

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

IDENTIFICACIÓN DE CLUSTERS ESPACIALES DE
EMPRESAS Y LA INFLUENCIA DE FACTORES
EXTERNOS EN SU CONSTITUCIÓN UTILIZANDO
TÉCNICAS DE MINERÍA DE DATOS

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO
MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

JORGE LUIS JÁCOME CORRALES

jorge.jacome@epn.edu.ec

DIRECTOR: PHD. MIGUEL ALFONSO FLORES SÁNCHEZ

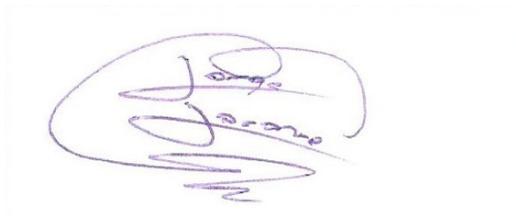
miguel.flores@epn.edu.ec

QUITO, OCTUBRE 2021

DECLARACIÓN

Yo JORGE LUIS JÁCOME CORRALES, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.



Jorge Luis Jácome Corrales

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por JORGE LUIS JÁCOME CORRALES, bajo mi supervisión.



PhD. Miguel Alfonso Flores Sánchez.
Director del Proyecto

AGRADECIMIENTOS

A Dios por guiar mi vida y ayudarme en mis metas.

A mis padres por todo el apoyo que me han otorgado.

A mi hermana Carla y mi cuñado Richard, que siempre me ayudan sin importar nada.

A mis sobrinos, por darme más motivos para no rendirme.

Al PhD. Miguel Flores por su guía durante toda la carrera.

A mis profesores que cada uno dio un granito para terminar este trabajo.

A mis amigos de la universidad con los cuales estudiaba sin dejar de lado la diversión.

Finalmente, a todos mis compañeros de trabajo cuyas insistencias me hacen mejorar día a día.

DEDICATORIA

A mis padres, a mi hermana y a mis sobrinos

Índice general

Resumen	IX
Abstract	X
1. Introducción	1
1.1. Antecedentes	1
1.2. Objetivos	5
1.3. Justificación	5
2. Análisis Exploratorio de Datos Espaciales	9
2.1. Datos Espaciales	9
2.2. Información a Utilizar y Procedencia	12
2.3. Tratamiento de Datos	13
2.3.1. Georeferenciación de Datos	13
2.3.2. Igualdad de Sistemas de Referencia	15
2.4. Visualización de Datos	16
3. Índices De Autocorrelación Espacial	19
3.1. Autocorrelación espacial	19
3.1.1. Matriz de Contigüidad y Matriz de Distancias	20
3.2. Indicador Global	21
3.2.1. Indicador de Moran	21
3.2.2. Prueba de Hipótesis y Valor-p	22
3.3. Indicador Local	24
3.3.1. Indicador de Moran	24

3.4. GeoDa	25
4. Modelado de Datos	27
4.1. Modelo de regresión lineal	27
4.2. Modelo de retardo espacial	28
4.3. Modelo de error espacial	29
4.4. Modelo de Orden Superior	30
4.5. Modelo espacial de Durbin	30
4.6. Prueba para Dependencia Espacial	31
4.7. Estimación	32
4.7.1. Estimación por Máxima Verosimilitud	32
4.7.2. Estimación por Variables Instrumentales	36
4.7.3. Comparativo de las Estimaciones	38
4.8. Interpretación de Parámetros Estimados	39
5. Resultados (Caso de Estudio)	41
6. Conclusiones y Recomendaciones	65
6.1. Conclusiones	65
6.2. Recomendaciones	67
Bibliografía	69
A. Anexo 1. Datos Espaciales	72
B. Anexo 2. Código de Georeferenciación	81
C. Anexo 3. Análisis Espacial por Actividades de las Empresas	85
C.1. Análisis a empresas de Comercio	86
C.2. Análisis a empresas Científico Técnicas	90
C.3. Análisis a empresas Administrativas	94
C.4. Análisis a empresas de Manufacturas	98
C.5. Análisis a empresas de Transporte	102
C.6. Análisis a empresas de Información	106
C.7. Análisis a empresas de Construcción	110

D. Anexo 4. Barrios con alta densidad de Empresas	114
E. Anexo 5. Análisis Espacial de las Variables	118
E.1. Población	119
E.2. Centros Educativos	121
E.3. Centros de Salud	122
E.4. Paradas de Buses	123
E.5. Edificios	124
E.6. Unidades de Policía Comunitaria	125
E.7. Áreas Verdes	126
F. Anexo 6. Mejor Modelo por Actividad Económica	127
F.1. Comercio. Modelo Robusto por Variables Instrumentales de Orden Superior	127
F.2. Científico Técnico. Modelo Robusto por Variables Instrumentales de Orden Superior	128
F.3. Administración. Modelo Robusto por Variables Instrumentales de Orden Superior	129
F.4. Manufactura. Modelo Robusto por Variables Instrumentales de Orden Superior	130
F.5. Transporte. Modelo Robusto por Variables Instrumentales de Orden Superior	131
F.6. Información. Modelo Robusto por Variables Instrumentales de Orden Superior	132
F.7. Construcción. Modelo Robusto por Variables Instrumentales de Orden Superior	132

Índice de figuras

2.1. Estructura de Datos Espaciales	10
2.2. Datos de Eventos	10
2.3. Datos de Línea	11
2.4. Datos de Polígono	11
2.5. Datos de Superficie	11
2.6. Barrios de Quito	17
2.7. Población de Quito	18
3.1. Tipos de Contigüidad	21
3.2. Gráfico de Dispersión de Moran	25
5.1. Distribución de las Empresas en los barrios de Quito	43
5.2. Resultados GeoDa. Autocorrelación en Total de Empresas	44
5.3. Resultados GeoDa. Gráfico de Dispersión de Moran	45
5.4. Clusters por Total de Empresas	45
5.5. Barrios con similares Características. Total de Empresas	47
5.6. Barrios por Actividades con alta densidad	48
5.7. Barrios Definidos como Altos	50
5.8. Barrios por Actividades con características similares	51
5.9. Distribución del residuo en el Espacio	53
5.10. Barrios por Actividades con alta densidad	61
A.1. Información Directorio de Empresas	73
A.2. Información de Población	74
A.3. Información de Centros Educativos	75

A.4. Información de Centros de Salud	76
A.5. Información de Paradas de Buses	77
A.6. Información de Edificios	78
A.7. Información de Unidades de Policia Comunitaria	79
A.8. Información de Áreas Verdes	80
C.1. Distribución de las Empresas de Comercio en los barrios de Quito . .	86
C.2. Resultados GeoDa. Autocorrelación Espacial de Comercio	87
C.3. Resultados GeoDa. Gráfico de Dispersión de Moran Comercio	87
C.4. Clusters por Comercio	88
C.5. Barrios con similares Características. Comercio	89
C.6. Distribución de las Empresas de Científico Técnico en los barrios de Quito	90
C.7. Resultados GeoDa. Autocorrelación Espacial de Científico Técnico . .	91
C.8. Resultados GeoDa. Gráfico de Dispersión de Moran Científico Técnico	91
C.9. Clusters por Científico Técnico	92
C.10.Barrios con similares Características. Científico Técnico	93
C.11.Distribución de las Empresas Administrativas en los barrios de Quito	94
C.12.Resultados GeoDa. Autocorrelación Espacial Administrativas	95
C.13.Resultados GeoDa. Gráfico de Dispersión de Moran Administrativas .	95
C.14.Clusters por Administrativas	96
C.15.Barrios con similares Características. Administrativas	97
C.16.Distribución de las Empresas Manufacturas en los barrios de Quito .	98
C.17.Resultados GeoDa. Autocorrelación Espacial Manufacturas	99
C.18.Resultados GeoDa. Gráfico de Dispersión de Moran Manufacturas . .	99
C.19.Clusters por Manufacturas	100
C.20.Barrios con similares Características. Manufacturas	101
C.21.Distribución de las Empresas de Transporte en los barrios de Quito .	102
C.22.Resultados GeoDa. Autocorrelación Espacial Transporte	103
C.23.Resultados GeoDa. Gráfico de Dispersión de Moran Transporte . . .	103
C.24.Clusters por Transporte	104
C.25.Barrios con similares Características. de Transporte	105
C.26.Distribución de las Empresas Información en los barrios de Quito . .	106

C.27.Resultados GeoDa. Autocorrelación Espacial Información	107
C.28.Resultados GeoDa. Gráfico de Dispersión de Moran Información . . .	107
C.29.Clusters por Información	108
C.30.Barrios con similares Características. Información	109
C.31.Distribución de las Empresas Construcción en los barrios de Quito . .	110
C.32.Resultados GeoDa. Autocorrelación Espacial Construcción	111
C.33.Resultados GeoDa. Gráfico de Dispersión de Moran Construcción . .	111
C.34.Clusters por Construcción	112
C.35.Barrios con similares Características. Construcción	113
E.1. Distribución de la Población	119
E.2. Autocorrelación Espacial de la Población	120
E.3. Distribución de los Centros Educativos	121
E.4. Autocorrelación Espacial de los Centros Educativos	121
E.5. Distribución de los Centros de Salud	122
E.6. Autocorrelación Espacial de los Centros de Salud	122
E.7. Distribución de las Paradas de Buses	123
E.8. Autocorrelación Espacial de las Paradas de Buses	123
E.9. Distribución de los Edificios	124
E.10.Autocorrelación Espacial de los Edificios	124
E.11.Distribución de las Unidades de Policía Comunitaria	125
E.12.Autocorrelación Espacial de las Unidades de Policía Comunitaria . . .	125
E.13.Distribución de las Áreas Verdes	126
E.14.Autocorrelación Espacial de las Áreas Verdes	126

Índice de tablas

2.1. Combinaciones de Variables de Locación	14
5.1. Distribución de las Empresas en Quito	42
5.2. Distribución de las Empresas en los barrios de Quito	42
5.3. Cantidad de Barrios por Cluster. Total de Empresas	44
5.4. Características del Cluster Alto-Alto. Total de Empresas	46
5.5. Barrios por Actividades con alta Densidad	47
5.6. Distribución de las Empresas con Alta Densidad	49
5.7. Distribución de las Empresas con Alta Densidad	49
5.8. Barrios por Actividad con características similares	50
5.9. Índices de Autocorrelación Espacial por Variable	52
5.10. Coeficientes de Factores Externos sobre el Total de empresas	62
5.11. Impacto de Factores Externos sobre el Total de empresas	62
5.12. Coeficiente del Intercepto por Actividad Económica	63
5.13. Impacto Directo de los Factores por Actividad Económica	63
5.14. Impacto Indirecto de los Factores por Actividad Económica	64
5.15. Impacto Total de los Factores por Actividad Económica	64
A.1. Información Directorio de Empresas	72
A.2. Información de Población	74
A.3. Información de Centros Educativos	75
A.4. Información de Centros de Salud	76
A.5. Información de Paradas de Buses	77
A.6. Información de Edificios	78
A.7. Información de Unidades de Policía Comunitaria	79

A.8. Información de Áreas Verdes	80
C.1. Cantidad de Barrios por Cluster. Comercio	87
C.2. Cantidad de Barrios por Cluster. Científico Técnico	91
C.3. Cantidad de Barrios por Cluster. Administrativas	95
C.4. Cantidad de Barrios por Cluster. Manufacturas	99
C.5. Cantidad de Barrios por Cluster. Transporte	103
C.6. Cantidad de Barrios por Cluster. Información	107
C.7. Cantidad de Barrios por Cluster. Construcción	111
D.1. Barrios de Quito con Alta Densidad de Empresas	117
F.1. Impacto de Factores Externos sobre las Empresas de Comercio	128
F.2. Impacto de Factores Externos sobre las Empresas Científico Técnicas	129
F.3. Impacto de Factores Externos sobre las Empresas de Administración	130
F.4. Impacto de Factores Externos sobre las Empresas de Manufactura . .	131
F.5. Impacto de Factores Externos sobre las Empresas de Transporte . . .	131
F.6. Impacto de Factores Externos sobre las Empresas de Información . .	132
F.7. Impacto de Factores Externos sobre las Empresas de Construcción . .	133

Resumen

El presente trabajo se centra en el uso de distintas técnicas de minería de datos, basadas en métodos estadísticos geospaciales para la identificación de patrones con respecto a las actividades económicas de las empresas, registradas en La Superintendencia de Compañías en el Distrito Metropolitano de Quito, al igual de definir factores externos que influyen en la constitución de estas.

Primero, para la creación de los clusters, se utilizan los indicadores locales de asociación espacial (LISA, Local Indicators of Spatial Association), los cuales definirán los barrios que posean alta densidad y se caracteriza a estos con variables auxiliares que describen la locación. Esto también ayuda a identificar potenciales barrios con similares características, pero sin una alta densidad de empresas. Después, se utilizan modelos de regresión espacial, para identificar la relación que existe entre las empresas y las variables auxiliares encontradas, analizando el coeficiente asociado a cada una.

Finalmente, al complementar ambos resultados se obtiene el listado de los barrios con alta densidad en los cuales se debería trabajar, con el factor de crecimiento o disminución de empresas de cada una de las variables auxiliares identificadas.

Palabras claves: Estadística Espacial, Indicadores Locales de Asociación Espacial, Modelos de Regresión Espacial.

Abstract

This work focuses on the use of different data mining techniques, based on geospatial statistical methods for the identification of patterns with respect to the economic activities of companies, registered in the Superintendencia de Compañías in the Metropolitan District of Quito, at as well as defining external factors that influence the constitution of these.

First, to create the clusters, the Local Indicators of Spatial Association (LISA) are used, which will define the neighborhoods that have high density and are characterized with auxiliary variables that describe the location. This also helps to identify potential neighborhoods with similar characteristics, but without a high density of businesses. Then, spatial regression models are used to identify the relationship that exists between the companies and the auxiliary variables found, analyzing the coefficient associated with each one.

Finally, by complementing both results, the list of high-density neighborhoods in which work should be done is obtained, with the factor of growth or decrease of companies for each of the auxiliary variables identified.

Keywords: Spatial Statistics, Local Indicators of Spatial Association, Spatial Regression Models.

Capítulo 1

Introducción

En este capítulo se establecen los antecedentes y metodología a utilizar dentro del trabajo. Primero, se define al ordenamiento territorial y la normativa establecida por el gobierno. Seguido, se presentan las metodologías alternas y su forma de aplicación para el estudio.

1.1. Antecedentes

La distribución en el espacio dentro del territorio nacional se ha vuelto un tema relevante para el estado ecuatoriano en los últimos años. Por esto, se puso en marcha la Ley Orgánica de Ordenamiento Territorial, Uso y Gestión de Suelo, que tiene como objetivo principal:

Fijar los principios y reglas generales que rigen el ejercicio de las competencias de ordenamiento territorial, uso y gestión del suelo urbano y rural, y su relación con otras que incidan significativamente sobre el territorio o lo ocupen, para que se articulen eficazmente, promuevan el desarrollo equitativo y equilibrado del territorio y propicien el ejercicio del derecho a la ciudad, al hábitat seguro y saludable, y a la vivienda adecuada y digna, en cumplimien-

to de la función social y ambiental de la propiedad e impulsando un desarrollo urbano inclusivo e integrador para el Buen Vivir de las personas, en concordancia con las competencias de los diferentes niveles de gobierno.(Asamblea Nacional, 2016)

Dentro de esta Ley se define como Ordenamiento Territorial al “proceso y resultado de organizar espacial y funcionalmente las actividades y recursos en el territorio, para viabilizar la aplicación y concreción de políticas públicas democráticas y participativas y facilitar el logro de los objetivos de desarrollo” (Asamblea Nacional, 2016), cuya planificación se encuentra a cargo de los Gobiernos Autónomos Descentralizados (GAD).

En el Distrito Metropolitano de Quito (DMQ), las secretarías de Territorio, Hábitad y Vivienda; y de Planificación diseñaron el Plan Metropolitano de Desarrollo y Ordenamiento Territorial. Donde se definen las actividades que logren un mejor estilo de vida a la población de la ciudad de Quito (Alcaldía de Quito, 2015).

El plan abarca varios ejes que van desde lo social hasta lo productivo con metas a cumplirse dentro de los próximos años. Uno de los ejes de la planificación se enfoca en el desarrollo económico y productivo, donde, las empresas y su distribución en el DMQ son un factor crucial (Alcaldía de Quito, 2015).

Una empresa se la puede definir como “la combinación más pequeña de unidades legales que constituye una unidad organizativa de producción de bienes y servicios y que disfruta de una cierta autonomía de decisión, principalmente a la hora de emplear los recursos corrientes de que dispone. La empresa ejerce una o más actividades en uno o varios lugares.” (CAN, 2010).

Dado que existe una gran variedad de empresas y establecimientos, es necesario poder clasificarlos de forma correcta, por ejemplo, de acuerdo a la actividad económica que realizan usando la Clasificación Industrial Internacional Uniforme (CIIU). Las CIIU constituyen una estructura de clasificación coherente y consistente de todas las actividades económicas que realizan las empresas, emitida por la Organización de Naciones Unidas (ONU), basada en un conjunto de conceptos, principios y normas

de clasificación (ONU, 2005).

Las CIIU cataloga las actividades económicas en varios niveles, partiendo de una clasificación general a una más detallada. Este trabajo se centra en la clasificación general de las CIIU, que se presenta a continuación:

A. Agricultura	G. Comercio	N. Administrativos
B. Minas y canteras	H. Transporte	O. Administración pública
C. Manufacturas	I. Alojamiento	P. Enseñanza
D. Suministro de energías	J. Información	Q. Salud humana
E. Distribución de agua	K. Financieras	R. Arte
F. Construcción	L. Inmobiliaria	S. Otros servicios
	M. Científico Técnico	

Otra forma de clasificarlo es considerando el tamaño de las empresas, en donde se toma en cuenta primordialmente la cantidad de empleados seguido del volumen de activos que posee. Los rangos usualmente utilizados en el país son: empresas pequeñas con menos de 50 trabajadores con activos hasta un millón de dólares; empresas medianas de 50 a menos de 200 trabajadores y activos hasta los cinco millones de dólares; y finalmente empresas grandes con más de 200 trabajadores y cinco millones de dólares en activos (INEC, 2018). Estos valores pueden cambiar o tomar en cuenta otras variables como volumen de ventas, lo cual es frecuentemente usado en la Unión Europea (UE) (Martínez Argudo, 2018) .

Dada la gran variedad de actividades económicas a las que se dedican las empresas y por el impacto que pueden llegar a tener dentro del ámbito productivo y económico del ordenamiento territorial de Quito, son un punto importante dentro de la planificación (Alcaldía de Quito, 2015).

En esta planificación, se realiza el diagnóstico de la situación de las empresas utilizando la información del censo económico, con un análisis descriptivo inicial de las empresas por tamaño y actividad económica dentro de cada administración zonal, con la finalidad de detectar aglomeración industrial a partir del cálculo de coeficientes de localización (Alcaldía de Quito, 2015).

Para generar este coeficiente se utiliza la cantidad de empleados dentro de una actividad y el total de empleados sobre una región de todas las actividades. La metodología empleada en la planificación es limitada, ya que no considera métodos más apropiados para la caracterización de estas aglomeraciones (cluster).

En este trabajo, se propone el uso de indicadores provenientes de la estadística espacial, los cuales ayudan a identificar centros de aglomeraciones utilizando su posición y no únicamente un espectro amplio como la administración zonal. Este indicador se ha utilizado en España para medir densidad de migración (Martori Cañas y Hoberg, 2007) y en Ecuador en el estudio de Rojas, 2015, donde se obtienen la densidad de población y empleo.

Dentro del objetivo de la planificación, se debe tomar en cuenta otros factores a parte del económico, como: el ambiental, social y de movilidad; y la relación generada entre ellos, debido a que coexisten en un espacio fijo con limitaciones. La alcaldía realiza el diagnóstico sobre estos ejes o factores de manera individual, y con los resultados busca obtener correlaciones entre estos (Alcaldía de Quito, 2015)..

En este trabajo, además se propone el uso de modelos de regresión espacial y el análisis sobre sus coeficientes para encontrar las distintas interacciones de los factores y la densidad de las empresas. Estos modelos se han utilizado en Argentina para describir factores que afectan a la asistencia escolar tomando en cuenta el departamento en el que residen (Cid, 2011).

1.2. Objetivos

Identificar las distintas tipologías y determinar la influencia de factores externos sobre las empresas del Distrito Metropolitano de Quito para mejorar el análisis de estas en el ordenamiento territorial de la ciudad.

Para alcanzarlo es necesario lograr los siguientes objetivos específicos:

- Realizar el análisis de datos espaciales inicial sobre las variables encontradas.
- Identificar las zonas que poseen una gran cantidad de empresas a través de estadística espacial.
- Ajustar el mejor modelo de regresión espacial sobre las empresas y los factores externos.

1.3. Justificación

La Ley Orgánica de Ordenamiento Territorial, Uso y Gestión del Suelo (2016) define a los GADs como los ejecutores del diagnóstico del espacio territorial, en este trabajo se proponen nuevas técnicas para realizar este análisis, utilizando como eje a las empresas.

El factor importante para el desarrollo de estas técnicas es la información espacial, es decir, información que permita su geolocalización¹. En Quito, estos datos son administrados por las distintas secretarías, los cuales deben ser proporcionados al GAD.

¹El término geolocalización no está incluido en el diccionario de la Real Academia Española (RAE). Sin embargo, este concepto se utilizan con frecuencia en nuestra lengua al hacer mención a la ubicación de algo o alguien en un sistema de coordenadas geográficas. (Pérez Porto y Merino, 2018a)

La aplicación de esta metodología contribuirá al análisis descriptivo desarrollado por el ente gubernamental, donde no solamente se aplicaría sobre las empresas, si no que pudiese extenderse sobre los demás ejes, siempre y cuando exista información disponible.

La metodología utilizada se fractura en dos partes, la primera en la creación de clusters a través de indicadores espaciales y la segunda, sobre los modelos de regresión espacial.

Para la **creación de estos clusters** es necesario entender la estadística espacial, la cual es una rama de la estadística general que utiliza los análisis descriptivos en conjunto con datos geográficos. Uno de los primeros estudios es de John Snow en el año 1854, donde se estudió los brotes de cólera en Londres.

El concepto básico en el que se basa la estadística espacial es el de la dependencia o autocorrelación espacial, que analiza los efectos en el cambio de una variable según su localización (Celemín, 2011). Los primeros índices formales para detectar la presencia de autocorrelación espacial se deben a Moran (1948) y Geary (1954). Este concepto se empezó a utilizar en distintas áreas, como ciencias sociales, econometría, geografía y medicina.

Otro punto a tratar es el relacionado con los clusters o patrones espaciales, que son un conjunto o zona en la cual se albergan objetos que poseen cierta característica en común. Al realizar un análisis de autocorrelación espacial se puede descubrir si una variable proviene de una distribución aleatoria, o si existe una asociación de valores similares en una locación. De este modo, si los valores altos o bajos tienden a agruparse indica una autocorrelación espacial positiva, por otro lado si los valores obtenidos son distintos entonces la autocorrelación es negativa. Si no se presenta ninguno de los dos casos no existe autocorrelación espacial (Luc, 1995).

Si se encuentra autocorrelación espacial se realiza el análisis de forma local, donde se calcula un indicador para cada zona, tomando en cuenta la información cercana y su alrededor. De esta manera el mapa se divide en diferentes zonas, aquellas que poseen autocorrelación espacial y otras que no. El indicador de autocorrelación que

se utilizará será el Indicador Local de Moran, que es un indicador local de asociación espacial, a los cuales se les conoce como Local Indicators of Spatial Association por sus siglas en inglés LISA (Luc, 1995).

Al utilizar un LISA se identifica los sectores por la densidad que poseen y además si pertenece a un nivel alto o bajo, dependiendo de la variable en cuestión. Así, si se toman todas aquellas que están en un nivel alto o bajo, se encuentran patrones espaciales o clusters (Luc, 1995).

Un documento referente a estos indicadores se encuentra dentro del estudio de la población migrante de España por Martori Cañas y Hoberg, 2007. En donde al tener la información de la vivienda de los migrantes y conociendo su país de procedencia, se logra clasificar los distintos sectores de Barcelona como alto o bajo en densidad. Dentro del ámbito de las empresas y en el contexto nacional, Rojas, 2015, en la revista Analitika, define posibles centros de empleo en el DMQ, el cual parte de los resultados de clusters de empresas y población utilizando LISA, para establecer posibles lugares en los cuales pueden ubicarse centros de población en relación con las distintas densidades de las variables utilizadas.

Otra metodología a usar son los modelos de regresión, los cuales son utilizados para estudiar la relación entre distintas variables. La regresión es el conjunto de técnicas usadas para explorar y cuantificar la relación de dependencia entre una variable cuantitativa llamada dependiente o respuesta y una o más variables independientes llamadas predictoras. El tipo de regresión más conocido es el lineal, ya sea simple o múltiple. Este tipo de regresión posee la siguiente forma funcional:

$$Y = \sum_{i=1}^N X_i \beta_i + \epsilon$$

donde: Y son las observaciones de la variable dependiente; X_i las observaciones de la variable independiente i tomando valores desde 1 hasta N posibles variables, β_i es el coeficiente de la regresión que mide la influencia de la variable i sobre la variable dependiente Y ; y finalmente ϵ es el error que se puede dar por variables no controladas.

Este tipo de regresión es posible aplicarla sobre datos espaciales, pero, si se demuestra que existe una autocorrelación espacial sobre las variables utilizadas en la regresión, se obtienen resultados atípicos, donde este comportamiento se presentaría dentro del término del error. Esto se puede identificar aplicando uno de los índices descritos anteriormente sobre el error, encontrando si este también tiene autocorrelación espacial. Con lo cual se puede aplicar modelos de regresión espacial.

Estos modelos parten de la dependencia espacial que puede encontrarse en las variables independientes, dependiente o en el error. Cuando la dependencia espacial se encuentra en la variable dependiente se denominan modelos de retardo espacial, por otro lado, si esta dependencia se presenta en los residuos son llamados modelos de error espacial (Luc, 2001).

También, existen modelos de orden superior que combinan ambos aspectos dentro de la dependencia espacial. Otro modelo es el de Durbin, el cual se utiliza cuando la dependencia espacial se encuentra dentro de la variable dependiente, ocasionada por la influencia de las variables independientes (Luc, 2001).

Con cualquiera de estos modelos se puede encontrar la influencia de cada variable independiente, al igual que la influencia que posee el espacio sobre la variable dependiente (Luc, 2001).

Los modelos de regresión espacial, encuentran su utilidad para definir la asistencia escolar. Cid, 2011, utiliza información del Censo Nacional de Argentina, con datos georeferenciados a nivel de cada departamento, relaciona el valor de la asistencia escolar con variables de los hogares como: nivel educativo del jefe de hogar, nivel de desocupación, cantidad de jóvenes, etc. Para encontrar la influencia de estas variables, se realiza varios modelos de regresión incluyendo modelos espaciales. Permitiendo futuras mejoras al poseer datos más precisos.

Capítulo 2

Análisis Exploratorio de Datos Espaciales

Este capítulo se centra en los datos espaciales, sus características y la información que se utiliza. En la primera sección se describen los datos espaciales, al igual que las formas más comunes de encontrarlos. Después se detalla la información disponible y su procedencia. Seguido se presenta el tratamiento que se le da a los datos previo a su trabajo. Por último, se muestra la forma más fácil e ilustrativa para entender estos datos.

2.1. Datos Espaciales

Los datos espaciales, a diferencia de los datos normales, poseen una estructura más compleja. Los componentes de esta estructura son dos: un elemento tabular o descriptivo en el que se representan los valores como cualquier base de datos antes conocida; y un elemento geométrico el cual almacena la información de locación con respecto a las variables dentro del elemento descriptivo (Benito Moreno, 2017).

Los datos espaciales son de tres tipos: de evento, objeto y superficie. Estos se

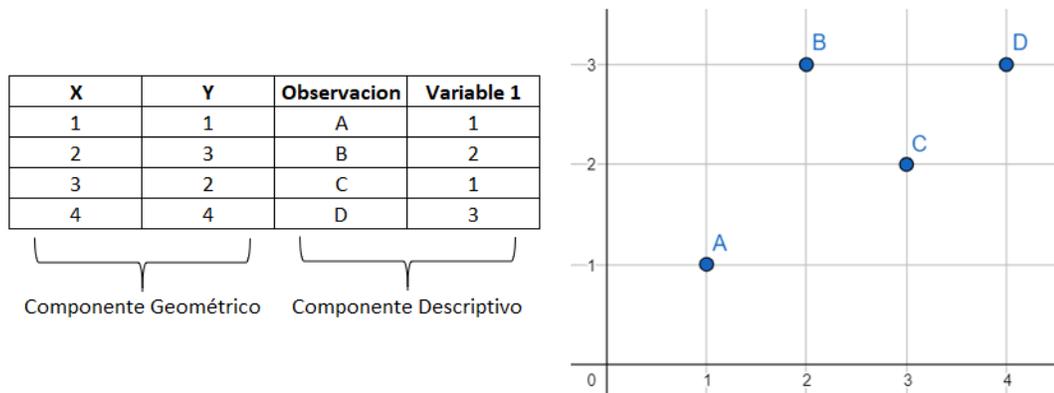


Figura 2.1: Estructura de Datos Espaciales

diferencian por la estructura asociada a la geometría dentro de cada objeto. A continuación, se describirá cada uno de ellos.

Los datos de eventos son aquellos representados por coordenadas geográficas de un punto de locación. Se utilizan para identificar la ubicación de edificaciones como tiendas, hoteles, restaurantes, etc. O eventos como accidentes, robos, etc.

Datos de Evento

X	Y	Observacion	Variable 1
2	1	A	3

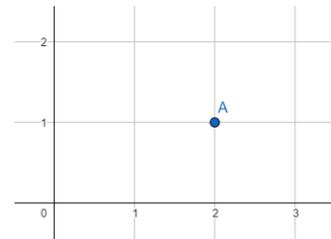


Figura 2.2: Datos de Eventos

Los datos de objeto son representaciones geométricas que a comparación de la anterior poseen más dimensiones, clasificándolos como datos de línea o polígono. Los datos de línea son aquellos constituídos por dos puntos y el trazo que los une, comúnmente usados para describir carreteras y ríos. Mientras que los datos de polígono tienen varios puntos como vértices para estructurarlos, empleados dentro de las divisiones administrativas, es decir, provincias, cantones, parroquias, etc.

Por último, los datos de superficie resultan al conectar información a unidades de

Datos de Línea

X1	Y1	X2	Y2	Observacion	Variable 1
1	2	2	1	A	3

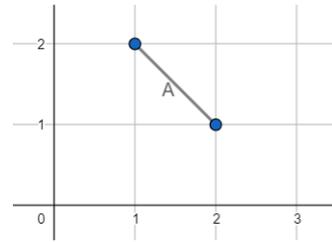


Figura 2.3: Datos de Línea

Datos de Polígono

X Centroide	Y Centroide	Área	Observacion	Variable 1
2	1.5	2	A	3

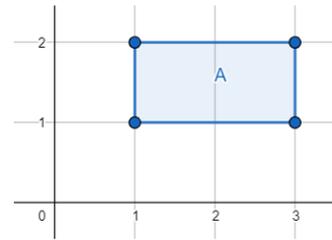


Figura 2.4: Datos de Polígono

igual área que se encuentran distribuidas dentro de un límite específico, es decir, se define los datos sobre una malla en el espacio donde cada cuadrante posee su propia información. Estos datos se utilizan para definir fenómenos de forma continua, por ejemplo, el clima.

Datos de Superficie

Observacion	Variable1
A	3
B	1
C	4
D	2
E	3
F	4

paso malla: 0.5

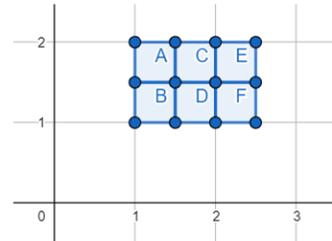


Figura 2.5: Datos de Superficie

2.2. Información a Utilizar y Procedencia

La información con la que se trabaja es proveniente de distintas entidades gubernamentales. A continuación se describe cada una de estas.

La **Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)** es el organismo técnico, con autonomía administrativa y económica, que vigila y controla la organización, actividades, funcionamiento, disolución y liquidación de las compañías y otras entidades en las circunstancias y condiciones establecidas por la Ley. Esta institución lleva un directorio actualizado de forma anual de todas las compañías del país. Hay que recalcar que no todas las empresas deben reportar su información a la superintendencia, mas bien solo aquellas que poseen una constitución más estable como sociedad anónima, laboral, etc. Se debe tomar en cuenta que los datos que proporcionan no son del índole espacial, por ende necesitan un tratamiento y una transformación previa para llegar a obtener la información deseada.

El **Sistema Nacional de Información (SNI)**, es el conjunto de varios actores de los cuales se accede, recoge, almacena y transforma sus datos en información accesible para entes gubernamentales y público en general. Este portal almacena información espacial de varios ministerios del país. Los datos que se usan se refieren a la locación de los centros educativos y centros de salud del país con corte al año 2014.

La página de **Gobierno Abierto de Quito** es un portal que sigue la iniciativa global en la cual, utilizando la tecnología, se busca transparentar la gestión de los entes gubernamentales conectando toda la información que se trabaje con el público y también recibir información de estos. En esta web se encuentra información de áreas verdes, paradas de buses y Unidades de Policía Comunitaria (UPC) en el DMQ con corte de información del año 2014 (Gobierno Abierto, 2020).

El **Instituto Nacional de Estadísticas y Censos (INEC)**, es la entidad encargada de planificar, normar y certificar la producción del Sistema Estadístico Nacional. Por ser la institución encargada de las estadísticas del país, esta posee gran cantidad de información la cual también es del índole espacial. Los datos encontrados son de

población del año 2010 y de edificios del año 2014.

Toda la información utilizada se describe una por una en la sección de los anexos (Anexo 1).

2.3. Tratamiento de Datos

Para el presente estudio, los datos se tratan, analizan, reestructuran y homologan con el fin de que se puedan trabajar de forma conjunta. Para poder trabajar con los datos se tuvo que tratarlos y transformarlos para que sean útiles dentro de este estudio. Se debe recalcar que no existe un procedimiento fijo y cada variable necesita un tratamiento específico para obtener el mejor resultado.

2.3.1. Georeferenciación de Datos

Inicialmente se desarrolla un procedimiento para transformar una base plana a una base de datos espacial. Esto se realiza debido a que el directorio de las empresas publicado por la SUPERCIAS contiene una base en la cual la locación está asociada a la dirección como una variable del tipo carácter.

Para transformarlo, se utiliza el software **R** y la librería “*ggmap*”. La transformación consiste en usar las variables del tipo carácter que definen su locación, como el barrio, edificio, calle principal, secundaria y el nombre de la empresa. Con la función *geocode* de la librería mencionada anteriormente se georeferencia su ubicación.

El concepto de georeferenciación está asociado al posicionamiento espacial de un cuerpo sobre un sistema de coordenadas. El sistema más común utilizado para la georeferenciación es a través de la longitud y latitud, es decir, obtener dos valores X Y los cuales ayudan a definir un punto dentro del globo terráqueo (Pérez Porto y Merino, 2018b).

La función geocode utiliza la API² de Google para poder conectarse a su servidor. Se realiza una búsqueda ingresando un texto específico el cual devuelve la longitud y latitud del resultado más parecido dentro de su mapa. Debido a que se posee más de una variable de texto para definir la locación se utiliza varias combinaciones de estos valores para obtener un buen resultado. Las combinaciones se estructuran como se muestra en la Tabla 2.1.

Orden de Combinación	Variable 1	Variable 2	Variable 3
1	Calle Principal	Intersección	Ciudad
2	Barrio	Calle Principal	Ciudad
3	Nombre de Compañía	Barrio	Ciudad
4	Edificio	Ciudad	

Tabla 2.1: Combinaciones de Variables de Locación

El procedimiento se basa en utilizar las distintas combinaciones en el orden dado y validarlo con los límites de la ciudad, obteniendo de resultado final el primer valor dentro de los límites fijados. A continuación se presenta el pseudocódigo del procedimiento.

```

1: n = número de Compañías
2: X = límites de longitud de Quito
3: Y = límites de latitud de Quito
4: for i = 1:n do
5:   resultado = geocode(Combinacion1(i))
6:   if resultado in X Y then
7:     resultadofinal[i] = resultado
8:   else
9:     resultado = geocode(Combinacion2(i))
10:    if resultado in X Y then
11:      resultadofinal[i] = resultado
12:    else
13:      resultado = geocode(Combinacion3(i))

```

²Una API (siglas de ‘Application Programming Interface’) es un conjunto de reglas (código) y especificaciones que las aplicaciones pueden seguir para comunicarse entre ellas: sirviendo de interfaz entre programas diferentes de la misma manera en que la interfaz de usuario facilita la interacción humano-software. (Merino, 2014)

```
14:         if resultado in X Y then
15:             resultadofinal[i] = resultado
16:         else
17:             resultado = geocode(Combinacion4(i))
18:             if resultado in X Y then
19:                 resultadofinal[i] = resultado
20:             else
21:                 resultadofinal[i] = Sin resultado
22:             end if
23:         end if
24:     end if
25: end if
26: end for
```

Debido a que se utilizan los nombres de las calles, no se obtiene la ubicación exacta de la compañía, pero si la esquina más cercana, dejando el error de locación en aproximadamente una manzana de distancia. Esto se puede dejar pasar debido a que las limitaciones de los barrios no cortan manzanas y se utiliza a los barrios como unidad de estudio.

2.3.2. Igualdad de Sistemas de Referencia

Se debe definir dos conceptos asociados a la cartografía que son las proyecciones cartográficas y los Sistemas de Referencia de Coordenandas (SRC).

- Las **proyecciones cartográficas** son el conjunto de técnicas utilizadas por los cartógrafos para mostrar la información de la superficie esférica del planeta sobre un plano de dos dimensiones. Las proyecciones cartográficas buscan tener gran precisión, para lo cual el tamaño y el propósito del mapa definen el tipo de proyección a utilizar (QGIS).
- Los **SRC** definen con coordenadas a cualquier punto de la tierra mediante tres números. En general, los SRC se pueden dividir en sistemas de referencia de

coordenadas proyectados (también denominados Cartesianos o rectangulares) y sistemas de referencia de coordenadas geográficas. (QGIS).

Estos conceptos son necesarios debido a que dentro de las bases de datos espaciales ya tienen asociada una proyección cartográfica, un SRC o ambos; esto es para definir con precisión la ubicación de cada uno de los datos. Como existen varios tipos de proyecciones o SRC, se debe transformar toda la información a un solo sistema o proyección. Esto de igual manera se realiza con el software **R** y la librería “*sp*”, utilizando la función *sptransform* para realizar la transformación deseada.

2.4. Visualización de Datos

Esta información puede ser visualizada o graficada en distintos mapas, debido a que ya se posee la ubicación de cada uno de los datos. A pesar de que estos tienen dicha característica en común, la forma en que estos se pueden mostrar varía dependiendo del tipo de dato espacial que se tenga.

En este trabajo se recopila toda la información dentro de los barrios del DMQ, por lo cual, resultando datos del tipo polígono. Sobre los barrios se ajusta toda la data y se la representa a través de mapas de calor, siendo estos los más utilizados.

Los valores que están asociados a cada barrio pueden tener un amplio rango, donde resulta más práctico juntarlos en intervalos. Lo comunmente usado son los percentiles o intervalos de igual tamaño, dependiendo del tipo que se escoja, puede esconder información o dar un mejor entendimiento.

En la Figura 2.6 se muestra la distribución de los barrios dentro del área urbana del DMQ. Seguido, en la Figura 2.7, se presenta la población en cada barrio utilizando intervalos de igual tamaño. Esto para poder reflejar de manera visual la distribución de la población dentro del DMQ.

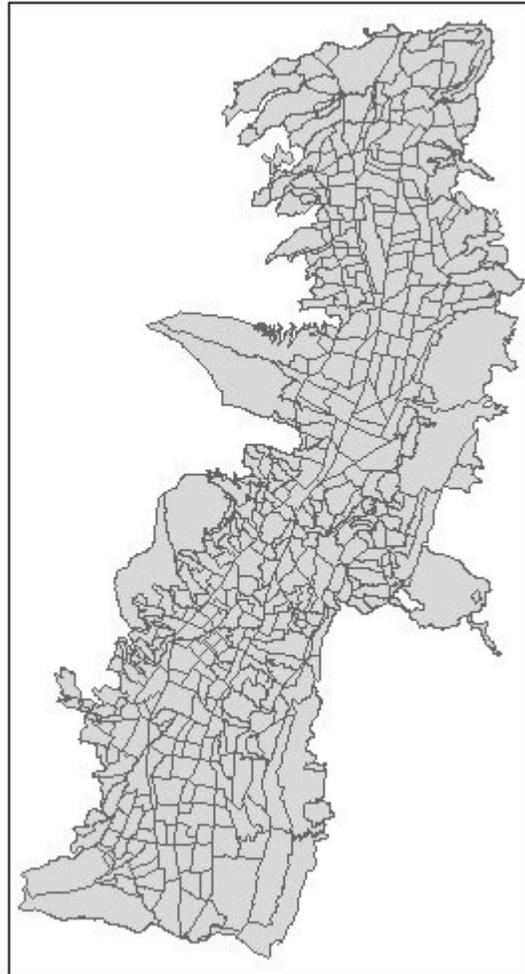


Figura 2.6: Barrios de Quito

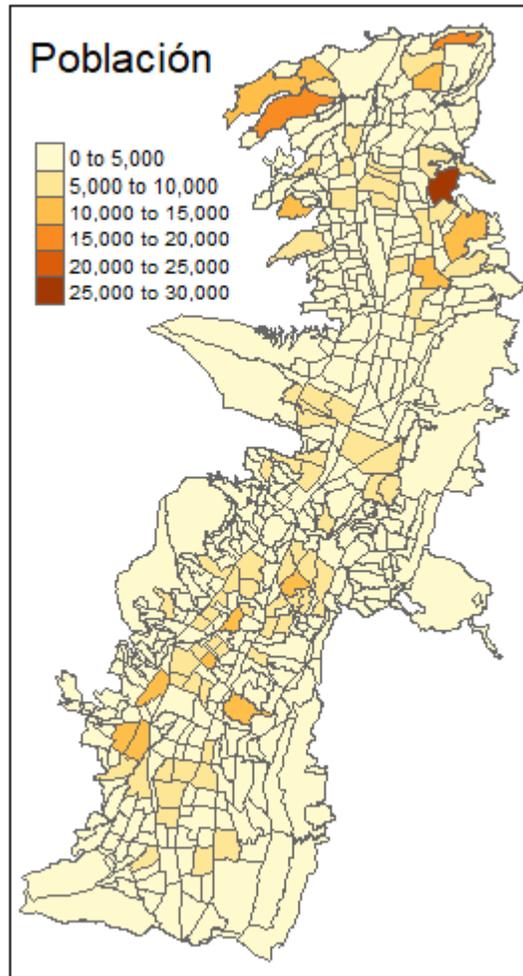


Figura 2.7: Población de Quito

Capítulo 3

Índices De Autocorrelación Espacial

Los índices de autocorrelación se dividen entre globales y locales. El global corresponde a calcular el índice utilizando todas las áreas de estudio, mientras que el local evalúa puntualmente un área con sus vecinos. A continuación, se obtiene la forma de estimarlos partiendo del concepto de autocorrelación espacial.

3.1. Autocorrelación espacial

De forma general, la autocorrelación espacial se refiere a la relación entre los objetos, actividades o variables de estudio en alguna locación con otros objetos ubicados cerca, buscando una diferencia o similitud. Este concepto se obtiene de [Goodchild, 1986](#), en donde también se menciona que son técnicas las cuales relacionan la información de la locación con el atributo en cuestión, como ya se menciona en el capítulo anterior.

Existen varios métodos para poder estimar a la autocorrelación espacial. Se utiliza el índice de Moran para definir el valor de autocorrelación sobre un conjunto de datos

espaciales. Primero se define una forma de medir la cercanía dentro de las distintas locaciones, para lo cual se utilizan matrices de contigüidad y de distancias.

3.1.1. Matriz de Contigüidad y Matriz de Distancias

Las dos matrices parten del mismo concepto, en las cuales al tener n locaciones se construye una matriz de tamaño $n \times n$ en donde cada entrada representa ya sea si existe contigüidad o la distancia entre las locaciones i, j , con i, j entre 1 y n , dependiendo del tipo de matriz. Por la forma de construcción, estas matrices poseen las siguientes propiedades:

- La diagonal principal contiene ceros, esto implica que un área no es vecina consigo misma.
- La matriz es simétrica.
- El número de vecinos de cada área está determinado por los valores distintos de ceros de cada fila de la matriz.

Las matrices de contigüidad tienen distintos tipos, dependiendo de la cantidad de conexiones que posean. Así, si se toma en cuenta únicamente a los contiguos como aquellos que comparte borde, se le nombra contigüidad tipo Torre. En cambio, si se toma solo aquellos que se conectan únicamente por el vértice, se le nombra contigüidad tipo Alfil. Y finalmente si toman todo tipo de conexión, se nombran tipo Reina. Todos estos nombres provienen del ajedrez y de los movimientos de cada pieza. (Celemín, 2011)

Ahora, tomando en cuenta las matrices de distancia, comúnmente se define una distancia máxima en los cuales se tomará la información. Esto para lograr enfocar de forma correcta los vecinos y que no se obtenga que todos son vecinos en un mapa. Una variante a esta matriz es utilizar la distancia inversa dentro de cada entrada, para definir mejor el efecto que se quiera tratar en el espacio, siendo más fuerte

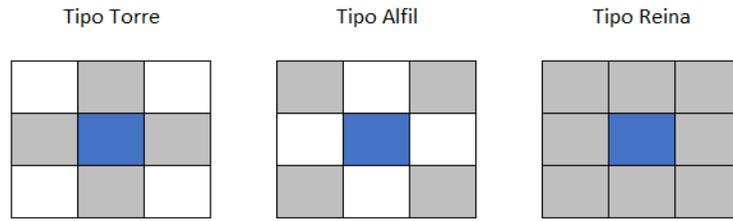


Figura 3.1: Tipos de Contigüidad

mientras más lejos se encuentre (matriz de distancias) o más fuerte mientras más cerca se encuentre (matriz de distancias inversa).

A partir de aquí, a estas se les denomina Matrices de Vecinos.

3.2. Indicador Global

3.2.1. Indicador de Moran

El indicador o índice de Moran es una medida proveniente de los años 1950 (Celemín, 2011), la cual a pesar de ser antigua se sigue manteniendo como una de las más utilizadas para el cálculo de autocorrelación espacial

La medida I de Moran viene dada por la siguiente expresión:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i - \bar{x})(z_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3.1)$$

donde: n es el número de áreas que se consideran; W_{ij} es una matriz de vecinos como cualquiera de las definidas anteriormente; los valores dados por $(x_i - \bar{x})$, $(x_j - \bar{x})$ siendo x_i y x_j el valor de la variable X del área i y j respectivamente, y siendo \bar{x} la media aritmética de todos los valores de las áreas. Una interpretación más sencilla a la ecuación es la relación entre la covarianza ponderada sobre la varianza de los valores.

El dominio del índice es $[-1, +1]$ en donde se considera que si: el indicador es cero, entonces no posee autocorrelación espacial, si el indicador es menor que cero posee autocorrelación negativa y si es mayor que cero posee autocorrelación positiva (Celemín, 2011). Estos conceptos de autocorrelación son los siguientes:

- Autocorrelación espacial positiva: los valores de las áreas vecinas son similares. Indica agrupamiento de las unidades espaciales.
- Autocorrelación espacial negativa: los valores de las áreas vecinas son distintos. Indica una tendencia a la dispersión de las unidades espaciales.
- Sin autocorrelación: no ocurre ninguna de las dos situaciones anteriores. Por lo tanto, los valores de las unidades espaciales vecinas presentan valores producidos de forma aleatoria.

El concepto para definir si posee o no autocorrelación se puede ampliar fijando un intervalo cercano al cero, en el cual, se define la no existencia de autocorrelación. El problema al definir esto es la gran cantidad de datos distintos en los que se pueda aplicar, dejando únicamente un intervalo fijo como algo insuficiente para validar el indicador. Por este motivo se trabajan sobre valores p para validar si es estadísticamente distinto de cero.

3.2.2. Prueba de Hipótesis y Valor-p

El proceso a utilizar se lo denomina como Aleatorización y Permutación (Celemín, 2011), que consiste en generar varias muestras donde los datos permutan entre las distintas locaciones, obteniendo el indicador de cada muestra y compararlo con el valor original, con el objetivo de que en cada área pueda ocurrir cualquier valor, dejando que el espacio no tenga influencia sobre este. Con esto la hipótesis nula y alternativa son las siguientes:

H_0 : autocorrelación espacial nula

H_1 : autocorrelación espacial no nula

Sobre la hipótesis nula, se obtiene que la esperanza del índice de Moran I es:

$$E(I) = \frac{-1}{(n-1)} \quad (3.2)$$

siendo n el número de áreas que se consideran. Y su varianza es:

$$\begin{aligned} Var(I) = & \frac{(n[(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2])}{(n-1)(n-2)(n-3)S_0^2} \\ & - \frac{k[(n^2 - n)S_1 - 2nS_2 + 6S_0^2]}{(n-1)(n-2)(n-3)S_0^2} - 2E(I) \end{aligned} \quad (3.3)$$

siendo:

$S_0 = \sum_{i=1}^n \sum_{j=1}^n W_{ij}$, la suma de todas las entradas de la matriz de vecinos.

$S_1 = \sum_{i=1}^n \sum_{j=1}^n (W_{ij} + W_{ji})^2$, donde, si la matriz es simétrica se tiene $S_1 = 2S_0$.

$S_2 = \sum_{i=1}^n (W_{i.} + W_{.i})^2$, 2 veces la suma de la i -ésima columna y la i -ésima fila.

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2}$$

El pséudo valor p desarrollado sobre este procedimiento, depende de, si el valor original pertenece a la muestra obtenida, así el cálculo es el siguiente:

$$p = (R + 1)/(M + 1) \quad (3.4)$$

donde R es la cantidad de muestras en las cuales su índice es mayor o igual al índice original y M el total de muestras. Si p es menor al criterio de rechazo, entonces no se puede aceptar la hipótesis nula (Luc, 2003).

3.3. Indicador Local

Los indicadores locales de asociación espacial (LISA), responden a la idea de trabajar en un área específica. Estos indicadores explican los grupos o clusters de valores similares alrededor del área. Las sumas de los indicadores locales de todas las áreas deben describir el indicador global (Luc, 1995). A continuación, se describe al índice local de Moran.

3.3.1. Indicador de Moran

El indicador local de Moran se obtiene al fragmentar el indicador global a una única área.

$$I_i = \frac{(x_i - \bar{x})}{m_2} \sum_{j=1}^n W_{ij}(x_j - \bar{x}) \quad (3.5)$$

donde m_2 es la varianza.

Este indicador posee las mismas cualidades que el indicador global, por ende, el concepto de autocorrelación y el proceso para el valor p se mantiene.

Dado que se puede obtener un índice sobre cada área, la información resultante es extensa, logrando tener más tipos de análisis, así se presenta una nueva herramienta. La figura 3.2 representa el gráfico de dispersión de Moran, este se divide en cuatro cuadrantes comenzando por el primero en la parte superior derecha y siguiendo en sentido de las agujas del reloj. En el eje de las x aparecen los valores estandarizados³ de la variable para cada área, y en el eje y los valores estandarizados del promedio de las unidades vecinas de la misma (Celemín, 2011).

Así, en el primer cuadrante se definen las áreas que poseen valores superiores a la media (eje x) y valores superiores a la media en las áreas vecinas (eje y), estos se denominan alto-alto o hotspots. La situación inversa se encuentra en el tercer cuadrante denominándose bajo-bajo o coldspots. En los cuadrantes restantes, se

³Se estandariza de la siguiente manera: a cada valor se le resta la media y se le divide para su varianza

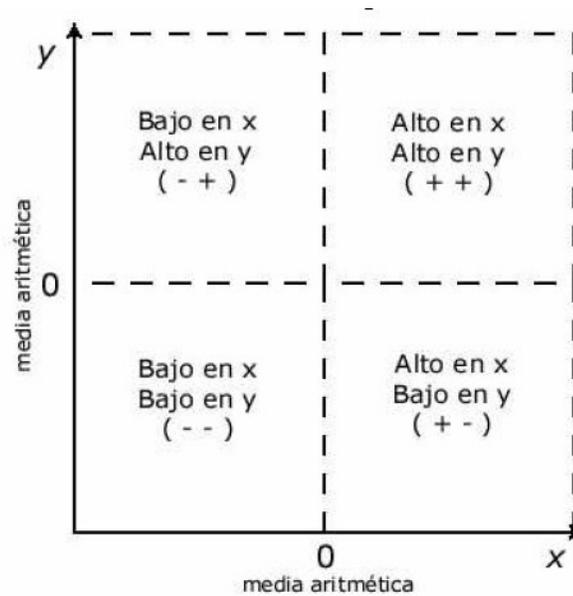


Figura 3.2: Gráfico de Dispersión de Moran

encuentran las áreas donde su valor propio es distinto al de sus vecinos, ubicando aquí outliers y denominándolos ya sea alto-bajo en el segundo cuadrante y bajo-alto en el cuarto cuadrante (Celemín, 2011).

Estos valores, utilizando aquellos que son significativos de acuerdo al valor p del indicador local, describen las áreas que pertenecen a los clusters alto y bajo, al igual de aquellas no significativas y atípicas.

3.4. GeoDa

Todo el análisis sobre los indicadores de asociación global y local se los realiza a travez de Geoda, que es una herramienta de software gratuita y de código abierto que sirve como introducción al análisis de datos espaciales. Está diseñado para facilitar nuevos conocimientos del análisis de datos mediante la exploración y el modelado de patrones espaciales (Luc, 2003).

GeoDa fue desarrollado por el Dr. Luc Anselin y su equipo. El programa propor-

ciona una interfaz gráfica y fácil de usar para métodos de análisis exploratorio de datos espaciales, como estadísticas de autocorrelación espacial para datos agregados (varios miles de registros) y análisis de regresión espacial básico para datos puntuales y poligonales (decenas de miles de registros) (Luc, 2003).

Capítulo 4

Modelado de Datos

En este capítulo se describen los modelos de regresión aplicables sobre datos espaciales. Estos se generan a partir del modelo de regresión lineal, para aplicar las respectivas correcciones y la forma de estimación de los parámetros.

4.1. Modelo de regresión lineal

Para este capítulo se parte del modelo de regresión lineal, la forma funcional del modelo es:

$$Y = \sum_{i=1}^N X_i \beta_i + \epsilon \quad (4.1)$$

donde: Y son las observaciones de la variable dependiente, X_i las observaciones de la variable independiente i tomando valores desde 1 hasta N posibles variables, β_i es el coeficiente de la regresión que mide la influencia de la variable i sobre la variable dependiente Y , y finalmente ϵ es el error que se puede dar por efectos no controladas. Reduciendo se obtiene:

$$Y = X\beta + \epsilon \quad (4.2)$$

Dentro del modelo, los valores dentro del componente del error ϵ_i son variables

independientes e idénticamente distribuidas a una distribución normal con media cero y varianza constante σ^2 (Amat Rodrigo, 2016).

También se debe validar que cada observación sea independiente de las otras, al igual que no exista relación entre las variables independientes (Amat Rodrigo, 2016).

El modelo es utilizado para entender la relación entre la variable dependiente y las variables explicativas, y así entender las posibles causas que ocasionan la variación de Y . Esto se puede obtener estimando los valores de β_i a partir de la información asociada, y también se puede predecir valores de Y a partir de valores de las variables independientes.

La interferencia del espacio dentro de los valores de los datos, provoca que los resultados con este modelo contengan inconsistencias, debido a que no se obtiene independencia entre los valores de cada observación. Por eso se realizan ajustes con distintos modelos, dependiendo del componente que este afectado por el espacio.

Estos modelos ayudan a entender estas violaciones a los supuestos siempre y cuando sean ocasionados por el espacio.

4.2. Modelo de retardo espacial

Este tipo de modelos se utilizan cuando la dependencia del espacio se localiza en la variable independiente (Luc, 2001). Estos modelos son extensión del modelo de regresión lineal en la cual la variable dependiente Y depende de los valores de las áreas vecinas. Este modelo de retardos espaciales es llamado modelo autoregresivo espacial (SAR, por sus siglas en ingles), con forma funcional:

$$Y = \rho WY + X\beta + \epsilon \quad (4.3)$$

donde: W representa la matriz de vecinos, y el escalar ρ determina el nivel de relación del área con sus vecinos en función de la matriz W cuyo valor debe ser estimado.

Tomando en cuenta a I_n la matriz identidad de tamaño n , el modelo puede ser reducido de la siguiente forma:

$$Y = (I_n - \rho W)^{-1}(X\beta + \epsilon) \quad (4.4)$$

Tomando en cuenta esta forma y considerando que se mantengan el término del error como variables independientes e idénticamente distribuidas a una distribución normal con media cero y varianza constante σ^2 , se puede definir su valor esperado mediante la expresión:

$$E[Y] = (I_n - \rho W)^{-1}(X\beta) \quad (4.5)$$

El término $(I_n - \rho W)^{-1}$ se denomina multiplicador espacial e indica que el valor esperado de cada observación y_i dependerá de una combinación lineal de valores X tomados por observaciones vecinas, escalado por el parámetro de dependencia ρ (Luc, 2001).

4.3. Modelo de error espacial

Este tipo de modelos se utilizan cuando la dependencia del espacio se localiza en el término del error (Luc, 2001). Esto puede ser causado por la interferencia de alguna variable que no está controlada en el modelo o por la estructura de los vecinos. Como posee varias razones para presentarse dentro del error, de igual manera se puede corregir de varias formas, donde la más común es representarla a través de sus vecinos de la siguiente manera:

$$\epsilon = \lambda W\epsilon + \mu \quad (4.6)$$

donde λ es el parámetro a estimar, y μ el término del error independiente e idénticamente distribuido a una distribución normal cero y varianza constante $\sigma^2 I_n$. Reduciendo esta expresión se obtiene:

$$\epsilon = (I_n - \lambda W)^{-1}\mu \quad (4.7)$$

Reemplazando este valor en el modelo de regresión lineal se obtiene:

$$Y = X\beta + (I_n - \lambda W)^{-1}\mu \quad (4.8)$$

Este modelo es denominado error espacial (SEM) (Luc, 2001).

4.4. Modelo de Orden Superior

En este se combina el modelo de retardo y el de error espacial, es decir, la afectación del espacio se da sobre la variable dependiente Y y el término del error (Luc, 2001). Este modelo, denominado SARMA, se lo describe de la siguiente manera:

$$Y = \rho WY + X\beta + \epsilon \quad (4.9)$$

$$\epsilon = \lambda W\epsilon + \mu$$

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

Dentro del modelo, la matriz de vecinos W no necesariamente debe ser la misma para las dos partes, por tanto, se utilizan matrices distintas que logren describir mejor cada efecto (Luc, 2001).

4.5. Modelo espacial de Durbin

Este modelo se usa en el caso de que la dependencia del espacio sobre la variable dependiente Y también sea causada por las variables explicativas (Luc, 2001). Partiendo del modelo de retardo espacial, se incluyen variables explicativas con retardos espaciales:

$$Y = \rho WY + X\beta + WX\gamma + \epsilon \quad (4.10)$$

Al realizar el siguiente ajuste como $Z = [X \ WX]$ y $\theta = [\beta' \ \gamma']'$ se transforma el

modelo en:

$$Y = \rho WY + Z\theta + \epsilon \quad (4.11)$$

obteniendo el modelo de retardos espaciales. De forma similar se puede extenderlo con el modelo de orden superior, obteniendo:

$$Y = \rho WY + X\beta + WX\gamma + \epsilon \quad (4.12)$$

$$\epsilon = \lambda W\epsilon + \mu$$

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

4.6. Prueba para Dependencia Espacial

La presencia de dependencia espacial en un modelo de regresión se puede detectar mediante pruebas de diagnóstico. A continuación, se describe la prueba utilizando el test de Moran, para detectar la dependencia espacial en términos de error para los modelos espaciales descritos en la sección anterior.

La forma de utilizar este estadístico es realizando el contraste a la hipótesis de existencia de autocorrelación espacial con el índice de Moran sobre los residuos que se obtengan del modelo de regresión lineal. El índice se puede reescribir de la siguiente forma:

$$I = \frac{n}{W_0} \frac{\epsilon' W \epsilon}{\epsilon' \epsilon} \quad (4.13)$$

con $W_0 = \sum_{i=1}^n \sum_{j=1}^n W_{ij}$ y $\epsilon' \epsilon$ la suma de los cuadrados de los residuos. Sobre este valor se aplica todo lo relacionado al índice global de Moran.

El rechazo de la hipótesis nula sobre la aleatoriedad implica la influencia del espacio dentro del modelo, pero no se especifica el componente que lo causa. Por este motivo, se utilizan los distintos modelos con el fin de obtener el mejor resultado posible (Herrera et al., 2012).

4.7. Estimación

Para estimar los parámetros del modelo de regresión lineal se utiliza mínimos cuadrados ordinarios, el cual busca reducir al máximo la suma de los cuadrados de los residuos. Dado que es el método más utilizado para resolver estos modelos, el parámetro se estima de la siguiente manera:

$$\beta = [X'X]^{-1} X'Y \quad (4.14)$$

Por las modificaciones realizadas sobre el modelo de regresión lineal, ya sea el método por mínimos cuadrados o el estimador descrito anteriormente, no son viables para utilizarlos con las nuevas regresiones descritas.

Para estimar los modelos de regresión espacial se utiliza el método de máxima verosimilitud (MV), el cual se basa en maximizar la probabilidad de distribución conjunta con respecto a los parámetros relevantes. Este método tiene propiedades teóricas asintóticas tales como consistencia, eficiencia o normalidad asintótica, y también se considera robusto para pequeñas variaciones de la suposición de normalidad (Herrera et al., 2012). También se estima utilizando variables instrumentales o más conocido como estimación de mínimos cuadrados en dos etapas (Luc, 2001).

4.7.1. Estimación por Máxima Verosimilitud

Partiendo del modelo de regresión de retardo espacial, para simplificar este modelo se escribe $A = I_n - \rho W$ obteniendo:

$$Y = A^{-1}X\beta + A^{-1}\epsilon \quad (4.15)$$

y considerando que el término del error sigue la distribución:

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

con σ^2 el parámetro de dispersión que también es desconocido. Con estos puntos descritos, se define que Y sigue la distribución:

$$Y \sim N(A^{-1}X\beta, \sigma^2 A^{-1}(A^{-1})') \quad (4.16)$$

La función de densidad de una normal $N(\mu, \sigma^2)$ es:

$$f(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) \quad (4.17)$$

así la función de verosimilitud partiendo de la densidad de Y es la siguiente:

$$\begin{aligned} L(Y|\rho, \beta, \sigma^2) &= \frac{1}{\sqrt{2\pi}} |\sigma^2 A^{-1}(A^{-1})'|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} (Y - A^{-1}X\beta)' A' A (Y - A^{-1}X\beta)\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} |A| \exp\left(-\frac{1}{2\sigma^2} (AY - X\beta)' (AY - X\beta)\right) \end{aligned} \quad (4.18)$$

La función de log-verosimilitud corresponde a:

$$l(Y|\rho, \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |A| - \frac{1}{2\sigma^2} (AY - X\beta)' (AY - X\beta) \quad (4.19)$$

donde los parametros a los cual esta atada son ρ, β y σ^2 . Así, al realizar las distintas derivadas parciales sobre estos términos se obtiene:

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \frac{1}{\sigma^2} [X'(AY - X\beta)] \\ \frac{\partial l}{\partial \rho} &= \frac{1}{\sigma^2} [Y'W(AY - X\beta)] - \frac{1}{|A|} \frac{\partial |A|}{\partial \beta} \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} [(AY - X\beta)' (AY - X\beta)] \end{aligned}$$

Con el sistema de ecuaciones obtenido se encuentra que no es lineal, imposibilitando la obtención de una solución analítica. Para resolverlo se procede a dejar dos parametros en función del tercero, así, solo se debe maximizar un único término y el

resto se calculan sobre este término.

$$\begin{aligned}
0 &= \frac{1}{\sigma^2} \left[X'(AY - X\hat{\beta}) \right] \\
0 &= X'AY - X'X\hat{\beta} \\
X'X\hat{\beta} &= X'AY \\
\hat{\beta} &= (X'X)^{-1}X'AY \\
\hat{\beta} &= (X'X)^{-1}X'(I_n - \rho W)Y \\
\hat{\beta} &= (X'X)^{-1}X'Y - \rho(X'X)^{-1}X'WY \\
\hat{\beta} &= \hat{\beta}_Y - \rho\hat{\beta}_{WY}
\end{aligned}$$

El parámetro $\hat{\beta}$ representa la estimación de β la cual depende de $\hat{\beta}_Y$ y $\hat{\beta}_{WY}$ las cuales son soluciones de los modelos de regresión de Y y WY . Con este resultado se puede obtener el error estimado mediante $\hat{\epsilon} = AY - X\hat{\beta}$, y así trabajar con el valor de dispersión σ^2 .

$$\begin{aligned}
0 &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \left[(AY - X\hat{\beta})'(AY - X\hat{\beta}) \right] \\
0 &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} [\hat{\epsilon}'\hat{\epsilon}] \\
\hat{\sigma}^2 &= \frac{\hat{\epsilon}'\hat{\epsilon}}{n} \\
\hat{\sigma}^2 &= \frac{(\hat{\epsilon}_Y - \rho\hat{\epsilon}_{WY})'(\hat{\epsilon}_Y - \rho\hat{\epsilon}_{WY})}{n}
\end{aligned}$$

donde $\hat{\epsilon}_Y$ y $\hat{\epsilon}_{WY}$ son los residuos resultantes sobre las regresiones de Y y WY . Aplicando estos ajustes, la función de verosimilitud depende de ρ , donde, al resolverlo se tiene la función de verosimilitud concentrada.

$$\ln L_{con}(\rho) = k + \ln |I_n - \rho W| - \frac{n}{2} \ln [(\hat{\epsilon}_Y - \rho\hat{\epsilon}_{WY})'(\hat{\epsilon}_Y - \rho\hat{\epsilon}_{WY})] \quad (4.20)$$

donde k es una constante que no depende de ρ .

Por último, se utiliza el mejor método para encontrar el valor que maximice la función de máxima verosimilitud concentrada.

Del valor encontrado se debe descartar que ρ es distinto de cero dado que se volvería al modelo de regresión lineal. Para esto se aplica una prueba de hipótesis, en la cual la hipótesis nula es $\rho = 0$, donde el multiplicador de Lagrange asociado es:

$$LM_{SAR} = \frac{\left(\frac{Y'W\hat{\epsilon}}{\hat{\sigma}^2}\right)^2}{\hat{\beta}'X'WMWX\hat{\beta}} \sim \chi_{(1)}^2 \quad (4.21)$$

con $M = I_n - X(X'X)^{-1}X'$

El siguiente modelo que se estudia es el de Durbin, del cual se menciona que con la respectiva asociación de parámetros se obtiene un modelo de retardos espaciales. Por este hecho, todo el proceso realizado se extiende a los modelos de Durbin.

A continuación se trabaja con los modelos de error espacial, en el cual sus estimadores se obtienen utilizando el mismo proceso con estimadores de máxima verosimilitud. Obteniendo:

$$\hat{\beta}_{ERR} = [X'(I_n - \lambda W)'(I_n - \lambda W)X]^{-1} X'(I_n - \lambda W)'(I_n - \lambda W)Y \quad (4.22)$$

$$\hat{\sigma}_{ERR}^2 = \frac{(\hat{\epsilon} - \lambda W\hat{\epsilon})'(\hat{\epsilon} - \lambda W\hat{\epsilon})}{n} \quad (4.23)$$

De forma similar se debe descartar que el valor de λ es distinto de cero, así, se define la hipótesis nula $\lambda = 0$ con el multiplicador de Lagrange:

$$LM_{ERR} = \frac{1}{2S_0} \left(\frac{\hat{\epsilon}'W\hat{\epsilon}}{\hat{\sigma}_{ERR}^2}\right)^2 \sim \chi_{(1)}^2 \quad (4.24)$$

siendo $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$.

Por último, se obtiene el modelo de orden superior, en el cual se combinan los resultados obtenidos en los modelos de retardo y error espacial. Lo que resta definir es que los valores de ρ y λ son distintos de cero, fijando la hipótesis nula $\rho = 0, \lambda = 0$

con el multiplicador de Lagrange:

$$LM_{SARMA} = \frac{\left[\left(\frac{Y'W\hat{\epsilon}}{\hat{\sigma}^2} \right) - \left(\frac{\hat{\epsilon}'W\hat{\epsilon}}{\hat{\sigma}^2} \right) \right]^2}{\frac{\hat{\beta}'X'WMWX\hat{\beta}}{\hat{\sigma}^2}} + \frac{1}{2S_0} \left(\frac{\hat{\epsilon}'W\hat{\epsilon}}{\hat{\sigma}^2} \right)^2 \sim \chi_{(2)}^2 \quad (4.25)$$

4.7.2. Estimación por Variables Instrumentales

Partiendo del modelo de regresión de retardo espacial (4.3), al realizar el siguiente ajuste $Z = [X \ WY]$ y $\theta = [\beta' \ \rho']'$ se transforma el modelo en:

$$Y = Z\theta + \epsilon \quad (4.26)$$

con $\epsilon \sim N(0, \sigma^2 I_n)$.

Dado que las variables independientes WY están correlacionadas con el término de error, no es posible estimar el modelo mediante métodos habituales, así se busca un instrumento H que no este correlacionado con el error, y a la vez, este fuertemente correlacionado con WY .

Antes de continuar, se debe definir la convergencia en probabilidad. Sea una sucesión de variables aleatorias X_n con $n \in \mathbb{N}$, converge en probabilidad a una variable aleatoria X si se cumple que: para todo ϵ mayor a 0 se tiene:

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$$

Para el caso de las matrices se intercambia el valor absoluto por la norma matricial y se lo representa como $plim_{n \rightarrow \infty} X_n = X$

Utilizando la definición anterior, el instrumento H debe cumplir con:

$$plim_{n \rightarrow \infty} n^{-1} H'WY = M_{HWY} \quad (4.27)$$

y

$$plim_{n \rightarrow \infty} n^{-1} H' \epsilon = 0 \quad (4.28)$$

siendo M_{HWY} una matriz finita no singular (Kelejian y Prucha, 1997).

Del modelo se tiene por (4.5) que:

$$E[Y] = (I_n - \rho W)^{-1} (X\beta)$$

expresando el término por la siguiente serie:

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

se obtiene que:

$$E[Y] = X\beta + \rho W X\beta + \rho^2 W^2 X\beta + \rho^3 W^3 X\beta + \dots$$

y

$$E[WY] = WX\beta + \rho W^2 X\beta + \rho^2 W^3 X\beta + \rho^3 W^4 X\beta + \dots \quad (4.29)$$

Por lo anterior, $E[WY]$ se relaciona linealmente con WX , W^2X , W^3X , ..., siendo instrumentos para las variables WY . que cumple (4.27). Habitualmente se utiliza los valores de WX y W^2X como instrumentos (Kelejian y Prucha, 1997).

Al aplicar esta idea en el modelo de retardos espaciales se obtiene que el estimador de variables instrumentales (IV) es:

$$\hat{\theta} = \left(\hat{Z}' Z \right)^{-1} \hat{Z}' Y \quad (4.30)$$

donde: $\hat{Z} = Q_H Z = [X \ WY]$, $\hat{W}Y = Q_H WY$ y $Q_H = H(H'H)^{-1}H'$ con los instrumentos $H = [X \ WX]$.

El estimador cumple con:

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow^d N \left(0, \sigma^2 plim_{n \rightarrow \infty} n [\hat{Z}' \hat{Z}]^{-1} \right)$$

La estimación IV de estos parámetros puede ser obtenida a través de un proceso de mínimos cuadrados en dos etapas, denominado 2SLS por sus siglas en inglés. Este procedimiento se describe en los siguientes pasos.

1.- Estimar la relación entre los instrumentos $H = [X \ WX]$ y $Z = [X \ WY]$ donde $Z = H\delta + \epsilon$ y $\delta = (H'H)^{-1}H'Z$, tal que los valores predichos del modelo son: $\hat{Z} = H\hat{\delta} = H(H'H)^{-1}H'Z = Q_H Z$ con $\epsilon \sim (0, \sigma^2 I)$ para cumplir (4.28).

2.- Estimar $Y = Z\theta + \epsilon$, utilizando los resultados del paso anterior:

$$\hat{\theta} = \left(\hat{Z}' \hat{Z} \right)^{-1} \hat{Z}' Y$$

Si se asume que el término del error es independiente pero heterocedástico, se modifica la matriz de varianzas-covarianzas asintótica de $\hat{\theta}$, y se obtiene la siguiente forma:

$$\left(\hat{Z}' \hat{Z} \right)^{-1} \hat{Z}' \hat{\Sigma} \hat{Z} \left(\hat{Z}' \hat{Z} \right)^{-1}$$

donde $\hat{\Sigma}$ es una matriz diagonal cuyos elementos i-ésimos son $\hat{\epsilon}_i^2$, con $\hat{\epsilon}_i = Y_i - Z_i \hat{\theta}$.

4.7.3. Comparativo de las Estimaciones

La diferencia entre estas dos alternativas radica en su funcionamiento bajo los supuestos sobre el término del error.

El estimador de máxima verosimilitud produce estimaciones consistentes en el caso de que los residuos sean independientes y esten idénticamente distribuidos a una normal, en el caso de existir heterocedasticidad el estimador se vuelve inconsistente.

En cambio, el estimador por variables instrumentales tiene la ventaja de que utiliza teoría asintótica sin necesidad del supuesto de normalidad de los residuos, y produce estimaciones consistentes, incluso en el caso de que se detecte heterocedasticidad.

4.8. Interpretación de Parámetros Estimados

Una virtud de la econometría espacial es la capacidad de adaptarse a estrategias de modelado extendidas que describen interacciones multiregionales. Sin embargo, este rico conjunto de información también aumenta la dificultad de interpretar las estimaciones resultantes.

Los modelos de regresión espacial aprovechan la complicada estructura de dependencia entre observaciones que representan países, regiones, condados, etc. Por ello, las estimaciones de los parámetros contienen una gran cantidad de información sobre las relaciones entre las observaciones. Un cambio en una sola observación asociada con cualquier variable explicativa afectará a la región misma (un impacto directo) y, potencialmente afectará a todas las demás regiones de manera indirecta (un impacto indirecto).

Los impactos se pueden definir de la siguiente manera:

1. **Impacto directo promedio.** El impacto de los cambios en la i -ésima observación ocasionada por la j -ésima variable, se denotará por X_{ij} , sobre y_i podría resumirse midiendo el promedio de los valores de la diagonal del producto entre el multiplicador espacial $(I_n - \rho W)^{-1}$ y el vector que contenga el coeficiente asociado a la variable $I_n \beta_j$. Se debe tener presente que, promediar el impacto directo asociado con todas las observaciones i , es similar a las interpretaciones típicas de coeficientes de regresión que representan la respuesta promedio de las variables dependientes a independientes sobre la muestra de observaciones.

2. **Impacto total promedio.** La suma de la i -ésima fila del producto entre el multiplicador espacial $(I_n - \rho W)^{-1}$ y el vector que contenga el coeficiente asociado a la variable $I_n \beta_j$, representaría el impacto total en la observación individual y_i resultante de cambiar la j -ésima variable explicativa en la misma cantidad en todas las n observaciones. Hay n de estas sumas, cada una por cada observación, obteniendo el impacto total del promedio de todas las sumas.

3. **Impacto indirecto promedio.** Es la diferencia entre el impacto total pro-

medio y el impacto directo promedio.

Se define $M(j)_{total}$, $M(j)_{directo}$ y $M(j)_{indirecto}$, que representan los impactos totales promedio, los impactos directos promedio y los impactos indirectos promedio de cambios en la variable j .

$$M(j)_{directo} = n^{-1}tr((I_n - \rho W)^{-1}I_n\beta_j) \quad (4.31)$$

$$M(j)_{total} = n^{-1}\mathbf{1}'_n [(I_n - \rho W)^{-1}I_n\beta_j] \mathbf{1}_n \quad (4.32)$$

$$M(j)_{indirecto} = M(j)_{total} - M(j)_{directo} \quad (4.33)$$

donde $\mathbf{1}_n$ es el vector de unos con tamaño n .

Capítulo 5

Resultados (Caso de Estudio)

En este capítulo se describen todos los resultados obtenidos con la metodología descrita. Se define la información que se utiliza para cada paso y los resultados obtenidos.

El foco de estudio del presente trabajo abarca el ámbito económico dentro del ordenamiento territorial del DMQ, tomando como eje principal a las empresas. La información que se usa de las empresas proviene del directorio de la Superintendencia de Compañías. Estos datos pasaron por un proceso previo de georreferenciación, que se describe en el Capítulo 2, para así obtener su ubicación dentro de la ciudad. Con este proceso se obtuvieron 16 117 empresas, distribuidas tal como se muestra en la Tabla 5.1.

En la Tabla 5.1 se define las 7 actividades económicas con mayor número de empresas dentro de la ciudad. Esto se lo hace para concentrar el estudio, y así descartar actividades que no tienen representatividad dentro de la ciudad. Todas estas empresas restantes se las asocia en Otros.

Como se menciona en el Capítulo 2, el proceso de georeferenciación tiene errores de precisión. Esto debido a que no se posee la dirección exacta de las empresas, sino únicamente se las puede ubicar dentro de la manzana, provocando un error

Actividad	Cantidad de Empresas	Porcentaje
Comercio	3899	24.19 %
Científico Técnico	3097	19.22 %
Administrativos	1440	8.93 %
Transporte	1232	7.64 %
Manufacturas	1215	7.54 %
Información	1160	7.20 %
Construcción	944	5.86 %
Otros	3130	19.42 %

Tabla 5.1: Distribución de las Empresas en Quito

de precisión de una manzana. Esto se corrige al asociar las empresas dentro de los barrios ubicados en la parte urbana de la Ciudad. Los barrios comprenden manzanas completas dentro de su constitución y no comprenden áreas grandes.

Los barrios a tomar en cuenta son 521, sobre los cuales se disponen las 16 mil empresas. La distribución de las empresas en los barrios se muestran en la Tabla 5.2 y en la figura 5.1.

Mínimo	0
Máximo	933
Promedio	30.93

Tabla 5.2: Distribución de las Empresas en los barrios de Quito

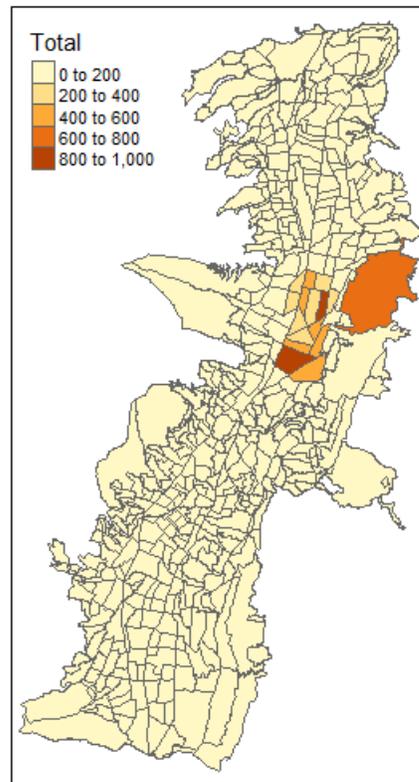


Figura 5.1: Distribución de las Empresas en los barrios de Quito

Con esta información, se calcula el índice de Moran global y se determina si el espacio es influyente sobre las empresas. Para este proceso se define una matriz de vecinos que ayude a identificarlos, para lo cual se utiliza la matriz de distancias inversas considerando que los centroides de cada barrio tengan un máximo de 2 kilómetros de separación. Se utiliza esta matriz pues los resultados son consistentes con la realidad a comparación de los otros tipos de matrices descritas anteriormente.

El cálculo del índice de Moran se realiza en el programa GeoDa. Cuyo resultado se muestra en la Figura 5.2, de aquí se obtiene que el índice es igual a 0.4261 con un valor-p de 0.001. Esto indica que no se puede aceptar la hipótesis nula de aleatoriedad, lo cual demuestra que el espacio es influyente dentro de la distribución de las empresas en los barrios.

El siguiente paso a realizar es el índice local, de forma similar se obtiene este

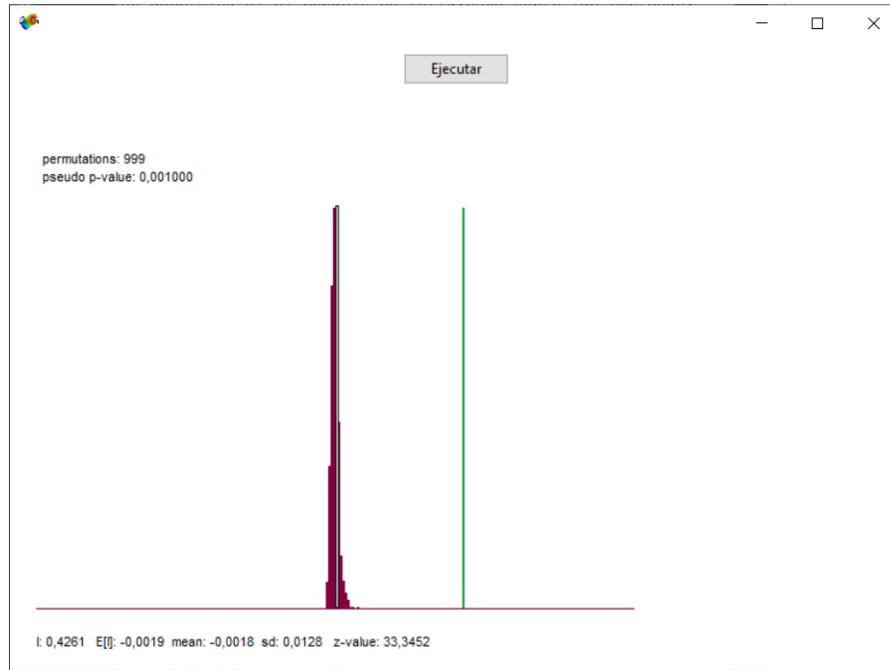


Figura 5.2: Resultados GeoDa. Autocorrelación en Total de Empresas

cálculo en GeoDa. Primero se utiliza el gráfico de dispersión de Moran para calificar a los barrios en los distintos clusters. Definido el cluster al que pertenece cada barrio, se utilizan únicamente aquellos que poseen su índice significativo, y así ver la distribución dentro del mapa.

Alto-Alto	48
Alto-Bajo	5
Bajo-Alto	15
Bajo-Bajo	288
No significativo	165

Tabla 5.3: Cantidad de Barrios por Cluster. Total de Empresas

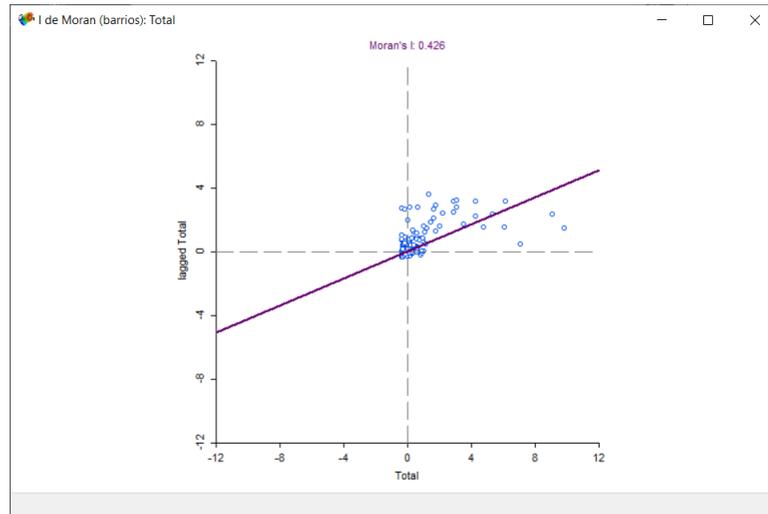


Figura 5.3: Resultados GeoDa. Gráfico de Dispersión de Moran

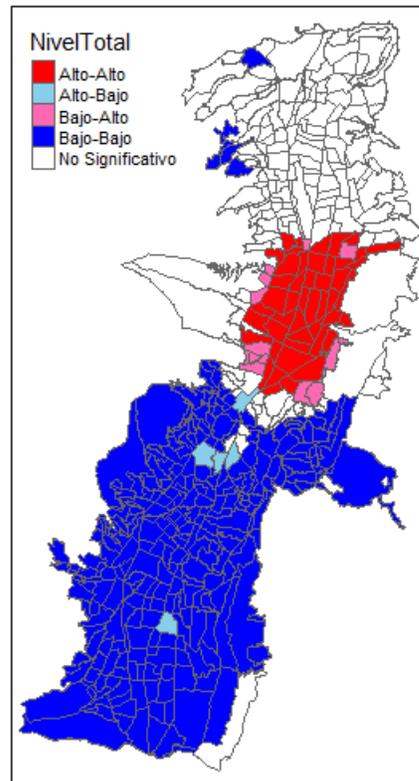


Figura 5.4: Clusters por Total de Empresas

La cantidad de barrios en cada cluster se muestra en la Tabla 5.3, y la locación en la Figura 5.4. De los distintos clusters identificados, se utilizan a los barrios marcados como Alto-Alto, a los cuales se les asocia variables sociodemográficas. Estos valores se muestran en la Tabla 5.4.

Variable	Min	Max
Población	42	8353
Área m2	133 209	1 259 712
Centros Educativos	0	3
Edificios	1	96
UPC	0	2

Tabla 5.4: Características del Cluster Alto-Alto. Total de Empresas

Se identifican a los barrios que se encuentran dentro de los rangos ya establecidos, por tanto, se obtienen 391 barrios cuya distribución se muestra en la Figura 5.5, estos representan el 75% del total de la muestra. Esto se debe a que se están tomando todas las empresas sin ningún tipo de filtro.

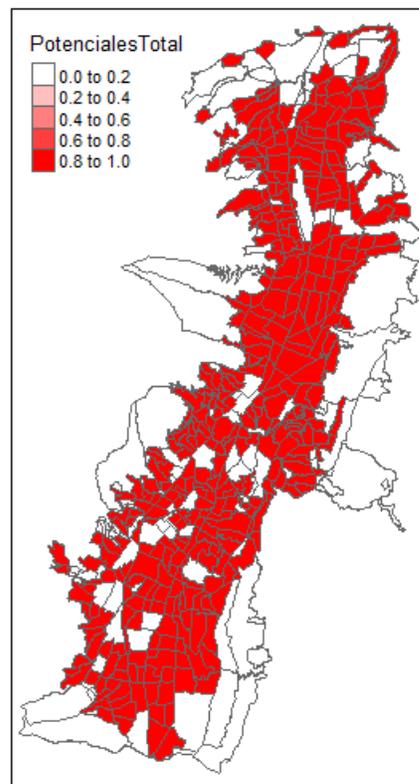


Figura 5.5: Barrios con similares Características. Total de Empresas

Por este motivo, se procede a realizar el mismo proceso sobre cada una de las actividades con mayor densidad dentro de Quito. Estos resultados se encuentran en el **Anexo 3**, con la diferencia de que se restringe más los datos, donde se toma como límites el percentil 5 % y 95 %.

En la Tabla 5.5 se presentan los resultados sobre la cantidad de barrios que fueron definidos con alta densidad al unir toda la información de las actividades económicas, y se los rankea por la cantidad de estas. En la Figura 5.6. se obtiene la distribución de estos barrios.

Actividades	0	1	2	3	4	5	6	7
Barrios	443	26	6	7	2	3	12	22

Tabla 5.5: Barrios por Actividades con alta Densidad

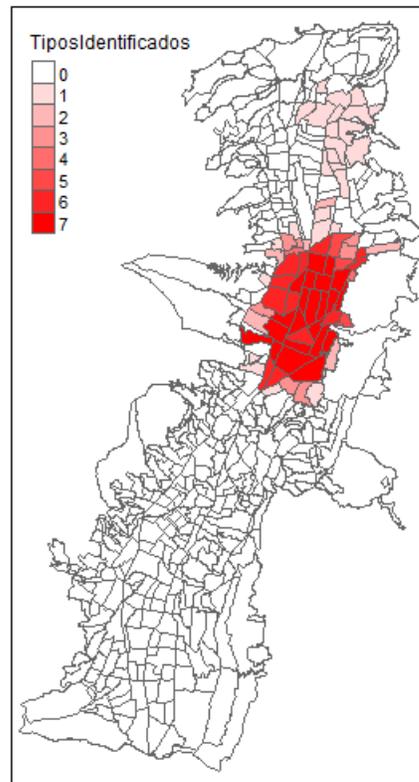


Figura 5.6: Barrios por Actividades con alta densidad

Se considera a los barrios que posean al menos una actividad económica para definir el cluster de barrios con alta densidad de empresas. Así se obtienen 78 barrios cuya disposición referente con las actividades se muestra en la tabla 5.6. Estos barrios se describen en el Anexo 4.

Actividad	Distribución
Comercio	23.61 %
Científico Técnico	21.50 %
Administrativos	8.99 %
Manufacturas	6.50 %
Transporte	4.25 %
Información	8.03 %
Construcción	5.82 %
Otros	21.30 %

Tabla 5.6: Distribución de las Empresas con Alta Densidad

Los barrios localizados se ubican en la parte central de la ciudad. Estos parten desde el barrio el Ejido hasta el sector de El Inca. Existe otro foco ubicado al norte de la ciudad, el cual parte desde La Kennedy hasta el terminal de Carcelén. Se muestra la ubicación en la Figura 5.7 al igual que su distribución en la Tabla 5.7.

Actividad	Distribución Sector 1	Distribución Sector 2
Comercio	22.91 %	33.04 %
Científico Técnico	22.46 %	8.70 %
Administrativos	9.38 %	3.78 %
Manufacturas	5.34 %	21.31 %
Transporte	3.86 %	9.46 %
Información	8.35 %	3.78 %
Construcción	5.87 %	5.17 %
Otros	21.83 %	14.76 %

Tabla 5.7: Distribución de las Empresas con Alta Densidad

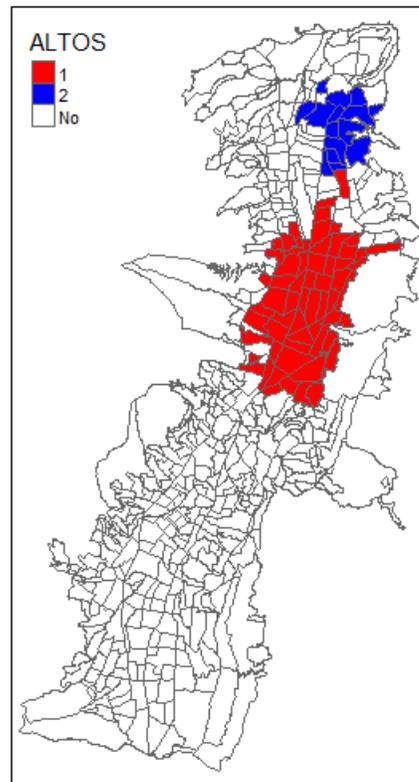


Figura 5.7: Barrios Definidos como Altos

De igual manera se trabaja con los barrios que poseen características similares con aquellos con alta densidad. Estos valores se muestran en la Tabla 5.8. y su distribución en la Figura 5.8.

Actividades	0	1	2	3	4	5	6	7
Barrios	270	16	35	40	14	23	48	75

Tabla 5.8: Barrios por Actividad con características similares

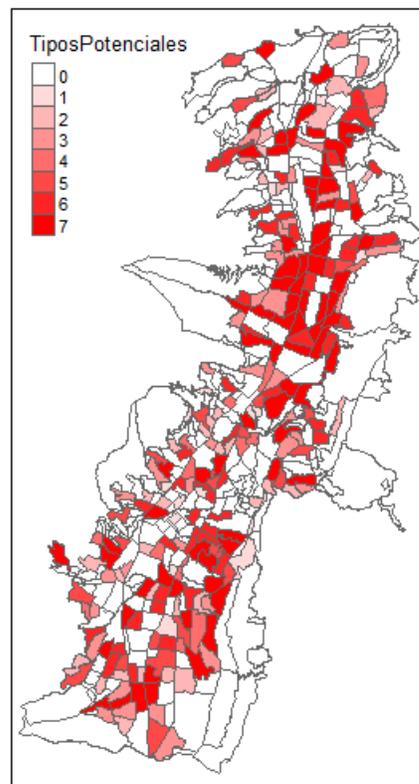


Figura 5.8: Barrios por Actividades con características similares

Una vez identificados los barrios con alta densidad, se estudia la relación entre las características sociodemográficas sobre la cantidad de empresas. Las variables que se utilizan se describen en el Anexo 1, y de forma similar se analiza la influencia del espacio sobre estas. Los resultados se encuentran en el Anexo 5 y se resume en la Tabla 5.9, los cuales nos indican que las variables poseen autocorrelación espacial debido a que sus valores p son menores al 0.05 y todas son positivas indicando que los valores dentro de los barrios son similares a los de sus vecinos.

Variable	Indicador	Valor p
Población	0.0674	0.001
Centros Educativos	0.0364	0.009
Centros de Salud	0.0451	0.007
Paradas de Buses	0.1038	0.001
Edificios	0.1765	0.001
UPC	0.0508	0.001
Áreas Verdes	0.1759	0.001

Tabla 5.9: Índices de Autocorrelación Espacial por Variable

Con estas variables y la cantidad de empresas se procede a estimar el modelo de regresión lineal. Esto se realiza con el software **R** mediante el siguiente modelo:

$$TotalEmpresas \sim Area + Poblacion + CentrosEducativos + Edificios + UPC$$

donde el resultado obtenido es:

Residuals:

Min	1Q	Median	3Q	Max
-351.78	-23.08	-5.12	7.12	717.38

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
Area	1.302e-05	3.799e-06	3.428	0.000656	***
Poblacn	-5.064e-03	1.286e-03	-3.936	9.41e-05	***
CEdctvs	-9.355e+00	4.111e+00	-2.276	0.023269	*
Edif	5.966e+00	3.522e-01	16.938	< 2e-16	***
UPC	-1.246e+01	6.163e+00	-2.022	0.043677	*

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

Residual standard error: 72.85 on 516 degrees of freedom

Multiple R-squared: 0.4348, Adjusted R-squared: 0.4294

F-statistic: 79.4 on 5 and 516 DF, p-value: < 2.2e-16

El resultado de los coeficientes nos indican que: a mayor área y/o edificios incrementa la cantidad de empresas, por el contrario, a mayor población, centros educativos y/o UPCs disminuye la cantidad de empresas.

El análisis siguiente se realiza sobre los residuos de la regresión, en donde al aplicar el índice de Moran se valida si estos residuos son aleatorios o si dependen del espacio. Así, el resultado es el siguiente:

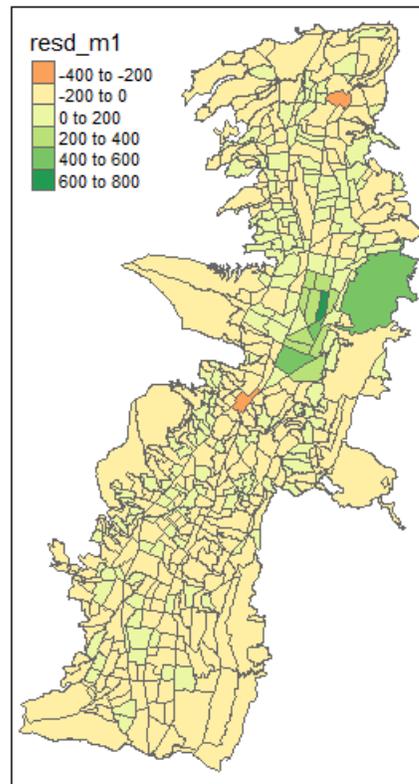


Figura 5.9: Distribución del residuo en el Espacio

Monte-Carlo simulation of Moran I

```
data: BarriosCL1$resd_m1
```

```
weights: mydm.lw
```

```
number of simulations + 1: 1000
```

statistic = 0.45174, observed rank = 831, p-value = 0.169
 alternative hypothesis: greater

Este resultado nos indica que el espacio influye sobre los residuos, en donde, para corregir este fenómeno se procede a utilizar los modelos descritos en el Capítulo 4. Primero se estima utilizando máxima verosimilitud el modelo de retardos espaciales, las variables que lo ajustan son:

$$TotalEmpresas \sim Area + CentrosEducativos + Edificios + UPC + 1$$

donde el resultado obtenido es:

Residuals:

Min	1Q	Median	3Q	Max
-202.215	-13.898	2.907	14.088	573.386

Type: lag

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.4511e+01	3.7321e+00	-6.5676	5.112e-11
Area	1.5395e-05	3.0839e-06	4.9922	5.970e-07
CEdctvs	-7.2573e+00	3.0316e+00	-2.3939	0.016672
Edif	3.7717e+00	2.7879e-01	13.5287	< 2.2e-16
UPC	-1.2503e+01	4.4416e+00	-2.8149	0.004879

Rho: 0.78199, LR test value: 266.5, p-value: < 2.22e-16

Asymptotic standard error: 0.03528

z-value: 22.165, p-value: < 2.22e-16

Wald statistic: 491.3, p-value: < 2.22e-16

Log likelihood: -2842.905 for lag model

ML residual variance (sigma squared): 3044.8, (sigma: 55.18)

```

Number of observations: 521
Number of parameters estimated: 7
AIC: 5699.8, (AIC for lm: 5964.3)
LM test for residual autocorrelation
test value: 1.8458, p-value: 0.17427

```

En este modelo se valida que el parámetro ρ sea distinto de cero para que los retardos sean aceptados. El valor p asociado al parámetro ρ es menor a $2,22e - 16$, lo que indica que ρ es distinto de cero. También al aplicar el índice de Moran se obtiene que los residuos son aleatorios, ya que el valor-p es 0.17427. Por tanto, el modelo de retardos espaciales corrige la influencia del espacio en el modelo planteado.

Para la correcta especificación del modelo estimado por máxima verosimilitud se debe validar que los residuos son normales idénticamente distribuidos, para esto se utilizó el test de Jarque-Bera de normalidad, obteniendo lo siguiente:

```
Jarque-Bera test for normality
```

```

data:  m1s$residuals
JB = 45378, p-value < 2.2e-16

```

Esto indica que el modelo no está correctamente especificado. El siguiente modelo a utilizar es el de error espacial estimado por máxima verosimilitud. Las variables que lo ajustan son:

$$TotalEmpresas \sim Area + CentrosEducativos + Edificios + UPC$$

donde el resultado obtenido es:

```

Residuals:
      Min       1Q   Median       3Q      Max
-192.33205  -13.79774   0.57455   9.93477  577.91289

```

Type: error

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
Area	1.5346e-05	3.1796e-06	4.8263	1.391e-06
CEdctvs	-7.1377e+00	3.0173e+00	-2.3656	0.018000
Edif	3.8764e+00	3.0441e-01	12.7340	< 2.2e-16
UPC	-1.2895e+01	4.4100e+00	-2.9241	0.003455

Lambda: 0.88001, LR test value: 257.54, p-value: < 2.22e-16

Asymptotic standard error: 0.033919

z-value: 25.944, p-value: < 2.22e-16

Wald statistic: 673.1, p-value: < 2.22e-16

Log likelihood: -2849.958 for error model

ML residual variance (sigma squared): 3053.8, (sigma: 55.261)

Number of observations: 521

Number of parameters estimated: 6

AIC: 5711.9, (AIC for lm: 5967.5)

De forma similar al anterior se valida que el parámetro λ es significativo, además, los residuos son aleatorios, obteniendo la corrección sobre el factor del espacio. Se aplica el test de Jarque-Bera sobre los residuos, con el siguiente resultado:

Jarque-Bera test for normality

data: mis\$residuals

JB = 45378, p-value < 2.2e-16

Esto nos indica que el modelo no está correctamente especificado. El siguiente modelo que se ajusta es el de orden superior por máxima verosimilitud, cuyas variables son:

$$TotalEmpresas \sim Area + CentrosEducativos + Edificios + UPC + 1$$

donde el resultado obtenido es:

Residuals:

	Min	1Q	Median	3Q	Max
	-203.3727	-13.5774	3.3672	13.5581	578.1486

Type: sac

Coefficients: (numerical Hessian approximate standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.4169e+01	4.3200e+00	-5.5947	2.210e-08
Area	1.6601e-05	3.2691e-06	5.0781	3.813e-07
CEdctvs	-7.2118e+00	3.0408e+00	-2.3717	0.017707
Edif	3.8370e+00	2.9164e-01	13.1566	< 2.2e-16
UPC	-1.3036e+01	4.4721e+00	-2.9150	0.003557

Rho: 0.73842

Approximate (numerical Hessian) standard error: 0.060928

z-value: 12.12, p-value: < 2.22e-16

Lambda: 0.19991

Approximate (numerical Hessian) standard error: 0.15633

z-value: 1.2788, p-value: 0.20098

LR test value: 268.14, p-value: < 2.22e-16

Log likelihood: -2842.084 for sac model

ML residual variance (sigma squared): 3052.5, (sigma: 55.249)

Number of observations: 521

Number of parameters estimated: 8

AIC: 5700.2, (AIC for lm: 5964.3)

Siguiendo el mismo esquema se valida que los parámetros ρ y λ son distintos de cero. Se aplica el test de Jarque-Bera sobre los residuos, con el siguiente resultado:

Jarque-Bera test for normality

data: miss\$residuals
 JB = 47083, p-value < 2.2e-16

Esto nos indica que el modelo no está correctamente especificado. El último modelo a probar por máxima verosimilitud es el de Durbin, y las variables a utilizar son:

$$TotalEmpresas \sim Area + CentrosEducativos + Edificios + UPC$$

donde el resultado obtenido es:

Residuals:

	Min	1Q	Median	3Q	Max
	-207.2945	-14.5671	1.0672	10.8466	571.3432

Type: mixed

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
Area	1.5474e-05	3.1966e-06	4.8407	1.294e-06
CEdctvs	-7.3942e+00	3.0448e+00	-2.4284	0.015165
Edif	3.7621e+00	3.0632e-01	12.2816	< 2.2e-16
UPC	-1.3480e+01	4.4518e+00	-3.0279	0.002463
lag.Area	-1.6547e-05	5.7068e-06	-2.8995	0.003737
lag.CEdctvs	-1.5840e+01	1.3424e+01	-1.1800	0.237985
lag.Edif	-5.6387e-01	8.1429e-01	-0.6925	0.488642
lag.UPC	-5.5722e+00	1.6333e+01	-0.3412	0.732985

Rho: 0.80419, LR test value: 166.13, p-value: < 2.22e-16

Asymptotic standard error: 0.045009

z-value: 17.867, p-value: < 2.22e-16

Wald statistic: 319.24, p-value: < 2.22e-16

```
Log likelihood: -2842.875 for mixed model
ML residual variance (sigma squared): 3030.4, (sigma: 55.049)
Number of observations: 521
Number of parameters estimated: 10
AIC: 5705.8, (AIC for lm: 5869.9)
LM test for residual autocorrelation
test value: 0.0050373, p-value: 0.94342
```

En este modelo se obtiene que el valor de ρ es distinto de cero y a la vez se obtiene coeficientes sobre las variables retardadas. En donde, a pesar de no ser significativas se conservan ya que su variable sin retardo si es significativa, y además, los residuos son aleatorios. Igualmente se aplica el test de Jarque-Bera sobre los residuos, con el siguiente resultado:

```
Jarque-Bera test for normality
```

```
data: mid$residuals
JB = 46341, p-value < 2.2e-16
```

Esto nos indica que el modelo no esta correctamente especificado.

Al haber probado los modelos con su estimación a través de máxima verosimilitud se encuentra que el mayor problema al momento de validarlos recae en la normalidad requerida en los residuos. En el caso de que se hubiera logrado satisfacer este supuesto en más de un modelo se utilizaría el Índice AIC para seleccionar el modelo final.

A continuación se utilizó la estimación del modelo de retardos espaciales pero con su estimación por variables instrumentales, y las variables a utilizar son:

$$TotalEmpresas \sim Area + CentrosEducativos + Edificios + UPC + 1$$

donde el resultado obtenido es:

Residuals:

	Min	1Q	Median	3Q	Max
	-201.0378	-13.6317	3.0775	14.1814	572.0330

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
Rho	7.8936e-01	5.4944e-02	14.3667	< 2.2e-16
(Intercept)	-2.4637e+01	3.8497e+00	-6.3996	1.558e-10
Area	1.5408e-05	3.0924e-06	4.9827	6.271e-07
CEdctvs	-7.2068e+00	3.0637e+00	-2.3523	0.01866
Edif	3.7533e+00	2.9923e-01	12.5432	< 2.2e-16
UPC	-1.2444e+01	4.4748e+00	-2.7809	0.00542

Residual variance (sigma squared): 3075.9, (sigma: 55.461)

De forma similar a los anteriores, para que el modelo sea correcto el valor que toma ρ debe ser distinto de cero, lo cual es correcto pues su valor-p es menor a $2,2e - 16$.

Para la validación del modelo se necesita que el residuo sea de media cero y en caso de detectarse heterocedasticidad, aplicar la corrección mencionada en el capítulo anterior. Para el primer punto, plantearemos la hipótesis $H_0: \bar{\epsilon} = 0$ contra $H_1: \bar{\epsilon} \neq 0$. Como no se conoce el valor de la varianza se utiliza el valor estimado por los residuos S y dado que se está trabajando con 521 barrios, por teorema central del límite se tiene que el estadístico:

$$Z = \frac{\bar{\epsilon}}{S/\sqrt{n}} \sim N(0, 1)$$

En este caso, el estadístico tomaría valor de $z_0 = \frac{-1,054877 \times 10^{-12}}{55,1935/\sqrt{521}} = -9,9575 \times 10^{-12}$ con un valor-p de 0,9999, por lo tanto, no se puede rechazar que el valor esperado de los residuos es cero.

Para detectar la homocedasticidad o heterocedasticidad se grafica el valor predicho por el modelo frente a los residuos obtenidos, generando el siguiente gráfico:

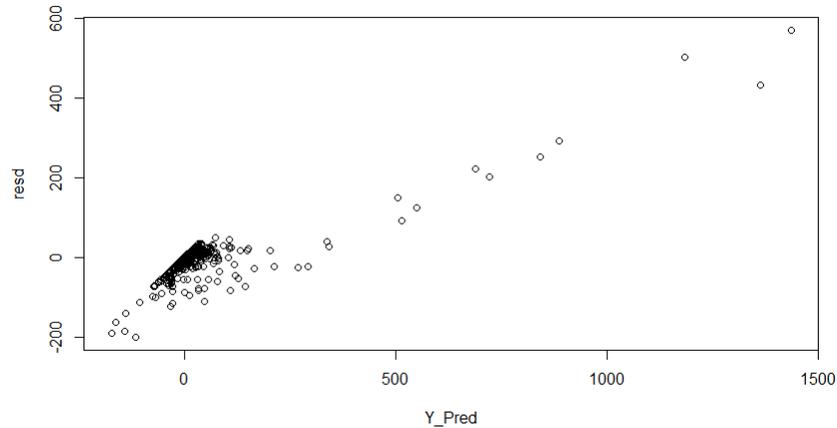


Figura 5.10: Barrios por Actividades con alta densidad

En la Figura 5.10 se detecta una aglomeración en los puntos cercanos al origen y crecimiento en los valores extremos. Se complementa con la aplicación del test Breusch-Pagan (Hernández y Mazo, 2020) para detectar la relación del residuo con las variables del modelo bajo el supuesto de homocedasticidad. De este se obtiene un valor-p de $2,527323 \times 10^{-21}$, por este motivo hay evidencias de que no se detecta homocedasticidad.

Se aplica la corrección mencionada para obtener resultados consistentes aún en frente de este efecto. Así las variables utilizadas son:

$$TotalEmpresas \sim CentrosEducativos + Edificios + UPC + 1$$

donde el resultado obtenido es:

Residuals:

Min	1Q	Median	3Q	Max
-205.3788	-11.6401	2.1152	11.2195	617.9470

Coefficients:

Estimate	HCO	std. Error	z value	Pr(> z)
----------	-----	------------	---------	----------

Rho	0.79017	0.11262	7.0163	2.279e-12
(Intercept)	-17.39316	4.59781	-3.7829	0.0001550
CEdctvs	-7.30582	3.86766	-1.8890	0.0588985
Edif	3.83538	1.10632	3.4668	0.0005267
UPC	-12.75386	5.00535	-2.5480	0.0108328

Residual variance (sigma squared): 3217.7, (sigma: 56.724)

De forma similar se valida que ρ es distinto a cero, y con un valor-p de 0,9999 asociado al valor esperado de los residuos no se puede rechazar que este es cero.

Con esto se ha encontrado un modelo correctamente especificado para el total de las empresas, en la Tabla 5.10 se describe los coeficientes que toman las variables.

Intercepto	-17.39316
Centros Educativos	-7.30582
Edificios	3.83538
UPC	-12.75386
ρ	0.79017

Tabla 5.10: Coeficientes de Factores Externos sobre el Total de empresas

Estos coeficientes no describen directamente la influencia de las variables sobre el total de las empresas debido a la influencia del espacio. A continuación se estiman los impactos promedios de las variables obteniendo:

Variable	Directo	Indirecto	Total
Centros Educativos	-7.900883	-26.91636	-34.81725
Edificios	4.147774	14.13045	18.27822
UPC	-13.792679	-46.98827	-60.78095

Tabla 5.11: Impacto de Factores Externos sobre el Total de empresas

Se realiza el mismo proceso para todas las actividades obteniendo el mejor modelo para cada uno (ver Anexo 5) y extrayendo el impacto por cada variable, los cuales se muestran en las Tablas 5.12, 5.13, 5.14 y 5.15.

Actividad	Intercepto
Comercio	-3.82655
Científico Técnico	-3.41149
Administrativos	-1.68829
Manufacturas	-0.95429
Transporte	-0.79558
Información	-1.31652
Construccion	-0.80750

Tabla 5.12: Coeficiente del Intercepto por Actividad Económica

	Comercio	C. Tecnico	Administrativo	Manufacturas	Transporte	Información	Construcción
Población		-0.0009	-0.0005	-0.0002			-0.0002
Centros Educativos	-2.2265			-0.6610			
Edificios	0.9368	0.8320	0.4439	0.3370	0.1949	0.2841	0.1803
Paradas de Buses		0.3521	0.1541				0.0834
UPC	-2.7442	-3.1865	-1.4189	-0.6452		-1.4646	-0.6973

Tabla 5.13: Impacto Directo de los Factores por Actividad Económica

	Comercio	C. Tecnico	Administrativo	Manufacturas	Transporte	Información	Construcción
Población		-0.0033	-0.0011	-0.0003			-0.0009
Centros Educativos	-8.1039			-1.3559			
Edificios	3.4095	3.1814	1.1177	0.6913	0.3940	1.2079	0.9953
Paradas de Buses		1.3462	0.3879				0.4604
UPC	-9.9881	-12.1842	-3.5724	-1.3235		-6.2263	-3.8491

Tabla 5.14: Impacto Indirecto de los Factores por Actividad Económica

	Comercio	C. Tecnico	Administrativo	Manufacturas	Transporte	Información	Construcción
Población		-0.0041	-0.0016	-0.0005			-0.0010
Centros Educativos	-10.3304			-2.0168			
Edificios	4.3462	4.0134	1.5616	1.0283	0.5889	1.4920	1.1756
Paradas de Buses		1.6983	0.5419				0.5438
UPC	-12.7323	-15.3706	-4.9913	-1.9687		-7.6910	-4.5464

Tabla 5.15: Impacto Total de los Factores por Actividad Económica

Los resultados de estos modelos muestran que los signos de los parámetros son similares a los obtenidos en primera instancia. Así, la cantidad de edificios y las paradas de buses incrementan la cantidad de empresas. Por el contrario, la población, los centros educativos y los UPCs disminuyen la cantidad de empresas. Se debe considerar que en ciertos casos, algunas variables no son significativas dependiendo de la actividad económica.

Capítulo 6

Conclusiones y Recomendaciones

6.1. Conclusiones

Dado que la alcaldía del DMQ utiliza técnicas descriptivas al momento de realizar el diagnóstico del ordenamiento territorial, en este trabajo se propone el uso de nuevas herramientas centradas en el análisis espacial.

Se analiza las variables de tipo espacial para comprender su distribución en la ciudad, de lo cual se obtiene que las variables población, centros educativos, centros de salud, paradas de buses, edificios, unidades de policias comunitarias y áreas verdes son influenciadas por el espacio de forma positiva.

La variable que explica la cantidad de edificios es la que posee el mayor índice de Moran, el cual al ser positiva nos indica que la cantidad de edificios dentro de un barrio es similar a los de sus vecinos, relacionado por la distancia que los separa. Este fenómeno es el mismo en todas las variables descritas anteriormente.

Al observar la cantidad de empresas sobre los barrios, se obtiene que son influenciados por el espacio con autocorrelación positiva, es decir, las empresas poseen el mismo efecto que las variables descritas anteriormente.

Se encontró que en el sector centro norte de la ciudad, partiendo desde el parque El Ejido hasta el sector del Inca, existe alta densidad de empresas. Esta sección de la ciudad se describe por las actividades de comercio, científico técnicas, administrativas, manufacturas, transporte, información y construcción. Existe otro conjunto de barrios ubicado al norte de la ciudad partiendo desde La Kennedy hasta el terminal de Carcelén, el cual se diferencia del anterior por poseer un incremento en empresas manufactureras (Tabla 5.7).

Al utilizar los modelos de regresión espacial se logró corregir el fenómeno generado por el espacio, en comparación a los resultados de la regresión lineal.

A cada actividad económica se le ajusta el mejor modelo espacial, seleccionando aquel que este correctamente especificado dentro de los descritos en el Capítulo 4. Se debe tomar en cuenta que el modelo que se aplique no es necesariamente el mismo para todos, ya sea, por el tipo de modelo o las variables que se utilizan.

La variable que siempre se toma en cuenta dentro de los modelos es la cantidad de edificios. Estos a pesar de que ocupan determinado espacio en el área de cada barrio, logran ampliarlo al ser construcciones de forma vertical. Es por eso que, en todos los modelos realizados, los edificios resultan ser significativos sobre cada una de las actividades. Donde esta variable siempre posee signo positivo, indicando mayor influencia sobre las actividades con mayor densidad (comercio y científico técnico).

Otro factor relevante son las paradas de buses. Este tiene asociado un coeficiente con valor positivo, sin embargo no es significativo para todas las actividades. Las paradas de buses tienen relación directa sobre las empresas científico técnicas, administrativas y de construcción.

El resto de los factores descritos poseen signos negativos (incluyendo el intercepto). Todos estos factores tienen en común que cubren grandes áreas como viviendas, centros educativos y unidades de policía comunitaria. También este valor negativo del intercepto indica que, sin ninguno de los factores no se debe tener empresas dentro de los barrios.

En cuanto a las variables área del barrio, centros de salud y áreas verdes, se

encuentra que no son significativas para ninguna de las actividades económicas estudiadas.

Finalmente, con los objetivos cumplidos se puede concluir que al implementar estos resultados en el análisis descriptivo que realiza la alcaldía del DMQ, se busca mejorar el ordenamiento territorial de la ciudad, y así evitar la aglomeración de empresas y recursos.

6.2. Recomendaciones

La información utilizada en el presente trabajo sobre la dirección de las empresas representa un limitante en la realización del estudio, debido a que únicamente se cuenta con el directorio de empresas de la SUPERCIAS actualizado al año 2019. Se vuelve necesario hacer un análisis pospandemia ya que se estima que cerraron algunas empresas, al poseer estos datos podría realizarse un nuevo estudio para medir este impacto, por lo cual es recomendable actualizar la información.

De la misma forma la variable población cuenta con datos actualizados al año 2010 y el resto de variables al año 2014. Por lo cual, es recomendable repetir el estudio cuando se tenga información más reciente.

Se recomienda recopilar información de las empresas hasta la más actualizada, para aplicar modelos de regresión espacio temporales y con ello identificar el movimiento de locación de las empresas.

En el trabajo se utilizan modelos de regresión espacial enfocados en la econometría, sin embargo, existen modelos similares a través de inferencia bayesiana utilizando aproximaciones de Laplace Anidada Integrada (INLA). Para mejorar el trabajo se recomienda utilizar estos modelos y comparar sus resultados. En este trabajo no se toman en cuenta por las deficiencias propias dentro de la estadística bayesiana.

Finalmente, se propone realizar estudios de índole espacial ya que con esta información el ente gubernamental como el empresarial pueden definir acciones según

sea la necesidad. Así, como invertir en cierto barrio; o decidir establecerse dentro o fuera de la aglomeración empresarial.

Bibliografía

- Joaquín Amat Rodrigo. Introducción a la regresión lineal múltiple. 2016. URL https://github.com/JoaquinAmatRodrigo/Estadistica-con-R/blob/master/PDF_format/25_Regresi%C3%B3n_lineal_m%C3%BAltiple.pdf.
- Manu Benito Moreno. Sig y datos espaciales. introducción y conceptos básicos. 2017.
- Secretaría General de la Comunidad Andina CAN. Manual sobre los directorios de empresas con fines estadísticos en la comunidad andina. 2010.
- Juan Pablo Celemín. Autocorrelación espacial e indicadores locales de asociación espacial. importancia, estructura y aplicación. 2011. URL https://www.researchgate.net/publication/317536284_Autocorrelacion_espacial_e_indicadores_locales_de_asociacion_espacial_Importancia_estructura_y_aplicacion?enrichId=rgreq-10c69359effe5306ba87c165745189af-XXX&enrichSource=Y292ZXJQYWdlOzMxNzUzNjI4NDtBUzo2Nzc3OTk0NzE5MDI3MjBAMTUzODYxMTM5NTUxOA%3D%3D&el=1_x_3&_esc=publicationCoverPdf.
- Juan Carlos Cid. Aplicación de un modelo de econometría espacial a datos agregados de asistencia escolar en la argentina. *Gobierno de la Provincia de Salta*, 2011.
- Alcaldía Metropolitana de Quito. Plan metropolitano de desarrollo y ordenamiento territorial. 2015. URL http://www7.quito.gob.ec/mdmq_ordenanzas/Sesiones%20del%20Concejo/2015/Sesi%C3%B3n%20Extraordinaria%202015-02-06/Plan%20Metropolitano%20de%20Desarrollo%20y%20Ordenamiento%20Territorial%202015-%202025/Volumen%20I/.

Gobierno Abierto de Quito. Portal gobierno abierto de quito. 2020. URL http://gobiernoabierto.quito.gob.ec/?page_id=988.

Asamblea Nacional del Ecuador. Ley orgánica de ordenamiento territorial, uso y gestión del suelo. 2016. URL <https://www.habitatyvivienda.gob.ec/wp-content/uploads/downloads/2016/08/Ley-Organica-de-Ordenamiento-Territorial-Uso-y-Gestion-de-Suelo1.pdf>.

Michael F. Goodchild. Spatial autocorrelation. 1986.

Freddy Hernández y Mauricio Mazo. Modelos de regresión con r. 2020.

Marcos Herrera, Jorge Paz, y Juan Cid. Introducción a la econometría espacial. una aplicación al estudio de la fecundidad en la argentina usando r. 2012. URL https://www.researchgate.net/publication/235799573_Introduccion_a_la_Econometria_Espacial_Una_Aplicacion_al_Estudio_de_la_Fecundidad_en_la_Argentina_usando_R.

INEC. Portal instituto nacional de estadísticas y censos. URL <https://www.ecuadorencifras.gob.ec/estadisticas/>.

Instituto Nacional de Estadística y Censo INEC. Resultados principales del directorio de empresas y establecimientos 2017. 2018. URL https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Economicas/DirectorioEmpresas/Directorio_Empresas_2017/Documentos_DIEE_2017/Documentos_DIEE_2017/Principales_Resultados_DIEE_2017.pdf.

Harry H. Kelejian y Ingmar R. Prucha. A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances1. 1997. URL <http://econweb.umd.edu/~kelejian/research/P071897.PDF>.

Anselin Luc. Local indicators of spatial association – lisa. 1995. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.1995.tb00338.x>.

- Anselin Luc. *A Companion to Theoretical Econometrics*, cap. 14. Blackwell Publishing Ltd, 2001.
- Anselin Luc. Documentacion geoda. 2003.
- Javier Martínez Argudo. Criterios de clasificación de empresas. 2018. URL <http://www.econosublime.com/2018/12/criterios-clasificacion-empresas.html>.
- Joan Carles Martori Cañas y Karen Hoberg. Nuevas técnicas de estadística espacial para la detección de clusters residenciales de población inmigrante. *Revista Electrónica de Geografía y Ciencias Sociales*, 2007. URL <http://www.ub.edu/geocrit/sn/sn-263.htm>.
- Marcos Merino. ¿qué es una api y para qué sirve? 2014. URL <https://www.ticbeat.com/tecnologias/que-es-una-api-para-que-sirve/>.
- Organización de las Naciones Unidas ONU. Clasificación industrial internacional uniforme de todas las actividades económicas (ciiu). 2005.
- Julián Pérez Porto y María Merino. Definición de geolocalización. 2018a. URL <https://definicion.de/geolocalizacion/>.
- Julián Pérez Porto y María Merino. Definición de georeferenciación. 2018b. URL <https://definicion.de/georeferenciacion/>.
- Documentación de QGIS. Sistema de referencia de coordenadas. URL https://docs.qgis.org/3.10/es/docs/gentle_gis_introduction/coordinate_reference_systems.html.
- Diego Rojas. Localización de centros de empleo y su influencia sobre la distribución de la población en el distrito metropolitano de quito. *Analítika*, 2015.
- SNI. Portal sistema nacional de información. URL <https://sni.gob.ec/inicio>.
- SUPERCIAS. Portal superintencia de compañías, valores y seguros. URL <https://www.supercias.gob.ec/portalscvsv/>.

Apéndice A

Anexo 1. Datos Espaciales

Información	Directorio de Empresas
Procedencia	Superintendencia de Compañías
Año de Información	2019
Tratamiento	Georeferenciación de Direcciones
Tipo de Dato	Evento
Nivel de Información	Distrito Metropolitano de Quito

Tabla A.1: Información Directorio de Empresas

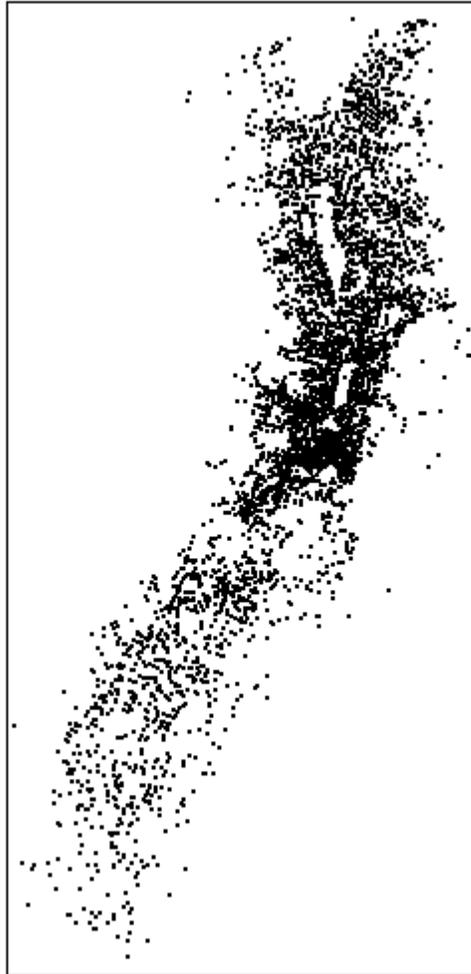


Figura A.1: Información Directorio de Empresas

Información	Población
Procedencia	Instituto de Estadísticas y Censos
Año de Información	2010
Tratamiento	Asociación de Datos con su Geometría
Tipo de Dato	Polígonos
Nivel de Información	Distrito Metropolitano de Quito

Tabla A.2: Información de Población



Figura A.2: Información de Población

Información	Centros Educativos
Procedencia	Sistema Nacional de Información
Año de Información	2014
Tratamiento	Selección de Datos necesarios
Tipo de Dato	Evento
Nivel de Información	Nacional

Tabla A.3: Información de Centros Educativos



Figura A.3: Información de Centros Educativos

Información	Centros de Salud
Procedencia	Sistema Nacional de Información
Año de Información	2014
Tratamiento	Selección de Datos necesarios
Tipo de Dato	Evento
Nivel de Información	Nacional

Tabla A.4: Información de Centros de Salud



Figura A.4: Información de Centros de Salud

Información	Paradas de Buses
Procedencia	Gobierno Abierto de Quito
Año de Información	2015
Tratamiento	Selección de Datos necesarios
Tipo de Dato	Evento
Nivel de Información	Distrito Metropolitano de Quito

Tabla A.5: Información de Paradas de Buses



Figura A.5: Información de Paradas de Buses

Información	Edificios
Procedencia	Instituto de Estadísticas y Censos
Año de Información	2014
Tratamiento	Selección de Datos necesarios
Tipo de Dato	Evento
Nivel de Información	Nacional

Tabla A.6: Información de Edificios

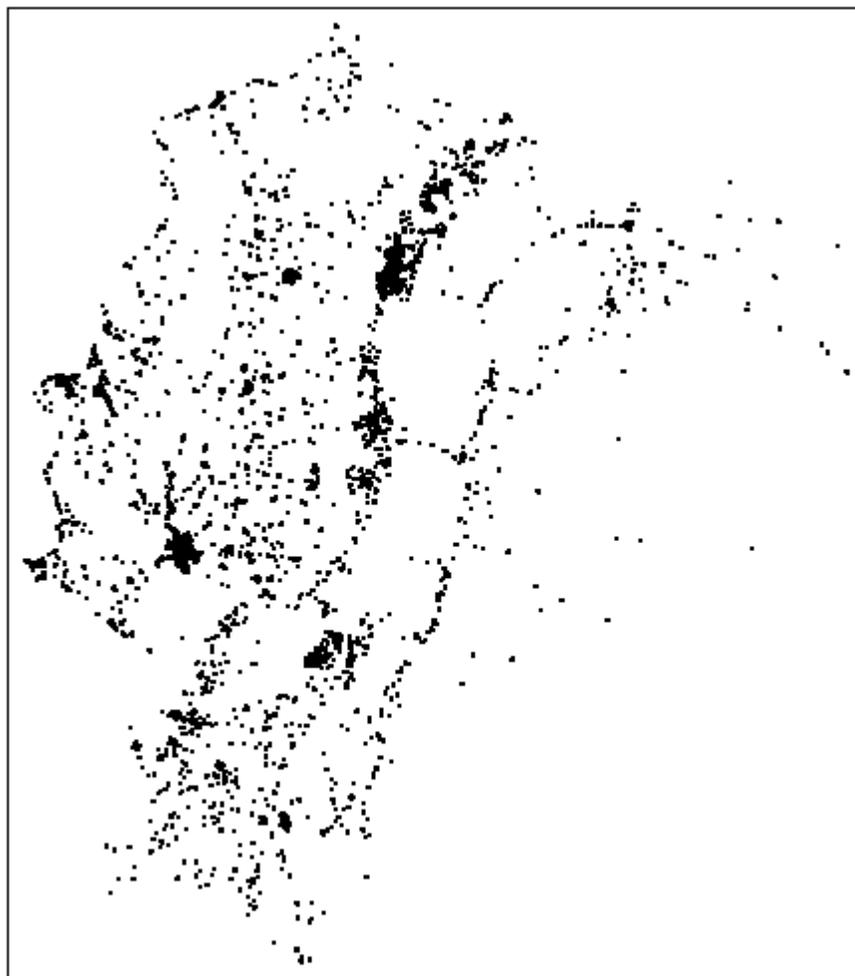


Figura A.6: Información de Edificios

Información	Unidades de Policia Comunitaria
Procedencia	Gobierno Abierto de Quito
Año de Información	2015
Tratamiento	Selección de Datos necesarios
Tipo de Dato	Evento
Nivel de Información	Distrito Metropolitano de Quito

Tabla A.7: Información de Unidades de Policia Comunitaria

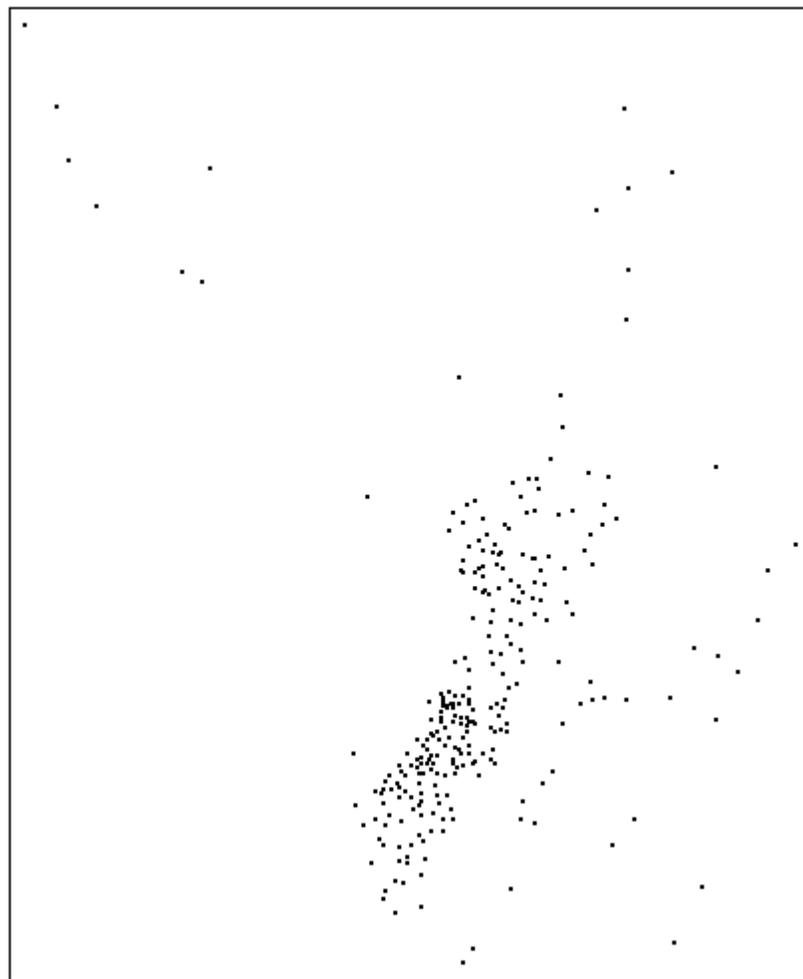


Figura A.7: Información de Unidades de Policia Comunitaria

Información	Áreas Verdes
Procedencia	Gobierno Abierto de Quito
Año de Información	2015
Tratamiento	Selección de Datos necesarios
Tipo de Dato	Polígono
Nivel de Información	Distrito Metropolitano de Quito

Tabla A.8: Información de Áreas Verdes

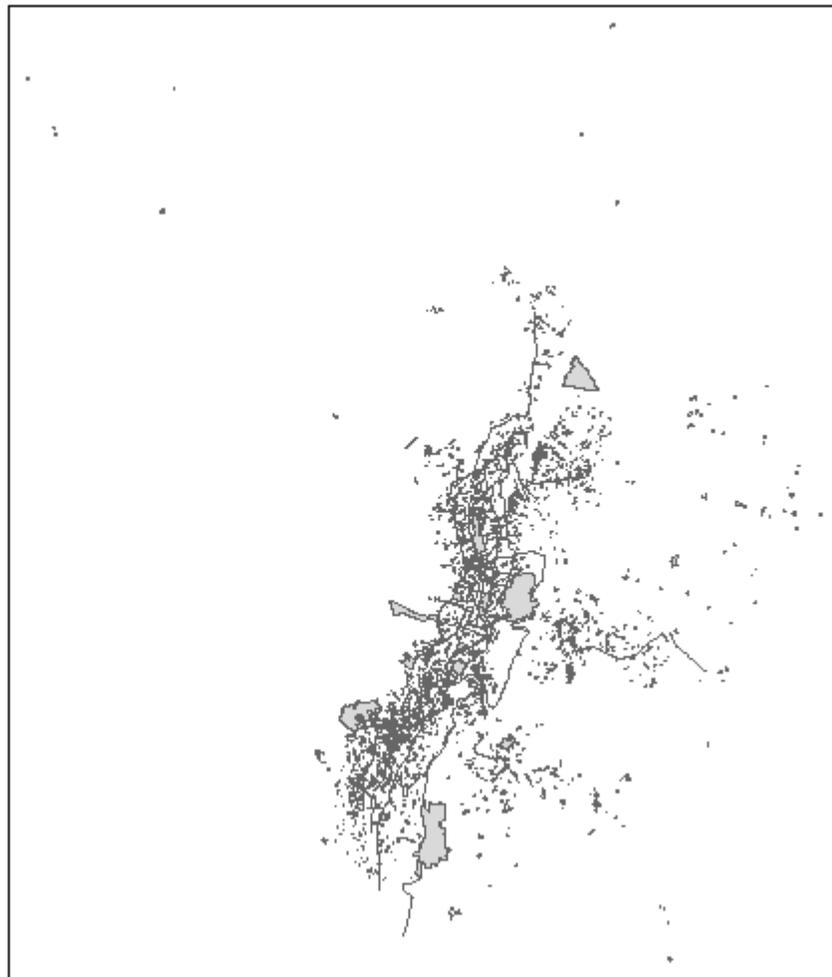


Figura A.8: Información de Áreas Verdes

Apéndice B

Anexo 2. Código de Georeferenciación

```
## Librerias
library(ggmap)
## Carga de la data
library(readxl)
baseC <- data.frame(read_excel("Directorio Resumido.xlsx"))

## Generacion de Direcciones
dir1 <- paste(Aux$CALLE,Aux$INTERSECCIÓN,Aux$CIUDAD,sep=" ")
dir2 <- ifelse(is.na(Aux$BARRIO)==T,
              paste(Aux$CALLE,Aux$INTERSECCIÓN,Aux$CIUDAD,sep=" "),
              paste(Aux$CALLE,Aux$INTERSECCIÓN,Aux$BARRIO,Aux$CIUDAD,sep=" "))
dir3 <- paste(Aux$NOMBRE.COMPAÑÍA,Aux$CIUDAD,sep=" ")
dir4 <- ifelse(is.na(Aux$EDIFICIO)==T,
              paste(Aux$NOMBRE.COMPAÑÍA,Aux$CIUDAD,sep=" "),
              paste(Aux$EDIFICIO,Aux$CIUDAD,sep=" "))

## Variables de Almacenamiento
res1 <- c()
res2 <- c()
```

```
resd <- c()
config <- c()
n<-dim(baseC)[1]

## Proceso Reiterativo de Georeferenciación
for (i in 1:n) {
  resultados <- geocode(dir1[i], output = "latlona", source = "google")
  lon <- as.numeric(resultados[1])#lon
  lat <- as.numeric(resultados[2])#lat
  if(lon> -78.982357 & lon< -78.191341 & lat> -0.636362 & lat< 0.223305
    & is.na(lon)==F & is.na(lat)==F){
    res1[i] <- lon
    res2[i] <- lat
    resd[i] <- resultados$address
    config[i] <- 1
  }else{
    if(dir1[i]!=dir2[i]){
      resultados <- geocode(dir2[i], output = "latlona", source = "google")
      lon <- as.numeric(resultados[1])#lon
      lat <- as.numeric(resultados[2])#lat
    }
    if(lon> -78.982357 & lon< -78.191341 & lat> -0.636362 & lat< 0.223305
      & is.na(lon)==F & is.na(lat)==F){
      res1[i] <- lon
      res2[i] <- lat
      resd[i] <- resultados$address
      config[i] <- 2
    }else{
      resultados <- geocode(dir3[i], output = "latlona", source = "google")
      lon <- as.numeric(resultados[1])#lon
      lat <- as.numeric(resultados[2])#lat
      if(lon> -78.982357 & lon< -78.191341 & lat> -0.636362 & lat< 0.223305
```

```
      & is.na(lon)==F & is.na(lat)==F){
    res1[i] <- lon
    res2[i] <- lat
    resd[i] <- resultados$address
    config[i] <- 3
  }else{
    if(dir3[i]!=dir4[i]){
      resultados <- geocode(dir4[i], output = "latlona", source = "google")
      lon <- as.numeric(resultados[1])#lon
      lat <- as.numeric(resultados[2])#lat
    }
    if(lon> -78.982357 & lon< -78.191341 & lat> -0.636362 & lat< 0.223305
      & is.na(lon)==F & is.na(lat)==F){
      res1[i] <- lon
      res2[i] <- lat
      resd[i] <- resultados$address
      config[i] <- 4
    }else{
      res1[i] <- 0
      res2[i] <- 0
      resd[i] <- "NADA"
      config[i] <- 0
    }
  }
}
}
}
}

## Construcción de una Base con los resultados obtenidos
puntos<-data.frame(Aux$EXPEDIENTE,Dir1=dir1,Dir2=dir2,Dir3=dir3,Dir4=dir4,
                  lon=res1,lat=res2,Res=resd,Sel=config)
write.csv(puntos,"Georeferenciacion.csv")
```

```
save(puntos,file = "Georeferenciacion.RData")
```

Apéndice C

Anexo 3. Análisis Espacial por Actividades de las Empresas

C.1. Análisis a empresas de Comercio

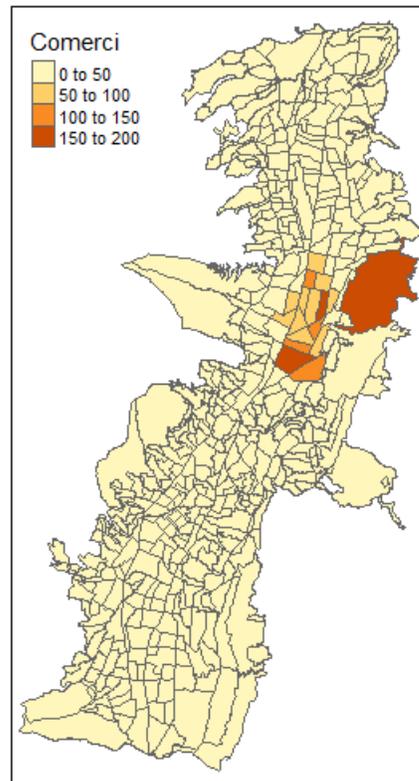


Figura C.1: Distribución de las Empresas de Comercio en los barrios de Quito

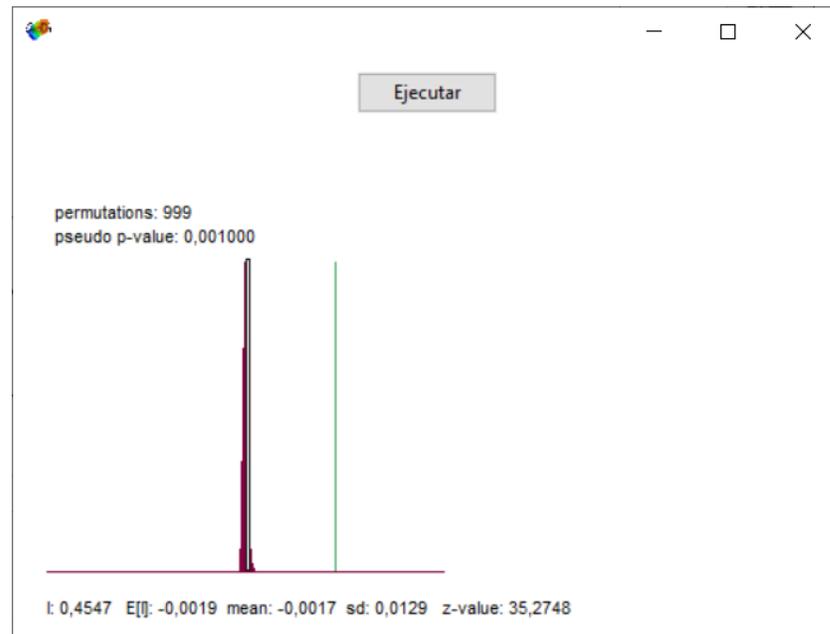


Figura C.2: Resultados GeoDa. Autocorrelación Espacial de Comercio

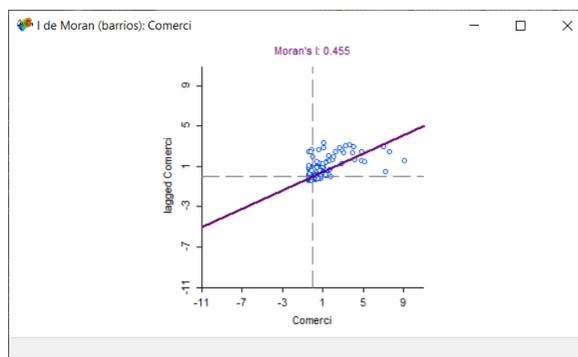


Figura C.3: Resultados GeoDa. Gráfico de Dispersión de Moran Comercio

Alto-Alto	53
Alto-Bajo	7
Bajo-Alto	16
Bajo-Bajo	287
Nosignificativo	158

Tabla C.1: Cantidad de Barrios por Cluster. Comercio

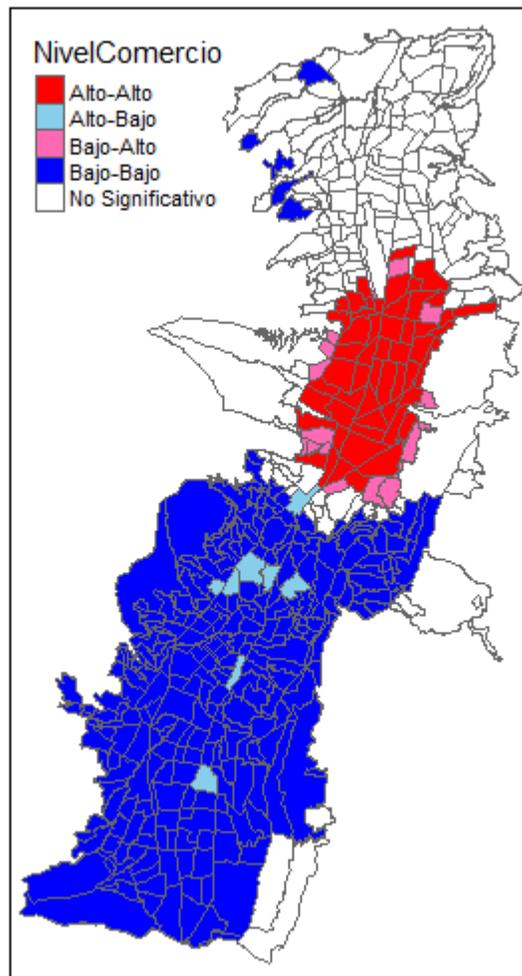


Figura C.4: Clusters por Comercio

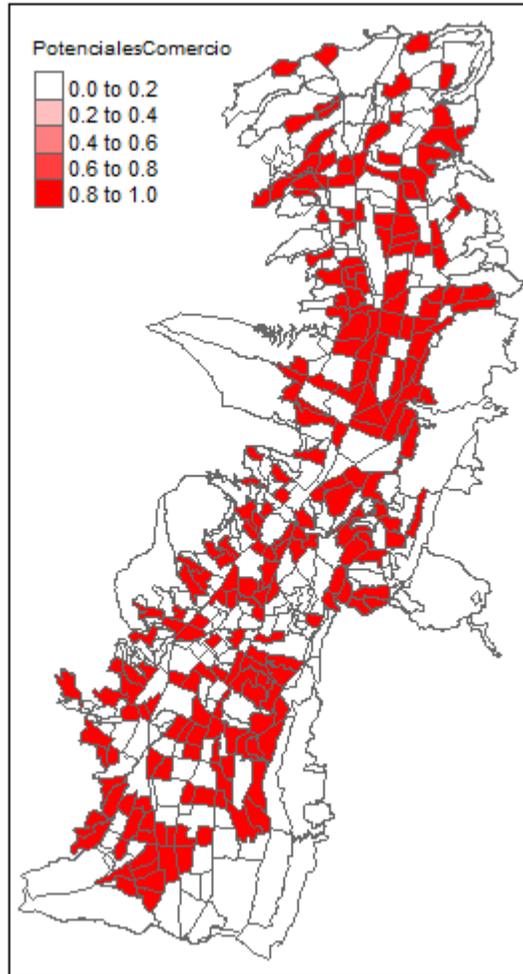


Figura C.5: Barrios con similares Características. Comercio

C.2. Análisis a empresas Científico Técnicas

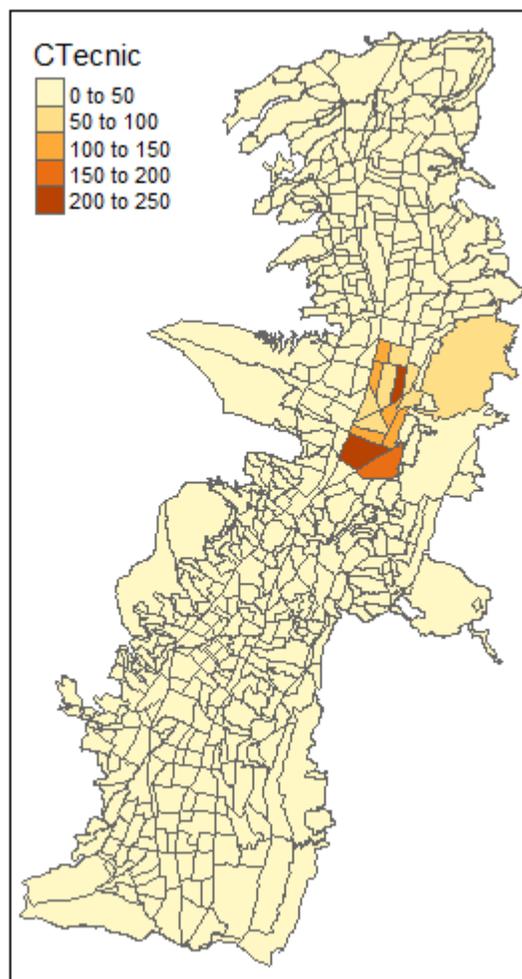


Figura C.6: Distribución de las Empresas de Científico Técnico en los barrios de Quito

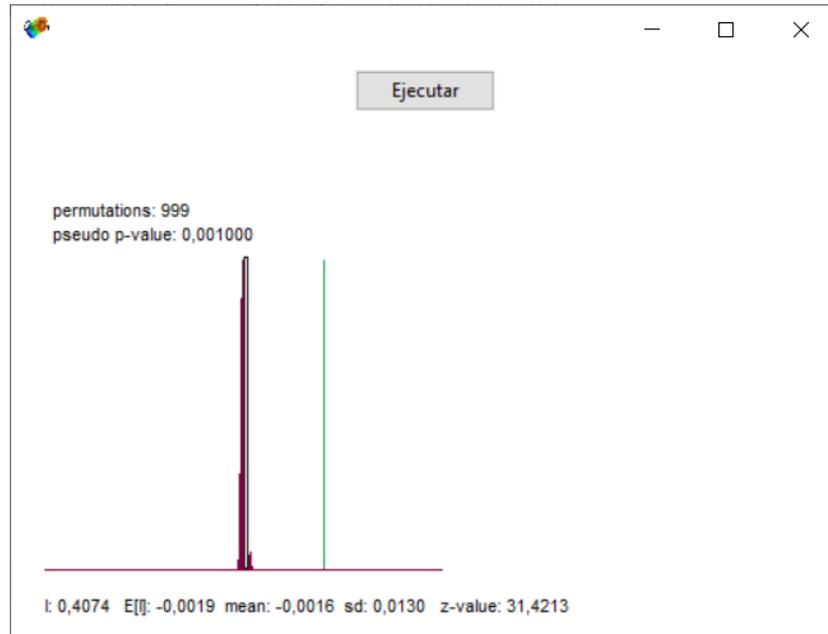


Figura C.7: Resultados GeoDa. Autocorrelación Espacial de Científico Técnico

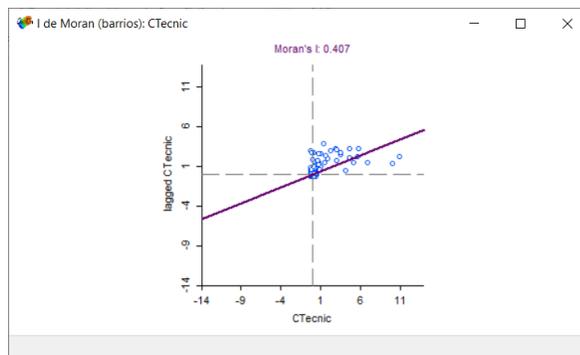


Figura C.8: Resultados GeoDa. Gráfico de Dispersión de Moran Científico Técnico

Alto-Alto	41
Alto-Bajo	3
Bajo-Alto	11
Bajo-Bajo	283
Nosignificativo	183

Tabla C.2: Cantidad de Barrios por Cluster. Científico Técnico

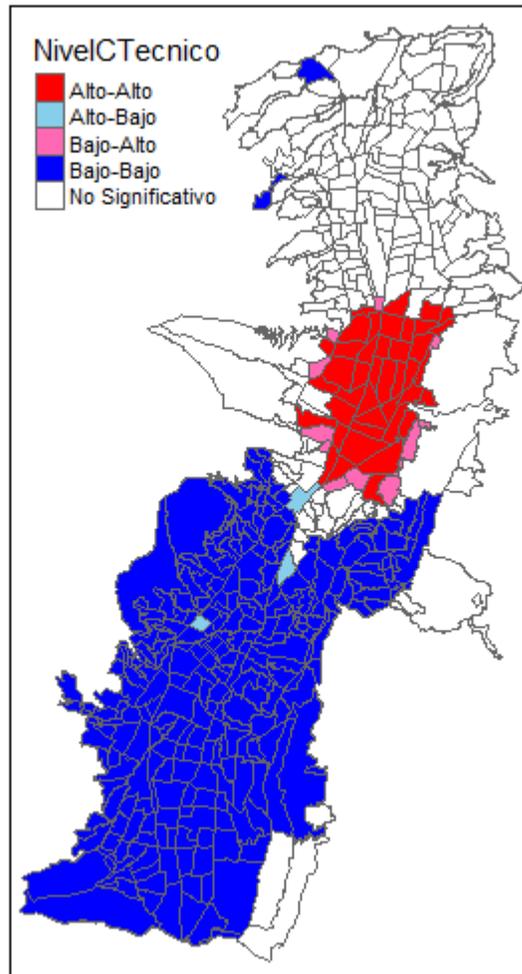


Figura C.9: Clusters por Científico Técnico

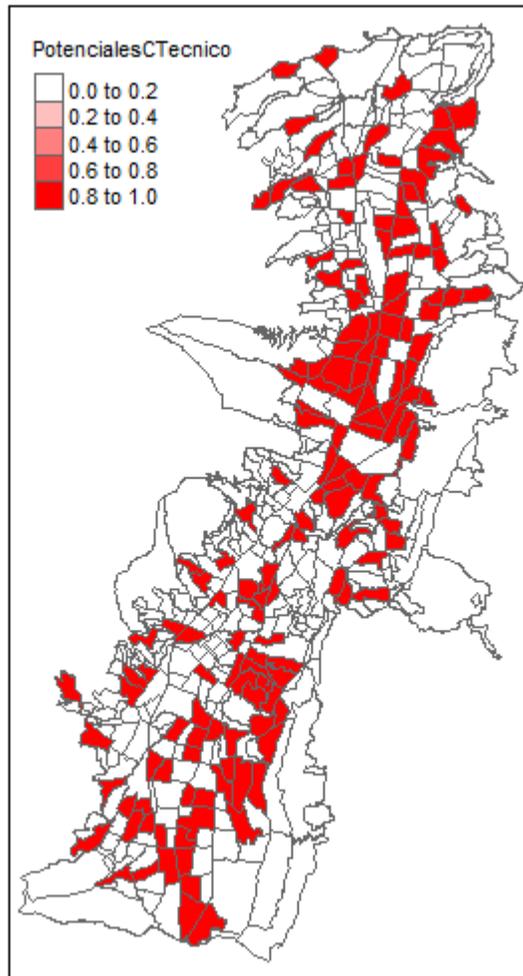


Figura C.10: Barrios con similares Características. Científico Técnico

C.3. Análisis a empresas Administrativas

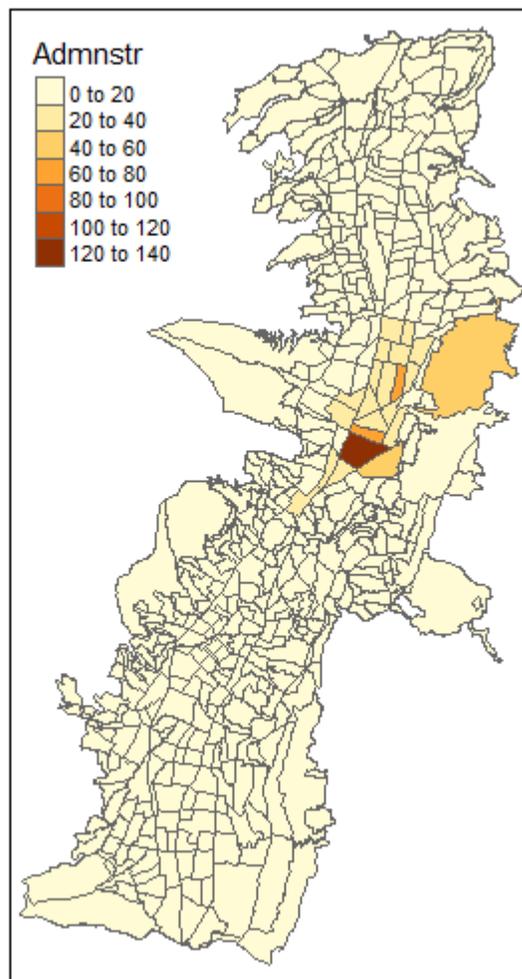


Figura C.11: Distribución de las Empresas Administrativas en los barrios de Quito

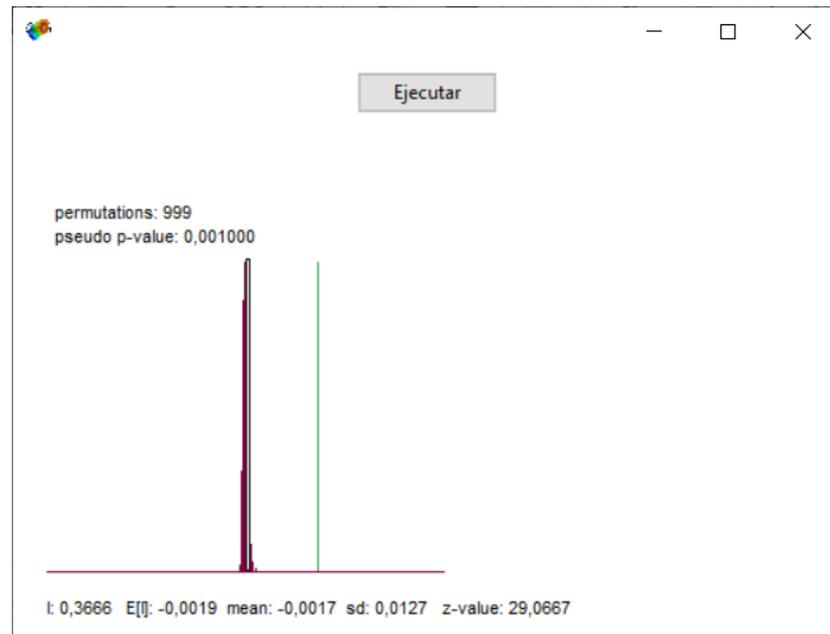


Figura C.12: Resultados GeoDa. Autocorrelación Espacial Administrativas

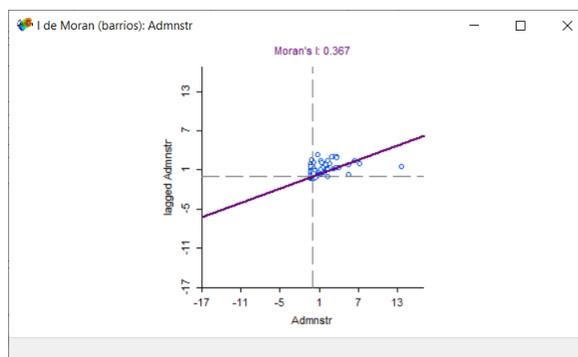


Figura C.13: Resultados GeoDa. Gráfico de Dispersión de Moran Administrativas

Alto-Alto	43
Alto-Bajo	9
Bajo-Alto	19
Bajo-Bajo	234
Nosignificativo	216

Tabla C.3: Cantidad de Barrios por Cluster. Administrativas

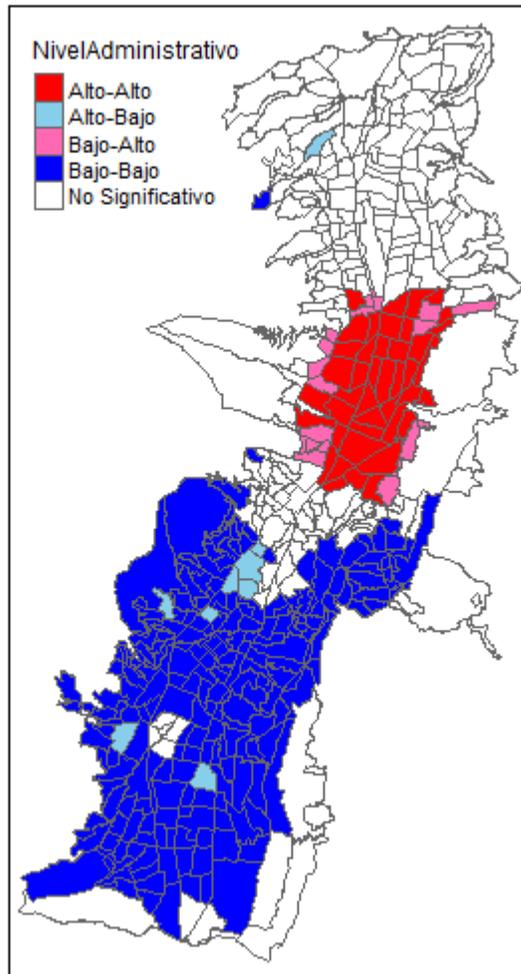


Figura C.14: Clusters por Administrativas

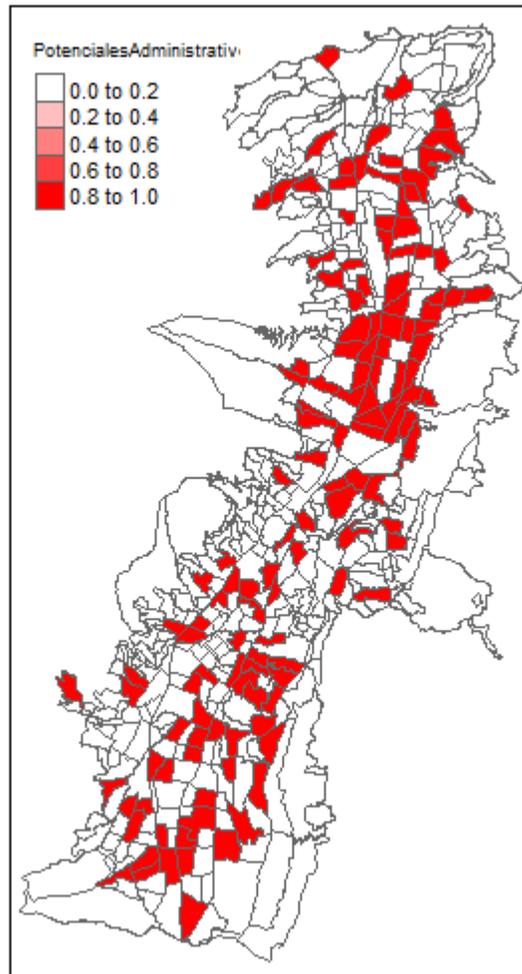


Figura C.15: Barrios con similares Características. Administrativas

C.4. Análisis a empresas de Manufacturas

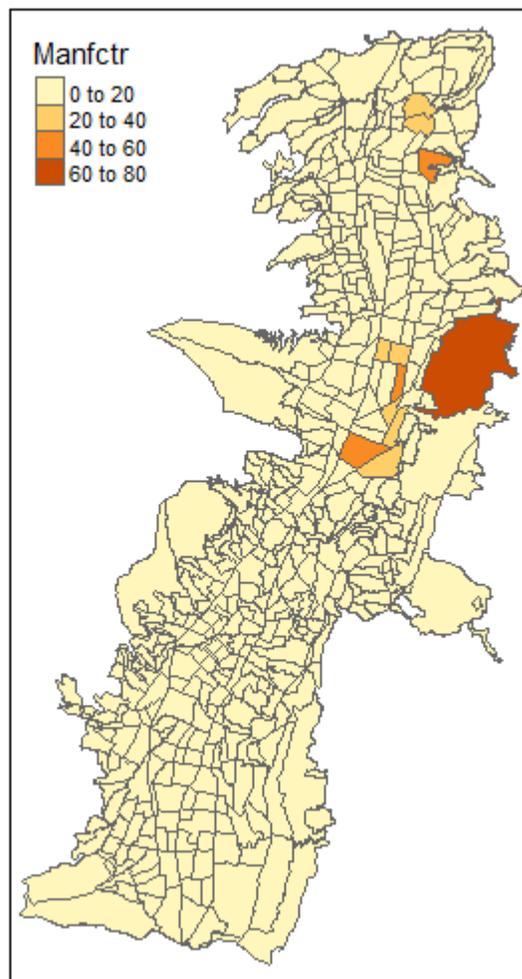


Figura C.16: Distribución de las Empresas Manufactureras en los barrios de Quito

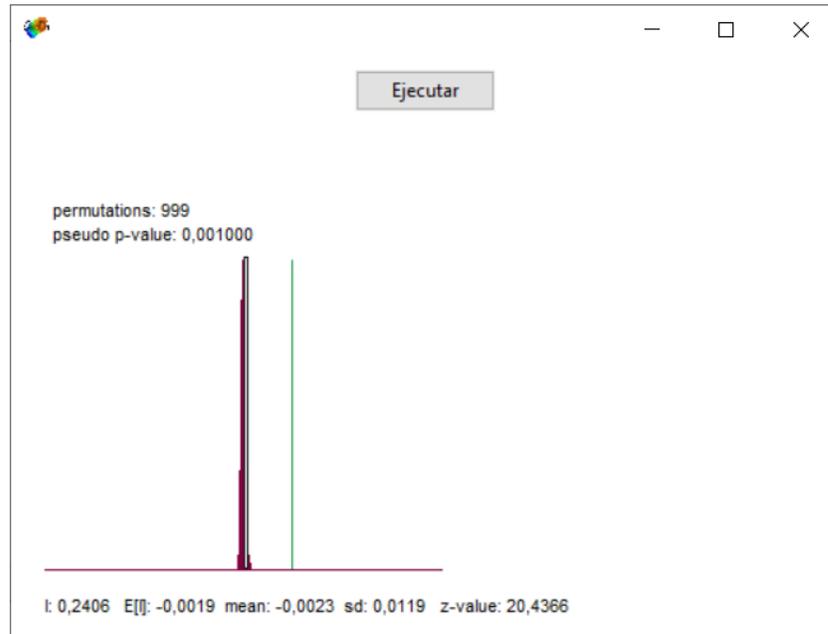


Figura C.17: Resultados GeoDa. Autocorrelación Espacial Manufacturas

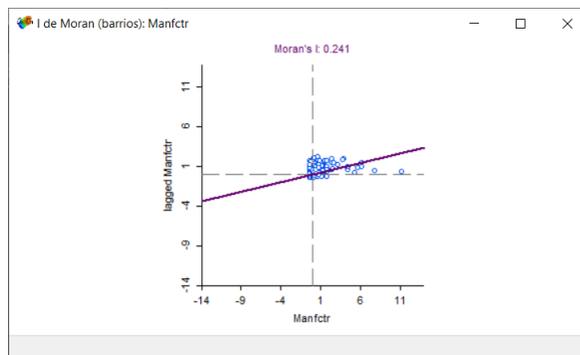


Figura C.18: Resultados GeoDa. Gráfico de Dispersión de Moran Manufacturas

Alto-Alto	58
Alto-Bajo	13
Bajo-Alto	28
Bajo-Bajo	227
Nosignificativo	195

Tabla C.4: Cantidad de Barrios por Cluster. Manufacturas

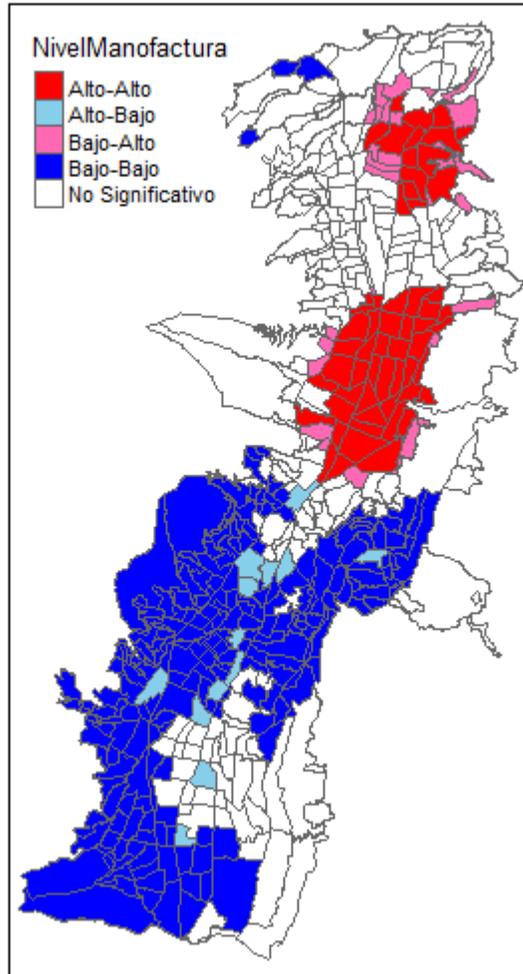


Figura C.19: Clusters por Manufacturas

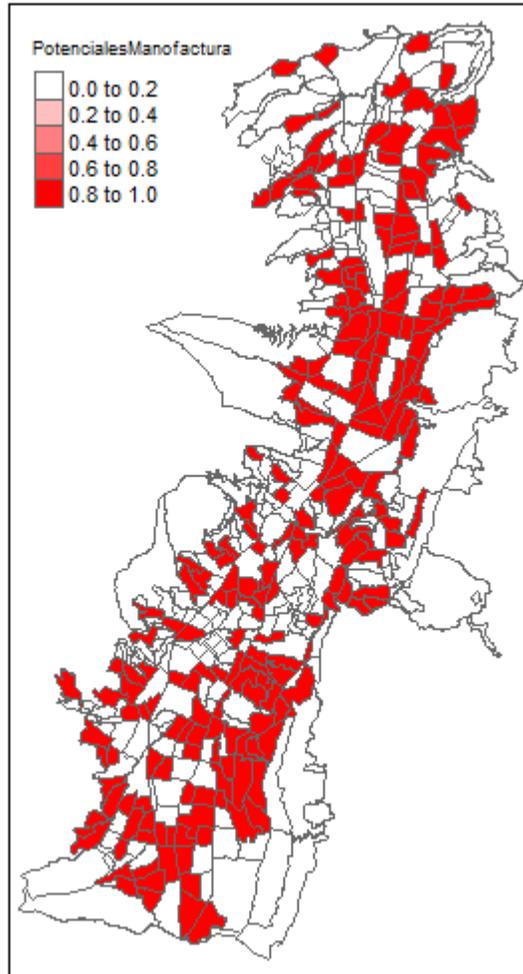


Figura C.20: Barrios con similares Características. Manufacturas

C.5. Análisis a empresas de Transporte

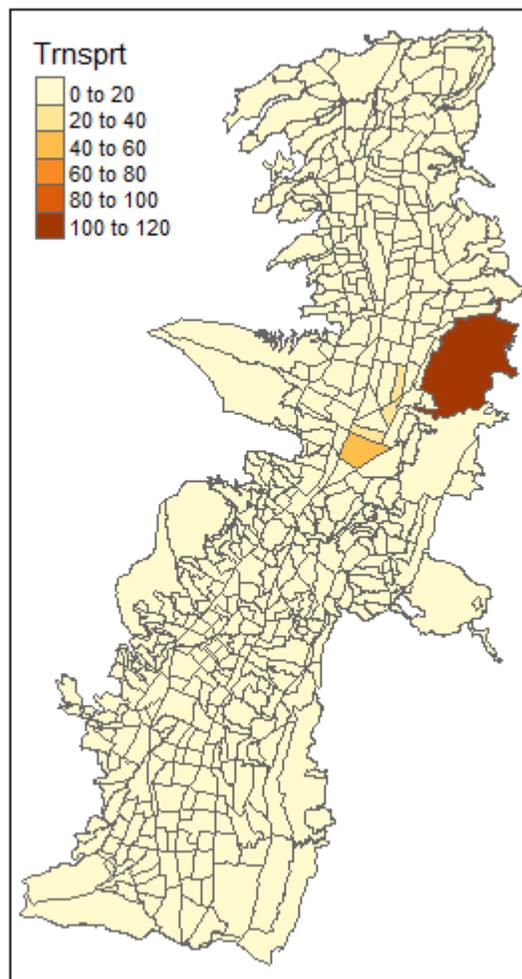


Figura C.21: Distribución de las Empresas de Transporte en los barrios de Quito

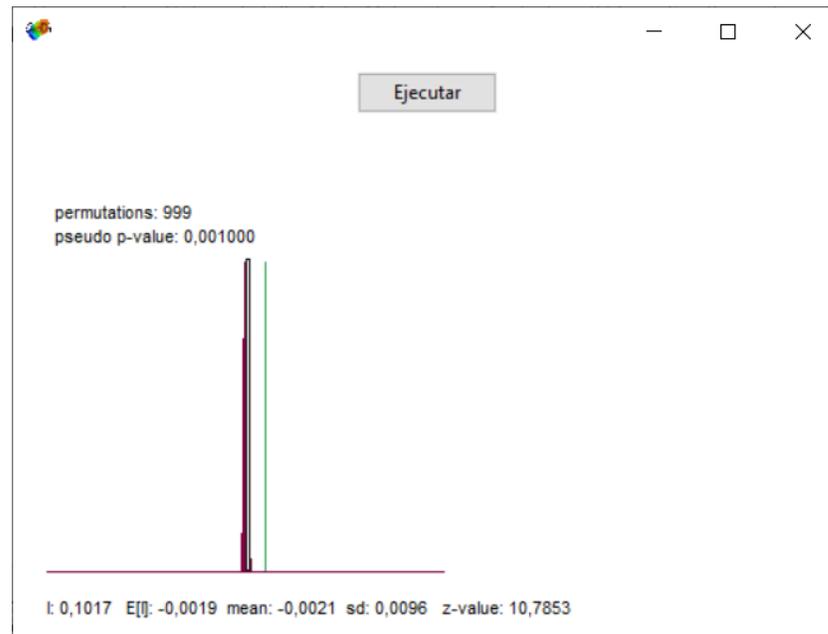


Figura C.22: Resultados GeoDa. Autocorrelación Espacial Transporte

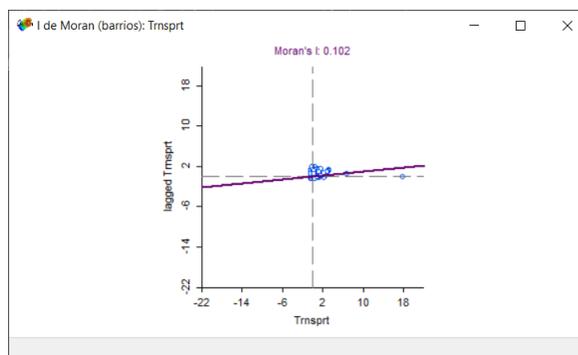


Figura C.23: Resultados GeoDa. Gráfico de Dispersión de Moran Transporte

Alto-Alto	26
Alto-Bajo	13
Bajo-Alto	13
Bajo-Bajo	131
Nosignificativo	338

Tabla C.5: Cantidad de Barrios por Cluster. Transporte

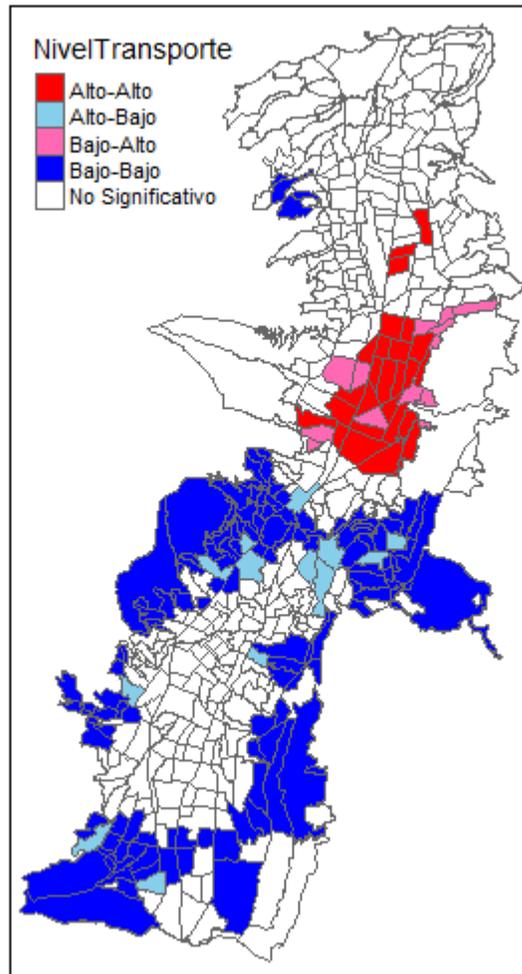


Figura C.24: Clusters por Transporte

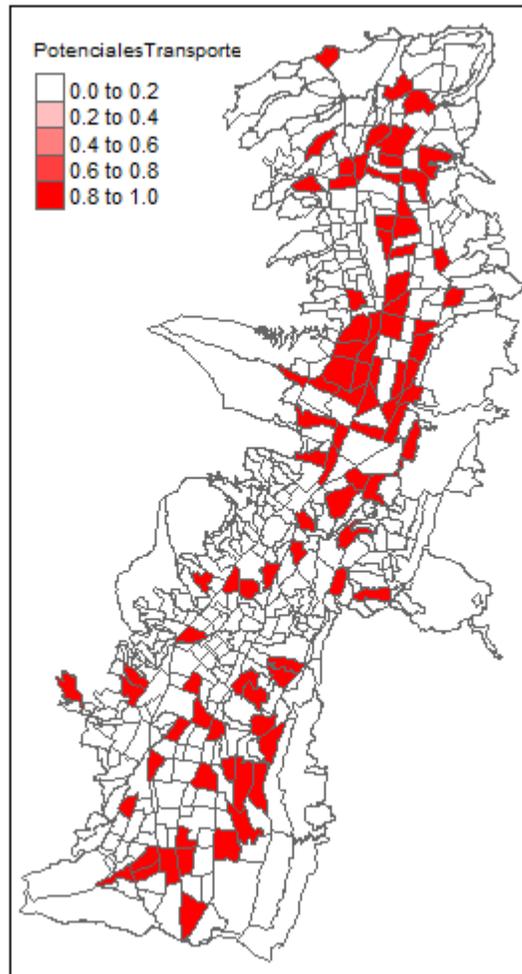


Figura C.25: Barrios con similares Características. de Transporte

C.6. Análisis a empresas de Información

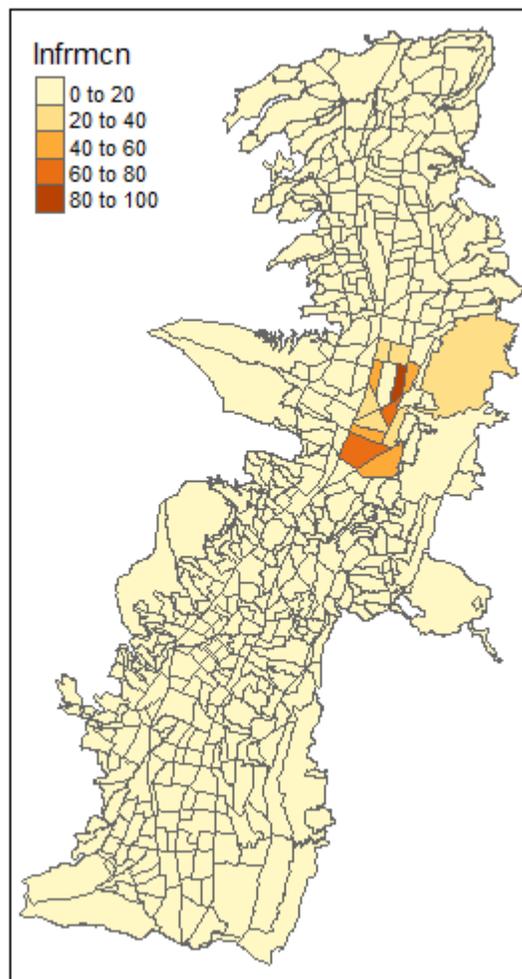


Figura C.26: Distribución de las Empresas Información en los barrios de Quito

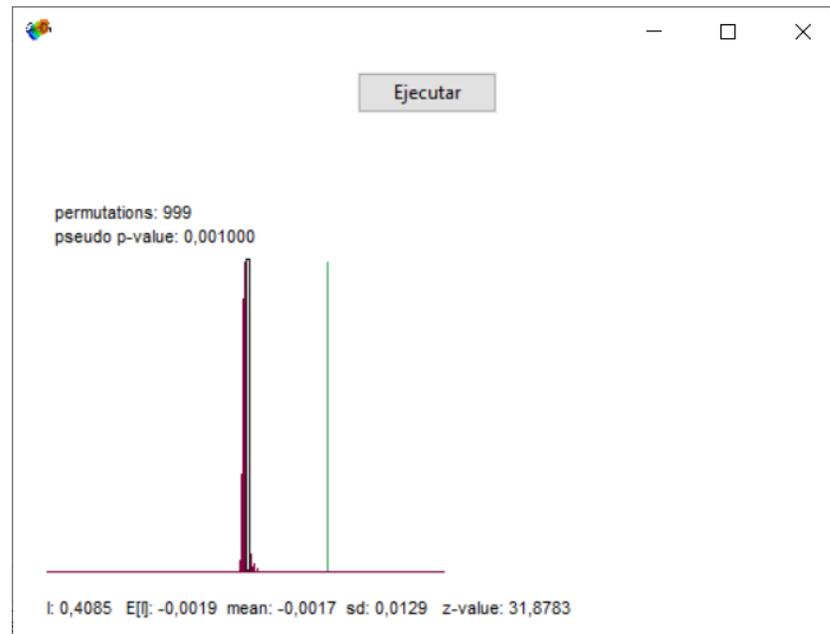


Figura C.27: Resultados GeoDa. Autocorrelación Espacial Información

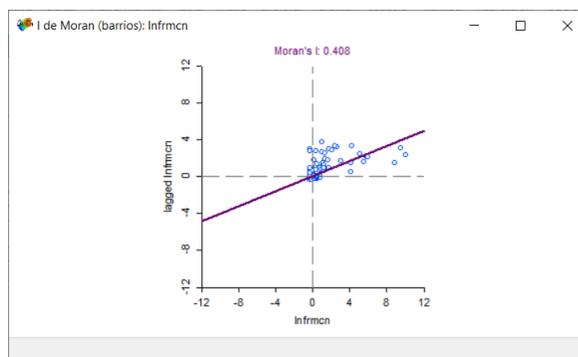


Figura C.28: Resultados GeoDa. Gráfico de Dispersión de Moran Información

Alto-Alto	40
Alto-Bajo	3
Bajo-Alto	16
Bajo-Bajo	257
Nosignificativo	205

Tabla C.6: Cantidad de Barrios por Cluster. Información

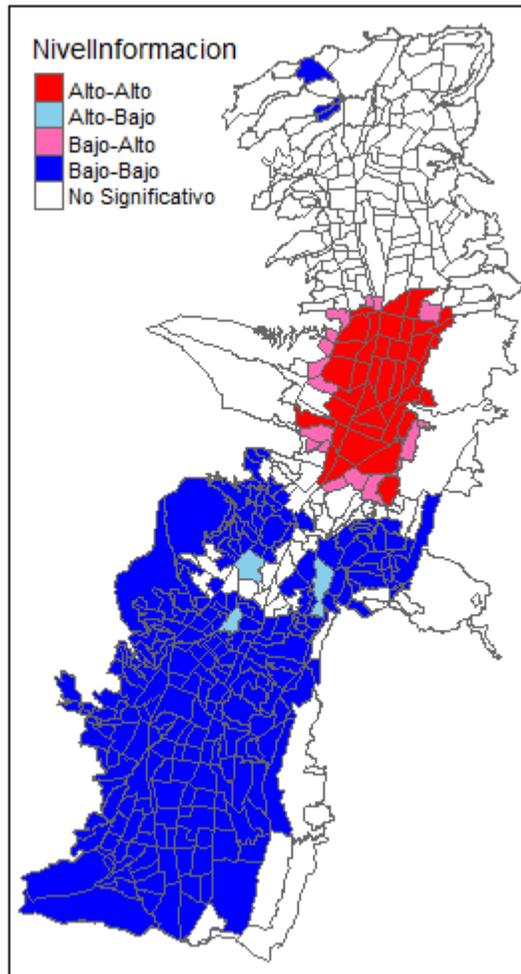


Figura C.29: Clusters por Información

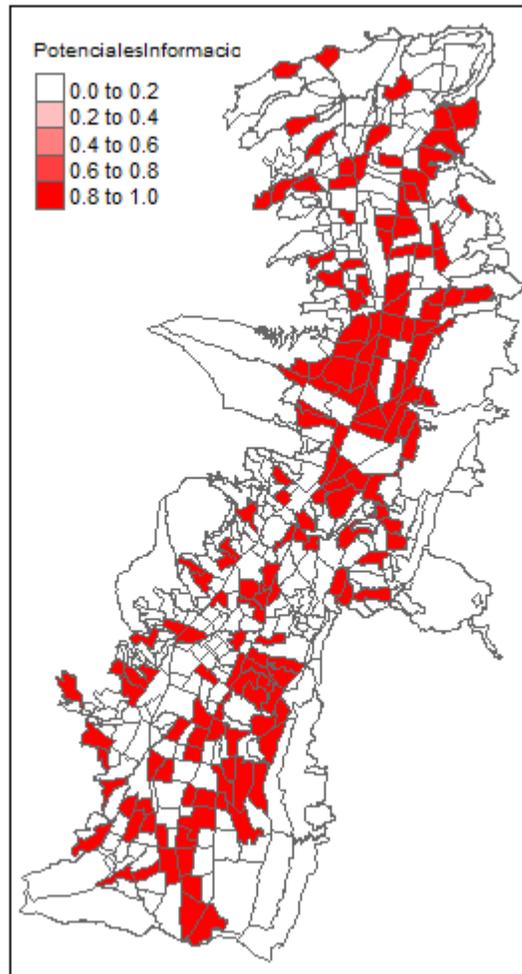


Figura C.30: Barrios con similares Características. Información

C.7. Análisis a empresas de Construcción

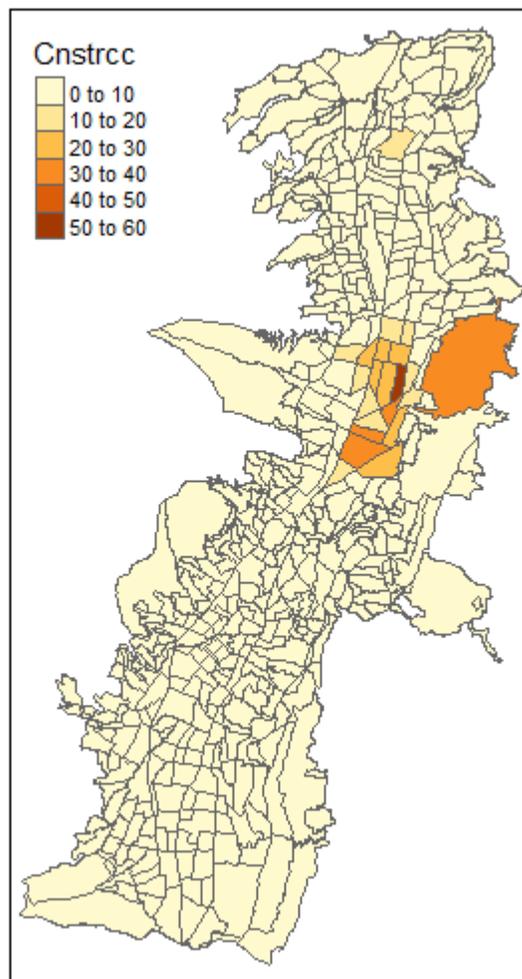


Figura C.31: Distribución de las Empresas Construcción en los barrios de Quito

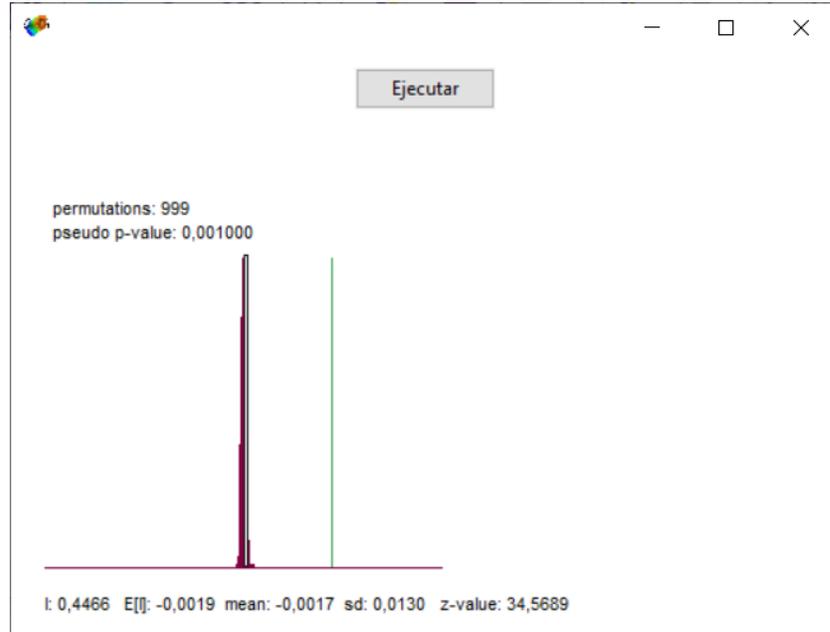


Figura C.32: Resultados GeoDa. Autocorrelación Espacial Construcción

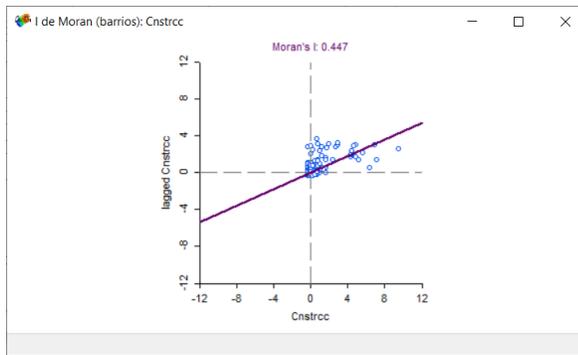


Figura C.33: Resultados GeoDa. Gráfico de Dispersión de Moran Construcción

Alto-Alto	47
Alto-Bajo	21
Bajo-Alto	12
Bajo-Bajo	254
Nosignificativo	187

Tabla C.7: Cantidad de Barrios por Cluster. Construcción

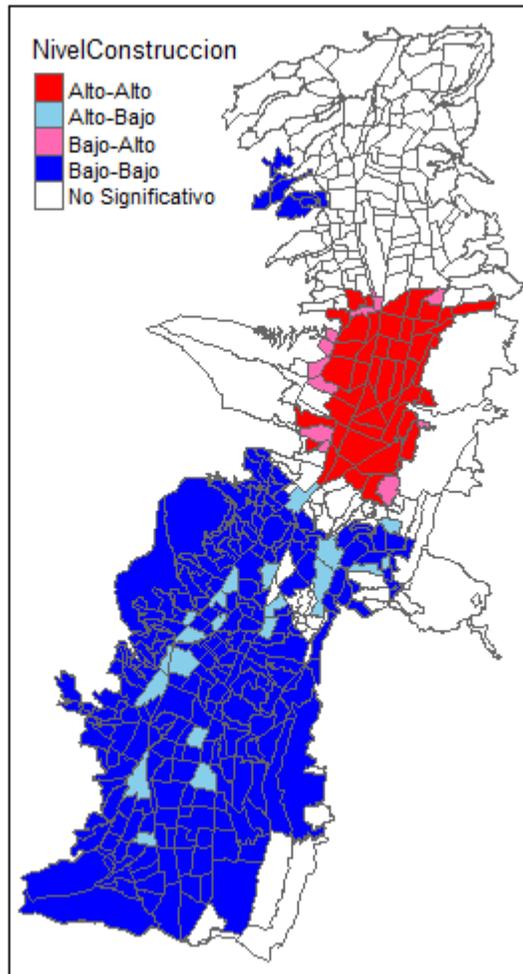


Figura C.34: Clusters por Construcción

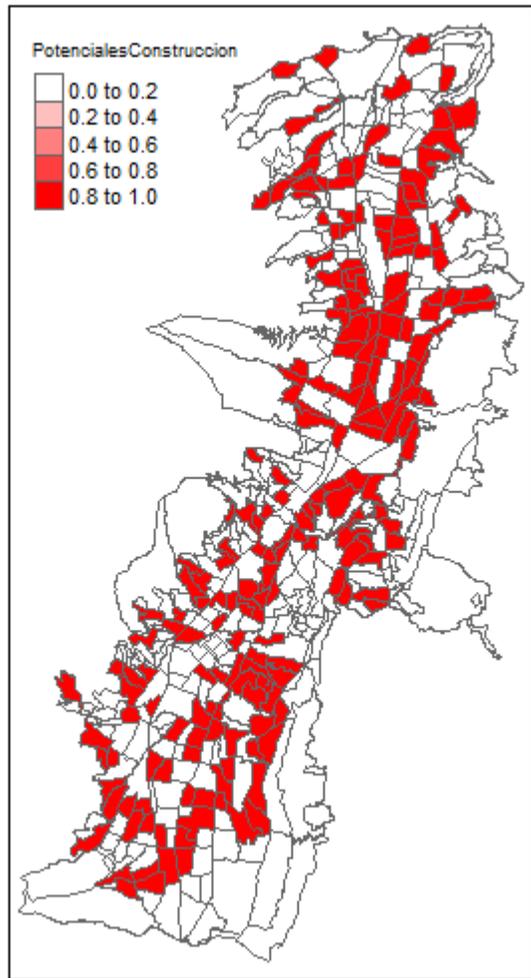


Figura C.35: Barrios con similares Características. Construcción

Apéndice D

Anexo 4. Barrios con alta densidad de Empresas

Nombre	Lat_y	Long_x	Comercio	C. Tecnico	Administrativo	Manufactura	Transporte	Informacion	Construccion
PROF MUNICIPALES	-0.157988842	-78.48988987	1	0	0	1	0	0	1
MALDONADO	-0.158789403	-78.48665459	1	0	0	1	0	0	0
JARDINES DEL BATAN	-0.171104482	-78.46897016	1	0	1	0	0	1	1
COND.S.PICHINCHA	-0.203606645	-78.5071671	0	0	0	0	0	0	1
COOP.MONSERRAT	-0.127910396	-78.47463794	0	0	0	1	0	0	0
MONGE DONOSO	-0.101089261	-78.47388272	0	0	0	1	0	0	0
EL BATAN	-0.172349751	-78.4736323	1	1	1	1	1	1	1
S.JOSE INCA	-0.158479642	-78.47369376	1	0	1	1	0	1	1
CRISTINIA 2	-0.113707347	-78.4693023	0	0	0	1	0	0	0
LIFE	-0.161449145	-78.49158758	1	1	0	1	0	1	0
DAMMER 1	-0.152272104	-78.4754178	1	0	0	0	0	0	0
ANA LUISA	-0.166312745	-78.46678925	1	1	1	1	0	1	1
UNION NACIONAL	-0.158619269	-78.49823139	1	0	0	0	0	0	0

Continúa en la página siguiente.

Nombre	Lat_y	Long_x	Comercio	C. Tecnico	Administrativo	Manufactura	Transporte	Informacion	Construccion
S.FRANCISCO NORTE	-0.105405396	-78.45972762	0	0	0	1	0	0	0
GABRIEL MARINA	-0.162665191	-78.46561913	1	1	1	1	0	1	1
C.T CONTRALORIA	-0.096370838	-78.48049894	0	0	0	1	0	0	0
LOS LAURELES	-0.156694423	-78.46868149	1	0	1	1	0	0	0
IÑAQUITO ALTO	-0.172439069	-78.50285495	0	1	0	0	0	0	0
LA ALAMEDA	-0.216217638	-78.50041027	0	0	1	0	0	0	0
BATAN ALTO	-0.179192774	-78.4730109	1	1	1	1	1	1	1
BELLAVISTA ALTO	-0.188694859	-78.47169255	0	1	1	1	0	1	1
STA.LUCIA BAJA	-0.106863876	-78.47496536	0	0	0	1	0	0	0
LA CONCEPCION	-0.153029695	-78.4913534	1	0	0	0	0	0	0
STA.LUCIA ALTA	-0.123661267	-78.47658572	0	0	0	1	0	0	0
6 DE DICIEMBRE	-0.166837679	-78.47753104	1	1	1	1	1	1	1
LA VICTORIA	-0.142259688	-78.47905148	1	0	0	0	1	0	0
CHAUPICRUZ	-0.165473421	-78.48806879	1	1	1	1	0	1	1
EL INCA	-0.16624755	-78.47246334	1	1	0	1	0	1	1
S CLARA S.MILAN	-0.200991382	-78.49891282	1	1	1	1	1	1	1
BATAN BAJO	-0.191880687	-78.4828484	1	1	1	1	1	1	1
LA CAROLINA	-0.181086107	-78.48781839	1	1	1	1	1	1	1
CRISTIANA	-0.113047404	-78.47648545	0	0	0	1	0	0	0
LUCIA ALBAN DER.	-0.126843629	-78.47889268	0	0	0	1	0	0	0
QUITO TENIS G.C	-0.163555737	-78.49910725	1	0	0	0	0	0	1
LA GRANJA	-0.184534301	-78.5041221	1	1	0	1	0	0	0
IÑAQUITO	-0.172462914	-78.48553647	1	1	1	1	1	1	1
UNION NACIONAL	-0.157531331	-78.49403409	1	0	1	0	0	0	1
VOZ DE LOS ANDES	-0.173513773	-78.4902888	1	1	1	1	1	1	1
ESTADIO ATAHUALPA	-0.181117841	-78.47667348	1	1	1	1	1	1	1
LA LUZ	-0.146934665	-78.48076191	0	0	0	0	1	0	0
AMERICA	-0.207623417	-78.50773929	1	0	0	0	0	0	0

Continúa en la página siguiente.

Nombre	Lat_y	Long_x	Comercio	C. Tecnico	Administrativo	Manufactura	Transporte	Informacion	Construccion
LARREA	-0.209477046	-78.50256746	1	1	1	1	0	1	1
LA COLON	-0.19889811	-78.49009542	1	1	1	1	1	1	1
BELLAVISTA	-0.187003761	-78.47669616	1	1	1	1	0	1	1
BENALCAZAR CD	-0.182923721	-78.48061294	1	1	1	1	1	1	1
GONZALES SUAREZ	-0.197589403	-78.47804177	1	1	1	1	1	1	1
EL INCA	-0.161258823	-78.47048831	0	1	0	1	0	0	1
BELLAVI CARRETAS	-0.100180066	-78.46673348	0	0	0	1	0	0	0
ZALDUMBIDE	-0.158980788	-78.48093764	1	1	1	1	0	1	1
MARIANA DE JESUS	-0.18768468	-78.4954798	1	1	1	1	1	1	1
JULIO MORENO D.	-0.214211062	-78.49326192	1	0	1	0	0	0	1
RUMIPAMBA	-0.18136242	-78.4923497	1	1	1	1	0	1	1
PONCIANO BAJO	-0.106123837	-78.4874002	0	0	0	1	0	0	0
EJIDO	-0.211442453	-78.49915677	1	1	1	1	0	1	1
BATAN BAJO	-0.173580564	-78.48011534	1	1	1	1	1	1	1
GRANDA CENTENO	-0.173402595	-78.4959435	1	1	1	1	0	1	1
LAS BROMELIAS	-0.160102498	-78.45653672	1	0	0	0	0	0	1
LA PRADERA	-0.195139864	-78.48876185	1	1	1	1	0	1	1
RODRIGUEZ AGUIRRE	-0.134441171	-78.47290421	0	0	0	0	1	0	0
LA VICENTINA	-0.217643491	-78.48825536	0	1	1	0	0	0	1
LA REPUBLICA	-0.189193472	-78.49032493	1	1	1	1	1	1	1
LA PAZ	-0.196971598	-78.48180199	1	1	1	1	1	1	1
SANTA LUCICIA 2	-0.21764322	-78.48344167	0	0	0	0	0	1	0
COLLALOMA 9 DE JUNIO	-0.120919859	-78.47535488	0	0	0	1	0	0	0
LA GASCA	-0.195151455	-78.50675056	1	1	1	1	1	1	1
GUAPULO	-0.204232794	-78.47800525	0	0	0	0	1	0	1
JIPIJAPA	-0.165818878	-78.48278894	1	1	1	1	1	1	1
LAS CASAS	-0.189119056	-78.50637624	1	0	1	0	0	0	0
LAS CASAS BAJO	-0.192828933	-78.49839913	1	1	1	1	1	1	1

Continúa en la página siguiente.

Nombre	Lat_y	Long_x	Comercio	C. Tecnico	Administrativo	Manufactura	Transporte	Informacion	Construccion
LAS ACACIAS	-0.153957253	-78.48114182	1	0	0	0	0	0	0
P.LA CAROLINA	-0.183401185	-78.48444254	1	1	1	1	1	1	1
TENNIS CLUB	-0.166874857	-78.49407646	1	1	1	1	0	1	1
CARRETAS	-0.107526992	-78.4663669	0	0	0	1	0	0	0
COMITE DEL PUEBLO	-0.121557955	-78.46648384	0	0	0	1	0	0	0
MARISOL	-0.107176794	-78.48143401	0	0	0	1	0	0	0
S.GABRIEL	-0.180679879	-78.49923418	1	1	1	1	0	1	1
LA FLORESTA	-0.207645872	-78.48580312	1	1	1	1	1	1	1
MARISCAL SUCRE	-0.204203098	-78.49190877	1	1	1	1	1	1	1

Tabla D.1: Barrios de Quito con Alta Densidad de Empresas

Apéndice E

Anexo 5. Análisis Espacial de las Variables

E.1. Población

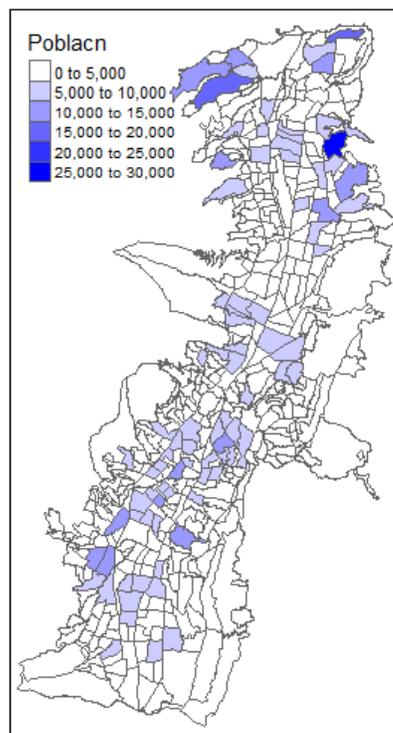


Figura E.1: Distribución de la Población

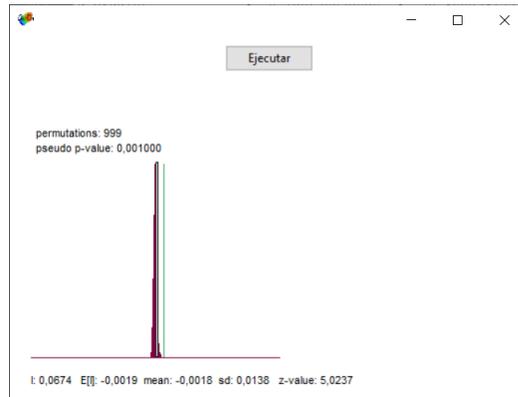


Figura E.2: Autocorrelación Espacial de la Población

E.2. Centros Educativos

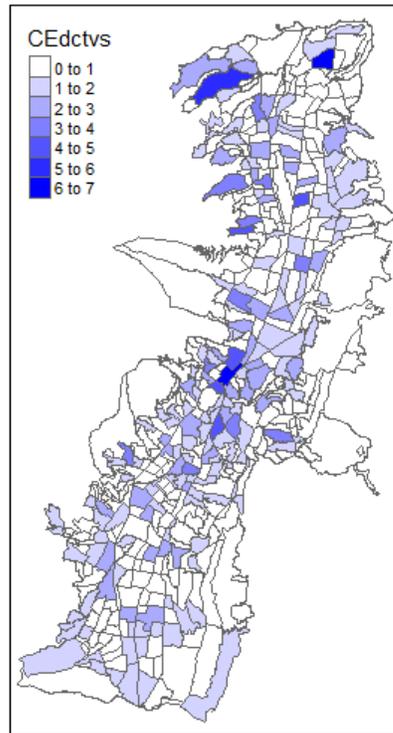


Figura E.3: Distribución de los Centros Educativos

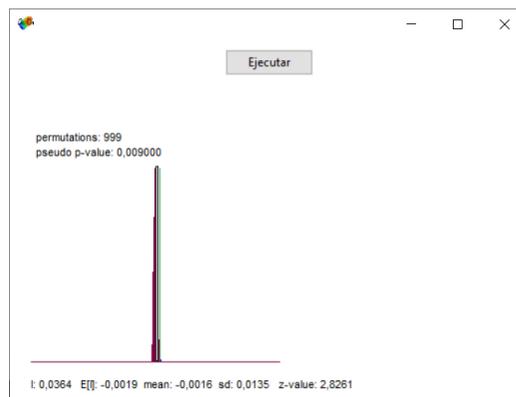


Figura E.4: Autocorrelación Espacial de los Centros Educativos

E.3. Centros de Salud

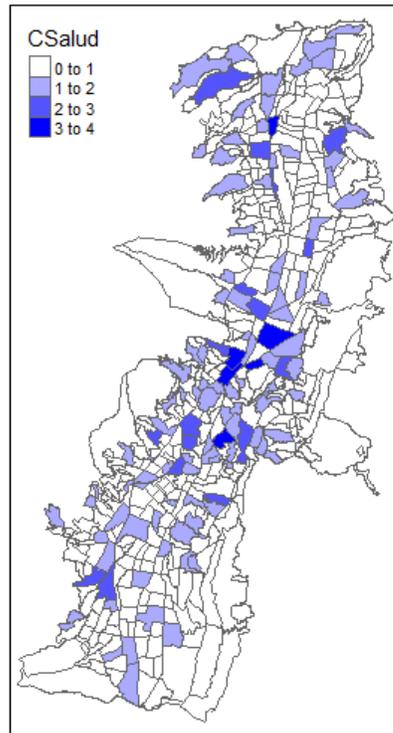


Figura E.5: Distribución de los Centros de Salud

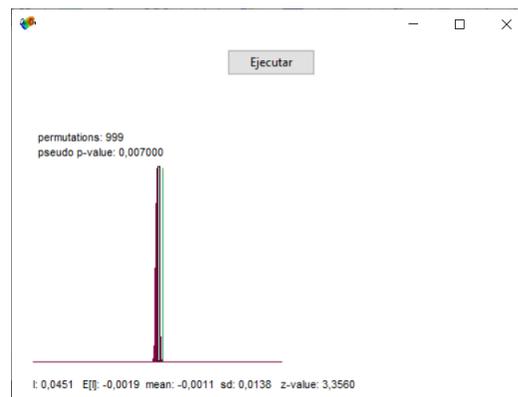


Figura E.6: Autocorrelación Espacial de los Centros de Salud

E.4. Paradas de Buses

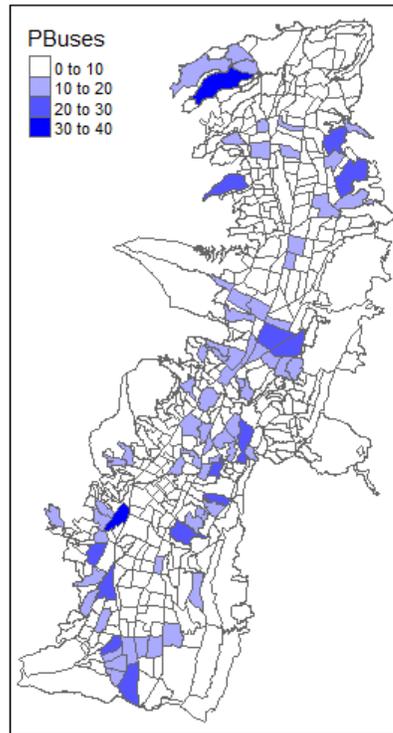


Figura E.7: Distribución de las Paradas de Buses

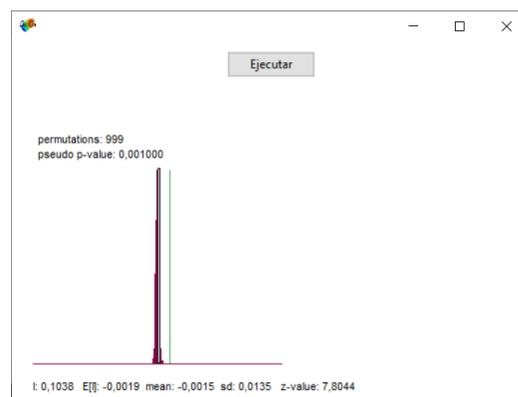


Figura E.8: Autocorrelación Espacial de las Paradas de Buses

E.5. Edificios

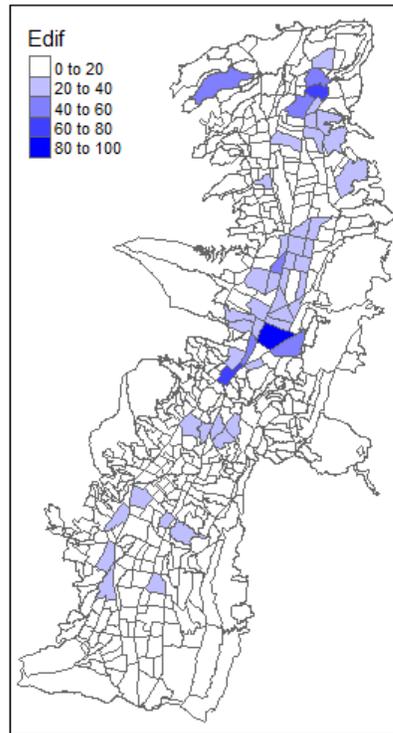


Figura E.9: Distribución de los Edificios

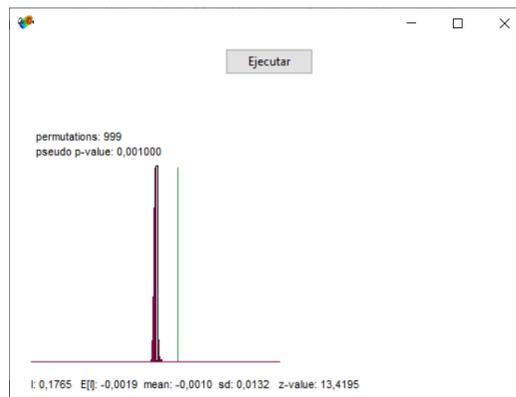


Figura E.10: Autocorrelación Espacial de los Edificios

E.6. Unidades de Policía Comunitaria

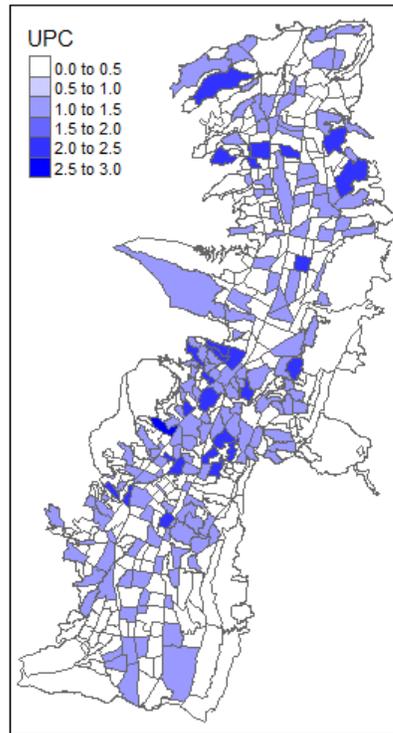


Figura E.11: Distribución de las Unidades de Policía Comunitaria

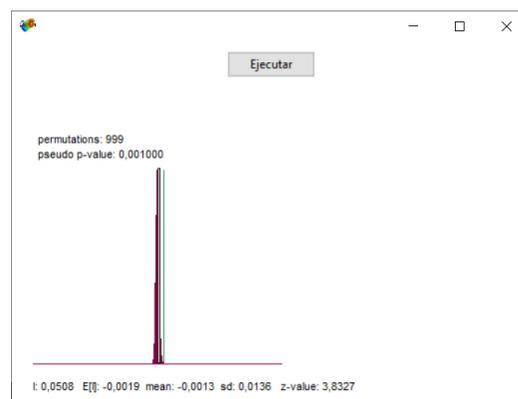


Figura E.12: Autocorrelación Espacial de las Unidades de Policía Comunitaria

E.7. Áreas Verdes

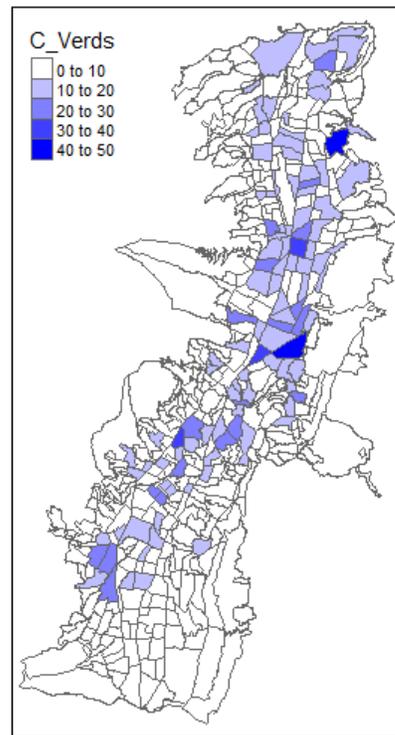


Figura E.13: Distribución de las Áreas Verdes

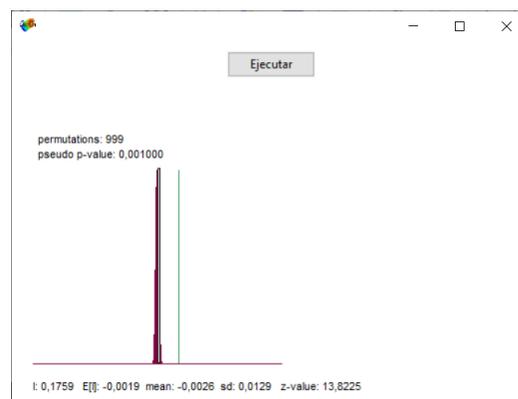


Figura E.14: Autocorrelación Espacial de las Áreas Verdes

Apéndice F

Anexo 6. Mejor Modelo por Actividad Económica

F.1. Comercio. Modelo Robusto por Variables Instrumentales de Orden Superior

$$\text{Comercio} \sim \text{CentrosEducativos} + \text{Edificios} + \text{UPC} + 1$$

Residuals:

	Min	1Q	Median	3Q	Max
	-43.422816	-2.758845	0.050823	2.756685	135.582352

Coefficients:

	Estimate	HCO	std. Error	z value	Pr(> z)
Rho	0.80160		0.10331	7.7593	8.438e-15
(Intercept)	-3.82655		0.85529	-4.4740	7.678e-06
CEdctvs	-2.04957		0.75134	-2.7279	0.006374
Edif	0.86230		0.19350	4.4563	8.339e-06
UPC	-2.52610		0.99987	-2.5264	0.011522

Residual variance (sigma squared): 138.23, (sigma: 11.757)

Impactos

Variable	Directo	Indirecto	Total
Centros Educativos	-2.2265184	-8.103858	-10.330377
Edificios	0.9367508	3.409491	4.346242
UPC	-2.7441981	-9.988056	-12.732255

Tabla F.1: Impacto de Factores Externos sobre las Empresas de Comercio

F.2. Científico Técnico. Modelo Robusto por Variables Instrumentales de Orden Superior

$$CTecnico \sim Poblacion + ParadasdeBuses + Edificios + UPC + 1$$

Residuals:

Min	1Q	Median	3Q	Max
-52.49142	-2.89444	0.76013	2.93420	178.83051

Coefficients:

	Estimate	HCO	std. Error	z value	Pr(> z)
Rho	0.80982604		0.12997825	6.2305	4.65e-10
(Intercept)	-3.41149403		1.07602222	-3.1705	0.001522
Poblacn	-0.00078241		0.00040814	-1.9170	0.055233
PBuses	0.32296556		0.16697490	1.9342	0.053087
Edif	0.76324178		0.29864935	2.5556	0.010599
UPC	-2.92309145		1.35351640	-2.1596	0.030802

Residual variance (sigma squared): 210.13, (sigma: 14.496)

Impactos

Variable	Directo	Indirecto	Total
Poblacion	-0.0008529051	-0.003261284	-0.004114189
Paradas de Buses	0.3520639625	1.346199867	1.698263829
Edificios	0.8320079853	3.181379403	4.013387388
UPC	-3.1864547788	-12.184163831	-15.370618610

Tabla F.2: Impacto de Factores Externos sobre las Empresas Científico Técnicas

F.3. Administración. Modelo Robusto por Variables Instrumentales de Orden Superior

$$\text{Administracion} \sim \text{Poblacion} + \text{ParadasdeBuses} + \text{Edificios} + \text{UPC} + 1$$

Residuals:

	Min	1Q	Median	3Q	Max
	-26.60923	-1.37029	0.41479	1.44981	70.20100

Coefficients:

	Estimate	HCO	std. Error	z value	Pr(> z)
Rho	0.73216407		0.11307962	6.4748	9.496e-11
(Intercept)	-1.68829075		0.52487338	-3.2166	0.001297
Poblacn	-0.00042656		0.00020150	-2.1169	0.034269
PBuses	0.14514609		0.07427475	1.9542	0.050680
Edif	0.41826378		0.15969164	2.6192	0.008814
UPC	-1.33685150		0.63562798	-2.1032	0.035448

Residual variance (sigma squared): 32.32, (sigma: 5.6851)

Impactos

Variable	Directo	Indirecto	Total
Poblacion	-0.0004527467	-0.001139869	-0.001592616
Paradas de Buses	0.1540567893	0.387864861	0.541921650
Edificios	0.4439415148	1.117700263	1.561641778
UPC	-1.4189224958	-3.572385087	-4.991307583

Tabla F.3: Impacto de Factores Externos sobre las Empresas de Administración

F.4. Manufactura. Modelo Robusto por Variables Instrumentales de Orden Superior

$$\text{Manufactura} \sim \text{Poblacion} + \text{CentrosEducativos} + \text{Edificios} + \text{UPC} + 1$$

Residuals:

	Min	1Q	Median	3Q	Max
	-13.518819	-1.284908	0.015475	0.833600	68.109552

Coefficients:

	Estimate	HCO	std. Error	z value	Pr(> z)
Rho	6.8784e-01		1.2467e-01	5.5172	3.445e-08
(Intercept)	-9.5429e-01		2.6709e-01	-3.5729	0.000353
Poblacn	-1.5444e-04		6.4975e-05	-2.3770	0.017456
CEdctvs	-6.2957e-01		2.5788e-01	-2.4413	0.014634
Edif	3.2099e-01		3.9995e-02	8.0258	1.110e-15
UPC	-6.1454e-01		2.9195e-01	-2.1049	0.035297

Residual variance (sigma squared): 21.78, (sigma: 4.6669)

Impactos

Variable	Directo	Indirecto	Total
Poblacion	-0.0001621433	-0.0003326078	-0.000494751
Centros Educativos	-0.6609670853	-1.3558550104	-2.016822096
Edificios	0.3369972257	0.6912891537	1.028286379
UPC	-0.6451868668	-1.3234847324	-1.968671599

Tabla F.4: Impacto de Factores Externos sobre las Empresas de Manufactura

F.5. Transporte. Modelo Robusto por Variables Instrumentales de Orden Superior

$$\text{Transporte} \sim \text{Edificios} + 1$$

Residuals:

Min	1Q	Median	3Q	Max
-10.14494	-1.22501	-0.36830	0.52309	108.01010

Coefficients:

	Estimate	HCO	std. Error	z value	Pr(> z)
Rho	0.684470		0.229877	2.9776	0.0029056
(Intercept)	-0.795575		0.296854	-2.6800	0.0073617
Edif	0.185807		0.048924	3.7979	0.0001459

Residual variance (sigma squared): 29.497, (sigma: 5.4311)

Impactos

Variable	Directo	Indirecto	Total
Edificios	0.1949283	0.3939457	0.588874

Tabla F.5: Impacto de Factores Externos sobre las Empresas de Transporte

F.6. Información. Modelo Robusto por Variables Instrumentales de Orden Superior

$$\text{Informacion} \sim \text{Edificios} + \text{UPC} + 1$$

Residuals:

	Min	1Q	Median	3Q	Max
	-19.04159	-1.01590	0.24333	0.98758	57.00969

Coefficients:

	Estimate	HC0 std. Error	z value	Pr(> z)
Rho	0.826645	0.146426	5.6455	1.647e-08
(Intercept)	-1.316524	0.425845	-3.0916	0.001991
Edif	0.258643	0.087537	2.9547	0.003130
UPC	-1.333262	0.544274	-2.4496	0.014301

Residual variance (sigma squared): 29.993, (sigma: 5.4766)

Impactos

Variable	Directo	Indirecto	Total
Edificios	0.2841258	1.207862	1.491988
UPC	-1.4646228	-6.226335	-7.690958

Tabla F.6: Impacto de Factores Externos sobre las Empresas de Información

F.7. Construcción. Modelo Robusto por Variables Instrumentales de Orden Superior

$$\text{Construccion} \sim \text{Poblacion} + \text{ParadasdeBuses} + \text{Edificios} + \text{UPC} + 1$$

Residuals:

Min	1Q	Median	3Q	Max
-13.12312	-0.84239	0.12466	0.66450	32.53399

Coefficients:

	Estimate	HCO	std. Error	z value	Pr(> z)
Rho	8.6322e-01		1.2379e-01	6.9733	3.096e-12
(Intercept)	-8.0750e-01		1.9500e-01	-4.1410	3.458e-05
Poblacn	-1.4228e-04		7.2856e-05	-1.9530	0.0508250
PBuses	7.4377e-02		3.4280e-02	2.1697	0.0300279
Edif	1.6080e-01		4.5717e-02	3.5172	0.0004361
UPC	-6.2186e-01		2.7170e-01	-2.2888	0.0220904

Residual variance (sigma squared): 10.956, (sigma: 3.31)

Impactos

Variable	Directo	Indirecto	Total
Poblacion	-0.0001595532	-0.0008806706	-0.001040224
Paradas de Buses	0.0834046075	0.4603603105	0.543764918
Edificios	0.1803128169	0.9952551410	1.175567958
UPC	-0.6973411522	-3.8490462221	-4.546387374

Tabla F.7: Impacto de Factores Externos sobre las Empresas de Construcción