

ALGUNOS ELEMENTOS PARA EL ANÁLISIS DE DATOS
ESPACIALES: TEORÍA Y APLICACIÓN

INGRID ACEVEDO BOHÓRQUEZ
NATALIA MARÍA GÓMEZ ALVAREZ

TRABAJO DE GRADO PRESENTADO COMO
REQUISITO PARCIAL PARA OPTAR AL TÍTULO DE
MAESTRÍA EN MATEMÁTICAS APLICADAS

ASESOR: Dr. HERMILSON VELÁSQUEZ CEBALLOS

MEDELLN
UNIVERSIDAD EAFIT
DEPARTAMENTO DE CIENCIAS BASICAS

Junio 2008

NOTA DE ACEPTACION

FREDDY HERNAN MARIN SANCHEZ

NOTA DE ACEPTACION

FRANCISCO IVAN ZULUAGA DIAZ

Medellín, 5 de junio de 2008

Resumen

La econometría espacial es una disciplina que se sirve de una serie de técnicas y herramientas estadísticas, algebraicas y del cálculo para evaluar la presencia de efectos espaciales en la explicación de fenómenos económicos y sociales y de formular modelos que tengan en cuenta dichos efectos. Para ello se identifican algunos efectos espaciales que pueden presentarse en los datos como son la dependencia espacial, autocorrelación espacial, heterogeneidad espacial, heteroscedasticidad espacial. Adicionalmente, el análisis exploratorio de datos espaciales, propone el uso de adaptaciones de algunos gráficos tradicionales que facilitan la identificación de los efectos para el análisis de datos espaciales, entre ellos se encuentran los box maps, histogramas, mapa y scatterplot de Moran, mapas e histogramas de frecuencias, por mencionar algunos para describir los patrones de relacionamiento entre los datos observados en diferentes unidades espaciales. Una parte muy importante dentro de este Análisis es la identificación de clusters que permite ver como se dan concentraciones de unidades espaciales con valores similares de la variable en estudio.

Para que los resultados de los análisis exploratorios de datos espaciales conduzcan a conclusiones contundentes, es necesario que en la toma de muestras de información en la que se base el estudio cumpla con unas características adecuadas de delimitación de las regiones, para ello se deben utilizar criterios definidos según el objetivo del estudio pues de lo contrario se pueden presentar problemas de medición que afectarían la confiabilidad de las conclusiones. Igual sucede con la matriz de contigüidad, herramienta que de no definirse de la manera más adecuada teniendo en cuenta el fin del estudio, puede distorsionar los resultados del análisis.

Un elemento que ha adquirido vital importancia para probar la presencia de autocorrelación espacial en los datos bajo estudio es el índice I de Moran, que siendo en

esencia un cociente de formas cuadráticas de residuos de regresión adaptado al contexto espacial, ofrece bondades de cálculo, manipulación y distribución que son una de las herramientas preferidas por los investigadores para describir la importancia de conocer la ubicación física de una observación si se quiere pronosticar el valor de un atributo en esa ubicación.

Finalmente, el análisis exploratorio de los datos de número de homicidios cometidos en cada municipio del departamento de Antioquia entre los años 2001 y 2006, confirma la presencia de efectos espaciales y se identifican clusters en el comportamiento del fenómeno social bajo análisis. Esta conclusión indica que un modelo que se formule con el fin de explicar o predecir el comportamiento de esta variable debe tener en cuenta en su especificación la ubicación geográfica de las observaciones, pues de lo contrario padecerá de errores de especificación y el efecto espacial quedará “encubierto” dentro del error del modelo pudiendo ser clara y correctamente definido como una variable adicional.

Palabras Clave: Econometría espacial, Análisis Exploratorio de Datos Espaciales, Clusters

Abstract

Spatial econometrics is a discipline that utilizes a series of techniques and tools of statistics, algebra, and calculus. Its aim is to evaluate the presence of spatial effects in scenarios of economical and social phenomena and to formulate models that take such effects into account. In order to do so, we identify spatial effects present among the data, such as spatial dependence, spatial autocorrelation, spatial heterogeneity and spatial heteroscedasticity.

Exploratory analysis of spatial data proposes that we use adaptations of several traditional graphics that ease the identification of spatial effects. Such graphics that allow discovering patterns between spatial data are box maps, histograms, Moran's maps and scatterplots, frequency maps and histograms.

A crucial part of this analysis is the identification of clusters which reveals the form of the concentration of spatial units with similar values for the variable under study. So that the results of the exploratory analysis of spatial data lead to trusty conclusions it is necessary that the employed sample techniques hold some accurate characteristics of region definition. To achieve this, we have to use specific criteria according to the objectives of the underlying study. Otherwise, we can face measurement problems that would affect the reliability of the conclusions. Likewise, the contiguity matrix can introduce some distortion to the results of the analysis when not properly defined. Moran's I statistic has became one of the most important elements to prove the presence of spatial autocorrelation between data. It is essentially a ratio of quadratic

forms of regression residuals adapted to a spatial context. Moran's I offers a number of calculation, manipulation and distribution advantages. That is the reason for researchers to prefer it over others when needing to describe the importance of knowing the geographic location of an observation in order to predict the value of an attribute in that location.

Finally, exploratory analysis conducted about the number of homicides occurred in each town of Antioquia between 2001 and 2006 confirm the presence of spatial effects and clusters related to the social phenomenon under study. This conclusion indicates that a model that aims to explain or predict the number of homicides should consider into its specification the geographic position of the spatial units (through a contiguity matrix). Otherwise, it will involve specification errors and the spatial effect will be hidden under the error of the model instead of being clearly defined as an additional variable of the model.

Índice general

Introducción	1
1. Preliminares Matemáticos y Estadísticos	7
1.1. Introducción	7
1.2. Estimación por el método de Máxima Verosimilitud	7
1.3. Función de verosimilitud para un modelo espacial	8
1.4. Derivadas parciales con respecto al parámetro ρ	14
1.5. Derivadas parciales con respecto al parámetro λ	18
1.6. Derivadas parciales con respecto al parámetro β	21
1.7. Derivadas parciales con respecto al parámetro α_p	22
1.8. Primeras derivadas parciales de la función de verosimilitud	25
1.9. Segundas derivadas parciales de la función de verosimilitud	27
1.9.1. Para los elementos de la diagonal	27
1.9.2. Para los términos cruzados	29
1.10. Distribución Normal Multivariada	38
2. Análisis Exploratorio de Datos Espaciales	49
2.1. Introducción	49
2.2. Efectos Espaciales	51
2.2.1. Dependencia Espacial	52

2.2.2.	Autocorrelación Espacial	54
2.2.3.	Heterogeneidad Espacial	55
2.2.4.	Heteroscedasticidad Espacial	57
2.3.	Matriz de Contigüidad	57
2.4.	Medidas de Similaridad	59
2.4.1.	Definición de métrica	60
2.4.2.	Medidas de distancia	61
2.4.3.	Coeficientes de correlación	63
2.4.4.	Coeficientes de asociación	64
2.4.5.	Coeficientes de probabilidad	67
2.5.	Análisis tradicional vs. Análisis espacial	68
2.6.	Visualización de los datos	68
2.7.	Presentación de los datos	69
2.7.1.	Identificación de valores atípicos	70
2.7.2.	Identificación de la distribución de la variable	72
2.7.3.	Identificación de efectos espaciales	73
2.8.	Transformaciones utilizadas en el AEDE	76
2.9.	Delimitación de regiones	80
2.10.	Identificación de <i>clusters</i>	83
3.	Índice I de Moran	86
3.1.	Introducción	86
3.2.	Evolución del índice <i>I</i> de Moran	88
3.3.	Construcción del I de Moran	90
3.4.	Distribución Exacta del I de Moran	100
4.	Caso de Aplicación	107
4.1.	Introducción	107

4.2. Departamento de Antioquia	108
4.3. Definición de variables	108
4.4. Resultados Obtenidos	109
4.4.1. Identificación de valores atípicos	109
4.4.2. Identificación de la distribución de la variable	110
4.4.3. Identificación de efectos espaciales	111
4.5. Identificación de <i>clusters</i>	112
5. Conclusiones	131
Bibliografía	133

Índice de cuadros

2.1. Técnicas de Análisis Exploratorio de Datos Espaciales	71
----------------------------------------------------------------------	----

Índice de figuras

2.1. Dependencia espacial y agregación	54
4.1. Subregiones del Departamento de Antioquia	108
4.2. <i>Clusters</i> identificados para el año 2006	113
4.3. Diagrama de Caja - Mapa para el año 2001	114
4.4. Diagrama de Caja - Mapa para el año 2002	114
4.5. Diagrama de Caja - Mapa para el año 2003	115
4.6. Diagrama de Caja - Mapa para el año 2004	115
4.7. Diagrama de Caja - Mapa para el año 2005	116
4.8. Diagrama de Caja - Mapa para el año 2006	116
4.9. Boxmap para el año 2001	117
4.10. Boxmap para el año 2002	118
4.11. Boxmap para el año 2003	118
4.12. Boxmap para el año 2004	119
4.13. Boxmap para el año 2005	119
4.14. Boxmap para el año 2006	120
4.15. Histograma para el año 2001 - Subregión Oriente resaltada	120
4.16. Histograma para el año 2002 - Subregión Norte resaltada	120
4.17. Histograma para el año 2003 - Subregión Nordeste resaltada	121
4.18. Histograma para el año 2004 - Subregión Urabá resaltada	121

4.19. Histograma para el año 2005 - Subregión Suroeste resaltada	121
4.20. Histograma para el año 2006 - Subregión Occidente resaltada	122
4.21. Mapa de Cuartiles para el año 2001	123
4.22. Mapa de Cuartiles para el año 2002	123
4.23. Mapa de Cuartiles para el año 2003	124
4.24. Mapa de Cuartiles para el año 2004	124
4.25. Mapa de Cuartiles para el año 2005	125
4.26. Mapa de Cuartiles para el año 2006	125
4.27. Cartograma para el año 2001	126
4.28. Cartograma para el año 2002	126
4.29. Cartograma para el año 2003	127
4.30. Cartograma para el año 2004	127
4.31. Cartograma para el año 2005	128
4.32. Cartograma para el año 2006	128
4.33. Diagrama de Dispersión de Moran para el año 2001	129
4.34. Diagrama de Dispersión de Moran para el año 2002	129
4.35. Diagrama de Dispersión de Moran para el año 2003	129
4.36. Diagrama de Dispersión de Moran para el año 2004	130
4.37. Diagrama de Dispersión de Moran para el año 2005	130
4.38. Diagrama de Dispersión de Moran para el año 2006	130

Introducción

En este trabajo se presentan de manera formal y didáctica algunos elementos del Análisis Exploratorio de Datos Espaciales y de la econometría espacial. El estudio y los resultados de esta disciplina son relativamente recientes y han cobrado mayor importancia en las últimas décadas ya que con el avance de las tecnologías en las comunicaciones y la globalización de la economía, los sucesos que ocurren en una ubicación específica tienen repercusiones sobre sus vecinos directos e incluso sobre otros, aparentemente remotos. En el estudio de cualquier fenómeno de carácter social o económico la ubicación geográfica de los agentes constituye un aspecto importante dentro de la especificación de los modelos econométricos, ya que puede existir algún efecto espacial, que de no ser incorporado en la especificación, podría afectar la validez del modelo. Ante esta realidad y gracias al desarrollo tecnológico de los sistemas de georreferenciación de datos, surge la necesidad de contar con herramientas apropiadas para el procesamiento, descripción y análisis de la información. Los métodos tradicionales de la estadística descriptiva no tienen en cuenta la localización geográfica de los datos, razón por la cual se ha conformado un conjunto de técnicas nuevas denominado Análisis Exploratorio de Datos Espaciales - AEDE, que incorpora el aspecto espacial como un factor importante para tener en cuenta en los análisis.

Teniendo en cuenta que la econometría tradicional no ha incorporado el efecto de dichas circunstancias y que la estadística espacial se ocupa de otro tipo de problemas, ha surgido una disciplina a la cual se le ha dado el nombre de econometría espacial.

Según Luc Anselin, uno de sus principales investigadores, “las actividades como la estimación de modelos espaciales de interacción, el análisis estadístico de la función de densidad urbana y la implementación empírica de modelos econométricos regionales, podrían ser considerados econometría espacial” [4].

Al utilizar información georreferenciada surge el interrogante sobre si se encuentra presente algún tipo de dependencia espacial entre los datos. Esta dependencia se denomina autocorrelación espacial y es el más importante de los efectos espaciales. Para contrastar su presencia el estadístico más utilizado fue propuesto por Moran en los años cincuenta y a partir de él se han diseñado nuevas propuestas.

El objetivo principal del presente trabajo es presentar de manera formal algunos elementos del Análisis Exploratorio de Datos Espaciales, presentar un algoritmo para realizar la identificación de *clusters* dentro de las unidades espaciales, hacer una revisión de la literatura existente sobre el estadístico I de Moran y realizar una aplicación de la teoría utilizando información sobre la variable homicidios en los municipios del departamento de Antioquia.

Para cumplir el objetivo se presenta un primer capítulo que corresponde al desarrollo matemático formal de algunos conceptos que constituyen la base de la econometría espacial. Se comienza con un modelo espacial y se deduce la forma que cobra la función de Máxima Verosimilitud que es la base para realizar la estimación de los parámetros del modelo. A continuación se presentan las derivadas parciales de cada uno de los elementos de la función de Verosimilitud, elementos que constituirán la matriz de información, la cual sirve para obtener los parámetros del modelo. Después se presenta el desarrollo de la función de densidad de probabilidad normal multivariada, la cual es elemento básico en la deducción de la función de verosimilitud.

El segundo capítulo de este trabajo está conformado por las definiciones de los términos usados en la econometría espacial; los procedimientos que permiten realizar adecuadamente el Análisis Exploratorio de Datos Espaciales - AEDE, y sus diferencias

fundamentales con el análisis de datos tradicional. Se hace énfasis en el trabajo gráfico del AEDE, que permite observar las características de los datos espaciales que no se perciben en el análisis tradicional. Al final del Capítulo se presenta la descripción de un algoritmo que permite identificar *clusters* de unidades espaciales con características similares.

Un aspecto fundamental en el análisis de datos espaciales lo constituye la autocorrelación espacial, por esta razón se incluye en el tercer capítulo una revisión del índice I de Moran y sus diferentes variantes desde el artículo inicial *The Interpretation of Statistical Maps* [19] hasta *On the asymptotic distribution of the Moran I test statistic with applications* [16].

Por último se incluye en el cuarto capítulo un caso aplicado con datos reales del departamento de Antioquia, donde además del Análisis Exploratorio de Datos Espaciales, se implementará un algoritmo en MATLAB que permite identificar *clusters* o agrupaciones de municipios del departamento, que presentan características similares de violencia.

Este trabajo reviste una importancia especial, pues al consultar se ha encontrado que en Colombia existen muy pocos trabajos aplicados de este tema, y se considera que realmente muchos de los modelos econométricos desarrollados hasta el momento presentan errores de especificación al no contemplar la localización geográfica como un elemento importante dentro de la modelación. El presente trabajo no pretende ser un manual exhaustivo del tema correspondiente a la Econometría Espacial sólomente busca ser una base para futuros trabajos de investigación de grupos que se interesen en el asunto. En este tópico es importante realizar valoraciones más profundas sobre los modelos que se pueden implementar o especificar.

Marco Teórico y Estado del Arte

Los recientes procesos de integración regional y los cada vez más frecuentes acuerdos comerciales entre naciones, plantean nuevos desafíos para la modelación en econometría que involucren las relaciones espaciales entre las regiones.

Este trabajo pretende servir como base para la realización de asesorías y consultorías en asuntos que involucren análisis cuantitativos regionales que se realicen en Colombia. Se espera iniciar un proceso continuo y sostenible que facilite la interconexión de grupos de investigación en el desarrollo de aplicaciones estadísticas, matemáticas y econométricas que tengan en común el espacio como elemento integrador. Con esto se espera contribuir al fortalecimiento del desarrollo integral de la investigación sobre temas económicos, sociales, políticos y demográficos.

El inicio de las investigaciones en econometría espacial se remonta a la década de los 70, cuando se empieza a hablar de la Econometría Espacial como una disciplina útil para el estudio de problemas regionales. En 1979, año que se ha adoptado como fecha de nacimiento de la econometría espacial, se publica *Spatial Econometrics* cuyos autores son Jean Paelink y Leo Klaassen, quienes afirman en el prólogo de su libro “el presente libro no es más que el prefacio de tal disciplina: en otras palabras, representa un balance provisional, basado esencialmente en nuestro propio trabajo, el cual nos permitirá organizar mejor nuestras ideas”.

En el año 1988 Luc Anselin publica su libro *Spatial Econometrics: Methods and Models* [4], el cual él mismo presenta de la siguiente manera: “En este libro combino diferentes resultados recientes para elaborar un enfoque amplio donde puedan incorporarse los efectos espaciales en Econometría. Mi interés principal es demostrar como estos efectos espaciales pueden considerarse casos especiales dentro del planteamiento econométrico general, y subrayar por que necesitan de un conjunto específico de métodos y técnicas, incluidas en el campo de la econometría espacial”. En este libro

Anselin adapta los procedimientos existentes dentro de la econometría tradicional, para conformar una metodología básica para la aplicación de la econometría espacial. En 1998 James LeSage publica en la red de la página de la Universidad de Toledo un libro *Spatial Econometrics CIRCULATED FOR REVIEW* [17], el cual “provee una introducción a la econometría espacial, así como también un conjunto de funciones MATLAB que implementan métodos de estimación en econometría espacial”, tal como lo manifiesta el autor en su prólogo.

En el año 2000 Rosina Moreno Serrano y Esther Vaya Valcarce publican el libro *Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial*[21], primera publicación sobre el tema que se realiza en idioma castellano, en la que las autoras afirman que la finalidad de su libro “se centra de forma explícita en los efectos espaciales y consiste en el conjunto de técnicas que permiten describir distribuciones espaciales, identificar localizaciones atípicas (outliers espaciales), descubrir esquemas de asociación espacial (cluster espacial) y sugerir diferentes regímenes espaciales u otras formas de inestabilidad espacial”.

En lo que tiene que ver específicamente con el estadístico I de Moran la historia se remonta más atrás. En el primer artículo publicado por Moran en 1948 *The Interpretation of Statistical Maps*[19], se discute sobre la influencia de la independencia o dependencia entre lo que ocurre en dos unidades espaciales en las pruebas de significancia.

En el artículo publicado por Moran en 1950 *Notes on Continuous Stochastic Phenomena* [20] aparece por primera vez el estadístico I de Moran. Pasan casi dos décadas sin que se vuelvan a encontrar trabajos sobre el Estadístico I de Moran hasta 1971, cuando Andrew Cliff y Keith Ord publican su artículo *Testing for Spatial Autocorrelation Among Regression Residuals* [10], en el cual generalizaron el estadístico I de Moran con el fin de obtener una prueba para la autocorrelación espacial en un modelo

de regresión lineal. Los resultados anteriores están sujetos a unas condiciones específicas con respecto a la distribución para muestras grandes del estadístico de prueba I de Moran, las cuales son restrictivas y excluyen muchas aplicaciones potenciales. Posteriormente en el año 1995 Luc Anselin y Raymond Florax publican su artículo *Small sample properties of tests for spatial dependence in regression models: some further results* [5].

En el año 1997 Luc Anselin y Harry Kelejian publican el artículo *Testing for spatial autocorrelation in the presence of endogenous regressors*. [6]

Por ultimo el artículo mas reciente es *On the asymptotic distribution of the Moran I test statistic with applications* de Harry Harry Kelejian, Ingmar R. Prucha. [16]

Capítulo 1

Preliminares Matemáticos y Estadísticos

1.1. Introducción

A continuación se presentan algunos resultados matemáticos y estadísticos que constituyen los fundamentos teóricos para el desarrollo formal de la econometría espacial.

1.2. Estimación por el método de Máxima Verosimilitud

La estimación por Máxima Verosimilitud es un método tradicional en estadística, el cual también se utiliza en econometría espacial, ya que en el estudio de esta, usualmente se hacen supuestos sobre el tipo de distribución para los términos de perturbación y por lo tanto este procedimiento constituye una manera natural de abordar el problema de estimación.

1.3. Función de verosimilitud para un modelo espacial

Se considera la siguiente especificación para un modelo espacial:

$$Y = \rho W_1 Y + X\beta + \epsilon \quad (1.1)$$

$$\epsilon = \lambda W_2 \epsilon + u \quad (1.2)$$

$$u \sim N(0, \Omega), \quad (1.3)$$

esto indica que:

- Las variables u_i son normales
- El vector de medias es cero, $E(u) = 0$
- La matriz de varianzas y de covarianzas del vector u es $\Sigma_u = \Omega$

Los elementos de la diagonal de la matriz de covarianza $\Sigma_u = \Omega$ son

$$\Omega_{ii} = h_i(z\alpha) \quad h_i > 0 \quad (1.4)$$

Donde

N : es el número de observaciones

Y : es un vector ($N \times 1$) correspondiente a la variable endógena

WY : corresponde al retardo espacial de la variable Y

$W\epsilon$: corresponde al retardo espacial del término de perturbación ϵ

ρ : es el coeficiente de la variable dependiente rezagada espacialmente

X : Matriz de variables exógenas ($N \times K$)

β : es un vector ($K \times 1$) de parámetros asociado con la matriz de variables exógenas

- ϵ : es un vector ($N \times 1$) de términos de perturbación
 λ : es el coeficiente en la estructura autorregresiva espacial de ϵ
 u : es el vector ($N \times 1$) de términos de perturbación, el cual distribuye normal.
 Ω : matriz ($N \times N$) de varianzas y covarianzas
 W_1 y W_2 : matrices de ponderaciones espaciales ($N \times N$)

Los elementos de la diagonal de la matriz de varianzas del término de error, tal como están especificados, permiten la existencia de heteroscedasticidad, siendo una función de $P + 1$ variables exógenas z , los cuales incluyen un término constante. Los P parámetros α están asociados con los términos no constantes, de tal forma que para $\alpha = 0$, se obtiene que

$h = \sigma^2$ que corresponde a la situación de homocedasticidad clásica.

Las matrices W_1 y W_2 , pueden ser estandarizadas o no y están respectivamente asociadas con un proceso autorregresivo espacial en la variable dependiente y en el término de perturbación.

Este modelo tiene $3 + K + P$ parámetros desconocidos, el cual se puede representar de forma vectorial como:

$$\theta = [\rho, \lambda, \sigma^2, \beta', \alpha']' \quad (1.5)$$

De manera alternativa el modelo se puede presentar usando la siguiente simplificación en la notación:

$$A = I - \rho W_1 \quad (1.6)$$

$$B = I - \lambda W_2 \quad (1.7)$$

De tal forma que la ecuación (1.1) se transforma así:

$$Y - \rho W_1 Y = X\beta + \epsilon \quad (1.8)$$

$$(I - \rho W_1)Y = X\beta + \epsilon \quad (1.9)$$

$$AY = X\beta + \epsilon \quad (1.10)$$

y la ecuación (1.2) queda así:

$$\epsilon - \lambda W_2 \epsilon = u \quad (1.11)$$

$$(I - \lambda W_2)\epsilon = u \quad (1.12)$$

$$B\epsilon = u \quad (1.13)$$

La matriz de varianzas y covarianzas del término de perturbación u es $\Sigma_u = \Omega$.

$$\Sigma_u = \Omega = E[(u - E(u))(u - E(u))'] \quad (1.14)$$

Como $E(u) = 0$, se obtiene que

$$\Sigma_u = \Omega = E(uu')$$

La matriz Ω es diagonal y definida positiva, por lo tanto existe una matriz C tal que

$$C'\Omega C = I$$

donde I es la matriz identidad.

Sea v un vector $(N \times 1)$ de variables aleatorias

$$\begin{aligned} v &= Cu \\ (Cu)' \Omega (Cu) &= u' C' \Omega Cu \\ (Cu)' \Omega (Cu) &= u' u \\ v' \Omega v &= u' u \end{aligned} \tag{1.15}$$

Como la matriz Ω es definida positiva se puede expresar como $(\Omega^{\frac{1}{2}})^2 = \Omega$
Reescribiendo la ecuación (1.15) tenemos

$$v' \Omega^{\frac{1}{2}} \Omega^{\frac{1}{2}} v = u' u \tag{1.16}$$

$$u = \Omega^{\frac{1}{2}} v \tag{1.17}$$

o alternativamente

$$v = \Omega^{-\frac{1}{2}} u \tag{1.18}$$

Con este resultado el término de perturbación de la ecuación (1.13) puede ser expresado como

$$\epsilon = B^{-1} \Omega^{\frac{1}{2}} v \tag{1.19}$$

Substituyendo el resultado de la ecuación (1.19) en la ecuación (1.10) obtenemos

$$AY = X\beta + B^{-1} \Omega^{\frac{1}{2}} v \tag{1.20}$$

o, alternativamente

$$v = \Omega^{-\frac{1}{2}} B (AY - X\beta) \tag{1.21}$$

En esta expresión no lineal (en los parámetros), v es un vector de términos de error independientes y con distribución normal estándar. Por lo tanto la ecuación (1.21) corresponde a la expresión usual para la forma implícita de modelos no lineales,

$$v = f(Y, X, \theta)$$

En la construcción de la función de verosimilitud para Y , el Jacobiano es el concepto que permite relacionar las distribuciones de Y y de v , a través de la expresión (1.20). El Jacobiano para la transformación del vector de variables aleatorias v en el vector de variables aleatorias Y es:

$$J = \left| \frac{\partial v}{\partial Y} \right|$$

$$J = \left| \frac{\partial(\Omega^{-\frac{1}{2}}B(A Y - X\beta))}{\partial Y} \right| \quad (1.22)$$

$$J = \left| \Omega^{-\frac{1}{2}}BA \right| = \left| \Omega^{-\frac{1}{2}} \right| |B| |A| \quad (1.23)$$

La función de densidad de la distribución normal multivariada está dada por la siguiente ecuación

$$f(x) = \frac{1}{\sqrt{|\Sigma|}(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}. \quad (1.24)$$

Donde

x : es el vector de p variables aleatorias,

$\mu = E(x)$: es el valor esperado de la variable aleatoria x

Σ_x : es la matriz de varianzas y de covarianzas de la variable aleatoria x .

Si se supone que $u \sim N_N(0, \Omega)$ la función de densidad de probabilidad conjunta

para el término de error u es:

$$f(u) = \frac{1}{|\Omega|^{\frac{1}{2}} (2\pi)^{\frac{1}{2}N}} e^{-\frac{1}{2}u'\Omega^{-1}u}. \quad (1.25)$$

Para realizar el cambio de variable a v reemplazamos en la expresión anterior el resultado de la expresión (1.17) y multiplicamos por el Jacobiano,

$$J = \left| \frac{\partial u}{\partial v} \right| = \left| \Omega^{\frac{1}{2}} \right|$$

con lo cual obtenemos

$$f(v) = \frac{\left| \Omega^{\frac{1}{2}} \right|}{|\Omega|^{\frac{1}{2}} (2\pi)^{\frac{1}{2}N}} e^{-\frac{1}{2}v'\Omega^{\frac{1}{2}}\Omega^{-1}\Omega^{\frac{1}{2}}v} \quad (1.26)$$

$$f(v) = \frac{\left| \Omega^{\frac{1}{2}} \right|}{|\Omega|^{\frac{1}{2}} (2\pi)^{\frac{1}{2}N}} e^{-\frac{1}{2}v'\Omega^{\frac{1}{2}}\Omega^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}\Omega^{\frac{1}{2}}v} \quad (1.27)$$

$$f(v) = \frac{1}{(2\pi)^{\frac{1}{2}N}} e^{-\frac{1}{2}v'v} \quad (1.28)$$

Realizamos la transformación de variable de v a y

$$f(y) = (2\pi)^{-\frac{1}{2}N} e^{-\frac{1}{2}v'v} \left| \frac{\partial v}{\partial y} \right| \quad (1.29)$$

$$f(y) = (2\pi)^{-\frac{1}{2}N} e^{-\frac{1}{2}v'v} \left| \Omega^{-\frac{1}{2}} \right| |B| |A| \quad (1.30)$$

La función log-verosimilitud L que se obtiene a partir de la expresión anterior es:

$$L = \ln(f(y)) \quad (1.31)$$

$$L = -\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2}v'v - \frac{N}{2} \ln 2 \quad (1.32)$$

con

$$v = \Omega^{-\frac{1}{2}}B(Ay - X\beta)$$

$$A = I - \rho W_1$$

$$B = I - \lambda W_2$$

Para la obtención de las derivadas parciales de esta expresión con respecto a los elementos del vector de parámetros, se comienza con unos resultados básicos.

1.4. Derivadas parciales con respecto al parámetro

ρ

$$\frac{\partial(\rho W_1)}{\partial \rho} = W_1 \tag{1.33}$$

$$\begin{aligned} \frac{\partial(A)}{\partial \rho} &= \frac{\partial(I - \rho W_1)}{\partial \rho} \\ \frac{\partial(A)}{\partial \rho} &= \frac{\partial I}{\partial \rho} - \frac{\partial(\rho W_1)}{\partial \rho} \\ \frac{\partial(A)}{\partial \rho} &= -W_1 \end{aligned} \tag{1.34}$$

Sea C la matriz de cofactores de la matriz A y $C'=(Adj A)$ la matriz adjunta de A . Como A es invertible $|A| \neq 0$

$$\begin{aligned} AC' &= |A| I \\ A(Adj A) &= |A| I \\ A^{-1}A(Adj A) &= |A| A^{-1} \\ (Adj A) &= |A| A^{-1} \end{aligned} \tag{1.35}$$

Ahora se considera la derivada del determinante, para lo cual se toma una función real definida como $f(Y) = |Y|$, donde $Y = (y_{ij})$ y c_{ij} es el cofactor de y_{ij} entonces se tiene que

$$f(Y) = |Y| = \sum_{i=1}^m c_{ij} y_{ij}$$

Como c_{1j}, \dots, c_{mj} no dependen de y_{ij} , tenemos que

$$\frac{\partial f(Y)}{\partial y_{ij}} = c_{ij}$$

De estas derivadas parciales se obtiene el siguiente diferencial

$$df(Y) = d|Y| = \sum_{i=1}^m \sum_{j=1}^m c_{ij} y_{ij} = tr(Adj Y) dY$$

Por lo tanto

$$d|A| = tr(Adj A) dA$$

Ahora si se reemplaza $(AdjA)$ por la expresión obtenida en la ecuación (1.35), se obtiene

$$d|A| = |A| tr A^{-1} dA$$

$$\begin{aligned} \frac{\partial(\ln |A|)}{\partial \rho} &= \frac{1}{|A|} \frac{\partial |A|}{\partial \rho} \\ \frac{\partial(\ln |A|)}{\partial \rho} &= \frac{1}{|A|} |A| tr A^{-1} \frac{\partial(A)}{\partial \rho} \\ \frac{\partial(\ln |A|)}{\partial \rho} &= tr A^{-1} \frac{\partial(A)}{\partial \rho} \\ \frac{\partial(\ln |A|)}{\partial \rho} &= tr A^{-1} (-W_1) \\ \frac{\partial(\ln |A|)}{\partial \rho} &= -tr A^{-1} W_1 \end{aligned} \tag{1.36}$$

$$\begin{aligned} \frac{\partial v}{\partial \rho} &= \frac{\partial(\Omega^{-\frac{1}{2}} B(Ay - X\beta))}{\partial \rho} \\ \frac{\partial v}{\partial \rho} &= \Omega^{-\frac{1}{2}} B \frac{\partial((Ay - X\beta))}{\partial \rho} \\ \frac{\partial v}{\partial \rho} &= \Omega^{-\frac{1}{2}} B \frac{\partial A}{\partial \rho} y \\ \frac{\partial v}{\partial \rho} &= \Omega^{-\frac{1}{2}} B (-W_1) y \\ \frac{\partial v}{\partial \rho} &= -\Omega^{-\frac{1}{2}} B W_1 y \end{aligned} \tag{1.37}$$

$$\begin{aligned}
\frac{\partial v'}{\partial \rho} &= \left(\frac{\partial v}{\partial \rho} \right)' \\
\frac{\partial v'}{\partial \rho} &= (-\Omega^{-\frac{1}{2}} B W_1 y)' \\
\frac{\partial v'}{\partial \rho} &= -y' W_1' B' \Omega^{-\frac{1}{2}}
\end{aligned} \tag{1.38}$$

$$\begin{aligned}
\frac{\partial v'v}{\partial \rho} &= v' \left(\frac{\partial v}{\partial \rho} \right) + \left(\frac{\partial v'}{\partial \rho} \right) v \\
\frac{\partial v'v}{\partial \rho} &= v'(-\Omega^{-\frac{1}{2}} B W_1 y) + (-\Omega^{-\frac{1}{2}} B W_1 y)' v \\
\frac{\partial v'v}{\partial \rho} &= v'(-\Omega^{-\frac{1}{2}} B W_1 y) + v'(-\Omega^{-\frac{1}{2}} B W_1 y) \\
\frac{\partial v'v}{\partial \rho} &= -2v' \Omega^{-\frac{1}{2}} B W_1 y
\end{aligned} \tag{1.39}$$

$$\begin{aligned}
I &= A A^{-1} \\
\frac{\partial I}{\partial \rho} &= \frac{\partial A A^{-1}}{\partial \rho} \\
0 &= \frac{\partial A}{\partial \rho} A^{-1} + A \frac{\partial A^{-1}}{\partial \rho} \\
A \frac{\partial A^{-1}}{\partial \rho} &= -\frac{\partial A}{\partial \rho} A^{-1} \\
\frac{\partial A^{-1}}{\partial \rho} &= -A^{-1} \frac{\partial A}{\partial \rho} A^{-1} \\
\frac{\partial A^{-1}}{\partial \rho} &= -A^{-1} (-W_1) A^{-1} \\
\frac{\partial A^{-1}}{\partial \rho} &= A^{-1} W_1 A^{-1}
\end{aligned} \tag{1.40}$$

$$\begin{aligned}
\frac{\partial(\text{tr}(A^{-1}W_1))}{\partial\rho} &= \text{tr} \left(\frac{\partial(A^{-1}W_1)}{\partial\rho} \right) \\
\frac{\partial(\text{tr}(A^{-1}W_1))}{\partial\rho} &= \text{tr} \left(\frac{\partial A^{-1}}{\partial\rho} W_1 + A^{-1} \frac{\partial W_1}{\partial\rho} \right) \\
\frac{\partial(\text{tr}(A^{-1}W_1))}{\partial\rho} &= \text{tr} (A^{-1}W_1 A^{-1}W_1 + 0) \\
\frac{\partial(\text{tr}(A^{-1}W_1))}{\partial\rho} &= \text{tr} (A^{-1}W_1 A^{-1}W_1)
\end{aligned} \tag{1.41}$$

1.5. Derivadas parciales con respecto al parámetro

λ

$$\frac{\partial(\lambda W_2)}{\partial\lambda} = W_2 \tag{1.42}$$

$$\begin{aligned}
\frac{\partial(B)}{\partial\lambda} &= \frac{\partial(I - \lambda W_2)}{\partial\lambda} \\
\frac{\partial(B)}{\partial\lambda} &= \frac{\partial I}{\partial\lambda} - \frac{\partial(\lambda W_2)}{\partial\lambda} \\
\frac{\partial(B)}{\partial\lambda} &= -W_2
\end{aligned} \tag{1.43}$$

$$\begin{aligned}
\frac{\partial(\ln |B|)}{\partial\lambda} &= \frac{1}{|B|} \frac{\partial |B|}{\partial\lambda} \\
\frac{\partial(\ln |B|)}{\partial\lambda} &= \frac{1}{|B|} |B| \operatorname{tr} B^{-1} \frac{\partial B}{\partial\lambda} \\
\frac{\partial(\ln |B|)}{\partial\lambda} &= \operatorname{tr} B^{-1} \frac{\partial(B)}{\partial\lambda} \\
\frac{\partial(\ln |B|)}{\partial\lambda} &= \operatorname{tr} B^{-1} (-W_2) \\
\frac{\partial(\ln |B|)}{\partial\lambda} &= -\operatorname{tr} B^{-1} W_2
\end{aligned} \tag{1.44}$$

$$\begin{aligned}
\frac{\partial v}{\partial\lambda} &= \frac{\partial(\Omega^{-\frac{1}{2}} B(Ay - X\beta))}{\partial\lambda} \\
\frac{\partial v}{\partial\lambda} &= \Omega^{-\frac{1}{2}} \frac{\partial B}{\partial\lambda} (Ay - X\beta) \\
\frac{\partial v}{\partial\lambda} &= \Omega^{-\frac{1}{2}} (-W_2) (Ay - X\beta) \\
\frac{\partial v}{\partial\lambda} &= -\Omega^{-\frac{1}{2}} W_2 (Ay - X\beta)
\end{aligned} \tag{1.45}$$

$$\begin{aligned}
\frac{\partial v'}{\partial\lambda} &= \left(\frac{\partial v}{\partial\lambda} \right)' \\
\frac{\partial v'}{\partial\lambda} &= (-\Omega^{-\frac{1}{2}} W_2 (Ay - X\beta))' \\
\frac{\partial v'}{\partial\lambda} &= -(Ay - X\beta)' W_2' \Omega^{-\frac{1}{2}}
\end{aligned} \tag{1.46}$$

$$\begin{aligned}
\frac{\partial v'v}{\partial\lambda} &= v' \left(\frac{\partial v}{\partial\lambda} \right) + \left(\frac{\partial v'}{\partial\lambda} \right) v \\
\frac{\partial v'v}{\partial\lambda} &= v' (-\Omega^{-\frac{1}{2}} W_2 (Ay - X\beta)) + (-\Omega^{-\frac{1}{2}} W_2 (Ay - X\beta))' v \\
\frac{\partial v'v}{\partial\lambda} &= v' (-\Omega^{-\frac{1}{2}} W_2 (Ay - X\beta)) + v' (-\Omega^{-\frac{1}{2}} W_2 (Ay - X\beta)) \\
\frac{\partial v'v}{\partial\lambda} &= -2v' \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta)
\end{aligned} \tag{1.47}$$

$$\begin{aligned}
I &= BB^{-1} \\
\frac{\partial I}{\partial \lambda} &= \frac{\partial BB^{-1}}{\partial \lambda} \\
0 &= \frac{\partial B}{\partial \lambda} B^{-1} + B \frac{\partial B^{-1}}{\partial \lambda} \\
B \frac{\partial B^{-1}}{\partial \lambda} &= -\frac{\partial B}{\partial \lambda} B^{-1} \\
\frac{\partial B^{-1}}{\partial \lambda} &= -B^{-1} \frac{\partial B}{\partial \lambda} B^{-1} \\
\frac{\partial B^{-1}}{\partial \lambda} &= -B^{-1}(-W_2)B^{-1} \\
\frac{\partial B^{-1}}{\partial \lambda} &= B^{-1}W_2B^{-1}
\end{aligned} \tag{1.48}$$

$$\begin{aligned}
\frac{\partial(\text{tr}(B^{-1}W_2))}{\partial \lambda} &= \text{tr} \left(\frac{\partial(B^{-1}W_2)}{\partial \lambda} \right) \\
\frac{\partial(\text{tr}(B^{-1}W_2))}{\partial \lambda} &= \text{tr} \left(\frac{\partial B^{-1}}{\partial \lambda} W_2 + B^{-1} \frac{\partial W_2}{\partial \lambda} \right) \\
\frac{\partial(\text{tr}(B^{-1}W_2))}{\partial \lambda} &= \text{tr} (B^{-1}W_2B^{-1}W_2 + 0) \\
\frac{\partial(\text{tr}(B^{-1}W_2))}{\partial \lambda} &= \text{tr} (B^{-1}W_2B^{-1}W_2)
\end{aligned} \tag{1.49}$$

1.6. Derivadas parciales con respecto al parámetro

β

$$\begin{aligned}
 \frac{\partial v}{\partial \beta} &= \frac{\partial(\Omega^{-\frac{1}{2}}B(Ay - X\beta))}{\partial \beta} \\
 \frac{\partial v}{\partial \beta} &= \Omega^{-\frac{1}{2}}B \frac{\partial((Ay - X\beta))}{\partial \beta} \\
 \frac{\partial v}{\partial \beta} &= \Omega^{-\frac{1}{2}}B(-X) \\
 \frac{\partial v}{\partial \beta} &= -\Omega^{-\frac{1}{2}}BX \tag{1.50}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial v}{\partial \beta'} &= \frac{\partial(\Omega^{-\frac{1}{2}}B(Ay - X\beta))}{\partial \beta'} \\
 \frac{\partial v}{\partial \beta'} &= \frac{\partial((Ay - X\beta)'B'\Omega^{-\frac{1}{2}})}{\partial \beta'} \\
 \frac{\partial v}{\partial \beta'} &= (-X)'B'\Omega^{-\frac{1}{2}} \\
 \frac{\partial v}{\partial \beta'} &= -X'B'\Omega^{-\frac{1}{2}} \tag{1.51}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial v'}{\partial \beta} &= \left(\frac{\partial v}{\partial \beta} \right)' \\
 \frac{\partial v'}{\partial \beta} &= (-\Omega^{-\frac{1}{2}}BX)' \\
 \frac{\partial v'}{\partial \beta} &= -X'B'\Omega^{-\frac{1}{2}} \tag{1.52}
 \end{aligned}$$

$$\begin{aligned}
\frac{\partial v'}{\partial \beta'} &= \left(\frac{\partial v}{\partial \beta'} \right)' \\
\frac{\partial v'}{\partial \beta'} &= (-X' B' \Omega^{-\frac{1}{2}})' \\
\frac{\partial v'}{\partial \beta'} &= -\Omega^{-\frac{1}{2}} B X
\end{aligned} \tag{1.53}$$

$$\begin{aligned}
\frac{\partial v'v}{\partial \beta} &= v' \left(\frac{\partial v}{\partial \beta} \right) + \left(\frac{\partial v'}{\partial \beta} \right) v \\
\frac{\partial v'v}{\partial \beta} &= v'(-\Omega^{-\frac{1}{2}} B X) + (-\Omega^{-\frac{1}{2}} B X)' v \\
\frac{\partial v'v}{\partial \beta} &= v'(-\Omega^{-\frac{1}{2}} B X) + v'(-\Omega^{-\frac{1}{2}} B X) \\
\frac{\partial v'v}{\partial \beta} &= -2v' \Omega^{-\frac{1}{2}} B X
\end{aligned} \tag{1.54}$$

$$\begin{aligned}
\frac{\partial v'v}{\partial \beta'} &= v' \left(\frac{\partial v}{\partial \beta'} \right) + \left(\frac{\partial v'}{\partial \beta'} \right) v \\
\frac{\partial v'v}{\partial \beta'} &= v'(-X' B' \Omega^{-\frac{1}{2}}) + (-\Omega^{-\frac{1}{2}} B X) v \\
\frac{\partial v'v}{\partial \beta'} &= v'(-X' B' \Omega^{-\frac{1}{2}}) + v'(-X' B' \Omega^{-\frac{1}{2}}) \\
\frac{\partial v'v}{\partial \beta'} &= -2v' X' B' \Omega^{-\frac{1}{2}}
\end{aligned} \tag{1.55}$$

1.7. Derivadas parciales con respecto al parámetro

α_p

Sea $\Omega = h_i(z\alpha)$ con $h_i = \sigma^2 + z_{i1}\alpha_1 + \dots + z_{ip}\alpha_p$. Esto puede ser expresado matricialmente así:

$$\Omega = \begin{bmatrix} h_1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & h_i & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & h_p \end{bmatrix} \quad (1.56)$$

$$z = \begin{bmatrix} z_{11} & \dots & z_{1i} & \dots & z_{1p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{i1} & \dots & z_{ii} & \dots & z_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{p1} & \dots & z_{pi} & \dots & z_{pp} \end{bmatrix} \quad (1.57)$$

$$\alpha' = [\alpha_1 \quad \dots \quad \alpha_i \quad \dots \quad \alpha_p]$$

De tal manera que la matriz Ω puede ser escrita como $\Omega = \sigma^2 I + z\alpha I$

Se define

$$\frac{\partial \Omega}{\partial \alpha_p} = H_p$$

$$\begin{aligned} \frac{\partial(Ln|\Omega|)}{\partial \alpha_p} &= tr\Omega^{-1} \frac{\partial \Omega}{\partial \alpha_p} \\ \frac{\partial(Ln|\Omega|)}{\partial \alpha_p} &= tr\Omega^{-1} H_p \end{aligned} \quad (1.58)$$

$$\begin{aligned}
\frac{\partial v}{\partial \alpha_p} &= \frac{\partial(\Omega^{-\frac{1}{2}}B(Ay - X\beta))}{\partial \alpha_p} \\
\frac{\partial v}{\partial \alpha_p} &= \frac{\partial(\Omega^{-\frac{1}{2}})}{\partial \alpha_p} B(Ay - X\beta) \\
\frac{\partial v}{\partial \alpha_p} &= -\frac{1}{2}\Omega^{-\frac{3}{2}}\frac{\partial \Omega}{\partial \alpha_p} B(Ay - X\beta) \\
\frac{\partial v}{\partial \alpha_p} &= -\frac{1}{2}\Omega^{-\frac{3}{2}}H_p B(Ay - X\beta)
\end{aligned} \tag{1.59}$$

$$\begin{aligned}
\frac{\partial v'}{\partial \alpha_p} &= \left(\frac{\partial v}{\partial \alpha_p} \right)' \\
\frac{\partial v'}{\partial \alpha_p} &= -\frac{1}{2} \left(\Omega^{-\frac{3}{2}} H_p B(Ay - X\beta) \right)' \\
\frac{\partial v'}{\partial \alpha_p} &= -\frac{1}{2} [H_p B(Ay - X\beta)]' \Omega^{-\frac{3}{2}}
\end{aligned} \tag{1.60}$$

$$\begin{aligned}
\frac{\partial v'v}{\partial \alpha_p} &= v' \frac{\partial v}{\partial \alpha_p} + \frac{\partial v'}{\partial \alpha_p} v \\
\frac{\partial v'v}{\partial \alpha_p} &= v' \left(-\frac{1}{2} \Omega^{-\frac{3}{2}} H_p B(Ay - X\beta) \right) - \frac{1}{2} \left(\Omega^{-\frac{3}{2}} H_p B(Ay - X\beta) \right)' v \\
\frac{\partial v'v}{\partial \alpha_p} &= -\frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B(Ay - X\beta) - \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B(Ay - X\beta) \\
\frac{\partial v'v}{\partial \alpha_p} &= -v' \Omega^{-\frac{3}{2}} H_p B(Ay - X\beta)
\end{aligned} \tag{1.61}$$

1.8. Primeras derivadas parciales de la función de verosimilitud

$$\begin{aligned}\frac{\partial L}{\partial \beta} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \beta} \\ \frac{\partial L}{\partial \beta} &= -\frac{1}{2} \left(\frac{\partial v'v}{\partial \beta} \right)\end{aligned}$$

Tomando el resultado obtenido en la expresión (1.54), se tiene

$$\begin{aligned}\frac{\partial L}{\partial \beta} &= -\frac{1}{2} (-2v'\Omega^{-\frac{1}{2}}BX) \\ \frac{\partial L}{\partial \beta} &= v'\Omega^{-\frac{1}{2}}BX\end{aligned}\tag{1.62}$$

$$\begin{aligned}\frac{\partial L}{\partial \beta'} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \beta'} \\ \frac{\partial L}{\partial \beta'} &= -\frac{1}{2} \left(\frac{\partial v'v}{\partial \beta'} \right)\end{aligned}$$

Sustituyendo el resultado obtenido en la expresión (1.55), se tiene

$$\begin{aligned}\frac{\partial L}{\partial \beta'} &= -\frac{1}{2} (-2v'X'B'\Omega^{-\frac{1}{2}}) \\ \frac{\partial L}{\partial \beta'} &= v'X'B'\Omega^{-\frac{1}{2}} \\ \frac{\partial L}{\partial \beta'} &= \Omega^{-\frac{1}{2}}BXv\end{aligned}\tag{1.63}$$

$$\begin{aligned}\frac{\partial L}{\partial \rho} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \rho} \\ \frac{\partial L}{\partial \rho} &= \frac{\partial (\ln |A| - \frac{1}{2} v'v)}{\partial \rho} \\ \frac{\partial L}{\partial \rho} &= \frac{\partial \ln |A|}{\partial \rho} - \frac{1}{2} \frac{\partial v'v}{\partial \rho}\end{aligned}$$

Reemplazando los resultados obtenidos en las expresiones (1.36) y (1.39), se tiene

$$\begin{aligned}\frac{\partial L}{\partial \rho} &= -\text{tr} A^{-1} W_1 - \frac{1}{2} (-2v' \Omega^{-\frac{1}{2}} B W_1 y) \\ \frac{\partial L}{\partial \rho} &= -\text{tr} A^{-1} W_1 + v' \Omega^{-\frac{1}{2}} B W_1 y\end{aligned}\tag{1.64}$$

$$\begin{aligned}\frac{\partial L}{\partial \lambda} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \lambda} \\ \frac{\partial L}{\partial \lambda} &= \frac{\partial (\ln |B| - \frac{1}{2} v'v)}{\partial \lambda} \\ \frac{\partial L}{\partial \lambda} &= \frac{\partial \ln |B|}{\partial \lambda} - \frac{1}{2} \frac{\partial v'v}{\partial \lambda}\end{aligned}$$

Incorporando los resultados obtenidos en las expresiones (1.44) y (1.47), se tiene

$$\begin{aligned}\frac{\partial L}{\partial \lambda} &= -\text{tr} B^{-1} W_2 - \frac{1}{2} (-2v' \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta)) \\ \frac{\partial L}{\partial \lambda} &= -\text{tr} B^{-1} W_2 + v' \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta)\end{aligned}\tag{1.65}$$

$$\begin{aligned}\frac{\partial L}{\partial \alpha_p} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \alpha_p} \\ \frac{\partial L}{\partial \alpha_p} &= \frac{\partial \left(-\frac{1}{2} \ln |\Omega| - \frac{1}{2} v'v \right)}{\partial \alpha_p} \\ \frac{\partial L}{\partial \alpha_p} &= -\frac{1}{2} \frac{\partial \ln |\Omega|}{\partial \alpha_p} - \frac{1}{2} \frac{\partial v'v}{\partial \alpha_p}\end{aligned}$$

Con los resultados obtenidos en las expresiones (1.61) y (1.58), se tiene

$$\begin{aligned}\frac{\partial L}{\partial \alpha_p} &= -\frac{1}{2}tr\Omega^{-1}H_p - \frac{1}{2}(-v'\Omega^{-\frac{3}{2}}H_pB(Ay - X\beta)) \\ \frac{\partial L}{\partial \alpha_p} &= -\frac{1}{2}tr\Omega^{-1}H_p + \frac{1}{2}v'\Omega^{-\frac{3}{2}}H_pB(Ay - X\beta)\end{aligned}\quad (1.66)$$

1.9. Segundas derivadas parciales de la función de verosimilitud

1.9.1. Para los elementos de la diagonal

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta \partial \beta'} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \beta \partial \beta'} \\ \frac{\partial^2 L}{\partial \beta \partial \beta'} &= \frac{\partial}{\partial \beta} \left(\frac{\partial L}{\partial \beta'} \right)\end{aligned}$$

Observando el resultado obtenido en la expresión (1.63), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta \partial \beta'} &= \frac{\partial(\Omega^{-\frac{1}{2}}BXv)}{\partial \beta} \\ \frac{\partial^2 L}{\partial \beta \partial \beta'} &= \Omega^{-\frac{1}{2}}BX \frac{\partial v}{\partial \beta}\end{aligned}$$

Aplicando el resultado de (1.50)

$$\begin{aligned}
\frac{\partial^2 L}{\partial \beta \partial \beta'} &= \Omega^{-\frac{1}{2}} B X (-\Omega^{-\frac{1}{2}} B X) \\
\frac{\partial^2 L}{\partial \beta \partial \beta'} &= -X' B' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} B X \\
\frac{\partial^2 L}{\partial \beta \partial \beta'} &= -X' B' \Omega^{-1} B X
\end{aligned} \tag{1.67}$$

$$\begin{aligned}
\frac{\partial^2 L}{\partial \rho^2} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2 \right)}{\partial \rho^2} \\
\frac{\partial^2 L}{\partial \rho^2} &= \frac{\partial}{\partial \rho} \left(\frac{\partial L}{\partial \rho} \right)
\end{aligned}$$

Sustituyendo el resultado obtenido en la expresión (1.64), se tiene

$$\begin{aligned}
\frac{\partial^2 L}{\partial \rho^2} &= \frac{\partial(-tr A^{-1} W_1 + v' \Omega^{-\frac{1}{2}} B W_1 y)}{\partial \rho} \\
\frac{\partial^2 L}{\partial \rho^2} &= \frac{\partial(-tr A^{-1} W_1)}{\partial \rho} + \frac{\partial(v' \Omega^{-\frac{1}{2}} B W_1 y)}{\partial \rho} \\
\frac{\partial^2 L}{\partial \rho^2} &= -\frac{\partial(tr A^{-1} W_1)}{\partial \rho} + \frac{\partial v'}{\partial \rho} \Omega^{-\frac{1}{2}} B W_1 y
\end{aligned}$$

Usando los resultados de (1.41) y (1.38)

$$\begin{aligned}
\frac{\partial^2 L}{\partial \rho^2} &= -tr (A^{-1} W_1 A^{-1} W_1) + (-y' W_1' B' \Omega^{-\frac{1}{2}}) \Omega^{-\frac{1}{2}} B W_1 y \\
\frac{\partial^2 L}{\partial \rho^2} &= -tr (A^{-1} W_1 A^{-1} W_1) - y' W_1' B' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} B W_1 y \\
\frac{\partial^2 L}{\partial \rho^2} &= -tr (A^{-1} W_1 A^{-1} W_1) - y' W_1' B' \Omega^{-1} B W_1 y \\
\frac{\partial^2 L}{\partial \rho^2} &= -tr (A^{-1} W_1 A^{-1} W_1) - (B W_1 y)' \Omega^{-1} B W_1 y
\end{aligned} \tag{1.68}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda^2} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \lambda^2} \\ \frac{\partial^2 L}{\partial \lambda^2} &= \frac{\partial}{\partial \lambda} \left(\frac{\partial L}{\partial \lambda} \right)\end{aligned}$$

Aplicando el resultado obtenido en la expresión (1.65), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda^2} &= \frac{\partial(-tr B^{-1} W_2 + v' \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta))}{\partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda^2} &= \frac{\partial(-tr B^{-1} W_2)}{\partial \lambda} + \frac{\partial(v' \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta))}{\partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda^2} &= -\frac{\partial(tr B^{-1} W_2)}{\partial \lambda} + \frac{\partial v'}{\partial \lambda} \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta)\end{aligned}$$

Reemplazando los resultados de (1.49) y (1.46)

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda^2} &= -tr (B^{-1} W_2 B^{-1} W_2) + -(Ay - X\beta)' W_2' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta) \\ \frac{\partial^2 L}{\partial \lambda^2} &= -tr (B^{-1} W_2 B^{-1} W_2) - (Ay - X\beta)' W_2' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta) \\ \frac{\partial^2 L}{\partial \lambda^2} &= -tr (B^{-1} W_2 B^{-1} W_2) - (Ay - X\beta)' W_2' \Omega^{-1} W_2 (Ay - X\beta) \\ \frac{\partial^2 L}{\partial \lambda^2} &= -tr (B^{-1} W_2 B^{-1} W_2) - [W_2 (Ay - X\beta)]' \Omega^{-1} W_2 (Ay - X\beta)\end{aligned}\tag{1.69}$$

1.9.2. Para los términos cruzados

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta \partial \rho} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \beta \partial \rho} \\ \frac{\partial^2 L}{\partial \beta \partial \rho} &= \frac{\partial}{\partial \beta} \left(\frac{\partial L}{\partial \rho} \right)\end{aligned}$$

Sustituyendo el resultado obtenido en la expresión (1.64), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta \partial \rho} &= \frac{\partial(-tr A^{-1} W_1 + v' \Omega^{-\frac{1}{2}} B W_1 y)}{\partial \beta} \\ \frac{\partial^2 L}{\partial \beta \partial \rho} &= \frac{\partial(-tr A^{-1} W_1)}{\partial \beta} + \frac{\partial(v' \Omega^{-\frac{1}{2}} B W_1 y)}{\partial \beta} \\ \frac{\partial^2 L}{\partial \beta \partial \rho} &= -\frac{\partial(tr A^{-1} W_1)}{\partial \beta} + \frac{\partial v'}{\partial \beta} \Omega^{-\frac{1}{2}} B W_1 y\end{aligned}$$

Usando el resultado de (1.52)

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta \partial \rho} &= 0 + (-X' B' \Omega^{-\frac{1}{2}}) \Omega^{-\frac{1}{2}} B W_1 y \\ \frac{\partial^2 L}{\partial \beta \partial \rho} &= -X' B' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} B W_1 y \\ \frac{\partial^2 L}{\partial \beta \partial \rho} &= -X' B' \Omega^{-1} B W_1 y \\ \frac{\partial^2 L}{\partial \beta \partial \rho} &= -(X B)' \Omega^{-1} B W_1 y\end{aligned}\tag{1.70}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \rho \partial \beta'} &= \frac{\partial\left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2\right)}{\partial \rho \partial \beta'} \\ \frac{\partial^2 L}{\partial \rho \partial \beta'} &= \frac{\partial}{\partial \rho} \left(\frac{\partial L}{\partial \beta'} \right)\end{aligned}$$

Tomando el resultado obtenido en la expresión (1.63), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \rho \partial \beta'} &= \frac{\partial(\Omega^{-\frac{1}{2}} B X v)}{\partial \rho} \\ \frac{\partial^2 L}{\partial \rho \partial \beta'} &= \Omega^{-\frac{1}{2}} B X \frac{\partial v}{\partial \rho}\end{aligned}$$

Aplicando la expresión (1.37)

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \rho \partial \beta'} &= \Omega^{-\frac{1}{2}} B X (-\Omega^{-\frac{1}{2}} B W_1 y) \\
 \frac{\partial^2 L}{\partial \rho \partial \beta'} &= -X' B' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} B W_1 y \\
 \frac{\partial^2 L}{\partial \rho \partial \beta'} &= -(X B)' \Omega^{-1} B W_1 y
 \end{aligned} \tag{1.71}$$

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \beta \partial \lambda} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2 \right)}{\partial \beta \partial \lambda} \\
 \frac{\partial^2 L}{\partial \beta \partial \lambda} &= \frac{\partial}{\partial \beta} \left(\frac{\partial L}{\partial \lambda} \right)
 \end{aligned}$$

Incorporando el resultado obtenido en la expresión (1.65), se tiene

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \beta \partial \lambda} &= \frac{\partial (-tr B^{-1} W_2 + v' \Omega^{-\frac{1}{2}} W_2 (A y - X \beta))}{\partial \beta} \\
 \frac{\partial^2 L}{\partial \beta \partial \lambda} &= \frac{\partial (-tr B^{-1} W_2)}{\partial \beta} + \frac{\partial (v' \Omega^{-\frac{1}{2}} W_2 (A y - X \beta))}{\partial \beta} \\
 \frac{\partial^2 L}{\partial \beta \partial \lambda} &= 0 + \frac{\partial v'}{\partial \beta} \Omega^{-\frac{1}{2}} W_2 (A y - X \beta) + v' \Omega^{-\frac{1}{2}} W_2 \frac{\partial (A y - X \beta)}{\partial \beta}
 \end{aligned}$$

Si se trae el resultado de (1.52) se obtiene

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \beta \partial \lambda} &= -X' B' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} W_2 (A y - X \beta) + v' \Omega^{-\frac{1}{2}} W_2 (-X) \\
 \frac{\partial^2 L}{\partial \beta \partial \lambda} &= -X' B' \Omega^{-1} W_2 (A y - X \beta) - v' \Omega^{-\frac{1}{2}} W_2 X \\
 \frac{\partial^2 L}{\partial \beta \partial \lambda} &= -(B X)' \Omega^{-1} W_2 (A y - X \beta) - v' \Omega^{-\frac{1}{2}} W_2 X
 \end{aligned} \tag{1.72}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda \partial \beta'} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \lambda \partial \beta'} \\ \frac{\partial^2 L}{\partial \lambda \partial \beta'} &= \frac{\partial}{\partial \lambda} \left(\frac{\partial L}{\partial \beta'} \right)\end{aligned}$$

Con el resultado obtenido en la expresión (1.63), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda \partial \beta'} &= \frac{\partial(\Omega^{-\frac{1}{2}} B X v)}{\partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda \partial \beta'} &= \Omega^{-\frac{1}{2}} \frac{\partial B}{\partial \lambda} X v + \Omega^{-\frac{1}{2}} B X \frac{\partial v}{\partial \lambda}\end{aligned}$$

Aplicando el resultado de (1.43) y (1.45)

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda \partial \beta'} &= \Omega^{-\frac{1}{2}} (-W_2) X v + \Omega^{-\frac{1}{2}} B X (-\Omega^{-\frac{1}{2}} W_2 (A y - X \beta)) \\ \frac{\partial^2 L}{\partial \lambda \partial \beta'} &= -v' \Omega^{-\frac{1}{2}} - W_2 X - (B X)' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} W_2 (A y - X \beta) \\ \frac{\partial^2 L}{\partial \lambda \partial \beta'} &= -(B X)' \Omega^{-1} W_2 (A y - X \beta) - v' \Omega^{-\frac{1}{2}} W_2 X\end{aligned}\tag{1.73}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta \partial \alpha_p} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \beta \partial \alpha_p} \\ \frac{\partial^2 L}{\partial \beta \partial \alpha_p} &= \frac{\partial}{\partial \beta} \left(\frac{\partial L}{\partial \alpha_p} \right)\end{aligned}$$

Reemplazando el resultado obtenido en la expresión (1.66), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta \partial \alpha_p} &= \frac{\partial \left(-\frac{1}{2} \text{tr} \Omega^{-1} H_p + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B (A y - X \beta) \right)}{\partial \beta} \\ \frac{\partial^2 L}{\partial \beta \partial \alpha_p} &= \frac{1}{2} \frac{\partial v'}{\partial \beta} \Omega^{-\frac{3}{2}} H_p B (A y - X \beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B \frac{\partial (A y - X \beta)}{\partial \beta}\end{aligned}$$

Usando el resultado de (1.52), se obtiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta \partial \alpha_p} &= \frac{1}{2}(-X' B' \Omega^{-\frac{1}{2}}) \Omega^{-\frac{3}{2}} H_p B (A y - X \beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B (-X) \\ \frac{\partial^2 L}{\partial \beta \partial \alpha_p} &= -\frac{1}{2} (B X)' \Omega^{-2} H_p B (A y - X \beta) - \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B X\end{aligned}\quad (1.74)$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \alpha_p \partial \beta'} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2 \right)}{\partial \alpha_p \partial \beta'} \\ \frac{\partial^2 L}{\partial \alpha_p \partial \beta'} &= \frac{\partial}{\partial \alpha_p} \left(\frac{\partial L}{\partial \beta'} \right)\end{aligned}$$

Sustituyendo el resultado obtenido en la expresión (1.63), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \alpha_p \partial \beta'} &= \frac{\partial(\Omega^{-\frac{1}{2}} B X v)}{\partial \alpha_p} \\ \frac{\partial^2 L}{\partial \alpha_p \partial \beta'} &= \frac{\partial(\Omega^{-\frac{1}{2}})}{\partial \alpha_p} B X v + \Omega^{-\frac{1}{2}} B X \frac{\partial v}{\partial \alpha_p}\end{aligned}$$

Tomando el resultado de (1.59)

$$\begin{aligned}\frac{\partial^2 L}{\partial \alpha_p \partial \beta'} &= -\frac{1}{2} \Omega^{-\frac{3}{2}} H_p B X v + \Omega^{-\frac{1}{2}} B X \left(-\frac{1}{2} \Omega^{-\frac{3}{2}} H_p B (A y - X \beta) \right) \\ \frac{\partial^2 L}{\partial \alpha_p \partial \beta'} &= -\frac{1}{2} \Omega^{-\frac{3}{2}} H_p B X v - \frac{1}{2} (B X)' \Omega^{-\frac{1}{2}} \Omega^{-\frac{3}{2}} H_p B (A y - X \beta) \\ \frac{\partial^2 L}{\partial \alpha_p \partial \beta'} &= -\frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B X - \frac{1}{2} (B X)' \Omega^{-2} H_p B (A y - X \beta)\end{aligned}\quad (1.75)$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \rho \partial \lambda} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2 \right)}{\partial \rho \partial \lambda} \\ \frac{\partial^2 L}{\partial \rho \partial \lambda} &= \frac{\partial}{\partial \rho} \left(\frac{\partial L}{\partial \lambda} \right)\end{aligned}$$

Aplicando el resultado obtenido en la expresión (1.65), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \rho \partial \lambda} &= \frac{\partial(-tr B^{-1} W_2 + v' \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta))}{\partial \rho} \\ \frac{\partial^2 L}{\partial \rho \partial \lambda} &= \frac{\partial(-tr B^{-1} W_2)}{\partial \rho} + \frac{\partial(v' \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta))}{\partial \rho} \\ \frac{\partial^2 L}{\partial \rho \partial \lambda} &= 0 + \frac{\partial v'}{\partial \rho} \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta) + v' \Omega^{-\frac{1}{2}} W_2 \frac{\partial(Ay - X\beta)}{\partial \rho}\end{aligned}$$

Con los resultados de (1.38) y (1.34) se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \rho \partial \lambda} &= (-y' W_1' B' \Omega^{-\frac{1}{2}}) \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta) + v' \Omega^{-\frac{1}{2}} W_2 (-W_1) y \\ \frac{\partial^2 L}{\partial \rho \partial \lambda} &= -(B W_1 y)' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} W_2 (Ay - X\beta) - v' \Omega^{-\frac{1}{2}} W_2 W_1 y \\ \frac{\partial^2 L}{\partial \rho \partial \lambda} &= -(B W_1 y)' \Omega^{-1} W_2 (Ay - X\beta) - v' \Omega^{-\frac{1}{2}} W_2 W_1 y \\ \frac{\partial^2 L}{\partial \rho \partial \lambda} &= -[W_2 (Ay - X\beta)]' \Omega^{-1} B W_1 y - v' \Omega^{-\frac{1}{2}} W_2 W_1 y\end{aligned}\tag{1.76}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda \partial \rho} &= \frac{\partial\left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2\right)}{\partial \lambda \partial \rho} \\ \frac{\partial^2 L}{\partial \lambda \partial \rho} &= \frac{\partial}{\partial \lambda} \left(\frac{\partial L}{\partial \rho} \right)\end{aligned}$$

Usando el resultado obtenido en la expresión (1.64), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda \partial \rho} &= \frac{\partial(-tr A^{-1} W_1 + v' \Omega^{-\frac{1}{2}} B W_1 y)}{\partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda \partial \rho} &= \frac{\partial(-tr A^{-1} W_1)}{\partial \lambda} + \frac{\partial(v' \Omega^{-\frac{1}{2}} B W_1 y)}{\partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda \partial \rho} &= 0 + \frac{\partial v'}{\partial \lambda} \Omega^{-\frac{1}{2}} B W_1 y + v' \Omega^{-\frac{1}{2}} \frac{\partial B}{\partial \lambda} W_1 y\end{aligned}$$

Sustituyendo los resultados de (1.46) y (1.43)

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \lambda \partial \rho} &= -(Ay - X\beta)' W_2' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} B W_1 y + v' \Omega^{-\frac{1}{2}} (-W_2) W_1 y \\
 \frac{\partial^2 L}{\partial \lambda \partial \rho} &= -(Ay - X\beta)' W_2' \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} B W_1 y - v' \Omega^{-\frac{1}{2}} W_2 W_1 y \\
 \frac{\partial^2 L}{\partial \lambda \partial \rho} &= -[W_2(Ay - X\beta)]' \Omega^{-1} B W_1 y - v' \Omega^{-\frac{1}{2}} W_2 W_1 y
 \end{aligned} \tag{1.77}$$

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \rho \partial \alpha_p} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2 \right)}{\partial \rho \partial \alpha_p} \\
 \frac{\partial^2 L}{\partial \rho \partial \alpha_p} &= \frac{\partial}{\partial \rho} \left(\frac{\partial L}{\partial \alpha_p} \right)
 \end{aligned}$$

Tomando el resultado obtenido en la expresión (1.66), se tiene

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \rho \partial \alpha_p} &= \frac{\partial \left(-\frac{1}{2} \text{tr} \Omega^{-1} H_p + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B (Ay - X\beta) \right)}{\partial \rho} \\
 \frac{\partial^2 L}{\partial \rho \partial \alpha_p} &= \frac{1}{2} \frac{\partial v'}{\partial \rho} \Omega^{-\frac{3}{2}} H_p B (Ay - X\beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B \frac{\partial (Ay - X\beta)}{\partial \rho}
 \end{aligned}$$

Con el resultado de (1.38) y (1.34)

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \rho \partial \alpha_p} &= \frac{1}{2} (-y' W_1' B' \Omega^{-\frac{1}{2}}) \Omega^{-\frac{3}{2}} H_p B (Ay - X\beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B (-W_1) y \\
 \frac{\partial^2 L}{\partial \rho \partial \alpha_p} &= -\frac{1}{2} (B W_1 y)' \Omega^{-2} H_p B (Ay - X\beta) - \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B W_1 y
 \end{aligned} \tag{1.78}$$

$$\begin{aligned}
 \frac{\partial^2 L}{\partial \alpha_p \partial \rho} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2 \right)}{\partial \alpha_p \partial \rho} \\
 \frac{\partial^2 L}{\partial \alpha_p \partial \rho} &= \frac{\partial}{\partial \alpha_p} \left(\frac{\partial L}{\partial \rho} \right)
 \end{aligned}$$

Aplicando el resultado obtenido en la expresión (1.64), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \alpha_p \partial \rho} &= \frac{\partial(-tr A^{-1} W_1 + v' \Omega^{-\frac{1}{2}} B W_1 y)}{\partial \alpha_p} \\ \frac{\partial^2 L}{\partial \alpha_p \partial \rho} &= \frac{\partial v'}{\partial \alpha_p} \Omega^{-\frac{1}{2}} B W_1 y + v' \frac{\partial \Omega^{-\frac{1}{2}}}{\partial \alpha_p} B W_1 y\end{aligned}$$

Usando el resultado obtenido en (1.60)

$$\begin{aligned}\frac{\partial^2 L}{\partial \alpha_p \partial \rho} &= -\frac{1}{2} [H_p B (Ay - X\beta)]' \Omega^{-\frac{3}{2}} \Omega^{-\frac{1}{2}} B W_1 y + v' \left(-\frac{1}{2} \Omega^{-\frac{3}{2}} H_p B W_1 y\right) \\ \frac{\partial^2 L}{\partial \alpha_p \partial \rho} &= -\frac{1}{2} [H_p B (Ay - X\beta)]' \Omega^{-2} B W_1 y - \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B W_1 y \\ \frac{\partial^2 L}{\partial \alpha_p \partial \rho} &= -\frac{1}{2} (B W_1 y)' \Omega^{-2} H_p B (Ay - X\beta) - \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B W_1 y\end{aligned}\tag{1.79}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda \partial \alpha_p} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v' v - \frac{N}{2} \ln 2\right)}{\partial \lambda \partial \alpha_p} \\ \frac{\partial^2 L}{\partial \lambda \partial \alpha_p} &= \frac{\partial}{\partial \lambda} \left(\frac{\partial L}{\partial \alpha_p} \right)\end{aligned}$$

Tomando el resultado obtenido en la expresión (1.66), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda \partial \alpha_p} &= \frac{\partial \left(-\frac{1}{2} tr \Omega^{-1} H_p + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p B (Ay - X\beta)\right)}{\partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda \partial \alpha_p} &= \frac{1}{2} \frac{\partial v'}{\partial \lambda} \Omega^{-\frac{3}{2}} H_p B (Ay - X\beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_p \frac{\partial B}{\partial \lambda} (Ay - X\beta)\end{aligned}$$

Aplicando el resultado de (1.46) y (1.43)

$$\begin{aligned}\frac{\partial^2 L}{\partial \lambda \partial \alpha_p} &= \frac{1}{2}(-(Ay - X\beta)'W_2'\Omega^{-\frac{1}{2}})\Omega^{-\frac{3}{2}}H_p B(Ay - X\beta) + \frac{1}{2}v'\Omega^{-\frac{3}{2}}H_p(-W_2)(Ay - X\beta) \\ \frac{\partial^2 L}{\partial \lambda \partial \alpha_p} &= -\frac{1}{2}[W_2(Ay - X\beta)]'\Omega^{-2}H_p B(Ay - X\beta) - \frac{1}{2}v'\Omega^{-\frac{3}{2}}H_p W_2(Ay - X\beta)\end{aligned}\tag{1.80}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \alpha_p \partial \lambda} &= \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \alpha_p \partial \lambda} \\ \frac{\partial^2 L}{\partial \alpha_p \partial \lambda} &= \frac{\partial}{\partial \alpha_p} \left(\frac{\partial L}{\partial \lambda} \right)\end{aligned}$$

Sustituyendo el resultado obtenido en la expresión (1.65), se tiene

$$\begin{aligned}\frac{\partial^2 L}{\partial \alpha_p \partial \lambda} &= \frac{\partial (-tr B^{-1}W_2 + v'\Omega^{-\frac{1}{2}}W_2(Ay - X\beta))}{\partial \alpha_p} \\ \frac{\partial^2 L}{\partial \alpha_p \partial \lambda} &= \frac{\partial v'}{\partial \alpha_p} \Omega^{-\frac{1}{2}}W_2(Ay - X\beta) + v' \frac{\partial (\Omega^{-\frac{1}{2}})}{\partial \alpha_p} W_2(Ay - X\beta)\end{aligned}$$

Reemplazando el resultado de (1.60)

$$\begin{aligned}\frac{\partial^2 L}{\partial \alpha_p \partial \lambda} &= \left(-\frac{1}{2}[H_p B(Ay - X\beta)]'\Omega^{-\frac{3}{2}}\right)\Omega^{-\frac{1}{2}}W_2(Ay - X\beta) + v' \left(-\frac{1}{2}\Omega^{-\frac{3}{2}}H_p \right)W_2(Ay - X\beta) \\ \frac{\partial^2 L}{\partial \alpha_p \partial \lambda} &= -\frac{1}{2}[H_p B(Ay - X\beta)]'\Omega^{-2}W_2(Ay - X\beta) - \frac{1}{2}v'\Omega^{-\frac{3}{2}}H_p W_2(Ay - X\beta) \\ \frac{\partial^2 L}{\partial \alpha_p \partial \lambda} &= -\frac{1}{2}[W_2(Ay - X\beta)]'\Omega^{-2}H_p B(Ay - X\beta) - \frac{1}{2}v'\Omega^{-\frac{3}{2}}H_p W_2(Ay - X\beta)\end{aligned}\tag{1.81}$$

$$\frac{\partial^2 L}{\partial \alpha_p \partial \alpha_q} = \frac{\partial \left(-\frac{N}{2} \ln(\pi) - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v - \frac{N}{2} \ln 2 \right)}{\partial \alpha_p \partial \alpha_q}$$

$$\frac{\partial^2 L}{\partial \alpha_p \partial \alpha_q} = \frac{\partial}{\partial \alpha_p} \left(\frac{\partial L}{\partial \alpha_q} \right)$$

Con el resultado obtenido en la expresión (1.66), se tiene

$$\frac{\partial^2 L}{\partial \alpha_p \partial \alpha_q} = \frac{\partial \left(-\frac{1}{2} \text{tr} \Omega^{-1} H_q + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_q B(Ay - X\beta) \right)}{\partial \alpha_p}$$

$$\frac{\partial^2 L}{\partial \alpha_p \partial \alpha_q} = -\frac{1}{2} \frac{\partial (\text{tr} \Omega^{-1})}{\partial \alpha_p} H_q - \frac{1}{2} \text{tr} \Omega^{-1} \frac{\partial H_q}{\partial \alpha_p} + \frac{1}{2} \frac{\partial v'}{\partial \alpha_p} \Omega^{-\frac{3}{2}} H_q B(Ay - X\beta)$$

$$+ \frac{1}{2} v' \frac{\partial (\Omega^{-\frac{3}{2}})}{\partial \alpha_p} H_q B(Ay - X\beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} \frac{\partial H_q}{\partial \alpha_p} B(Ay - X\beta)$$

Usando el resultado de (1.60)

$$\frac{\partial^2 L}{\partial \alpha_p \partial \alpha_q} = -\frac{1}{2} \text{tr} \frac{\partial (\Omega^{-1})}{\partial \alpha_p} H_q - \frac{1}{2} \text{tr} \Omega^{-1} H_{pq} + \frac{1}{2} \left(-\frac{1}{2} [H_p B(Ay - X\beta)]' \Omega^{-\frac{3}{2}} \right) \Omega^{-\frac{3}{2}} H_q B(Ay - X\beta)$$

$$+ \frac{1}{2} v' \left(-\frac{3}{2} \Omega^{-\frac{5}{2}} H_p \right) H_q B(Ay - X\beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_{pq} B(Ay - X\beta)$$

$$\frac{\partial^2 L}{\partial \alpha_p \partial \alpha_q} = \frac{1}{2} \text{tr} \Omega^{-2} H_p H_q - \frac{1}{2} \text{tr} \Omega^{-1} H_{pq} - \frac{1}{4} [H_p B(Ay - X\beta)]' \Omega^{-\frac{3}{2}} H_q B(Ay - X\beta)$$

$$- \frac{3}{4} v' \Omega^{-\frac{5}{2}} H_p H_q B(Ay - X\beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_{pq} B(Ay - X\beta)$$

$$\frac{\partial^2 L}{\partial \alpha_p \partial \alpha_q} = -\frac{1}{2} \text{tr} (-\Omega^{-2}) H_p H_q - \frac{1}{2} \text{tr} \Omega^{-1} H_{pq} - \frac{1}{4} [B(Ay - X\beta)]' \Omega^{-\frac{3}{2}} H_q H_p B(Ay - X\beta)$$

$$- \frac{3}{4} v' \Omega^{-\frac{5}{2}} H_p H_q B(Ay - X\beta) + \frac{1}{2} v' \Omega^{-\frac{3}{2}} H_{pq} B(Ay - X\beta) \quad (1.82)$$

1.10. Distribución Normal Multivariada

A continuación se presenta la construcción de la función de densidad de la distribución normal multivariada, para un vector de variables aleatorias x , a partir de la función de densidad de la distribución normal univariada. Se parte de la siguiente

igualdad

$$ke^{-\frac{1}{2}\alpha(x-\beta)^2} = ke^{-\frac{1}{2}(x-\beta)\alpha(x-\beta)} \quad (1.83)$$

Donde

α : corresponde a una constante positiva

k : es un coeficiente, escogido de tal forma, que la integral de la ecuación (1.83) sobre todo el eje x , sea igual a la unidad

x : es una variable aleatoria.

La función de densidad de una distribución normal multivariada se obtiene considerando x un vector de variables aleatorias, X_1, \dots, X_p

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$$

la constante escalar β es reemplazada por un vector

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix}$$

y la constante positiva α es reemplazada por una matriz definida positiva (simétrica)

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}$$

El cuadrado $\alpha(x - \beta)^2 = (x - \beta)\alpha(x - \beta)$ se reemplaza por la forma cuadrática

$$(x - b)'A(x - b) = \sum_{i,j=1}^p a_{ij}(x_i - b_i)(x_j - b_j) \quad (1.84)$$

Por lo tanto la función de densidad de una distribución normal p -variada es:

$$f(x_1, \dots, x_p) = K e^{-\frac{1}{2}(x-b)'A(x-b)}, \quad (1.85)$$

donde K es escogido de tal forma que

$$K = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_p) dx_p \dots dx_1 = 1 \quad (1.86)$$

$f(x_1, \dots, x_p)$ es no negativa

Dado que la Matriz A es definida positiva,

$$(x - b)'A(x - b) > 0, \quad (1.87)$$

y por lo tanto la densidad está acotada; es decir que,

$$f(x_1, \dots, x_p) \leq K. \quad (1.88)$$

$$\text{Sea } K^* = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-b)'A(x-b)} dx_p \dots dx_1. \quad (1.89)$$

Como la matriz A es definida positiva, existe una matriz no singular C tal que

$$C'AC = I, \quad (1.90)$$

donde I denota la matriz identidad y C' la transpuesta de la matriz C .

Sea

$$x - b = Cy, \quad (1.91)$$

donde

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$$

Luego

$$(x - b)'A(x - b) = (Cy)'ACy = y'C'ACy = y'y \quad (1.92)$$

Dado que se trata de una transformación de variable, es necesario hallar el Jacobiano de dicha transformación, el cual corresponde a la expresión

$$J = \text{mod } |C|, \quad (1.93)$$

$\text{mod } |C|$ es el valor absoluto del determinante de la matriz C . Así la ecuación (1.89) se convierte en

$$K^* = \text{mod } |C| \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}y'y} dy_p \dots dy_1. \quad (1.94)$$

Se tiene que

$$e^{-\frac{1}{2}y'y} = e^{(-\frac{1}{2}\sum_{i=1}^p y_i^2)} = \prod_{i=1}^p e^{-\frac{1}{2}y_i^2}, \quad (1.95)$$

luego, la ecuación (1.94) puede ser escrita como

$$K^* = \text{mod } |C| \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_1^2} \dots e^{-\frac{1}{2}y_p^2} dy_p \dots dy_1 \quad (1.96)$$

$$K^* = \text{mod } |C| \prod_{i=1}^p \left\{ \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2} dy_i \right\}$$

Como

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = 1. \quad (1.97)$$

Por lo tanto,

$$K^* = \text{mod } |C| \prod_{i=1}^p \left\{ \sqrt{2\pi} \right\} \quad (1.98)$$

$$K^* = \text{mod } |C| (2\pi)^{\frac{1}{2}p} \quad (1.99)$$

Retomando la ecuación (1.90), se tiene

$$|C'| \cdot |A| \cdot |C| = |I| \quad (1.100)$$

Dado que

$$|C'| = |C| \quad (1.101)$$

y que $|I| = 1$

$$\text{mod } |C| = 1/\sqrt{|A|}. \quad (1.102)$$

Por lo tanto

$$K = 1/K^* = \sqrt{|A|}(2\pi)^{-\frac{1}{2}p}. \quad (1.103)$$

Reemplazando K en la expresión (1.85) se obtiene

$$\frac{\sqrt{|A|}}{(2\pi)^{\frac{1}{2}p}} e^{\frac{1}{2}(x-b)'A(x-b)}. \quad (1.104)$$

Se verá ahora cual es el significado real de los parámetros A y b encontrando los primeros dos momentos de X_1, \dots, X_p , para lo cual se toman estas variables aleatorias como elementos constitutivos del vector aleatorio

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}$$

Se define a continuación el valor esperado de una matriz aleatoria, y se aplica el resultado obtenido al vector anterior, recordando que un vector aleatorio no es otra cosa que una matriz con una sola columna.

Definición 1. Sea Z una matriz de variables aleatorias, $Z = (Z_{gh})$, $g = 1, 2, \dots, m$; $h = 1, 2, \dots, n$.

Si las variables aleatorias Z_{11}, \dots, Z_{mn} pueden tomar solamente un número finito de valores, la matriz aleatoria Z puede ser una de un número finito de matrices, es decir $Z(1), \dots, Z(q)$. Si la probabilidad de $Z = Z(i)$ es p_i , es posible definir $E(Z) = \sum_{i=1}^q Z(i)p_i$. Es claro entonces que $E(Z) = E(Z_{gh})$.

Si las variables aleatorias Z_{11}, \dots, Z_{mn} tienen una función de densidad conjunta, operando con sumas de Riemann se puede definir $E(Z)$ como el límite, en caso de que éste exista, de aproximar las sumatorias de la clase ocurren en el caso discreto,

entonces de nuevo $E(Z) = [E(Z_{gh})]$. Por lo tanto en general se usa la siguiente definición:

Definición 2. *El valor esperado de una matriz aleatoria Z es*

$$E(Z) = [E(Z_{gh})], \quad g = 1, \dots, m; h = 1, \dots, n.$$

En particular si Z es el vector aleatorio X definido anteriormente, el valor esperado de X es

$$E(X) = \mu \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{bmatrix}$$

$$\text{Si } Z = (X - \mu)(X - \mu)', \quad E(Z) = E\{(X - \mu)(X - \mu)'\} = \Sigma_x$$

El i -ésimo elemento de la diagonal de Σ_x corresponde a

$$E(X_i - \mu_i)^2 = \sigma_{X_i} = \text{Var}(X_i),$$

y el ij -ésimo elemento de la matriz Σ_x , que no pertenece a la diagonal, es decir que $i \neq j$,

$$E(X_i - \mu_i)(X_j - \mu_j) = \text{Cov}(X_i, X_j), \quad \text{es la covarianza entre } X_i \text{ y } X_j$$

$$\begin{aligned} \Sigma_x &= \text{Cov}(X) = E(XX' - \mu X' - X \mu' + \mu \mu') \\ &= E(XX') - \mu \mu' - \mu \mu' + \mu \mu' \\ &= E(XX') - \mu \mu' \end{aligned} \tag{1.105}$$

La operación de tomar el valor esperado de una matriz aleatoria (o vector aleatorio)

satisface ciertas reglas que se resumen en los siguientes lemas:

Lema 1. Sean Z una matriz aleatoria de tamaño $m \times n$, D una matriz real de tamaño $l \times m$, K una matriz real de tamaño $n \times q$ y F una matriz real de tamaño $l \times q$, entonces:

$$E(DZK + F) = D[E(Z)]K + F \quad (1.106)$$

Demostración. El elemento en la i -ésima fila y la j -ésima columna de la matriz $E(DZK + F)$ es

$$E \left(\sum_{h,g} d_{ih} Z_{hg} k_{gj} + f_{ij} \right) = \sum_{h,g} d_{ih} [E(Z_{hg})] k_{gj} + f_{ij} \quad (1.107)$$

El cual corresponde al elemento en la i -ésima fila y la j -ésima columna de la matriz $D[E(Z)]K + F$ □

Lema 2. Si $Y = DX + f$, donde X es un vector aleatorio, entonces

$$E(Y) = D(E(X)) + f \quad (1.108)$$

$$Cov(Y) = D(Cov(X))D' \quad (1.109)$$

Demostración. La primera de las afirmaciones procede directamente del lema 1. La segunda se presenta a continuación

$$Cov(Y) = E[Y - E(Y)][Y - E(Y)]' \quad (1.110)$$

$$Cov(Y) = E[DX + f - (D(E(X)) + f)][DX + f - (D(E(X)) + f)]'$$

$$Cov(Y) = E[D(X - E(X))][D(X - E(X))]'$$

$$Cov(Y) = E[D(X - E(X))][(X - E(X))'D']$$

Lo cual proporciona el lado derecho de la expresión (1.109), por lema 1. \square

Cuando la transformación corresponde a la expresión (1.91), que es $X = CY + b$, entonces $E(X) = CE(Y) + b$. La función de densidad de Y es proporcional a la expresión (1.95), obteniendo que

$$\frac{1}{(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}y'y} = \prod_{j=1}^p \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2} \right) \quad (1.111)$$

El valor esperado de la i -ésima componente de Y es

$$E(Y_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i \prod_{j=1}^p \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2} \right) dy_1 \cdots dy_p \quad (1.112)$$

$$E(Y_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i e^{-\frac{1}{2}y_i^2} dy_i \prod_{\substack{j=1 \\ j \neq i}}^p \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2} dy_j \right)$$

$$E(Y_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i e^{-\frac{1}{2}y_i^2} dy_i$$

$$E(Y_i) = 0$$

La última igualdad se obtiene porque $y_i e^{-\frac{1}{2}y_i^2}$ es una función impar de y_i y además se trata del valor esperado de una variable aleatoria normal estándar. Por lo tanto si $E(Y) = 0$, la media de X , denotada por μ , es

$$\mu = E(X) = b \quad (1.113)$$

Tomando la expresión (1.109) y aplicándola a la transformación $X = CY + b$, se obtiene que $Cov(X) = C[E(Y Y')]C'$. Por lo tanto el ij -ésimo elemento de la matriz

$E(Y Y')$ es: .

$$E(Y_i Y_j) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} y_i y_j \prod_{h=1}^p \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_h^2} \right) dy_1 \dots dy_p \quad (1.114)$$

porque la función de densidad de Y esta dada por la expresión(1.111) .Si se toma $i = j$, se tiene

$$E(Y_i^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i^2 e^{-\frac{1}{2} y_i^2} dy_i \prod_{\substack{h=1 \\ h \neq i}}^p \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_h^2} dy_h \right) \quad (1.115)$$

$$E(Y_i^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i^2 e^{-\frac{1}{2} y_i^2} dy_i$$

$$E(Y_i^2) = 1$$

Esta última igualdad se obtiene, porque se trata del valor esperado del cuadrado de una variable normalmente distribuida con media 0 y varianza 1. Si $i \neq j$ la expresión (1.114) se vuelve

$$E(Y_i Y_j) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i e^{-\frac{1}{2} y_i^2} dy_i \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_j e^{-\frac{1}{2} y_j^2} dy_j \cdot \prod_{\substack{h=1 \\ h \neq i, j}}^p \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_h^2} dy_h \right) \quad (1.116)$$

$$E(Y_i Y_j) = 0 \quad i \neq j$$

dado que la primera integral da como resultado cero. Las expresiones (1.115) y (1.116) pueden ser resumidas como

$$E(Y Y') = I \quad (1.117)$$

$$E(X - \mu)(X - \mu)' = C I C' = C C' \quad (1.118)$$

Tomando la expresión (1.90) $C'AC = I$, se obtiene $A = (C')^{-1}C^{-1}$. Luego

$$A^{-1} = [(C')^{-1}C^{-1}]^{-1} = CC' \quad (1.119)$$

Por lo tanto, la matriz de covarianzas de X es

$$\Sigma_x = E [(X - \mu)(X - \mu)'] = A^{-1} \quad (1.120)$$

De la expresión (1.119) se ve que la matriz Σ_x es definida positiva.

Si la expresión (1.104), corresponde a la función de densidad de un vector aleatorio X p -dimensional, el valor esperado de X es b y la matriz de covarianza es A^{-1} . Inversamente, dado un vector μ y una matriz definida positiva Σ , existe una función de densidad normal multivariada,

$$f(x; \mu, \Sigma_x) = (2\pi)^{-\frac{1}{2}p} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \quad (1.121)$$

tal que el valor esperado del vector aleatorio con esta densidad es μ y la matriz de covarianza es Σ_x .

Capítulo 2

Análisis Exploratorio de Datos Espaciales - AEDE

2.1. Introducción

Antes de comenzar describiendo cualquier método estadístico de análisis de datos espaciales es necesario definir que se debe entender como datos espaciales. Un dato espacial puede ser definido como la observación de una variable asociada a una localización del espacio geográfico. Esta localización puede ser expresada mediante tres tipos de objetos espaciales: puntos, líneas y polígonos (Anselin, 2001B, MapInfo, 1995).

- Los puntos se encuentran determinados por las coordenadas terrestres de latitud y longitud, pudiendo corresponderse con personas, individuos, empresas, ciudades, municipios.
- Las líneas son objetos que cubren una distancia dada y unen varios puntos o nodos, se determinan por las coordenadas de los puntos o nodos que comunican. Éste sería el caso de las líneas de transmisión, infraestructuras viales, calles de

una ciudad, etc.

- Los polígonos son figuras planas conectadas por distintas líneas u objetos cerrados que cubren un área determinada, se determinan mediante las coordenadas de todos los vértices del polígono. A este tipo de objeto espacial le pertenecen los países, departamentos, áreas municipales, comunas, unidades censales o áreas comerciales. (Chasco, 2003)

Cuando se tienen observaciones georreferenciadas, se deben utilizar herramientas que permitan detectar ciertas características dentro de los datos, como son tendencia, valores atípicos, esquemas de asociación y dependencia espacial, concentración espacial o puntos calientes/fríos, entre otros.

Aunque en la actualidad se tiene gran cantidad de información georreferenciada, estos datos suelen ser tratados con herramientas del análisis de series temporales (o de corte transversal, no espacial), sin usar técnicas adecuadas para el análisis estadístico espacial.

Los métodos que permiten extraer dichas características de los datos georreferenciados se conocen con el nombre de análisis exploratorio de datos espaciales (AEDE) y se conciben como una disciplina dentro del análisis estadístico más general, diseñada para el tratamiento específico de los datos geográficos. El AEDE se utiliza para identificar relaciones sistemáticas entre variables, o dentro de una misma variable, cuando no existe un conocimiento claro sobre su distribución en el espacio geográfico (Chasco 2006).

Al tratar de analizar los datos georreferenciados surge la pregunta. ¿Por qué varían las relaciones sobre el espacio? La razón más simple es que existen variaciones espaciales en las relaciones observadas debidas a variaciones muestrales aleatorias. Una segunda razón puede atribuirse a que las relaciones en sí pueden ser diferentes a través del espacio, tal vez porque existen variaciones espaciales en las actitudes o preferencias

de la población o existen asuntos administrativos, políticos o de otros contextos que producen respuestas diferentes ante el mismo estímulo. Una tercera razón puede ser la omisión o una representación funcional incorrecta de una o más variables relevantes para la explicación del modelo (Haining 2003).

El objetivo principal del análisis exploratorio de datos espaciales está relacionado con la identificación de excepciones locales o tendencias generales, ya sea en los datos o en las relaciones. De ahí que técnicas tales como el encadenamiento de ventanas y la suavización de datos permitan que la información sea examinada interactivamente de tal forma que los puntos que aparecen como atípicos puedan ser automáticamente localizados en un mapa. Realmente no importa qué técnica exploratoria se use, el objetivo es identificar puntos inusuales en los datos y se concentra en las excepciones en lugar de la tendencia general.

El Análisis Exploratorio de Datos Espaciales debe ser la etapa inicial de cualquier estudio econométrico que involucre datos georeferenciados.

2.2. Efectos Espaciales

En el presente numeral se hace una breve descripción de los efectos espaciales que pueden presentarse cuando se trabaja con datos georeferenciados, estos efectos son los que impiden que los métodos de la econometría estándar sean una buena herramienta para su modelación.

La información obtenida para uso de la ciencia regional posee características que provienen de su ubicación geográfica. Estas características se han denominado efectos espaciales, los cuales pueden ser divididos en dos tipos: dependencia espacial y heterogeneidad espacial.

De los dos, la dependencia espacial o la autocorrelación espacial es la más conocida. La dependencia espacial significa la ausencia de independencia que con frecuencia

está presente entre observaciones en conjuntos de datos. Esta dependencia puede ser expresada según la primera ley de la geografía de Tobler (1979), en la cual “todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes”.

El segundo tipo de efecto espacial, la heterogeneidad espacial, está relacionada con la ausencia de estabilidad en el comportamiento o las relaciones bajo estudio. Más precisamente, esto implica que los parámetros y formas funcionales varían con la ubicación y no son homogéneos en los conjuntos de datos.

En resumen, el campo de la econometría espacial consiste de aquellos métodos y técnicas que, basados en una representación formal de la estructura espacial de dependencia y heterogeneidad, proveen los medios de realizar una especificación, estimación, prueba de hipótesis y predicción adecuados.

2.2.1. Dependencia Espacial

La dependencia espacial en una colección de observaciones de datos muestrales se refiere al hecho de que una observación asociada con una localización la cual se puede denominar i , depende de otras observaciones asociadas con localizaciones $j \neq i$. Esto puede ser formalmente expresado como

$$y_i = f(y_j) \quad i = 1, \dots, n \quad j \neq i \quad (2.1)$$

Donde cada observación de una variable y en $i \in S$ (con S como el conjunto que contiene todas las unidades espaciales de observación), está relacionada formalmente a través de la función f , con las magnitudes de la variable en otra unidad espacial en el sistema.

Esta simple expresión, por si misma, no es muy útil en una situación empírica, dado que resultaría un sistema inidentificable, con muchos más parámetros (potencialmente

$N^2 - N$) que observaciones (N). Imponiendo una estructura en las relaciones funcionales involucradas en f , es decir, una forma particular para el proceso espacial, un número limitado de características de la dependencia espacial podría ser estimado y probado empíricamente.

Surge así el siguiente cuestionamiento. ¿Porque se esperaría que los datos observados en un punto del espacio sean dependientes de valores observados en otros puntos? Dos tipos de condiciones podrían conducir a una situación como la descrita.

La primera puede ser producto de los errores de medición para las observaciones en unidades espaciales contiguas. Ejemplos de estos problemas son: La delimitación arbitraria de las unidades espaciales de observación, las cuales no recogen adecuadamente el proceso generador de los datos muestrales.

La segunda está relacionada con problemas de agregación espacial, ya que en muchas situaciones encontradas en la práctica, la información es recolectada sólo a una escala agregada, por lo tanto, podría haber muy poca correspondencia entre el alcance del fenómeno agregado bajo estudio y la delimitación de las unidades espaciales de observación.

Los errores de medición tienden a saltarse las barreras de las unidades espaciales y como resultado, es probable que los errores para una observación en la unidad espacial i , estén relacionados con los errores en la unidad vecina j .

En un contexto de regresión, esto podría llevar fácilmente a términos de perturbación no esféricos y errores en la medición de las variables como se ilustra en la figura 2.1. Las verdaderas escalas espaciales de las variables bajo consideración son las áreas A , B y C , mientras que las observaciones están agregadas para las unidades definidas como 1 y 2. Como resultado la variable observada Y_1 será un agregado de Y_A y parte de Y_B y la variable observada Y_2 será una agregación de Y_C y la parte restante de Y_B :

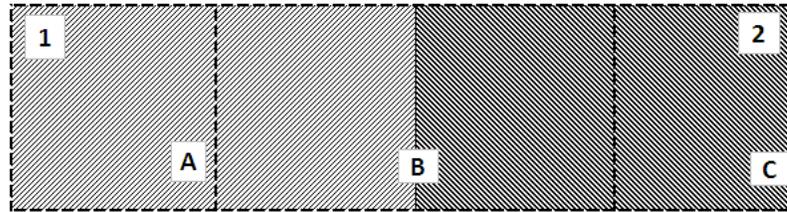


Figura 2.1: Dependencia espacial y agregación

$$Y_1 = Y_A + \lambda Y_B$$

$$Y_2 = Y_C + (1 - \lambda)Y_B$$

Esta agregación posiblemente contendrá los errores en la valoración del parámetro de ponderación λ , el cual está presente en Y_1 y en Y_2 . Estos errores de medición generarán un patrón que exhibe dependencia espacial.

Una segunda y tal vez más importante razón que podría causar dependencia espacial, proviene de la importancia del espacio como elemento estructural fundamental en explicaciones sobre el comportamiento humano y las actividades económicas. La ciencia regional está basada en la premisa de que la ubicación y la distancia son fuerzas importantes en trabajos de geografía humana y actividades de mercado. Todas estas nociones han sido formalizadas en la teoría de la ciencia regional que se apoya en las nociones de interacción espacial, procesos de difusión y jerarquías espaciales.

2.2.2. Autocorrelación Espacial

Definida de manera simple, la autocorrelación espacial es la concentración o dispersión de los valores de una variable en un mapa. Dicho de otra manera, la autocorrelación espacial refleja el grado en que objetos o actividades en una unidad geográfica son similares a otros objetos o actividades en unidades geográficas próximas (Goodchild, 1987).

La dependencia espacial se produce cuando “el valor de la variable dependiente en una unidad espacial es parcialmente función del valor de la misma variable en unidades vecinas” (Flint, Harrower y Edsall, 2000: 4). En el análisis de datos agregados geográficamente es frecuente encontrar que los valores de las variables estén autocorrelacionados espacialmente o sean espacialmente dependientes. La diferencia entre autocorrelación espacial y dependencia espacial está, fundamentalmente, en el uso de las palabras y estriba en que el primer caso se refiere simultáneamente a un fenómeno y técnica estadística, y el segundo, a una explicación teórica (Vilalta, 2004).

2.2.3. Heterogeneidad Espacial

El término heterogeneidad espacial se refiere a la variación en las relaciones sobre el espacio. En el más general de los casos se podría considerar que una relación diferente se presente para cada punto en el espacio. Formalmente una relación lineal se representaría así

$$y_i = X_i\beta_i + \epsilon_i \quad i = 1, \dots, n \quad (2.2)$$

La estimación de este tipo de modelos lleva implícito problemas relacionados con los grados de libertad, simplemente no se tienen suficientes datos con los cuales producir estimadores para cada punto en el espacio.

Para proceder con el análisis, se debe proveer una especificación para la variación sobre el espacio. Por lo tanto para llevar a cabo estimaciones e inferencias con soporte formal y asegurar la identificabilidad del modelo, es necesario imponer algunas restricciones a la expresión general.

En la literatura de ciencia regional y geografía económica, hay una amplia evidencia de la falta de uniformidad de los efectos del espacio. Muchos factores, tales como jerarquías del sitio central, la existencia de regiones líderes y regiones rezagadas, efectos

de cosecha en el crecimiento urbano, etc.; son argumentos que requieren de estrategias que tomen en cuenta las características particulares de cada ubicación o unidad espacial. En el trabajo econométrico esto se puede llevar a cabo mediante la consideración de parámetros variables, coeficientes aleatorios, o varias formas de cambio estructural.

Además, de esta falta de estabilidad estructural de varios de los fenómenos en el espacio, las unidades espaciales en observación son distantes de ser homogéneas. Por ejemplo, las unidades de los censos tienen diferente área y forma, los complejos urbanos tienen poblaciones o niveles de ingreso desiguales, y las regiones tienen varios grados de desarrollo tecnológico diferentes. En la medida en que estos aspectos de heterogeneidad se reflejen en los errores de medición, podrían resultar en heterocedasticidad.

A diferencia del caso de la dependencia espacial, los problemas causados por la heterogeneidad pueden ser, en la mayor parte, resueltos por los medios de las técnicas de la econometría estándar. Específicamente, métodos que pertenecen a parámetros variantes, coeficientes aleatorios e inestabilidad estructural pueden ser fácilmente adaptados para tomar en cuenta las variaciones en el espacio. Sin embargo, en muchos casos, la estructura espacial inherente en los datos puede llevar a procedimientos más eficientes. También, la compleja interacción, que resulta de la estructura espacial y de los flujos espaciales, podría generar dependencia en combinación con la heterogeneidad. En tal situación, el problema de distinguir entre dependencia espacial y heterogeneidad espacial es altamente complejo.

2.2.4. Heteroscedasticidad Espacial

La heteroscedasticidad consiste en la ausencia de estabilidad en la dispersión de un fenómeno, como sucede muchas veces con los residuos de una regresión y puede representarse como:

$$\text{Var}(u_i) = \sigma_i^2 \quad (2.3)$$

donde σ_i^2 indica que la varianza de la perturbación aleatoria es diferente para cada observación muestral i .

2.3. Matriz de Contigüidad

Se denomina matriz de contigüidad o de conectividad al arreglo W donde tanto cada una de las filas como de las columnas representa una región en el espacio objeto de estudio. Esta matriz representa la relación que tiene cada una de las regiones con las demás regiones del espacio en estudio, tal como se vería en un mapa. La forma más sencilla de construcción de esta matriz es utilizando notación binaria, donde 1 representa la presencia de contigüidad espacial entre dos unidades y 0 la ausencia de contigüidad espacial entre dos unidades. Una matriz construida de esta manera es simétrica.

Existe un desconcertante gran número de formas para definir la presencia o ausencia de contigüidad, algunas de las cuales se definen a continuación.

- Contigüidad lineal: En la cual se define $W_{ij} = 1$ para unidades que comparten un borde común a la izquierda o a la derecha, es decir la unidad i se encuentra inmediatamente al lado de la unidad j
- Contigüidad de torre: en la cual se define $W_{ij} = 1$ para unidades que comparten

un lado común con la región de interés a la izquierda, a la derecha, arriba o abajo.

- Contigüidad de alfil: en la cual se define $W_{ij} = 1$ para unidades que compartan un vértice común con la región de interés.
- Contigüidad lineal doble: Para dos unidades a la derecha inmediata o a la izquierda de la región de interés se define $W_{ij} = 1$.
- Contigüidad de torre doble: Para dos unidades a la derecha, a la izquierda, al norte o al sur de la región de interés se define $W_{ij} = 1$.
- Contigüidad de reina: Para unidades que compartan un lado en común o un vértice con la región de interés se define $W_{ij} = 1$.

Aunque parezca increíble existen muchas más formas en las cuales se podría procesar la matriz de Contigüidad.

En algunos textos la contigüidad lineal doble y la contigüidad doble de torre son referidas como contigüidad de segundo orden, mientras que las demás son denominadas de primer orden.

Algunas definiciones más elaboradas recaen en la longitud de los bordes comunes.

La guía principal para seleccionar una definición sobre las demás, sería la naturaleza del problema que se pretende modelar y tal vez la información adicional disponible diferente de las observaciones muestrales. Ejemplo de esto puede ser la conexión entre regiones por medio de autopistas, conocimiento de relaciones comerciales entre regiones, etc., que nos ayudan a decidir cual de las definiciones de contigüidad se ajusta más a las regiones objeto de estudio.

Una de las definiciones más usadas es la de la torre, con la cual se define la matriz de contigüidad de primer orden, ya que algunas veces solo es necesario establecer la localización de las unidades en el mapa que tienen bordes comunes con longitudes

positivas.

En el trabajo aplicado a menudo se lleva a cabo la transformación de la matriz de contigüidad de tal forma que los elementos de cada uno de los renglones sumen uno. Esta nueva matriz se denomina la matriz de contigüidad de primer orden estandarizada $C = W_z$.

La motivación para la estandarización puede ser vista considerando lo que ocurre si multiplicamos la matriz estandarizada por un vector de observaciones en las diferentes regiones de una variable asociada y . Esta nueva matriz producto $y^* = Cy$ representa una nueva variable igual a la media de las observaciones de las regiones contiguas.

2.4. Medidas de Similaridad

En la construcción de matrices de contigüidad que se ajusten al problema de estudio puede que la matriz de contigüidad binaria no provea una buena especificación, por lo cual es necesario tener otras posibles matrices. Para la construcción de estas matrices de contigüidad es necesario identificar la similaridad o disimilaridad existente entre los datos y las características inherentes a las variables en estudio. Estas medidas de similaridad se pueden clasificar en dos tipos:

- Las que reúnen propiedades de métrica, que son aquellas que son utilizadas con datos cuantitativos,

- Los coeficientes de asociación, que son aquellos empleados para datos expresados en una escala nominal.

2.4.1. Definición de métrica

Una métrica d sobre un conjunto Ω , es una aplicación

$$d : \Omega \times \Omega \rightarrow \mathfrak{R}$$

$$(x, y) \rightarrow d(x, y)$$

con las siguientes propiedades:

1. No negatividad: $d(x, y) \geq 0$
2. $d(x, y) = 0$, sí y solo si, $x = y$
3. Simetría: $d(x, y) = d(y, x)$, $\forall x, y \in \mathfrak{R}$
4. $d(x, y) \leq d(x, z) + d(z, y)$, $\forall x, y, z \in \mathfrak{R}$. Esto quiere decir que la longitud de uno de los lados del triángulo es menor o igual que la suma de las longitudes de los otros dos lados.

$d(x, y)$ es llamada la distancia entre x y y .

Definición 3. *Sea d una métrica para la cual no se cumple la desigualdad triangular pero se cumple*

$$d(x, y) \leq \max\{d(x, z), d(z, y)\} \quad \text{para todo } x, y \text{ y } z$$

a este tipo de distancia se le denomina ultramétrica, la cual es muy importante en los métodos de clasificación automática.

Las medidas de similaridad de uso más frecuente son:

- Medidas de distancia
- Coeficientes de correlación

- Coeficientes de asociación
- Medidas probabilísticas de similaridad

Antes de determinar cuál de estas medidas se va a emplear, es necesario identificar las variables que mejor representen el concepto de similaridad para el caso particular del estudio a desarrollar. Idealmente, dichas variables deberán escogerse dentro del marco conceptual que se usa explícitamente para la clasificación. La teoría en cada campo, es la base racional para la selección de las variables a usar en cada estudio. No se debe subestimar la importancia de la teoría para la selección de las variables, pues resulta muy peligroso caer en un “empirismo ingenuo”, por la facilidad con que los algoritmos forman grupos, sin importar el número y el tipo de variables, ya que por la naturaleza heurística de las técnicas de agrupamiento, se ha contaminado un poco su aplicación. Para la aplicación de esta técnica también se debe considerar la necesidad de estandarizar las variables, su transformación, o la asignación de un peso o ponderación para el cálculo de la medida de similaridad y la conformación de los conglomerados (Aldenderfer y Blashfield, 1984)

2.4.2. Medidas de distancia

Las medidas de distancia de uso más frecuente son:

- La distancia euclidiana, definida por la siguiente expresión

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{kj})^2}$$

- La distancia D^2 de Mahalanobis, también llamada distancia generalizada, la

cual se define por la siguiente expresión :

$$D^2 = d_{ij} = (X_i - X_j)' \Sigma^{-1} (X_i - X_j)$$

donde Σ es la matriz de varianzas y covarianzas de los datos, y X_i y X_j son los vectores de las mediciones que identifican los dos objetos i y j .

- La distancia de Manhattan, que se define como:

$$d_{ij} = \sum_{k=1}^p |X_{ik} - X_{kj}|$$

- La medida distancia de Minkowski

$$d_{ij} = \left(\sum_{k=1}^p |X_{ik} - X_{kj}|^r \right)^{\frac{1}{r}} \quad \text{con } r = 1, 2, \dots$$

También es posible utilizar otros conceptos de distancia mucho más intuitivos que los anteriores, que podrían resultar más aplicables si el problema que se está trabajando corresponde a mediciones espaciales.

A continuación, se presentan algunos de ellos:

- Bodson y Peeters (1975) proponen la utilización de un criterio de accesibilidad general, expresado entre 0 y 1, que combina en una función logística la frecuencia de varios canales de comunicación y transporte entre regiones como podrían ser las carreteras, vías férreas y demás medios de transporte y comunicación. La función aparece representada por la siguiente expresión:

$$w_{ij} = \sum_{n=1}^N K_n \left(\frac{a}{1 + b \exp(-c_j d_{ij})} \right)$$

Siendo K_n la importancia relativa del medio de comunicación n , d_{ij} la proximidad física entre dos regiones i y j y N el número total de medios de comunicación considerados, a , b y c_j son parámetros a estimar. Incorporar parámetros en el cálculo de las distancias es un problema importante ya que son determinados a priori, separadamente del resto del análisis espacial. Esto hace que la validez de las estimaciones esté condicionada a la estimación correcta de los parámetros.

- Case y Moreno (1993) propone la utilización de una distancia económica, sugiriendo, por ejemplo, la definición de w_{ij} como $w_{ij} = \frac{1}{|x_i - x_j|}$, donde las x son observaciones de características socioeconómicas, como por ejemplo el ingreso per cápita. Así, se supone que a mayor diferencia en los niveles de ingreso de dos unidades de observación, las relaciones entre las mismas son menores y viceversa.
- Vayá, López-Bazo y Moreno, proponen recoger el grado de intercambio comercial entre las regiones analizadas.

Para la contrastación y estimación, es importante que, independientemente del criterio que se use, se garantice que las distancias y los respectivos pesos sean considerados exógenos.

2.4.3. Coeficientes de correlación

Debido a su interpretación geométrica, los coeficientes de correlación son llamados frecuentemente medidas angulares. El coeficiente de correlación más popular es el coeficiente producto momento de Pearson, el cual determina el grado de correlación

o asociación lineal entre casos. Este coeficiente está definido por:

$$r_{jk} = \frac{\sum_i (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{\sqrt{\sum_i (X_{ij} - \bar{X}_j)^2} \sqrt{\sum_i (X_{ik} - \bar{X}_k)^2}}, \quad \text{con } i = 1, 2, \dots, p$$

X_{ij} : Es el valor en la región i para la variable j

\bar{X}_j : es la media de la variable j .

Esta medida se emplea para variables en escala al menos de intervalo.

El coeficiente de producto momento es sensible a la forma, esto significa que dos perfiles pueden tener correlación de +1,0 y no ser idénticos; gráficamente corresponde a líneas poligonales paralelas con alturas diferentes.

Una limitación del coeficiente de correlación es que no siempre satisface la desigualdad triangular y esto puede limitar la comparación entre perfiles. Otra limitación es su cálculo, pues debe obtenerse la media a través de diferentes tipos de variables, y no a través de casos como corresponde a su definición estadística; de cualquier modo, el coeficiente demuestra ser bueno frente a otros coeficientes de similaridad en el análisis de conglomerados, por cuanto reduce el número de clasificaciones incorrectas.

2.4.4. Coeficientes de asociación

Cuando los datos están expresados en escalas nominales, es apropiado utilizar coeficientes de asociación. Cada variable toma los valores de 0 (ausencia) y 1 (presencia) de un atributo; una tabla de doble entrada resume toda la información (como una matriz de diseño).

Al comparar estos objetos se tienen cuatro posibilidades (Crisci y López, 1983):

1. Que ambos tengan presente el carácter comparado (1,1).
2. Que ambos tengan ausente el carácter comparado (0,0).
3. Que el primero tenga el carácter presente y el segundo ausente (1,0).

4. Que el primero tenga el carácter ausente y el segundo presente (0,1).

La frecuencia con que se presentan esta cuatro características se resume en la siguiente tabla

	Objeto B	
Objeto A	1	0
1	(a)	(b)
0	(c)	(d)

Donde:

- (a) Número de atributos en los cuales el mismo estado es poseído por los dos objetos.
- (b) Frecuencia de caracteres en los cuales el primer objeto lo posee y el segundo no.
- (c) Número de caracteres en los que un estado está ausente en el primer objeto y presente en el segundo.
- (d) Número de caracteres en los cuales el mismo estado está ausente en ambos objetos.

- Coeficiente de asociación simple (S): Es la medida de similaridad más sencilla, entre los objetos i y j , se calcula mediante la siguiente fórmula:

$$S_{(i,j)} = \frac{a + d}{a + b + c + d}$$

sus valores están entre 0 y 1, este coeficiente toma en cuenta la ausencia de una variable para los dos objetos en consideración.

- El coeficiente de Jaccard (J), definido como

$$J_{(i,j)} = \frac{a}{a + b + c},$$

resuelve el problema de las ausencias conjuntas de una variable en el cálculo de la similaridad. Los biólogos anotan que con el empleo del coeficiente de asociación simple, algunos casos aparecerán como muy similares por el hecho de no poseer ningún atributo en común; es algo así como decir que una guayaba se parece a una naranja porque con ninguna de las dos se puede hacer jugo de mango.

- Rogers y Tanimoto (RT): le da prelación a las diferencias, como en el caso de los dos anteriores coeficientes donde sus valores oscilan entre 0 y 1; es decir, valores de mínima y máxima similitud, respectivamente. Su cálculo se hace mediante la siguiente expresión:

$$RT_{(i,j)} = \frac{a + d}{a + 2b + 2c + d}$$

- Sørensen o Dice (SD): este coeficiente le confiere mayor importancia a las coincidencias en estado de presencia, se expresa como

$$SD_{(i,j)} = \frac{2a}{2a + b + c}$$

Los valores de este coeficiente varían entre 0 y 1; y representan valores de mínima y máxima similitud, respectivamente.

- Sokal y Sneath (SS): este coeficiente tiene más en cuenta las coincidencias, tanto por presencia como por ausencia de los atributos. Sus valores se obtienen calculando

$$SS_{(i,j)} = \frac{2(a + d)}{2(a + d) + b + c}$$

y toma valores entre 0 y 1 que equivalen a la mínima y máxima semejanza respectivamente.

- Hamann (H): considera importante las diferencias entre coincidencias y no coincidencias. Los valores de similitud están en el rango de -1 a 1, mínima y máxima similitud, respectivamente. Se expresa así

$$H_{(i,j)} = \frac{(a + d) - (b + c)}{a + b + c + d}$$

Aunque en la literatura de taxonomía numérica se encuentran otros coeficientes, con los presentados anteriormente se brinda una idea general de esta estrategia para medir similaridad entre objetos.

Una objeción que se le puede hacer a los coeficientes de asociación es su aplicación sólo a respuestas dicotómicas; aunque, los datos continuos se pueden transformar a valores de tipo 0 y 1, el problema se reduce a decidir a qué valores se les asigna como 0 y a cuales como 1, esta transformación hace que se pierda información pues no tiene en cuenta la intensidad de los atributos.

2.4.5. Coeficientes de probabilidad

Son bastante diferentes a los anteriores, este tipo de coeficientes de medida trabaja directamente sobre los datos originales. Al construir conglomerados se considera la ganancia de información al combinar dos casos; se fusionan los dos casos que suministren la menor ganancia de información. Una limitación de estas medidas probabilísticas es su utilización únicamente para variables dicótomas.

2.5. Análisis tradicional vs. Análisis espacial

El análisis exploratorio de datos tradicional se ha definido como una colección de técnicas para resumir las propiedades de los datos (estadística descriptiva), detectar patrones en los datos, identificar características inusuales en los datos y formular hipótesis desde los datos. Estas técnicas también se usan para examinar los resultados del modelo y proporcionar evidencia de si los supuestos del modelo son satisfechos. El conjunto de técnicas aplicadas son gráficas (gráficos y figuras) y/o numéricas (resúmenes cuantitativos de los datos).

De la misma manera, el análisis exploratorio de datos espaciales está constituido por diversas técnicas que permiten explorar los datos espaciales, resumir las propiedades espaciales de los datos, detectar patrones en los datos, formular hipótesis que se refieren a la presencia de fenómenos espaciales dentro de los datos, identificar casos o subconjuntos de casos que son inusuales dada su localización en el mapa, etc. A diferencia del análisis tradicional, en el análisis exploratorio de datos espaciales los mapas cobran especial importancia, ya que permiten responder preguntas tales como: ¿Dónde caen en el mapa los casos atípicos observados en el histograma? ¿Dónde caen los valores de cierto atributo de una parte del mapa en el *scatterplot*? ¿Cuáles son los patrones espaciales y las asociaciones espaciales en este conjunto de datos?, etc.

2.6. Visualización de los datos

La visualización de datos espaciales emplea herramientas cartográficas (diferentes formas de presentación de mapas) y enlaces entre la cartografía y los diferentes gráficos estadísticos. Algunas utilidades de estos enlaces son entre otras:

- Enlace gráfico-mapa: permite identificar la localización en el área de estudio de

casos particulares como los *outliers*.

- Enlace mapa-gráfico: permite identificar si un área determinada es distinta en términos de sus atributos.
- Enlace dinámico gráfico-mapa: permite moverse sobre un polígono del gráfico y obtener las regiones resaltadas en la ventana del mapa.
- Enlace dinámico mapa-gráfico: permite moverse sobre el mapa y obtener los casos resaltados en la ventana del gráfico, o calcular los estadísticos para los casos que caen dentro de las regiones resaltadas.
- El cartograma: es una presentación del mapa donde las áreas de las regiones son proporcionales a la población de cada región y puede proveer un mejor marco espacial del atributo analizado. (Haining 2003)

A continuación detallaremos tanto los métodos visuales de análisis exploratorio de datos como los métodos de análisis numérico:

2.7. Presentación de los datos

Las herramientas gráficas juegan un papel muy importante en la visualización de los datos en el AEDE y los mapas se convierten en un elemento esencial en la visualización. La interacción entre las gráficas, los mapas, las tablas y los parámetros de las gráficas son tan importantes que se requiere obtener rápidamente los mapas apropiados, así como una modificación de los parámetros de manera interactiva para poder proveer diferentes tipos de visualizaciones, lo cual se logra con enlaces entre la cartografía y las presentaciones estadísticas.

Los mapas y los gráficos estadísticos deben ser analizados conjuntamente ya que se

complementan unos a otros en el análisis exploratorio de datos espaciales. La presentación de los datos hace referencia a lo que se muestra en un gráfico y qué tipo de gráfico se construye. Los gráficos para el caso univariado incluyen entre otros los histogramas, los boxplots, los gráficos cuantil-cuantil y los gráficos de probabilidad. Los gráficos para el caso multivariado incluyen el *scatterplot* (donde los casos son representados como puntos), los gráficos de ejes coordenados paralelos, los gráficos en tres dimensiones (caso de tres variables) y los glifos (donde los casos son representados como símbolos complejos, como el caso de los rostros de Chernoff).

Teniendo en cuenta que las diferentes presentaciones de los datos sirven para identificar características dentro de los datos, presentamos algunos de los diferentes tipos de gráficos que se pueden construir, de acuerdo con las características que se pueden obtener de ellos. La visualización de este tipo de gráficos es posible en la mayoría de los programas que existen en la actualidad para Análisis Exploratorio de Datos Espaciales, que permiten la interacción entre el gráfico y el mapa. En la tabla 2.1 se resumen las características que pueden ser extraídas de cada una de los diferentes tipos de gráficos que se tienen.

2.7.1. Identificación de valores atípicos

Diagrama de Caja - Mapa

El diagrama de caja, caja de bigotes ó *boxplot* es un gráfico que corresponde a la representación gráfica de cinco índices numéricos de las observaciones realizadas, que son en orden ascendente, valor mínimo observado, cuartil inferior o primer cuartil, mediana, cuartil superior o tercer cuartil y valor máximo observado. Los bigotes se encuentran en los valores máximo y mínimo de las observaciones y la caja se encuentra trazada entre los dos cuartiles, con una línea que corresponde a la localización de la mediana. Una representación más refinada de los diagramas de caja localizan los

	Perspectiva Econometría Espacial
Visualización de distribuciones espaciales	<i>Box map</i> , Histograma, Análisis de la varianza exploratorio espacial
Visualización de asociación espacial global	Gráficos del retardo espacial, Mapa y <i>Scatterplot</i> de Moran
Visualización de asociación espacial local	Mapas LISA, <i>Outliers</i> en el <i>Scatterplot</i> de Moran
Asociación espacial multivariante	<i>Scatterplot</i> multivariante de Moran
Heterogeneidad Espacial	Mapa, Histograma de Frecuencias, Diagrama de Dispersión

Fuente: Anselin 1998

Cuadro 2.1: Técnicas de Análisis Exploratorio de Datos Espaciales

bigotes en las siguientes ubicaciones:

$$\text{Bigote inferior} = \min(x_{\min}, Q_1 - 1,5(Q_3 - Q_1)) = \min(x_{\min}, Q_1 - 1,5RIQ)$$

$$\text{Bigote superior} = \max(x_{\max}, Q_3 + 1,5(Q_3 - Q_1)) = \max(x_{\max}, Q_3 + 1,5RIQ)$$

Donde:

Q_1 : Es el primer cuartil

Q_2 : Es la mediana

Q_3 : Es el tercer cuartil

$Q_3 - Q_1 = RIQ$: Es el rango intercuartílico

x_{\min} y x_{\max} : Son los valores mínimo y máximo observados.

De esta forma se consideran valores atípicos o *outliers* las observaciones que se encuentren por encima del bigote superior o por debajo del bigote inferior. El rango de

1.5 veces el recorrido intercuartílico puede ser cambiado por 3.0, el cual constituye un criterio más estricto para definir una observación como valor atípico.

Al trabajar con un paquete estadístico espacial y seleccionar en el diagrama de caja las observaciones atípicas, se podrá ver en el mapa a qué unidades espaciales corresponden dichas observaciones. (En el capítulo 4 se presentan los gráficos para el caso aplicado)

Mapa de Caja o *Boxmap*

Un mapa de caja es una versión del mapa de cuartiles, en el cual aparecen resaltados los valores atípicos o *outliers* que se encuentran por encima o por debajo de 1.5 (o 3.0) veces el rango intercuartil. La clasificación es la misma que la realizada en el diagrama de caja o *Boxplot*.

Diagrama de Dispersión - Mapa

En el caso multivariado para dos variables es útil realizar el gráfico de dispersión o *scatterplot* donde cada una de las variables es representada en cada uno de los ejes y cada punto corresponderá a una unidad espacial. Los puntos que se encuentran alejados de la nube de puntos serán los atípicos y utilizando los gráficos interactivos, las unidades espaciales correspondientes a dichos puntos, podrán ser identificados fácilmente en el mapa.

2.7.2. Identificación de la distribución de la variable

Histograma - Mapa

El histograma es la gráfica que representa las frecuencias de los valores observados y es una aproximación discreta de la función de densidad de probabilidad de una variable aleatoria.

Al construir el histograma se puede observar la distribución de la variable observada en el área de estudio, al trabajar con diagramas interactivos se puede observar en el mapa las unidades espaciales que pertenecen a cada uno de los intervalos. Adicionalmente si se quiere ver la distribución de la variable en una sub-región perteneciente al área de estudio, al seleccionar la subregión se puede observar en el histograma resaltada la distribución de la variable para un subconjunto de unidades espaciales, es decir el histograma regional. Si al comparar el histograma del área de estudio con el histograma regional resaltado estos difieren entre sí, este hecho sugeriría la posibilidad de la existencia del fenómeno espacial denominado heterogeneidad espacial.

Mapa de cuantiles

Es un mapa donde aparecen la distribución de las unidades espaciales de acuerdo con los cuantiles, es decir que las unidades que pertenecen al mismo cuantil tienen el mismo color. Puede ser elaborado con cuartiles, quintiles o percentiles.

2.7.3. Identificación de efectos espaciales

Cartograma

El cartograma es una representación gráfica del mapa donde cada unidad espacial es reemplazada por un círculo cuya área es proporcional al valor de la variable en dicha unidad espacial. El cartograma permite identificar valores atípicos y patrones espaciales en las unidades espaciales del área en estudio, ya que las unidades espaciales con los valores atípicos se resaltan con colores diferentes del resto de las unidades.

Diagrama de dispersión de Moran

El *scatterplot* de Moran es un diagrama de dispersión donde se representa la variable en estudio y el retardo espacial de dicha variable. El diagrama se obtiene representando en un plano cartesiano n pares de valores.

$$\left\{ \sum_{j \in N(i)} w^*(i, j) z(j), z(i) \right\}_{i=1, \dots, n}$$

Donde:

$w^*(i, j)$: denota la (i, j) -ésima entrada de la matriz estandarizada de ponderaciones.

$N(i)$: denota el conjunto de vecinos de la unidad i ($i \notin N(i)$).

El eje vertical es usado para los valores de los vecinos espacialmente promediados y el eje horizontal para la unidad espacial que se encuentra en el centro del promedio espacial. Anselin ha denominado este como el *scatterplot* de Moran (Anselin 1998). El estadístico I de Moran de la autocorrelación espacial es el estimador de la pendiente de la regresión por mínimos cuadrados ordinarios de $\left\{ \sum_{j \in N(i)} w^*(i, j) z(j), z(i) \right\}_{i=1, \dots, n}$ en $\{z(i)\}_{i=1, \dots, n}$. Cuanto mayor sea la pendiente obtenida mayor será el grado de autocorrelación espacial.

Este diagrama de dispersión divide el tipo de asociación espacial en cuatro categorías:

- Dos para autocorrelación espacial positiva, localizadas en los cuadrantes I , donde unidades espaciales con valores altos de la variable analizada están rodeadas de unidades espaciales con valores también altos en la característica analizada; y III, donde unidades con valores bajos de la variable están rodeadas de unidades con valores bajos de la variable.
- Dos para autocorrelación espacial negativa, localizadas en los cuadrantes II, donde unidades espaciales con valores bajos de la variable en estudio están rodeadas de unidades espaciales con valores altos de la variable: y IV en el cual

unidades espaciales con valores altos en la variable se encuentran rodeadas de unidades espaciales cuyos valores de la variable son bajos.

Boxplots Rezagados

Los *Boxplots* calculan separadamente los subconjuntos espaciales y luego los superponen, su comparación puede indicar diferencias entre las áreas. Las unidades espaciales que componen cada sub-área pueden no ser equivalentes, por ejemplo pueden variar en tamaño, por lo tanto la ponderación podría ser considerada de la siguiente manera. Sea $z(1), \dots, z(n(R))$, $(n(r))$: denota los valores para la región r y $w(1), \dots, w(n(R))$ las correspondientes ponderaciones para cada una de las unidades espaciales, con $r = 1, \dots, R$.

El propósito es hacer comparaciones sobre R subregiones con $R \geq 2$.

Para encontrar la mediana ponderada se ordenan las observaciones de tal forma que $z([1]) \leq z([2]) \leq \dots \leq z(n[R])$. Las ponderaciones se reordenan de tal forma que correspondan con el orden en $\{z(i)\}$.

La suma acumulada de las ponderaciones reordenadas es:

$$S(k) = \sum_{j=1}^k w([j]) \quad k = 1, 2, \dots, n(r) \quad (2.4)$$

El índice t de la mediana ponderada es el valor más pequeño de k , tal que $S(k)$ es al menos la mitad de la suma total de las ponderaciones:

$$t = \text{mín}\{k | S(k) \geq S(n(r))/2\}$$

Si $S(t)$ es estrictamente mayor que $S(n(r))/2$ la mediana ponderada es $z(t)$. Si $S(t)$ es igual a $S(n(r))/2$ la mediana ponderada es el promedio entre $z(t)$ y $z(t + 1)$. Los

cuartiles superior e inferior pueden ser calculados usando el mismo principio y el boxplot construido.

Las medianas ponderadas son utilizadas en varios contextos que incluyen la suavización de los datos espaciales. Las ponderaciones se pueden realizar basándose en el área, en la distancia hasta un punto dado o el tamaño de la población entre otras. Las ponderaciones mayores pueden ser asignadas a datos con errores estándar más pequeños, los cuales corresponden a las áreas con mayor población. La Función de densidad conjunta sugerida por Majure et al. (1996) incluye las ponderaciones:

$$F_{n(r)}(z; r) = \frac{\sum_{j=1, \dots, n(r)} w(j) I(z(j) \leq z)}{\sum_{j=1, \dots, n(r)} w(j)} \quad (2.5)$$

$$\text{donde } I(z(j) \leq z) = \begin{cases} 1 & \text{si } z(j) \leq z \\ 0 & \text{en otro caso} \end{cases} \quad (2.6)$$

Los boxplots espacialmente rezagados son una serie ordenada de boxplots. El orden es con respecto a una área o sitio determinado. Cada boxplot usa el subconjunto de los datos que están a una distancia dada del área escogida. Cada boxplot puede basarse en estadísticos ponderados como se describió anteriormente.

2.8. Transformaciones utilizadas en el AEDE

Si se obtienen observaciones para estudios sociales, económicos, demográficos, etc. las unidades espaciales corresponden en la mayoría de los casos a unidades geográficas o divisiones políticas, como son los municipios, departamentos, países, condados, estados, etc. Los datos para las diferentes unidades espaciales a menudo vienen totalizados, y las unidades espaciales difieren en tamaño, por lo tanto en la etapa exploratoria es necesario ajustar los datos de tal forma que el tamaño de la población no distorsione

las comparaciones (Haining 2003). Para realizar este ajuste es necesario convertir las observaciones en tasas o ratas, donde la escogencia del denominador debe ser bien definida.

La conversión teniendo en cuenta los diferentes tamaños poblacionales no es suficiente para asegurar comparabilidad de los datos dado que:

- Las tasas de unidades espaciales con poblaciones pequeñas son menos robustas que las calculadas para unidades espaciales con poblaciones grandes, puesto que al añadir o sustraer un caso el efecto será mayor en la tasa correspondiente a una unidad espacial con una población pequeña que en la tasa proveniente de una unidad espacial con una población grande.
- Las tasas calculadas para unidades espaciales con poblaciones pequeñas tienen mayores varianzas que las tasas calculadas para unidades espaciales con poblaciones grandes y existe una dependencia varianza-media en las tasas.

Sean

$n(i)$: el tamaño de la población,

$O(i)$: número de individuos en la unidad i que poseen la característica en evaluación

La tasa para la unidad i será

$$SR(i) = \frac{O(i)}{n(i)}$$

Si se considera que $O(i)$ es un proceso binomial entonces su valor esperado y su varianza están dados por:

$$E \left[\frac{O(i)}{n(i)} \right] = \left(\frac{1}{n(i)} \right) \quad E [O(i)] = p(i) \quad (2.7)$$

$$Var \left[\frac{O(i)}{n(i)} \right] = \left(\frac{1}{n(i)} \right)^2 \quad Var [O(i)] = \frac{p(i)(1-p(i))}{n(i)} \quad (2.8)$$

$p(i)$: es la probabilidad de que un individuo de la unidad i tenga la característica que está siendo evaluada.

Se observa que a medida que $n(i)$ se incrementa la $Var[O(i)/n(i)]$ decrece.

La media y la varianza de las expresiones (2.7) y (2.8) no son independientes. Para valores de $p(i)$ cercanos a 0 o a 1 y dado $n(i)$, la varianza es proporcional a la media (Haining 2003). Con el fin de remover esta dependencia media-varianza surge la necesidad del uso de una transformación estabilizadora de la varianza (especialmente para valores pequeños de $n(i)$). Cressie y Read (1989) sugieren que para datos dependientes espacialmente la transformación raíz cuadrada Freeman-Tukey para $O(i)$:

$$\left(\frac{O(i)}{n(i)}\right)^{1/2} + \left(\frac{(O(i) + 1)}{n(i)}\right)^{1/2} \quad (2.9)$$

es preferible a la usual transformación raíz cuadrada:

$$\left[\frac{(O(i) + 1)}{n(i)}\right]^{1/2} \quad (2.10)$$

Y se obtiene que el error estándar del estimado de la tasa $p(i)$ es:

$$\left[\frac{p(i)(1 - p(i))}{n(i)}\right]^{1/2}, \quad (2.11)$$

el cual está inversamente relacionado con el tamaño de la población. De aquí se obtiene que cualquier variación real espacial en las tasas puede ser confundida por variaciones en el $n(i)$ (tamaño de la muestra) o alternativamente variaciones espaciales en las tasas podrían ser producto de la estructura espacial en $n(i)$. Sin embargo la transformación Freeman-Tukey no iguala las varianzas (asunto importante en la regresión), para lo cual serán necesarias ponderaciones posteriores (Haining 2003).

Para muchos tipos de mapas se utilizan las razones estandarizadas, particularmente para clasificar unidades con relación a $E(i)$, la más sencilla de ellas es

$$SR(i) = O(i)/E(i)$$

$E(i)$: es el valor esperado en la unidad i y se asume fijo.

También se han propuesto otras razones alternativas que tienen mejores propiedades en ciertas situaciones. Un ejemplo de ellas es la sugerida por Visvalingam en 1983 basada en diferencias cuadradas.

$$\frac{[O(i) - E(i)]^2}{E(i)} \quad \text{y} \quad (2.12)$$

$$\frac{[|O(i) - E(i)| - 0,5]^2}{E(i)} \quad (2.13)$$

Otros estadísticos basados en procesos de Poisson son usados en epidemiología (Wray y otros 1999)

$$[O(i) - E(i)]^2 - O(i) \quad (\text{para } O(i) \geq E(i)) \quad (2.14)$$

$$\left[\frac{O(i)(O(i) - 1)}{E(i)^2} \right] \quad (2.15)$$

$$\left[\frac{O(i)(O(i) - 1)}{E(i)} \right] - E(i) \quad (2.16)$$

Estos estadísticos reducen el riesgo de obtener valores extremos en la distribución producidos por unidades con poblaciones pequeñas.

Otro método es asignar intervalos de confianza a cada razón estandarizada. Por ejemplo si se toma un modelo donde se asume que $O(i)$ es independiente e idénticamente distribuido Poisson, el parámetro es $\lambda(i) = E(i)r(i)$, entonces de las propiedades de la distribución Poisson se sigue que:

$$Var\left(\frac{O(i)}{E(i)}\right) = \left(\frac{1}{E(i)}\right)^2 Var(O(i)) = \left(\frac{1}{E(i)}\right)^2 (E(i)r(i)) = \frac{O(i)}{E(i)^2} \quad (2.17)$$

De donde se obtiene que el error estándar de la razón estandarizada es $O(i)^{1/2}/E(i)$.

Usando una aproximación normal para la distribución muestral de $SR(i)$, se puede

calcular el intervalo de confianza del 95 %:

$$SR(i) \pm 1,96 \left[\frac{O(i)^{1/2}}{E(i)} \right] \quad (2.18)$$

Existen otros métodos para calcular el intervalo de confianza pero también presentan problemas. El error estándar tiende a ser grande para unidades con poblaciones pequeñas y pequeño para unidades con poblaciones grandes, debido al efecto del tamaño de la población en el término $O(i)^{1/2}$ comparado con el efecto causado en el término $E(i)$. (Haining 2003)

2.9. Delimitación de regiones

La delimitación de las unidades espaciales con el fin de realizar la recolección de datos para los censos, no siempre refleja los procesos sociales y económicos, los cuales se encuentran más relacionados con los vecindarios o comunidades, que con las unidades censales. Esta delimitación se convierte en un filtro artificial de los datos. Por lo tanto la perspectiva que se puede obtener de los datos regionales dependerá de la partición del área de estudio, en cuanto a la escala como a la configuración particular con respecto a la realidad subyacente que represente. En este punto puede surgir un nuevo cuestionamiento: ¿Se están particionando áreas homogéneas al delimitar las unidades espaciales o se están agrupando varias comunidades que son muy diferentes en términos de las variables de estudio?

La agregación de pequeñas unidades espaciales puede ser necesaria ya que áreas con poblaciones grandes producen tasas más robustas que permiten la comparación significativa entre áreas. Adicionalmente, unir unidades espaciales con valores similares en las variables de estudio puede facilitar la visualización del mapa (Wang et al., 1997; Jaakkola, 1998). Por lo tanto si la agregación es requerida al menos en la etapa

exploratoria, sería deseable construir diferentes agregaciones de acuerdo con criterios específicos, con el fin de asegurar la sensibilidad de los resultados y así permitir la escogencia de la configuración más apropiada.

La regionalización es una forma especial de clasificación, donde las unidades espaciales básicas son agrupadas de acuerdo con un conjunto de criterios definidos. La diferencia entre la clasificación y la regionalización consiste en que esta última introduce la contigüidad espacial o restricciones de adyacencia. Uno de los principales problemas existentes con la clasificación y la regionalización es el gran número de particiones que es necesario considerar para encontrar la mejor de ellas.(Haining 2003)

1. Enfoques no espaciales para la construcción de regiones

Rutinas ordinarias de clasificación no espaciales pueden ser usadas para contruir regiones. Las observaciones son agrupadas en términos de su similaridad en uno o más atributos con el fin de minimizar la varianza dentro del grupo (homogeneidad intragrupal) y maximizar la varianza entre los grupos (heterogeneidad intergrupala). Las clasificaciones pueden tomar ya sea enfoques heurísticos o jerárquicos. Los clasificadores jerárquicos comienzan asignando las n unidades a un gran grupo, y de acuerdo con las medidas de similaridad entre las observaciones se van separando en diferentes grupos. Cada movimiento es el mejor del conjunto de posibles movimientos, pero la clasificación final, no siempre es la óptima. Los métodos no jerárquicos o heurísticos intentan encontrar una partición óptima. El más conocido es el de las k -medias en el cual se comienza con una partición de n observaciones en K grupos y luego se intercambian las observaciones individuales empleando una función objetiva para determinar si un grupo es mejor que otro.

Los clasificadores no espaciales inducen una regionalización en la que pueden resultar más de K regiones, algunas de las cuales pueden contener solamente

una o dos de las unidades espaciales básicas. Esta regionalización puede ser modificada basándose en la similaridad de regiones adyacentes en términos de la evidencia cuantitativa así como en el conocimiento local.

2. Enfoques espaciales para la construcción de regiones

Otros algoritmos de regionalización se basan en las mismas rutinas de los métodos anteriores excepto que la mezcla o separación de unidades y el intercambio no pueden violar restricciones de contigüidad espacial asociadas con las unidades espaciales (Webster and Burrough, 1972; Perruchet, 1983). Estos algoritmos son apropiados para la implementación con los sistemas geográficos de información, dado que estos sistemas almacenan información sobre la adyacencia espacial.

Openshaw (1978) desarrolló un algoritmo que consiste en comenzar con una regionalización inicial aleatoria del área de estudio con un número determinado de regiones, luego se selecciona aleatoriamente una de las regiones. Las unidades espaciales que rodean la región seleccionada son muestreadas y movidas dentro de la región, si la función objetiva no sufre deterioro alguno. El proceso es repetido región por región hasta que no puede lograrse ninguna mejora en la función objetiva.

Cualquier método utilizado puede ser sensible a la escogencia de la regionalización inicial, por lo que puede ser mejor comenzar con una regionalización inicial que reconozca la presencia de algunas áreas homogéneas en el área de estudio. Se han propuesto muchos métodos diferentes para establecer las regiones. Oliver y Webster presentan un método que busca un atributo de disimilaridad $d(i, j)$ de dos unidades espaciales i y j como la diferencia ponderada de los valores de sus atributos. La ponderación puede ser ajustada para reflejar la importancia de cada uno de los atributos usados en la medida de la disimilaridad. Esta medida de disimilaridad $d(i, j)$ es multiplicada por una función de su separación

geográfica para obtener un nuevo índice de disimilaridad $d^*(i, j)$ así:

$$d^*(i, j) = d(i, j)f(|s(i) - s(j)|) \quad (2.19)$$

Donde:

$|s(i) - s(j)| = u(i, j)$: es la distancia de separación entre las dos unidades cuyas localizaciones están dadas por $s(i)$ y $s(j)$.

f : puede ser la función geoestadística del modelo que mejor describa la autocorrelación espacial entre los datos.

3. Criterios de diseño para construir regiones

La regionalización puede ser problemática si existen muchos criterios de diseño. Los principales criterios pueden ser: Que las regiones sean homogéneas en términos de un conjunto de valores de atributos, que sean aproximadamente de igual tamaño en términos de la población y que sean relativamente compactas.

Los métodos de regionalización para el análisis exploratorio de datos espaciales deben ser relativamente rápidos de implementar, en particular si se desea ver los datos a través de diferentes particiones y a través de diferentes escalas.

2.10. Identificación de *clusters*

Un *cluster* puede ser definido como la agrupación de unidades espaciales localizadas en un área específica, que concentran valores más elevados o bajos de lo que cabría esperar en caso de encontrarnos ante una distribución homogénea, dominando la aleatoriedad en el resto del territorio estudiado.

Una de las dificultades con las que se han encontrado los analistas de fenómenos espaciales ha sido la identificación de *clusters* espaciales en los mapas. Existen diversos métodos para realizar la identificación de estos *clusters*, en el presente trabajo

se recogen las ideas centrales del desarrollado realizado por Jared Aldstadt y Arthur Getis, quienes propusieron un algoritmo denominado AMOEBA (A Multidirectional Optimum Ecotope-Based Algorithm), para la construcción de clusters espaciales teniendo como referencia las colas de la distribución. Este método es multidireccional, ya que se comienza en una unidad espacial que se considera “semilla” y la búsqueda se realiza en todas las direcciones haciendo comparaciones entre el conjunto total de unidades espaciales. Este algoritmo comparte algunas características con los métodos de segmentación de imágenes, cuyo objetivo es partir un gran número de píxeles en regiones homogéneas a partir de un número pequeño de píxeles semilla.

El algoritmo utiliza el estadístico local de Getis-Ord, el cual para una unidad espacial i está definido como:

$$G_i^* = \frac{\sum_{j=1}^N w_{ij}x_j - \bar{x} \sum_{j=1}^N w_{ij}}{S \sqrt{\frac{[\sum_{j=1}^N w_{ij}^2 - (\sum_{j=1}^N w_{ij})^2]}{N-1}}} \sim N(0, 1) \quad (2.20)$$

$$S = \sqrt{\frac{\sum_{j=1}^N x_j^2}{N} - (\bar{x})^2} \quad (2.21)$$

Donde

N es el número total de unidades espaciales

x_j es el valor del fenómeno en estudio en la unidad j

\bar{x} es la media de todos los valores

$$w_{ij} = \begin{cases} 1 & j \text{ es vecina de } i \\ 0 & \text{En otro caso} \end{cases} \quad (2.22)$$

Los elementos utilizados para una prueba de hipótesis son:

H_0 : No existe autocorrelación espacial

H_a : Existe autocorrelación espacial

$$G_i^* \sim N(0, 1)$$

El algoritmo AMOEBA se realiza para cada unidad i y sus pasos se exponen a continuación.

Paso 1. Se calcula el estadístico G_i^* para la unidad i , el cual se denota como $G_i^*(0)$, en este momento el *cluster* estaría conformado únicamente por la unidad i . Si este valor $G_i^*(0)$ es mayor que cero indica que el valor en la unidad espacial i es mayor que la media de todas las unidades y un valor menor que cero indica que el valor en dicha unidad es más pequeño que la media.

Paso 2. Se calcula el valor $G_i^*(1)$ para cada posible subregión que contenga i y todas las posibles combinaciones de las unidades espaciales vecinas contiguas. Si el estadístico $G_i^*(1)$ de alguna de las combinaciones es mayor en valor absoluto que el estadístico $G_i^*(0)$, se agregan las unidades que hacen parte de dicha combinación al *cluster*. En caso contrario termina el procedimiento para la unidad i .

Paso 3. Si se encontró una combinación cuyo estadístico maximiza el estadístico anterior, además de agregar las unidades al *cluster*, se eliminan todas aquellas unidades espaciales consideradas en las demás combinaciones, que no hacían parte de la mayor en valor absoluto. En este caso se repite el procedimiento con los vecinos contiguos de las unidades espaciales incluidas en el *cluster* y se evalúan todas las posibles combinaciones. El procedimiento continúa de la misma forma hasta no encontrar un estadístico mejor en valor absoluto.

Paso 4. El *cluster* estará formado por todas las unidades que conforman la combinación que produjo el mayor valor del estadístico G_i^* .

Capítulo 3

Índice I de Moran

3.1. Introducción

La distribución no aleatoria de los fenómenos en el espacio tiene varias consecuencias para el análisis estadístico convencional. Por ejemplo, los parámetros estimados con base en muestras que no se distribuyen aleatoriamente en el espacio son sesgados hacia valores prevalentes en las regiones favorecidas en el muestreo. Como resultado, muchos de los supuestos que se deben hacer sobre los datos antes de aplicar pruebas estadísticas no son válidos en estos casos.

Otro punto de vista al respecto es que la autocorrelación espacial introduce “redundancia” entre los datos, tal que cada unidad adicional aporta menos información nueva. Esto afecta el cálculo de los intervalos de confianza, etc. Tales efectos implican que sea recomendable evaluar el grado de autocorrelación espacial en un conjunto de datos espaciales antes de realizar cualquier análisis estadístico convencional basado en esta información.

Entre las medidas de diagnóstico de autocorrelación disponibles están los estadísticos de conteo conjunto, el índice I de Moran, el índice C de Geary y la nube de variograma. Estas son técnicas que ayudan a describir qué tan útil es conocer la ubicación de

una observación si se desea determinar un valor probable de un atributo medido en esa ubicación.

Indiscutiblemente, la herramienta más utilizada con este fin es el índice I de Moran. El mismo es una adaptación de una medida de correlación no-espacial a un contexto espacial y se aplica normalmente a unidades espaciales donde hay disponibilidad de información en forma de razones o intervalos. Una de las presentaciones del índice de Moran es

$$I = \frac{n}{\sum_{i=1}^n (y_i - \bar{y})^2} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad (3.1)$$

Se puede analizar cada uno de los factores que intervienen en el índice. Primero el numerador de la segunda fracción, es decir, $\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})$, que se reconoce como el término de varianza, de hecho es una covarianza. Los subíndices i y j se refieren a diferentes unidades o zonas espaciales en el estudio y y_i es el valor de la observación para cada una de ellas. Al calcular el producto de la diferencia de las observaciones de dos zonas con la media general (\bar{y}), se determina hasta dónde varían las observaciones conjuntamente. Si tanto y_i como y_j están al mismo lado de la media, este producto es positivo; si por el contrario, están ubicados a lados diferentes de la media, el producto es negativo y el tamaño absoluto del valor resultante depende de qué tan cercanos sean los valores observados a la media general. Los términos de covarianza son multiplicados por w_{ij} , este es un elemento de la matriz de ponderaciones W . En el caso más sencillo,

$$w_{ij} = \begin{cases} 1 & \text{si la región } i \text{ y la región } j \text{ son adyacentes} \\ 0 & \text{En caso contrario} \end{cases}$$

Los demás elementos de la fórmula normalizan el valor de I respecto al número de zonas en consideración, el número de adyacencias del problema y el rango de valores

en y .

El divisor $\sum \sum w_{ij}$ introduce el número de relaciones en el mapa.

El multiplicador $\frac{n}{\sum_{i=1}^n (y_i - \bar{y})^2}$ es en realidad una división por la varianza general del conjunto de datos, lo que asegura que I no es grande porque sencillamente los valores y variabilidad en y son altos. La ecuación del I de Moran puede escribirse en términos matriciales de forma más compacta como

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{y'Wy}{y'y} \quad (3.2)$$

Donde:

y : es el vector columna cuyas entradas son cada diferencia $y_i - \bar{y}$

El índice I de Moran y el índice Durbin y Watson son estructuralmente equivalentes al ser ambos un cociente de formas cuadráticas de residuos de regresión, su diferencia está en las matrices de pesos que especifican los enlaces entre las observaciones. La eficiencia y propiedades de los estimadores así como de otros estadísticos dependen en general de si los términos de error de un modelo son correlacionados espacialmente. De allí que sea importante probar la existencia de correlación espacial.

3.2. Evolución del índice I de Moran

La prueba más ampliamente utilizada para la detección de correlación espacial es aquella basada en la prueba estadística I de Moran. En esencia, esta prueba está formulada como una forma cuadrática de las variables a las que se les está midiendo la correlación espacial. Su definición original estandariza las variables restándoles la media muestral y reduciéndolas a través de la aplicación de un factor adecuado.

Cliff y Ord, en sus publicaciones de 1972, 1973 y 1981, generalizaron el estadístico de prueba I de Moran con el objetivo de derivar una prueba para correlación espacial

que funcionara en modelos de regresión lineal. La generalización que desarrollaron está formulada en términos de una forma cuadrática de los residuales estimados e incluye la fórmula original del índice como un caso especial correspondiente a un error donde el único regresor se relaciona con el intercepto. Asumiendo que las innovaciones son independientes e idénticamente distribuidas siguiendo una normal, Cliff y Ord derivaron la distribución del índice de prueba I para muestras grandes al igual que los momentos para muestras pequeñas.

En 1976, Sen dedujo la distribución del estadístico I de Moran para muestras grandes bajo el supuesto débil de que las variables son independientes e idénticamente distribuidas dado que se cumplan ciertas condiciones de los momentos. Su análisis sólo cubre la formulación original del índice y no la generalización introducida por Cliff y Ord.

Como se ha notado, todos los desarrollos mencionados se limitan a las muestras grandes, son deducidas bajo la hipótesis de correlación espacial nula y esto reduce considerablemente el espectro de las aplicaciones potenciales.

En 1981 Burridge demostró que el estadístico de prueba de Moran tiene una interpretación como multiplicador de LaGrange. King, en su publicación de 1981 y Anselin y Florax, en la suya de 1995, estudiaron las propiedades de diferentes pruebas de correlación espacial aplicadas en casos de muestras pequeñas. Anselin ha estudiado las pruebas de multiplicadores de LaGrange para dependencia y heterogeneidad espacial (1988), pruebas locales alternativas de correlación espacial (1995) y en compañía de Kelejian, en 1997, estudió el comportamiento si se utilizaban como regresores variables dependientes rezagadas espacialmente. Brett y Pinske propusieron una prueba no paramétrica para dependencia espacial.

En 1999, Pinske estableció un conjunto de condiciones débiles bajo las cuales se pueden usar algunas variaciones del índice I de Moran (el índice original y una variante transversal) para probar la presencia de correlación espacial. Sus principales aportes

comprenden la definición de un conjunto general de condiciones débiles bajo las cuales la prueba de Moran es válida en modelos de variable dependiente limitada y variable dependiente no lineal, igualmente, innova en cuanto a que una de las variables cuya correlación interesa probar puede tener media diferente de cero, lo cual es particularmente interesante cuando esta variable se define como una función de parámetros desconocidos.

En 2001, Kelejian y Prucha, basando sus demostraciones en supuestos menos exigentes, establecieron la distribución límite del estadístico I de Moran para variados modelos de variables dependientes y para modelos lineales transversales con la presencia de una variable dependiente rezagada espacialmente. Los resultados obtenidos en ese trabajo permiten la presencia de heteroscedasticidad entre las innovaciones, rezagos espaciales en la variable dependiente, errores estimados que dependen de manera no lineal de los datos y parámetros estimados e incluso la naturaleza triangular de los datos. Algunos resultados están relacionados con la estimación consistente de la varianza asintótica de formas cuadráticas.

3.3. Construcción del I de Moran

Sea

y : un vector de n variables aleatorias

X : una matriz no estocástica de tamaño $(n \times k)$

ε : un vector de perturbaciones aleatorias, $\varepsilon \sim N(0, \sigma^2 V)$

β : un vector de parámetros $(k \times 1)$

Si se considera la especificación

$$y = X\beta + \varepsilon \tag{3.3}$$

Se tiene

$$E(y) = E(X\beta + \varepsilon) = E(X\beta) + E(\varepsilon) = X\beta$$

La matriz de varianzas y de covarianzas de y

$$\begin{aligned}\Sigma_y &= E[(y - E(y))(y - E(y))'] \\ &= E[(y - X\beta)(y - X\beta)'] \\ &= E(\varepsilon\varepsilon') = \Sigma_\varepsilon = \sigma^2V\end{aligned}$$

por tanto $y \sim N(X\beta, \sigma^2V)$

El objetivo aquí es contrastar la validez de la hipótesis nula $H_0 : V = I$, dada la estructura de la regresión.

Frente a la hipótesis alternativa general de que $H_a : V \neq I$

I : es la matriz identidad.

En el caso bajo consideración, esta autocorrelación se llama normalmente “autocorrelación espacial” dada la naturaleza misma de los datos.

Para series de tiempo, a menudo, la hipótesis alternativa asume que los términos de perturbación aleatoria del modelo siguen un proceso de Markov.

Este supuesto se puede extender a los errores de regresión en contextos espaciales como se expone a continuación:

Se supone que los errores de regresión se ajustan al siguiente proceso de Markov de primer orden:

$$\varepsilon_i = \rho d_i \sum_j c(i, j) \varepsilon_j + u_i \quad i = 1, \dots, n \quad (3.4)$$

$$\varepsilon_i = \rho d_i [c(i, 1)\varepsilon_1 + c(i, 2)\varepsilon_2 + c(i, 3)\varepsilon_3 + \dots] + u_i$$

$$\varepsilon_1 = \rho d_1 [c(1, 1)\varepsilon_1 + c(1, 2)\varepsilon_2 + c(1, 3)\varepsilon_3 + \dots] + u_1$$

$$\varepsilon_2 = \rho d_2 [c(2, 1)\varepsilon_1 + c(2, 2)\varepsilon_2 + c(2, 3)\varepsilon_3 + \dots] + u_2$$

$$\varepsilon_3 = \rho d_3 [c(3, 1)\varepsilon_1 + c(3, 2)\varepsilon_2 + c(3, 3)\varepsilon_3 + \dots] + u_3$$

$$\vdots$$

$$\varepsilon_n = \rho d_n [c(n, 1)\varepsilon_1 + c(n, 2)\varepsilon_2 + c(n, 3)\varepsilon_3 + \dots] + u_n$$

Esto se puede expresar matricialmente de la siguiente manera:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \rho \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & d_n \end{bmatrix} \begin{bmatrix} c(1, 1) & c(1, 2) & c(1, 3) & \dots & c(1, n) \\ c(2, 1) & c(2, 2) & c(2, 3) & \dots & c(2, n) \\ c(3, 1) & c(3, 2) & c(3, 3) & \dots & c(3, n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c(n, 1) & c(n, 2) & c(n, 3) & \dots & c(n, n) \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix} \quad (3.5)$$

En forma compacta

$$\varepsilon = \rho DC\varepsilon + u \quad (3.6)$$

Donde:

D : es una matriz diagonal ($n \times n$) de términos genéricos d_i ,

C : es una matriz ($n \times n$) de constantes conocidas genéricas $c(i, j)$, que está dada por las relaciones de contigüidad entre las unidades espaciales i y j .

$$\begin{aligned}
\varepsilon &= \rho DC\varepsilon + uu = \varepsilon - \rho DC\varepsilon \\
u &= (I - \rho DC)\varepsilon \\
\varepsilon &= (I - \rho DC)^{-1}u
\end{aligned} \tag{3.7}$$

$$\begin{aligned}
\sigma^2 V &= \Sigma_y = E(\varepsilon\varepsilon') = \Sigma_\varepsilon \\
\Sigma_\varepsilon &= E \left[(I - \rho DC)^{-1}u [(I - \rho DC)^{-1}\mu]' \right] \\
\Sigma_\varepsilon &= E \left[(I - \rho DC)^{-1}uu' [(I - \rho DC)^{-1}]' \right] \\
\Sigma_\varepsilon &= E \left[(I - \rho DC)^{-1}uu' [(I - \rho DC)']^{-1} \right] \\
\Sigma_\varepsilon &= \sigma^2 [(I - \rho DC)'(I - \rho DC)]^{-1} \\
\Sigma_\varepsilon &= \sigma^2 [I - \rho(DC - C'D)' + \rho^2 C'D^2C]^{-1}
\end{aligned}$$

haciendo $W = (DC - C'D)'$ y $G = C'D^2C$ se tiene que

$$\begin{aligned}
\sigma^2 V &= \sigma^2 (I - \rho W + \rho^2 G)^{-1} \\
V &= (I - \rho W + \rho^2 G)^{-1} \\
V^{-1} &= I - \rho W + \rho^2 G
\end{aligned} \tag{3.8}$$

Si se define el elemento genérico de G como

$$g(i, j) = \begin{cases} 1, & i = j = 2, \dots, n-1 \\ 0, & \text{En cualquier otro caso} \end{cases}$$

y $W = A + A'$ donde los elementos genéricos de la matriz A se definen como

$$a_{ij} = \begin{cases} 1, & j = i - 1 \\ 0, & \text{En cualquier otro caso} \end{cases}$$

La ecuacion 3.8 sirve como base para realizar el contraste $H_0 : \rho = 0$ versus $H_a : \rho = \rho_1$.

Moran en su artículo “Notes on Continuous Stochastic Phenomena”, propuso una prueba para el contraste de correlación entre unidades espaciales “adyacentes”.

Esta prueba está definida sobre mn variables independientes x_{ij} de la siguiente manera:

$$mn\bar{x} = \sum_{i,j} x_{ij}$$

$$z_{ij} = x_{ij} - \bar{x}$$

$$r_{11} = \left(\frac{mn}{2mn - m - n} \right) \frac{\sum_i^m \sum_j^{n-1} z_{ij} z_{i,j+1} + \sum_i^{m-1} \sum_j^n z_{ij} z_{i+1,j}}{\sum_{ij} z_{ij}^2}$$

$$r_{11} = \frac{(mn) [\sum_i^m (z_{i1} z_{i2} + z_{i2} z_{i3} + \dots + z_{i,n-1} z_{in}) + \sum_i^{m-1} (z_{i1} z_{i+1,1} + z_{i2} z_{i+1,2} + \dots + z_{in} z_{i+1,n})]}{(2mn - m - n) [\sum_i (z_{i1}^2 + z_{i2}^2 + \dots + z_{in}^2)]}$$

$$r_{11} = \left(\frac{mn}{2mn - m - n} \right) [(z_{11} z_{12} + z_{12} z_{13} + \dots + z_{1,n-1} z_{1n} + z_{21} z_{22} + z_{22} z_{23} + \dots + z_{2,n-1} z_{2n} + \dots + z_{m1} z_{m2} + z_{m2} z_{m3} + \dots + z_{m,n-1} z_{mn}) + (z_{11} z_{21} + z_{12} z_{2,2} + \dots + z_{1n} z_{2n} + z_{21} z_{31} + z_{22} z_{32} + \dots + z_{2n} z_{3,n} + \dots + z_{m-1,1} z_{m1} + z_{m-1,2} z_{m2} + \dots + z_{m-1,n} z_{mn})] /$$

$$(z_{11}^2 + z_{12}^2 + \dots + z_{1n}^2 + z_{21}^2 + z_{22}^2 + \dots + z_{2n}^2 + \dots + z_{m1}^2 + z_{m2}^2 + \dots + z_{mn}^2) \quad (3.9)$$

$$r_{11} = \left(\frac{mn}{2mn - m - n} \right) \frac{\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix} \begin{bmatrix} z_{12} & z_{22} & \cdots & z_{m2} \\ z_{13} & z_{23} & \cdots & z_{m3} \\ \vdots & \vdots & \vdots & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{mn} \end{bmatrix}}{\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix} \begin{bmatrix} z_{11} & z_{21} & \cdots & z_{m1} \\ z_{12} & z_{22} & \cdots & z_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{mn} \end{bmatrix}}$$

$$r_{11} = \left(\frac{mn}{2mn - m - n} \right) \frac{\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix}}{\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix} \begin{bmatrix} z_{11} & z_{21} & \cdots & z_{m1} \\ z_{12} & z_{22} & \cdots & z_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{mn} \end{bmatrix}} \begin{bmatrix} z_{12} & z_{22} & \cdots & z_{m2} \\ z_{13} & z_{23} & \cdots & z_{m3} \\ \vdots & \vdots & \vdots & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{mn} \end{bmatrix}$$

Sea

$$e = \begin{bmatrix} z_{11} & z_{21} & \cdots & z_{m1} \\ z_{12} & z_{22} & \cdots & z_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{mn} \end{bmatrix}$$

$$r_{11} = \left(\frac{mn}{2mn - m - n} \right) \frac{e'}{e'e} \begin{bmatrix} z_{12} & z_{22} & \cdots & z_{m2} \\ z_{13} & z_{23} & \cdots & z_{m3} \\ \vdots & \vdots & \vdots & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{mn} \end{bmatrix}$$

$$r_{11} = \left(\frac{mn}{2mn - m - n} \right) \frac{e'}{e'e} \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} z_{11} & z_{21} & \cdots & z_{m1} \\ z_{12} & z_{22} & \cdots & z_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{mn} \end{bmatrix}$$

Sea

$$W = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$r_{11} = \left(\frac{mn}{2mn - m - n} \right) \frac{e'W e}{e'e}$$

Definición 4. El índice I de Moran se define a partir de la construcción anterior como

$$I = \frac{e'W e}{e'e} \tag{3.10}$$

$$r_{11} = \left(\frac{mn}{2mn - m - n} \right) I$$

Este estadístico se aproxima a la razón estadística de verosimilitud cuando $\rho \rightarrow 0$ y se puede mostrar que es consistente para $H_0 : \rho = 0$ contra alternativas generales que especifican valores de $\rho \neq 0$.

La ventaja más importante de esta representación del estadístico es que, bajo la hipótesis nula y asumiendo que los errores se distribuyen normalmente, $e'e$ se distribuye como una variable χ^2 centrada, lo cual es útil en otras instancias.

La ecuación 3.10 es equivalente matemáticamente a una regresión de mínimos cuadrados ordinarios de We sobre e y no de e sobre We , la cual correspondería a un proceso autorregresivo espacial.

La distribución asintótica del estadístico I de Moran calculado para residuos de regresión fue desarrollada por Cliff y Ord (1972, 1973, 1981).

Cuando el estadístico ha sido debidamente transformado, se puede ajustar a una distribución normal estándar.

Definición 5. Se llama matriz de proyección o de ortogonalización a $M = I - X(X'X)^{-1}X'$

Proposición 1. M es una matriz simétrica es decir $M' = M$

Proposición 2. M es una matriz idempotente, es decir $MM = M$

Se considera la especificación

$$\begin{aligned}
 Y &= X\beta + \varepsilon \\
 \hat{\varepsilon} &= Y - \hat{Y} = Y - X\beta \\
 &= Y - X(X'X)^{-1}X'Y \\
 &= [I - X(X'X)^{-1}X']Y \\
 &= [I - X(X'X)^{-1}X'](X\beta + \varepsilon) \\
 &= X\beta + \varepsilon - X(X'X)^{-1}(X'X)\beta - X(X'X)^{-1}X'\varepsilon \\
 &= X\beta + \varepsilon - X\beta - X(X'X)^{-1}X'\varepsilon \\
 &= [I - X(X'X)^{-1}X']\varepsilon
 \end{aligned}$$

Sea $M = I - X(X'X)^{-1}X'$

entonces $\hat{\varepsilon} = MY = M\varepsilon$

$$\hat{\varepsilon} = MY = M(X\beta + \varepsilon) \quad \text{ó} \quad \hat{\varepsilon} = M\varepsilon = M(Y - X\beta) \quad \text{respectivamente}$$

De esta manera, la matriz M proyecta Y y de igual manera a ε en el espacio de los residuales $\hat{\varepsilon}$.

Proposición 3. *Si los términos de perturbación ε son independientes e idénticamente distribuidos, con $E(\varepsilon) = 0$ y matriz de varianzas y covarianzas $\Sigma_\varepsilon = \sigma^2 I$, entonces los residuales $\hat{\varepsilon}$ están correlacionados.*

Demostración.

$$\begin{aligned}
 \Sigma_{\hat{\varepsilon}} &= E[(\hat{\varepsilon} - E(\hat{\varepsilon}))(\hat{\varepsilon} - E(\hat{\varepsilon}))'] = E[(M\varepsilon)(M\varepsilon)'] \\
 &= E(M\varepsilon\varepsilon'M') = M\sigma^2 IM' = \sigma^2 MM = \sigma^2 M
 \end{aligned}$$

□

Para hallar las expresiones correspondientes a los dos primeros momentos de I , se pueden utilizar diversos métodos, entre ellos se puede seguir un procedimiento similar a los empleados por Durbin y Watson para validar la presencia de autocorrelación entre errores en el caso de series de tiempo.

Los momentos de I son evaluados en términos de los valores propios de la matriz MW y teniendo en cuenta las características de M si suponemos que el rango de M es $(n - k - 1)$, sus valores propios que son diferentes de cero se pueden escribir como $v_1 \leq v_2 \leq \dots \leq v_{n-k-1}$.

Un procedimiento similar al empleado por Durbin-Watson permite establecer que

$$E(I) = \mu = (n - k - 1)^{-1} \sum v_i = \bar{v}$$

$$Var(I) = \sigma^2 = 2 \sum \frac{(v_i - \bar{v})^2}{(n - k - 1)(n - k + 1)}.$$

(Moran [20])

En ocasiones, se puede tener un valor de n muy grande y k finito, en este caso se utiliza un acumulador estandarizado de I , que es aproximadamente:

$$l_j = \frac{k_j}{\sigma_j} = 2^{\frac{j}{2-1}} (j - 1)! \frac{\sum (v_i - \bar{v})^j}{[\sum (v_i - \bar{v})^2]^{\frac{j}{2}}} \quad (3.11)$$

La ecuación 3.11 es útil para mostrar la normalidad asintótica del índice I .

La evaluación explícita de los v_i no es directa, Durbin y Watson consideraron los valores propios de W , los que se llamarán λ_i . En su trabajo, Durbin y Watson demostraron que $\lambda_i \leq v_i \leq \lambda_{i+k+1}$, lo cual permitió acotar los valores de v_i y, como consecuencia, los valores de los momentos de I .

De hecho, si $\xi_1, \dots, \xi_{n-k-1}$ representan $(n - k - 1)$ variables aleatorias independientes

e idénticamente distribuidas con media cero, se sigue que:

$$I = \sum_{i=1}^{n-k-1} v_i \xi_i^2 / V \quad (3.12)$$

$$\text{Donde } V = \sum_{i=1}^{n-k-1} \xi_i^2, \text{ y } I_I \leq I \leq I_S,$$

$$\text{donde } I_I = \sum_{i=1}^{n-k-1} \lambda_i \xi_i^2 / V \text{ y } I_S = \sum_{i=1}^{n-k-1} \lambda_{i+k+1} \xi_i^2 / V$$

Los valores de I_I e I_S pueden usarse para encontrar las cotas inferior y superior de los puntos porcentuales del estadístico.

En el caso de series de tiempo, W tiene una forma especial con valores propios bien definidos, sin embargo, en el caso espacial W puede tomar cualquier forma dependiendo del problema que se analice. Por lo tanto, a pesar de que el método general puede aplicarse al caso espacial, no es posible construir las tablas generales de puntos porcentuales para el índice.

Con los valores $E(I)$ y $V(I)$ y teniendo en cuenta el supuesto sobre la distribución de las variables X_i podemos estandarizar.

$$Z = \frac{I - E(I)}{\sqrt{V(I)}} \sim_{asy} N(0, 1) \quad (3.13)$$

Se verá más adelante que el I de Moran se puede escribir en términos de los residuales.

3.4. Distribución Exacta del I de Moran

El estadístico I de Moran se distribuye asintóticamente normal para la mayoría de las estructuras espaciales. Hallar la distribución exacta del estadístico es una tarea compleja que ha sido abordada desde enfoques algebraicos y numéricos.

Es necesario resaltar que el índice I de Moran puede transformarse en una suma de

cuadrados de variables normales. Las ponderaciones de esta suma difieren entre las distribuciones exacta y límites y están basadas en los valores propios de la matriz de la estructura espacial. Poder determinar esos valores propios, permite hacer importantes deducciones sobre la distribución del índice, tales como sus momentos y su posible rango.

Sean

- $1_{n \times 1}$: un vector columna de n unos
 $n = 1'1$: es el número de unidades espaciales
 $A = 1'_{1 \times n} C_{n \times n} 1_{n \times 1}$: es la conectividad total del sistema espacial
 $M = I - X(X'X)^{-1}X'$: es la matriz de proyección

El índice I de Moran se puede expresar en su forma más general como:

$$I = \frac{n}{A} \frac{y' M C M y}{y' M M y} \quad (3.14)$$

Estandarizando cada fila de C al dividirla por la suma de la fila (también conocida como conectividad, valencia o grado de la unidad espacial), se construye

$W = \text{diag}(C 1_{n \times 1})^{-1} C$ que se conoce como la contraparte estocástica de C .

Se define C_s la matriz de conectividad global estandarizada como

$$C_s = \frac{n}{A} C \quad (3.15)$$

La definición de esta matriz de conectividad global estandarizada permite comparar el estadístico resultante de diversas estructuras espaciales C , sin importar que tengan diferente número de unidades espaciales y diferente conectividad total A del sistema geográfico.

Teniendo en cuenta que $\hat{\varepsilon} = M\varepsilon$ el índice I de Moran se puede escribir en términos

de los residuales como

$$I = \frac{\hat{\varepsilon}' C_s \hat{\varepsilon}}{\hat{\varepsilon}' \hat{\varepsilon}} \quad (3.16)$$

Recordemos que el estadístico d de Durbin y Watson está definido como

$$d = \frac{\sum_{t=2}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\varepsilon}_t^2} = \frac{\hat{\varepsilon}' A \hat{\varepsilon}}{\hat{\varepsilon}' \hat{\varepsilon}} = \frac{Y' M A M Y}{Y' M M Y} \quad (3.17)$$

$$\text{con } A = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \quad (3.18)$$

En este punto se hace evidente la equivalencia estructural entre el índice I de Moran y el índice d de Durbin y Watson. Se trata en ambos casos de un cociente de formas cuadráticas de residuos de regresión. Su diferencia radica en la especificación de los enlaces entre las observaciones, es decir, en C_s para el caso espacial, y su equivalente A para el caso serial.

Lema 3. Sean $\hat{\varepsilon}$ y ε dos vectores $n \times 1$ normales aleatorios tales que:

$\hat{\varepsilon} = M\varepsilon$, $M = I - X(X'X)^{-1}X'$, X es una matriz no estocástica ($n \times p$) y de rango p e $I = \frac{\hat{\varepsilon}' C_s \hat{\varepsilon}}{\hat{\varepsilon}' \hat{\varepsilon}}$. Entonces

1. Existe una transformación ortogonal $\varepsilon = H\delta$ tal que las variables aleatorias δ_i

son independientes e idénticamente distribuidas normales $N(0, 1)$. De tal manera que el índice I de Moran se puede escribir como:

$$I = \frac{\sum_{i=1}^{n-p} \gamma_i \delta_i^2}{\sum_{i=1}^{n-p} \delta_i^2} = \frac{\delta' H' M C_s M H \delta}{\delta' H' M M H \delta} = \frac{\delta' \Gamma \delta}{\delta' I_{n-p} \delta} \quad (3.19)$$

Donde:

Γ es una matriz diagonal ($n \times n$) con los valores propios de la matriz $M C_s M$

H : es una matriz ($n \times$) con los vectores propios de la matriz $M C_s M$

I_{n-p} : es una matriz ($n \times n$) que se asemeja a una matriz identidad con solo $n - p$ unos en su diagonal y el resto de sus elementos iguales a cero.

2. Si s combinaciones lineales independientes de las columnas de X están perfectamente correlacionadas cada una con un vector propio de C_s diferente y si los valores propios restantes se enumeran de tal manera que

$$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n-s},$$

entonces $\lambda_i \leq \gamma_i \leq \lambda_{p+i-s} \quad i = 1, 2, \dots, n - p \quad (3.20)$

De este lema se puede deducir el siguiente corolario

Corolario 1.

$$I_I \leq I \leq I_S, \quad \text{donde}$$

$$I_I = \frac{\sum_{i=1}^{n-p} \lambda_i \delta_i^2}{\sum_{i=1}^{n-p} \delta_i^2} \quad e \quad I_S = \frac{\sum_{i=1}^{n-p} \lambda_{p+i-s} \delta_i^2}{\sum_{i=1}^{n-p} \delta_i^2} \quad (3.21)$$

Cabe resaltar que hasta el momento se ha encontrado la forma de expresar el índice I de Moran como la suma de variables aleatorias independientes e idénticamente distribuidas normales multiplicadas por los valores propios de la matriz $M C_s$. El rango de valores del índice se ha limitado al intervalo (γ_1, γ_{n-p}) . Los momentos teóricos del

índice se pueden calcular a partir de los valores propios $\gamma_1, \dots, \gamma_{n-p}$ y por lo tanto no dependen exclusivamente de la estructura espacial C_s , sino también de la matriz de diseño X a través de la matriz de proyección M . Adicionalmente, se ha hallado una relación entre los valores propios $\gamma_1, \dots, \gamma_{n-p}$ de los residuos distribuidos espacialmente. Esto permite derivar las distribuciones de I_I e I_S que limitan la distribución exacta del I de Moran y que son independientes de la matriz de regresión X .

Hasta el momento se han realizado numerosos intentos por derivar la función de distribución de cocientes de formas cuadráticas. Este es un problema que sólo se ha logrado resolver a través del método analítico propuesto por Imhof en 1961 y la transformación algebraica propuesta por Koerts y Abrahamse en 1968. Pan Jic-jian (1968) y Sen (1990) propusieron algunos enfoques directos más recientemente, sin embargo sus conclusiones no fueron lo suficientemente prácticas y consecuentemente no han sido muy aplicadas.

Imhof determinó la distribución exacta de una forma cuadrática de variables normales, es decir, de variables distribuidas según una χ^2 , basándose en su función característica a través de la transformada integral inversa de Fourier, la cual evaluó finalmente usando una integración numérica sobre valores reales. Koerts y Abrahamse propusieron una transformación algebraica directa sobre la ecuación 3.19 para convertir ese cociente de formas cuadráticas en una forma cuadrática sencilla

$$\begin{aligned} Pr(I \leq \frac{I_O}{H_O}) &= Pr(Q_O \leq \frac{O}{H_O}) \\ \Leftrightarrow Pr\left(\frac{\sum_{i=1}^{n-p} \gamma_i \delta_i^2}{\sum_{i=1}^{n-p} \delta_i^2} \leq \frac{I_O}{H_O}\right) &= Pr\left[\sum_{i=1}^{n-p} (\gamma_i - I_O) \delta_i^2 \leq \frac{O}{H_O}\right] \end{aligned} \quad (3.22)$$

El término $Q_O = \sum_{i=1}^{n-p} (\gamma_i - I_O) \delta_i^2$ es una suma ponderada de $n - p$ variables χ^2 centrales, cada una con un grado de libertad.

Su función de distribución $F\left(\frac{I_O}{H_O}\right) = Pr\left(Q_O \leq \frac{O}{H_O}\right)$ puede ser calculada por el

método de Imhof, que la simplifica para variables con distribución χ^2 en la integral.

$$F\left(\frac{I_O}{H_O}\right) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\text{sen}[\theta(u)]}{u\rho(u)} du \quad (3.23)$$

donde
$$\theta(u) = \frac{1}{2} \sum_{i=1}^{n-p} \tan^{-1}[(\gamma_i - I_O)u], \quad \text{y} \quad \rho(u) = \prod_{i=1}^{n-p} [1 + (\gamma_i - I_O)^2 u^2]^{1/4}$$

La integral 3.23 se puede aproximar por integración numérica y teniendo en cuenta que

$$\lim_{u \rightarrow 0} \frac{\text{sen}[\theta(u)]}{u\rho(u)} = \frac{1}{2} \sum_{i=1}^{n-p} (\gamma_i - I_O) \quad (3.24)$$

Los parámetros de integración numérica, a saber, tamaño de subintervalo y error de truncado apropiados para resolver la integral infinita 3.23 dependen del comportamiento de la función $\frac{\text{sen}[\theta(u)]}{u\rho(u)}$.

El error de truncado, expresado en este caso como

$$\epsilon_t = \frac{1}{\pi} \int_R^\infty \frac{\text{sen}[\theta(u)]}{u\rho(u)} du \quad (3.25)$$

tiene un límite superior ϵ_t , es decir $|\epsilon_t| < \epsilon_t$, donde

$$T = \left[\epsilon_t \pi \frac{n-p}{2} \prod_{i=1}^{n-p} |\gamma_i - I_O|^{1/2} \right]^{-2/(n-p)} \quad (3.26)$$

Es importante anotar que el error de truncado no está definido en $\gamma_i = I_O$ con $(i = 1, \dots, n-p)$. Esta discontinuidad se puede corregir como lo sugirió Farebrother en 1980.

De esta manera, la integral se podría aproximar como

$$F\left(\frac{I_O}{H_O}\right) \approx \frac{1}{2} - \left[\frac{1}{\pi} \left(\frac{1}{2} \sum_{i=1}^{n-p} (\gamma_i - I_O) \right) + \left(\epsilon_t \pi \frac{n-p}{2} \prod_{i=1}^{n-p} |\gamma_i - I_O|^{1/2} \right)^{-2/(n-p)} \right] \quad (3.27)$$

Numéricamente, el método de Imhof cuenta con varias ventajas:

1. Su simplicidad, al ser una integral real comparada con la complejidad de la transformada integral inversa original de Fourier con partes imaginarias,
2. su generalidad, al no tener casos especiales y particularidades,
3. sus propiedades matemáticas bien conocidas.

La calidad del procedimiento numérico aplicado puede ser evaluado comparando los momentos teóricos basados en los valores propios con los momentos obtenidos de la función de distribución calculada. Esta calidad depende del tamaño de los subintervalos y del error de truncado T , adicionalmente la suavidad de la función de distribución calculada depende del número de puntos de soporte escogidos en el intervalo (γ_1, γ_{n-p}) . Para estructuras espaciales pequeñas, se ha encontrado que los valores teóricos y calculados de variancia, desviación y curtosis son idénticos hasta el cuarto dígito de significancia, con una tendencia a mejorar en la medida en que las estructuras espaciales crecen; las medias se observan iguales.

Capítulo 4

Caso de Aplicación

4.1. Introducción

En este capítulo se presenta un caso de aplicación del Análisis Exploratorio de Datos espaciales con los datos suministrados por la Secretaria de Gobierno del Departamento de Antioquia. La variable a analizar es el número de homicidios en cada uno de los 125 municipios que conforman el departamento durante los años 2001 a 2006.

En los estudios de fenómenos sociales la ubicación geográfica juega un papel relevante que debe ser tomada en cuenta al construir modelos en los cuales intervengan variables relacionadas con la ubicación geográfica. Estudios de este tipo de fenómenos que tengan en cuenta de manera explícita el concepto de territorio son muy escasos en Colombia. Por lo tanto se considera que realizar aplicaciones con datos de Colombia es un paso muy importante para futuros trabajos.

Al realizar esta propuesta utilizando de manera explícita la información geográfica, se muestra la importancia que tiene el concepto de región para análisis con los datos de Colombia, esto se ve reflejado en el hecho de que a través del estudio se muestra que los fenómenos de violencia en Colombia no son procesos que se produzcan de manera aleatoria sino que obedecen a patrones.

4.2. Departamento de Antioquia

El Departamento de Antioquia está localizado en la zona noroccidental de la República de Colombia, tiene un área de $63,612 \text{ km}^2$ y ocupa el sexto puesto en tamaño dentro de los departamentos de Colombia. El departamento está conformado por 125 municipios los cuales se han agrupado de acuerdo con su ubicación en nueve subregiones. En la figura 4.1 se presenta el mapa de Antioquia con sus municipios y sus subregiones

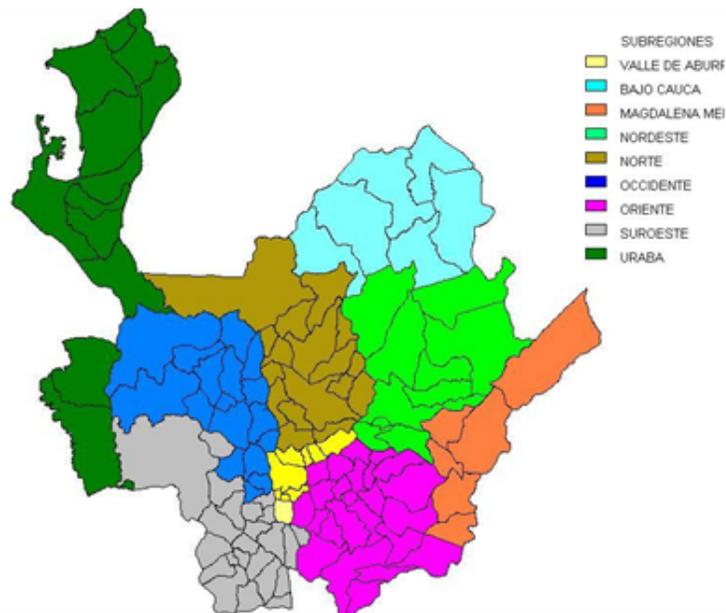


Figura 4.1: Subregiones del Departamento de Antioquia

4.3. Definición de variables

X_i : Es el número de homicidios en el municipio i . $i = 1, \dots, 125$. Se tuvo en cuenta en el análisis la información de la variable X_i para los años 2001 a 2006.

Dado que los municipios tienen áreas disímiles entre si y con el fin de asegurar la

comparabilidad de los valores de la variable en estudio en los diferentes municipios, se convirtió el número de homicidios en una tasa de homicidios por cada 100.000 habitantes de acuerdo con la siguiente expresión:

$$TasaHomicidios_i = \frac{X_i * 100,000}{Poblacion_i} \quad (4.1)$$

La información sobre la variable objeto de estudio fue procesada utilizando el programa GeoDa (Geodata Analysis Software), desarrollado por Luc Anselin y sus colaboradores en la Universidad de Illinois, para realizar Análisis Exploratorio de Datos Espaciales.

4.4. Resultados Obtenidos

A continuación se presentan el reporte de los resultados que arroja el GeoDa y se realizan los análisis correspondientes.

4.4.1. Identificación de valores atípicos

Diagrama de Caja - Mapa

En las figuras 4.3 a 4.8 se presentan los diagramas de Caja-Mapas para los años 2001 a 2006. En estos diagramas los municipios que aparecen resaltados en el mapa son aquellos que tuvieron las mayores tasas de homicidios y corresponden a los valores atípicos o *outliers*, a su vez esta información aparece en los puntos resaltados con amarillo en el diagrama de caja del lado derecho de la figura. En la parte inferior se observa la tabla donde aparecen los nombres de los municipios resaltados en el mapa y la subregión a la cual pertenecen.

Un análisis de los reportes permite afirmar que el fenómeno de violencia presenta una estructura en la cual el fenómeno aparece agrupado en municipios con altos índices

de violencia. Los municipios con mayores tasas de homicidios se encuentran en la misma zona para casi todos los años observados y algunos municipios se repiten en el transcurso de estos, esto nos lleva a conjeturar que el fenómeno de violencia no se distribuye en forma aleatoria.

Mapa de Caja o *Boxmap*

En las figuras 4.9 a 4.14 se presentan los BoxMap para cada uno de los años considerados. En este gráfico aparecen sombreados con seis colores diferentes los municipios, de acuerdo con su ubicación en cada una de las regiones del Boxplot. De este reporte se puede concluir como se están distribuyendo los municipios de acuerdo con la zona del Boxplot a la cual pertenecen. En muchos casos municipios que pertenecen a una de las zonas del boxplot se encuentran cercanos a municipios que pertenecen a la misma zona del mismo. Esto significa que no solamente se encuentra agrupación de municipios con valores atípicos, sino también de municipios que pertenecen a otra localización dentro del Boxplot.

4.4.2. Identificación de la distribución de la variable

Histograma - Mapa

En las figuras 4.15 a 4.20 se presentan los histogramas de la variable Tasa de Homicidios para cada uno de los años considerados. El histograma es una aproximación de la distribución de la variable en la zona de estudio y en este tipo de reporte es posible obtener también la aproximación a la distribución de la variable en una subregión de la zona de estudio, esta distribución aparece resaltada en color amarillo sobre el histograma de la zona de estudio. Al comparar los dos histogramas, el de la zona de estudio y el de la región se puede detectar heterogeneidad entre las dos cuando

el histograma regional no presenta una estructura similar al histograma general, esta heterogeneidad espacial significa que no hay uniformidad de los efectos espaciales entre la zona de estudio y la región. En los reportes se han resaltado los municipios que conforman algunas regiones con el fin de comparar el histograma del área de estudio con el histograma regional. Por ejemplo en la figura 4.19, la distribución de la subregión suroeste tiene la misma estructura que la distribución del total de municipios, por lo cual se puede concluir que no hay heterogeneidad espacial entre la zona de estudio y la región; en cambio en la figura 4.15 la distribución aproximada de la subregión oriente no presenta una estructura similar a la del departamento, por lo cual podemos concluir que esta región presenta heterogeneidad espacial con respecto al total del departamento.

Mapa de cuantiles

En las figuras 4.21 a 4.26 se presentan los mapas de cuantiles donde aparecen la distribución de las unidades espaciales de acuerdo con los cuantiles. Los municipios que pertenecen al mismo cuartil tienen el mismo color. La interpretación de este reporte es similar a la del BoxMap, ya que se puede ver como se agrupan los municipios que pertenecen a cada uno de los cuantiles, dentro de la zona de estudio.

4.4.3. Identificación de efectos espaciales

Cartograma

Las figuras 4.27 a 4.32 corresponden a los cartogramas para los años de estudio, donde cada unidad espacial es reemplazada por un círculo cuya área es proporcional al valor de la variable en dicha unidad espacial. En el cartograma las unidades espaciales con los valores atípicos se resaltan con colores diferentes del resto de las unidades, en este caso aparecen resaltados con color amarillo. El cartograma permite la identificación

de las unidades con valores atípicos y adicionalmente permite comparar visualmente la relación que tienen las unidades con valores atípicos y las unidades con valores no atípicos.

Diagrama de dispersión de Moran

El *scatterplot* de Moran es un diagrama de dispersión donde se representa la variable en estudio y el retardo espacial de dicha variable. El estadístico de prueba I de Moran para contrastar la autocorrelación espacial es el estimador de la pendiente de la regresión por mínimos cuadrados ordinarios. En las figuras 4.33 a 4.38 se presentan los diagramas de dispersión de Moran de cada uno de los años estudiados. En estos reportes se observa en la parte superior del diagrama de dispersión el valor del estadístico de prueba I de Moran. En la parte derecha del diagrama aparece el cálculo de la media de I $E(I)$, la desviación estándar $V(I)$ y el P valor de la prueba. La conclusión para cada uno de los años considerados es el rechazo de la hipótesis nula de no autocorrelación espacial, es decir existe autocorrelación espacial entre los datos considerados, lo cual implica que modelos econométricos que se construyan utilizando en su especificación la variable tasa de homicidios, y en los cuales no se considere la ubicación del municipio dentro de la zona de estudio, quedarán mal especificados con las implicaciones que ello conlleva.

4.5. Identificación de *clusters*

A continuación se presenta el gráfico correspondiente a los *clusters* encontrados, después de correr un algoritmo del tipo AMOEBA con los datos para el año 2006. Este algoritmo fue programado en MATLAB.

En la figura 4.2 se pueden identificar 17 *clusters* con los municipios del departamento de Antioquia, los *clusters* son agrupaciones de unidades espaciales, en nuestro

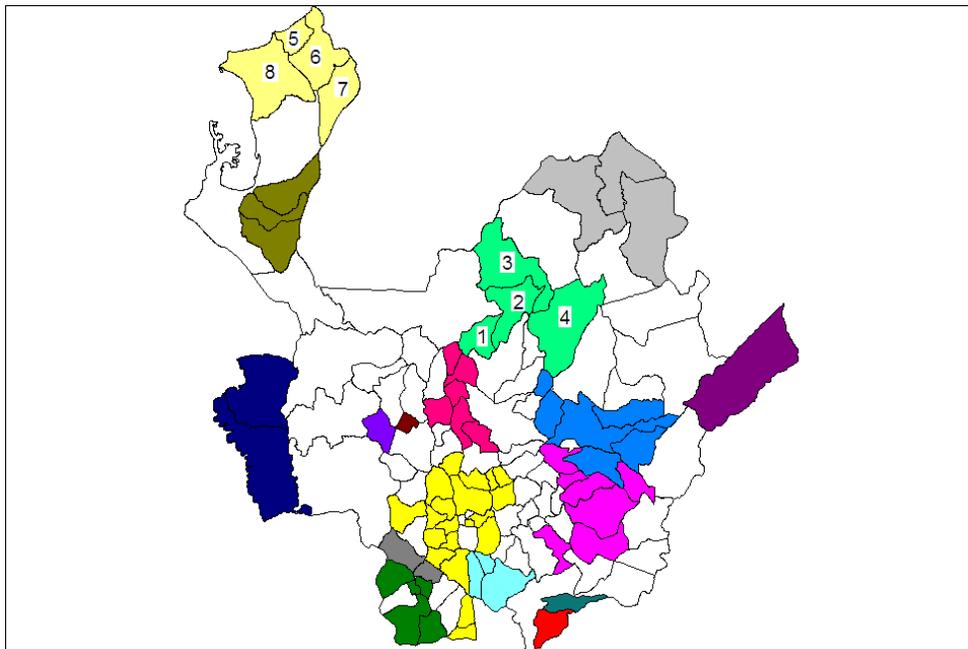


Figura 4.2: *Clusters* identificados para el año 2006

caso de municipios que tienen características similares en la variable observada. Se encontraron *clusters* de diferentes tipos, dentro de los que sobresalen, aquellos en los que municipios con tasas de homicidios altas están rodeados de municipios con tasas de homicidios también altas como es el *cluster* formado por Tarazá (3), Anorí (4), Valdivia (2) y Briceño (1) y municipios con tasas de homicidios bajas están rodeados de municipios con tasas de homicidios bajas como corresponde al *cluster* conformado por Arboletes (6), San Juan de Urabá (5), Necoclí (8) y San Pedro de Urabá (7).

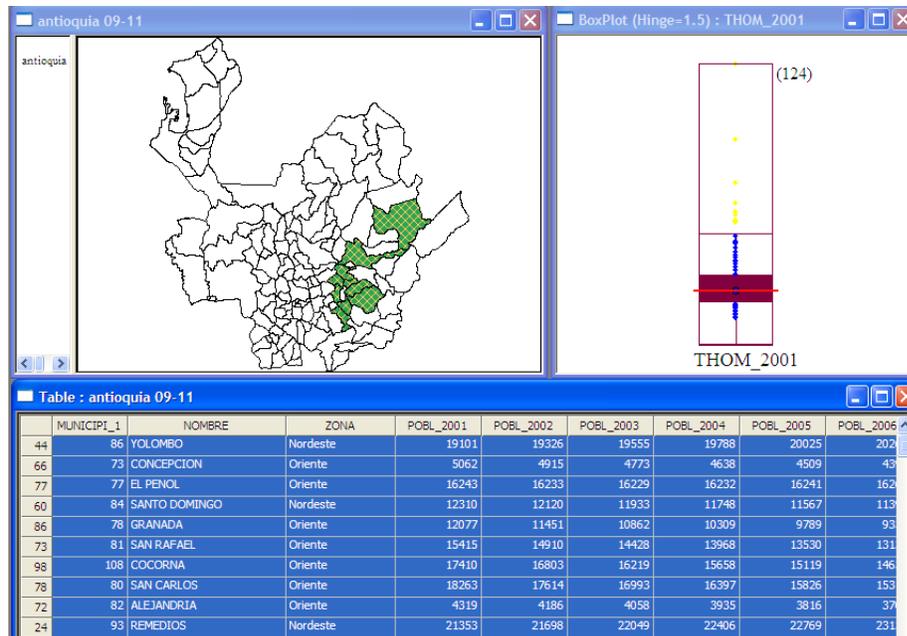


Figura 4.3: Diagrama de Caja - Mapa para el año 2001

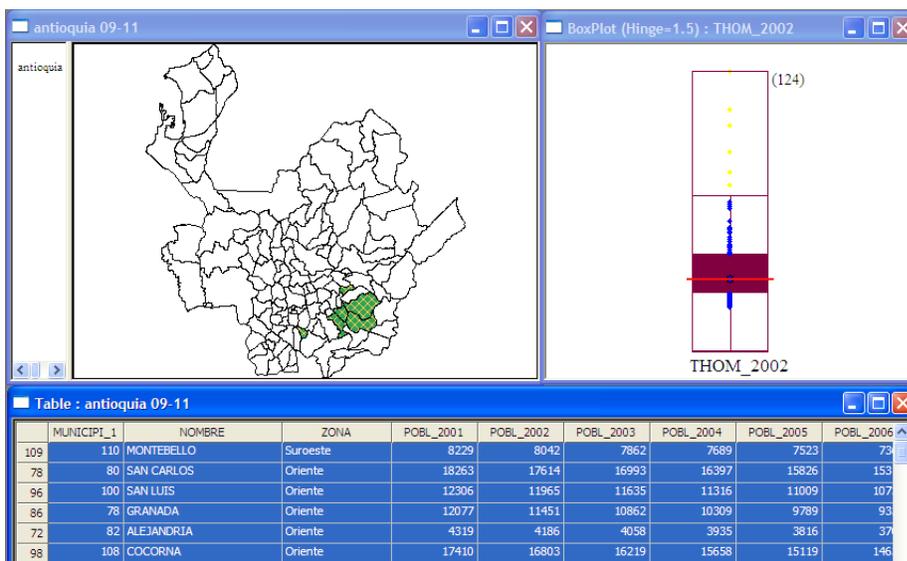


Figura 4.4: Diagrama de Caja - Mapa para el año 2002

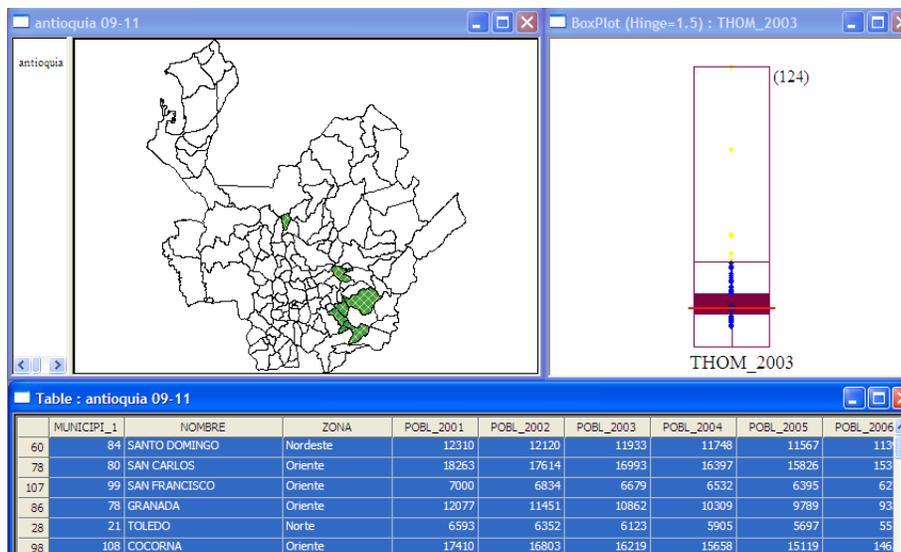


Figura 4.5: Diagrama de Caja - Mapa para el año 2003

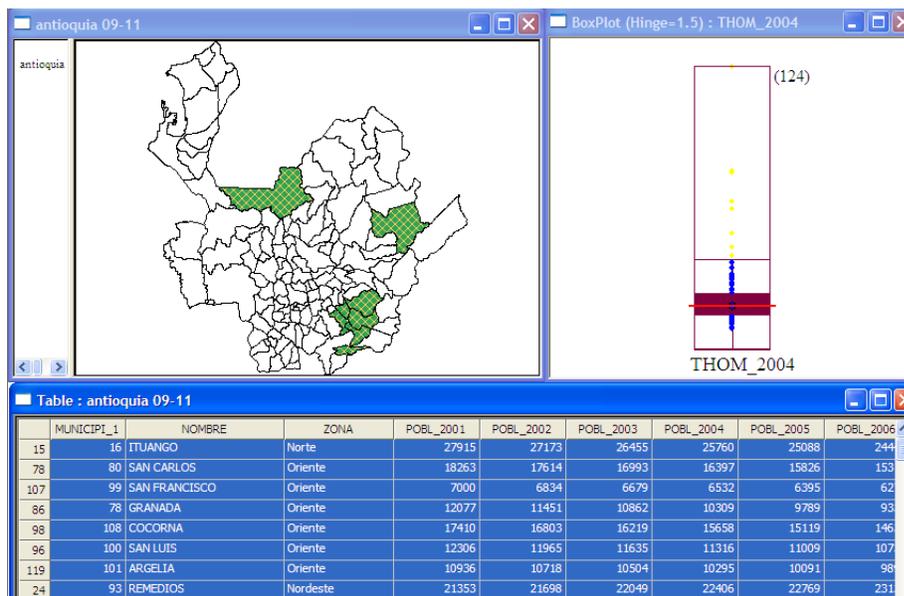


Figura 4.6: Diagrama de Caja - Mapa para el año 2004

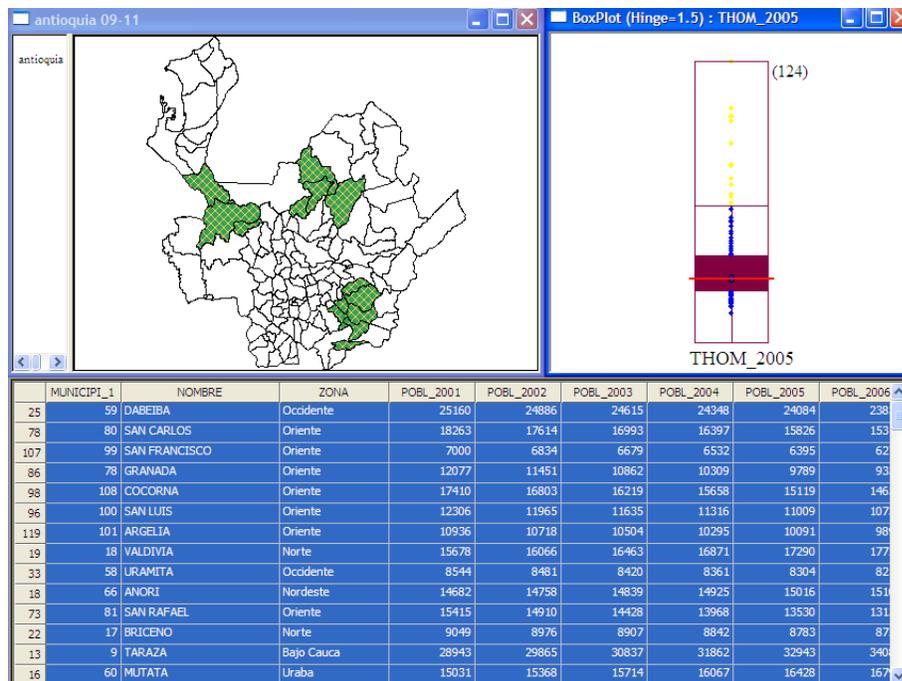


Figura 4.7: Diagrama de Caja - Mapa para el año 2005

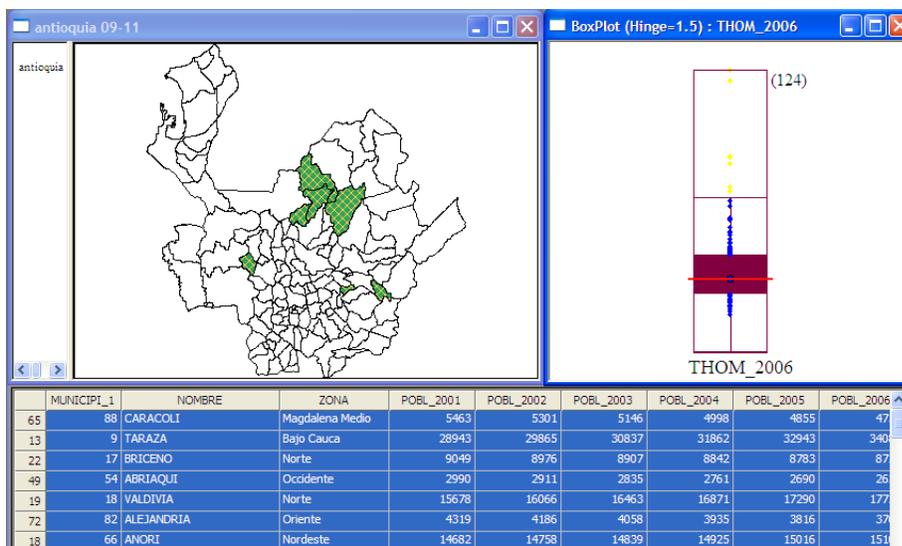


Figura 4.8: Diagrama de Caja - Mapa para el año 2006

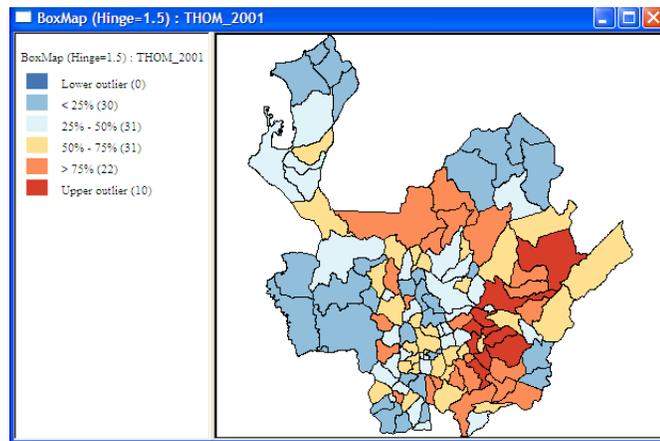


Figura 4.9: Boxmap para el año 2001

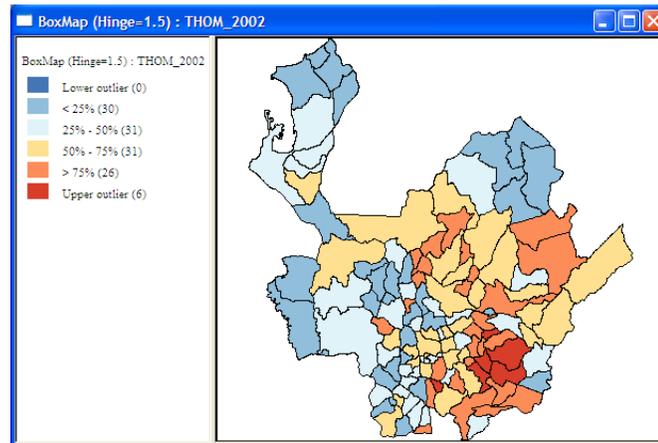


Figura 4.10: Boxmap para el año 2002

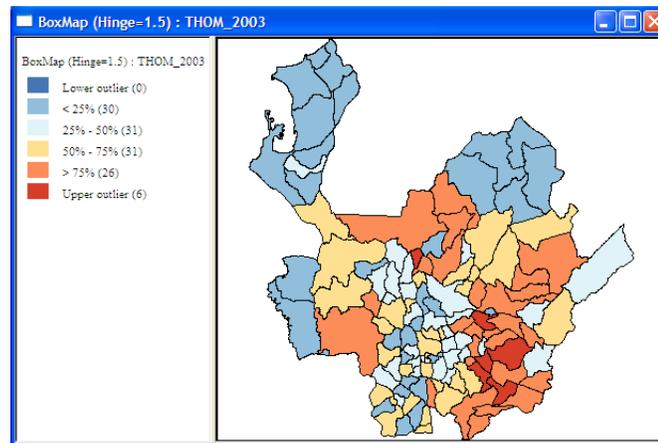


Figura 4.11: Boxmap para el año 2003

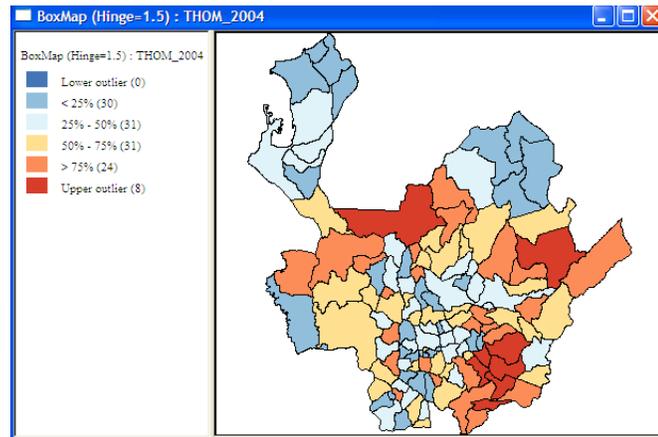


Figura 4.12: Boxmap para el año 2004

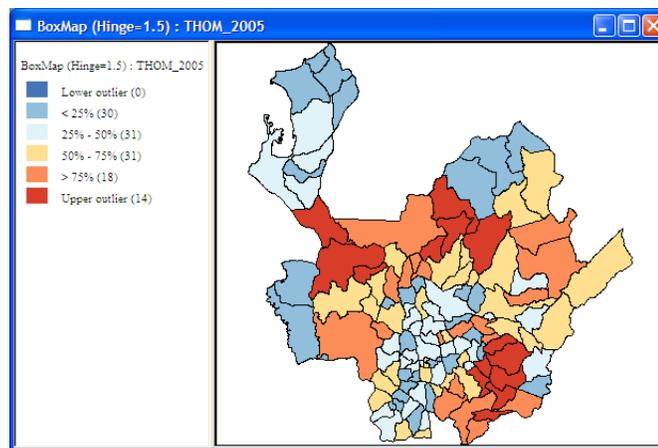


Figura 4.13: Boxmap para el año 2005

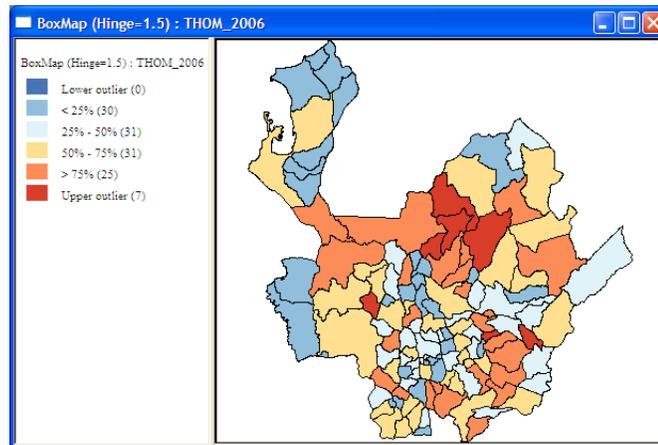


Figura 4.14: Boxmap para el año 2006

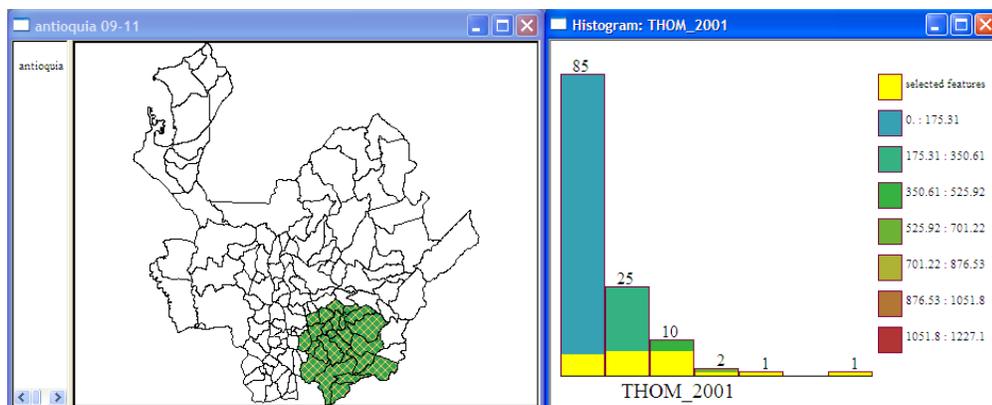


Figura 4.15: Histograma para el año 2001 - Subregión Oriente resaltada

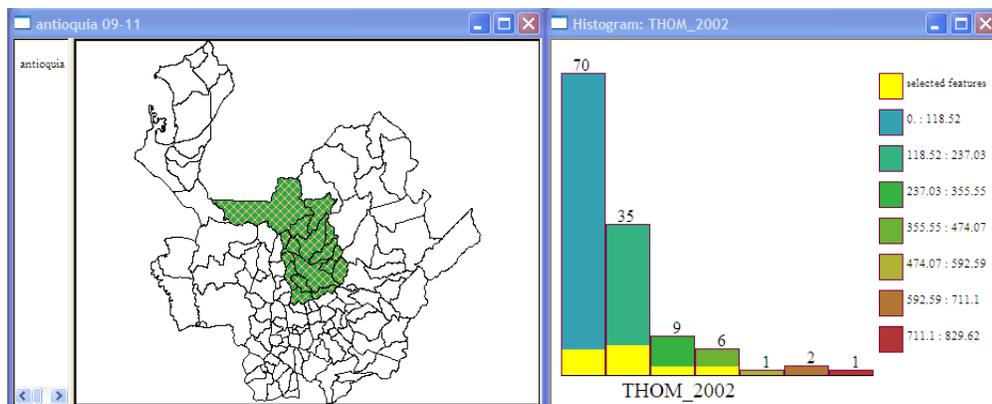


Figura 4.16: Histograma para el año 2002 - Subregión Norte resaltada

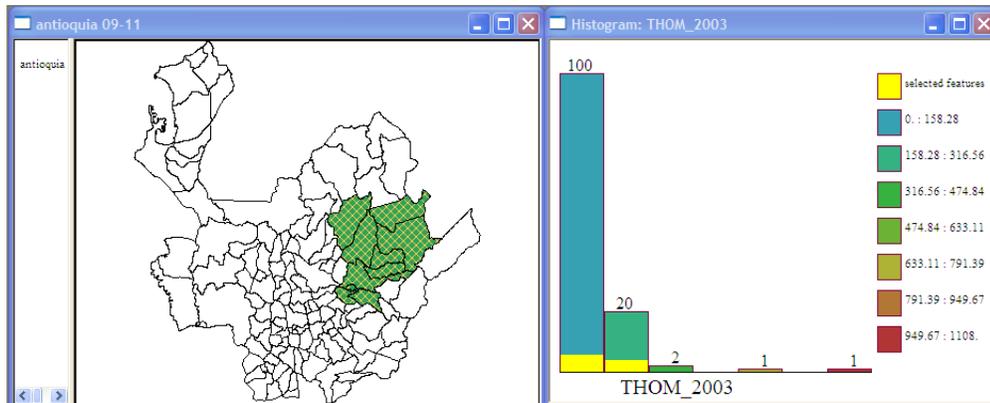


Figura 4.17: Histograma para el año 2003 - Subregión Nordeste resaltada

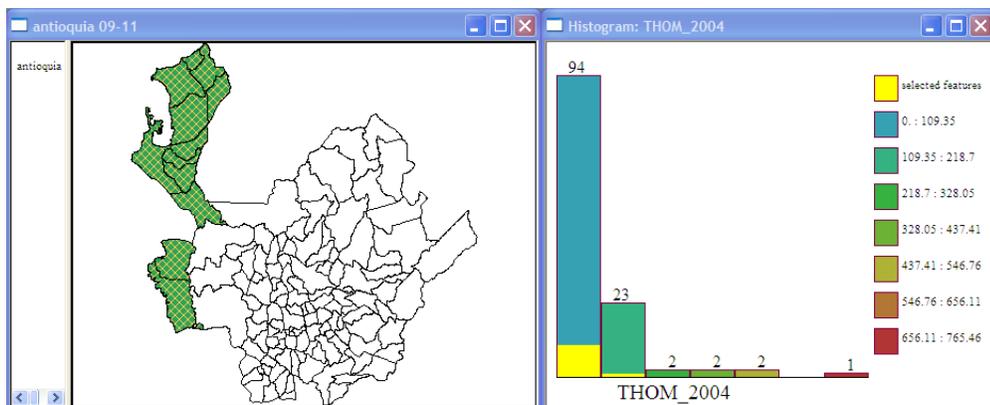


Figura 4.18: Histograma para el año 2004 - Subregión Urabá resaltada

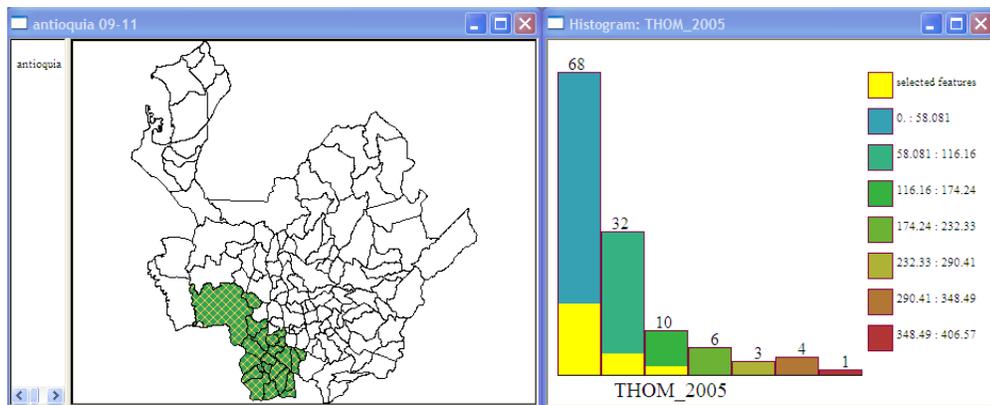


Figura 4.19: Histograma para el año 2005 - Subregión Suroeste resaltada

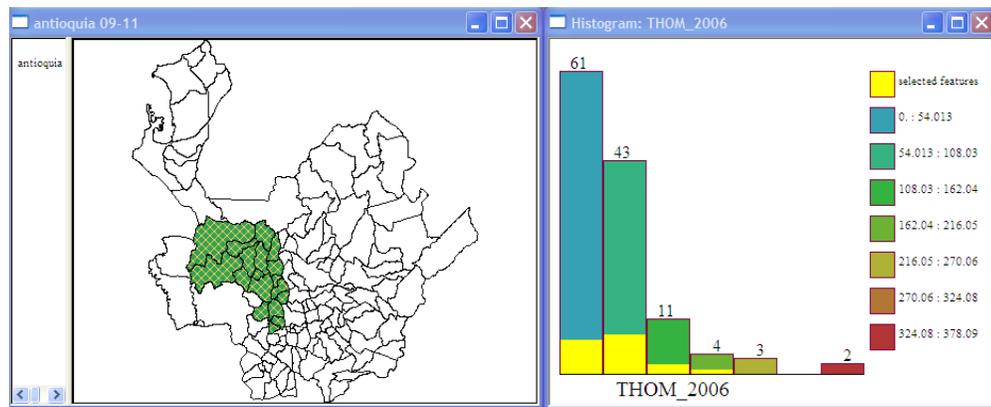


Figura 4.20: Histograma para el año 2006 - Subregión Occidente resaltada

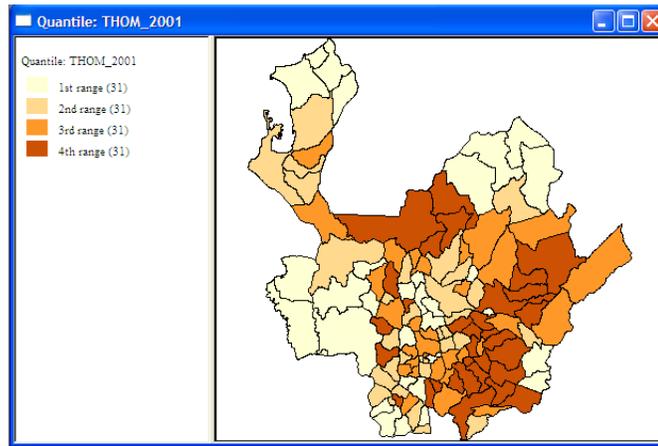


Figura 4.21: Mapa de Cuartiles para el año 2001

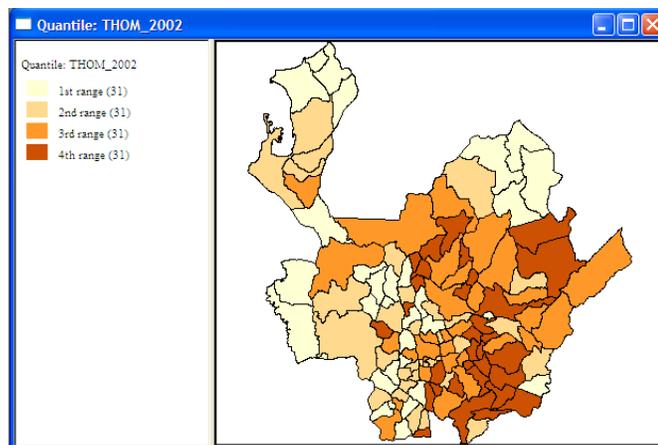


Figura 4.22: Mapa de Cuartiles para el año 2002

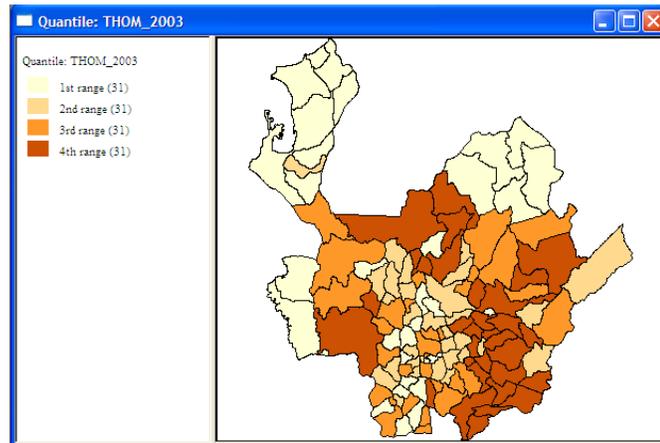


Figura 4.23: Mapa de Cuartiles para el año 2003

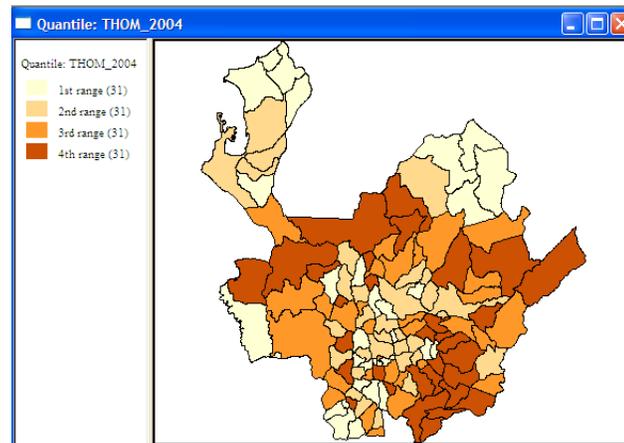


Figura 4.24: Mapa de Cuartiles para el año 2004

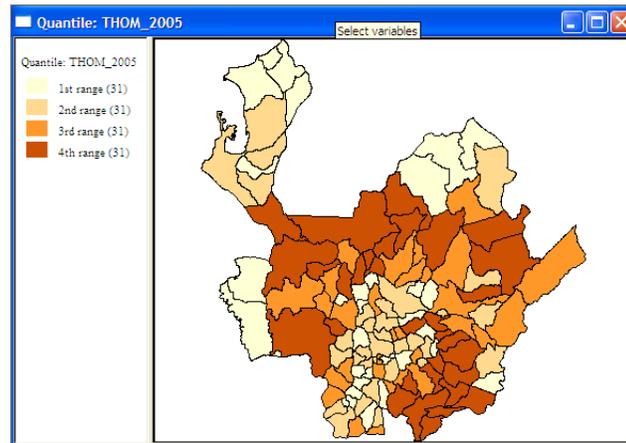


Figura 4.25: Mapa de Cuartiles para el año 2005

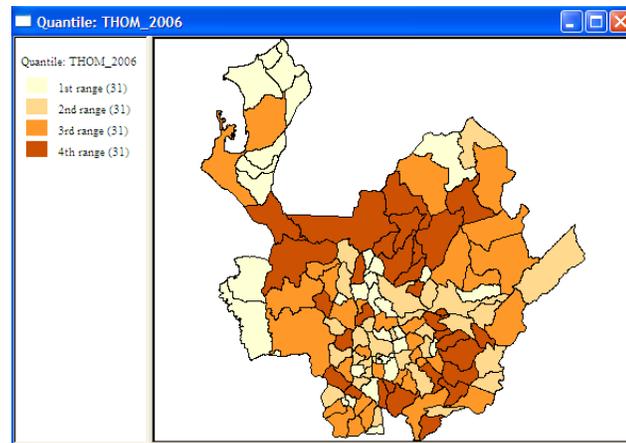


Figura 4.26: Mapa de Cuartiles para el año 2006

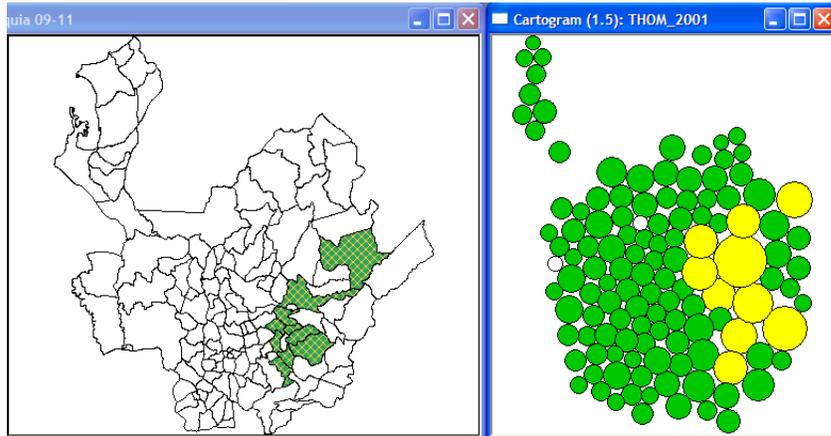


Figura 4.27: Cartograma para el año 2001

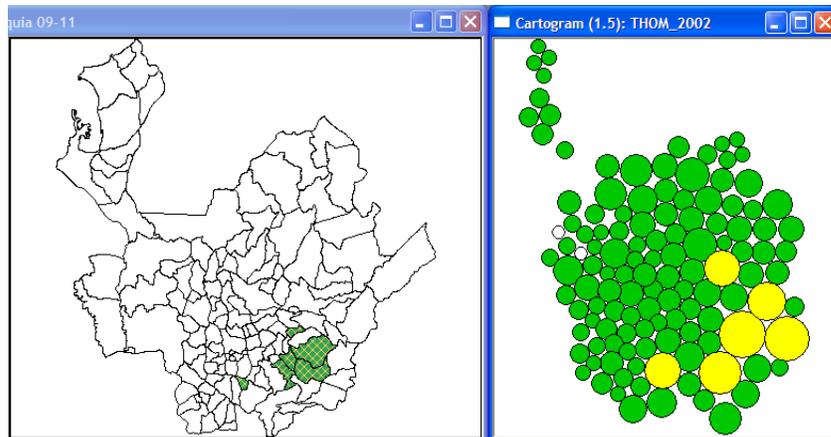


Figura 4.28: Cartograma para el año 2002

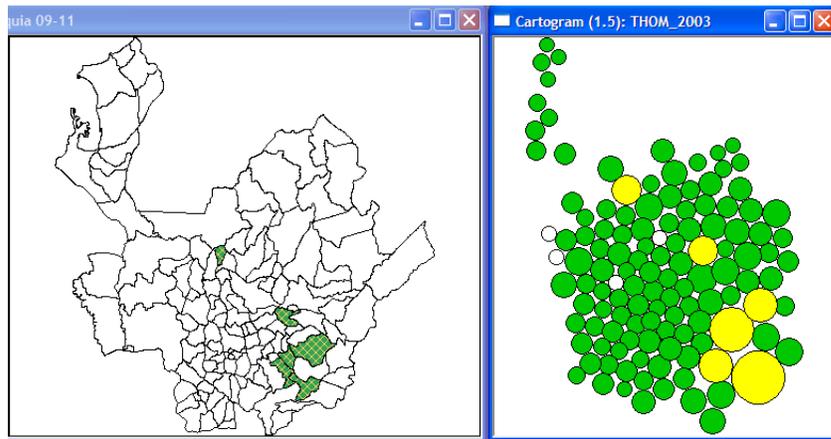


Figura 4.29: Cartograma para el año 2003

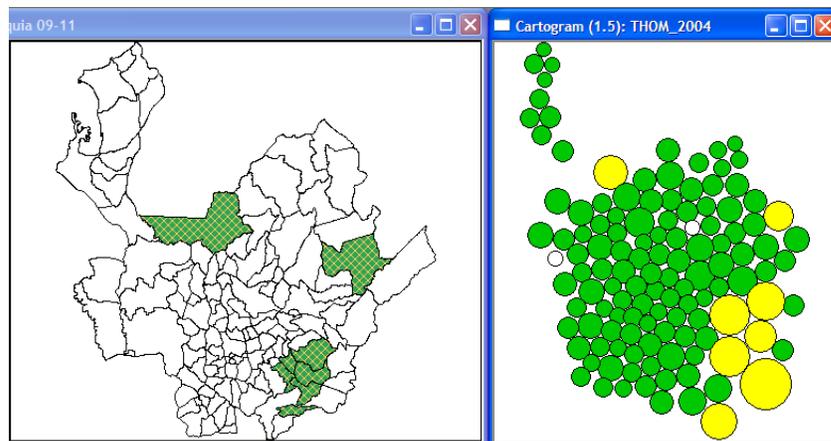


Figura 4.30: Cartograma para el año 2004

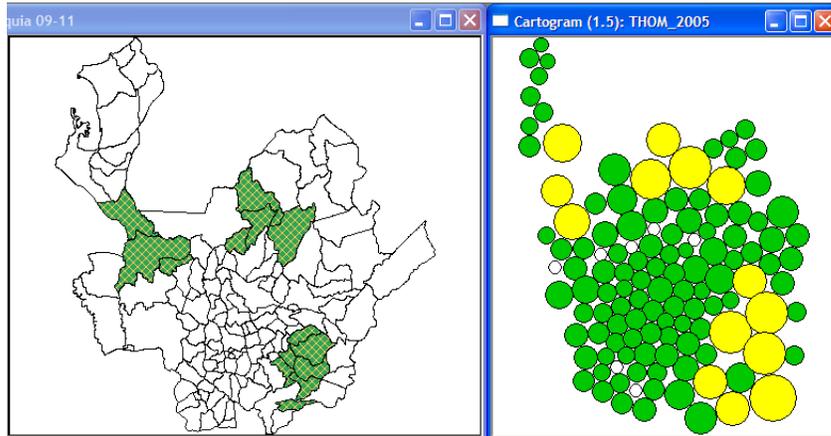


Figura 4.31: Cartograma para el año 2005

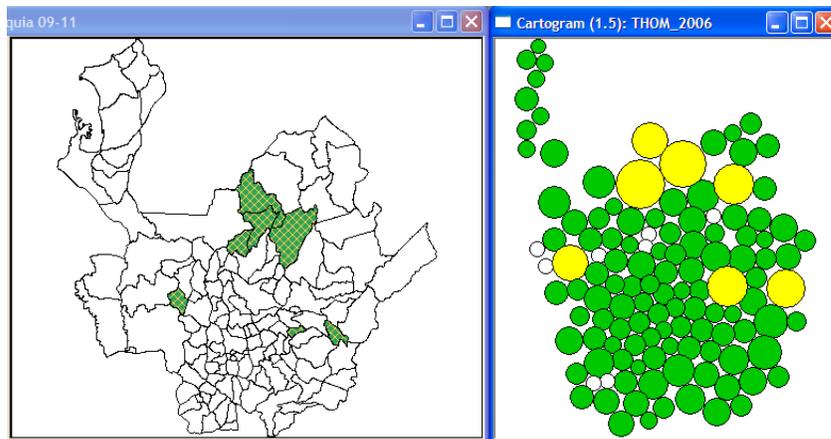


Figura 4.32: Cartograma para el año 2006

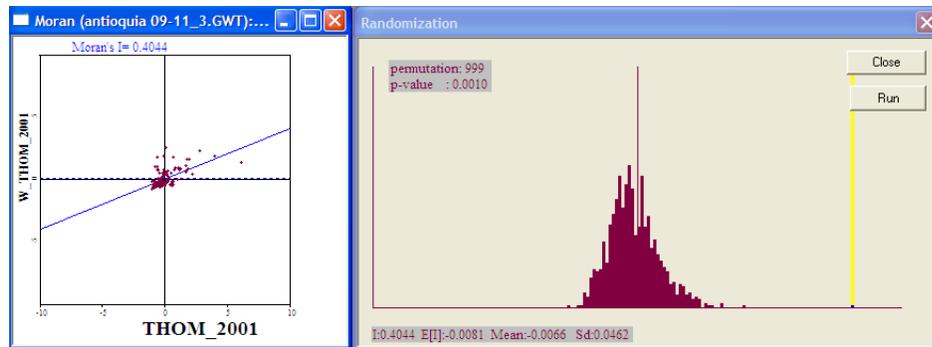


Figura 4.33: Diagrama de Dispersión de Moran para el año 2001

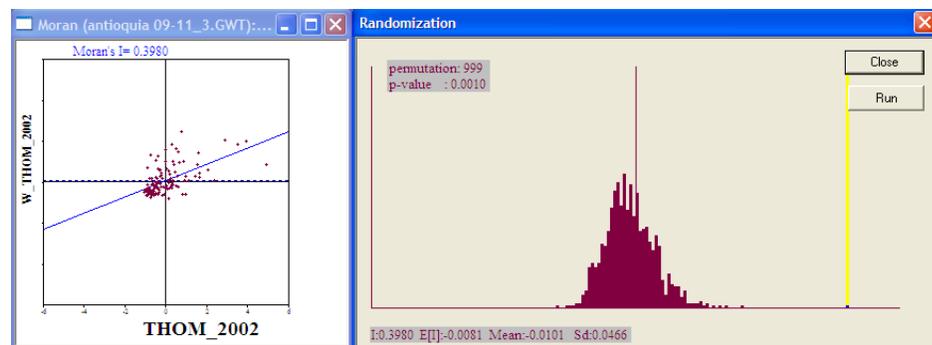


Figura 4.34: Diagrama de Dispersión de Moran para el año 2002

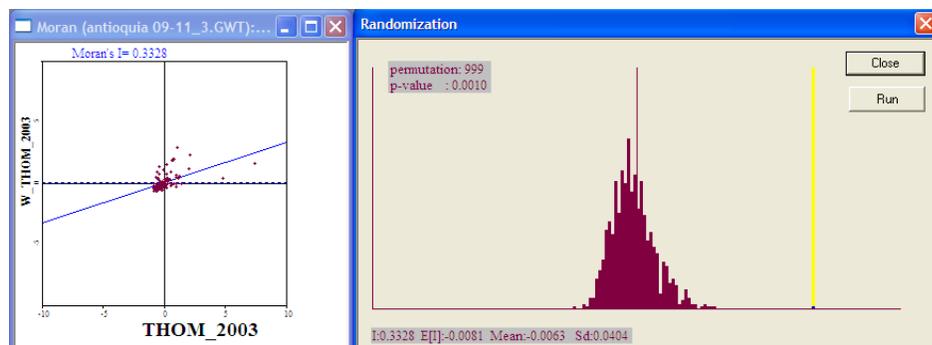


Figura 4.35: Diagrama de Dispersión de Moran para el año 2003

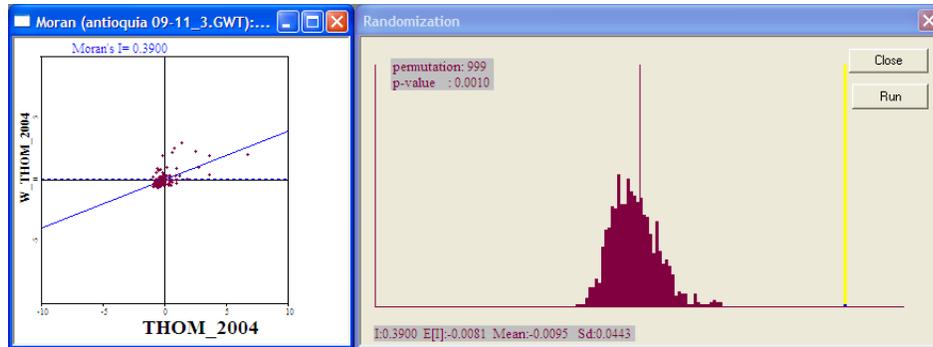


Figura 4.36: Diagrama de Dispersión de Moran para el año 2004

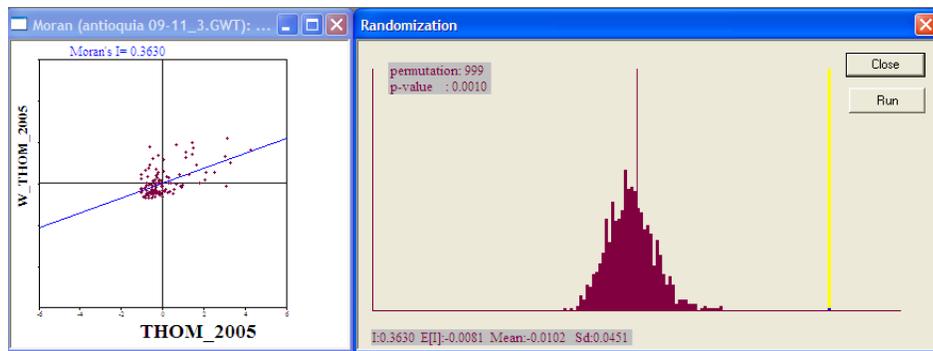


Figura 4.37: Diagrama de Dispersión de Moran para el año 2005

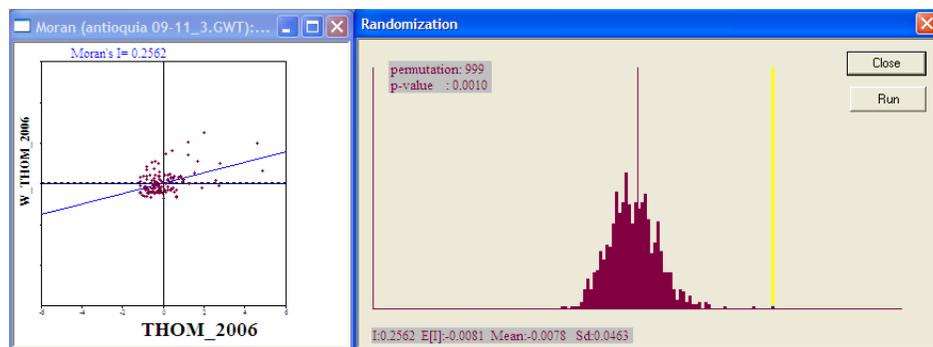


Figura 4.38: Diagrama de Dispersión de Moran para el año 2006

Capítulo 5

Conclusiones

De acuerdo con el desarrollo del trabajo y los resultados obtenidos podemos concluir que si se tienen datos georreferenciados que se pretende usar para construir modelos, el análisis de la información debe realizarse en la primera etapa con las técnicas del Análisis Exploratorio de Datos Espaciales para comprobar si existe algún tipo de efecto espacial entre los datos. En caso de existir algún efecto espacial y de no realizarse este AEDE los modelos construidos con estos datos tendrán errores de especificación. Adicionalmente si se tiene en cuenta que el espacio es un elemento estructural fundamental en explicaciones sobre el comportamiento humano y las actividades económicas, las variables de estudio relacionadas con los fenómenos mencionados se deben analizar mediante AEDE y no mediante análisis de datos tradicional.

Por otra parte los resultados del caso de estudio sobre la violencia en Antioquia indican que este fenómeno de violencia no tiene una distribución aleatoria sino que presenta patrones espaciales de comportamiento, donde la violencia en un municipio contagia a los municipios vecinos. Esto nos lleva a concluir que existen muchos fenómenos sociales en el país que podrían ser analizados mediante estas técnicas y así encontrar patrones de comportamiento espacial que sirvan a los gobiernos de carácter regional y nacional para proponer estrategias políticas orientadas hacia esos patrones.

Si se hiciera una búsqueda exhaustiva de estudios previos donde se hayan construido modelos sobre este tipo de fenómenos, se encontraría que dichos modelos deben ser reevaluados a la luz de la posibilidad de la existencia de efectos espaciales entre los datos.

El contenido del presente trabajo constituye la etapa inicial del trabajo con datos georreferenciados, como son el AEDE y la prueba de hipótesis de autocorrelación espacial. Consideramos que en etapas posteriores se pueden emprender proyectos de investigación en el área de la econometría espacial cuyos tópicos entre otros podrían ser: planteamiento de diferentes modelos teniendo en cuenta las condiciones del problema a estudiar y su estimación; construcción de matrices de contigüidad usando sistemas diferentes al binario y criterios diferentes al de la vecindad, por ejemplo utilizando diferentes medidas de similaridad que respondan adecuadamente al problema presentado; estudio de los efectos de la agregación espacial en los datos; características por considerar en el muestreo para obtener mayor eficacia en datos georreferenciados.

Bibliografía

- [1] Aldstadt J., Getis A., *Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters*. Geographical Analysis ISSN 0016-7363. The Ohio State University **38** (2006) , 327–343.
- [2] Anderson T.W., *An introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Mathematical Statistics, Second Edition 1984
- [3] Anselin Luc, *Exploring Spatial Data with GeoDaTM : A Workbook*, Spatial Analysis Laboratory Department of Geography University of Illinois,(2005), <http://sal.agecon.uiuc.edu/>
- [4] Anselin Luc, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht/Boston/London , 1988.
- [5] Anselin Luc, Florax Raymond J.G.M. *Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results*, Anselin, L., Florax, R. (Eds.), *New Directions in Spatial Econometrics*. Springer Verlag, New York, (1995) pp. 21–74.
- [6] Anselin Luc, Kelejian Harry H., *Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors*, *International Regional Science Review* 20, 1& 2 (1997), pp. 153–182

- [7] Chasco Yrigoyen C., *Métodos gráficos del análisis exploratorio de datos*, www.uam.es/coro.chasco/investigacion/aede03.pdf
- [8] Chasco Yrigoyen C., *Tesis Doctoral: Econometría Espacial Aplicada a la Predicción-Extrapolación de Datos Microterritoriales*, Consejería de Economía e Innovación Tecnológica, (2003)
- [9] Chasco Yrigoyen C., *Análisis estadístico de datos geográficos en geomarketing: el programa GeoDa*, Distribución y Consumo, Marzo-Abril 2006, pp. 34-45
- [10] Cliff A. And J. Ord, *Testing for Spatial Autocorrelation Among Regression Residuals*, *Geographical Analysis*, 4 1972, 267 – 284.
- [11] Díaz M. Luis G., *Estadística Multivariada, inferencia y métodos*, Facultad de Ciencias, Universidad Nacional de Colombia, 2002
- [12] Fotheringham A. Stewart, Brunsdon Chris y Charlton Martin, *Quantitative Geography*, Sage Publications, London, 2000
- [13] Getis, A. *Reflections on spatial autocorrelation*. *Reg. Sci. Urban Econ.* (2007), doi:10.1016/j.regsciurbeco.2007.04.005
- [14] Haining R., *Spatial Data Analysis, Theory and Practice*. Cambridge, Cambridge University Press, UK, 2003.
- [15] Hendry David F., *Advanced Text in Econometrics, Dynamic Econometrics*, Oxford University Press, 1995
- [16] Kelejian Harry H., Prucha Ingmar R. *On the asymptotic distribution of the Moran I test statistic with applications*, *Journal of Econometrics* 104 (2001) 219–257
- [17] LeSage James P. , *Spatial Econometrics*, Department of Economics University of Toledo, CIRCULATED FOR REVIEW December, 1998

- [18] Magnus Jan R., Neudecker Heinz, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley Series in Probability and Statistics, Edición revisada
- [19] Moran P. A. P. *The Interpretation of Statistical Maps*, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 10, No. 2. (1948), pp. 243–251
- [20] Moran P. A. P., *Notes on Continuous Stochastic Phenomena*, Biometrika, Vol. 37, No. 1/2. (Jun., 1950), pp. 17–23.
- [21] Moreno R. y Vayá, E., *Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial*, Universitat de Barcelona, 2000, Barcelona
- [22] O’Sullivan D. y Unwin D. *Geographical Information Analysis*, John Wiley & sons, Inc., 2002
- [23] Ord, J., *Estimation methods for Models of Spatial Interaction*, Journal of the American Statistical Association, 1975, **70**, 120–26
- [24] Vilalta y Perdomo Carlos Javier , *Cómo enseñar autocorrelación espacial*, Economía, Sociedad y Territorio, Volumen v **18**, 2005, 323–333
- [25] Wasserman, *All of Statistics, A concise Course in Statistical Inference*, Springer, 2003