448  
 Mean dependent var : 0.935715 Number of Variables : 6  
 S.D. dependent var : 1.20949 Degrees of Freedom : 442

R-squared : 0.255516 F-statistic : 30.34  
 Adjusted R-squared : 0.247094 Prob(F-statistic) : 1.57146e-026  
 Sum squared residual: 487.911 Log likelihood : -654.801  
 Sigma-square : 1.10387 Akaike info criterion : 1321.6  
 S.E. of regression : 1.05065 Schwarz criterion : 1346.23  
 Sigma-square ML : 1.08909  
 S.E of regression ML: 1.04359

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-998318.7	109453.2	-9.120963	0.00000
XCNTRD	-0.372861	0.05064634	-7.362053	0.00000
YCNTRD	0.229138	0.02465908	9.292236	0.00000
X2	-6.428842e-008	9.379594e-009	-6.854073	0.00000
Y2	-1.332773e-008	1.420438e-009	-9.382828	0.00000
XY	4.739156e-008	6.331878e-009	7.484597	0.00000

## REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 14851036.790657

## TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	6.5835	0.03719

## DIAGNOSTICS FOR HETEROSKEDASTICITY

## RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	20.1567	0.00117
Koenker-Bassett test	5	15.6142	0.00804

## SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	20	N/A	N/A

## DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : ROBTHIESENABR8\_1105.gal

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.4159	15.7447	0.00000
Lagrange Multiplier (lag)	1	220.9518	0.00000
Robust LM (lag)	1	1.7147	0.19038
Lagrange Multiplier (error)	1	220.8820	0.00000
Robust LM (error)	1	1.6448	0.19966
Lagrange Multiplier (SARMA)	2	222.5967	0.00000

## COEFFICIENTS VARIANCE MATRIX

CONSTANT	XCNTRD	YCNTRD	X2	Y2
11980004703.710421	3563.341462	-2678.695193	0.000445	0.000151
3563.341462	0.002565	-0.000914	0.000000	0.000000
-2678.695193	-0.000914	0.000608	-0.000000	-0.000000
0.000445	0.000000	-0.000000	0.000000	0.000000
0.000151	0.000000	-0.000000	0.000000	0.000000
-0.000426	-0.000000	0.000000	-0.000000	-0.000000

## XY

-0.000426
-0.000000
0.000000
-0.000000
-0.000000
0.000000

OBS	ltot141	PREDICTED	RESIDUAL
1	-1.72917	-0.00609	-1.72309
2	-1.93442	0.16051	-2.09493
3	0.09243	0.44704	-0.35461
4	1.22704	0.62609	0.60095
5	1.30644	0.49199	0.81445
6	2.46771	0.72802	1.73969
7	0.72835	0.27331	0.45504
8	0.64077	0.30771	0.33306
9	0.79287	-0.11827	0.91114
10	0.14438	-0.40515	0.54954
.			
.			
.			
440	-0.34925	-0.19365	-0.15560
441	0.85530	1.27891	-0.42362
442	1.34072	1.64720	-0.30648
443	4.14510	1.49569	2.64942
444	0.12751	-0.23903	0.36654
445	2.46526	1.44869	1.01657
446	2.81201	1.49980	1.31221
447	1.89997	1.37432	0.52565
448	1.01591	1.40849	-0.39259

===== END OF REPORT =====

Cuadro 3.11: Quadratic Trend Surface Tercer Semestre

Del análisis de los dos tipos de regresiones presentadas (linear trend surface y quadratic trend surface) si bien en el primer tipo de modelo se tiene un ajuste que es aceptable, al analizar los resultados del modelo quadratic trend surface se tiene un valor para  $R^2$  ajustado en cada semestre

mejor que los correspondientes al modeo (linear trend surface), lo mismo se puede decir del p-valor.

Con respecto al análisis de multicolinealidad se tiene un problema manifiesto de ello en los dos tipos de modelos, mostrándose un mayor valor para el modelo cuadrático lo cual es bastante obvio puesto que las regresoras  $X^2$ ,  $Y^2$  y  $XY$  están entre sí correlacionadas.

En las regresiones de tipo lineal, el test de normalidad (Jarque - Bera) para los sus valores obtenidos indican que se está transgrediendo el supuesto de normalidad para modelos de Mínimos Cuadrados Ordinarios. Algo similar ocurre en la regresión cuadrática.

Con respecto a los test de heteroscedasticidad, en el modelo lineal (para los tres semestres) se rechaza la hipótesis nula de homocedasticidad, mientras que en los modelos de tipo cuadrático, en el primer semestre no existe evidencia suficiente para rechazar esta hipótesis nula, en el segundo semestre los valores obtenidos no son concluyentes por lo que tampoco se puede rechazar esta hipótesis únicamente en el tercer semestre se tendría evidencia que permitiría rechazar la hipótesis de homocedasticidad. Concluyéndose por tanto que el modelo cuadrático es el que mejor se ajusta a este análisis.

Para visualizar en forma gráfica estas tendencias, se utiliza los denominados polígonos de Thiessen que presentan en deciles la distribución de la variable robos para cada semestre, pudiéndose apreciar en ellos que se tiene una tendencia pronosticada de un incremento porcentual de la variable estudiada hacia el norte de la ciudad de Quito. Además el modelo de regresión cuadrática (Quadratic trend surface predicted), ver figura (3.28) sugiere una tendencia de crecimiento radial desde los polígonos más centrales los cuales presentan altos índices de robos hacia sus contornos.

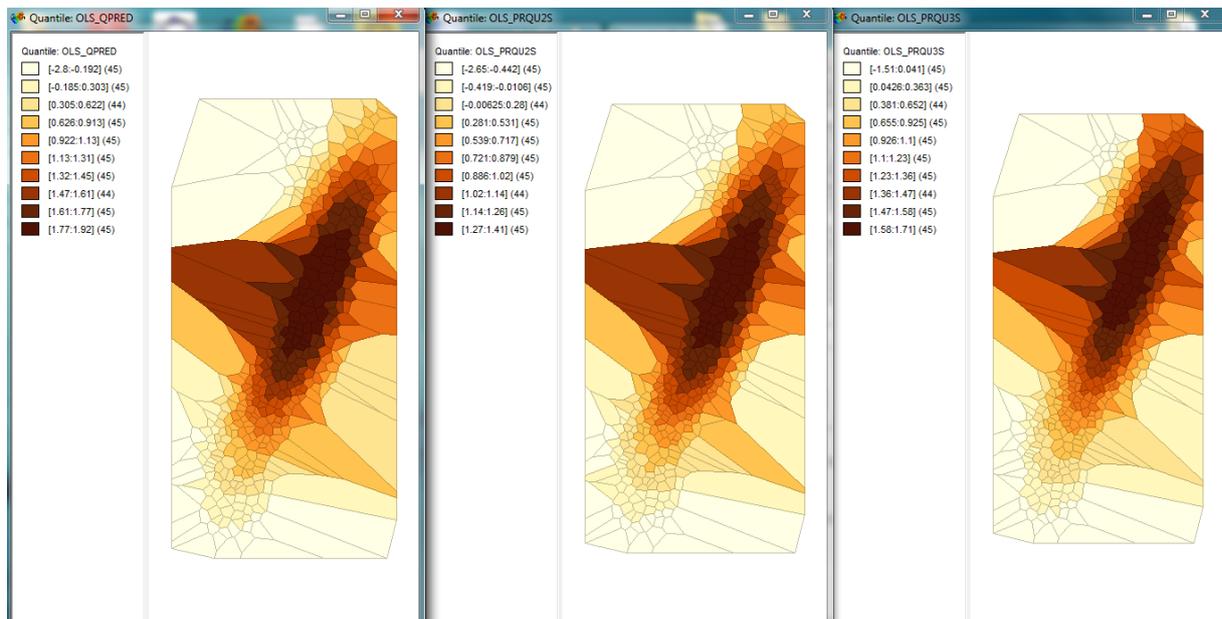


Figura 3.28: Quadratic trend surface predicted value map contiguidad de orden 2

### 3.12.2. Mapas de residuos y valores pronosticados

Para el proceso de especificación del modelo espacial Anselin (2005) sugiere como se ha indicado anteriormente partir de una regresión MCO considerando las variables exógenas propuestas generar mapas que incluyan, las variables correspondientes, los valores pronosticados y los residuos. A decir de Anselin (2005) Los mapas con valores pronosticados pueden ser pensados como mapas suavizados, en el sentido de que la variabilidad aleatoria debido a factores distintos de los incluidos en el modelo se han suavizado. Un mapa residual a su vez puede dar una indicación de exceso o baja predicción sistemática en regiones particulares, lo que puede ser evidencia de autocorrelación espacial.

### 3.12.3. Mapas de residuos

Aprovechando la funcionalidad de GeoDa y puesto que se han generado en la tabla nuevos datos como los valores pronosticados y los residuos, estos valores permiten explorar también la existencia de patrones. El mapa de residuos más útil probablemente es el mapa de desviación estándar. Utilizando entonces los valores de los residuos obtenidos en las regresiones de superficie de tendencia cuadrática para los tres semestres antes mencionados figura (3.29) en ellos se puede apreciar claramente la formación de polígonos que sirven para ilustrar la formación de patrones con sobre o baja predicción, así como la magnitud de los residuos, especialmente aquellos mayores

a  $\pm 2$  desviaciones estándar. Tener en cuenta los amplios patrones con sobre predicción (residuos negativos, o tonos azules) en las zonas periféricas y subpredicción (residuos positivos, o tonos marrones) en el núcleo. La inspección visual sugiere la presencia de autocorrelación espacial, pero esto requiere de pruebas formales antes de poder afirmar concluyentemente. También tener en cuenta varios residuos de gran tamaño (el marrón muy oscuro y azul). Esto no debe sorprender, ya que el *modelo* sólo contiene la ubicación como una variable y se considera que no hay otras características distintivas de los robos.

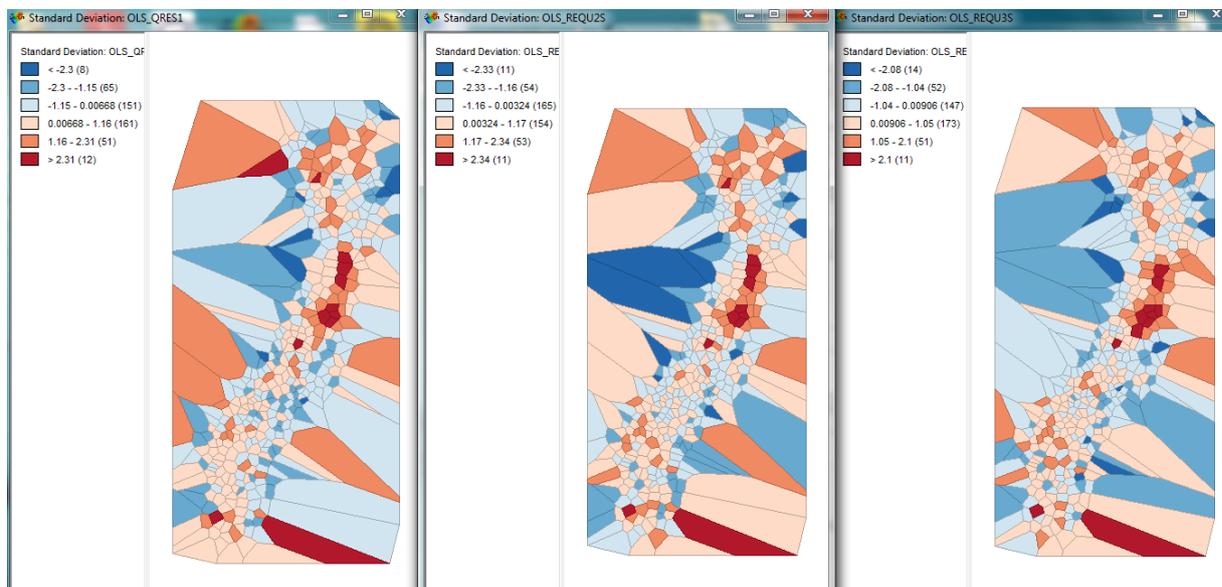


Figura 3.29: Mapa de Residuos, Superficie de Tendencia Cuadrática (Quadratic Trend Surface) contiguidad de orden 2

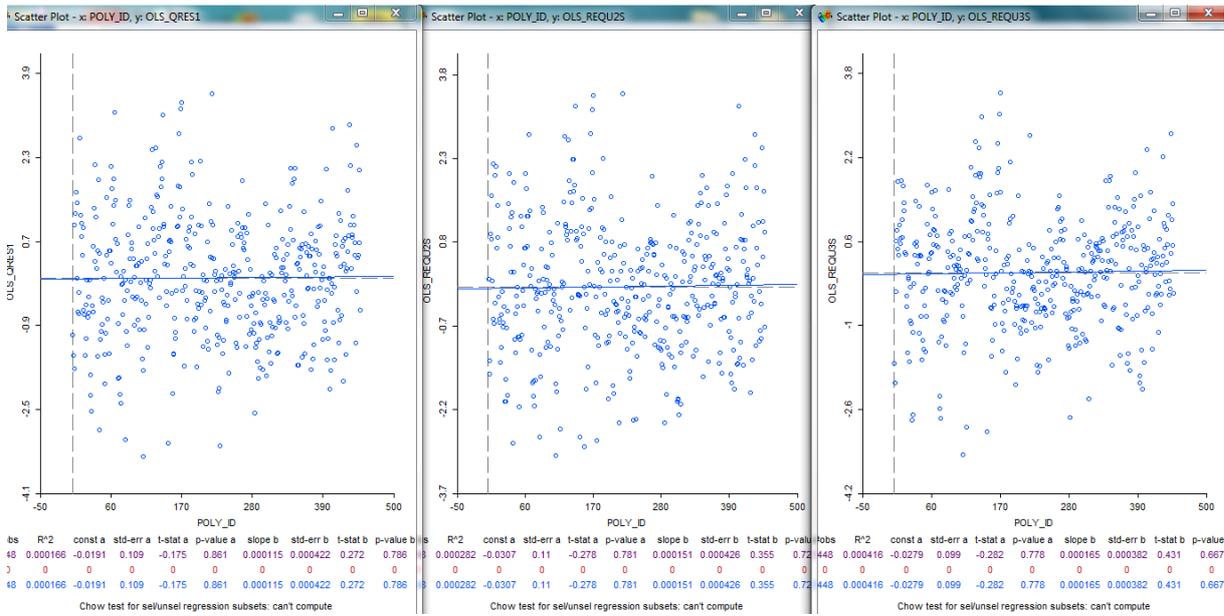


Figura 3.30: Quadratic trend surface residual /fitted value plot (contiguidad de orden 2)

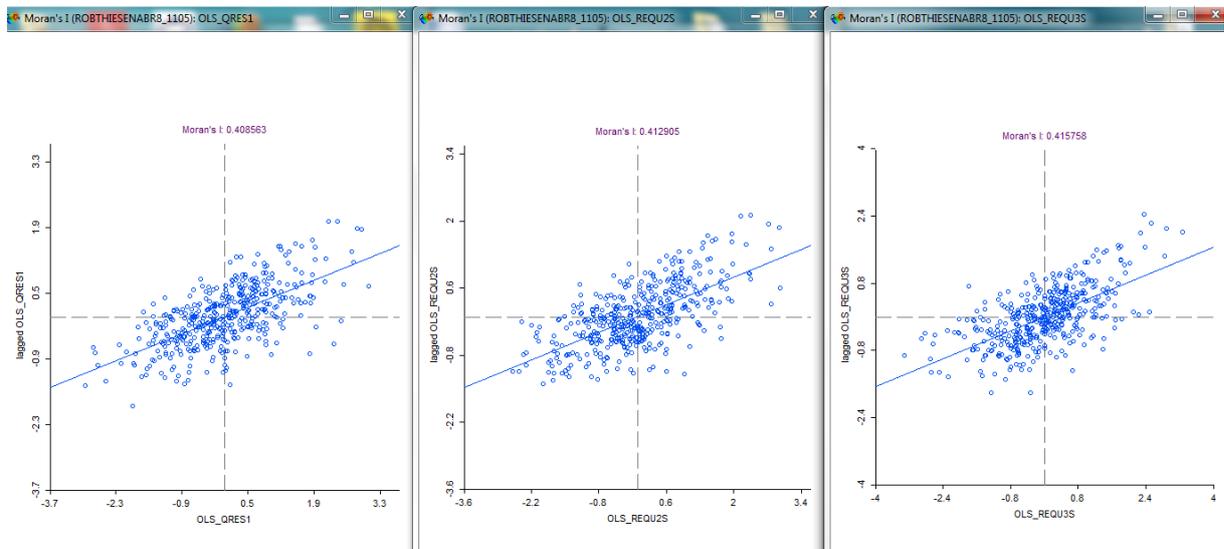


Figura 3.31: Moran scatter plot para Quadratic Trend Surface Residual (contiguidad de orden 2)

### 3.12.4. Máxima Verosimilitud en el modelo de rezagos espaciales

En modelos con rezagos espaciales se debe hacer una distinción entre los residuos y el error de predicción, el segundo es la diferencia entre el valor observado y el valor pronosticado que sólo

utiliza variables exógenas en lugar de tratar el retardo espacial  $W_y$  como valor observado.

El reporte de resultados en el modelo con rezagos espaciales presentado en este trabajo consta de las estimaciones de los coeficientes y sus errores asintóticos, pruebas-t, medidas de ajuste (Akaike Inf Criterion AIC, log likelihood y Schwarz Criterion SC), prueba de heterosedasticidad de Breusch-Pagan y la prueba de razón de verosimilitud para el parámetro de rezago espacial.

## SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : SHAPE\_ORD2\_5VAR\_ABR10\_1736  
 Spatial Weight : datos\_panel\_robos.gal  
 Dependent Variable : lprob13 Number of Observations: 448  
 Mean dependent var : 1.3911 Number of Variables : 7  
 S.D. dependent var : 1.46128 Degrees of Freedom : 441  
 Lag coeff. (Rho) : 0.67148

R-squared : 0.654210 Log likelihood : -592.553  
 Sq. Correlation : - Akaike info criterion : 1199.11  
 Sigma-square : 0.738381 Schwarz criterion : 1227.84  
 S.E of regression : 0.859291

Variable	Coefficient	Std.Error	z-value	Probability
W_lprob13	0.6714796	0.03797949	17.68006	0.00000
CONSTANT	-138.4382	17.81991	-7.768736	0.00000
larea	0.3624583	0.07600914	4.768615	0.00000
lpemig	-0.3985076	0.1638764	-2.431757	0.01503
lpmuj	5.682644	2.13897	2.65672	0.00789
lpnanlf	14.94505	2.908974	5.137565	0.00000
lnviv	-0.3243317	0.1697712	-1.910405	0.05608

## REGRESSION DIAGNOSTICS

## DIAGNOSTICS FOR HETEROSKEDASTICITY

## RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	33.3015	0.00000

## DIAGNOSTICS FOR SPATIAL DEPENDENCE

## SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : datos\_panel\_robos.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	200.2179	0.00000

## COEFFICIENTS VARIANCE MATRIX

CONSTANT	larea	lpemig	lpmuj	lpnanlf
317.549298	-0.409663	0.812629	-8.672660	-39.307266
-0.409663	0.005777	0.001388	0.002693	0.049299
0.812629	0.001388	0.026855	-0.067293	-0.076226
-8.672660	0.002693	-0.067293	4.575194	-2.850401
-39.307266	0.049299	-0.076226	-2.850401	8.462132
1.126368	-0.003385	0.001003	-0.013278	-0.176010
0.365390	-0.000697	0.000598	-0.014615	-0.039419
lnviv	W_lprob13			
1.126368	0.365390			
-0.003385	-0.000697			
0.001003	0.000598			
-0.013278	-0.014615			
-0.176010	-0.039419			
0.028822	0.000156			
0.000156	0.001442			

OBS	lprob13	PREDICTED	RESIDUAL	PRED ERROR
1	-0.63056	-0.09779	0.69141	-0.53277
2	-1.9344	0.19084	-1.19315	-2.12525
3	-1.8535	0.68386	-1.95847	-2.53734
4	1.6499	1.46964	0.63052	0.18026
5	1.5382	2.08013	-0.16047	-0.54188
6	2.3565	2.26039	0.26605	0.09609
7	0.092359	1.01016	0.43152	-0.91780
8	2.0545	1.49980	1.57951	0.55466
9	-1.605	1.07216	-1.77244	-2.67719
10	-2.0528	0.63060	-1.27510	-2.68344
.				
.				
.				
440	0.16158	-0.04244	-0.14337	0.20401
441	0.8553	0.55239	-0.25376	0.30291
442	1.8629	2.50674	-0.57838	-0.64383
443	4.7404	3.76865	0.37255	0.97174
444	0.74655	0.61773	0.10073	0.12882
445	2.5641	2.97232	-0.39588	-0.40822
446	3.847	2.73865	1.25624	1.10836
447	2.4431	2.22188	0.19071	0.22117
448	1.0159	1.84664	-0.68246	-0.83073

=====  
===== END OF REPORT =====

Cuadro 3.12: Resultados del modelo de rezagos espaciales-Estimación de Máxima Verosimilitud

# Capítulo 4

## MODELOS ESPACIALES PARA PANEL DE DATOS

La literatura de panel ha considerado recientemente los modelos de regresión de panel con perturbaciones espacialmente autocorrelacionados, tanto en el contexto de especificaciones tanto de efectos fijos (FE), así como efectos aleatorios (ARE) . En un entorno de componentes de error, Baltagi et al. (2003) introducen un modelo (también considerada en Anselin, 1988), donde los errores son idiosincráticos autocorrelacionados espacialmente, mientras que los efectos individuales no lo son. La matriz de varianza de un modelo de este tipo es complicado y el cálculo de su inversa computacionalmente es exigente. Kapoor et al. (2007) consideran un modelo en el que la correlación espacial en ambos tanto los componentes individuales como los componentes del error comparten el mismo parámetro espacial; y, por lo tanto, la expresión de la matriz de varianza es más simple y el cálculo de su inversa es más fácil. **spml** tiene en cuenta tanto las especificaciones y varios métodos para la estimación de los coeficientes de la regresión.

En el presente trabajo se describe la implementación de máxima verosimilitud de ambos modelos (es decir, con efectos individuales que pueden o no ser autocorrelacionado espacialmente). Se consideran modelos de efectos fijos, así como modelos de efectos aleatorios en el contexto de un modelo de tipo general espacial Cliff-Ord que incluye una variable dependiente rezagada espacialmente y un término de error espacialmente autocorrelacionado.

Además, **spml** presenta las características de los estimadores de momentos generalizados de un modelo tipo Cliff-Ord donde los efectos individuales son espacialmente autocorrelacionados . Una vez más, se implementan modelos con efectos aleatorios, así como con efectos fijos.

## 4.1. Estructuras de datos

Los datos de panel se refieren a una sección transversal de observaciones (individuos, grupos, países, regiones) repetidas durante varios períodos de tiempo. Cuando el número de observaciones de la sección transversal es constante a través de períodos de tiempo se dice que el panel es balanceado. El presente trabajo se centra únicamente en dichos paneles balanceados. En una configuración de panel espacial, las observaciones están asociadas con una posición particular en el espacio. Los datos pueden ser observados ya sea en lugares puntuales (por ejemplo, los datos de la vivienda) o agregados en áreas regulares o irregulares (por ejemplo, países, regiones, estados, ciudades, etc). La estructura de las interacciones entre cada par de unidades espaciales se representa por medio de una matriz de ponderaciones espaciales.

La matriz de ponderaciones espaciales  $W$  es una matriz positiva  $N \times N$ . Las observaciones aparecen tanto en filas como en columnas. Aquí, los elementos no nulos de la matriz indican si dos lugares son vecinos. Como consecuencia, el elemento  $w_{ij}$  indica la intensidad de la relación entre el elemento de la fila  $i$  y el de la columna  $j$ . Por convención, los elementos diagonales  $w_{ii}$  se ajustan a cero para excluir vecinos de si mismos. La matriz de pesos se utiliza generalmente en forma estandarizada por filas.

Una posible fuente de confusión en el desarrollo de rutinas ad-hoc se deriva de la diferente notación que caracteriza a los modelos de datos de panel espaciales en comparación con los modelos de datos de panel tradicionales. Por un lado, los datos de panel son generalmente ordenados primero por sección transversal y luego por período de tiempo (es decir, siendo el tiempo el índice más rápido). Por otro lado, los datos de panel espaciales se apilan primero por período de tiempo y luego por sección transversal. En **spml**, es tratado por el usuario. El orden interno de las funciones de estimación es por lo general (pero no siempre) los datos del panel espacial del primer caso. No obstante, los datos pueden ser suministrados de acuerdo con las convenciones implementadas en el paquete de **plm** para datos de panel de econometría (Croissant y Millo 2008). Existen tres posibilidades:

1. un `data.frame` cuyas dos primeras variables son los índices individuales y de tiempo. El argumento de índice se debe dejar en el valor predeterminado (es decir NULL)
2. un `data.frame` y un vector de caracteres indican los índices de las variables
3. un objeto de la clase `pdata.frame`

`pdata.frames` son objetos especiales creados para tratar con datos de panel. Son parte de una infraestructura general disponible en `pml` y hecha para manejar las operaciones de rezago y diferencia (de series). Los métodos disponibles en **spml** están orientados a los paneles estáticos; no obstante, definiendo los datos como `pdata.frame` podría simplificar el tiempo de cálculo de los rezagos de las variables explicativas.

La matriz de ponderaciones espaciales  $W$  puede ser una matriz objeto (con los estimadores realizando una mínima comprobación de compatibilidad de dimensión) o un objeto `listw` de la clase definida en **spdep** (Bivand 2011) La clase es un formato eficiente y tiene la ventaja de estar bien establecida en el ambiente R. Funcionalidades para cambiar entre los dos formatos están disponibles como funciones `listw2mat` y `mat2listw` del paquete **spdep**.

## 4.2. Clases y métodos para modelos de panel espacial

Los dos objetivos principales de **spml** son la estimación y la prueba de modelos de datos de panel espaciales. Por un lado, la información proporcionada en la salida de los procedimientos de prueba es similar a un objeto de clase **htest**; y, por lo tanto, produce un informe de salida similar. Por otro lado, los modelos de paneles espaciales requieren diferentes estructuras y métodos de las clases disponibles en **plm**. Por lo general, esto se debe a que los modelos de paneles espaciales implican la estimación de coeficientes adicionales (por ejemplo, el coeficiente para el término de rezago espacial en el modelo de rezago espacial de efectos fijos o el coeficiente de error correlacionado y los componentes de la varianza en las especificaciones de los efectos aleatorios).

La nueva clase **spml** hereda la estructura general de los objetos `lm`. El objeto **spml** es una lista de varios elementos, entre ellos: los coeficientes estimados, el vector de residuos y los `fitted.values` y un modelo de elemento que contiene los datos utilizados en la estimación. Como es común para la mayoría de los modelos que se estiman por máxima verosimilitud, **spml** también comprende un componente `logLik` con el valor de log-verosimilitud en el parámetro óptimo. Esto se puede extraer y volver a utilizar para fines de prueba o de selección de modelos.

Algunos elementos del objeto `lm` se ha pensado sean excluidos. Estas omisiones se deben en parte a la naturaleza del proceso de estimación (el cual no utiliza, por ejemplo, la descomposición "qr"). Se han añadido elementos específicos para adaptarse a los parámetros espaciales y de covarianza. Además del elemento usual `vcov` que da los coeficientes de la matriz de varianza covarianza,

el elemento **vcov.errcomp** contiene la matriz de covarianza de los coeficientes estimados de la covarianza del error .

Una nueva clase se define para los resúmenes de los objetos de **spml**. En consistentes con los objetos *lm* y *plm*, el método proporciona tablas de diagnóstico para los elementos de los objetos de **spml**. Métodos de impresión también están disponibles con una descripción mínima del objeto modelo (incluyendo llamadas, coeficientes y parámetros de covarianza). Además, los métodos extractores han sido definidos por algunos elementos relevantes de los objetos del modelo. Junto con los estándar *coef*, *residuos*, y *vcov*, se proporcionan métodos extractores para las matrices de covarianza del coeficiente espacial autorregresivo estimado y componentes de covarianza.

La disponibilidad de estos extractores es coherente con el marco de modelización del proyecto **R** y favorece la interoperabilidad de los objetos de **spml** con diagnósticos genéricos basados en pruebas de Wald. En particular se refiere a las funciones de *waldtest* (para cero-restricciones conjuntas) en **lmtest** (Zeileis y Hothorn 2002) y *linearHypothesis* (para restricciones lineales genéricas) en **car** (Fox y Weisberg 2010).

Por último, un método de extracción para efectos fijos y un procedimiento sumario para la visualización de ellos también están disponibles.

A lo largo de este análisis, todas las funciones se ilustran mediante el conocido conjunto de datos Munnell (1990) sobre la productividad del capital público en 48 estados de USA. observado durante 17 años (disponibles en R en el paquete *Ecdat*, Croissant 2011). Una matriz binaria de contiguidad de pesos espaciales para USA. está incluida en el paquete.

```
R > data("Produc", package = "Ecdat")
```

```
R > data("usaww")
```

Munnell (1990) especifica una función de producción Cobb-Douglas que relaciona el producto social bruto(*gsp*) de un estado dado a la entrada de capital público (*pcap*), el capital privado (*pc*), el empleo (*emp*) y la tasa de desempleo del estado(*unemp*) añadido para capturar efectos de ciclo económico. La fórmula del modelo es definida una vez e incluye un término constante.

```
R > fm <- -log(gsp) log(pcap) + log(pc) + log(emp) + unemp
```

Se transformará además la matriz de pesos a un objeto **listw** usando la infraestructura del paquete **spdep**:

```
R > library("spdep")
R > usalw <- mat2listw(usaww)
```

### 4.3. Modelos Espaciales de Panel de Datos

Los modelos espaciales de panel de datos capturan interacciones espaciales a través de las unidades espaciales y del tiempo. Existe información bastante amplia tanto para modelos estáticos, como para modelos dinámicos, en este trabajo se tratará únicamente como modelos estáticos.

Se empieza con un modelo de panel que incluye un rezago espacial de la variable dependiente y perturbaciones espaciales autorregresivas:

$$y = \lambda(I_T \otimes W_N)y + X\beta + u \quad (4.1)$$

Donde:

$y$  vector de observaciones  $NT \times 1$  de la variable dependiente.

$X$  matriz de observaciones de  $NT \times k$  de regresoras no estocásticas.

$I_T$  matriz identidad de dimension  $T$ .

$W_N$  matriz de pesos espaciales de constantes conocidas cuyos elementos de la diagonal son cero y

$\lambda$  el correspondiente parámetro espacial.

El vector de perturbaciones es la suma de dos términos:

$$u = (\iota_T \otimes I_N)\mu + \varepsilon \quad (4.2)$$

Donde:

$\iota_T$  vector de unos  $T \times 1$ .

$I_N$  matriz identidad  $N \times N$ .

$\mu$  vector invariante en el tiempo de efectos específicos individuales (no espacialmente autocorrelacionado).

$\varepsilon$  vector de innovaciones espacialmente autocorrelacionadas que sigue un proceso autorregresivo

de la forma siguiente:

$$\varepsilon = \rho(I_T \otimes W_N)\varepsilon + \nu \quad (4.3)$$

Con  $\rho$  ( $|\rho| < 1$ ) como el parámetro espacial autorregresivo,  $W_N$  la matriz de pesos espaciales,  $\nu_{it} \sim IID(0, \sigma_\nu^2)$  y  $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$ <sup>1</sup>

Como en la literatura clásica de pane de datos, los efectos individuales pueden ser tratados como fijos o aleatorios. En un modelo de efectos aleatorios, el uno está implícitamente asumiendo que los efectos individuales no observables no están correlacionados con las otras variables explicativas en el modelo. En este caso,  $\mu_i \sim IID(0, \sigma_\mu^2)$  el término del error puede ser re-escrito como:

$$\varepsilon = (I_T \otimes B_N^{-1})\nu \quad (4.4)$$

Donde  $B_N = (I_N - \rho W_N)$ . Como una consecuencia, el término del error se convierte en:

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B_N^{-1})\nu \quad (4.5)$$

Y la matriz de varianza-covarianza para  $\varepsilon$  es:

$$\Omega_u = \sigma_\mu^2(\iota_T \iota_T^T \otimes I_N) + \sigma_\nu^2[I_T \otimes (B_N^T B_N)^{-1}] \quad (4.6)$$

Al obtener varias pruebas de multiplicador de Lagrange (LM) , Baltagi et al. (2003) consideran un modelo de regresión de datos de panel que es un caso especial del modelo presentado anteriormente en que no incluye el retardo espacial de la variable dependiente. Elhorst (2003, 2009) define una taxonomía de modelos de datos de panel espaciales, asumiendo efectos fijos y efectos aleatorios. Siguiendo la distinción típica hecha en modelos de corte transversal, Elhorst (2003, 2009) define versiones de paneles de datos de errores espaciales y de modelos espaciales con rezagos, tanto para efectos fijos, así como también para efectos aleatorios . Sin embargo, él no considera un modelo que incluya tanto el retardo espacial de la variable dependiente y un término de error espacialmente autocorrelacionado. Por lo tanto, los modelo revisados en Elhorst (2003,2009) también pueden ser vistos como un caso especial de esta especificación más general.

---

<sup>1</sup>notar que la matriz de pesos espaciales en la ecuación de regresión y el término del error pueden diferir en muchas de sus implementaciones. Sin embargo en la discusión de los modelos se asume que son los mismos por simplicidad. Esto es asumiendo que  $I_N - \rho W_N$  es no singular donde  $I_N$  es una matriz identidad de dimensión  $N$

Una segunda especificación para las perturbaciones es considerado en Kapoor et al. (2007). Ellos asumen que la correlación espacial se aplica tanto a los efectos individuales y al resto de los componentes del error. Aunque los datos que generan los dos procesos parecen similares, ello implica mecanismos diferentes de contagio espacial que se rigen por una diferente estructura de la matriz de varianza covarianza implícita. En este caso, el término de las perturbaciones sigue un proceso autorregresivo espacial de primer orden de la forma:

$$u = \rho(I_T \otimes W_N)u + \varepsilon \quad (4.7)$$

Donde  $W_N$  es la matriz de pesos espaciales y  $\rho$  el correspondiente parámetro espacial autorregresivo. Para permitir además que las innovaciones se correlacionen con el tiempo, el vector de innovaciones en la Ecuación (4.8) sigue una estructura de componentes del error

$$\varepsilon = (\iota_T \otimes I_N)\mu + \nu \quad (4.8)$$

Donde  $\mu$  es el vector de efectos específicos de una sección transversal,  $\nu$  vector de innovaciones que varía tanto sobre las unidades de la sección transversal como por períodos de tiempo,  $\iota_T$  es un vector de unos e  $I_N$  una matriz identidad de  $N \times N$ . Al obtener una prueba de Hausman para un Modelo espacial de panel de datos Cliff y Ord, Mutl y Pfaffermay (2011) consideran el modelo presentado arriba y discuten la estimación de variables instrumentales bajo ambas especificaciones las de efectos fijos y las de efectos aleatorios. Se extienden en la obra de Kapoor et al. (2007) quienes no incluyen el rezago espacial de la variable dependiente en la ecuación de regresión. Bajo el supuesto de efectos aleatorios que los efectos individuales son independientes de las regresoras del modelo, se puede re-escribir la ecuación (4.9) como:

$$u = [I_T \otimes (I_N - \rho W_N)^{-1}]\varepsilon \quad (4.9)$$

De esto se sigue que la matriz de varianza-covarianza de  $u$  es:

$$\Omega_u = [I_T \otimes (I_N - \rho W_N)^{-1}]\Omega_\varepsilon[I_T \otimes (I_N - \rho W_N^T)^{-1}] \quad (4.10)$$

Donde:

$$\Omega_\varepsilon = \sigma_\nu^2 Q_0 + \sigma_1^2 Q_1, \quad \text{con}$$

$$\sigma_1^2 = \sigma_\nu^2 + T\sigma_\mu^2,$$

$$Q_0 = (I_T - \frac{J_T}{T}) \otimes I_N,$$

$$Q_1 = \frac{J_T}{T} \otimes I_N$$

$J_T = \iota_T \iota_T^T$ , es la matriz de varianza-covarianza típica de un modelo de componente de error de una vía adaptada a los diferentes ordenes de los datos.

Como debe quedar claro de la discusión anterior, estos dos modelos de paneles difieren en términos de sus matrices de varianza. La matriz de varianza en la Ecuación (4.6) es más complicado que el de la Ecuación (4.10), y, por lo tanto, su inverso es más difícil de calcular.

En el presente trabajo, se considera la implementación de ambas especificaciones de los términos del error. Para la primera especificación, se implementa la estimación de máxima verosimilitud para los modelos de efectos aleatorios, así como de efectos fijos. Para la segunda especificación, se implementa tanto máxima verosimilitud y la estimación de variables instrumentales bajo los supuestos de efectos aleatorios y de efectos fijos. La siguiente sección está dedicada a la discusión de la implementación de ML para los dos modelos.

## 4.4. Implementación ML

Ambos modelos de efectos fijos y aleatorios en  $R$  se implementan dentro de la misma estructura del software. `spml` es la función envolvente general y el argumento `model` controla la especificación. De acuerdo con la sintaxis `plm`, `model` toma el valor “*within*” para efectos fijos, “*random*” para efectos aleatorios, y “*pooling*” para el caso en que no se consideren ni efectos fijos ni efectos aleatorios. La estructura espacial se especifica mediante la combinación de los argumentos lógicos `lag` (que, si es verdadero (*TRUE*), añade un término autorregresivo espacial en la variable dependiente) y `spatial.error`. Este último argumento toma tres valores posibles: “*b*” (“Baltagi”) para la especificación en la ecuación (4.3), “*kkp*” (“Kapoor, Kelejian y Prucha”) para la especificación en la ecuación (4.7), y “*none*” para ninguna correlación espacial del error.

### 4.4.1. Modelo de Efectos Aleatorios

Para un modelo con componentes del error auto-correlacionados espacialmente, mínimos cuadrados ordinarios (MCO) es ineficiente incluso cuando  $\sigma_\mu^2 = 0$ . Análogamente, MCO en un modelo de efectos aleatorios (incluso sin componentes espaciales) es también ineficiente. Una manera alternativa (es decir, más eficiente) de la estimación del modelo es a través de máxima verosimilitud. En la presente sección se discute el enfoque de estimación de la especificación completa, es decir, en el que se tiene un retardo espacial, efectos aleatorios y correlación espacial de la forma especificada en la Ecuación (4.3).

Escalando la matriz de covarianza del error por la varianza del error idiosincrático  $\sigma_\varepsilon^2$ ,

y denotando

$$\phi = \sigma_\mu^2 / \sigma_\varepsilon^2, \bar{J}_T = J_T / T, E_T = I_T - \bar{J}_T \quad \text{y} \quad A_N = (I_N - \lambda W_N),$$

las expresiones para la matriz de covarianza escalada del error  $\Sigma$ ,

su inversa  $\Sigma^{-1}$  y su determinante  $|\Sigma|$  pueden ser escritos respectivamente como:

$$\Sigma = \phi(J_T \otimes I_N) + I_T \otimes (B^T B)^{-1}$$

$$\Sigma^{-1} = \bar{J}_T \otimes ((T\phi I_N + (B^T B)^{-1})^{-1} + E_T \otimes B^T B)$$

$$|\Sigma| = |T\phi I_N + (B^T B)^{-1}| |(B^T B)^{-1}|^{T-1}$$

Sustituyendo en la fórmula general dada en Anselin (1988 Cap. 6), se deriva la ecuación de verosimilitud:

$$L(\beta, \sigma_\varepsilon^2, \phi, \lambda, \rho) = -\frac{NT}{2} 2\pi - \frac{NT}{2} \ln \sigma_\nu^2 + T \ln |A| - \frac{1}{2} \ln |T\phi I_N + (B^T B)^{-1}| + (T-1) \ln |B| - \frac{1}{2\sigma_\nu^2} u^T \Sigma^{-1} u$$

Se implementa un proceso interactivo para obtener las estimaciones de máxima verosimilitud. Comenzando con valores iniciales para  $\lambda$ ,  $\rho$  y  $\phi$ , se obtienen los estimados para  $\beta$  y  $\sigma_\nu^2$  desde las condiciones de primer orden:

$$\beta = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} A y$$

$$\sigma_\nu^2 = (A y - X \beta)^T \Sigma^{-1} (A y - X \beta) / NT$$

La verosimilitud se puede concentrar y maximizar con respecto a  $\lambda$ ,  $\rho$ , y  $\phi$ . Los valores estimados de  $\lambda$ ,  $\rho$ , y  $\phi$  a su vez se utilizan para actualizar la expresión  $A$  y  $\Sigma^{-1}$ . Estos pasos se repiten hasta alcanzar el criterio de convergencia. En otras palabras, para un  $\Sigma$  específico la estimación puede ser operativa mediante un procedimiento iterativo de dos pasos que alterna entre mínimos cuadrados generalizados (GLS para  $\beta$  y  $\sigma_\nu^2$ ) y la verosimilitud concentrada (para los parámetros restantes) hasta la converjencia.<sup>2</sup> Desde el punto de vista de la implementación según algunos investigadores existen al menos un par de diferentes maneras de proceder. En primer lugar se ha decidido incluir el paso GLS con el objetivo de maximizar la función (es decir la función que se utiliza como argumento para el optimizador). En otras palabras el paso GLS es parte del proceso

<sup>2</sup>Tener en cuenta que estos pasos siguen siendo válidos cuando el modelo a ser estimado es una de las especificaciones anidadas donde por ejemplo uno de los coeficientes espaciales se restringe a cero.

de optimización de verosimilitud,<sup>3</sup> se obtienen errores estándar para  $\beta$  de GLS y empleando un Hessiano numérico para realizar inferencia estadística sobre los componentes del error.<sup>4</sup>

#### 4.4.2. Modelo de Efectos Fijos

Estimaciones consistentes de efectos fijos individuales, para  $N$  grandes no son posibles debido a problemas de parámetros incidentales. (Elhorst 2003) señala que cuando el interés está principalmente en el vector los parámetros  $\beta$  de la regresión una extensión del modelo de efectos fijos a un contexto espacial puede ser todavía apropiado, considerando modelos con rezagos y errores espaciales por separado, pero no da especificaciones para modelos que incluyan ambas autocorrelación espacial de los términos del error y un rezago espacial de la variable dependiente.

Un modelo de efectos fijos con rezagos espaciales puede ser escrito como:

$$y = \lambda(I_T \otimes W_N)y + (\iota_T \otimes I_N)\mu + X\beta + \varepsilon \quad (4.11)$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

Donde

$\lambda$  coeficiente autorregresivo espacial

$W_N$  matriz no estocástica de pesos espaciales

$\iota_T$  vector columna de unos de dimensión  $T$

$I_N$  matriz identidad de  $N \times N$

La teoría general de estimación por máxima verosimilitud se asemeja al caso de sección transversal. La presencia de rezago espacial introduce una forma de endogeneidad que viola el supuesto de los modelos de regresión estándar (es decir las regresoras no están correlacionadas con los términos del error). (Elhorst 2003) sugiere transformar las variables en la ecuación (4.11) eliminando el tiempo de los efectos individuales invariante y utilizar estas variables transformadas para la función de máxima verosimilitud. La transformación se obtiene restando el promedio para cada sección transversal a través del tiempo. como consecuencia, los efectos fijos y el término constante (así como las otras variables que no varían con el tiempo) son eliminados del modelo. Formalmente la transformación se puede escribir como:

<sup>3</sup>Existen muchos optimizadores disponibles en "R"

<sup>4</sup>El hessiano numérico es evaluado en los valores de los parámetros de ML usando diferencias finitas.

$$y^* = \lambda(I_T \otimes W_N)y^* + (\iota_T \otimes I_N)\mu + X^*\beta + \varepsilon^* \quad (4.12)$$

Donde  $y^* = Q_0y$ ,  $X^* = Q_0X$ ,  $\varepsilon^* = Q_0\varepsilon$ , y  $Q$  fueron definidos anteriormente. La función log-verosimilitud de la ecuación (4.11) será:

$$L = -\frac{NT}{2}\ln(2\pi\sigma_\varepsilon^2) + T\ln|I_N - \lambda W_N| - \frac{NT}{2\sigma_\varepsilon^2}e^T e \quad (4.13)$$

Donde  $e = y - \lambda(I_T \otimes W_N)y - X\beta$  y  $\ln|I_N - \lambda W_N|$  es el Jacobiano del determinante.<sup>5</sup> (Elhorst 2009) sugiere un enfoque de verosimilitud concentrada para maximización, ecuación (4.13). El procedimiento de estimación es sustancialmente análogo al empleado en casos de sección transversal. después de la transformación, dos regresiones auxiliares de  $y^*$  y  $(I_N \otimes W_N)y^*$  sobre  $X^*$  son realizadas. Los correspondientes residuos ( $e_0^*$ , y  $e_1^*$ ) son combinados para obtener la verosimilitud concentrada.

$$L = C + T\ln|I_N - \lambda W_N| - \frac{NT}{2}\ln[(e_0^* - \lambda e_1^*)^T(e_0^* - \lambda e_1^*)] \quad (4.14)$$

con una constante  $C$  no dependiente de  $\lambda$ . Un procedimiento de optimización numérica es necesario para obtener el valor de  $\lambda$  que maximice la ecuación (4.14). Finalmente los estimados de  $\beta$  y  $\sigma_\varepsilon^2$  son obtenidos desde las condiciones de primer orden de la función de verosimilitud reemplazando  $\lambda$  con su valor estimado por ML. Análogo al modelo de sección transversal, el estimador para  $\beta$  puede también ser visto como mínimos cuadrados generalizados de un modelo de regresión lineal con matriz de varianza de disturbios  $\sigma_\varepsilon^2 Q_0$ .<sup>6</sup> Inferencia estadística de los parámetros del modelo puede basarse en la expresión para la matriz de varianza-covarianza asintóticas derivadas en Elhorst (2009) y Elhorst and Freret (2009):

<sup>5</sup>A veces la verosimilitud está expresada en términos del log Jacobiano  $\sum_i \ln(1 - \lambda\omega_i)$  donde  $\omega_i$  son los valores propios de la matriz de pesos espaciales. El método de cálculo del Jacobiano por defecto en R, está basado en la descomposición de valores propios usando la función *eigenw*.

<sup>6</sup>Anselin et al. (2008) señalan que varios aspectos de los efectos fijos de modelos con rezago espacial merecen mayor investigación. El principal problema está relacionado con las propiedades de  $Q_0$ . Por definición  $Q_0$  es singular y por tanto  $|Q_0|$  no existe. Si bien esto no es un problema en casos no espaciales, el log verosimilitud para modelos espaciales puede estar basado en la normalidad multivariante del término del error. Por lo tanto debido a las propiedades de  $Q_0$  la probabilidad incondicional de la unión llega a degenerarse. Aunque teóricamente relevantes, estas consideraciones pueden no ser un problema en la práctica. Para hacer frente a esto, Lee y Yu (2010) sugieren usar una transformación diferente basada en la matriz ortonormal de  $Q_0$ .

$$AsyVar(\beta, \lambda, \sigma_\varepsilon^2) = \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2} X^{*T} X^* & \frac{1}{\sigma_\varepsilon^2} X^{*T} (I_T \otimes \widetilde{W}) X^* \beta \\ \frac{1}{\sigma_\varepsilon^2} \beta^T X^{*T} (I_T \otimes \widetilde{W}^T \widetilde{W}) X^* \beta + Ttr(\widetilde{W} \widetilde{W} + \widetilde{W}^T \widetilde{W}) & \frac{1}{\sigma_\varepsilon^2} tr(\widetilde{W}) \\ \frac{1}{\sigma_\varepsilon^2} tr(\widetilde{W}) & \frac{NT}{2\sigma_\varepsilon^4} \end{bmatrix}^{-1} \quad (4.15)$$

donde  $\widetilde{W} = W(I_N - \lambda W)^{-1}$  y los elementos que faltan que no pueden ser llenados por simetría son ceros. La carga computacional implicada en el cálculo del error estándar asintótico del parámetro espacial puede ser costoso para problemas de grandes dimensiones (principalmente debida a la inversa de la matriz de  $N \times N$  involucrada en el cálculo). El bloque de coeficientes de la matriz de covarianza relativa al vector  $\beta$  de parámetros no presenta dificultades de cálculo particulares. Los efectos fijos pueden ser recuperados por:

$$\mu_i = \frac{1}{T} \sum_{t=1}^T (y_{it} - \lambda \sum_{j=1}^N w_{ij} y_{jt} - x_{it} \beta) \quad (4.16)$$

Promediando en todas las observaciones también se puede recuperar el intercepto bajo la restricción de que los efectos individuales suman cero.

Un modelo de efectos fijos del error espacial se puede escribir como:

$$\begin{aligned} y &= (\iota_T \otimes I_N) \mu + X \beta + u \\ u &= \rho (I_T \otimes W_N) u + \varepsilon \end{aligned} \quad (4.17)$$

donde  $\rho$  es el coeficiente de autocorrelación espacial y  $\varepsilon$  es un término de error de comportamiento bueno

La estrategia de estimación para modelos de error espacial de sección transversal puede ser fácilmente extendido a contexto de panel. Una vez más un enfoque de verosimilitud concentrada se puede considerar pero se necesita un procedimiento iterativo para estimar los parámetros del modelo de error espacial. La idea genaral es iterar entre ML y mínimos cuadrados generalizados (GLS) hasta que se cumpla un criterio de convergencia. El modelo se transforma según la ecuación (4.12) para eliminar los efectos fijos. Más formalmente la función log-verosimilitud para el modelo de la ecuación (4.17) puede escribirse como:

$$L = -\frac{NT}{2} \ln(2\pi\sigma_\varepsilon^2) + T \ln |B_N| - \frac{1}{2\sigma_\varepsilon^2} e^T [I_T \otimes (B_n^T B_N)] e \quad (4.18)$$

con  $e = y - X\beta$  y  $B_N = (I_N - \rho W)$  dado  $\rho$  estimadores para  $\beta$  y  $\sigma_\varepsilon^2$  son derivados de las condiciones de primer orden como:

$$\beta = [X^T(I_T \otimes B_N^T B_N)X]^{-1} X^T(I_T \otimes B_N^T B_N)y \quad (4.19)$$

y

$$\sigma_\varepsilon^2 = \frac{e(\rho)^T e(\rho)}{NT} \quad (4.20)$$

Donde la notación indica la dependencia explícita de los residuos de  $\rho$ . Sustituyendo la ecuación (4.19) y la ecuación (4.20) en la ecuación (4.18), la función de log-verosimilitud concentrada puede derivarse como:

$$L = C - \frac{NT}{2} \ln[e(\rho)^T e(\rho)] + T \ln|B_N| \quad (4.21)$$

Donde  $C$  es una constante no dependiente de  $\rho$  y  $B_N$  está definida abajo. El proceso de estimación puede ser resumido como sigue. Estimación de residuos por MCO (del modelo transformado) se pueden usar para obtener una estimación inicial de  $\rho$ . el estimado inicial de  $\rho$  a su vez puede ser utilizado para calcular un estimador factible GLS (espacial) (FGLS) de los coeficientes de la regresión, la varianza del error y un nuevo conjunto de residuos estimados GLS. Un procedimiento iterativo puede ser empleado: la verosimilitud concentrada y los estimadores GLS son alternativamente calculados hasta que converjan. La matriz asintótica de varianza-covarianza de los parámetros es:

$$AsyVar(\beta, \lambda, \sigma_\varepsilon^2) = \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2} X^{*T} X^* & & \\ & T tr(\widetilde{W}\widetilde{W} + \widetilde{W}^T \widetilde{W}) & \\ & \frac{T}{\sigma_\varepsilon^2} tr(\widetilde{W}) & \frac{NT}{2\sigma_\varepsilon^4} \end{bmatrix}^{-1} \quad (4.22)$$

donde  $\widetilde{W} = W(I_N - \lambda W)^{-1}$ . Consideraciones hechas para el caso de rezagos espaciales se aplican también y los efectos individuales pueden ser recuperados por:

$$\mu_i = \frac{1}{T} \sum_{t=1}^T (y_{it} - x_{it}\beta) \quad (4.23)$$

## 4.5. Pruebas

A continuación se pone en práctica algunas pruebas de multiplicadores de Lagrange (ML) para los modelos de regresiones de datos de panel con errores espaciales correlacionados que son

consideradas en Baltagi *et al.* (2003). También se implementará un test de Hausman.

#### 4.5.1. Pruebas LM

Las pruebas por multiplicadores de Lagrange (LM) son ampliamente empleados para pruebas de efectos aleatorios y correlación serial o transversal en modelos de datos de panel. Requiriendo solamente la estimación de la especificación de la restricción. Las pruebas LM son particularmente utilizadas para efectos espaciales aleatorios debido a las dificultades de cálculo relacionados con la estimación del modelo completo.

Baltagi *et al.* (2003) derivan pruebas conjuntas, pruebas marginales y condicionales para todas las combinaciones de efectos aleatorios y correlación espacial para la especificación de los modelos presentados. En particular las pruebas que se consideran son:

1.  $H_0^a : \lambda = \sigma_\mu^2 = 0$  bajo la hipótesis alternativa de que al menos un componente no es cero
2.  $H_0^b : \sigma_\mu^2 = 0$  (asumiendo  $\lambda = 0$ ) bajo la alternativa de un sólo lado que el componente de la varianza es mayor que cero
3.  $H_0^c : \lambda = 0$  asumiendo efectos no aleatorios  $\sigma_\mu^2 = 0$  bajo la alternativa de dos lados de que el coeficiente de autocorrelación espacial es diferente de cero
4.  $H_0^d : \lambda = 0$  asumiendo la posible existencia de efectos aleatorios ( $\sigma_\mu^2$  puede o no ser cero) bajo la alternativa de dos lados de que el coeficiente de autocorrelación espacial es diferente de cero
5.  $H_0^e : \sigma_\mu^2 = 0$  asumiendo la posible existencia de autocorrelación espacial ( $\lambda$  puede o no ser cero) y la alternativa de un lado de que el componente de la varianza es mayor que cero

La prueba LM conjunta para la primera hipótesis de efecto no aleatorios y no autocorrelación espacial ( $H_0^a$ ) está dada por:

$$LM_j = \frac{NT}{2(T-1)}G^2 + \frac{N^2T}{b}H^2 \quad (4.24)$$

donde,

$$G = \tilde{\mu}^T (J_T \otimes I_N) \tilde{\mu} / \tilde{\mu}^T \tilde{\mu} - 1,$$

$$H = \tilde{\mu}^T (I_T \otimes (W + W^T)/2) \tilde{\mu} / \tilde{\mu}^T \tilde{\mu},$$

$$b = \text{tr}(W + W^T)^2/2, \text{ y } \mu \text{ denota residuos MCO.}$$

La ecuación (4.24) es también el punto de partida para la obtención de las pruebas LM marginales usadas para verificar  $H_0^b$  y  $H_0^c$ . La versión estandarizada de la prueba LM marginal de efectos no aleatorios asumiendo ninguna correlación espacial está dada por:

$$SLM_1 = \frac{LM_1 - E(LM_1)}{\sqrt{Var(LM_1)}} \quad (4.25)$$

donde,  $LM_1$  es la raíz cuadrada del primer término en la ecuación (4.24). Análogamente la versión estandarizada de la prueba LM marginal de autocorrelación no espacial suponiendo que no hay efectos aleatorios está dada por:

$$SLM_2 = \frac{LM_2 - E(LM_2)}{\sqrt{Var(LM_2)}} \quad (4.26)$$

donde,  $LM_2$  es la ahora la raíz cuadrada del segundo término en la ecuación (4.24). Notar que ambas ecuaciones (4.25) y (4.26) deben ser asintóticamente distribuidas normalmente con  $N \rightarrow \infty$  (para  $T$  fijo) bajo  $H_0^b$  y  $H_0^c$  respectivamente. Basados en las ecuaciones (4.25) y (4.26), son usadas para pruebas estadísticas de un lado para  $H_0^a : \lambda = \sigma_\mu^2 = 0$  puede obtenerse como:

$$LM_H = (LM_1 + LM_2)/\sqrt{2} \quad (4.27)$$

que se distribuye asintóticamente  $N(0, 1)$ . Una prueba para la hipótesis nula conjunta puede por lo tanto estar basada en la siguiente regla de decisión:

$$\chi_m^2 = \begin{cases} LM_1^2 + LM_2^2 & \text{if } LM_1 > 0, LM_2 > 0 \\ LM_1^2 & \text{if } LM_1 > 0, LM_2 \leq 0 \\ LM_2^2 & \text{if } LM_1 \leq 0, LM_2 > 0 \\ 0 & \text{if } LM_1 \leq 0, LM_2 \leq 0 \end{cases}$$

bajo la hipótesis nula del estadístico de prueba  $\chi_m^2$  como una distribución mixta  $\chi^2$  dada por:

$$\chi_m^2 = (1/4)\chi^2(0) + (1/2)\chi^2(1) + (1/4)\chi^2(2) \quad (4.28)$$

Al usar  $LM_2$  es uno asumiendo que no existen efectos aleatorios individuales. Sin embargo especialmente cuando el componente de la varianza es grande, esto puede llevar a inferencias incorrectas. Por esta razón que Baltagi *et al.* (2003) deriva una prueba condicional LM contra el coeficiente de autocorrelación espacial siendo cero suponiendo que el componente de la varianza puede o no puede ser cero. La expresión de esta prueba asume la siguiente forma:

$$LM_\lambda = \frac{\widehat{D}(\lambda)^2}{[(T-1) + \hat{\sigma}_\nu^4/\hat{\sigma}_1^4]b} \quad (4.29)$$

donde,

$$\widehat{D}(\lambda)^2 = \frac{1}{2}\hat{u}^T\left[\frac{\hat{\sigma}_\nu^4}{\hat{\sigma}_1^4}\right](\bar{J}_T \otimes (W^T + W)) + \frac{1}{\hat{\sigma}_\nu^4}(E_T \otimes (W^T + W))\hat{u}, \text{ también}$$

$$\sigma_1^4 = \hat{u}^T(\bar{J}_T \otimes I_N)\hat{u}/N,$$

$$\sigma_\nu^4 = \hat{u}^T(E_T \otimes I_N)\hat{u}/N(T-1)$$

y contrario a las pruebas anteriores que utilizan residuos MCO los residuos  $\hat{u}$  son ML. La desventaja comparativa de esta última prueba es que su aplicación es ligeramnete más complicada, ya que se basa en los residuos ML. Una prueba unilateral se obtiene simplemente tomando la raíz cuadrada de la ecuación (4.29). El estadístico de prueba resultante debe estar distribuido asintóticamente  $N(0, 1)$ . Del mismo modo al utilizar  $LM_1$  es uno suponiendo que no hay correlación del error espacial. Esta suposición puede conducir a inferencias incorrectas sobre todo cuando  $\lambda$  no está muy cerca de cero. Una prueba LM condicional asumiendo la posible existencia de la correlación espacial del error puede ser derivada como:

$$LM_\mu = (\widehat{D}_\mu)^2 \left( \frac{2\hat{\sigma}_\nu^4}{T} \right) (TN\hat{\sigma}_\nu^4 ec - N\hat{\sigma}_\nu^4 d^2 - T\hat{\sigma}_\nu^4 g^2 e + 2\hat{\sigma}_\nu^4 gh d - \hat{\sigma}_\nu^4 h^2 c)^{-1} \times (N\hat{\sigma}_\nu^4 c - \hat{\sigma}_\nu^4 g^2) \quad (4.30)$$

donde,

$$g = \text{tr}[(W^T \hat{B} + \hat{B}^T W)(\hat{B}^T \hat{B})^{-1}]$$

$$h = \text{tr}[(\hat{B}^T \hat{B})]$$

$$d = \text{tr}[(W^T \hat{B} + \hat{B}^T W)]$$

$$c = \text{tr}[(W^T \hat{B} + \hat{B}^T W)(\hat{B}^T \hat{B})^{-1}]^2 \quad \text{y}$$

$$e = \text{tr}[(\hat{B}^T \hat{B})^2]$$

una prueba unilateral se puede definir sacando la raíz cuadrada de la ecuación (4.30) basada en residuos ML. El estadístico de prueba debe ser distribuido asintóticamente  $N(0, 1)$

### 4.5.2. Prueba espacial de Hausman

La prueba de Hausman (Hausman 1978) compara los efectos estimados aleatorios y fijos y prueba si el supuesto de que los efectos aleatorios son soportados o no por los datos. Mutl y Pfaffermayr (2011) muestran como extender este procedimiento al marco espacial. El estadístico de prueba de Hausman toma la forma:

$$H = NT(\hat{\theta}_{FGLS} - \hat{\theta}_W)^T (\hat{\Sigma}_W - \hat{\Sigma}_{FGLS})^{-1} (\hat{\theta}_{FGLS} - \hat{\theta}_W) \quad (4.31)$$

donde,  $\hat{\theta}_{FGLS}$  y  $\hat{\theta}_W$  son la regresión espacial GLS y los estimadores *within* respectivamente,  $\hat{\Sigma}_W$  y  $\hat{\Sigma}_{FGLS}$  las estimaciones correspondientes de los coeficientes de las matrices de varianza covarianza.  $H$  está asintóticamente distribuida  $\chi^2$  con  $k$  grados de libertad, donde  $k$  es el número de regresoras en el modelo.

## 4.6. Análisis de Modelos Espaciales para Datos de Panel

### 4.6.1. Modelo Pooling (Modelo1)

```
> modelo1 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="pooling", lag=TRUE, spatial.error="none")
> summary(modelo1)
Spatial panel random effects ML model
```

Call:

```
sprem1(formula = formula, data = data, index = index, w = listw2mat
(listw), w2 = listw2mat(listw2), lag = lag, errors = errors, cl = cl)
```

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.770	-0.316	0.504	0.474	1.290	4.980

Spatial autoregressive coefficient:

	Estimate	Std. Error	t-value	Pr(> t )
lambda	0.567000	0.027141	20.891	< 2.2e-16 ***

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	808.75531	124.83841	6.4784	9.269e-11 ***
lat131	-2.19158	0.64817	-3.3812	0.0007218 ***
lon131	10.30252	1.59139	6.4739	9.550e-11 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### 4.6.2. Modelo Pooling (Modelo2)

```
> modelo2 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="pooling", lag=TRUE, spatial.error="b")
> summary(modelo2)
Spatial panel random effects ML model
```

```

Call:
spreml(formula = formula, data = data, index = index, w = listw2mat
(listw), w2 = listw2mat(listw2), lag = lag, errors = errors, cl = cl)

Residuals:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-4.150 -1.060  -0.181  -0.259   0.622   4.200

Error variance parameters:
  Estimate Std. Error t-value Pr(>|t|)
rho 0.811401  0.027219  29.81 < 2.2e-16 ***

Spatial autoregressive coefficient:
  Estimate Std. Error t-value Pr(>|t|)
lambda -0.48766  0.07251 -6.7254 1.751e-11 ***

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept) 1696.5429   272.4234  6.2276 4.736e-10 ***
lat131      -2.0274     1.4955 -1.3557  0.1752
lon131      21.5998     3.4728  6.2197 4.981e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

### 4.6.3. Modelo Pooling (Modelo3)

```

> modelo3 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="pooling", lag=TRUE, spatial.error="kkp")
> summary(modelo3)
Spatial panel random effects ML model

Call:
spreml(formula = formula, data = data, index = index, w = listw2mat
(listw), w2 = listw2mat(listw2), lag = lag, errors = errors, cl = cl)

Residuals:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-4.150 -1.060  -0.181  -0.259   0.622   4.200

Error variance parameters:
  Estimate Std. Error t-value Pr(>|t|)
rho 0.811401  0.027219  29.81 < 2.2e-16 ***

Spatial autoregressive coefficient:
  Estimate Std. Error t-value Pr(>|t|)
lambda -0.48766  0.07251 -6.7254 1.751e-11 ***

```

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )	
(Intercept)	1696.5429	272.4234	6.2276	4.736e-10	***
lat131	-2.0274	1.4955	-1.3557	0.1752	
lon131	21.5998	3.4728	6.2197	4.981e-10	***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 4.6.4. Modelo Within (Modelo4)

```
> modelo4 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="within", lag=TRUE, spatial.error="none")
> summary(modelo4)
Spatial panel fixed effects lag model
```

Call:

```
spml(formula = ltot131 ~ lat131 + lon131, data = datos_panel,
      listw = lw, model = "within", lag = TRUE, spatial.error = "none")
```

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-1.5100	-0.2680	0.0052	0.2580	1.7500

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )	
lambda	0.452180	0.030443	14.8534	<2e-16	***
lat131	-0.270898	14.780606	-0.0183	0.9854	
lon131	-6.476952	16.175448	-0.4004	0.6888	

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 4.6.5. Modelo Within (Modelo5)

```
> modelo5 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="within", lag=TRUE, spatial.error="b")
> summary(modelo5)
Spatial panel fixed effects sarar model
```

Call:

```
spml(formula = ltot131 ~ lat131 + lon131, data = datos_panel,
      listw = lw, model = "within", lag = TRUE, spatial.error = "b")
```

Residuals:

	Min.	1st Qu.	Median	3rd Qu.	Max.
	-2.11000	-0.40700	-0.00137	0.35900	2.32000

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )	
lambda	-0.710012	0.066653	-10.6524	<2e-16	***
rho	0.796465	0.025083	31.7527	<2e-16	***
lat131	2.617562	11.831852	0.2212	0.8249	
lon131	-11.988451	13.466644	-0.8902	0.3733	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 4.6.6. Modelo Within (Modelo6)

```
> modelo6 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="within", lag=TRUE, spatial.error="kkp")
```

```
> summary(modelo6)
```

Spatial panel fixed effects sarar model

Call:

```
spml(formula = ltot131 ~ lat131 + lon131, data = datos_panel,
      listw = lw, model = "within", lag = TRUE, spatial.error = "kkp")
```

Residuals:

	Min.	1st Qu.	Median	3rd Qu.	Max.
	-2.11000	-0.40700	-0.00137	0.35900	2.32000

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )	
lambda	-0.710012	0.066653	-10.6524	<2e-16	***
rho	0.796465	0.025083	31.7527	<2e-16	***
lat131	2.617562	11.831852	0.2212	0.8249	
lon131	-11.988451	13.466644	-0.8902	0.3733	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 4.6.7. Modelo Random (Modelo7)

```
> modelo7 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="random", lag=TRUE, spatial.error="none")
```

```
> summary(modelo7)
```

Spatial panel random effects ML model

Call:

```
spreml(formula = formula, data = data, index = index, w = listw2mat
(listw), w2 = listw2mat(listw2), lag = lag, errors = errors, cl = cl)
```

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.850	-0.385	0.458	0.420	1.230	4.920

Error variance parameters:

	Estimate	Std. Error	t-value	Pr(> t )
phi	3.42495	0.30875	11.093	< 2.2e-16 ***

Spatial autoregressive coefficient:

	Estimate	Std. Error	t-value	Pr(> t )
lambda	0.501590	0.029195	17.181	< 2.2e-16 ***

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	876.4675	199.8998	4.3845	1.162e-05 ***
lat131	-2.3614	1.0404	-2.2697	0.02323 *
lon131	11.1647	2.5482	4.3813	1.180e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 4.6.8. Modelo Random (Modelo8)

```
> modelo8 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="random", lag=TRUE, spatial.error="b")
Error in if (e1 == 0) return(spam(0, nrow(e2), ncol(e2))) else { :
  missing value where TRUE/FALSE needed
> summary(modelo8)
Error in summary(modelo8) :
  error in evaluating the argument 'object' in selecting a method for
  function 'summary': Error: object 'modelo8' not found
```

#### 4.6.9. Modelo Random (Modelo9)

```
> modelo9 = spml(ltot131~lat131+lon131, data=datos_panel, listw = lw,
+ model="random", lag=TRUE, spatial.error="kcp")
> summary(modelo9)
```

Spatial panel random effects ML model

Call:

```
spreml(formula = formula, data = data, index = index, w = listw2mat
(listw), w2 = listw2mat(listw2), lag = lag, errors = errors, cl = cl)
```

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-4.170	-1.140	-0.309	-0.379	0.473	4.090

Error variance parameters:

	Estimate	Std. Error	t-value	Pr(> t )
phi	3.43967	0.31217	11.019	< 2.2e-16 ***
rho	0.80276	0.02633	30.488	< 2.2e-16 ***

Spatial autoregressive coefficient:

	Estimate	Std. Error	t-value	Pr(> t )
lambda	-0.620082	0.070531	-8.7916	< 2.2e-16 ***

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	1468.3297	408.1724	3.5973	0.0003215 ***
lat131	-1.2362	2.2739	-0.5437	0.5866724
lon131	18.6895	5.2033	3.5919	0.0003283 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 4.6.10. Modelo Pooling (Modelo plm)

```
modelop = plm(ltot131~lag(ltot131,listw = lw)+lat131+lon131,
data=datos_panel, model="pooling", lag=TRUE)
```

```
> summary(modelop)
```

Oneway (individual) effect Pooling Model

Call:

```
plm(formula = ltot131 ~ lag(ltot131, listw = lw) + lat131 + lon131,
data = datos_panel, model = "pooling", lag = TRUE)
```

Balanced Panel: n=448, T=2, N=896

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-2.9000	-0.4920	0.0813	0.5120	2.1500

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	100.025535	108.326042	0.9234	0.3561

```
lag(ltot131, listw = lw)    0.750483    0.018519  40.5248    <2e-16 ***
lat131                    0.846500    0.547159   1.5471    0.1222
lon131                    1.269160    1.380826   0.9191    0.3583
---
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
Total Sum of Squares:    1526.3
Residual Sum of Squares: 491.79
R-Squared                : 0.67778
  Adj. R-Squared        : 0.67476
F-statistic: 625.442 on 3 and 892 DF, p-value: < 2.22e-16
```

En esta regresión (Modelo Pooling) no se considera efectos fijos ni aleatorios y se puede observar que en este caso es significativo el rezago de la variable dependiente (ltot131), por tanto se entiende que existe autocorrelación, entre la variable dependiente y su rezago.

#### 4.6.11. Modelo Within (Modelo plm)

```
> modelow = plm(ltot131~slag(ltot131,listw = lw)+lat131+lon131,
data=datos_panel, model="within", lag=TRUE)
```

```
> summary(modelow)
```

Oneway (individual) effect Within Model

Call:

```
plm(formula = ltot131 ~ slag(ltot131, listw = lw) + lat131 +
     lon131, data = datos_panel, model = "within", lag = TRUE)
```

Balanced Panel: n=448, T=3, N=1344

Residuals :

```
  Min. 1st Qu.  Median 3rd Qu.    Max.
-1.6600 -0.2620  0.0112  0.2590  1.7700
```

Coefficients :

```
                Estimate Std. Error t-value Pr(>|t|)
slag(ltot131, listw = lw) 0.670712   0.047047 14.2563  <2e-16 ***
lat131                   0.518868  17.905159  0.0290  0.9769
lon131                   -7.440598  19.595079 -0.3797  0.7042
---
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
Total Sum of Squares:    301.5
Residual Sum of Squares: 245.59
R-Squared                : 0.18545
  Adj. R-Squared        : 0.12322
F-statistic: 67.7714 on 3 and 893 DF, p-value: < 2.22e-16
```

En esta regresión (Modelo Within) solamente con efectos fijos se puede observar que en este caso también es significativo el rezago espacial de la variable dependiente (ltot131), por tanto se

entiende que existe autocorrelación, entre la variable dependiente y su rezago espacial.

#### 4.6.12. Modelo Random (Modelo plm)

```
modelor = plm(ltot131~slag(ltot131,listw = lw)+lat131+lon131,
data=datos_panel, model="random", lag=TRUE)
```

```
> summary(modelor)
```

```
Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)
```

Call:

```
plm(formula = ltot131 ~ slag(ltot131, listw = lw) + lat131 +
lon131, data = datos_panel, model = "random", lag = TRUE)
```

Balanced Panel: n=448, T=3, N=1344

Effects:

	var	std.dev	share
idiosyncratic	0.2750	0.5244	0.236
individual	0.8915	0.9442	0.764
theta:	0.6947		

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-1.9700	-0.3290	0.0597	0.3740	1.5000

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t )	
(Intercept)	561.049910	200.262324	2.8016	0.005159	**
slag(ltot131, listw = lw)	0.735801	0.036594	20.1072	< 2.2e-16	***
lat131	-1.456215	1.019839	-1.4279	0.153558	
lon131	7.147429	2.552728	2.7999	0.005185	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 501.06

Residual Sum of Squares: 369.16

R-Squared : 0.26325

Adj. R-Squared : 0.26246

F-statistic: 159.598 on 3 and 1340 DF, p-value: < 2.22e-16

En esta regresión (Modelo Random) solamente con efectos aleatorios se puede observar que en este caso son significativos el rezago espacial de la variable dependiente (ltot131) y la variable independiente longitud (lon131), por tanto se entiende que existe autocorrelación, entre la variable dependiente, su rezago espacial y la variable independiente longitud.

#### 4.6.13. Pruebas Sobre los Modelos

En la siguiente prueba se comparan los modelos con efectos fijos

y el modelo sin efectos fijos ni aleatorios, el reporte indica un p-value = NA. Sin embargo el resultado indica que la hipótesis alternativa (modelo de efectos fijos) es significativo.

```
pFtest(modelow,modelop)
F test for individual effects
```

```
data:  ltot131 ~ slag(ltot131, listw = lw) + lat131 + lon131
F = -895.2449, df1 = -1, df2 = 893, p-value = NA
alternative hypothesis: significant effects
```

```
phtest(modelow,modelor) # Prueba de Hausman (para efectos aleatorios):
# Ho: El modelo de efectos aleatorios es mejor que el de efectos fijos
Hausman Test
```

```
data:  ltot131 ~ slag(ltot131, listw = lw) + lat131 + lon131
chisq = 5.3364, df = 3, p-value = 0.1488
alternative hypothesis: one model is inconsistent
```

```
plmtest(modelop, type=c("bp")) # Prueba de Breusch-Pagan Lagrange
Multiplier (LM) para efectos aleatorios Ho: No hay diferencia
significativa entre las varianza de las unidades.
Esto es, no hay efecto de panel effect; MCO es mejor que
efectos aleatorios.
```

```
De acuerdo al reporte se rechaza Ho porque p-value es 0.1488 entonces
el modelo de efectos aleatorios seria el adecuado
Lagrange Multiplier Test - (Breusch-Pagan)
```

```
data:  ltot131 ~ lag(ltot131, listw = lw) + lat131 + lon131
chisq = 84.3389, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects
```

```
effects.$\textbf{spml}\$(modelo4)
```

```
Intercept:
```

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	-508.18	1268.60	-0.4006	0.6887

```
Spatial fixed effects:
```

	Estimate	Std. Error	t-value	Pr(> t )
1	-1.6476e+00	1.2689e+03	-0.0013	0.9990
2	-2.3327e+00	1.2688e+03	-0.0018	0.9985
3	-1.7175e+00	1.2687e+03	-0.0014	0.9989
4	2.4356e-01	1.2687e+03	0.0002	0.9998
5	2.0920e-01	1.2686e+03	0.0002	0.9999
6	1.1126e+00	1.2686e+03	0.0009	0.9993
7	-4.4416e-02	1.2687e+03	0.0000	1.0000

8	8.0852e-01	1.2687e+03	0.0006	0.9995
9	-1.1277e+00	1.2688e+03	-0.0009	0.9993
10	-1.2667e+00	1.2688e+03	-0.0010	0.9992
.	.	.	.	.
100	5.3214e-01	1.2682e+03	0.0004	0.9997
101	6.7578e-01	1.2682e+03	0.0005	0.9996
102	1.2327e+00	1.2681e+03	0.0010	0.9992
103	7.5441e-01	1.2680e+03	0.0006	0.9995
104	1.8789e-01	1.2681e+03	0.0001	0.9999
105	4.8760e-01	1.2681e+03	0.0004	0.9997
106	1.2758e+00	1.2682e+03	0.0010	0.9992
107	1.2645e+00	1.2681e+03	0.0010	0.9992
108	5.5100e-01	1.2681e+03	0.0004	0.9997
109	8.2616e-01	1.2682e+03	0.0007	0.9995
110	1.1496e+00	1.2681e+03	0.0009	0.9993
.	.	.	.	.
200	-8.1191e-01	1.2682e+03	-0.0006	0.9995
201	-5.6726e-01	1.2682e+03	-0.0004	0.9996
202	-7.6502e-01	1.2680e+03	-0.0006	0.9995
203	7.0943e-01	1.2681e+03	0.0006	0.9996
204	-1.3601e-01	1.2682e+03	-0.0001	0.9999
205	1.8461e+00	1.2680e+03	0.0015	0.9988
206	5.2725e-01	1.2680e+03	0.0004	0.9997
207	-5.7319e-01	1.2680e+03	-0.0005	0.9996
208	-1.4848e+00	1.2681e+03	-0.0012	0.9991
209	-6.4983e-01	1.2681e+03	-0.0005	0.9996
210	-1.3443e+00	1.2681e+03	-0.0011	0.9992
.	.	.	.	.
300	-8.5957e-01	1.2685e+03	-0.0007	0.9995
301	-9.0717e-02	1.2685e+03	-0.0001	0.9999
302	8.8243e-01	1.2686e+03	0.0007	0.9994
303	2.2672e-01	1.2687e+03	0.0002	0.9999
304	-9.8172e-01	1.2686e+03	-0.0008	0.9994
305	-4.8366e-01	1.2685e+03	-0.0004	0.9997
306	-2.4897e-01	1.2685e+03	-0.0002	0.9998
307	1.1787e+00	1.2685e+03	0.0009	0.9993
308	-5.8570e-01	1.2686e+03	-0.0005	0.9996
309	-9.5477e-01	1.2686e+03	-0.0008	0.9994
310	1.2998e-01	1.2686e+03	0.0001	0.9999
.	.	.	.	.
430	-5.6336e-01	1.2690e+03	-0.0004	0.9996
431	-1.0880e+00	1.2690e+03	-0.0009	0.9993
432	-7.7450e-01	1.2691e+03	-0.0006	0.9995
433	-1.1143e+00	1.2692e+03	-0.0009	0.9993
434	-2.1079e+00	1.2693e+03	-0.0017	0.9987
435	-1.8356e-01	1.2692e+03	-0.0001	0.9999

436	-4.4611e-01	1.2690e+03	-0.0004	0.9997
437	-6.3477e-01	1.2690e+03	-0.0005	0.9996
438	1.5461e+00	1.2691e+03	0.0012	0.9990
439	8.5013e-01	1.2690e+03	0.0007	0.9995
440	-1.1564e+00	1.2692e+03	-0.0009	0.9993
441	-1.0438e+00	1.2691e+03	-0.0008	0.9993
442	-8.5640e-01	1.2691e+03	-0.0007	0.9995
443	-1.2318e+00	1.2690e+03	-0.0010	0.9992
444	-6.8765e-02	1.2689e+03	-0.0001	1.0000
445	-1.3682e+00	1.2689e+03	-0.0011	0.9991
446	-1.3499e+00	1.2690e+03	-0.0011	0.9992
447	-1.5242e+00	1.2690e+03	-0.0012	0.9990
448	-1.8499e+00	1.2691e+03	-0.0015	0.9988

Como puede comprobarse en la tabla anterior ninguno de los coeficientes es significativo entonces el modelo de efectos fijos no es el adecuado, por tanto se puede pensar en utilizar los modelos de MCO o de efectos aleatorios.

Se procede a continuación con la misma prueba pero aplicándola al **modelo5**: con rezago espacial , efectos fijos, tipo de error espacial (b) "Baltagi"

```
> effects.$\textbf{spml}$ (modelo5)
```

Intercept:

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	-939.32	1114.77	-0.8426	0.3994

Spatial fixed effects:

	Estimate	Std. Error	t-value	Pr(> t )
1	-3.3052e+00	1.1150e+03	-0.0030	0.9976
2	-3.6081e+00	1.1150e+03	-0.0032	0.9974
3	-2.5532e+00	1.1149e+03	-0.0023	0.9982
4	2.8271e-01	1.1148e+03	0.0003	0.9998
5	2.2707e-01	1.1148e+03	0.0002	0.9998
6	1.2728e+00	1.1148e+03	0.0011	0.9991
7	-2.2076e+00	1.1148e+03	-0.0020	0.9984
8	-5.4044e-01	1.1149e+03	-0.0005	0.9996
9	-2.5547e+00	1.1150e+03	-0.0023	0.9982
10	-3.6835e+00	1.1150e+03	-0.0033	0.9974
.				
.				
110	9.2708e-01	1.1143e+03	0.0008	0.9993
111	6.3261e-01	1.1144e+03	0.0006	0.9995
112	6.4239e-01	1.1143e+03	0.0006	0.9995
113	-9.1514e-01	1.1142e+03	-0.0008	0.9993
114	-2.5304e+00	1.1142e+03	-0.0023	0.9982

115	-9.7131e-01	1.1142e+03	-0.0009	0.9993
116	-5.4402e-01	1.1141e+03	-0.0005	0.9996
117	1.3926e+00	1.1141e+03	0.0012	0.9990
118	1.6208e-01	1.1142e+03	0.0001	0.9999
119	9.8808e-01	1.1142e+03	0.0009	0.9993
120	6.1104e-01	1.1142e+03	0.0005	0.9996
.				
.				
200	-1.1594e+00	1.1144e+03	-0.0010	0.9992
201	-1.3471e+00	1.1144e+03	-0.0012	0.9990
202	-3.8625e-01	1.1143e+03	-0.0003	0.9997
203	1.9628e-01	1.1143e+03	0.0002	0.9999
204	-3.0265e-01	1.1144e+03	-0.0003	0.9998
205	9.8578e-01	1.1142e+03	0.0009	0.9993
206	2.5366e-01	1.1143e+03	0.0002	0.9998
207	-5.4369e-01	1.1143e+03	-0.0005	0.9996
208	-1.2691e+00	1.1143e+03	-0.0011	0.9991
209	-5.1510e-01	1.1143e+03	-0.0005	0.9996
210	-1.6556e+00	1.1143e+03	-0.0015	0.9988
.				
.				
300	-8.1470e-01	1.1147e+03	-0.0007	0.9994
301	-2.9675e-01	1.1147e+03	-0.0003	0.9998
302	4.9598e-01	1.1148e+03	0.0004	0.9996
303	3.3770e-03	1.1148e+03	0.0000	1.0000
304	-1.1020e+00	1.1148e+03	-0.0010	0.9992
305	-4.6263e-01	1.1147e+03	-0.0004	0.9997
306	-3.2049e-01	1.1147e+03	-0.0003	0.9998
307	7.4883e-01	1.1147e+03	0.0007	0.9995
308	-4.8261e-01	1.1147e+03	-0.0004	0.9997
309	-1.1593e+00	1.1148e+03	-0.0010	0.9992
310	-4.9030e-01	1.1148e+03	-0.0004	0.9996
.				
.				
420	-2.3464e+00	1.1149e+03	-0.0021	0.9983
421	-1.8138e+00	1.1149e+03	-0.0016	0.9987
422	-2.3817e-01	1.1149e+03	-0.0002	0.9998
423	-7.5615e-01	1.1149e+03	-0.0007	0.9995
424	-1.7922e-03	1.1150e+03	0.0000	1.0000
425	-8.1713e-01	1.1151e+03	-0.0007	0.9994
426	-1.3911e+00	1.1149e+03	-0.0012	0.9990
427	-8.0697e-01	1.1150e+03	-0.0007	0.9994
428	-2.2134e+00	1.1151e+03	-0.0020	0.9984
429	-3.2371e-01	1.1151e+03	-0.0003	0.9998
430	-9.3966e-01	1.1151e+03	-0.0008	0.9993
431	-2.2133e+00	1.1152e+03	-0.0020	0.9984
432	-1.0709e+00	1.1152e+03	-0.0010	0.9992
433	-2.7370e+00	1.1153e+03	-0.0025	0.9980
434	-2.4052e+00	1.1154e+03	-0.0022	0.9983
435	-1.3452e+00	1.1153e+03	-0.0012	0.9990
436	-7.1549e-01	1.1151e+03	-0.0006	0.9995

437	-6.5201e-01	1.1151e+03	-0.0006	0.9995
438	1.2110e+00	1.1152e+03	0.0011	0.9991
439	5.8671e-01	1.1151e+03	0.0005	0.9996
440	-1.7898e+00	1.1153e+03	-0.0016	0.9987
441	-3.9024e-01	1.1152e+03	-0.0003	0.9997
442	3.0291e-01	1.1152e+03	0.0003	0.9998
443	1.4903e-03	1.1151e+03	0.0000	1.0000
444	-3.3441e-01	1.1150e+03	-0.0003	0.9998
445	-1.9373e-01	1.1151e+03	-0.0002	0.9999
446	-7.1638e-03	1.1151e+03	0.0000	1.0000
447	-7.8760e-01	1.1152e+03	-0.0007	0.9994
448	-1.3647e+00	1.1152e+03	-0.0012	0.9990

La tabla anterior permite observar que ninguno de los coeficientes es significativo concluyéndose entonces que el modelo pooling (Sin efectos fijos ni aleatorios) es mejor que el modelo de efectos fijos (tipo de error espacial(b) "Baltagi")

Ahora se procede a realizar la misma prueba pero con el **modelo6**: con rezago espacial , efectos fijos, tipo de error espacial (kkp) "Kapoor,Kelejian,Prucha"

```
effects.$\textbf{spml}$ (modelo6)
```

Intercept:

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	-939.32	1114.77	-0.8426	0.3994

Spatial fixed effects:

	Estimate	Std. Error	t-value	Pr(> t )
1	-3.3052e+00	1.1150e+03	-0.0030	0.9976
2	-3.6081e+00	1.1150e+03	-0.0032	0.9974
3	-2.5532e+00	1.1149e+03	-0.0023	0.9982
4	2.8271e-01	1.1148e+03	0.0003	0.9998
5	2.2707e-01	1.1148e+03	0.0002	0.9998
6	1.2728e+00	1.1148e+03	0.0011	0.9991
7	-2.2076e+00	1.1148e+03	-0.0020	0.9984
8	-5.4044e-01	1.1149e+03	-0.0005	0.9996
9	-2.5547e+00	1.1150e+03	-0.0023	0.9982
10	-3.6835e+00	1.1150e+03	-0.0033	0.9974
.				
.				
100	7.3960e-02	1.1144e+03	0.0001	0.9999
101	9.6063e-01	1.1144e+03	0.0009	0.9993
102	1.7754e+00	1.1143e+03	0.0016	0.9987
103	1.5275e+00	1.1142e+03	0.0014	0.9989
104	7.5205e-01	1.1143e+03	0.0007	0.9995

105	5.9072e-01	1.1144e+03	0.0005	0.9996
106	1.6940e+00	1.1144e+03	0.0015	0.9988
107	2.0140e+00	1.1144e+03	0.0018	0.9986
108	6.1460e-01	1.1143e+03	0.0006	0.9996
109	4.1123e-01	1.1144e+03	0.0004	0.9997
110	9.2708e-01	1.1143e+03	0.0008	0.9993
.				
.				
200	-1.1594e+00	1.1144e+03	-0.0010	0.9992
201	-1.3471e+00	1.1144e+03	-0.0012	0.9990
202	-3.8625e-01	1.1143e+03	-0.0003	0.9997
203	1.9628e-01	1.1143e+03	0.0002	0.9999
204	-3.0265e-01	1.1144e+03	-0.0003	0.9998
205	9.8578e-01	1.1142e+03	0.0009	0.9993
206	2.5366e-01	1.1143e+03	0.0002	0.9998
207	-5.4369e-01	1.1143e+03	-0.0005	0.9996
208	-1.2691e+00	1.1143e+03	-0.0011	0.9991
209	-5.1510e-01	1.1143e+03	-0.0005	0.9996
210	-1.6556e+00	1.1143e+03	-0.0015	0.9988
.				
.				
300	-8.1470e-01	1.1147e+03	-0.0007	0.9994
301	-2.9675e-01	1.1147e+03	-0.0003	0.9998
302	4.9598e-01	1.1148e+03	0.0004	0.9996
303	3.3770e-03	1.1148e+03	0.0000	1.0000
304	-1.1020e+00	1.1148e+03	-0.0010	0.9992
305	-4.6263e-01	1.1147e+03	-0.0004	0.9997
306	-3.2049e-01	1.1147e+03	-0.0003	0.9998
307	7.4883e-01	1.1147e+03	0.0007	0.9995
308	-4.8261e-01	1.1147e+03	-0.0004	0.9997
309	-1.1593e+00	1.1148e+03	-0.0010	0.9992
310	-4.9030e-01	1.1148e+03	-0.0004	0.9996
.				
.				
420	-2.3464e+00	1.1149e+03	-0.0021	0.9983
421	-1.8138e+00	1.1149e+03	-0.0016	0.9987
422	-2.3817e-01	1.1149e+03	-0.0002	0.9998
423	-7.5615e-01	1.1149e+03	-0.0007	0.9995
424	-1.7922e-03	1.1150e+03	0.0000	1.0000
425	-8.1713e-01	1.1151e+03	-0.0007	0.9994
426	-1.3911e+00	1.1149e+03	-0.0012	0.9990
427	-8.0697e-01	1.1150e+03	-0.0007	0.9994
428	-2.2134e+00	1.1151e+03	-0.0020	0.9984
429	-3.2371e-01	1.1151e+03	-0.0003	0.9998
430	-9.3966e-01	1.1151e+03	-0.0008	0.9993
431	-2.2133e+00	1.1152e+03	-0.0020	0.9984
432	-1.0709e+00	1.1152e+03	-0.0010	0.9992
433	-2.7370e+00	1.1153e+03	-0.0025	0.9980
434	-2.4052e+00	1.1154e+03	-0.0022	0.9983
435	-1.3452e+00	1.1153e+03	-0.0012	0.9990
436	-7.1549e-01	1.1151e+03	-0.0006	0.9995

437	-6.5201e-01	1.1151e+03	-0.0006	0.9995
438	1.2110e+00	1.1152e+03	0.0011	0.9991
439	5.8671e-01	1.1151e+03	0.0005	0.9996
440	-1.7898e+00	1.1153e+03	-0.0016	0.9987
441	-3.9024e-01	1.1152e+03	-0.0003	0.9997
442	3.0291e-01	1.1152e+03	0.0003	0.9998
443	1.4903e-03	1.1151e+03	0.0000	1.0000
444	-3.3441e-01	1.1150e+03	-0.0003	0.9998
445	-1.9373e-01	1.1151e+03	-0.0002	0.9999
446	-7.1638e-03	1.1151e+03	0.0000	1.0000
447	-7.8760e-01	1.1152e+03	-0.0007	0.9994
448	-1.3647e+00	1.1152e+03	-0.0012	0.9990

Como puede comprobarse en la tabla anterior ninguno de los coeficientes es significativo entonces el modelo pooling (Sin efectos) es mejor que el modelo de efectos fijos (tipo de error espacial (kkp) "Kapoor,Kelejian,Prucha")

# Capítulo 5

## CONCLUSIONES

La obtención de la información fue posible gracias a la utilización de programas que las instituciones involucradas actualmente disponen para su análisis tales como REDATAM, ARGIS, R, SPSS, EXCEL, entre otros, así como del empleo del software de distribución libre GeoDA, cuyo principal característica es que nos permite vincular o transferir los resultados específicos tanto del análisis descriptivo como del análisis espacial de tal manera que los resultados numéricos pueden ser visualizados inmediatamente sobre mapas geográficos, además el grupo de herramientas disponibles son apropiadas y hasta imprescindibles para la planificación, y que serían útiles en las diferentes instituciones encargadas del control y seguridad pública.

Con este estudio se propone a las Instituciones del Estado, compatibilicen o uniformización la información, pues podemos mencionar como ejemplo, que los circuitos y subcircuitos, divisiones territoriales empleadas por la Policía, no son compatibles con las divisiones geográficas que emplean el ex CEASI y el INEC. Pues corresponden a tamaños y formas geográficas diferentes elaboradas para uso de la Policía, como parte de este trabajo se intentó compatibilizar la ubicación geográfica de las diferentes UPCs del Distrito Metropolitano de Quito, con la base de datos generada, sin embargo a pesar de las solicitudes presentadas a la Policía del Distrito 9, no se obtuvo dicha información. Posiblemente se deba a cuestiones de seguridad o tal vez a la confidencialidad de la misma. De haberse contado con ella, creemos que habría formado parte del modelo matemático presentado, ya que se estima que la presencia policial incidiría notablemente en la tasa de robos. La nueva información conformada por tablas y mapas cartográficos, puede servir como base o como alternativa para dicha compatibilización.

El empleo de zonas censales obedeció al hecho de que contribuye favorablemente a nuestro análisis y son unidades geograficas correspondientes a regiones administrativas que si disponen

de datos, y que a través de ellas, se puede proveer servicios, distribuir fondos públicos, y lograr eficacia y eficiencia en el tratamiento de la información.

A partir del estudio de los robos en la ciudad de Quito, a los que se han aplicado las técnicas del análisis espacial se puede observar los diferentes comportamientos de los valores asociados al número de robos registrados en cada zona censal (grupo de población) así como de las diferentes variables estudiados.

Con este estudio se promueve la investigación de los delitos, en particular la tasa de robos, perpetrados durante el año 2013 y el primer semestre del año 2014, en los diferentes sectores del Distrito Metropolitano de Quito, además se pudo identificar tendencias y patrones asociados a la inseguridad en los diferentes sectores de esta ciudad. Se ha generado una base de datos de información relevante que puede ser de utilidad para los organismos e instituciones encargadas de la seguridad pública. Así mismo se ha tratado de presentar una muestra de los resultados que se pueden obtener, al emplear análisis de estadística espacial a través del empleo de GeoDa, dentro de estudios o investigaciones útiles para los sistema de seguridad integral y para la comunidad en general.

Los mapas así como los valores numéricos asociados a ellos obtenidos como producto de este análisis espacial, permitirán a las autoridades locales, visualizar los sectores donde la tasa de robos se ha concentrada mayoritariamente, y en función de ellos, distribuir sus recursos tanto humanos como materiales, que creemos deben, estar acorde y de forma proporcional a las necesidades de cada zona, teniendo en cuenta también la cantidad de habitantes, la actividad comercial, el tamaño de la unidad geográfica, la existencia de UPCs (presencia policial), pues se estima que la probabilidad de que la delincuencia se incremente, en áreas, donde las redes de control público, fallan, es significativa (Bursik y Grasmick, 1993 ).

Los mapas presentados sugieren que la mayoría de los robos se están ejecutando en zonas censales con mayor actividad financiera de la ciudad, especialmente en el centro norte de la ciudad. Sin embargo también se puede observar que la delincuencia en los sectores denominados zonas rojas, si se toma en cuenta la tasa de robos frente al número de habitantes presentes en ellas, aún sigue siendo significativo. Se puede visualizar así mismo que en los sectores periféricos que se consideran como los menos atendidos por las autoridades locales y gubernamentales, la tasa de robos parece decrecer, esto posiblemente pueda explicarse si se considera el hecho de que

los delincuentes así como sus víctimas no necesariamente son residentes de las zonas donde se perpetraron los robos.

Se trata de insentivar la ejecución de nuevos estudios, investigaciones o diagnósticos que a su vez permitan la formulación de políticas públicas relacionadas a solventar este tipo problemáticas sociales y de seguridad pública. Con la ampliación y/o continuación de estudios similares se podrá monitorear y evaluar los resultados de políticas, estrategias o programas que se puedan implementar, con el fin de solventar problemas de esta naturaleza.

Se debe aunar esfuerzos para establecer una red técnica que homologue, estandarice y procese la información, así como se debe propender a mejorar la fluidez en la comunicación y coordinación entre las instituciones gubernamentales.

Varios artículos científicos relacionados con análisis de datos espaciales como el propuesto, sugieren que la selección de la matriz de pesos espaciales, debe estar acorde a los datos cuya contigüedad de primer orden corresponda al criterio de la Reina, toda vez que los resultados obtenidos de esta manera son más desables para estos investigadores, (Anselin & Serge 1994).

Los índices de autocorrelación espacial tanto locales como globales y las mapas cartográficos obtenidos con GeoDa, son adecuados para detectar clusters, patrones espaciales, datos atípicos y a la vez abren la posibilidad de plantearnos varios modelos matemáticos, hasta lograr obtener el modelo que mejor se ajusta a nuestros requerimientos.

Para el presente análisis y una vez que hemos empleado las diferentes herramientas de estadística espacial, en particular el suavizado de Bayes, lo que se pretendió con ello, es evitar posibles errores que pueden generarse si se ignora el efecto de la autocorrelación espacial, en el análisis de regresión, pues ello conduciría a que los coeficientes sean ineficientes para mostrar la magnitud de la relación entre las variables. Por otro lado si se ignora la heteroscedasticidad espacial, además de que los coeficientes se vuelven ineficientes, las pruebas de significancia estadística sobre los coeficientes, serían cuestionables, debido a la inflación de los errores estándar (Anselin, 1988). El efecto que esto tendría en el análisis teórico-social sería, que el contexto geográfico, tiene un efecto parcial, en el comportamiento humano (O'Loughlin & Anselin 1992).

Finalmente se puede concluir que este análisis de datos espaciales, permite establecer que las va-

riables socioeconómicas que se han considerado, por sí solas, no necesariamente explican la concentración geográfica de los delitos en ciertas zonas censales del Distrito Metropolitano de Quito. De hecho al considerar ciertas variables que presumíamos podían estar relacionadas con la tasa de robos los resultados no mostraron ser significativos y en ciertos casos hasta parecían contradictorios. Más bien se puede apreciar que ellos, se deben a una gran cantidad de factores externos, que no son tema del presente estudio. Sin embargo podemos mencionar que estadísticamente de algún modo los delitos si están relacionados con las variables socioeconómicas utilizadas y que están presentes en el modelo matemático obtenido.

## **Limitaciones del estudio**

Como se ha indicado antes, el presente trabajo una vez que se logró obtener y procesar la información que actualmente está disponible en varias Instituciones del Estado, su análisis se orientó, al estudio de la tasa de robos en las diferentes zonas censales del sector urbano del Distrito Metropolitano de Quito correspondiente al año 2013 y primer semestre del año 2014, con la finalidad de poder determinar como están los datos distribuidos espacialmente y como se pueden describir sus respectivos valores de manera estadística, como éstas tasas están relacionadas entre sí, como se ven afectadas o no por las tasas de robos en las zonas censales vecinas, se investiga además la presencia de patrones, la existencia de correlaciones y autocorrelaciones, la formación de agrupamientos (clústers), la presencia de valores atípicos y finalmente se realizan varias regresiones espaciales hasta obtener el modelo matemático que trata de explicar como ciertas variables obtenidas de indicadores socioeconómicos se relacionan o no con estas tasas de robos.

## **Recomendaciones para futuros estudios**

Es preciso hacer notar que la gran cantidad de datos disponibles tanto en el INEC, ex CEASI, Instituto de la Ciudad, Observatorio Metropolitano de Seguridad Ciudadana (OMSC), el Sistema Nacional de Información (SNI), Policía Nacional Zona 9 del DMQ, instituciones todas ellas gubernamentales que generalmente realizan estudios o están relacionadas de alguna manera con este tipo de temáticas, imposibilita la realización estudios que contemplen la totalidad de ellos. Esto ahunado al hecho de que la información a más de extensa aún no se encuentra compatibilizada (cada institución maneja sus propios datos en formatos diferentes con dimensiones y tamaños de polígonos geográficos diferentes y en ciertos casos se pudo detectar, incompatibilidad entre la información de una institución otra. Por lo que este campo de estudio aún sigue siendo muy amplio

y variado, el sólo hecho de poder compatibilizar la información, requerirá de mucho tiempo, esfuerzo y recursos.

El campo de estudio de la Estadística Espacial es tan vasto, así como las necesidades individuales o particulares de tal manera que se podrán generar tantos estudios como requerimientos se presenten.

# Capítulo 6

## REFERENCIAS BIBLIOGRÁFICAS

# Bibliografía

- [1] Anderson, D. Sweeney, D. y Williams, T. (2009). Estadística para Administración y Economía. CENGAGE Learning Editores, SA. 10ma Ed.
- [2] Anselin, L. (1988). Spatial Econometrics. Methods and Models. Kluwer Academic Publisher, Dordrecht
- [3] Anselin, L. and Getis, A. (1992). Spatial statistical analysis and geographic information systems. The Annals of Regional Science.
- [4] Anselin, L. (1989) What is Special about Spatial Data? Alternative Perspectives on Spatial Data Analysis. Technical Report 89-4, National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA [contained on CD-Rom Fundamental Research in Geographic Information and Analysis, Santa Barbara, Ca: NCGIA]
- [5] Anselin, L. (1990a) Spatial Dependence and Spatial structural Instability in Applied Regression Analysis. Journal of Regional Science 30:185-207
- [6] Anselin, L. (1990b) Some Robust Approaches to testing and Estimation in Spatial Econometrics. Regional Science and Urban Economics 20:141-163
- [7] Anselin, L. (1994). Exploratory Spatial Data Analysis and Geographic Information Systems. In M. Painho (Ed.), New Tools for Spatial Analysis, Eurostat, Luxembourg, 1994.
- [8] Anselin, L (1995). Local Indicators of Spatial Association - LISA, Geographical Analysis.
- [9] Anselin, L (1996). The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In M. Fischer, H. Scholten, and D. Unwin (Eds.), Spatial Analytical Perspectives on GIS. London: Taylor and Francis.
- [10] Anselin, L. (1996). Interactive Techniques and exploratory spatial data analysis . In Longley, P. A., Goodchild, M. J. , and Rhind, D. W., editors, Geographical Information Systems: Principles, Techniques, Management and Applications. John Wiley, New York.

- [11] Anselin, L. Kelejian, H. (1997) Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors. *International Regional Science Review* 20:153-182
- [12] Anselin, L. Bera, A. (1998) Spatial Dependence in Linear Regression Models With an Introduction to Spatial Econometrics . In A. Ullah and D. Giles, Eds., *Handbook of Applied Economic Statistics*, pp 237-289. New York: Marcel Dekker.
- [13] Anselin, L (2003). *GeoDa 0.9 User's Guide*. Spatial Analysis Laboratory Department of Agricultural and Consumer Economics. University of Illinois, Urbana-Champaign Urbana, IL 61801, from <http://sal.agecon.uiuc.edu/>
- [14] Anselin, L (2004). *GeoDa 0.9.5-i Release Notes*. Spatial Analysis Laboratory Department of Agricultural and Consumer Economics. University of Illinois, Urbana-Champaign Urbana, IL 61801, from <http://sal.agecon.uiuc.edu/>
- [15] Anselin, L (2005). *Exploring Spatial Data with GeoDa: A Workbook*. Spatial Analysis Laboratory Department of Geography. University of Illinois, Urbana-Champaign Urbana, IL 61801, from <http://sal.uiuc.edu/>
- [16] Anselin, L. (1988). *Spatial Econometrics Methods and Model*. Studies Regional Science. Kluwer Academic Publishers. 101 Philip Drive, Norwell, MA 20261, U.S.A.
- [17] Anselin, L. Le Gallo, J. Jayet, H. (2008). *Spatial Panel Econometrics*. In: Matyas L, Sevestre P (eds) *The econometrics of panel data, fundamentals and recent developments in theory and practice*, 3rd edn. Kluwer, Dordrecht, pp 901 -969.
- [18] Arbia, G. (2006). *Spatial Econometrics. Statistical Foundations and Applications to Regional Convergence*. *Advances Spatial Science*. University 'G. d'Annunzio' of Chieti Viale Pindaro, Italy. Springer Verlag Berlin Heidelberg. Germany.
- [19] Baltagi, H. (1981) Simultaneous Equations with Error Components. *Journal of Econometrics*, 17:21-49
- [20] Baltagi, B. Levin, D. (1992) Cigarette taxation: raising revenues and reducing consumption *Struct Change Econ Dyn* 3(2): 321-335
- [21] Baltagi, B. Song, SH. Koh, W. (2003) Testing panel error correlation. *J Econometrics* 117:123-150

- [22] Baltagi, B. Levin, D. (2004) Prediction in the panel data model with spatial autocorrelation. In: Anselin, L. Florax RJGM. Rey, SJ. (eds) *Advances in spatial econometrics: methodology, tools and applications*. Springer, Berlin, pp 283-295
- [23] Baltagi, B. (2005) *Econometric and analysis of panel data*, 3rd edn. Wiley, Chichester
- [24] Baltagi, B. Song, SH. Jung, BC. Koh, W. (2007) Testing for Serial Correlation, Spatial Autocorrelation and Random Effects using Panel Data. *J Econometrics* 140:5-51
- [25] Baltagi, B. (2008a) *Econometric Analysis of Panel Data*. 4th ed. John Wiley & Sons, New York
- [26] Baltagi, B. (2008b) Testing for random Effects and Spatial Lag Dependence in Panel Data Models. *Statistics and Probability Letters*, 78:3304-3306
- [27] Baltagi, B. Liu, L. (2011) Instrumental Variable Estimation of a spatial Autoregressive Panel Model with Random Effects. *Economics Letters*, 111:135-137
- [28] Baltagi, B. Egger, P. Pfaffermayr, M. (2012) A generalized spatial panel data model with random effects. CESifo, Syracuse University. Working Paper Series No. 3930. Available at SSRN: <http://ssrn.com/abstract=2145816> or <http://EconPapers.RePEc:max:cprwps:113>
- [29] Brusch, T. Pagan, A. (1980). The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics. *Review of Econometric Studies*. 47: 239-253
- [30] Bursik Jr., Robert J. *Neighborhood-based networks and the control of crime and delinquency*. En: BARLOW, Hugh D. (Ed.) *Crime and Public Policy: Putting Theory to Work*. Boulder, CO, Westview Press. 1995.
- [31] Bursik Jr. Robert J. & GRASMICK, Harold G. *Neighborhoods and crime: the dimensions of effective community control*. New York, Lexington. 1993.
- [32] CEPAL. UNICEF(2010). *Pobreza infantil en América Latina y el Caribe*.
- [33] Cliff, A. y Ord, J. K. (1972). Testing for Spatial Autocorrelation among Regression Residuals *Geographical Analysis*, 4:267-284
- [34] Cliff, A. y Ord, J. K. (1973). *Spatial Autocorrelation*. London: Pion
- [35] Cliff, A. y Ord, J. K. (1981). *Spatial Processes: Models and Applications*. London: Pion.
- [36] Cressie, N. (1993). *Statistic for Spatial Data (Revised Version)*. New York. Wiley

- [37] Croissant, Y. Millo, G. (2008) Panel Data Econometrics in R. The plm Package. *Journal of Statistical Software*, 27(2), 1-43. URL <http://www.jstatsoft.org/v27/i02/>.
- [38] Croissant, Y. (2011) Ecdat: Data Sets for Econometrics. R package version 0.1-6.1, URL <http://CRAN.R-project.org/package=Ecdat>
- [39] Davidson, J. (1994). *Stochastic Limit Theory* Oxford: Oxford University Press
- [40] Davidson, R. y Mackinnon J. (1984). Model Specification Test Based on Artificial Regressions. *International Economic Review*
- [41] Durbin, R. (1988) Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms. *Review of Econometrics and Statistics* 70:466-474
- [42] Elhorst, J. (2001) Dynamic models in space and time. *Geogr Anal* 33:119-140
- [43] Elhorst, J. (2003) Specification and Estimation of Spatial Panel Data Models. *International Regional Sciences Review*, 26(3):244-268
- [44] Elhorst, J. (2005) Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geogr Anal* 37(1):62-83
- [45] Elhorst, J. (2008a). Serial and Spatial Error Correlation. *Economics Letters*, 100:422-424
- [46] Elhorst, J. (2008b). A spatiotemporal analysis of aggregate labour force behaviour by sex and age across the European union. *J Geogr Syst* 10(2):167-190
- [47] Elhorst, J. (2009). Spatial Panel Data Models. In MM Fisher, A Getis (eds), *Handbook of Applied Spatial Analysis*. Springer-Verlag
- [48] Elhorst, J. (2010a). Applied spatial econometrics; raising the bar. *Spat Econ Anal* 5:9-28
- [49] Elhorst, J. (2010b). Dynamic panel with endogenous interaction effects when T is small. *Reg Sci Urban Econ* 40(5):272-282
- [50] Elhorst, J. Piras, G. Arbia, G. (2010). Growth and convergence in a multi-regional model with space-time dynamics. *Geogr Anal* 42(3):338-355
- [51] Elhorst, J. Zandberg, E. de Haan, J. (2013). The impact of interaction effects among neighboring countries on financial liberalization and reform; a dynamic spatial panel data approach. *Spat Econ Anal*. doi: 10.1080/17421772.2012.760136
- [52] Elhorst, J. (2014) *Spatial Econometrics from Cross-Sectional Data to Spatial Panels*. Springer

- [53] Fox, J. Weisberg, S. (2010). *An R Companion to Applied Regression*. 2nd ed. Sage Publications, Thousand Oaks.
- [54] Flint, C., Harrower M., & Edsall, R. (2000). *But How Does Place Matter? Using Bayesian Networks to Explore a Structural Definition of Place*. Paper presented at the New Methodologies for The Social Sciences Conference. University of Colorado at Boulder.
- [55] Goodchild, M. (1987). A Spatial Analytical Perspective on Geographical Information Systems. *International Journal Of Geographical Information Systems*, 1, 327-334.
- [56] Gujarati, D. y Porter, D. (2010). *Econometría*. Mc Graw Hill. 5ta Ed.
- [57] Hausman, A. (1978). Specification Tests in Econometrics. *Econometrica*, 46: 1251-1271
- [58] Kapoor, M. Kelijian, H. Prucha, I. (2007) Panel Data Model with Spatially Correlated Error Components. *J Econometrics* 140:97-130. TSP and Stata software available at [http://econweb.umd.edu/prucha/Research\\_Prog3.htm](http://econweb.umd.edu/prucha/Research_Prog3.htm).
- [59] Kelijian, H. Prucha, I. (1997). Estimation of Spatial Regression Models with Autoregressive Errors by Stage Least Squares Procedures: A Serious Problem. *International Regional Science. Review* 20:103-111
- [60] Kelijian, H. Prucha, I. (1998). A Generalized Spatial two Stage Least Squares Procedure for Estimation a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* 17:99-121
- [61] Kelijian, H. Prucha, I. (1999). On the Asymtotic Distribution of the Moran I test Statistic with Applicatons. Working Paper , Department of Economics, University of Maryland, College Park, MD.
- [62] Kelijian, H. Prucha, I. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J Econometrics* 157:53-67
- [63] Keith, H. (1999). *Mapping Crime Principle and Parctice*. National Institute of Justice. Crime Mapping Research Center. Departamento de Justicia de EEUU oficina de Programas de Justicia. Washington, DC.
- [64] Lee, L. Yu, J (2010a) Some Recent Development in Spatial Panel Data Models. *Reg Sci Urban Economics* 40:255-271

- [65] Lee, L. Yu, J (2010b) Efficient GMM Estimation of Spatial Dynamic Panel Data Models with Fixed Effects. <http://www.economics.smu.edu.sg/events/Paper/LungfeiLee.pdf>
- [66] Lee, L. Yu, J (2010c) A Spatial Dynamic Panel Data Model with both time and Individual Fixed Effects. *Econometric Theor* 26:564-597
- [67] Lee, L. Yu, J (2010d) Estimation of Spatial Autoregressive Panel data Models with Fixed Effects. *Journal of Econometrics* 1554:165-185
- [68] LeSage, J. (1999) *Spatial Econometrics Using MATLAB*. Retrieved August, 2001, from: [www.econ.utoledo.edu](http://www.econ.utoledo.edu).
- [69] LeSage, J. Pace, R. (2009a) *Introduction To Spatial Econometrics*. Chapman & Hall/CRC Boca Raton
- [70] LeSage, J. Pace, R. (2009b) *Introduction to Spatial Econometrics*. STATISTICS: Textbooks and Monographs D.B. Owen. CRC Press. Boca Raton.
- [71] Lind, D. Marchall, W. y Wathen, S. (2008). *Estadística Aplicada a los Negocios y la Economía*. Mc Graw Hill. 13va Ed.
- [72] O'Loughlin, J. & Anselin, L. (1992). *Geography of International Conflict and Cooperation: Theory and Methods*. In M. Ward (Ed.) *The New Geopolitics*, 11-38.
- [73] Martori, J.C. y Hoberg, K. (2008, 15 de abril). *Nuevas Técnicas de Estadística Espacial para la detección de Clusters Residenciales de Población Inmigrante*. *Revista Scripta Nova*. Universidad de Barcelona. Vol XII, núm 263.
- [74] Montgomery, D. Runger, G. (2010). *Probabilidad y Estadística aplicadas a la Ingeniería*. Mc Graw Hill. 5ta Ed.
- [75] Moran, P. (1950). *Notes on continuous stochastic phenomena*. *Biometrika*.
- [76] Moreno, R. y Vayá, E. (2000). *Técnicas Econométricas para el Tratamiento de Datos Espaciales: La Econometría Espacial*, Edicions Universitat de Barcelona, Barcelona.
- [77] Moreno, R. y Vayá, E. (2004). *Econometría espacial: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas*. *Investigaciones Regionales*. Sección Artículos. pp 83-126.
- [78] Mutl, J. (2006) *Dynamic Panel data Models with Spatially Autocorrelated Disturbances*. PhD Thesis, University of Maryland, College Park

- [79] Mutl, J. Pfaffermayr, M. (2011) The Hausman Test in a Cliff and Ord Panel Data. *Econometrics Journal*, 14:48-76
- [80] Pace, R. and Barry, R. (1997a). Sparse Spatial Autoregressions. *Statistics and Probability Letters*.
- [81] Pace, R. and Barry, R. (1997b). Quick Computation of Saptial Autoregressive Estimators. *Geographical Analysis*.
- [82] Parent, O. LeSage, J. (2010) A spatial dynamic panel data with random effects applied to commuting times . *Transp Res Part B* 44:633-645
- [83] R Development Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [84] Millo, G. Piras, G. (2012) spml: Spatial Panel Data Models in R. *Journal of Statistical Software*. <http://www.jstatsoft.org>
- [85] Yang, Z. Li, C. Tse, Y. (2006) Functional form and spatial dependence in spatial panels. *Econ Lett* 91:138-145
- [86] Yu, J. de Jong, R. Lee, L. (2008) Quasi-maximum likelihood estimators for spatial dynamic panel
- [87] Yu, J. de Jong, R. Lee, L. (2012) Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration. *J Econometrics* 167:16-37
- [88] Villalta C. J. (2005). Sobre la Espacialidad de los Procesos Electorales Urbanos y una Comparación entre las Técnicas de Regresión OLS y SAM. *Estudios Demográficos y Urbanos. Tecnológico de Monterrey. Ciudad de México*.
- [89] Wooldridge, J. (2002) *Econometric Analysis of Cross-Section and Panel Data*. MIT Press

# Capítulo 7

## ANEXOS

### ANEXO A

GeoDa es marca registrada de Luc Anselin. Es una colección de software con herramientas diseñadas para análisis exploratorio de datos espaciales (ESDA). Puede obtenerse desde <http://geodacenter.asu.edu/software/downloads>. Referirse a guía de usuario de GeoDa en <http://geodacenter.asu.edu/software/documentation>.

R es un entorno de software libre para computación y gráficos estadísticos. Se compila y ejecuta en una amplia variedad de plataformas UNIX, Windows y MacOS. Está disponible en <http://www.r-project.org/>

### ANEXO B

#### 7.1. Definición de términos

**Estadística Espacial:** Técnicas estadísticas desarrolladas para ser empleadas con datos geográficos.

**Georeferenciar:** Posicionar una entidad en una ubicación geográfica única.

**Entidad:** Elemento u objeto de análisis (zonas censales) con la misma representación espacial.

**Variables:** Características presentes en cada elemento (tasa de robos, hogares con extrema pobreza, etc.)

**Homologar:** Poner las variables en relación de igualdad o equivalencia.

**Dataset:** Colección de datos normalmente tabulados o recuperados de bases de datos y sirven para mapear objetos.

**Geodatabase:** Base de datos espacial para almacenamiento de geoinformación. Es una colección de datasets geográficos, administrados como tablas por filas y columnas.

**SIG:** Sistemas de información geográfica.

**Punto:** En estadística espacial, proporciona información de la localización geográfica.

**Línea:** Proporciona información de la localización geográfica más la información de los atributos.

**Polígono:** Contiene las coordenadas geográficas y la geometría (forma) que define al polígono.<sup>1</sup>

**Shapefile:** Formato de archivo informático de datos espaciales (SHP), empleado para el intercambio de información geográfica.

**Patrones:** Conjunto de elementos que forman una unidad diferenciada y que generalmente se repite.

**Ráster:** Imagen de mapa de píxeles

**ESDA:** Análisis Exploratorio de Datos Espaciales.<sup>2</sup>

**LISA:** Análisis Local de Autocorrelación Espacial.<sup>3</sup>

**Dependencia Espacial (DE):** El valor de una variable depende o es función de la misma variable en unidades geográficas vecinas

**Autocorrelación Espacial (AE):** Concentración o dispersión de valores de una variable que

<sup>1</sup>Las coordenadas geográficas son entendidas también, como coordenadas proyectadas sobre una superficie plana.

<sup>2</sup>ESDA Exploratory Spatial Data Analysis.

<sup>3</sup>LISA, Local Spatial Autocorrelation Analysis, esta basado en el índice local de Moran (Anselin, 1995), corresponde a la medida de autocorrelación espacial para cada ubicación geográfica de manera individual. Puede ser univariado o bivariado.

puede visualizarse en un mapa.

**Heterogeneidad espacial (HE):** Variación de las relaciones entre las variables en el espacio (LeSage, 1999.)

**heteroscedasticidad:** Corresponde a la existencia de una varianza no constante en las perturbaciones aleatorias de un modelo de regresión.

**DPA:** División Político Administrativa del País.

**D.M.Q.:** Distrito Metropolitano de Quito.

**Shape file:** Archivo principal que se utiliza en GeoDa, el mismo que está formado por varios archivos, cuyas extensiones son:

**dbf:** tabla de datos

**prj:** proyección

**sbn**

**sbx**

**shp:** información gráfica

**shx:** relación entre la parte gráfica y la tabla de datos.

**Tabla:** Conjunto de datos organizados en filas y columnas.

**Filas:** Cada fila de la tabla representará un registro único (zona censal).

**Columna:** Cada columna representará un campo (variable).

**Atributo:** Información no espacial acerca de una entidad geográfica, almacenada en una tabla. Ejemplo nombre, longitud.

**Unir:** Anexar los campos de una tabla a las de otra a través de un atributo o campo común.

**ID de objeto:** Valor administrado por el sistema que identifica de forma única a un registro.

