

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

# Análisis epidemiológico-espacial de los determinantes de la neumonía en Bogotá (2004-2014)

**David Enrique Payares García**
**Norgith Bibiana Quintero Alonso**

Universidad Distrital Francisco José Caldas
Facultad de Ingeniería, Ingeniería Catastral y Geodesia
Bogotá, Colombia
2018

# Ánalisis epidemiológico-espacial de los determinantes de la neumonía en Bogotá (2004-2014)

**David Enrique Payares García**
**Norgith Bibiana Quintero Alonso**

Tesis o trabajo de grado presentado como requisito parcial para optar al título de:
**Ingeniero Catastral & Geodesta**

Director:
PhD. Carlos Eduardo Melo Martinez

Línea de investigación:
Estadística Espacial

Universidad Distrital Francisco José de Caldas
Facultad de Ingeniería, Ingeniería Catastral & Geodesia
Bogotá D.C., Colombia
2018

A nuestras madres

# Agradecimientos

Nos gustaría agradecer especialmente al director de este trabajo, Carlos Eduardo Melo Martínez, por la paciente orientación, apoyo y consejos que nos brindó durante el tiempo como sus estudiantes y como autores del presente proyecto. Hemos sido increíblemente afortunados de tener un profesor que se preocupara tanto por nuestro trabajo, y que respondiera a todas nuestras preguntas e inquietudes tan prontamente. Toda la pasión, curiosidad y entusiasmo que tiene por la estadística espacial han sido contagiosos y motivadores para nosotros.

Queremos darle un reconocimiento a la Secretaría de Salud Distrital por permitirnos adquirir y procesar la información aquí palpada, sin su pronta respuesta y diligencia, no existiría la presente investigación. De la misma manera, a nuestra Universidad Francisco José de Caldas y nuestro proyecto curricular Ingeniería Catastral y Geodesia, por mostrarnos académica, profesional y éticamente el camino que debemos seguir.

Expresamos nuestra gratitud a nuestras familias por todo su amor y aliento. A nuestras madres que nos educaron con amor a la ciencia y fueron nuestro apoyo en todas nuestras búsquedas; su integridad y modestia nos ha impulsado a ser mejores personas día a día.

Finalmente, estamos inmensamente agradecidos con nosotros mismos; sin el amor, ánimo, apoyo y paciencia que nos hemos brindado el uno al otro probablemente esta etapa de nuestras vidas no hubiera sido tan alegre y satisfactoria. El presente documento expresa el respeto y admiración que sentimos mutuamente.

# Contenido


**1. INTRODUCCIÓN** **1**

**2. PLANTEAMIENTO DEL PROBLEMA** **2**
2.1. Formulación del Problema . . . . . 2
2.2. Justificación . . . . . 3

**3. OBJETIVOS** **4**
3.1. Objetivo General . . . . . 4
3.2. Objetivos específicos . . . . . 4

**4. ANTECEDENTES Y MARCO REFERENCIAL** **5**
4.1. Epidemiología . . . . . 6
4.2. Neumonía . . . . . 6
4.2.1. Epidemiología . . . . . 7
4.2.2. Factores de riesgo . . . . . 7
4.3. Estadística Espacial: Datos de Área . . . . . 8
4.3.1. Econometría Espacial . . . . . 8
4.3.2. Autocorrelación espacial . . . . . 8
4.3.3. Heterogeneidad espacial . . . . . 9
4.3.4. Análisis Exploratorio de Datos Espaciales . . . . . 9
4.3.5. Matriz de Pesos Espaciales . . . . . 9
4.3.6. Contrastes de Autocorrelación Espacial . . . . . 10
4.3.7. Análisis de Cooordenadas Principales de Matrices de Pesos Espaciales(PCNM) . . . . . 13
4.3.8. Modelo de Regresión Espacial . . . . . 13
4.3.9. Modelos Espaciales Alternativos . . . . . 15
4.3.10. Regresión geográficamente ponderada (GWR) . . . . . 16

**5. METODOLOGÍA** **18**
5.1. Área de Estudio . . . . . 18
5.2. Materiales . . . . . 19
5.3. Métodos . . . . . 26
5.3.1. Solicitud de Datos . . . . . 28
5.3.2. Selección de la enfermedad . . . . . 28
5.3.3. Filtrado de la Información . . . . . 28
5.3.4. Búsqueda de Variables . . . . . 29
5.3.5. Análisis Exploratorio de Datos . . . . . 29
5.3.6. Construcción Modelo Autorregresivo Espacial . . . . . 32
5.3.7. Selección Modelo Autorregresivo Espacial . . . . . 33
5.3.8. Validación Modelo Autorregresivo Espacial . . . . . 34
5.3.9. Interpretación del Modelo . . . . . 34
5.3.10. Generación de Impactos . . . . . 34

5.3.11. Diagnóstico de la Neumonia . . . . . 35

**6. RESULTADOS** **36**
6.1. Ánálisis Exploratorio Espacial (AEDE) . . . . . 36
6.1.1. Selección de Matriz de Pesos Espaciales . . . . . 41
6.1.2. Ánálisis Exploratorio Global . . . . . 44
6.1.3. Ánálisis Exploratorio Local . . . . . 47
6.1.4. Análisis Bivariado . . . . . 51
6.2. Modelo Regresivo Clásico . . . . . 53
6.3. Modelo Regresivo Espacial . . . . . 61
6.4. Impactos . . . . . 69
6.5. Regresión Geográficamente Ponderada . . . . . 72

**7. DISCUSIÓN** **83**

**8. CONCLUSIONES Y RECOMENDACIONES** **86**

**A. Anexo: Datos para el cálculo de la SMR por neumonía** **87**

**B. Anexo: Correlogramas bivariados** **89**

**BIBLIOGRAFÍA** **91**

# Lista de Figuras


1. Ubicación espacial de la región de estudio. . . . . 18
2. Diagrama de flujo de los métodos a utilizar . . . . . 27
3. Distribución espacial de la SMR por neumonía en las localidades de Bogotá: 2004 - 2007. . . . . 36
4. Distribución espacial de la SMR por neumonia en las localidades de Bogotá: 2011 - 2014. . . . . 37
5. Histograma, Gráfico Q-Q y Boxplot de la SMR: 2004. . . . . 38
6. Histograma, Gráfico Q-Q y Boxplot de la SMR: 2007. . . . . 39
7. Histograma, Gráfico Q-Q y Boxplot de la SMR: 2011. . . . . 39
8. Histograma, Gráfico Q-Q y Boxplot de la SMR: 2014. . . . . 40
9. Gráficos criterios seleccionados por año . . . . . 43
10. Correlogramas . . . . . 44
11. Mapa cluster LISA . . . . . 48
12. Mapa Significancia LISA . . . . . 49
13. Getis Cluster local . . . . . 50
14. Getis Significancia local . . . . . 51
15. Distribución espacial de los R-cuadrados locales: GWR 2004 . . . . . 73
16. Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2004: Asistencia Escolar y Calidad de Vida. . . . . 74
17. Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2004: Hogares donde no se consumió ninguna de las tres comidas diarias . . . . . 75
18. Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2004: Cobertura de vacunación con anti Haemophilus influenzae tipo B y Temperatura Ambiente. . . . . 75
19. Distribución espacial de los R-cuadrados locales: GWR 2007 . . . . . 76
20. Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2007: Asistencia Escolar e índice de prácticas deportivas. 77
21. Distribución espacial de los R-cuadrados locales: GWR 2011 . . . . . 78
22. Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2011: Calidad de Vida y número de IPS. . . . . 79
23. Distribución espacial de los R-cuadrados locales: GWR 2014 . . . . . 80
24. Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2014: Número de IPS y calidad de la vivienda. . . . . 81
25. Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2014: Cobertura de vacunación y Temperatura Ambiente. 81
26. Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2014: Contaminación por PM10 y Hogares que no consumieron las tres comidas diarias. . . . . 82
27. Moran bivariado 2004 . . . . . 89
28. Moran bivariado 2007 . . . . . 89
29. Moran bivariado 2011 . . . . . 90

30. Moran bivariado 2014 . . . . . 90

# Lista de Tablas


1. Selección variable respuesta. . . . . 20
2. Variables explicativas seleccionadas para el estudio. . . . . 21
3. Resumen de medidas estadísticas de la variable SMR para cada año. . 41
4. Criterios de Selección de matriz W 2004-2007 . . . . . 42
5. Criterios de Selección de matriz W 2011-2014 . . . . . 42
6. Contrastes globales de variables endógena y exógenas año 2004 . . . . 45
7. Contrastes globales de variables endógena y exógenas año 2007 . . . . 46
8. Contrastes globales de variables endógena y exógenas año 2011 . . . . 47
9. Contrastes globales de variables endógena y exógenas año 2014 . . . . 47
10. Moran bivariado 2004-2007 . . . . . 52
11. Moran bivariado 2011 . . . . . 52
12. Estimación modelo de regresión clásico para el año 2004. . . . . 54
13. Supuestos del modelo regresivo clásico: 2004. . . . . 55
14. Supuesto de multicolinealidad (Inflación de la varianza): 2004. . . . . 55
15. Estimación modelo de regresión clásico para el año 2007. . . . . 56
16. Supuestos del modelo regresivo clásico: 2007. . . . . 57
17. Supuesto de multicolinealidad (Inflación de la varianza): 2007. . . . . 57
18. Estimación modelo de regresión clásico para el año 2011. . . . . 58
19. Supuestos del modelo regresivo clásico: 2011. . . . . 58
20. Supuesto de multicolinealidad (Inflación de la varianza): 2011. . . . . 59
21. Estimación modelo de regresión clásico para el año 2014. . . . . 59
22. Supuestos del modelo regresivo clásico: 2014. . . . . 60
23. Supuesto de multicolinealidad (Inflación de la varianza): 2014. . . . . 60
24. Test Global de I de Moral para los residuales de una regresión. . . . . 62
25. Diagnóstico para dependencia espacial: Multiplicadores de Lagrange. 62
26. Comparación de Modelos espaciales: Criterio de Akaike . . . . . 63
27. Estimación modelo de regresión espacial para el año 2004: SARAR . . 65
28. Estimación modelo de regresión espacial para el año 2007: GNS . . . 66
29. Estimación modelo de regresión espacial para el año 2011: SAR . . . 67
30. Estimación modelo de regresión espacial para el año 2014: GNS . . . 69
31. Impactos: 2004 . . . . . 70
32. Impactos: 2007 . . . . . 70
33. Impactos: 2011 . . . . . 71
34. Impactos: 2014 . . . . . 71
35. Estimación regresión geográficamente ponderada: 2004 . . . . . 73
36. Estimación Regresión geográficamente ponderada: 2007 . . . . . 76
37. Estimación Regresión geográficamente ponderada: 2011 . . . . . 77
38. Estimación regresión geográficamente ponderada: 2014 . . . . . 79
39. Datos para hallar SMR 2004-2007 . . . . . 87
40. Datos para hallar SMR 2011-2014 . . . . . 88

# 1. INTRODUCCIÓN

Se han propuesto diferentes conceptos al término de salud a lo largo del tiempo, entre estos se encuentran "ausencia de enfermedad" y "estado de completo bienestar físico, mental y social" formulado por la Organización Mundial de la Salud. Por esto, al existir una enfermedad que altera las condiciones de bienestar de los individuos de una zona determinada produce el surgimiento de dificultades en salud pública que autoridades gubernamentales y profesionales expertos en dichos temas no pueden ignorar, ya que su principal función es la de efectuar control y prevención de las enfermedades (Quah, 2017).

La epidemiología espacial es una herramienta que permite el análisis de la distribución espacial y variación geográfica de enfermedades con el fin de describir, explicar y cuantificar estas variaciones con relación a diferentes factores de riesgo, para entender el comportamiento de una enfermedad y sus procesos de ocurrencia en una región y tiempo específico. Las poblaciones vulnerables deben ser concebidas en sus ambientes sociales, económicos y culturales debido a que estos elementos influyen en la propagación de enfermedades y afectan las condiciones de salud. Los estudios epidemiológicos que tienen en cuenta el componente espacio temporal ayudan a comprender por qué las distribuciones de una enfermedad no son uniformes, las causas que cambian la incidencia de la enfermedad y cómo pueden predecirse estos cambios (Waller y Gotway, 2004).

Por consiguiente, evaluar el impacto de una enfermedad es fundamental para realizar programas y políticas que busquen intervenir a la población con mayor índice de vulnerabilidad y observar los efectos positivos y negativos que obtengan en la salubridad de los individuos. Así mismo, al identificar las consecuencias con respecto a los factores de riesgo posibilita la intervención en el territorio y mejor toma de decisiones. Para el caso de estudio en Bogotá, se propone analizar por medio del análisis de áreas o lattices las muertes por neumonía. Se desarrollará un modelo de autorregresión espacial que modele la variabilidad de los datos y que identifique los principales factores que determinan la neumonia y sus muertes en las diferentes localidades del territorio distritalino.

Se realizará un estudio retrospectivo de las muertes por neumonía en 19 de las 20 localidades de Bogotá, Colombia en los años 2004 - 2007 - 2010 y 2014, identificando la dependencia espacial entre los datos. Los modelos de regresión espacial se estimarán usando una serie de predictores -o variables –ambientales, sociales, económicos, conductuales y sanitarios. Se espera que los resultados ilustren cómo cada variable y su relación espacial con las demás inciden de manera positiva o negativa en la tasa de mortalidad por neumonía en Bogotá y cómo éstas pueden identificar las necesidades de la población afectada con el fin de desarrollar políticas beneficiosas para la comunidad teniendo en cuenta estos factores.

# 2. PLANTEAMIENTO DEL PROBLEMA

## 2.1. Formulación del Problema

Según la Secretaría de Salud Distrital las enfermedades respiratorias ocupan el tercer lugar en provocar la muerte de la población bogotana, después de las enfermedades isquémicas del corazón y las enfermedades cerebrovasculares. De estas enfermedades respiratorias, la neumonía es la enfermedad pionera en casos de hospitalización y aporta la mayor mortalidad del grupo especialmente en los menores de un año y adultos mayores. Debido a las constantes aglomeraciones, el número creciente de establecimientos manufactureros y automotores, los extenuantes horarios de trabajo, el sedentarismo y los demás factores que son frecuentes en una capital como la Bogotana, la neumonía adquiere especial acogida entre los habitantes. Socialmente esta situación se ve reflejada en el aumento anual de más del 5 % de muertes por neumonía y el estado de alerta de los organismos de control sanitarios. El crecimiento exponencial, en materia económica, social y poblacional de la ciudad de Bogotá, ha causado que las medidas de control que existen para mitigar el impacto de la mortalidad por neumonía sean opacadas por el incremento en la ocurrencia de dicha enfermedad. El desarrollo de nuevas metodologías que permitan identificar los determinantes que llevan a la muerte por esta enfermedad, resulta necesario en lo que concierne a posibilitar a las autoridades sanitarias información sobre los factores inmersos en la mortalidad por neumonía en el distrito capital colombiano en un periodo discreto de diez años (2004-2014) con el fin de adoptar estrategias de gobierno que deriven acciones favorables para la salud de la población y que moderen la actuación de los factores conforme la ciudad crece.

Investigaciones previas sobre los determinantes de las muertes por neumonía en Bogotá se han centrado en el actuar de los factores de riesgo conductuales y biológicos de los individuos, arrojando explicaciones parciales y enfoques en su mayoría sobre medidas para reducir el impacto de la enfermedad en mortalidad de la población (Taboada, Leal, Caicedo, Camargo, y Roa, 2015). Aun así, estas investigaciones han subestimado una de las interrelaciones más importantes de los determinantes; la dependencia espacial.

Debido a la constante variabilidad de la enfermedad y sus factores y al gran número de ocurrencias de mortalidad dada por la enfermedad en todo el territorio bogotano, se plantea la utilización de la técnica estadística espacial del análisis de datos de área, que es empleada para el análisis regresivo en datos agregados (frecuencias) (Cressie, 1993), y ampliamente desarrollado en estudios epidemiológicos dado que este método permite resumir e identificar los determinantes de una enfermedad distribuida discretamente a través del espacio (Waller y Gotway, 2004), siendo así más fácil su manipulación y por lo tanto, la obtención de conclusiones (Anselin, 1988b). También, por la naturaleza de los datos del presente estudio, esta técnica es

ideal pues sus resultados permitirán la formulación y el mejoramiento de los planes y políticas de salubridad existentes para que velen por una mejor calidad de vida de los habitantes de la ciudad de Bogotá.

## 2.2. Justificación

Bogotá, al estar localizada en el centro del país y poseer la mayor cantidad de habitantes, lucha contra enfermedades endémico-epidémicas de fácil transmisión que conducen a enormes problemas de salud pública. Uno de los tipos de enfermedades que presenta relevante preocupación en el distrito es el de las enfermedades respiratorias cuyo comportamiento cíclico en sus dinámicas y su estrecha relación con los fenómenos socioeconómicos, climáticos y ambientales de la ciudad amenazan el bienestar de la población y redundan en tasas de mortalidad alarmantes. La neumonía, es actualmente la primera causa de mortalidad de las enfermedades respiratorias en los bogotanos, siendo su principal pero no único objetivo los más vulnerables: niños menores a cinco años y adultos mayores, quienes al verse expuestos a dicha enfermedad poseen altas probabilidades de fallecer.

Según la fundación neumológica colombiana 1 de cada 5 pacientes con neumonía muere sin un temprano tratamiento y el panorama empeora en los estratos más bajos en donde las condiciones nutricionales, sociales y ambientales son claramente deficientes. La Secretaria de Salud Distrital aunque ha adoptado medidas de control y prevención de la enfermedad, no ha avanzado significativamente en ello pues los estudios sobre la neumonía solo se centran en el aspecto epidémico desestimando el aspecto social, económico y espacial. Asimismo, la identificación independiente de los determinantes de la enfermedad ha llevado a análisis que si bien funcionan, se aíslan del comportamiento integral de la misma.

De esta manera, se convierte imperativo el estudio de esta enfermedad desde una perspectiva espacial que integre diversas variables ambientales, sociales, económicas y epidémicas que interactúan a través del espacio y que pueden describir posibles patrones que permitirán identificar, controlar y mitigar el impacto de la neumonía en la población bogotana y en sus tasas de mortalidad.

Este estudio tiene como finalidad examinar los patrones geográficos de la mortalidad gracias a la neumonía y el papel que esbozan los factores ecológicos, sociales y económicos en determinarla (Buzzoa et al., 2013).

# 3. OBJETIVOS

## 3.1. Objetivo General

Identificar los factores y cuantificar su incidencia en la mortalidad de la neumonia en las localidades de Bogotá utilizando análisis de datos de área para el periodo discreto 2004-2014 (Estadística espacial).

## 3.2. Objetivos específicos

- Modelar estadística y espacialmente la mortalidad de la neumonía con respecto a factores sociales, ambientales, conductuales y de asistencia médica.
- Analizar la distribución geográfica de la neumonía en un periodo determinado de la ciudad de Bogotá.
- Hallar y modelar las relaciones espaciales de los factores de la neumonía.
- Identificar las relaciones espaciales entre la neumonía y variables sociales, económicas, conductuales, ambientales y de asistencia médica.

# 4. ANTECEDENTES Y MARCO REFERENCIAL

El estudio de la neumonía ha sido objeto de investigación en todo el mundo, investigaciones como la adquisición de la enfermedad (Baik et al., 2000), la mortalidad (Fine et al., 1997) y la determinación de algunos factores de riesgo como la edad, el género, la clase social (Kaplan et al., 2003) han permitido a varias entidades estatales en los principales centros del mundo desarrollar políticas de contención y control de la enfermedad (European Respiratory Society, 2016). Los grupos de edad extremos (niños y adultos mayores), han sido encontrados como las poblaciones con el mayor riesgo de adquirir la enfermedad y ser más susceptibles a morir por la misma (Loeb, 2003). Enfermedades asociadas a la neumonía incluyen el asma y las enfermedades cardiovasculares (Kaplan et al., 2003).

Si bien el papel de estos factores son significativos para determinar la neumonía , aún existe una varianza considerable que sigue siendo inexplicable (Loeb, 2003). Previos estudios se han hecho con enfoques médicos, pero no suelen abordar la influencia de otros determinantes como los factores sociales, económicos, ambientales y físicos (Evans y Stoddart, 1990). Varios estudios han examinado las relaciones entre la neumonía y un rango de variables socioeconómicas, que si bien han generado resultados, éstos no llegan a las mismas conclusiones (Farr, Bartlett, y Miller, 2000); Wood et al. (1999) encontraron una asociación negativa significativa entre las variables socioeconómicas y la mitigación de las muertes por neumonía. Los estándares más deficientes de las variables ambientales, físicas y sociales que afectan directamente a un individuo portador de la neumonía, aumentan las posibilidades de mortalidad (Loeb, 2003). El conocimiento de la variación y distribución a través del espacio de las enfermedades endémico-epidémicas ha aumentado significativamente el entendimiento del desarrollo y factores de las mismas, pero siguen siendo muy escasos los estudios y en algunos casos incluso inexplorables (Cliff, Hagget, y Ord, 1986).

En Colombia, hasta la actualidad no se ha explorado asiduamente el campo de los estudios epidémico-espaciales. Se han encontrado algunos estudios inherentes a enfermedades infecciosas como el Dengue en la Guajira (Ortiz, Rúa-Uribe, Suarez, y Mafla, 2013) y la tuberculosis en Antioquia (Pamplona, Jaramillo, y Alba, 2015) los cuales se han aproximado en la identificación de los factores sociales, económicos y ambientales desde un ámbito espacial. Específicamente en Bogotá, no se han encontrados muchos estudios epidemiológicos aunque la ciudad los demande constantemente debido a las fuertes dinámicas de las enfermedades que afectan su territorio. No obstante, se estima que la capital cuenta con varios estudios epidemiológicos espaciales que hasta la fecha se encuentran en desarrollo o no son del conocimiento público.

## 4.1. Epidemiología

La epidemiología se define como el estudio de la distribución y determinantes de los eventos o estados relacionados con la salud en una población específica, y la aplicación de este estudio en el control de los problemas sanitarios (Last, 1995). Esto implica que la epidemiología no solo concierne las muertes, enfermedades y discapacidades, sino también estados positivos de salud y los medios para mejorarla.

El objetivo de un estudio en epidemiología generalmente es la población humana. Una población, en epidemiología, puede definirse geográficamente o, en otros términos; un ejemplo puede ser los trabajadores de una mina o un grupo de pacientes de un hospital. Un grupo común usado en epidemiología es aquel dado por un área geográfica específica en un tiempo específico (Cameron y Jones, 1983). Esto forma la base para definir subgrupos que pueden aludir al sexo, edad, entre otros. La arquitectura de la población varía en función de su geografía y periodos de tiempo. La epidemiología debe tener en cuenta estas variaciones (Waller y Gotway, 2004).

Actualmente en el campo de la salud pública, la epidemiología es usada en gran número de formas. Los primeros estudios en epidemiología estaban preocupados por la etiología de las enfermedades transmisibles, y éstos han sido esenciales desde entonces pues llevan a la identificación de métodos preventivos. En consecuencia, la epidemiología es una ciencia básica médica cuyo objetivo final es la mejora de las condiciones de salud de la población.

Las causales de algunas enfermedades pueden vincularse primariamente a factores genéticos, pero suelen ser más comúnmente el resultado entre la interacción de factores ambientales y genéticos; entendiéndose el ambiente como los factores biológicos, químicos, psicologícos u otros que puedan afectar la salud (Cameron y Jones, 1983). El comportamiento y el estilo de vida son de gran importancia en esta conexión, y la epidemiología es altamente utilizada para estudiar su influencia e intervención preventiva en la salud.

La epidemiología es a menudo usada para describir las condiciones de salud en los grupos poblacionales. El conocimiento de las enfermedades que agobian la población es esencial para las autoridades sanitarias, quienes buscan los mejores posibles efectos identificando programas de salud primarios para la prevención y cuidado. En áreas especializadas, como en epidemiología ambiental y ocupacional, el énfasis está dado en los estudios de población con tipos particulares de exposición.

## 4.2. Neumonía

Una variedad de síndromes pulmonares ocurren después de realizar la aspiración, definiéndola como contenido gástrico en la laringe y en las vías respiratorias inferiores, esto depende de factores como la cantidad de material que se aspira, la frecuencia

de la aspiración, el mecanismo de defensa del individuo y la respuesta del huésped al material. Entre estos síndromes se encuentran la neumonitis por aspiración o síndrome de Mendelson, que es una neumonitis química causada por la aspiración de materiales y la neumonía por aspiración, que es un proceso infeccioso causado por la aspiración de las secreciones orofaríngeas colonizada por bacterias patógenas. Esta última se desarrolla posteriormente debido a la aspiración de patógenos de una orofaringe previamente colonizada que es la principal vía por la cual las bacterias entran en los pulmones. De hecho, Haemophilus influenzae y Streptococcus pneumoniae colonizan primero la naso / orofaringe antes de ser aspirados y causan neumonía adquirida en la comunidad. Sin embargo, cuando se utiliza el término neumonía por aspiración, se refiere al desarrollo de un infiltrado radiográfico con factores de riesgo que aumentan la aspiración orofaríngea. Cualquier condición que aumenta el volumen y/o la carga bacteriana de la secreción orofaríngea en donde está alterado el mecanismo de defensa del huésped puede conducir a la neumonía por aspiración (Vincent, Jesse, y Vincent, 2012).

### 4.2.1. Epidemiología

Diversos estudios demuestra que la neumonía por aspiración genera la neumonía adquirida por la comunidad del 5 al 15 % de los casos, este tipo de neumonia es la principal causa de morbilidad y mortalidad en los ancianos. Estudios epidemiológicos han demostrado que la incidencia de neumonía aumenta con el envejecimiento, el riesgo es casi seis veces mayor en adultos que alcanzan más de los 75 años, en comparación con los menores de 60 años de edad. Además, la mortalidad por neumonía aumenta sorprendentemente con el envejecimiento. La tasa de ataque por neumonía es más alta entre los que viven en hogares geriátricos (Vincent et al., 2012).

### 4.2.2. Factores de riesgo

- La disfagia es frecuente después de un accidente cerebrovascular. En los pacientes con un accidente cerebrovascular agudo, la incidencia de disfagia oscila entre 40 % y 70 %. Aproximadamente el 40-50 % de los pacientes que han tenido un accidente cerebrovascular y disfagia, han aspirado y tienen mayor riesgo de desarrollar neumonía.

- Las personas de la tercera edad están en riesgo de tener un infarto cerebral y se ha demostrado que tienen cinco veces más el riesgo de desarrollar neumonía que los pacientes con tomografía computarizada de cabeza normal.

- Casi todos los pacientes con enfermedades degenerativas del sistema nervioso central adquieren disfagia, también personas con Alzheimer, esclerosis y Parkinson, sin existir una relación directa con la enfermedad neurológica.

- Considerando que el riesgo de padecer enfermedades neurológicas degenerativas y cerebrovasculares es mayor en los hogares geriátricos, se ha resportado que la incidencia de disfagia en esta población está entre el 50 % y el 75 % que explica las tasas altas de neumonía.

- Los niños menores de 5 años son susceptibles a adquirir la enfermedad debido a que su mecanismo de defensa no se encuentra bien desarrollado o tienen problemas de desnutrición crónica.

- Contaminantes ambientales externos como los que se encuentran en el aire donde abundan inmensas cantidades de material particulado sedimentado que logran entrar por las vías respiratorias debido a la constante exposición, y contaminantes internos como el tabaquismo que produce fumadores activos y pasivos.

## 4.3. Estadística Espacial: Datos de Área

### 4.3.1. Econometría Espacial

Se encarga de identificar y reconocer los efectos espaciales, así como del tratamiento de la interacción, estructura y autocorrelación espacial en modelos de regresión con datos de corte transversal y panel de datos (Aroca y Bosch, 2000).

Los datos lattice o reticulares son aquellos que hacen referencia a regiones y zonas con formas irregulares en el espacio, como por ejemplo los municipios, departamentos, provincias, etc (Cressie, 1993). La estructura que predomina en la econometría espacial es la de grilla o lattice; es necesaria la topología y orden espacial de los datos en forma de una matriz de pesos espaciales (Serrano y Valcarce, 2000).

### 4.3.2. Autocorrelación espacial

La autocorrelación o dependencia espacial está definida como el agrupamiento o dispersión de los valores de una variable en una zona geográfica determinada, permite estudiar la manera en la que un fenómeno se esparce a través de las unidades espaciales (Perdomo, 2005), aparece debido a una relación funcional entre los eventos de dicho fenómeno que ocurren en una región específica del espacio y en otras contiguas (Moreno y Vayá, 2000). Cuando exista concentración de valores similares y próximos indica la presencia de autocorrelación positiva, en caso contrario, cuando estos valores sean disímiles y dispersos la autorrelación es negativa, y cuando no ocurren ninguno de los dos casos mencionados no existe autocorrelación espacial debido a que los valores de la variable se distribuyen de forma aleatoria (Martori, Hoberg, y Madariaga, 2008).

### 4.3.3. Heterogeneidad espacial

Se refiere a las variaciones de las relaciones espaciales por la ausencia de estabilidad del comportamiento de la variable analizada, esto produce que el fenómeno varíe según su localización geográfica (Chasco Yrigoyen, 2003). Los problemas que causa la heterogeneidad espacial se pueden resolver por medio de técnicas de la econometría estándar (Acevedo y Velásquez, 2008).

### 4.3.4. Análisis Exploratorio de Datos Espaciales

Es el conjunto de diferentes técnicas que proporcionan la exploración de los datos espaciales, detectando patrones de comportamiento, formulando hipótesis debido a la presencia de fenómenos espaciales por su ubicación geográfica del conjunto de datos. Este análisis permite conocer la localización de casos atípicos en el mapa, las asociaciones espaciales del conjunto de datos que pueden ser globales o locales y sugiere estructuras de heterogeneidad espacial en la región geográfica (Serrano y Valcarce, 2000).

Asimismo el AEDE tiene como finalidad explicar las diferentes especificaciones que incluyen un efecto de autocorrelación espacial para diseñar el fenómeno en un modelo, estos esquemas de autocorrelación en modelos regresivos requieren de una matriz W de pesos espaciales que almacene las influencias que presentan las unidades espaciales (Moreno y Vayá, 2000). Se analizan los factores de dependencia sustantiva que aparece por omitir el retardo espacial de manera incorrecta, dependencia residual que ocurre cuando la dependencia espacial no es causada por la omisión incorrecta del retardo espacial de cualquier variable sistemática, estos factores están presentes en los modelos de regresión espacial debido a la especificación del mismo.

Para cada tipo de dependencia espacial Anselin en 1988 expresó que existen diferentes modelo, cuando exista autocorrelación espacial univariante el tipo de modelo sugerido es el autoregresivo de regresión espacial, para la relación estructural entre variantes el modelo que se propone es el básico de regresión espacial, para la dependencia espacial residual el modelo de error espacial y para la dependencia espacial sustantiva, el modelo de retardo espacial.

### 4.3.5. Matriz de Pesos Espaciales

Es un arreglo matemático que define la longitud de las posibles interrelaciones entre los elementos que se encuentran en un sistema espacial estableciendo relaciones de interdependencia. Esta matriz permite plasmar las características multidireccionales de los datos espaciales y abarca las parejas posibles de realizaciones de la variable con el fin de determinar la influencia que presentan entre sí.

La matriz se supone conocida y tiene que estar definida con anterioridad debido a que entra al proceso de estimación como una constante.

La matriz $W$ presenta las siguientes características:

$$W = \begin{pmatrix} 0 & w_{12} & ... & w_{1N} \\ w_{12} & 0 & ... & w_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1N} & w_{2N} & \vdots & 0 \end{pmatrix} \tag{1}$$

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

### 4.3.6. Contrastes de Autocorrelación Espacial

Los contrastes existentes son de tipo global y local. Los globales se encargan de investigar la hipótesis de estacionariedad de una variable sobre la totalidad de la muestra y los locales verifican estacionariedad intraregional (Bohorquez, 2009).

#### 4.3.6.1 Estadísticos Globales

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

Estas pruebas de asociación espacial tienen las siguientes hipótesis:

$H_0$ : Ausencia de autocorrelación espacial.

$H_a$ : Presencia de autocorrelación espacial.

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

$$I = \frac{N}{S_0} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{2}$$

Donde

$N$: Cantidad de observaciones o tamaño muestral.

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

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

$\bar{y}$: Valor esperado de la variable y.

Se calcula un estadístico de prueba (3) que tiene una distribución normal con el fin de aceptar o rechazar la hipótesis nula.

$$Z(I) = \frac{I - E(I)}{\sigma(I)} \tag{3}$$

Un valor no significativo de Z(I) provoca el no rechazo de la hipótesis nula, aceptando la ausencia de autocorrelación espacial (distribución aleatoria en el espacio), mientras que con un valor significativo se rechaza la hipótesis nula. Cuando éste sea positivo expresa la existencia de concentración espacial de valores similares (clúster) y negativo muestra la presencia de outliers espaciales, es decir, existe concentración de valores disímiles.

- El test global **C de Geary** está representado por la siguiente ecuación:

$$C = \frac{N-1}{2S_0} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij}(y_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{4}$$

Donde

$N$: Cantidad de observaciones o tamaño muestral.

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

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

$\bar{y}$: Valor esperado de la variable y.

Tiene como estadístico de prueba la ecuación (5) que se distribuye como una normal tipificada.

$$Z(C) = \frac{C - E(C)}{\sigma(C)} \tag{5}$$

Si el valor del estadístico no es significativo se acepta la hipótesis nula donde hay ausencia de autocorrelación espacial. Cuando este valor es significativo y positivo se

rechaza la hipótesis nula y expresa la presencia de autocorrelación espacial negativa, si es negativo existe autocorrelación espacial positiva.

- En el test **G(d) de Getis y Ord** dos pares de regiones (i,j) serán vecinas cuando se encuentre dentro de una distancia d específica. Está descrito en la siguiente ecuación: (Chasco Yrigoyen, 2003)

$$G(d) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}(d)y_i y_j}{\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j}; para j \neq i \tag{6}$$

El estadístico de prueba es calculado de la misma forma que los anteriores constrastes globales y es interpretado así: si Z(G) es significativo y es mayor a cero se rechaza la hipótesis nula y se comprueba la existencia de valores elevados alrededor de i, si es menor a cero existen valores bajos alrededor de i, y si el estadístico no es significativo no se rechaza la hipótesis nula y se concluye que existe completa aletoriedad en la región.

#### 4.3.6.2 Estadísticos Locales

- Indicador local de Asociación espacial (LISA) es utilizado para detectar aglomeraciones (clústers) mediante la medición de la autocorrelación espacial donde se localiza cada observación, los mapas LISA ilustran los valores significativos y representan gráficamente los valores atípicos espaciales (Moreno y Vayá, 2000).

El estadístico local más utilizado es el I de Moran es un indicador de dependencia local basado en el contraste I de moran, está expresado de la siguiente forma:

$$I = \frac{z_0}{\sum_i \frac{z_i^2}{n}} \sum_{j \in J_i} W_{ij} Z_j \tag{7}$$

Donde $z$ es el valor de la variable bajo la normal estándar y $J_i$ el conjunto de unidades que define a los vecinos de $i$ (Bohorquez, 2009).

- El estadístico de autocorrelación espacial de Getis y Ord proviene de una lógica de análisis de patrones puntuales. En su primera formulación, consistía en una relación entre el número de observaciones dentro de un rango dado de un punto y el total de puntos. Este se aplica, de una manera general, a los valores en ubicaciones vecinas (Anselin, 1995).

Este estadístico consiste en una relación del promedio ponderado de los valores en las ubicaciones vecinas, a la suma de todos los valores, sin incluir el valor en la

ubicación ($x_i$). En contraste, el estadístico local incluye el valor $x_i$ en el numerador y denominador como se expresa en la ecuación 8.

$$G_i = \frac{\sum_j w_{ij} x_j}{\sum_j x_i} \tag{8}$$

### 4.3.7. Análisis de Cooordenadas Principales de Matrices de Pesos Espaciales(PCNM)

La prueba de autocorrelación global del I de Moran puede ser descrita de forma matricial como lo muestra la ecuación (9).

$$I(x) = \frac{n}{1'W1} \frac{x'\Omega x}{(I - 11'/n)x} \tag{9}$$

$$\Omega = (I - 11'/n)W(I - 11'/n) \tag{10}$$

Esta metodología consiste en descomponer espectralmente $\Omega$ por medio de una matriz W de pesos espaciales. El valor máximo que adquiere $I(x)$ es $I(u_i) = (n/1'W1)\lambda_{max}$ y el mínimo $I(u_i) = (n/1'W1)\lambda_{min}$, donde $\lambda_{max}$ y $\lambda_{min}$ son los valores propios mínimos y máximos de $\Omega$; con esto se garantiza que los vectores propios $u_i$ maximicen $I(x)$.

Los Mapas de Eigenvectores (MEM definición por sus siglas en inglés) son el conjunto de $u_i$ y son usados como variables explicativas en los modelos de regresión. Para hallar la mejor matriz W de pesos espaciales se genera una regresión entre la variable respuesta y los vectores propios asociados a cada matriz de pesos. El modelo seleccionado será el que posea el menor criterio de Akaike (AIC) (Dray, Legendre, y Peres-Neto, 2006).

### 4.3.8. Modelo de Regresión Espacial

El análisis de regresión trata de la dependencia de las variables explicativas, con el objeto de estimar y/o predecir la media o valor promedio poblacional de la variable dependiente en terminos de los valores conocidos o fijos de las variables explicativas (Canavos, 1992). Se trata de encontrar un método para hallar una recta que se ajuste de una manera adecuada a la nube de puntos definida por todos los pares de valores muestrales. Este método de estimación se fundamenta en una serie de supuestos, que hacen posible que los estimadores poblacionales que se obtienen a partir de una muestra adquieran propiedades que permitan señalar que los estimadores obtenidos sean los mejores.

El modelo general anidado espacial (GNS definición por sus siglas en inglés) se expresa como:

$$y = \rho W_1 y + X\beta_1 + W_2 R\beta_2 + \epsilon \tag{11}$$
$$\epsilon = \lambda W_2 \epsilon + u$$

Donde:

$y$ es el vector de la variable explicada.

$W_1 y$ es el retardo espacial de la variable de estudio.

$X$ es una matriz de variables independientes no rezagadas espacialmente.

$R$ es una matriz de variables explicativas no necesariamente con los mismos elementos de $X$ rezagadas espacialmente.

Los parámetros $\lambda$ y $\rho$ miden el grado de la dependencia espacial.

Con $u \sim N(0, \Omega)$, $\Omega$ es matriz de covarianza generalizada con elementos en la diagonal dados por $\Omega_{ii} = h_i(z\alpha)h_i$.

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

- Modelo del error espacial autoregresivo (SEM por sus siglas en inglés)

$$y = X\beta_1 + (I - \lambda W_2)^{-1} u \tag{12}$$

- Modelo del rezago espacial (SAR por sus siglas en inglés)

$$y = \rho W y + X\beta + u \tag{13}$$
$$u \sim N(0, \Omega)$$

- Modelo mixto autoregresivo regresivo espacial con error espacial autoregresivo (SARAR por sus siglas en inglés)

$$y = \rho W_1 y + X\beta_1 + \epsilon \tag{14}$$
$$\epsilon = (I - \lambda W_2)^{-1} u$$
$$\text{Con } u \sim N(0, \Omega)$$

- Modelo mixto regresivo - regresivo espacial (SDM por sus siglas en inglés)

$$y = \rho W_1 y + X\beta_1 + W_2 R\beta_2 + u \tag{15}$$

- Modelo mixto regresivo cruzado - regresivo espacial (SLX por sus siglas en inglés)

$$y = X\beta_1 + W_2 R\beta_2 + u \tag{16}$$

- Modelo mixto regresivo cruzado - regresivo espacial con error espacial autorregresivo (SDEM por sus siglas en inglés)

$$y = X\beta_1 + W_2 R\beta_2 + (I - \lambda W_2)^{-1}u \tag{17}$$

### 4.3.9. Modelos Espaciales Alternativos

Si bien, los modelos paramétricos o modelos de regresión espacial tienen un amplio espectro para modelar la dependencia y variabilidad espacial de las variables, es necesario considerar posibles alternativas en las cuales el comportamiento de los modelos discrepa de los métodos estadísticos tradicionales.

Varias alternativas de la familia de los modelos de regresión espacial han sido desarrolladas , entre las que se destacan los modelos *no paramétricos* y *semiparamétricos* así como los *modelos bayesianos jerárquicos autorregresivos espaciales de efectos individuales.*

#### 4.3.9.1 Aproximaciones No Paramétricas y Semiparamétricas

El modelamiento inadecuado de la heterogeneidad espacial puede redundar en dependecia espacial (McMillen, 2003). Los modelos espaciales no paramétricos y semiparamétricos intentan modelar esta heterogeneidad espacial. La función de las aproximaciones semiparamétricas consiste en modelar información de corte transversal en donde la dependencia espacial se origina en su fuente. El modelo de regreción lineal ponderado (LWR) modela la dependencia espacial basándose en modelos separados estimados usando una submuestra de la información apoyado en observaciones cerca de cada observación (McMillen, 1996).

En la ecuación (7) se observa un tipo de modelo no paramétrico que modela la dependencia espacial. $\Omega(i)$ representa una matriz diagonal de $nxn$ que contiene los pesos espaciales basados en la distancia para la observación $i$ que refleja la distancia entre la observación $i$ y las demás observaciones.

$$\Omega(i)^{1/2}y = \Omega(i)^{1/2}X_{\gamma_i} + \Omega(i)^{1/2}X_{\epsilon_i} \tag{18}$$

El subindice $i$ en $\gamma_i$ indica que el vector parámetro $kx1$ está asociado a la región $i$. El modelo LWR produce $n$ vecstores de parámetros estimados, uno para cada región/observación. Estos estimados son calculados usando la expresión:

$$\hat{\gamma_i} = (X'\Omega(i)X)^{-1}X'\Omega(i)Y \tag{19}$$

Pace y LeSage (2004) explican que los métodos LWR al aumentar el tamaño de la muestra producen valores estimados menos volátiles e incrementan la dependencia espacial. Seleccionar una muestra pequeña reduce la variabilidad espacial, pero a cambio la variabilidad de los parámetros aumenta impidiendo la detección de la variación a través del espacio de los parámetros en patrones sistemáticos. Así, Pace y LeSage introdujeron un esquema de estimación local autorregresivo espacial (SALE). El método SALE modifica la volatilidad de la dependencia espacial del LWR adicionando un rezago espacial de la variable dependiente (Pace y LeSage, 2004).

Otros métodos no paramétricos se centran en la estimación de la estructura de covarianza/correlación espacial. En este caso, la dependencia radica en la estructura espacial de la conectividad de las observaciones de la muestra en lugar de tomar el enfoque convencional que especifica esta estructura en función de la configuración espacial de las observaciones. También es típico que estos enfoques se centren en modelos derivados de las familias de modelos autorregresivos al restringir el parámetro $\rho = 0$, produciendo estructuras de covarianza espacial entre las unidades observadas (Conley y Ligon, 2002).

#### 4.3.9.2 Autorregresivo Espacial de Efectos Individuales

Existe una extensión a los modelos de regresión básicos que incorpora un vector de efectos individuales siguiento un proceso autorregresivo espacial. El modelo produce efectos individuales estimados para regiones espaciales agregadas de la muestra. Por ejemplo, si $n$ observaciones es el vector $y$ y la matriz $X$ representa departamentos, el vector de efectos individuales $\theta$ puede reflejar $m$ municipios, donde la matriz $\triangle$ contiene la agregación necesaria de $n$ departamentos a $m$ municipios. Ésto permite que el error asociado a los efectos del vector indivial $\theta$ exhiba varianzas heterogéneas, con un parámetro escalar de varianza por cada municipio (Pace y LeSage, 2004).

$$y = X\beta + \triangle\theta + \epsilon \tag{20}$$

$$\theta = \rho W\theta + \eta \tag{21}$$

Esta especificación refleja una aproximación más parmoniosa para modelar los efectos individuales que la típica aproximación que puede depender de la introducción de $m$ variables dicótomas en el modelo.

### 4.3.10. Regresión geográficamente ponderada (GWR)

Tiene como finalidad estimar un vector de parámetros diferente para cada observación, para que este vector dependa de la ubicación espacial de cada observación. Se

realiza la estimación debido a la inestabilidad de los parámetros y la ausencia de estacionariedad espacial (Bohorquez, 2009).

$$y_i = X_i\beta(v_i, v_i) + u \tag{22}$$

Donde

$i = 1, 2, ..., n$

$(v_i, v_i)$ coordenadas geográficas

La estimación de $\beta$ se realiza de la siguiente forma:

$$\widehat{\beta}(v_i, v_i) = [X'W(v_i, v_i)X]^{-1}X'W(v_i, v_i)y \tag{23}$$

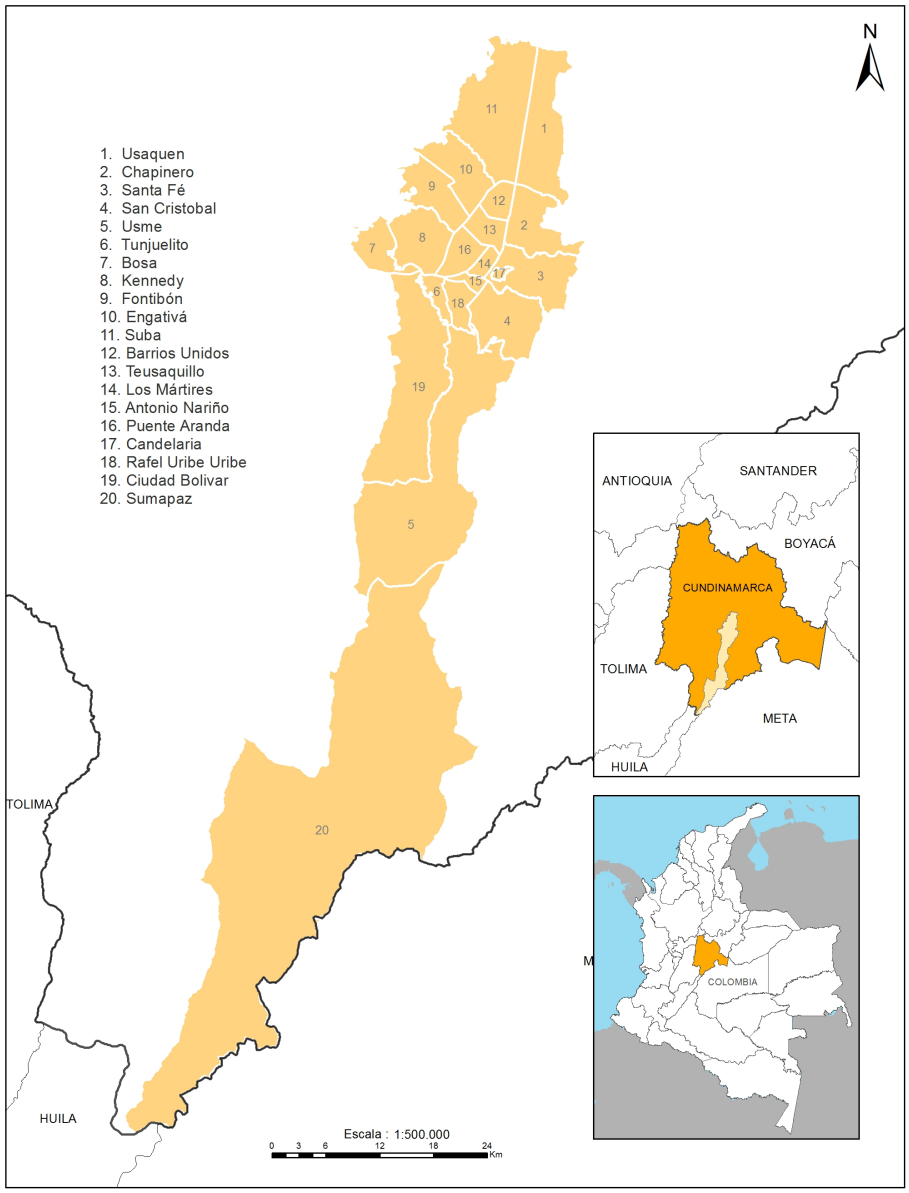
Donde $W(v_i, v_i) = Diag(\alpha_i j)$ y $\alpha_i j$ es una ponderación determinada en términos de la distancia de la observación espacial $i$ a la observación espacial $j$ ($j = 1, 2, ...n$).

# 5. METODOLOGÍA

## 5.1. Área de Estudio

La información obtenida se compiló en la ciudad de Bogotá en el departamento de Cundinamarca, Colombia (Figura 1).

Figura 1: Ubicación espacial de la región de estudio.




*Fuente: Elaboración propia*

La elección de esta ubicación espacial tuvo como base la constante vigilancia de la neumonía en Bogotá debido a su impacto en el territorio capitalino y su creciente relevancia en la morbilidad y mortalidad de su población, especialmente, en los residentes vulnerables (niños menores de 5 años y adultos mayores de 65)(Secretaria Distrital de Salud de Bogotá, 2004). Asimismo, Bogotá al ser la capital del país, abarca fuertes dinámicas económicas, sociales, conductuales y ambientales que sugieren la existencia de factores que pueden aumentar el riesgo de adquirir la enfermedad y morir a causa de ella. La región posee 1775 $km^2$ que constituyen el 7,3 % de la superficie del departamento de Cundinamarca, adicionalmente, se encuentra dividida en 20 localidades (Usaquén, Chapinero, Santa Fe, San Cristóbal, Usme, Tunjuelito, Bosa, Kennedy, Engativá, Suba. Barrios Unidos, Teusaquillo, Los Mártires, Antonio Nariño, Puente Aranda, La Candelaria, Rafael Uribe Uribe, Ciudad Bolívar, Sumapaz). Yace sobre una altura media de 2.625 m.s.n.m. Presenta una población actual aproximada de 8'081.000 habitantes y se espera llegue a los 8'380.000 para el año 2020 (Alcaldía Mayor de Bogotá, 2018).

Existían aproximadamente 7,8 millones de residentes en Bogotá en el 2014 según el Departamento Nacional de Estadística (DANE). El rango de las poblaciones por localidad va desde 24.000 en La Candelaria, al noroeste (Figura 1), a más de 1.1 millones en Suba, al norte. El centro de Bogotá es el área menos poblada debido a la poca superficie que ocupan las localidades en éste, mientras que el norte y sur compuesto por áreas metropolitanas presenta el mayor acopio de pobladores. Algunas zonas del sur y el oeste de la ciudad presentan áreas escasamente pobladas debido a la agricultura rural puntual y la falta de servicios domiciliarios (Alcaldía Mayor de Bogotá, 2018). Para los pobladores de Bogotá, el acceso a los servicios de atención médica está sujeto al régimen y a la entidad prestadora a los que pertenezcan. Sin embargo, la mayor parte de los centros asistenciales se encuentran distribuidos en el centro y norte de la ciudad por lo que los residentes ubicados al sur, oeste y partes del este deben recorrer una mayor distancia para acceder a los servicios de salud en comparación con sus contrapartes urbanas.

## 5.2. Materiales

Para desarrollar el proyecto se obtuvo como principal insumo la base de datos en Excel brindada por la Secretaría Distrital de Salud (SDS) con información por causas de muertes en las localidades de Bogotá del 2004-2014. Una vez identificada la enfermedad de la Neumonía, se procedió a conformar la Base de Datos Espacial en formato Shapefile con las localidades y variables sociales, ambientales, conductuales y de asistencia médica encontradas en diversas fuentes para el estudio.

La base de datos de la Secretaría Distrital de Salud de Bogotá de las muertes en Bogotá por las principales enfermedades fue usada para obtener la información de las muertes por neumonía como principal diagnóstico, segregadas por número de fallecimientos por localidad. Esta base de datos registra las muertes en todos los

hospitales de Bogotá, la edad y régimen de salud al que pertenecían las victimas cubriendo un periodo de 10 años (2004-2014); información más reciente no se obtuvo debido a que se encuentra en procesamiento por la secretaría. En la Tabla 1 se observan las variables provistas por la SDS.

Tabla 1: Selección variable respuesta.

| **Tipo** | **Variable** | **Descripción** |
|---|---|---|
| Descriptivo | Locnombre | Nombre de la localidad |
| | Loccodigo | Identificador númerico de la localidad |
| | Locarea | Superficie en $km^2$ de la localidad |
| | POB | Población por localidad |
| **Variable Respuesta** | M_* | Muertes por neumonia * Género, Edad y Regimén de salud |
| | TASA | Tasa de mortalidad específica por cada 10.000 habitantes |
| | **SMR** | Tasa de mortalidad estandarizada |

Aunque las muertes por la enfermedad se encontraban segregadas por edad, género y régimen de salud se estandarizó el total de muertes por localidad (SMR) con el fin de representar los patrones espaciales de la distribución, establecer cuántas personas por miles de la población morirá en un año dado y cuáles son las causas de estas defunciones. La estandarización de las muertes (SMR) será la variable respuesta del presente estudio.

La selección de las variables explicativas fue guiada por el marco conceptual y los factores incidentes en la neumonía. El presente estudio se concentra en los factores ambientales, sociales, económicos y conductuales menos estudiados espacialmente asociados a la enfermedad. Las variables explicativas son descritas en la Tabla 2 y en futuros párrafos. Para las variables explicativas se seleccionaron los años **2004, 2007, 2010** y **2014**, debido a que son representativos durante el periodo de análisis y principalmente, por la disponibilidad de los datos obtenidos en las diferentes fuentes de información, que si bien registran varios factores, se realizan con periodicidad de cada 4 años aproximadamente tanto a nivel nacional como distrital.

Las variables explicativas fueron derivadas de varias fuentes, todas ellas estatales. Las Encuestas Nacionales de Calidad de Vida (ECV) y las Encuestas Multipropósito de Bogotá (EMB) fueron las que mayor información sobre las variables proveyeron. De igual forma, algunas secretarias distritales, la red de monitoreo de calidad del aire y el instituto meteorológico y de estudios ambientales aportaron información sobre las variables ambientales y conductuales.

Tabla 2: Variables explicativas seleccionadas para el estudio.

| Tipo | Variable | Descripción | Fuente |
|---|---|---|---|
| Ambiental | TEM | Temperatura Media Anual por año | IDEAM |
| | CPM | Contaminación media movil anual por PM10 | RMCAB |
| Social y Económica | IDD | Tasa de desempleo | ECV, EMB, DANE |
| | CVV | Índice de Calidad de vivienda | ECV, EMB, DANE |
| | CVD | Índice de Calidad de Vida | ECV, EMB, DANE |
| | NBI | % de personas pobres por NBI | ECV, EMB, DANE |
| | ESC | % Asistencia Escolar | ECV, EMB, DANE |
| | ACU | % Cobertura del servicio de acueducto | EAB |
| Conductual | DEP | Indicador de practicas deportivas | SISCRED |
| | NUT | % de hogares donde algún miembro del hogar no consumió ninguna de las tres comidas | ECV, EMB, DANE |
| Asistencia médica | IPSE | Número de IPS por localidad | REPS |
| | VAC | Cobertura de vacunación con anti haemophilus influenzae tipo B | SDS |

**IDEAM:** Instituto de Hidrología, Meteorología y Estudios Ambientales; **RMCAB:** Red de Monitoreo de Calidad del Aire de Bogotá; **ECV:** Encuesta nacional de calidad de vida; **EMB:** Encuesta multipropósito de Bogotá; **SISCRED:** Secretaría de Cultura, Recreación y Deporte; **REPS:** Registro Especial de Prestadores de Servicios de Salud; **SDS:** Secretaria Distrital de Salud

La información cuantitativa de las variables se obtuvo para los años seleccionados; las variables ambientales para los 4 años fueron adquiridas en su totalidad por el IDEAM y el RMCAB; las variables socioeconómicas y la variable nutricional para los años 2004 y 2007 se extrajeron de boletines del DANE 2004 y ECV 2007, incluyendo algunas proyecciones de CV para el 2004. En tanto la información para el 2011 y 2014 se encontró en su totalidad en EMB 2011 y EMB 2014. La cobertura de acueducto para todos los años se hallo en la Empresa de acueducto y alcantarillado de Bogota (EAB). Las variables conductuales se reportaron en el SISCRED y en las ECV, EMB y el DANE. Finalmente los factores de asistencia médica se obtuvieron de la SDS y de la REPS.

### Descripción de las Variables

- **Tasa de mortalidad estandarizada (SMR):** La SMR es un estadístico muy difundido para representar los patrones espaciales de la distribución de una enfermedad. Estandariza los datos, re-expresándolos como la proporción entre el número de casos observados y el número de casos esperados (Waller y Gotway, 2004). Es la variable endógena o la variable respuesta del presente proyecto que se pretende relacionar espacial y econométricamente con las variables explicativas o exógenas ambientales, sociales, económicas, conductuales

y de asistencia médica.

- **Temperatura (°C)**: Valor dado en grados centígrados obtenido a partir del promedio de temperaturas medias registradas para los doce meses de un año para cada localidad de Bogotá. Es una variable exógena de tipo ambiental, cuyos valores, aunque no fluctúan considerablemente a través del territorio capitalino, sí presentan pequeñas variaciones que podrían ser relevantes para identificar las muertes por la neumonía. A menor temperatura ambiente la neumonía prevalece con mayor fuerza en los pacientes y a mayor temperatur ambiente las bacterias causantes de algunos tipos de neumonía se reproducen con mayor celeridad (Davis et al., 2010).

- **Contaminación por PM10**: Valor de la media móvil de PM10 entendido como la cantidad de partículas en suspensión que tienen un tamaño menor o igual a 10 $\mu$m constituidas por una multitud de contaminantes diferentes. Los efectos de estas partículas sobre la salud humana dependen de qué tanto las partículas penetran el tracto respiratorio produciendo irritación de las vías respiratorias y otros efectos negativos en la salud. Es una variable exógena de tipo ambiental, que representa suma importancia en las políticas de salud pública del distrito capital; en los últimos años ha aumentado en Bogotá y se estima que ha repercutido directamente en la morbilidad y mortalidad de las enfermedades respiratorias, en especial en la neumonía (Cliff et al., 1986).

- **Tasa de Desempleo**: Valor que mide el nivel de desocupación en relación con la población activa. En otras palabras, es la parte de la población que estando en edad, condiciones y disposición de trabajar (población activa) se encuentra sin una ocupación o empleo. Es una variable exógena de tipo socioeconómico, que en algunos estudios se ha encontrado como un factor de riesgo en la mortalidad por neumonía y una fuerte relación negativa entre la clase social y la incidencia de la enfermedad (Izquierdo et al., 2016). En Bogotá, dado el desequilibrio económico de la población, esta variable adquiere relevancia y podría de manera directa explicar parte de la mortalidad por neumonía.

- **Índice de Calidad de Vivienda:** Valor que determina si una vivienda es o no adecuada para su ocupación basándose en sus características físicas y de entorno. Este valor oscila entre 0 y 12,9 siendo el valor más alto la vivienda bajo las mejores condiciones físicas y ambientales. En cuanto a las características físicas se tienen en consideración los materiales de construcción, los acabados y dimensiones de la vivienda; en las características de entorno se concentra en la cobertura de servicios públicos domiciliarios y relación de vecindad. Es una variable independiente de tipo socioeconómico, fielmente ligada a la neumonía. Viviendas con calidades precarias (poca ventilación, pobre cobertura de servicios públicos, humedad, etc.) pueden crear ecosistemas en donde la neumonía puede prosperar causando mayores índices en morbilidad y mortalidad.

- **Índice de Calidad de Vida:** Este índice mide el estándar de vida mediante la

suma de variables de capital humano, variables de acceso potencial a bienes físicos y variables que describen la composición del hogar. Este índice asigna a cada hogar un puntaje que varía entre 0 y 100, aumentando cuando mejoran las condiciones de vida del hogar. En otras palabras, a mayor puntaje, mejores son las condiciones de vida de la población analizada (Izquierdo et al., 2016). Es una variable socioeconómica exógena que evidentemente se relaciona con la enfermedad de estudio: a mayores niveles de calidad de vida se espera la mortalidad de la neumonía disminuya, pues las condiciones son propicias para mitigar el impacto de la enfermedad; inversamente proporcional, menores niveles de calidad de vida se traducen en indiferencia hacia la enfermedad y por tanto a un mayor número de muertes a causa de ésta.

- **Pobreza por NBI:** Representa el porcentaje ( %) de personas consideradas pobres debido a las deficiencias en las Necesidades Básicas Insatisfechas NBI. Se le dice "pobre" a una persona si pertenece a un hogar que presenta carencias en la satisfacción de sus necesidades básicas, siendo éstas, en vivienda, salud, educación y empleo. Viviendas con características inadecuadas, servicios inadecuados, alta dependencia económica, niveles bajos de educación y/o hogares en hacinamientos críticos son particularidades propias de pobreza por NBI. Si bien es una variable exógena, que se relaciona íntimamente con otras como la calidad de vida, la calidad de la vivienda y el porcentaje de escolaridad, representa un valor único que relaciona porcentualmente la pobreza con el estado de la población. Es una variable socioeconómica, que en varios estudios (Izquierdo et al. (2016), Loeb (2004) y Flory et al. (2009)) ha sido encontrada como un factor que incrementa la susceptibilidad del sistema inmune a enfermedades como la neumonía, pues se relaciona con factores como el nivel de estrés, una dieta inadecuada o factores de vida asociados con la privación. A mayores índices de pobreza se espera sea mayor el número de casos de morbilidad y mortalidad por la enfermedad respiratoria.

- **Asistencia escolar:** Este valor simboliza el porcentaje ( %) de personas de 5 años o más que asisten a un establecimiento educativo en proporción al total de la población en edad de estudiar. Uno de los objetivos más importantes de cualquier nación es que todos los niños y adolescentes de su jurisdicción, independientemente de su origen social, recorran una trayectoria educativa; sin embargo, esta trayectoria no sucede en todos los casos; al producirse esta desigualdad escolar, también se producen cambios en la calidad de vida de la población incluyendo su estado de salud. Es una variable exógena de tipo socioeconómico, que se espera actúe de forma inversamente proporcional a la mortalidad por la neumonía: si aumenta la asistencia escolar, debe disminuir la mortalidad por neumonía, y viceversa. Esta afirmación se hace basándose en estudios que han asociado la neumonía al nivel educativo y asistencia escolar de los pacientes de la enfermedad (Jasti, Mortesen, Obrosky, Kapoor, y Fine, 2008).

- **Cobertura de Acueducto:** En este indicador se relaciona el porcentaje ( %) de cubrimiento de las redes de acueducto en la Ciudad, teniendo en cuenta el número de usuarios registrados legalmente ante la Empresa de Acueducto y Alcantarillado de Bogotá (EAAB). Es una variable explicativa de tipo socioeconómico, que según Hennessy et al. (2008) en un estudio realizado en Alaska, las regiones con menor proporción del servicio de acueducto representan tasas significativamente más altas de hospitalizaciones por influenza y neumonía. Asimismo, en la prevención de la neumonía se destacan la importancia del agua, la sanidad e higiene.

- **Indicador de prácticas deportivas:** Es el porcentaje ( %) de personas de 13 años y más que practican algún tipo de deporte. Se divide el total de personas de la población de estudio (total Bogotá, hombres, mujeres) que practican algún deporte, entre el total de población de estudio. Estas prácticas deportivas corresponden a la ejecución individual y colectiva de las diferentes disciplinas deportivas. Es una variable independiente de tipo conductual o de comportamiento, que siempre ha sido relacionada con la salud pública de una población. Se ha hablado que el sedentarismo afecta negativamente la esperanza de vida y es un factor de riesgo para el desarrollo de enfermedades respiratorias en poblaciones vulnerables (niños menores a cinco años y ancianos). Se espera una relación inversamente proporcional con la variable endógena del estudio.

- **Hogares sin consumo de tres comidas:** Esta variable simboliza el porcentaje ( %) de hogares en los que por falta de dinero algún miembro del hogar no consumió ninguna de las tres comidas, uno o más días de la semana. Es una variable exógena de tipo conductual, asociada a niveles de desnutrición en el hogar. Un estudio realizado por Desai, Sucher, y Chen (2004) encontró una relación entre los malos hábitos de nutrición y la neumonía sugiriendo que éstos tiene un impacto en estadías prolongadas en el hospital en pacientes con neumonía y en casos severos, la muerte por la enfermedad.

- **Número de IPS:** Este valor alude al número de Institutos prestadores de salud (IPS) reportados en una localidad por cada 100.000 habitantes. Es una variable de asistencia médica. Este factor alude a la cobertura de salud que se presenta en una localidad determinada. Se espera que a mayor cobertura, sea más fácil tratar la neumonía y por ende los índices de mortalidad sean más bajos.

- **Cobertura de vacunación con anti haemophilus influenzae tipo B:** Mide la proporción en porcentaje ( %) de niños menores de un año que han recibido las dosis de vacuna establecidas en el esquema, y que por consiguiente quedan vacunados. La Haemophilus influenzae es un cocobacilo causante de infecciones graves como neumonía, meningitis, artritis, otitis y sepsis. Sin esta vacuna los pacientes pueden desarrollar la enfermedad de la neumonía y morir a causa de este a pesar de un tratamiento. Es una variable explicativa de tipo asistencial médico, está relacionada con el contagio y mortalidad de la neumonía. A

mayores índices en la cobertura de esta vacuna, se espera el contagio por la misma sea menor redundando en mortalidades mínimas.

Si bien todas las variables previamente mencionadas se seleccionaron teniendo como base el marco conceptual y previos estudios de la enfermedad, también se seleccionan pensando en cómo éstas pueden mejorarse con el fin de mitigar el impacto de la enfermedad en la población, para este estudio, en Bogotá. Las variables ambientales, aunque no se pueden controlar directamente por los humanos, sí depende de factores (como la emisión de gases debido al número de fabricas, vehículos, aerosoles, etc.) que pueden regularse a través de políticas del estado e intervenciones gubernamentales. Las variables socioeconómicas, objetivo de múltiples políticas públicas, son sin duda los factores de mayor relevancia para el Estado y máxime cuando su distribución es heterogénea en el territorio; éstas deben estar contantemente monitoreadas pues no solo repercuten en la calidad de vida de los pobladores sino en múltiples dinámicas como lo es la salud de los mismos. Las variables conductuales o de comportamiento, que perjudican particular y colectivamente a los individuos, pueden generar acciones que promuevan las buenas prácticas en hábitos saludables. Las variables de asistencia médica, que son necesarias para garantizar las condiciones de salubridad y sanidad de los residentes de un territorio, sobre éstas debe reposar mayor atención por parte del gobierno si se habla de control, prevención y remediación de las principales enfermedades que afligen a los bogotanos.

**Softwares Utilizados**

El porcesamiento y análisis de los datos se realizó en software estadísticos y softwares geográficos.

- **R v3.4.4:** Es un software libre para realizar procesos estadísticos y gráficos. En este programa se realizó la mayor parte del proceso de la generación del modelo econométrico espacial y la validación del mismo; procedimientos como la selección de la matriz de pesos espaciales, el modelo autorregresivo, los estadísticos de prueba, las estadísticas descriptivas, entro otros procesos fueron desarrollados en el software. De la misma manera, la mayor parte de los gráficos fueron generados en este programa. La ventaja de este programa frente a otros radica en que al ser de libre licencia permite la colaboración de varios usuarios, que aportan librerías y funciones que agilizan el procesamiento de los datos; por otro lado, también permite la programación y generación de funciones aún no disponibles en las librerías.

- **GeoDa v1.12:** Desarrollado por la Universidad de Chicago, GeoDa es un software que introduce el análisis de datos espaciales enfocado a la rama de la estadística espacial llamada econometría espacial. El software es de uso libre y permitió la generación de gráficos de autocorrelación espacial, análisis exploratorio de datos espaciales y algunos estadístico como el I de Moran. Su mayor ventaja consiste en sus múltiples herramientas para analizar y modelar los datos espaciales de dominio fijo y discreto (datos de área).

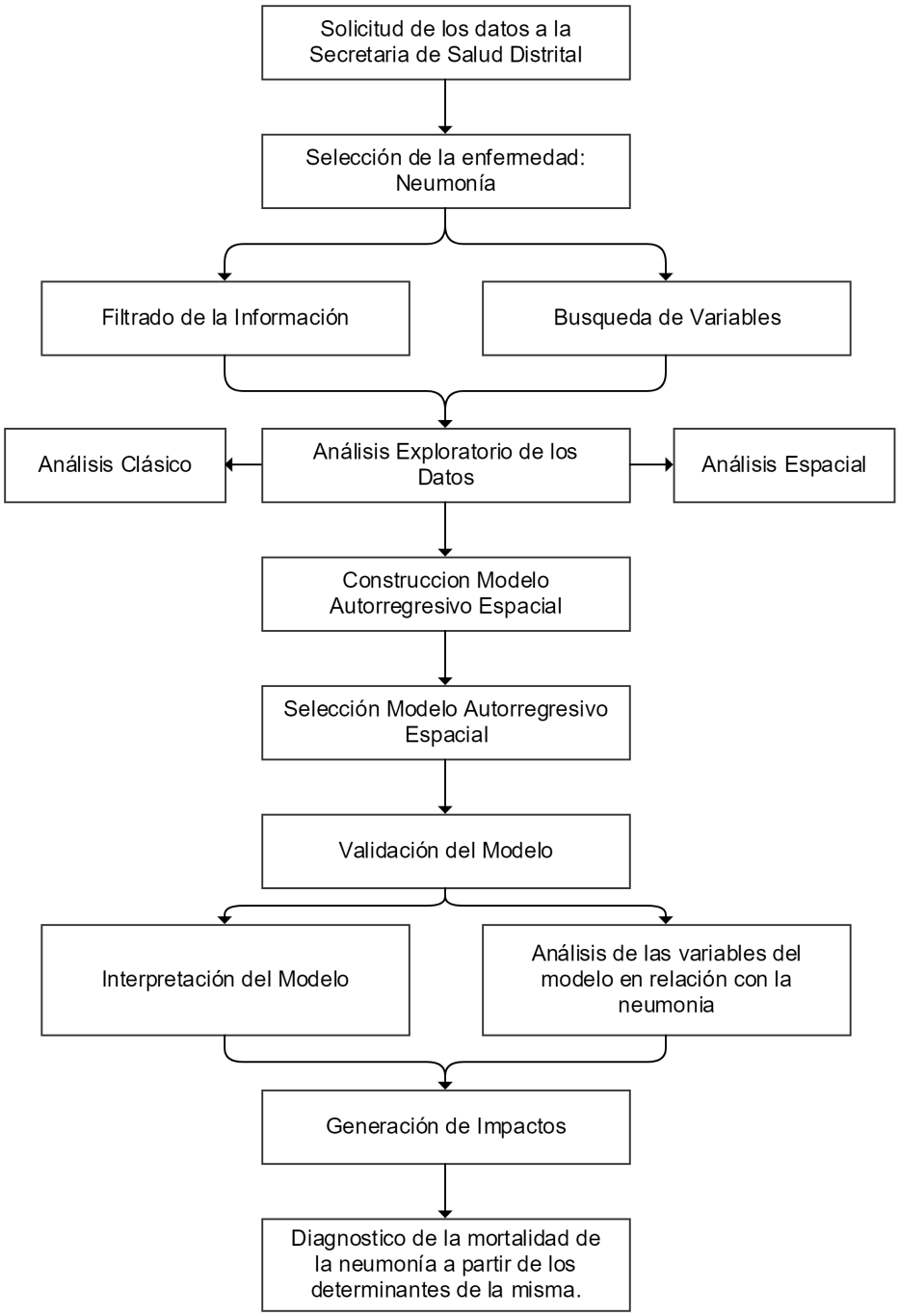
- **QGis v3.0.1:** Este software libre orientado a la creación, edición, visualización, análisis y publicación de información geoespacial. Utilizado en el presente estudio mayormente para la visualización de la información espacial, la adición y depuración de las variables a un formato geográfico y algunos análisis espaciales; así como la producción de mapas temáticos alusivos al estudio. Su mayor ventaja consta de ser un software libre que manipula la información espacial y permite la creación de plugins a través de Python para facilitar el análisis de los datos.

## 5.3. Métodos

El objetivo principal de este documento es mirar más allá del enfoque biomédico tradicional de los determinantes de la neumonía, identificar en cambio qué papel social, económico, ambiental, conductual y asistencial médico pueden tener éstos factores al determinar las tasas de mortalidad de la neumonía a través del espacio y el tiempo; para cumplir con este objetivo se debe identificar y modelar la dependencia espacial entre los diferentes factores y la variable respuesta. En la figura 2 se presenta el flujo de trabajo para el trato y análisis de los datos del presente estudio y en párrafos posteriores se explicará cada componente y sus subdivisiones.

Cabe mencionar que los métodos que se emplearan en el desarrollo del trabajo están enmarcados en el análisis de datos de área para datos epidemiológicos, en este caso de la neumonía.

Figura 2: Diagrama de flujo de los métodos a utilizar




*Fuente: Elaboración propia*

### 5.3.1. Solicitud de Datos

Se realizó una solicitud formal a la Secretaría distrital de Salud de Bogotá con el fin de obtener información de la morbilidad y mortalidad por las principales enfermedades incidentes en el territorio capitalino en el periodo 2000-2016. La SDS proveyó un archivo Excel que relacionaba las 20 principales enfermedades de fallecimientos en Bogotá; la información originalmente se encontraba segregada por localidad, edad, sexo y régimen del número de difuntos para los años comprendidos entre 2004 y 2014 (para los años anteriores y posteriores no existían registros o aún no habían sido procesados).

### 5.3.2. Selección de la enfermedad

Después de analizar las 20 enfermedades disponibles en el registro entregado por la SDS se seleccionó la neumonía como la enfermedad de estudio para el presente proyecto. En primera instancia por ser el principal diagnóstico con mayor índice de fallecimientos a través del periodo y en segundo lugar por ser una enfermedad considerablemente estudiada en el mundo (Farr et al., 2000) asociada a múltiples factores sociales, económicos y ambientales.

### 5.3.3. Filtrado de la Información

El insumo principal fue la base de datos en formato excel brindada por la Secretaría Distrital de Salud, en ésta se encuentran los conteos de muertes de diferentes enfermedades con intervalos de edad que abarca menores a 1 año hasta mayores a 80 años, en cada uno se discrimina la cantidad de hombres y mujeres fallecidos. Para obtener únicamente las muertes por Neumonía se utilizó la herramienta de filtrado del software.

La variable respuesta corresponde a la Tasa de Mortalidad Estandarizada, es un estadístico que representa patrones espaciales de distribución de una enfermedad, estandarizando los datos y re-expresándolos como la proporción entre el número de casos observados y el número de casos esperados (Galeano, Parra, Santa Guzmán, y Rodríguez, 2017).

$$SMR_i = \frac{Total_i}{E_i} \tag{24}$$

$$E_i = \frac{\sum Total_i}{\sum POB_i} POB_i \tag{25}$$

Donde $Total_i$ Es el número de casos observados de la Neumonía en una región i y $POB_i$ es la población en riesgo de contraer la enfermedad en dicha región i.

En las tablas 39 y 40 se encuentra la información que se utilizó para el cálculo la variable endógena y los resultados obtenidos a partir de la ecuación (24).

### 5.3.4. Búsqueda de Variables

La selección de las variables explicativas se realizó con base en previos estudios, el comportamiento de la capital y la distribución de las muertes. Se obtuvieron variables de diversas fuentes estatales basándose en los años escogidos para el estudio y la disponibilidad de la información. En la sección 5.2. Materiales se describen detalladamente las variables encontradas, el por qué de su elección y su fuente.

### 5.3.5. Análisis Exploratorio de Datos

El análisis exploratorio de datos espaciales (AEDE) incluye la visualización y exploración de datos en un espacio geográfico. El AEDE utiliza varios métodos, gráficos y mapas estadísticos para describir la información y hallar posibles relaciones entre la misma. Existen varias preguntas que AEDE puede responder por ejemplo si un grupo de observaciones en un diagrama de dispersión está relacionado con sus ubicaciones espaciales o si un área geográfica particular exhibe características diferentes a otra área geográfica teniendo en cuenta el mismo estadístico. Además, AEDE puede ayudar a crear nuevas hipótesis sobre los datos espaciales que posteriormente serán investigados con más detalles en análisis posteriores. Asimismo, los métodos de AEDE deben aplicarse antes de cualquier modelado avanzado y prueba de hipótesis estadísticas.

En esta sección, se encontrará que las aplicaciones de estadísticas espaciales para la neumonía, sus variables y la ubicación se vuelven más realistas, sobre todo porque los métodos utilizados y las estructuras de dependencia espacial se parecen más a la econometría en el dominio del espacio y del tiempo. La comprensión de la estacionariedad y la isotropía propias de la estadística espacial se trasladan a los estudios de datos de área, incluidos los intentos de detectar el rango espacial en el que los efectos del vecindario, los efectos secundarios, se hacen evidentes. Otras estadísticas descriptivas de la estadística básica también coadyuvan al entendimiento de la variable respuesta y su interacción con sus explicativas. Algunos mapas de cuantiles permiten el entendimiento de la variable a través del espacio y el periodo analizado.

#### 5.3.5.1 Visualización y exploración básica

El primer paso para explorar los datos espaciales consiste en mostrar la información en un mapa para identificar posibles patrones en el comportamiento de la variable respuesta. Para el caso de la neumonía, se realizó un mapa de cuantiles para la SMR en cada año. Este mapa permite determinar las locaciones espaciales en que la tasa de mortalidad es mayor y esbozar relaciones iniciales entre los valores de la SMR y las variables económicas, sociales, ambientales, conductuales y de asistencia médica. En este gráfico inicial se puedo apreciar autocorrelación espacial global debido a aglomeraciones de valores similares en las localidades de Bogotá y algunos valores atípicos en localidades de la ciudad. También se identificaron algunas variables exógenas que podrían relacionarse directamente con las altas y bajas tasas de SMR pues la distribución es similar en el territorio de Bogotá.

De igual forma, se produjeron las estadísticas descriptivas básicas para la variable respuesta como medidas de tendencia central y dispersión, así como histogramas, boxplot y gráfico normal. Las variables de tendencia central y los gráficos estadísticos permitieron dar una primera aproximación a los datos atípicos, distribución y forma de la tasa de mortalidad. Así, fue posible establecer qué localidades representan valores alarmantes de fallecimientos por la enfermedad, cuáles se encuentran por debajo de la media y cuáles valores son normales para Bogotá. Los estadísticos de dispersión facultaron la interpretación de qué tan compactos son los datos y cómo, desde una vista somera, se podrían agrupar los datos espacialmente.

Las estadísticas bivariadas y mutlivariadas se realizaron con el fin de cuantificar las relaciones entre la variable respuesta y las variables explicativas. De esta forma se buscaba mitigar la redundancia de la información, identificar la dependencia e interdependencia y la contribución de los factores en la variable evento. Con estas estadísticas fue posible ver las relaciones de linealidad entre la SMR y las variables independientes, qué variables son parecidas entre sí y de cuales se podría prescindir en el estudio.

#### 5.3.5.2 Matriz de pesos espaciales

La característica multidireccional de los datos espaciales es plasmada en la matriz de pesos espaciales o matriz de contigüidad, donde se involucran todos los posibles pares de observaciones y a través de la cual se debe decidir la influencia que tienen entre sí cada uno de éstos. Es imperativo que esta matriz sea conocida antes de estimar el modelo autorregresivo, pues hace parte del mismo como una constante. En consecuencia, es un factor muy importante la correcta especificación de esta matriz pues explica la importancia relativa entre los individuos.

Un punto importante a considerar radica en que no existe un algoritmo para definir los pesos o ponderaciones espaciales; por el contrario se deja libertad a quién estudia

el fenómeno pues se asume que éste conoce el comportamiento de su variable y la relación de la misma a través del espacio. La elección de la correcta matriz de contigüidad es de suma importancia, pues de ella dependerán las relaciones espaciales de los datos. Existen múltiples criterios para generar el valor de las entradas. Las matrices de pesos binarias asignan el valor de uno a las unidades contiguas y cero a las demás. Las matrices de pesos estandarizadas se construyen a partir de las relaciones binarias y en donde cada componente de la matriz se pondera por la suma de la fila a la que pertenece. Existen medidas extendidas de interrelaciones espaciales, asociadas a las matrices generalizadas de pesos. Con el fin de encontrar la mejor matriz de pesos espaciales se realiza una regresión entre la variable de estudio y los vectores propios asociados a cada matriz de contigüidad. Dray et al. (2006) consideran un conjunto amplio de matrices de pesos espaciales y se selecciona la que presente menor criterio de Akaike.

Algunos de los criterios utilizados para hallar la matriz de pesos basada en el análisis de coordenadas principales fueron: la **Triangulación Delaunay** que considera como vecino de un polígono a aquellos que pueden estar conectados mediante triángulos; así, cada polígono debe poseer el menos un vértice que conecte. El **Criterio Gabriel** que define que la distancia entre los polígonos debe ser menor o igual a la mínima distancia entre dos vértices de dos polígonos adyacentes. Los **Vecinos Relativos** que propone que una arista pertenece al grafo de vecinos relativos si sus extremos son vecinos relativos entre sí. La **Esfera de influencia** instala un conjunto de puntos finitos y un radio entre un polígono y su vecino más cercano; de esta forma se genera una circunferencia centrada en los polígonos y bajo este criterio las unidades se consideran vecinos si las circunferencias de cada una de éstas se intersecan en al menos dos puntos. Y los **K-Vecinos más cercanos** que considera relaciones de distancia no simétricas bajo matrices de pesos estandarizadas. El criterio de la vecindad de un polígono sobre $k$-vecinos más cercanos se define a partir de distancias euclidianas, las demás $k+1$,$k+2$,....,$k+n$ se consideran no vecinas del polígono. Una vez generadas las matrices con los criterios previamente mencionados y los criterios clásicos de contigüidad Efecto reina y efecto Torre se determinó cuál era la mejor para el estudio basándose en el menor valor del AIC para el periodo de estudio.

Este proceso se realizó en el software R usando la librería *Adespatial* y *spdep*. Ambas librerías arrojaron los valores del AIC con su nivel de significancia y permitieron la visualización de los criterios. Se calcularon los vectores propios de cada uno de los criterios de contigüidad evaluados bajo la metodología PCNM y los correspondientes estadísticos estandarizados de Moran. También, se generaron los gráficos de cada criterio, con el fin de inferir con mayor precisión qué relaciones de contigüidad representan la autocorrelación de los polígonos vecinos y se establecieron tablas comparativas entre los criterios para cada año.

#### 5.3.5.3 Análisis Global y Local

Con una matriz de pesos espaciales definida, ya es posible comprobar si existe autocorrelación espacial en una variable, así como determinar el signo de la relación. Los contrastes son de dos tipos, los globales y los locales. Los primeros indagan sobre la hipótesis de estacionariedad de una variable sobre la totalidad de la muestra, mientras que las pruebas locales verifican estacionariedad intrarregional.

En cuanto a los **estadísticos globales** para detectar autocorrelación espacial se utilizó el contraste Global I de moran que según Moreno y Vayá (2000) es una medida de la correlación entre cada región y sus vecinas. Si su estadístico de prueba enseña valores positivos y significativos significa la concentración de datos similares en regiones vecinas; por el contrario, si se muestran valores negativos y significativos revela que en las regiones vecinas existe disparidad. Otros criterioss utilizado para comprobar la autocorrelación espacial fueron el test de Geary y el de Getis y Ord. Los test mencionados arrojan un estadístico junto con su P-Valor que permite determinar el grado de autocorrelación global de la SMR a través de las localidades de Bogotá. Gráficos de dispersión de Moran, para cada variable, también fueron realizados con el fin de encontrar la presencia de autocorrelación espacial en la región de estudio en el periodo 2004-2014.

Los criterios locales permiten detectar aglomeraciones de unidades vecinas con valores extraordinariamente altos o bajos con relación en el promedio, es decir, autocorrelación significativa en solo una sección más pequeña de la región total de estudio. Los **estadísticos locales** utilizados fueron el I de Moran Local y el G de Getis y Ord. El uso de estos estadísticos locales permitió identificar en un mapa los clusters regionales y los valores atípicos de SMR en el espacio que presenta el I de Moran. De igual manera, se realizó el análisis LISA que descompone el índice global de autocorrelación (I de Moran) y verifica en qué forma contribuye cada unidad espacial a la formación del valor general, lo cual permite estimar un valor de significancia para cada clúster o agrupación conformado de valores similares de cada unidad espacial y sus vecinos. Para la variable independiente se pudo observar qué localidades estaban rodeadas de valores similares y cuáles de valores diferentes de SMR.

### 5.3.6. Construcción Modelo Autorregresivo Espacial

Mediante el análisis exploratorio se determina la existencia de autocorrelación espacial del SMR a nivel de localidad, sin embargo éste no permite decidir sobre la estructura de esta dependencia espacial; si se incorpora como rezago, en el término del error o en rezagos espaciales de otras variables explicativas.

A partir de la estimación por mínimos cuadrados ordinarios de la mortalidad explicada por las variables exógenas se realizó un modelo clásico de regresión que representará el comportamiento de SMR en función de los factores ambientales,

socioeconómicos, conductuales y asistenciales médicos. Se generó la estimación de diversos modelos de regresión lineal para obtener un modelo que cumpliese con las especificaciones y supuestos claves en la formulación de un modelo regresivo. Test estadísticos como la prueba de significancia global, prueba de significancia de los coeficientes, prueba de normalidad, prueba de heteroscedasticidad, prueba de multicolinealidad y prueba de autocorrelación en los residuos se realizan con el fin de determinar qué tan bueno y preciso es el modelo y si este representa las relaciones entre la variable independiente y sus dependientes.

Es de suma importancia la prueba de autocorrelación espacial en los residuos por medio del estadístico del I de Moran pues de esto dependerá la modelización de las variables; si el resultado arroja autocorrelación espacial es posible la búsqueda de una estructura de modelo espacial que capte el efecto del retardo espacial de la variable dependiente (modelo de rezago espacial) o bien, el retardo espacial en la perturbación aleatoria (modelo de error espacial); si esto no sucede, se deben buscar alternativas no paramétricas y semiparamétricas que permitan identificar y modelar la dependencia espacial de las variables.

### 5.3.7. Selección Modelo Autorregresivo Espacial

El siguiente paso consiste en seleccionar el modelo espacial más adecuado para cada año. Para ello se aplica el test de multiplicadores de LaGrange (LM) para cada uno de los posibles modelos espaciales. Esta prueba determina cuál puede ser el mejor modelo que personifica la dependencia espacial y la relación de las variables.

El test muestra los diferentes tipos de modelos espaciales (rezago espacial o error espacial) asociados a un parámetro y un P-valor de significancia. Se escoge el modelo cuyo valor P es el más significativo pues logra apreciar los casos de autocorrelación espacial, sea esta residual o sustancial o ambas, en las localidades será mucho mejor que los demás modelos, teniendo como consecuencia una mejor estimación de los parámetros y un modelo consistente. Para este cálculo debe haber una correcta especificación de la matriz de pesos espaciales y del modelo clásico de regresión.

Finalmente, se obtiene un modelo econométrico espacial para cada año que consolida el comportamiento de la tasa de mortalidad por la neumonía. Si bien, los multiplicadores de LaGrange determinan en gran medida cuál modelo es el más apropiado, es necesario apoyarse en los resultados del anisáis global y local del AEDE. Si la SMR se relaciona puntualmente entre localidades de Bogotá, podría tratarse de un modelo espacial SMA; pero si el comportamiento de la variable independiente es significativo a través de todo el territorio Bogotano se podría hablar de un modelo SAR. Estos análisis junto con las pruebas de autocorrelación espacial permitirán consolidar un buen modelo evitando falsas deducciones y errores posteriores con el modelo espacial seleccionado.

### 5.3.8. Validación Modelo Autorregresivo Espacial

Una vez se han establecido los modelos autorregresivos espaciales, éstos se deben someter de nuevo a validaciones diagnósticas formales como la multicolinealidad, significancia y heteroscedasticidad. De la misma manera, se deben analizar los residuos en busca de autocorrelaciones y normalidad. Si después de estos análisis los modelos aún son consistentes, se determinarán como los modelos más óptimos para modelar la SMR por neumonía a través del espacio junto con sus variables explicativas para las localidades de Bogotá. Este proceso debe realizarse para cada uno de los años propuestos y analizado en conjunto para identificar posibles relaciones temporales entre el evento.

Se debe reparar detalladamente en las variables que sobreviven el proceso del modelamiento, pues éstas, según su nivel de significancia están directamenente asociadas a la variable respuesta. Incrementos o disminuciones en estas variables estimularan la perturbación negativa o positiva de la SMR de muertes por neumonía y cómo esta perturbación es transmitida a las localidades vecinas.

### 5.3.9. Interpretación del Modelo

Con los modelos previamente establecidos y verificados ya es posible modelar la tasa de mortalidad por neumonía para cada año en función de las variables sociales, económicas, ambientales y de asistencia médica que permanecieron a través de la aplicación de pruebas y estadísticos para un modelo autorregresivo espacial. Se describe minuciosamente cómo y por qué las variables repercuten en el comportamiento de la variable endógena a través de la zona de estudio y en qué proporción los cambios relativos de éstas repercuten en las tasas de mortalidad en una localidad específica y en sus vecinas.

### 5.3.10. Generación de Impactos

Los efectos o impactos producidos por el modelo dependerán en primer lugar del tipo de modelo que se estableció para cada año y de las variables que lo representan. Como su nombre lo indica, los impactos enseñan los cambios relativos en la variable respuesta en función de las variables explicativas a través del espacio, es decir, que si el valor de una variable independiente aumenta, la variable respuesta debería reaccionar positiva o negativamente no solo para la localidad en análisis, sino también para aquellas vecinas.

Este proceso se realiza para verificar la calidad del modelo autorregresivo y para identificar qué variables tienen mayor radio e intensidad de alteración para la tasa de mortalidad y su distribución en las localidades de la capital colombiana.

### 5.3.11. Diagnóstico de la Neumonia

Por último, se realizó un análisis detallado de la mortalidad de la enfermedad para el periodo de tiempo establecido basado en los modelos autorregresivos, las variables y la población afectada. Se hizo un énfasis en los factores que resultaron en los modelos y se explicó por qué, para el periodo de tiempo y el espacio estudiado, son relevantes. Así mismo, se sugirieron medidas de control, prevención y reparación de las variables explicativas con el fin de mitigar la incidencia de la enfermedad en Bogotá en periodos posteriores.

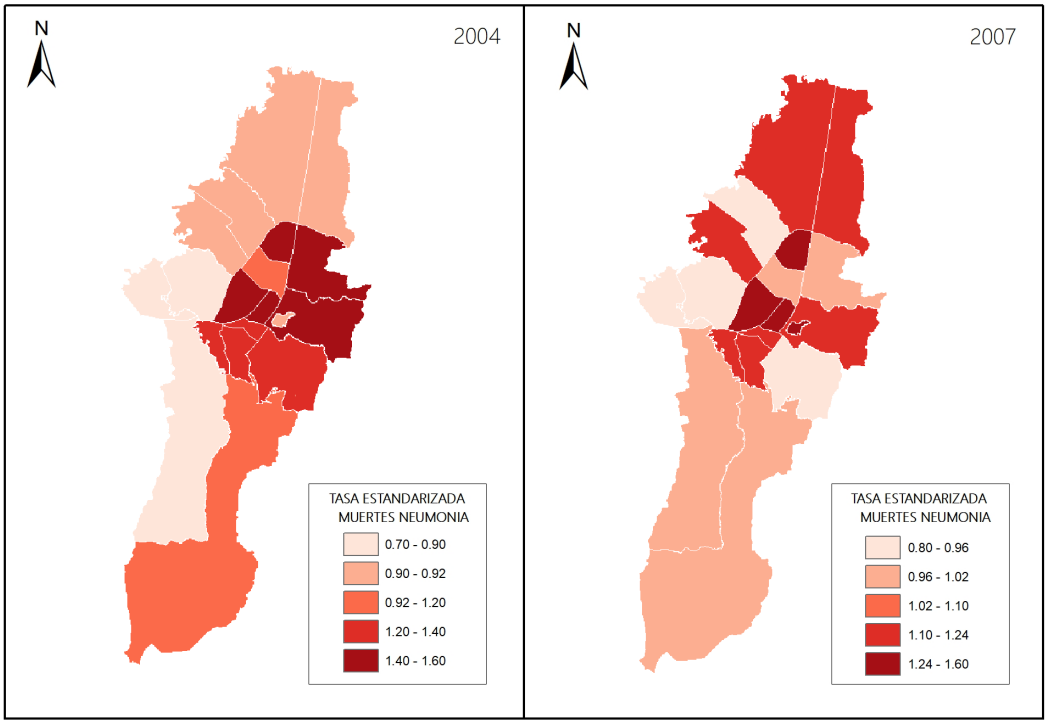
# 6. RESULTADOS

## 6.1. Ánalisis Exploratorio Espacial (AEDE)

Debido a que en el fenómeno estudiado existen datos espaciales con variación discreta, en otras palabras, datos lattice en localizaciones irregulares (Schabenberger y Gotway, 2017), se procedió a realizar el análisis exploratorio para lograr entender de manera básica la información, mediante métodos de estadística descriptiva y análisis exploratorio multivariado para hallar las relaciones existentes entre las variables respuesta y explicativas (Skinner et al., 2002).

Con el fin de analizar la distribución geográfica de la variable estudiada se realizan los mapas de cuantiles que agrupan el mismo número de valores de datos a cada clase. En las figuras 3 y 4 se encuentran los mapas de cuantiles para la variable endógena de los años de estudio, las localidades con color claro representan los valores mínimos de la variable y en color rojo oscuro los valores más altos.

Figura 3: Distribución espacial de la SMR por neumonía en las localidades de Bogotá: 2004 - 2007.



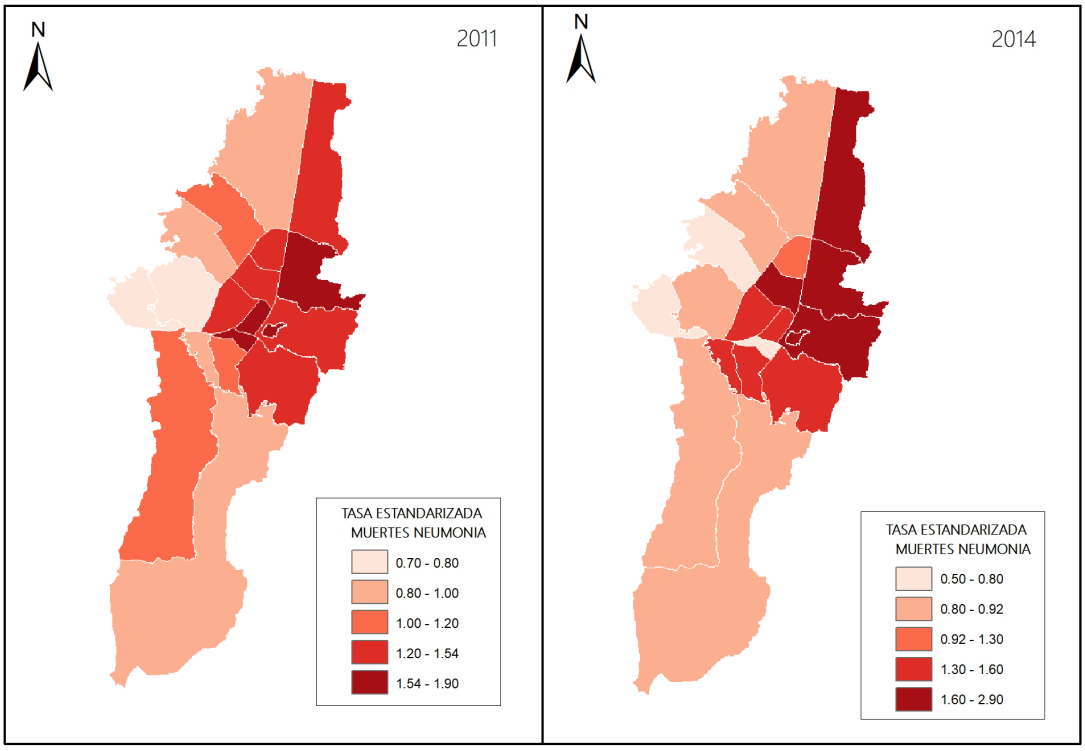

*Fuente: Elaboración propia*

Para el año 2004 los valores más altos de la SMR varían entre 1.4 a 1.6 que representan las localidades de Barrios Unidos, Chapinero, Puente Aranda, Los mártires y Santa Fé, y las tasas más bajas están presentes en las localidades de Bosa y Ciudad Bolívar.

En el mapa del año 2007 se observa que las localidades con tasas de mortalidad altas son Candelaria, Los Mártires, Puente Aranda y Barrios Unidos; estos valores varían entre 1.24 a 1.6, y las menos afectadas son Engativá, Kennedy, San Cristobal y Bosa. Los valores de la tasa de mortalidad estandarizada para estos dos años son similares alcanzado un máximo de 1.6 donde localidades como Barrios Unidos, Los Mártires y Puente Aranda son las más afectados por la variable respuesta y las regiones menos afectadas son Bosa y Kennedy.

También, se observa que Usaquen, Suba, Fontibon y Ciudad Bolívar presentaron un incremento en la tasa de mortalidad en el 2007 y Tunjuelito, Rafael Uribe Uribe y Antonio Nariño permanecen con valores de tasas que se encuentran en rangos de medio-altos con tasas superiores a 1.2.

Figura 4: Distribución espacial de la SMR por neumonia en las localidades de Bogotá: 2011 - 2014.




*Fuente: Elaboración propia*

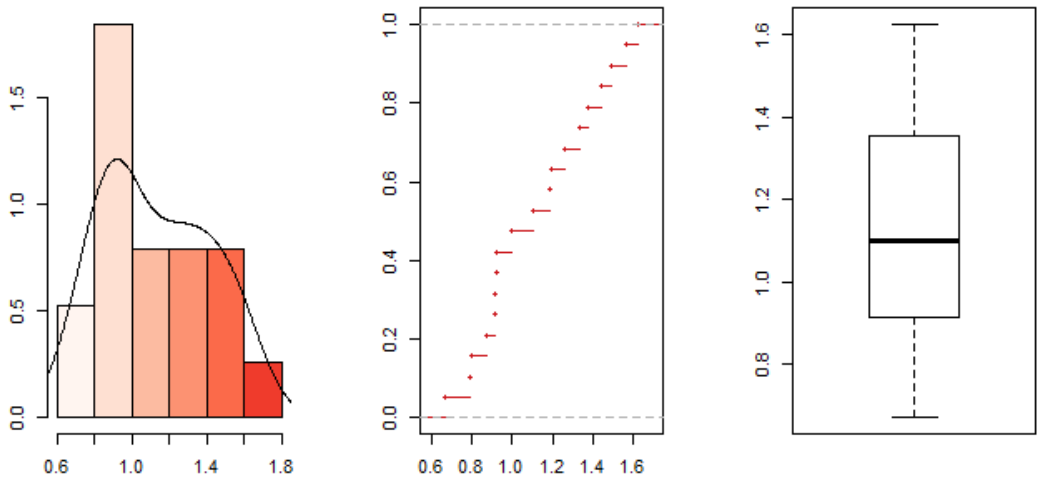
De igual manera, para el año 2011 continúa afectada la localidad de Los Mártires y

aparecen nuevas regiones con intervalos altos como Chapinero, Candelaria y Antonio Nariño; el valor máximo que alcanza la SMR en este año es de 1.9, y las zonas con valores mínimos son la localidad de Bosa y Kennedy.

En el 2014 los valores más grandes varían entre 1.6 a 2.9, donde se encuentran las regiones de Usaquen, Chapinero, Santa Fé, Candelaria y Teusaquillo, y en el intervalo mínimo de 0.6 - 0.9 Bosa y Fontibón. En estos dos años continúan con valores de tasa medio-bajos las localidades de Suba y Usme, además muestra que la tasa de mortalidad disminuyó en las regiones de Ciudad Bolívar, Rafael Uribe Uribe, Los Mártires y Antonio Nariño, y permanece con valores bajos de tasa de mortalidad la localidad de Bosa.

Finalmente, las regiones más afectadas en el periodo de estudio son Chapinero, Los Mártires, Santa Fé y Puente Aranda teniendo tasas de mortalidad altas y medio-altas con valores superiores a 1.1. En el caso de Usaquén y Teusaquillo, se encuentra que en un principio la tasa presentaba valores bajos para el año 2004 y fue elevándose hasta obtener valores altos en el año 2014. La regiones de Bosa y Kennedy tuvieron los valores más bajos de la variable en los cuatro años siendo estos menores a 0.9.

Figura 5: Histograma, Gráfico Q-Q y Boxplot de la SMR: 2004.



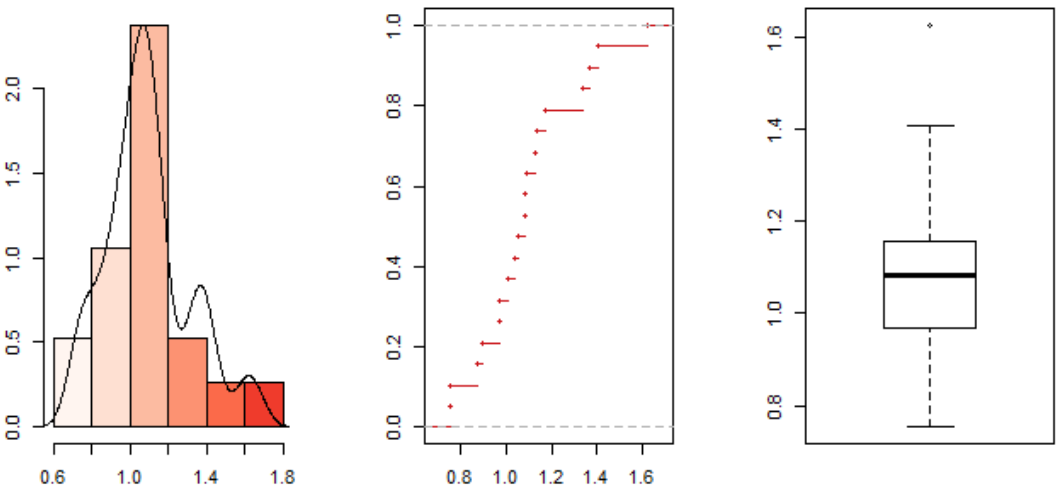

*Fuente: Elaboración propia*

Para examinar si la variable respuesta sigue una distribución normal, se analizan los diferentes gráficos como histograma, diagrama de distribución empírica y diagrama de caja de la variable para cada año. También, se expone en la Tabla 3 un resumen de las medidas de tendencia central de la variable endógena, por cada periodo en estudio, que describe el comportamiento del conjunto de los datos.

La SMR en el año 2004 presenta asimetría unimodal a la derecha ya que los valores tienden a reunirse a la izquierda de la media (Figura 5). Tiene un valor de la media

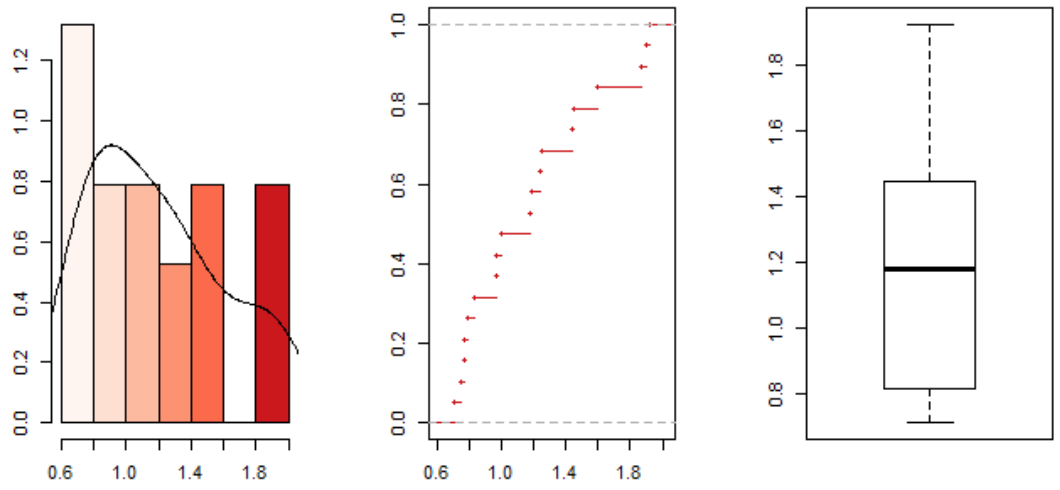
de 1.13 y mediana 1.1, el valor de la kurtosis muestra una concentración de valores en torno a la media. Se observa que en gráfico Q-Q los puntos se ajustan de manera aproximada a una recta, donde se concluye que la distribución tiende a ajustarse a una normal. El Boxplot expone que la tasa de mortalidad estandarizada no tiene datos atípicos, pero esta caja no es símetrica respecto a la mediana ya que los datos se ubican encima de ésta.

Figura 6: Histograma, Gráfico Q-Q y Boxplot de la SMR: 2007.




*Fuente: Elaboración propia*

Figura 7: Histograma, Gráfico Q-Q y Boxplot de la SMR: 2011.



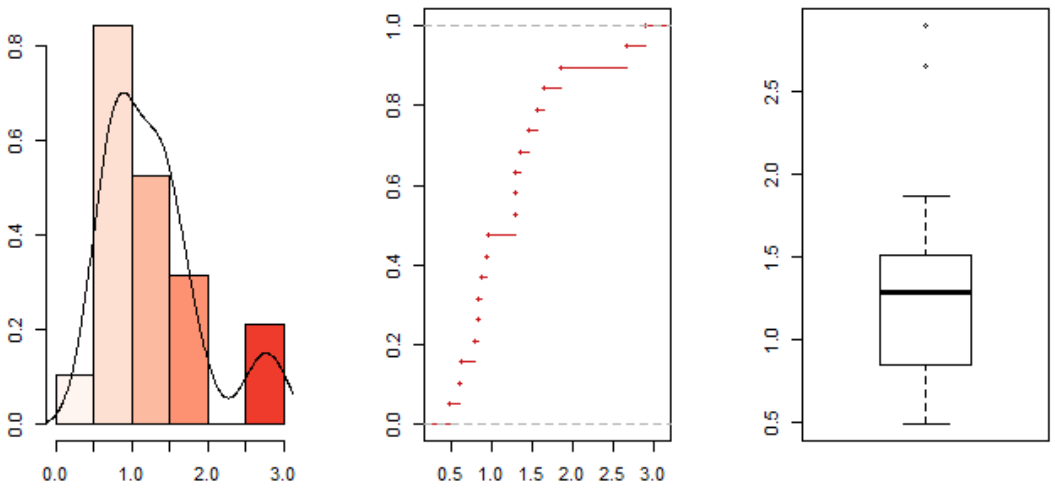

*Fuente: Elaboración propia*

El histograma del año 2007 presenta un pico alto en la tercera clase, se identifica

que gran parte de la información se agrupa en el lado izquierdo de éste provocando que los datos no se distribuyan uniformemente y muestre una distribución unimodal asimétrica positiva. Los datos tienden a ubicarse antes de la mediana como lo muestra el Boxplot (Figura 6) lo que genera un sesgo en la información. Se evidencia una media de 1.09 y un coeficiente de variación de 20.42 % que indica la presencia de datos atípicos. La tasa de mortalidad alcanza el mismo valor máximo en 2004 y 2007 siendo este de 1.63.

De la misma manera, para el año 2011 (Figura 7) los datos se agrupan al lado derecho del histograma evidenciando que la distribución es asimétrica a la derecha. En este año hubo un aumento de la tasa de mortalidad ya que se evidencia un valor máximo de 1.92. La mayoría de la información se ubica antes de la mediana, donde son más las localidades que presentan tasas con valores bajos que las que se encuentran más aquejadas por mortalidad a causa de la neumonía. El diagrama de caja está simétrico para el primer y tercer cuartil pero no para sus extremos.

Figura 8: Histograma, Gráfico Q-Q y Boxplot de la SMR: 2014.




*Fuente: Elaboración propia*

Por último, se encuentra que la distribución del histograma del año 2014 es unimodal con una asimetría positiva donde los valores tienden a reunirse más a la izquierda que a la derecha de la media. El coeficiente de variación es el más grande que en el resto de años estudiados siendo este valor de 50.87 % lo que significa que los datos están muy dispersos. La tasa alcanza el valor más pequeño y más grande en el 2014 que corresponden a 0.49 y 2.89, respectivamente. En el gráfico Q-Q se observa que los datos no se ajustan uniformemente ni se aproximan a una normal. En el Boxplot se evidencia que existen valores atípicos y la caja no se encuentra simétrica respecto a la mediana.

Tabla 3: Resumen de medidas estadísticas de la variable SMR para cada año.

| Medida estadística | 2004 | 2007 | 2011 | 2014 |
|---|---|---|---|---|
| Mínimo | 0.67 | 0.75 | 0.71 | 0.49 |
| Máximo | 1.63 | 1.63 | 1.92 | 2.89 |
| Media | 1.13 | 1.09 | 1.19 | 1.28 |
| Mediana | 1.10 | 1.08 | 1.18 | 1.28 |
| Desviacion estándar | 0.29 | 0.22 | 0.41 | 0.65 |
| Desviacion mediana | 0.23 | 0.12 | 0.35 | 0.40 |
| Asimetría | 0.22 | 0.57 | 0.53 | 1.08 |
| Kurtosis | 1.63 | 2.82 | 1.88 | 3.40 |
| Coef. Var. Promedio( %) | 25.53 | 20.42 | 34.26 | 50.87 |
| Coef. Var. Mediana( %) | 20.59 | 10.62 | 29.30 | 30.82 |

*Fuente: Elaboración propia en R*

### 6.1.1. Selección de Matriz de Pesos Espaciales

La matriz W de pesos espaciales introduce la dependencia entre las unidades y especifica los modelos espaciales ya que es un atributo característico de los modelos econométricos que usan datos georeferenciados. Para escoger la mejor matriz se analizan distintos criterios de contigüidad que permitirán determinar qué regiones tienen una influencia particular sobre la región en consideración expresada en nociones topológicas (Anselin, 1988b); si no se analizan diferentes tipos de contigüidad no se tendría en cuenta el principio de alotopía que establece que "a menudo lo que ocurre en una región se encuentra relacionado con otro fenómeno localizado en otras partes diferentes y remotas del espacio"(Ancot et al., 1982).

Las matrices de contigüidad comparadas son de criterios basados en contigüidad; Reina y Torre, basados en geometrías; Triangulación Delaunay, Esfera de Influencia, Gabriel, Vecinos relativos y basados en n-vecinos más cercanos con n=1, n=2 y n=4. Se selecciona la matriz de contigüidad que tenga el menor valor del coeficiente del criterio de Información de Akaike (AIC, por sus siglas en inglés).

En las tablas 4 y 5 se encuentran los resultados obtenidos para cada criterio del I de Moran y del AIC, el test del I de Moran se realiza para analizar la presencia de autocorrelación espacial de acuerdo a la matriz que posea el menor AIC.

Para el año 2004 se selecciona la matriz de contigüidad estilo reina que posee un AIC de -68.7727, se observa que la variable endógena para este año tiene una estructura de dependencia espacial ya que el estadístico del I de Moran es significativo y su p valor es menor al 5 %. En el año 2007 el menor valor del coeficiente AIC lo posee el de kn-vecinos cercanos con kn=2, pero al evaluar la dependencia espacial de la variable endógena se encuentra que sigue una distribución aleatoria y esta matriz no es significativa al elaborar los modelos espaciales. Por esta razón, se selecciona el

criterio de kn-vecinos cercanos con kn=1, en donde se observa que tiene el estadístico con valor significativo en el I de Moran.

Tabla 4: Criterios de Selección de matriz W 2004-2007

| Año | 2004 | | | | 2007 | | | |
|---|---|---|---|---|---|---|---|---|
| Criterio | I | P-valor | Z(I) | AIC | I | P-valor | Z(I) | AIC |
| Reina | 0.2341 | 0.0233 | 1.9897 | -68.7727 | 0.0986 | 0.1342 | 1.1068 | -58.9756 |
| Torre | 0.2864 | 0.012 | 2.2547 | -61.1534 | 0.0833 | 0.1692 | 0.9572 | -61.1347 |
| .T. Delaunay | 0.1071 | 0.0966 | 1.3011 | -61.1935 | 0.0888 | 0.1136 | 1.2075 | -66.0532 |
| Gabriel | 0.2166 | 0.0635 | 1.526 | -55.853 | 0.1339 | 0.1334 | 1.1106 | -63.0706 |
| Esfera | 0.2772 | 0.0271 | 1.9246 | -53.7099 | 0.1313 | 0.1293 | 1.1296 | -71.8989 |
| V. Relativos | 0.3409 | 0.0276 | 1.9161 | -55.7115 | 0.15008 | 0.1494 | 1.0392 | -71.2933 |
| Kn1 | 0.357 | 0.0829 | 1.3852 | -60.7074 | 0.5802 | 0.0128 | 2.2304 | -67.0918 |
| Kn2 | 0.2499 | 0.0517 | 1.6282 | -67.4836 | 0.2211 | 0.0617 | 1.5401 | -83.3805 |
| Kn4 | 0.1724 | 0.0309 | 1.8669 | -52.6594 | 0.0969 | 0.0963 | 1.3029 | -66.6514 |

*Fuente: Elaboración propia en R*

El menor valor del AIC en el año 2011 lo tiene el criterio basado en geometría de Triangulación Delaunay, muestra también el estadístico del I de Moran con mayor significancia teniendo un p valor muy cercano a cero, comprobándose la presencia de asociación espacial. La matriz de contigüidad seleccionada del año 2014 es la de kn-vecinos cercanos (kn=2) teniendo un AIC de -42.9033, al igual que en el 2007 la tasa de mortalidad estandarizada no presenta autocorrelación espacial con esta matriz seleccionada; pero fue la más indicada para realizar los modelos espaciales, ya que aparecieron más variables significativas que las presentadas en los modelos obtenidos a partir de los criterios de contigüidad kn=4 y Gabriel, que muestran p-valores significativos en el I de Moran.

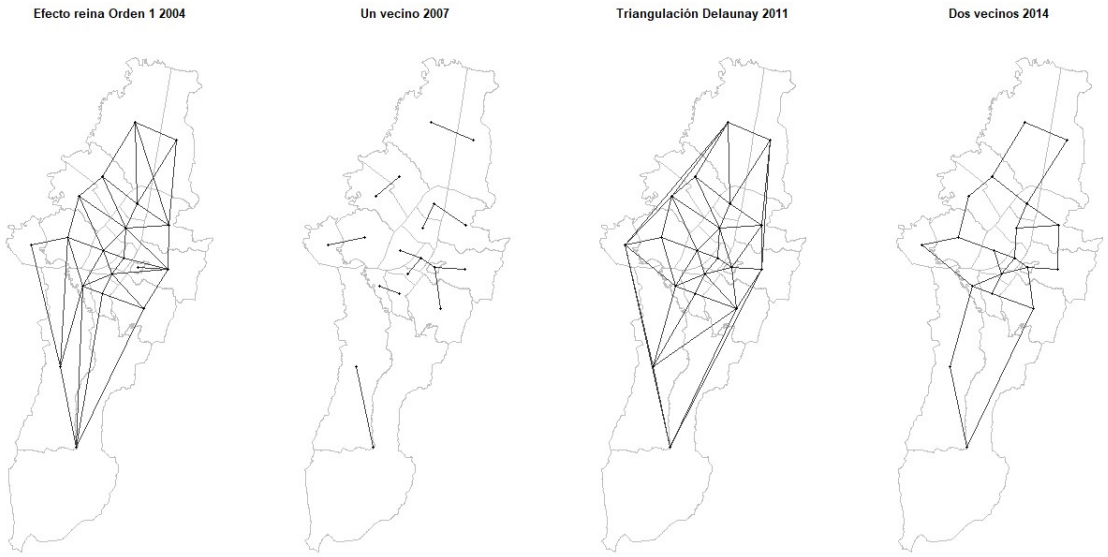
Tabla 5: Criterios de Selección de matriz W 2011-2014

| Año | 2011 | | | | 2014 | | | |
|---|---|---|---|---|---|---|---|---|
| Criterio | I | P-valor | Z(I) | AIC | I | P-valor | Z(I) | AIC |
| Reina | 0.2289 | 0.0243 | 1.972 | -46.1606 | 0.1995 | 0.0304 | 1.8748 | -31.8595 |
| Torre | 0.2777 | 0.0132 | 2.2177 | -48.8688 | 0.2479 | 0.0161 | 2.1409 | -25.3955 |
| T. Delaunay | 0.4473 | $2.46x10^5$ | 4.0594 | -62.9298 | 0.2841 | 0.0018 | 2.9095 | -23.5326 |
| Gabriel | 0.4773 | 0.0012 | 3.0147 | -57.8476 | 0.3381 | 0.009 | 2.3631 | -27.438 |
| Esfera | 0.4024 | 0.0037 | 2.6721 | -55.7801 | 0.2725 | 0.0211 | 2.0297 | -24.9781 |
| V. Relativos | 0.3775 | 0.0173 | 2.1123 | -59.3651 | 0.1781 | 0.1133 | 1.2091 | -35.2253 |
| Kn1 | 0.6134 | 0.0117 | 2.266 | -56.3651 | 0.1702 | 0.2087 | 0.8107 | -20.2536 |
| Kn2 | 0.4107 | 0.006 | 2.5075 | -58.4859 | 0.0851 | 0.2114 | 0.8015 | -42.9033 |
| Kn4 | 0.3941 | 0.0001 | 3.7143 | -59.0707 | 0.1955 | 0.0141 | 2.1928 | -29.2983 |

*Fuente: Elaboración propia en R*

En la figura 9 se muestran las gráficas de los criterios seleccionados para cada año, observando que Efecto Reina (2004) y Triangulación Delaunay (2011) presentan la mayor cantidad de links, debido a que en el primer criterio se comparte un lado o vértice y en el segundo se considera vecino de una región aquellas regiones que se puedan conectar a partir de trángulos, provocando que se relacionen la mayoría de los polígonos que están visibles entre sí por medio de la distancia euclidiana, generando relaciones entre las localidades más lejanas. A partir del gráfico de kn1 se determina la distancias mínima de las conectividades de los vecinos más cercanos.

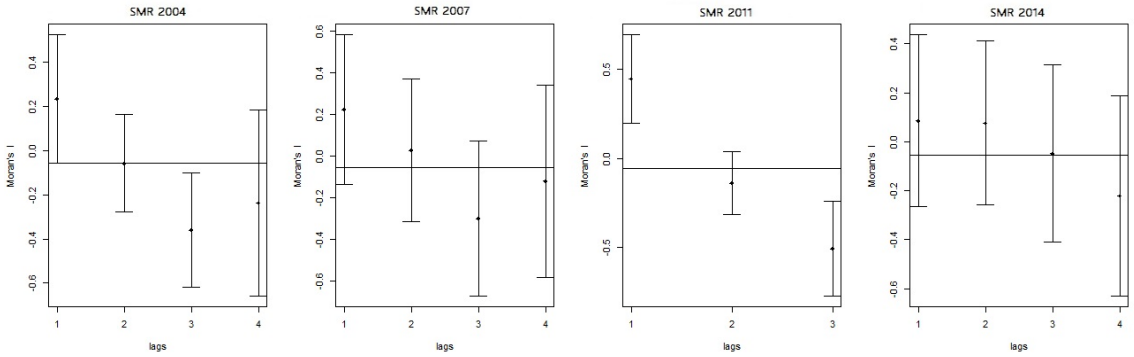
Figura 9: Gráficos criterios seleccionados por año




*Fuente: Elaboración propia*

Con el fin de encontrar estructuras de autocorrelación espacial en la variable con la matriz de contigüidad seleccionada, se analizan los correlogramas para cada año (Figura 10) éstos permiten examinar patrones de autocorrelación espacial en los datos o modelar residuales, mostrando la correlación entre parejas de observaciones espaciales cuando se aumenta la distancia o el rezago entre ellos. Se encuentra que en todos los años ningún rezago es significativo, excepto para el 2011 donde es significativo el de primer orden.

Figura 10: Correlogramas




*Fuente: Elaboración propia*

### 6.1.2. Ánalisis Exploratorio Global

Una vez seleccionada la matriz de pesos espaciales para cada año se procede a analizar la dependencia espacial de cada variable (endógena y exógenas) en la región por medio de los contrastes I de Moran, C de Geary y G de Getis y Ord con el fin establecer la presencia de un efecto de difusión de las variable e identificar las relaciones espaciales que presentan de manera global. Cada test arrojó resultados con una probabilidad asociada a un intervalo de confianza del 95 %. Además, este análisis permite determinar si la variable respuesta se ha transmitido sobre las regiones vecinas que en un principio presentaban bajos niveles de mortalidad.

En la tabla 6 se encuentran los valores obtenidos para cada test del año 2004 realizado con la matriz de pesos espaciales efecto reina. Se evidencia que la variable endógena y variables explicativas como índice de desempleo, contaminación, nutrición, índice de necesidades básicas insatisfechas, índice de calidad de vivienda, índice de calidad de vida y número de IPS presentan esquemas de autocorrelación espacial positiva cuando se aplica el I de moran, estas variables poseen un p-valor menor a 0.05 y un estadístico significativo, que corresponden a una concentración de valores similares en las localidades. Sin embargo, en el test G de Getis y Ord que pretende confirmar la presencia de autocorrelación espacial, las variables de índice de desempleo, nutrición y necesidades básicas insatisfechas aparecen con estadísticos no significativos. Para las demás variables se acepta la hipótesis nula y se concluye que no están relacionadas espacialmente obedeciendo esquemas de aletoriedad espacial.

Tabla 6: Contrastes globales de variables endógena y exógenas año 2004

| Test | I de Moran | | | C de Geary | | | G de Getis y Ord | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | I | P value | Z(I) | C | P value | Z(C) | G | P value | Z(G) |
| SMR | 0.2341 | 0.0233 | 1.9897 | 0.718 | 0.0258 | -1.9454 | 0.269 | 0.0068 | 2.4654 |
| IDD | 0.4363 | 0.0002 | 3.4428 | 0.5908 | 0.0027 | -2.7774 | 0.2442 | 0.5566 | -0.1423 |
| TEM | -0.0883 | 0.5922 | -0.233 | 1.0326 | 0.5867 | 0.219 | 0.2483 | 0.2324 | 0.7308 |
| CPM | 0.4005 | 0.0008 | 3.1373 | 0.5558 | 0.0011 | -3.0599 | 0.2755 | 0.0217 | 2.0189 |
| DEP | -0.0148 | 0.3867 | 0.288 | 0.9462 | 0.3588 | -0.3617 | 0.246 | 0.4628 | 0.0934 |
| NUT | 0.5538 | $1.4x10^5$ | 4.1841 | 0.4282 | $3.9x10^5$ | -3.9462 | 0.2533 | 0.3275 | 0.4469 |
| NBI | 0.4822 | $9.7x10^5$ | 3.7247 | 0.5506 | 0.001 | -3.0776 | 0.2666 | 0.2141 | 0.7922 |
| CVV | 0.416 | 0.0005 | 3.2414 | 0.5591 | 0.0011 | -3.0397 | 0.2477 | 0.0319 | 1.8524 |
| CVD | 0.6043 | $1.9x10^6$ | 4.6165 | 0.3899 | $1.7x10^5$ | -4.1427 | 0.248 | 0.0623 | 1.535 |
| ESC | -0.0589 | 0.5092 | -0.023 | 0.9898 | 0.472 | -0.0703 | 0.246 | 0.4334 | 0.1676 |
| IPSE | 0.1675 | 0.0058 | 2.5216 | 0.6661 | 0.0326 | -1.8438 | 0.5708 | 0.0314 | 1.8603 |
| VAC | -0.1317 | 0.7025 | -0.5316 | 1.0314 | 0.5847 | 0.2138 | 0.248 | 0.3701 | 0.0878 |

*Fuente: Elaboración propia en R*

Para el año 2007 (Tabla 7) las variables que indican autocorrelación positiva en el espacio según el I de moran son SMR, CPM, NUT, NBI, CVV, y CVD, la variable ESC no presenta un estadístico significativo pero su p-valor es menor a 0.05. Por esta razón, se procede a evaluar los resultados del test C de Geary que confirma que esta variable no presenta asociación espacial. En el contraste G de Getis y Ord se rechaza la hipótesis nula para tres variables SMR, DEP y NUT, y para las demás se concluye que presentan completa aletoriedad espacial.

En el año 2011 la variable endógena está espacialmente correlacionada en todos los test realizados, las variables explícativas de índice de desempleo, contaminación, nutrición, necesidades básicas insatisfechas, calidad de vida, calidad de vivienda y número de IPS muestran valores agrupados en el espacio. En los resultados obtenidos a partir del contraste C de Geary, aparece autocorrelacionada el índice de prácticas deportivas; pero en el G de Getis y Ord solo es posible encontrar concentración de valores similares en índice de desempleo y número de IPS.

Tabla 7: Contrastes globales de variables endógena y exógenas año 2007

| Test | I de Moran | | | C de Geary | | | G de Getis y Ord | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | I | P value | Z(I) | C | P value | Z(C) | G | P value | Z(G) |
| SMR | 0.5802 | 0.0128 | 2.2304 | 0.7425 | 0.2045 | -0.8255 | 0.0618 | 0.0031 | 2.7357 |
| IDD | 0.3849 | 0.0628 | 1.5315 | 0.4366 | 0.0342 | -1.8218 | 0.0579 | 0.2063 | 0.8192 |
| TEM | 0.306 | 0.104 | 1.2592 | 0.7027 | 0.1685 | -0.9601 | 0.055 | 0.6869 | -0.487 |
| CPM | 0.6912 | 0.0057 | 2.5258 | 0.3077 | 0.0106 | -2.3021 | 0.0563 | 0.3893 | 0.2812 |
| DEP | 0.3848 | 0.0572 | 1.5781 | 0.6255 | 0.1193 | -1.1785 | 0.0584 | 0.0486 | 1.6583 |
| NUT | 0.4473 | 0.0445 | 1.7005 | 0.5173 | 0.0541 | -1.6055 | 0.0687 | 0.02 | 2.0524 |
| NBI | 0.5389 | 0.0194 | 2.065 | 0.3068 | 0.0124 | -2.2438 | 0.0672 | 0.1038 | 1.26 |
| CVV | 0.6054 | 0.0117 | 2.2659 | 0.337 | 0.0148 | -2.1741 | 0.0553 | 0.77 | -0.7388 |
| CVD | 0.6095 | 0.0117 | 2.2651 | 0.2536 | 0.0068 | -2.4639 | 0.0556 | 0.4193 | 0.2036 |
| ESC | 0.4508 | 0.0431 | 1.7147 | 0.6522 | 0.1241 | -1.1548 | 0.0547 | 0.8999 | -1.281 |
| IPSE | -0.0503 | 0.4896 | 0.0259 | 0.7196 | 0.2294 | -0.7407 | 0.0527 | 0.5302 | -0.0757 |
| VAC | 0.0563 | 0.3399 | 0.4128 | 1.1535 | 0.6815 | 0.472 | 0.0531 | 0.946 | -1.6069 |

*Fuente: Elaboración propia en R*

La tasa de mortalidad estandarizada en el año 2014 no obedece un esquema de autocorrelación espacial debido a que los resultados en los diferentes contrastes practicados no son significativos ni presentan p valores menores a 5 %. Las variables exógenas como IDD, TEM, DEP, NBI, CVV, CVD y VAC indican concentración de valores parecidos en regiones vecinas según el I de Moran. Sin embargo, en el contraste G de Getis y Ord en todas las variables existe ausencia de autocorrelación espacial. En el contraste C de Geary la contaminaciòn aparece con un p-valor menor al 5 % pero su estadístico no es significativo, para la variable índice de nutrición en este test se confirma la presencia de asociación espacial. En cuanto a la calidad de vivienda, se acepta la hipòtesis nula debido a que su estadístico no es significativo.

Por otra parte, los resultados obtenidos exponen que la variable respuesta sigue una estructura de completa aletoriedad espacial solo en el año 2014, para el resto de años estudiados se concluye que los valores de la tasa en una localidad específica no dependen sólo de sí mismos sino también de la ocurrencia en regiones vecinas.

Tabla 8: Contrastes globales de variables endógena y exógenas año 2011

| Test | I de Moran | | | C de Geary | | | G de Getis y Ord | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | I | P value | Z(I) | C | P value | Z(C) | G | P value | Z(G) |
| SMR | 0.4473 | $2.4x10^5$ | 4.0594 | 0.5312 | $5.2x10^5$ | -3.878 | 0.3077 | 0.0092 | 2.3568 |
| IDD | 0.3012 | 0.0015 | 2.9564 | 0.6654 | 0.0027 | -2.7791 | 0.2901 | 0.2418 | 0.7004 |
| TEM | 0.0864 | 0.599 | -0.2507 | 1.0183 | 0.5602 | 0.1515 | 0.2861 | 0.562 | -0.1559 |
| CPM | 0.2774 | 0.0024 | 2.8078 | 0.6839 | 0.0042 | -2.6328 | 0.2914 | 0.2313 | 0.7344 |
| DEP | 0.14367 | 0.0383 | 1.7704 | 0.7422 | 0.0152 | -2.1628 | 0.2862 | 0.5299 | -0.075 |
| NUT | 0.3963 | 0.0001 | 3.6018 | 0.5687 | 0.0001 | -3.5604 | 0.3028 | 0.0783 | 1.4165 |
| NBI | 0.4624 | $1.7x10^5$ | 4.1363 | 0.5432 | $8.0x10^5$ | -3.7722 | 0.3224 | 0.0221 | 2.0119 |
| CVV | 0.2731 | 0.0037 | 2.6721 | 0.683 | 0.0043 | -2.6252 | 0.2869 | 0.2861 | 0.5648 |
| CVD | 0.3604 | 0.0003 | 3.3751 | 0.5775 | 0.0002 | -3.4984 | 0.2867 | 0.388 | 0.2845 |
| ESC | 0.1154 | 0.0813 | 1.3962 | 0.8128 | 0.0603 | -1.5514 | 0.2883 | 0.0851 | 1.371 |
| IPSE | 0.2164 | 0.0005 | 3.2778 | 0.7366 | 0.0112 | -2.2821 | 0.5236 | 0.0003 | 3.3901 |
| VAC | -0.1781 | 0.8387 | -0.989 | 1.1334 | 0.8652 | 1.1038 | 0.2852 | 0.8677 | -1.1155 |

*Fuente: Elaboración propia en R*

Tabla 9: Contrastes globales de variables endógena y exógenas año 2014

| Test | I de Moran | | | C de Geary | | | G de Getis y Ord | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | I | P value | Z(I) | C | P value | Z(C) | G | P value | Z(G) |
| SMR | 0.0851 | 0.2114 | 0.8015 | 0.9369 | 0.3841 | -0.2946 | 0.1161 | 0.2996 | 0.5255 |
| IDD | 0.4888 | 0.0014 | 2.9874 | 0.3671 | 0.0008 | -3.1281 | 0.1097 | 0.612 | -0.2845 |
| TEM | 0.2244 | 0.0415 | 1.733 | 0.605 | 0.0464 | -1.68 | 0.1121 | 0.2354 | 0.7211 |
| CPM | 0.1788 | 0.093 | 1.3223 | 0.6482 | 0.0478 | -1.6666 | 0.111 | 0.5036 | -0.0089 |
| DEP | 0.3586 | 0.0111 | 2.2838 | 0.4781 | 0.0052 | -2.5596 | 0.1121 | 0.3302 | 0.4394 |
| NUT | 0.2264 | 0.0629 | 1.5308 | 0.5615 | 0.0136 | -2.2066 | 0.1116 | 0.4753 | 0.0618 |
| NBI | 0.3801 | 0.0092 | 2.3551 | 0.5708 | 0.0147 | -2.1766 | 0.1199 | 0.3415 | 0.4083 |
| CVV | 0.2859 | 0.0174 | 2.1098 | 0.7105 | 0.1087 | -1.2334 | 0.1115 | 0.2433 | 0.6958 |
| CVD | 0.5939 | 0.0002 | 3.4928 | 0.2878 | 0.0001 | -3.6448 | 0.1115 | 0.1708 | 0.951 |
| ESC | 0.1741 | 0.1059 | 1.2485 | 0.8122 | 0.1729 | -0.9428 | 0.1099 | 0.8638 | -1.0976 |
| IPSE | -0.064 | 0.5437 | -0.1097 | 0.998 | 0.4975 | -0.0062 | 0.0815 | 0.5827 | -0.2087 |
| VAC | 0.3785 | 0.0078 | 2.4145 | 0.5901 | 0.0236 | -1.9828 | 0.1099 | 0.7696 | -0.7375 |

*Fuente: Elaboración propia en R*

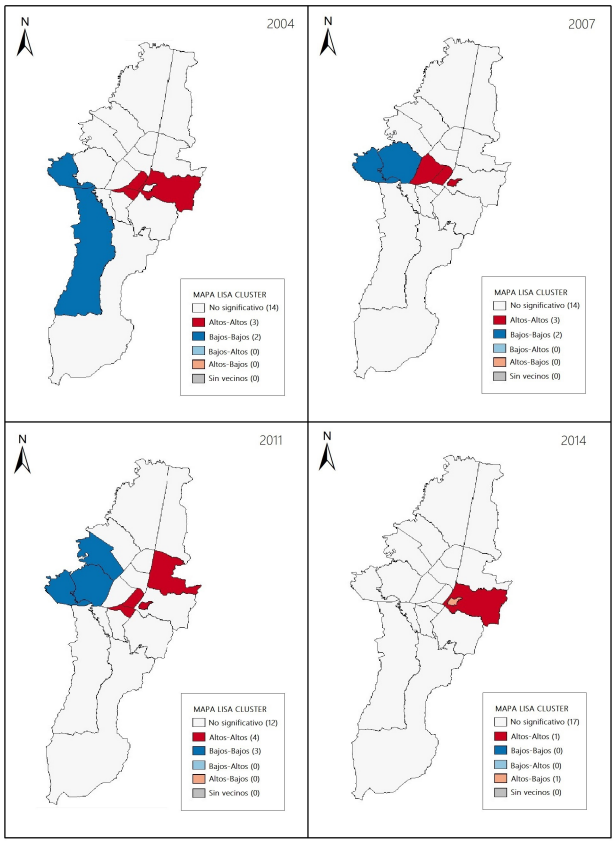
### 6.1.3. Ánalisis Exploratorio Local

Se procede a determinar la existencia de estructuras de dependencia espacial local de la variable endógena mediante mapas LISA para el periodo de estudio, en este análisis sobresalen las regiones con valores significativos a partir del estadístico de autocorrelación espacial local del I de Moran, para identificar la ubicación de los conglomerados espaciales destacando los puntos calientes y atípicos en la zona.

La figura 11 muestra que en el año 2004 existen clusters espaciales de valores altos rodeados de altos en las regiones de Chapinero, Los Mártires y Antonio Nariño, y de bajos rodeados de bajos en Bosa y Ciudad Bolívar. Para el 2007, se agrupan espacialmente las localidades de Candelaria, Los Mártires y Puente Aranda (altos-altos), Bosa y kennedy presenta agrupación de valores bajos rodeados de bajos. Existen conglomerados espaciales de puntos calientes altos rodeados de altos en las regiones de Chapinero, Candelaria, Los Mártires y Antonio Nariño, y de bajos rodeados de bajos en Bosa, Kennedy y Fontibón (año 2011).

Debido a que en el año 2014 no se encontró un esquema de autocorrelación espacial global por presencia de heterogeneidad espacial en los valores de la tasa de mortalidad estandarizada, en el mapa cluster LISA de este año se evidencia que existe un cluster espacial de valores altos rodeados de altos en la localidad de Santa Fe y un valor atípico de altos rodeados de bajos en Candelaria.

Figura 11: Mapa cluster LISA

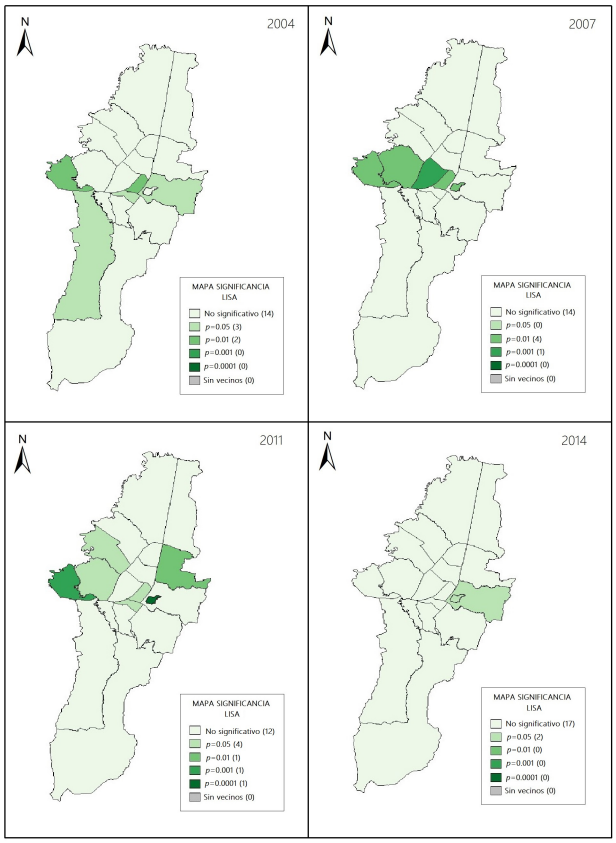



*Fuente: Elaboración propia*

Los mapas de significancia señalan la forma en que las variables están distribuidas en la región y las posibles aglomeraciones, mostrando la probabilidad de que las relaciones de contigüidad se generen de manera aleatoria. En el mapa de significancia

LISA del 2004 se evidencia que con un p valor al 5 % son significativas las localidades de Santa Fe, Antonio Nariño y Ciudad Bolivar, y significativos espacialmente al 1 % Bosa y Los Mártires. Se destacan 5 localidades significativas localmente en el año 2007, al 1 % están Candelaria, Los Mártires, Bosa y Kennedy y al 0.1 % Puente Aranda. En el año 2007 se encontraron 7 regiones con diferentes significancias, éstas son Los Mártires, Antonio Nariño, Kennedy y Fontibón con un valor del 5 %, Chapinero (1 %), Bosa (0.1 %) y Candelaria (0.01 %). Por útlimo, para el año 2014 solo se encontró una significancia al 5 % para las localidades de Santa Fe y Candelaria. Las regiones de Santa Fe y Antonio Nariño aparecen con una significancia del 5 % en dos años de estudio.

Figura 12: Mapa Significancia LISA



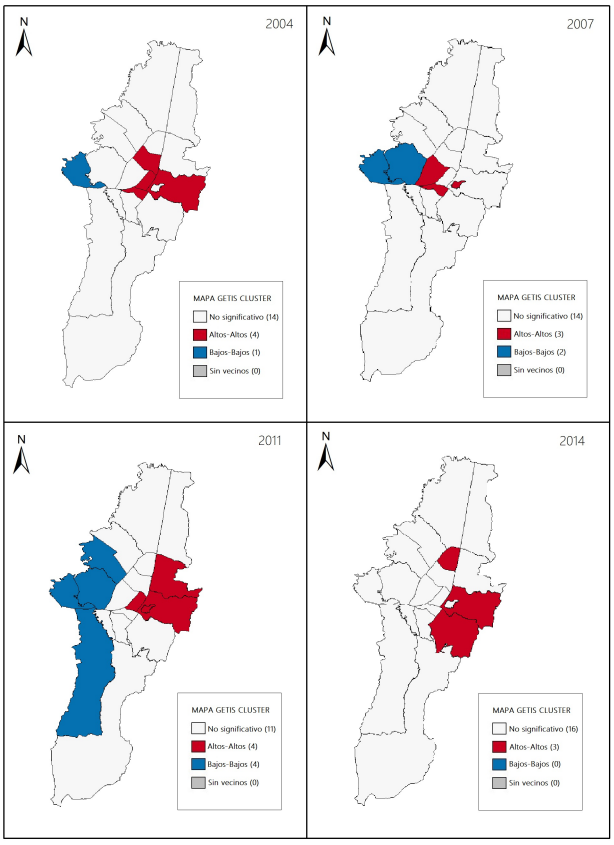

*Fuente: Elaboración propia*

Tambièn, se examinaron los mapas de autocorrelación local G de Getis y Ord que indican como hipótesis nula que no existe clúster espacial, éstos mapas sólo muestran la presencia de clusters espaciales. En las figuras 13 y 14 se observan los mapas de clúster y significancia respectivamente.

En el primero se identifican puntos calientes para el año 2004 en las localidades de Chapinero, Los Mártires y Antonio Nariño como se vislumbró en el estadístico local

anterior, y a diferencia de este aparece la región de Teusaquillo con cluster local espacial. Sólo se muestra como punto frío la localidad de Bosa. En el año 2007 los puntos calientes los presenta las localidades de Antonio Nariño, Puente Aranda y Candelaria, estas dos últimas también se observan como cluser espacial en el mapa LISA 2007, y de la misma manera aparecen como puntos fríos las regiones de Bosa y Kennedy. Para el año 2011 se observan valores bajos rodeados de bajos en las localidades de Fontibón, Kennedy, Bosa y Ciudad Bolivar y valores altos rodeados de altos en Chapinero, Candelaria, Santa Fe y Los Mártires. Finalmente, para el año 2014 al igual que en el mapa LISA se identifica un cluster espacial de valores altos rodeados de altos para la localidad de Santa Fe, pero aparecen las regiones de San Cristobal y Barrios Unidos, en este año no se identifican puntos fríos. La localidad de Santa Fe aparece en 3 años con valores altos rodeados de altos, al igual que Bosa con valores bajos rodeados de bajos.

Figura 13: Getis Cluster local

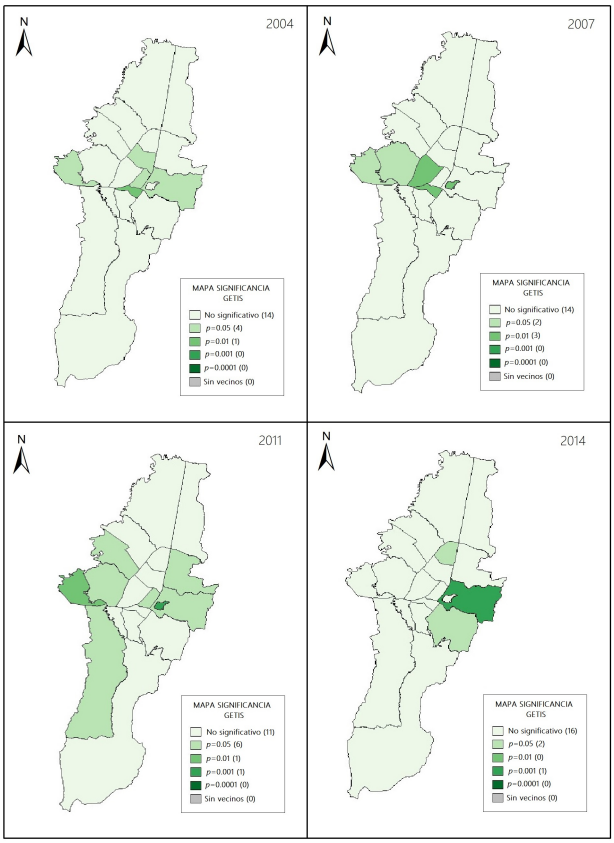



*Fuente: Elaboración propia*

En el mapa de significancia de Getis y Ord del 2004 se distinguen las regiones de Santa Fe, Teusaquillo, Los Mártires y Bosa con significancia del 5 % y Antonio Nariño con 1 %. Para el 2007 se manifiestan con una signifcancia del 1 % las regiones de Puente Aranda, Antonio Nariño y Candelaria, al 5 % Bosa y Kennedy. El año

2011 tiene mayor número de localidades significativas, se encuentran 6 regiones con significancia del 5 % éstas son Chapinero, Santa Fe, Los Mártires, Kennedy, Fontibon y Ciudad Bolivar, con 1 % aparece Bosa y 0.1 % Candelaria. Finalmente, para el año 2014 se encontraron significantes las localidades de Barrios Unidos y San Cristobal con 5 % y Santa Fe con 0.1 %.

Figura 14: Getis Significancia local




*Fuente: Elaboración propia*

### 6.1.4. Análisis Bivariado

Esta sección tiene como propósito encontrar las variables independientes que presentan estructuras de autocorrelación espacial con la variable endógena, para desarrollar modelos que describan el comportamiento de la tasa de mortalidad estandarizada de la neumonía. Este análisis bivariado a partir de la Función I de Moran Bivariado, permite determinar qué tanto se ve afectado el comportamiento de la variable dependiente con respecto a los valores de las variables explicativas. Para todos los años se utilizó la herramienta I Moran Bivariado de GeoDa, excepto para el 2011 donde se utilizó la función en R ya que en este año se eligió la matriz de contigüidad Triangulación Delaunay y no aparece entre las opciones de matriz de pesos espaciales del

software GeoDa.

En la tabla 10 se visualizan los resultados del análisis para los años 2004 y 2007; en el año 2004 se encuentra la variable índice de desempleo con un estadístico significativo de -2.49 y un p valor de 0.014, que confirma la asociación espacial con la variable endógena. También, aparece la variable de necesidades básicas insatisfechas con un estadístico significativo y un p valor menor al 5 % mostrando una estructura de autocorrelación espacial. Las variables de temperatura y vacunación manifiestan la presencia de autocorrelaciòn espacial negativa y la tasa de escolaridad, positiva.

Tabla 10: Moran bivariado 2004-2007

| Año | 2004 | | | 2007 | | |
|---|---|---|---|---|---|---|
| Variable | I | P valor | Z(I) | I | P valor | Z(I) |
| IDD | -0.2581 | 0.014 | -2.4982 | -0.0435 | 0.424 | -0.1994 |
| TEM | 0.1267 | 0.109 | 1.1655 | -0.4149 | 0.032 | -1.8772 |
| CPM | 0.0304 | 0.351 | 0.3094 | -0.1562 | 0.264 | -0.6358 |
| DEP | 0.0405 | 0.422 | 0.233 | 0.3005 | 0.081 | 1.3628 |
| NUT | -0.0363 | 0.359 | -0.3298 | 0.0342 | 0.406 | 0.1733 |
| NBI | -0.2124 | 0.022 | -2.0124 | -0.2058 | 0.196 | -0.8587 |
| CVV | 0.0901 | 0.193 | 0.8776 | 0.0278 | 0.455 | 0.1228 |
| CVD | 0.1754 | 0.05 | 1.7386 | 0.271 | 0.131 | 1.1436 |
| ESC | 0.1289 | 0.163 | 1.0293 | 0.6282 | 0.009 | 2.4485 |
| IPSE | 0.079 | 0.212 | 0.7435 | 0.1935 | 0.169 | 0.8749 |
| VAC | -0.059 | 0.354 | -0.4885 | -0.6075 | 0.01 | -2.6976 |

*Fuente: Elaboración propia en R*

Tabla 11: Moran bivariado 2011

| Año | 2011 | | | 2014 | | |
|---|---|---|---|---|---|---|
| Variable | I | P valor | Z(I) | I | P valor | Z(I) |
| IDD | -0.1458 | 0.4662 | -0.7287 | -0.2489 | 0.043 | -1.6916 |
| TEM | -0.1457 | 0.4667 | -0.7279 | -0.142 | 0.162 | -0.9856 |
| CPM | -0.3789 | 0.0090 | -2.6105 | -0.1489 | 0.159 | -1.0448 |
| DEP | -0.1002 | 0.718 | -0.3610 | 0.1587 | 0.133 | 1.128 |
| NUT | 0.0149 | 0.5694 | 0.5689 | -0.1308 | 0.172 | -0.9284 |
| NBI | -0.1212 | 0.596 | -0.5301 | -0.41 | 0.007 | -2.6972 |
| CVV | -0.1693 | 0.3585 | -0.9182 | -0.0021 | 0.394 | -0.2035 |
| CVD | 0.3418 | 0.0013 | 3.2078 | 0.2267 | 0.076 | 1.4968 |
| ESC | -0.2213 | 0.1807 | -1.3385 | -0.3533 | 0.015 | -2.3799 |
| IPSE | 0.2657 | 0.0095 | 2.5931 | 0.0106 | 0.309 | 0.3029 |
| VAC | -0.0218 | 0.7857 | 0.2719 | -0.3238 | 0.025 | -2.1296 |

*Fuente: Elaboración propia en R*

En la tabla 11 se observan los resultados para el año 2011 y 2014. A partir de los

valores del p valor que sean menores a 5 % se considera que poseen un esquema de autocorrelaciòn espacial con respecto a la variable dependiente las variables de contaminación, calidad de vida y número de instituciones prestadoras de servicios. Para el 2014, se encuentran asociadas en el espacio las variables de NBI, ESC y VAC.

Por último, se analizaron los correlogramas bivariados de la variable dependiente con las variables independientes que presentan autocorrelación espacial según los resultados obtenidos en el I de Moran Bivariado, éstos se encuentran en los anexos de la sección B, con el fin de encontrar rezagos de ordenes superiores que sean significativos. Pero sólo se encontró para el año 2011 que las variables CPM y CVD tienen rezagos significativos de orden 1 y 3.

## 6.2. Modelo Regresivo Clásico

Después de haber realizado el análisis exploratorio de la variable respuesta (SMR) se ha comprobado que ésta enseña un esquema significativo de dependencia espacial en cada uno de los años analizados. Sin embargo, este no permite, en primer lugar identificar las razones que expliquen este fenómeno y en segundo lugar, determinar la estructura de ésta dependencia (Moreno y Vayá, 2002). Variables exógenas como las prácticas deportivas (DEP), la asistencia escolar (ESC) , el número de centros de prestación de servicios de salud (IPSE) entre otras demostraron estar relacionadas en el espacio con la variable independiente y por ello deben ser consideradas para el modelamiento espacial pues informan de problemas de dependencia espacial en una ecuación.

El punto de partida previo a abordar los modelos espaciales consiste en la especificación del modelo propuesto como un modelo clásico de regresión lineal sin efectos espaciales, estimado por mínimos cuadrados ordinarios (MCO). El propósito básico de todo análisis de regresión lineal es encontrar una relación entre una variable dependiente y un conjunto de variables independientes o explicativas, que si es estimado por MCO, es capaz de modelar los valores de predicción y los valores observados con un buen ajuste bajo ciertas hipótesis básicas (Chasco Yrigoyen, 2003). Las variables exógenas que sobrevivan a la regresión clásica y a todos sus supuestos pueden convertirse en posibles candidatas para un modelo espacial (además de aquellas identificadas previamente en el AEDE) (Bohorquez, 2009).

Para cada año, las variables independientes identificadas fueron incluidas en una regresión clásica estimada por mínimos cuadrados ordinarios. Se aplico el método de stepwise hacia adelante en el año 2004, y en ambas direcciones para los años 2007, 2011 y 2014 con un nivel de significancia de $p < 0{,}05$ y con el criterio de Akaike (AIC) como un test parcial para determinar qué variables explicativas deben permanecer en el modelo final. Posteriormente del método stepwise se incluyeron las variables que guardan correlación espacial con la SMR y se probó qué tan significativo era su

aporte al modelo de regresión clásico.

Para el año 2004 la regresión stepwise se realizó hacia adelante teniendo como criterio el nivel de significancia; se utilizó el criterio del $p - valor$ debido a que conservaba más variables que el criterio AIC (es más adecuado para fines de pronóstico, pues se centra en la minimización de la pérdida de cuadrados) que las eliminaba todas. Las variables del mejor modelo de regresión encontrado se observan en la Tabla 12, en donde según éste, la SMR del 2004 puede explicarse moderadamente en un 52 % por la asistencia escolar (ESC), la calidad de vida (CVD), los hogares que no consumieron las tres comidas (NUT), la cobertura de vacunación contra la anti haemophilus influenzae tipo B (VAC) y la temperatura ambiente (TEM). Si bien no todas las variables son significativas al $p < 0{,}05$ todas aportan para explicar la SMR y su combinación presenta el $R - cuadrado$ más alto de todas las posibles combinaciones.

Tabla 12: Estimación modelo de regresión clásico para el año 2004.

| Variables | Estimado | Pvalor |
|---|---|---|
| Intercepto | -6.951905 | 0.03271 |
| ESC | -0.066563 | 0.02386 |
| CVD | 0.086622 | 0.00858 |
| NUT | 0.073027 | 0.01808 |
| VAC | 0.007959 | 0.05197 |
| TEM | 0.076035 | 0.10129 |
| R-cuadrado | | 0.5213 |
| R-cuadrado Ajustado | | 0.3372 |

*Fuente: Elaboración propia en R*

Aunque el AEDE identificó al índice de desempleo (IDD) y la pobreza por NBI (NBI) como variables relacionadas espacialmente con la SMR, éstas no fueron significativas en la elección por stepwise. Sin embargo, estas variables serán incorporadas en el modelo espacial y se determinará si son o no significativas para el mismo. La ecuación 26 representa el modelo regresivo lineal clásico para el año 2004, en donde los $\beta_n$ aluden a los valores estimados para cada variable en la tabla 12.

$$SMR_{2004} = \beta_0 + \beta_1 ESC + \beta_2 CVD + \beta_3 NUT + \beta_4 VAC + \beta_5 TEM \tag{26}$$

Se efectuaron pruebas de diagnósticos formales para el modelo clásico del 2004. En la Tabla 13 se observan los supuestos del modelo regresivo con su respectiva prueba y resultados de la misma. El supuesto de heteroscedasticidad, que propone que la varianza no es la misma para las variables del modelo, fue validado a través del test de Breusch-Pagan cuya hipótesis nula $H_o$ sostiene que la varianza no depende de los regresores auxiliares, es decir, existe homoscedasticidad. Debido a que el P-valor para

la prueba es mayor al nivel de significancia ($p < 0{,}05$) no hay evidencia suficiente para rechazar la hipótesis nula, por ende el modelo es homoscedástico. De la misma manera no se rechaza la hipótesis nula en la prueba de Shapiro Wilk realizada a los residuos; ésta dice que los datos siguen una distribución normal. Las hipótesis nulas de la prueba de RESET y de Rainbow que aluden a la buena especificación del modelo y la linealidad del mismo respectivamente no son rechazadas basándose en el nivel de significancia, lo que significa que las variables del modelo ayudan a explicar la variable respuesta y que la relación entre éstas es lineal (Kramer y Sonnberger, 1986).

Tabla 13: Supuestos del modelo regresivo clásico: 2004.

| Supuesto | Test | Contraste | P-valor |
|---|---|---|---|
| Heteroscedasticidad | Breush-Pagan | 1.0335 | 0.06628 |
| Normalidad | Shapiro Wilk | 0.95536 | 0.4849 |
| Especificación | RESET | 0.6336 | 0.549 |
| Linealidad | Rainbow | 2.0419 | 0.3028 |

*Fuente: Elaboración propia en R*

En cuanto al supuesto de multicolinealidad, el cual propone que una variable predictora puede ser linealmente predicha por las otras, se aplicó el estadístico de factores de inflación de varianza (VIF, por sus siglas en inglés). La regla general consiste en que los VIF que exceden valores de 4 justifican una investigación adicional, mientras que los VIF que exceden 10 son signos de multicolinealidad grave que requieren corrección. En la Tabla 14 se observan los VIF para cada variable, pese a que ninguno excede el valor de 10, la variable de hogares que no comieron ninguna de las tres comidas (NUT) debe ser tenida en cuenta pues puede estar afectando el modelo regresivo. La SMR para el 2004 es explicada por variables sociales, conductuales, ambientales y de asistencia médica.

Tabla 14: Supuesto de multicolinealidad (Inflación de la varianza): 2004.

| Variable | VIF |
|---|---|
| ESC | 1.270671 |
| CVD | 3.926196 |
| NUT | 4.148068 |
| VAC | 1.368438 |
| TEM | 1.279161 |

*Fuente: Elaboración propia en R*

En el año 2007 la regresión stepwise se realizó en ambas direcciones teniendo como criterio el AIC. Las variables que prevalecieron para este año fueron la asistencia escolar (ESC), el índice de prácticas deportivas (DEP) y el número de centros asistenciales (IPSE) y explicaban en un 59 % la SMR. Cuando se realizaron las pruebas

diagnósticas se encontró que los residuales del modelo no cumplían el supuesto de normalidad ($p-valor = 0.03312$) bajo el test de Shapiro-Wilk así que se realizó una serie de transformaciones con el fin de mitigar este error. De las transformaciones efectuadas la transformación logarítmica de las variables exógenas logró corregir este supuesto. En la tabla 15 se muestran el valor estimado y la significancia de las variables del modelo. Aunque el supuesto de normalidad fue corregido, se sacrificó en un 7 % el R-cuadrado del modelo, que terminó siendo de 52 %. Para este año las variables temperatura ambiente (TEM) y la cobertura de vacunación contra la anti haemophilus influenzae tipo B (VAC) identificadas en el AEDE no lograron ingresar al modelo regresivo clásico a excepción de la asistencia escolar (ESC) que en análisis bivariado presento significancia con la variable respuesta y sí logro estar presente en el modelo regresivo. Esto se debe a que la interacción de múltiples variables puede cambiar la significancia de sus vecinas (Florax y Folmer, 1992).

Tabla 15: Estimación modelo de regresión clásico para el año 2007.

| Variables | Estimado | Pvalor |
|---|---|---|
| Intercepto | 3.677626 | 0.000197 |
| log(ESC) | -0.096858 | 0.000681 |
| log(DEP) | 0.013310 | 0.044334 |
| log(IPSE) | -0.001983 | 0.100762 |
| R-cuadrado | | 0.5282 |
| R-cuadrado Ajustado | | 0.4338 |

*Fuente: Elaboración propia en R*

La ecuación 27 representa el modelo regresivo lineal clásico para el año 2007, en donde los $\beta_n$ aluden a los valores estimados para el logaritmo de cada variable en la tabla 15.

$$SMR_{2007} = \beta_0 + \beta_1 ln(ESC) + \beta_2 ln(DEP) + \beta_3 ln(IPSE) \tag{27}$$

Para los supuestos de heteroscedasticidad, normalidad en los residuos, especificación y linealidad se rechaza la hipótesis nula al 95 % de confianza (Véase Tabla 16) indicando que el modelo es homoscedástico, existe normalidad en los residuos, tiene buena especificación y la relación de linealidad es óptima para explicar el comportamiento de la variable endógena con sus variable exógenas.

Tabla 16: Supuestos del modelo regresivo clásico: 2007.

| Supuesto | Test | Contraste | P-valor |
|---|---|---|---|
| Heteroscedasticidad | Breush-Pagan | 0.8196 | 0.8448 |
| Normalidad | Shapiro Wilk | 0.91001 | 0.07405 |
| Especificación | RESET | 1.5199 | 0.2552 |
| Linealidad | Rainbow | 0.52134 | 0.8219 |

*Fuente: Elaboración propia en R*

A las variables de este modelo, también se les aplicó el estadístico VIF presentado en la tabla 17 para realizar el análisis de multicolinealidad en el modelo, en donde se obtuvo que ninguna variable representa un grado grave o moderado de multicolinealidad. Al final, la SMR solo es explicada por variables sociales, conductuales y de asistencia médica. Con estos supuestos validados, es posible tener un modelo clásico básico para realizar el modelo espacial.

Tabla 17: Supuesto de multicolinealidad (Inflación de la varianza): 2007.

| Variable | VIF |
|---|---|
| log(ESC) | 1.205692 |
| log(DEP) | 1.079389 |
| log(IPSE) | 1.286272 |

*Fuente: Elaboración propia en R*

Para el 2011 la regresión stepwise se realizó en ambas direcciones tomando como criterio de selección de las variables el AIC. La calidad de vida (CVD), la cobertura de vacunación contra la anti haemophilus influenzae tipo B (VAC), el número de centros asistenciales (IPSE) y la calidad de la vivienda (CVV) fueron las mejores candidatas para integrar el modelo regresivo clásico para este año con un R-cuadrado del 53 %. Al momento de realizar los supuestos de un modelo regresivo se halló que los residuales no venían de una distribución normal ($p-valor = 0.03585$) por lo que, al igual que en año 2007, se realizaron una serie de transformaciones a las variables para corregir este supuesto. La transformación que corrigió el supuesto de normalidad fue la transformación Box-Cox aplicada a la variable respuesta; según Box y Cox (1964) esta transformación permite corregir los sesgos en la distribución de los errores, las varianzas desiguales y la no linealidad redundando en datos normales. En la tabla 18 se muestran las variables junto con su estimado y su valor $p$; aunque no todas las variables son significativas para el modelo, contribuyen a una explicación de 58 % del mismo. Un aumento del 5 % con respecto al modelo sin la variable endógena transformada.

Tabla 18: Estimación modelo de regresión clásico para el año 2011.

| Variables | Estimado | Pvalor |
|---|---|---|
| Intercepto | -0.788623 | 0.8179 |
| CVD | 0.048973 | 0.0611 |
| CVV | -0.351091 | 0.1033 |
| IPSE | 0.003154 | 0.0403 |
| VAC | 0.018986 | 0.2964 |
| R-cuadrado | | 0.586 |
| R-cuadrado Ajustado | | 0.4677 |

*Fuente: Elaboración propia en R*

La variable número de centros asistenciales (IPSE), Calidad de vida (CVD) y calidad de vivienda (CVV) presentaron correlación espacial con la SMR según en AEDE previamente realizado y también permanecieron en el modelo clásico. El índice de desempleo (IDD), la temperatura ambiente (TEM), la contaminación por PM10 (CPM) y la asistencia escolar (ESC) no lograron permanecer en el modelo clásico final. No obstante, se incluirán en modelo espacial y se determinará si aportan o no información que modele la SMR. La ecuación 28 presenta el modelo regresivo lineal para el año 2011, en donde los $\beta_n$ representan los valores estimados de las variables en la tabla 18 y la variable SMR se encuentra modificada a través de la transformación Box-Cox.

$$BoxSMR_{2011} = \beta_0 + \beta_1 CVD + \beta_2 VAC + \beta_3 IPSE + \beta_4 CVV \tag{28}$$

Los diagnósticos formales presentados en la tabla 19 exponen la no existencia de heteroscedasticidad, la normalidad en los residuos, una correcta especificación y la linealidad del modelo como una decisión acertada teniendo como base el $p - valor$ mayor al 5 % para todas las pruebas de contraste, rechazando así las hipótesis alternas.

Tabla 19: Supuestos del modelo regresivo clásico: 2011.

| Supuesto | Test | Contraste | P-valor |
|---|---|---|---|
| Heteroscedasticidad | Breush-Pagan | 0.5266 | 0.9713 |
| Normalidad | Shapiro Wilk | 0.95384 | 0.458 |
| Especificación | RESET | 1.0528 | 0.3791 |
| Linealidad | Rainbow | 2.5301 | 0.1923 |

*Fuente: Elaboración propia en R*

Asimismo, la prueba de factor de inflación de la varianza muestra que ninguna variable supone un problema de multicolinealidad pues los valores asociados a sí

mismas no superan el valor de 10 de VIF. El modelo regresivo final solo es explicado por variables sociales y de asistencia médica.

Tabla 20: Supuesto de multicolinealidad (Inflación de la varianza): 2011.

| Variable | VIF |
|---|---|
| CVD | 1.169160 |
| CVV | 1.105576 |
| IPSE | 1.301852 |
| VAC | 1.218251 |

*Fuente: Elaboración propia en R*

El año 2014 es uno de los años con mayor cantidad de variables exógenas incorporadas al modelo regresivo lineal. La regresión stepwise hecha en ambas direcciones y con criterio AIC determinó que el número de centros asistenciales (IPSE), la calidad de la vivienda (CVV), la cobertura de vacunación contra la anti haemophilus influenzae tipo B (VAC), la temperatura ambiente (TEM), la contaminación por PM10 (CPM), los hogares que no consumieron las tres comidas (NUT), la pobreza por NBI y el índice de desempleo (IDD) son las variables que mejor modelan la SMR para el 2014. Todas las variables excepto el NBI y el IDD son significativas a un $p < 0{,}05$ y explican el modelo en un 93 %, el valor más alto para todos los años de análisis.

Tabla 21: Estimación modelo de regresión clásico para el año 2014.

| Variables | Estimado | Pvalor |
|---|---|---|
| Intercepto | 1.762e+01 | 1.39e-05 |
| IPSE | 3.626e-04 | 0.00225 |
| CVV | -1.161e+00 | 2.42e-05 |
| VAC | -3.071e-02 | 0.00581 |
| TEM | 1.409e-01 | 0.02722 |
| CPM | -1.499e-02 | 0.01251 |
| NUT | -1.128e-01 | 0.00768 |
| NBI | 2.623e-02 | 0.21539 |
| IDD | -3.235e-02 | 0.47660 |
| R-cuadrado | | 0.9341 |
| R-cuadrado Ajustado | | 0.8814 |

*Fuente: Elaboración propia en R*

De las 8 variables en el modelo regresivo, 3 (NBI, IDD y VAC) presentaban correlación espacial con la variable respuesta. La asistencia escolar, aunque en el análisis bivariado presentó correlación con la SMR, no participa en el actual modelo regresivo para el 2014 pero puede que su significancia sea mayor cuando se introduzca en

los modelos regresivos espaciales. La ecuación 29 exhibe el modelamiento regresivo para la SMR del 2014, siendo los $\beta_n$ los estimados de cada variable en la tabla 21.

$$SMR_{2014} = \beta_0 + \beta_1 IPSE + \beta_2 CVV + \beta_3 VAC + \beta_4 TEM + \beta_5 CPM + \beta_6 NUT + \beta_7 NBI + \beta_8 IDD \quad (29)$$

Los supuestos del modelo se pueden observar en la tabla 22 en donde, basados en el p-valor de cada test, se deduce que el modelo es homoscedástico, que los residuales siguen una distribución normal, que existe buena especificación del modelo y que la relación de linealidad es la más optima para modelar la SMR.

Tabla 22: Supuestos del modelo regresivo clásico: 2014.

| Supuesto | Test | Contraste | P-valor |
|---|---|---|---|
| Heteroscedasticidad | Breush-Pagan | 6.058 | 0.6407 |
| Normalidad | Shapiro Wilk | 0.98052 | 0.9482 |
| Especificación | RESET | 1.5919 | 0.2618 |
| Linealidad | Rainbow | 2.6510 | 0.7105 |

*Fuente: Elaboración propia en R*

El VIF para ninguna variable supera el valor de 10 (Véase Tabla 23) por lo que se puede inferir que no existen problemas de multicolinealidad graves que afecten el modelo. Finalmente, la SMR del 2014 es explicada por variables ambientales, sociales, conductuales y de asistencia médica siendo su modelamiento por MCO el mejor de los 4 años analizados.

Tabla 23: Supuesto de multicolinealidad (Inflación de la varianza): 2014.

| Variable | VIF |
|---|---|
| IPSE | 1.609282 |
| CVV | 1.868124 |
| VAC | 1.819097 |
| TEM | 1.598612 |
| CPM | 1.954274 |
| NUT | 2.383013 |
| NBI | 3.062311 |
| IDD | 3.315079 |

*Fuente: Elaboración propia en R*

Ahora que se conocen las variables que explican la SMR por neumonía para cada año, es necesario calcular los contrastes de autocorrelación espacial para determinar si los modelos presentados anteriormente explican correctamente la variable respuesta (la

SMR depende de factores propios de la localidad) o si por el contrario, existe algún tipo de interdependencia espacial entre las localidades y el modelo clásico estimado por MCO no es el ideal para representarla (la SMR guarda alguna relación espacial que afecta su comportamiento en una localidad y/o en sus vecinas).

## 6.3. Modelo Regresivo Espacial

El modelo clásico, como se ha mencionado previamente, es la base para seleccionar un modelo espacial apropiado. El modelo clásico estimado por mínimos cuadrados no es otra cosa que la simplificación de modelos espaciales en ausencia de dependencia espacial en donde existe una matriz general de varianza parametrizada para el termino del error (Anselin, 1988b). El principal obstáculo en la utilización de los modelos tradicionales radica en que al no considerar la presencia de autocorrelación espacial en los residuos, estos modelos pueden mostrar problemas de especificación; un efecto relevante es el riesgo de estimar coeficientes ineficientes en los que las pruebas de significancia estadística serían cuestionables debido a la inflación de los errores estándar. De este modo, la existencia de autocorrelación espacial representa una violación a los supuestos de los MCO por lo que es menester someter los residuos de las regresiones a pruebas estadísticas que comprueben su aleatoriedad espacial (Anselin, 1988a).

En el AEDE se determinó que la SMR se encontraba autocorrelacionada en el espacio, por lo que pensar que un modelo de regresión clásico podría funcionar es infundado. Si bien existe autocorrelación espacial, aún es incierto si ésta debe incorporarse en el modelo como rezago en términos de error o en rezagos espaciales de las variables explicativas. Para identificar la estructura de dependencia espacial de la SMR como primer paso se realiza la prueba de I de Moran a los residuos de la regresión por MCO; la detección de correlación espacial en los residuales implicaría una relación no linear entre la variable exógena y sus variables endógenas, la omisión de uno o más regresores espacialmente correlacionados o que una estructura de error autorregresivo es la más apropiada (Florax y Folmer, 1992). Aunque el test de Moran funciona bien como una prueba general, el diagnóstico por multiplicadores de Lagrange es más especifico; éste provee una base para la elección de un modelo regresivo espacial apropiado (Anselin, 1988a). Como segundo paso, se realiza el diagnóstico del Multiplicador de Lagrange (LM) con el fin de identificar si la variable se encuentra rezagada espacialmente, existe autocorrelación espacial en el error o ambos fenómenos suceden simultáneamente. Tanto la prueba de I de Moran como el LM se basan en los residuales del modelo por MCO con o sin efecto espacial (Chasco Yrigoyen, 2003).

En la Tabla 24 se muestra el estadístico de I de moran aplicado a los residuales de los modelos regresivos clásicos seleccionados en la sección anterior. Para el test se utilizaron las matrices de pesos espaciales previamente seleccionadas. Los años 2004, 2007 y 2014 no presentan autocorrelación espacial en los residuos basados en

su P-valor ($p - valor > 0{,}05$) pero el año 2011 sí exhibe signos de autocorrelación espacial. Dicho esto, el año 2011 debe ser analizado con mayor precaución pues el test de I de Moran podría indicar el problema de la variable omitida, que se corrige a través de la inclusión del rezago espacial en el modelo; para los demás años aunque no se observe autocorrelación en los residuos, se sabe a priori que la SMR está autocorrelacionada en el espacio y por ende un modelo espacial debe ser calculado para éstos.

Tabla 24: Test Global de I de Moral para los residuales de una regresión.

| Año | Matriz W | P-Valor | Z(I) |
|---|---|---|---|
| 2004 | Reina 1 | 0.1074 | 1.240219 |
| 2007 | K-vecinos 1 | 0.2861 | 0.564747 |
| 2011 | Triangulación | 0.0022 | 2.839594 |
| 2014 | K-vecinos 2 | 0.4471 | 0.133069 |

*Fuente: Elaboración propia en R*

El diagnóstico del Multiplicador de Lagrange para los modelos clásicos de cada año se expone en la tabla 25. Frente a cada año se encuentra el $p-valor$ asociado al valor del diagnóstico. Para el año 2004, 2007 y 2014, el modelo más adecuado podría ser un modelo Spatial Lag (SAR) y para el año 2011 la decisión estaría entre un modelo Spatial Lag (SAR) y un efecto de dependencia espacial de forma autorregresiva en la variable endógena con un esquema de medias móviles en las perturbaciones aletaorias (SARMA). No obstante, este diagnóstico es solo una sugerencia o una primera mirada para identificar la dependencia espacial en el modelo. Este diagnóstico en su versión clásica y robusta propuesto por Anselin (1988a) es poco claro en qué tan poderoso es en la presencia o inclusión del rezago espacial de las variables exógenas ($WX$). En vista de ello, es necesario calcular todos los posibles modelos espaciales (SAR, SEM, SDM, SARAR, SDEM, GNS, SLX) y determinar cuál de éstos, basado en el AIC, es el que mejor personifica la dependencia espacial y explica la SMR por neumonía para cada año.

Tabla 25: Diagnóstico para dependencia espacial: Multiplicadores de Lagrange.

| | P-Valor | | | | |
|---|---|---|---|---|---|
| Año | LMerror | LMlag | RLMerror | RLMlag | SARMA |
| 2004 | 0.8904 | 0.2812 | 0.0712 | 0.0360 | 0.1099 |
| 2007 | 0.9289 | 0.0495 | 0.2411 | 0.1085 | 0.2747 |
| 2011 | 0.1575 | 0.0194 | 0.3698 | 0.0402 | 0.0437 |
| 2014 | 0.5540 | 0.0205 | 0.7834 | 0.0240 | 0.0657 |

*Fuente: Elaboración propia en R*

Teniendo como base las variables explicativas del modelo clásico y aquellas identificadas en el análisis bivariado del AEDE, se determinó la composición de los modelos

básicos para construir los modelos espaciales. Se calculó cada uno de los siete modelos espaciales al modelo básico para cada año depurando las variables que no aportaban información a los mismos y afectaban negativamente el criterio de Akaike (AIC).

En la tabla 26 se compara el AIC de los siete modelos que incorporan la dependencia espacial en cada uno de los años de estudio. Para el año 2004 el modelo más adecuado, seleccionando como criterio el valor AIC, es un modelo autorregresivo espacial combinado, es decir, la variable endógena y el error del modelo se encuentran rezagados en el espacio. Tanto para el año 2007 como el año 2014 el mejor modelo regresivo espacial es el modelo general o modelo anidado general (GNS); este modelo posee todas las interacciones espaciales posibles: la variable endógena, las variables exógenas y el error se encuentran autocorrelacionados en el espacio. El modelo autorregresivo espacial (SAR) se identificó como el modelo más óptimo para representar el comportamiento de la SMR del 2011. El uso de criterio de Akaike para la selección de los modelos espaciales radica en que a diferencia de otros criterios (BIC, TIC, QAIC, Log-likelihood) el AIC se desempeña mejor con muestras de cualquier tamaño y provee una medida simple y efectiva para la selección del modelo más aproximado a el modelo verdadero (Hyeyoung y Sujit K., 2009).

Tabla 26: Comparación de Modelos espaciales: Criterio de Akaike

| | AIC | | | | | | |
|---|---|---|---|---|---|---|---|
| Año | SAR | SEM | SDM | SARAR | SDEM | GNS | SLX |
| 2004 | 3.6786 | 6.2452 | 1.6084 | 0.1838 | 3.4800 | 2.4605 | 2.8984 |
| 2007 | -5.6669 | -4.5179 | -0.4961 | -4.7907 | -1.6075 | -22.6418 | -2.4566 |
| 2011 | -1.0277 | 2.0781 | 0.4080 | 0.3159 | -0.2369 | 0.8180 | 9.0457 |
| 2014 | 26.5216 | 26.6751 | 10.0551 | 27.1536 | 8.3556 | 4.6172 | 10.6552 |

**SAR:** Modelo Spatial lag; **SEM:** Modelo Spatial error; **SDM:** Modelo Spatial Durbin; **SARAR:** Modelo combinado autorrregresivo espacial; **SDEM:** Modelo Spatial Durbin error; **GNS:** Modelo general anidado espacial; **SLX:** Modelo Spatial Lag of X

Es importante mencionar que para los 4 años de estudio la variable dependiente presenta autocorrelación espacial sustantiva ratificando las conclusiones halladas en el AEDE. Además, los años 2004, 2007 y 2014 también poseen autocorrelación espacial residual no obstante el test I de Moran para los residuos haya deducido que no existía autocorrelación espacial en los mismos. Aunque el diagnóstico del Multiplicador de Lagrange no arrojó los mismo resultados que el criterio de AIC para los modelos, este primero permitió determinar que todos los modelos debían tener la variable respuesta rezagada en el espacio, conclusión confirmada por el segundo.

Teniendo en cuenta la taxonomía de modelos de regresión lineal propuesta por Anselin (1988b) y por Florax y Folmer (1992) se especificaron los modelos para la SMR en cada año teniendo como base el análisis exploratorio, el modelo clásico

y el modelo autorregresivo espacial seleccionados para cada año.

Se identifica un modelo SARAR como el más adecuado para el año 2004. Que exista una estructura de autocorrelación espacial en la SMR indica que las perturbaciones de primer orden (Efecto reina) en una localidad serán de tipo contagioso para las localidades vecinas, en otras palabras, el hecho de tener un foco de mortalidad por neumonía en una localidad es suficiente para que ésta se propague a sus colindantes. De igual modo, la autocorrelación espacial residual en el modelo indica que el crecimiento azaroso de la tasa de mortalidad en una localidad no solo afectaría a ésta, sino también a otras localidades de Bogotá.

$$\begin{aligned} SMR = \beta_0 + \rho W_1 SMR + W_3 SMR + X\beta + u, \\ u = \lambda W_1 u + \epsilon \end{aligned} \tag{30}$$

La ecuación 30 presenta el modelo autorregresivo espacial del año 2004, en donde, $\beta_0$ alude al intercepto, $\rho$ y $\lambda$ a sus correspondientes coeficientes en la tabla 27, $W$ a la matriz de pesos espaciales efecto reina de orden 1 y orden 3 y $X\beta$ al vector de variables explicativas (asistencia escolar y temperatura ambiente) con su correspondiente estimado. La adición del término $W_3 SMR$ se realizó con base en el AEDE ya que es un término significativo para la explicación de la SMR; la reducción de variables explicativas con relación en el modelo clásico se debe a que cuando interactuaban la CVD, la NUT y la VAC con el componente espacial, el modelo presentaba errores de especificación y el ajuste disminuía considerablemente.

En promedio el 82 % de las perturbaciones en la SMR de la neumonía se transmiten a las localidades vecinas y el coeficiente $\lambda$ indica que las perturbaciones aleatorias también afectan las localidades contiguas, así como el cambio en las variable explicativas de una localidad, pueden afectar la SMR en otras regiones. El descenso de la asistencia escolar y el incremento de la temperatura ambiente estimulan el crecimiento de la tasa de mortalidad de la neumonía. De la misma manera, la relación de tercer orden con la SMR repercute positivamente en la tasa de mortalidad, es decir, que la SMR de una localidad disminuirá si la tasa de sus vecinas de tercer orden en efecto reina aumenta. Los valores de test de Breusch-Pagan, la prueba Shapiro-Wilk y el I de Moran, al 5 % no rechazan la hipótesis nula por lo que se puede afirmar que no existe heteroscedasticidad, los residuos siguen una distribución normal y no se encuentran autocorrelacionados en el espacio. La variabilidad del modelo es explicada en un 79 % por el Pseudo R-cuadrado y en un 63 % por el Nagerlkerke R-cuadrado, que es una versión ajustada del Pseudo R-cuadrado.

Tabla 27: Estimación modelo de regresión espacial para el año 2004: SARAR

| Variables | Estimado | Pvalor |
|---|---|---|
| $\rho$ | 0.82258 | 2.2236e-11 |
| $\lambda$ | 0.93627 | 7.8875e-05 |
| Intercepto | 2.487154 | 4.277e-05 |
| ESC | -0.077297 | 3.668e-07 |
| TEM | 0.089846 | 8.590e-05 |
| W3SMR | -1.007261 | 1.108e-06 |
| Breush Pagan | | 0.4991 |
| Shapiro-Wilk | | 0.5945 |
| I de Moran: Residuos | | 0.6027 |
| Pseudo R-cuadrado | | 0.7915 |
| Nagelkerke R-cuadrado | | 0.6390 |

*Fuente: Elaboración propia en R*

El modelo general anidado fue la mejor opción para el año 2007. Los cambios que ocurran en la SMR por neumonía de una localidad se verán afectados por las dinámicas de mortalidad de la enfermedad, las variables explicativas y el error aleatorio de sus localidades adyacentes. Tener localidades afectadas por la neumonía, la escolaridad, las prácticas deportivas, el número de IPS y la cobertura de vacunación podría contagiar a las localidades vecinas del sistema. De la misma manera, los choques aleatorios o variables omitidas no solo afectarán la localidad en donde se produjeron, sino que se transmitirá a todas las divisiones de Bogotá. En la ecuación 31 se puede observar el modelo general anidado del 2007. $\beta_0$ representa el intercepto, $\rho$ y $\lambda$ los coeficientes, $W$ la matriz de pesos espaciales de K-un vecino, $ln(X)\beta$ el logaritmo del vector de variables explicativas (asistencia escolar, prácticas deportivas, número de IPS y cobertura de vacunación) con su correspondiente estimado y $Wln(X\Theta)$ alude al retardo espacial del logaritmo de las variables explicativas; los valores estimados de los componentes mencionados anteriormente se encuentran en la tabla 28.

$$\begin{aligned} SMR = \beta_0 + \rho WSMR + ln(X)\beta + Wln(X\Theta) + u, \\ u = \lambda Wu + \epsilon \end{aligned} \tag{31}$$

En el modelo espacial del 2007, se adicionó la variable cobertura de vacunación (VAC) pues presentaba correlación espacial significativa con la SMR; esta adición refinó el ajuste del modelo espacial y contribuyó a aumentar el nivel de explicación de la variable respuesta. No se encontraron pruebas de autocorrelación espacial de otros órdenes en el AEDE por lo que no fue posible incluirlos en el modelo espacial del año en estudio.

Aproximadamente el 90 % de los cambios en la SMR de la neumonía en una localidad son adquiridos por las localidades vecinas en forma positiva, es decir, el aumento de la SMR en una localidad hace que ésta misma variable disminuya en las localidades contiguas. A medida que aumenta el error aleatorio en una región específica, también hace que aumente en las regiones colindantes. El descenso porcentual de la escolaridad, el número de IPS y la cobertura de la vacuna contra la anti haemophilus influenzae tipo B y el aumento porcentual de las prácticas deportivas provocan un aumento en la SMR por neumonía. Del mismo modo, la disminución porcentual de la escolaridad, el número de centros médicos y la cobertura de vacunación y el aumento de las prácticas deportivas en las localidades adyacentes de una localidad redunda en menores tasas de mortalidad para ésta. El modelo es homoscedástico, y los residuos siguen una distribución normal y no están autocorrelacionados en el espacio según el $p-valor$ de los estadísticos mostrados en la tabla 28. La variabilidad del modelo se explica en un 99 % aunque el R-cuadrado de Nagelkerke expone que solo lo hace en un 89 %. Sin embargo, ambos valores son altos y demuestran una buena bondad en el ajuste.

Tabla 28: Estimación modelo de regresión espacial para el año 2007: GNS

| Variables | Estimado | Pvalor |
|---|---|---|
| $\rho$ | -0.90281 | $<$ 2.2e-16 |
| $\lambda$ | 0.9511 | $<$ 2.2e-16 |
| Intercepto | 28.7401 | $<$ 2.2e-16 |
| log(ESC) | -4.3753 | $<$ 2.2e-16 |
| log(DEP) | 0.7153 | $<$ 2.2e-16 |
| log(IPSE) | -0.0971 | $<$ 2.2e-16 |
| log(VAC) | -0.1139 | 0.0039 |
| Wlog(ESC) | -4.3045 | $<$ 2.2e-16 |
| Wlog(DEP) | 0.7030 | $<$ 2.2e-16 |
| Wlog(IPSE) | -0.1017 | $<$ 2.2e-16 |
| Wlog(VAC) | -0.2020 | 5.141e-05 |
| Breush Pagan | | 0.3593 |
| Shapiro-Wilk | | 0.4877 |
| I de Moran: Residuos | | 0.3996 |
| Pseudo R-cuadrado | | 0.9947 |
| Nagelkerke R-cuadrado | | 0.8930 |

*Fuente: Elaboración propia en R*

En el año 2011 el modelo más adecuado es aquel en el que el valor que adopta la SMR en una localidad depende no solo de los condicionantes externos o variables explicativas en la misma localidad, sino también del valor de la SMR en otras localidades vecinas; dicho en otras palabras, un modelo de retardo espacial (Spatial Lag Model). En este modelo la SMR pone en manifiesto un proceso de propagación espacial, de forma que los valores altos y bajos de la SMR en una localidad específica,

realmente están incrementando la probabilidad de ocurrencia de tasas altas y bajas en lugares vecinos.

$$BoxSMR = \beta_0 + \rho WSMR + X\beta + \epsilon \tag{32}$$

Si bien el AEDE identificó algunas variables correlacionadas espacialmente con la SMR y además autocorrelación de tercer orden tanto en la variable endógena como en una variable exógena (CVD), éstas fueron adicionadas junto con el modelo clásico a los modelos espaciales demostrando que muchas no tenían significancia y que el modelo debía prescindir de ellas. Finalmente, luego de varias combinaciones entre las variables candidatas solo la Calidad de vida (CVD) y la vacunación contra la influenza tipo B (VAC) fueron relevantes para la modelación espacial de la SMR. La ecuación 32 presenta la estructura del modelo regresivo espacial para el año 2011, estando la variable respuesta transformada por la ecuación de Box-Cox con un parámetro de transformación $\lambda = -0{,}2329$ ($BoxSMR$), $\beta_0$ el intercepto, $\rho$ el coeficiente de correlación, $W$ la matriz de contigüidad por el criterio de triangulación de Delaunay y $X\beta$ el vector de variables explicativas con su valor estimado mostrados en la tabla 29.

Se estima que el 84 % de las alteraciones de la SMR en una localidad se transmiten a las localidades adyacentes. El aumento marginal en la calidad de vida y la cobertura de vacunación se traducen en la reducción de las tasas de mortalidad por neumonía en una localidad determinada. Los estadísticos de homoscedasticidad, normalidad y autocorrelación en los residuos infieren que el modelo no sufre de heteroscedasticidad y que los residuos siguen una distribución normal corrigiendo la autocorrelación de los mismos en el espacio identificada en el análisis de residuos del modelo estimado por MCO. El Pseudo R-cuadrado presenta un valor de 68 % aunque su versión ajustada exhibe un 79 %; aunque estos valores no son altos, son aceptables a la hora de describir la variabilidad del modelo espacial.

Tabla 29: Estimación modelo de regresión espacial para el año 2011: SAR

| Variables | Estimado | Pvalor |
|---|---|---|
| $\rho$ | 0.8405 | 0.0007694 |
| Intercepto | -4.2287 | 0.00158 |
| CVD | -0.0265 | 0.03392 |
| VAC | -0.0203 | 0.01880 |
| Breush Pagan | | 0.8334 |
| Shapiro-Wilk | | 0.1024 |
| I de Moran: Residuos | | 0.1294 |
| Pseudo R-cuadrado | | 0.6887 |
| Nagelkerke R-cuadrado | | 0.7919 |

*Fuente: Elaboración propia en R*

Como sucedió en el año 2007, el mejor modelo para el 2014 también fue el modelo general anidado (GNS). Este modelo que explica la tasa de mortalidad estandarizada por neumonía, en un punto del espacio geográfico o localidad, en función de ciertas características propias del número de centros médicos (IPSE), la calidad de la vivienda (CVV), la cobertura de vacunación (VAC) y la asistencia escolar (ESC) así como estas mismas variables y la SMR situadas en localidades cercanas. Además, las perturbaciones aleatorias siguen un esquema de autocorrelación espacial indicando que existen variables omitidas espacialmente correlacionadas y que tiene un impacto sobre la SMR en algunas localidades de Bogotá aunque éste sea trasladado al término del error.

$$\begin{aligned} Y = \rho WY + X\beta + WX\Theta + u, \\ u = \lambda Wu + \epsilon \end{aligned} \tag{33}$$

Pese a que el modelo clásico identificó 8 variables explicativas candidatas para el modelo espacial, estando 2 de estas autocorrelacionadas en el espacio con la variable respuesta según en AEDE, la mitad fueron removidas del modelo espacial; en primera instancia, cuando todas estas variables interactuaban en el modelo espacial perturbaban negativamente el criterio de Akaike y en segundo lugar, su posibles combinaciones no eran significativas para el modelo. Por otro lado, no se encontró prueba significativa de que la SMR se encontrara rezagada en el espacio en órdenes superiores, por lo que solo el rezago de primer orden aparece en el modelo. En la ecuación 33 se puede observar el modelo general anidado del 2014. $\beta_0$ representa el intercepto, $\rho$ y $\lambda$ los coeficientes, $W$ la matriz de pesos espaciales de K-dos vecinos, $(X)\beta$ el vector de variables explicativas (número de IPS, calidad de vida, cobertura de vacunación y asistencia escolar) con su correspondiente estimado y $W(X\Theta)$ alude al retardo espacial de las variables explicativas; los valores estimados de los componentes mencionados anteriormente se encuentran en la tabla 30.

Según el $\rho$ del GNS, al rededor del 61 % de las variaciones de la SMR son extendidas a las localidades en un grado de k-2 vecinos; también las perturbaciones aleatorias surgidas en una localidad son contagiadas a sus vecinas. El aumento en las instituciones de asistencia médica y la asistencia escolar influyen en el crecimiento de la SMR en tanto el aumento de la calidad de vida y la cobertura de vacunación disminuyen la tasa. El efecto de estas variables en las localidades adyacentes repercute inversamente en la SMR de la localidad: el aumento de los centros asistenciales y de la asistencia escolar reduce el SMR y el aumento de la calidad de vida y la cobertura de vacunación aumenta los índices de mortalidad. Según el test de Breush-Pagan y con base en el $p-valor$ el modelo es homoscedástico, la prueba de Shapiro-Wilk afirma que los residuos siguen una distribución normal y el test de I de Moran para los residuos confirma que no existe autocorrelación espacial en los residuos del modelo. El Pseudo R-cuadrado presenta un valor del 96 % aproximadamente lo que indica una muy buena bondad del ajuste. El R-cuadrado de Nagelkerke ratifica este valor

con un 94 %. El modelo ajusta bastante bien y se puede decir que la variabilidad del modelo es explicada en una proporción considerable.

Tabla 30: Estimación modelo de regresión espacial para el año 2014: GNS

| Variables | Estimado | Pvalor |
|---|---|---|
| $\rho$ | 0.61273 | 1.7396e-05 |
| $\lambda$ | -1.1021 | $<$ 1.6431e-13 |
| Intercepto | 8.6365e+00 | 0.0008267 |
| IPSE | 6.3359e-04 | $<$ 2.2e-16 |
| CVV | -6.5153e-01 | 1.110e-15 |
| VAC | -1.6851e-02 | 0.0347001 |
| ESC | 1.1958e-01 | 9.601e-06 |
| WIPSE | -4.2588e-04 | 3.256e-05 |
| WCVV | 4.4800e-01 | 0.0361124 |
| WVAC | 3.1851e-02 | 0.0110096 |
| WESC | -3.5561e-01 | 2.220e-16 |
| Breush Pagan | | 0.3836 |
| Shapiro-Wilk | | 0.3581 |
| I de Moran: Residuos | | 0.6571 |
| Pseudo R-cuadrado | | 0.9668 |
| Nagelkerke R-cuadrado | | 0.9472 |

*Fuente: Elaboración propia en R*

Conociendo los modelos de dependencia espacial para los años analizados ya es posible inferir a partir de sus resultados qué impactos directos, indirectos y totales se están presentando en el sistema de localidades de Bogotá. Aunque se podrían analizar directamente los coeficientes de regresión, es ineficiente debido a las externalidades generadas entre los términos en el procesamientos de la información (LeSage y Pace, 2010).

## 6.4. Impactos

A causa de que el impacto de los cambios en una variable explicativa difiere a través de todas las regiones, es necesario encontrar una medida que resuma estas variaciones. Pace y LeSage (2006) determinaron algunas medidas escalares que pueden ser usadas para medir estos impactos a través de todas las regiones.

El **impacto directo** que registra la suma de los impactos en la región que experimenta un cambio, el **impacto indirecto** que captura la suma de los impactos debido a los cambios en otras regiones y el **impacto total** que no es más que la suma del impacto directo e indirecto, es decir, el impacto de una región frente a su cambio y el percibido debido al cambio en otras regiones (Dall'erba y Llamosas-Rosas, 2006).

Gracias a que los modelos autorregresivos espaciales identificados para los años del estudio poseen el retardo espacial de la variable endógena fue posible calcular los tres tipos de impactos basados en los coeficientes de las variables explicativas por cada modelo.

Para el año 2004 en término de los impactos, un aumento del 100 % en la asistencia escolar reduciría la SMR por neumonía en un 43.5 % en el territorio bogotano, un 10 % en la misma localidad y un 33 % aproximadamente en las localidades vecinas. Desde otro ángulo, el aumento en un 100 % de la temperatura conduciría a un aumento del 50 % en el sistema de localidades de Bogotá, el 12 % en la propia localidad y el 38 % en las localidades que guarden relaciones de contigüidad con la localidad en donde se genera el cambio. Los cambios que sucedan en las localidades en un orden 3 del efecto reina, afectaran positivamente la SMR.

Tabla 31: Impactos: 2004

| Variable | Directo | Indirecto | Total |
|---|---|---|---|
| ESC | -0.1077877 | -0.3278840 | -0.4356718 |
| TEM | 0.1252863 | 0.3811138 | 0.5064001 |
| $W_3SMR$ | -1.4045826 | -4.2726595 | -5.772421 |

*Fuente: Elaboración propia en R*

El incremento porcentual de la asistencia escolar, el número de centros asistenciales, la cobertura de vacunación sugieren que se verá un descenso en la tasa de mortalidad para todo Bogotá en el año 2007; siendo el impacto de los cambios en la asistencia escolar mayor en la localidad donde se generó que aquel trasladado sus vecinas. Un fenómeno inverso sucede con la cobertura de vacunación y el número de IPS de una localidad, el impacto indirecto es mayor que el impacto directo en la localidad donde éste es registrado. Es claro que el incremento en el índice de prácticas deportivas en una localidad, se traduce en el incremento de la SMR en la misma en mayor proporción que el contagio a sus vecinas.

Tabla 32: Impactos: 2007

| Variable | Directo | Indirecto | Total |
|---|---|---|---|
| log(ESC) | -3.10041695 | -1.4611606 | -4.5615775 |
| log(DEP) | 0.50955804 | 0.2358837 | 0.7454417 |
| log(IPSE) | -0.04670181 | -0.0577914 | -0.1044932 |
| log(VAC) | 0.24280582 | -0.4088640 | -0.1660581 |

*Fuente: Elaboración propia en R*

Para el año 2011 es claro que el impacto en los cambios de la variación de la calidad de vida y la cobertura de vacunación en una localidad, son sentidos con mayor fuerza en aquellas adyacentes que en sí misma. Un aumento en la calidad de vida

y la cobertura de vacunación reducirían, en todo el territorio capitalino, un 16 % y un 12 % la tasa de mortalidad estandarizada transformada, que si bien resulta de la transformación Box-Cox, debido a la especificación de linealidad del modelo se puede interpretar como la variable sin transformar (Box y Cox, 1964).

Tabla 33: Impactos: 2011

| Variable | Directo | Indirecto | Total |
|---|---|---|---|
| CVD | -0.0365849 | -0.12978095 | -0.1663658 |
| VAC | -0.02800175 | -0.09933317 | -0.1273349 |

*Fuente: Elaboración propia en R*

La SMR del 2014 aminora en Bogotá en un 23 % aproximadamente si la calidad de la vivienda y la asistencia escolar aumentan. Se produce un aumento del 3 % en las localidades de Bogotá cuando el número de IPS y la cobertura de vacunación aumentan. El efecto contagioso para las localidades vecinas indica que a mayor número de centros asistenciales y de asistencia escolar en una localidad reducirá su tasa de mortalidad por neumonía. Por otro lado, características propias de una localidad como la calidad de vivienda y la cobertura de vacunación afectan directamente a la misma: a mayores índices en estas variables menor es la tasa de SMR.

Tabla 34: Impactos: 2014

| Variable | Directo | Indirecto | Total |
|---|---|---|---|
| IPSE | 0.0006247319 | -8.838281e-05 | 0.0005363491 |
| CVV | -0.6400526561 | 1.145056e-01 | -1.682354604 |
| VAC | -0.0117875963 | 5.052093e-02 | 0.0387333321 |
| ESC | 0.0531755410 | -6.626466e-01 | -0.6094710407 |

*Fuente: Elaboración propia en R*

Como se puede observar en las tablas expuestas anteriormente el impacto en la SMR en las localidades de Bogotá puede afectar la estructura mutua entre éstas junto con sus variables, en tanto el cambio en esta estructura mutua puede afectar el impacto de la SMR bogotana.

También es posible realizar el análisis de impactos a nivel local, sin embargo estos solo son observables en los vecinos de una localidad escogida ya que dependen de un modelo global (modelo autorregresivo espacial); con el fin de evaluar este impacto local se propone el uso de la regresión geográficamente ponderada que permitirá cuantificar la proporción y el nivel de significancia en que las variables explicativas afectan localmente las localidades.

## 6.5. Regresión Geográficamente Ponderada

Debido al supuesto de estacionariedad espacial, los parámetros derivados de los modelos autorregresivos espaciales generalmente son aplicados globalmente a través del espacio en donde las observaciones son tomadas. Sin embargo, bajo la inexistencia de estacionariedad espacial estos parámetros podrían no representar la situación verdadera, obviando la consideración de la variación local sobre el espacio (Foody, 2003). En este sentido, para identificar la variación local a través del espacio, es apropiado realizar un análisis local (Regresión geográficamente ponderada) en lugar de un análisis global (modelos autorregresivos espaciales).

La regresión geográficamente ponderada (GWR) es una técnica de regresión basada en el análisis local en donde es posible observar y cuantificar las variaciones en el espacio de las variables explicativas con respecto a su variable respuesta. Se basa en tres principios fundamentales: i) la información espacial no es siempre estacionaria, ii) la estructura espacial de la información influye considerablemente la estimación de las relaciones entre las variables de estudio y iii) esta relación puede ser localizada y alterada en el espacio (Fotheringham, Charlton, y Brunsdon, 1998). Según Brunsdon, Fotheringham, y Charlton (1996) el método consiste en ajustar tantas regresiones como unidades espaciales hayan, teniendo como criterio las distancias entre las mismas, atribuyendo mayor peso a aquellas más cercanas y menor peso a aquellas más lejanas a través de una función kernel. El uso de la GWR es ventajoso no solo por el hecho de ser una perspectiva local del fenómeno en estudio, sino también porque permiten determinar coeficientes independientes y R-cuadrados para cada región mitigando en mayor medida los errores de estimación y problemas de autocorrelación que surgen en los modelos globales (Fotheringham et al., 1998).

El uso de esta técnica para el presente estudio permitió entender en una mayor medida la influencia de los determinantes de la tasa de mortalidad para cada año y cómo ésta afecta las diferentes localidades de Bogotá. Identificada la autocorrelación espacial en el AEDE de la variable respuesta y las variables explicativas se realizó el análisis de correlación entre las mismas concluyéndose que ninguna superaba el nivel de riesgo de $\rho = 0{,}7$ (Rogerson, 2001).

Justo como se realizó con el modelo clásico y el modelo autorregresivo espacial se buscó explicar en gran medida la SMR en función de sus variables explicativas a través de la combinación de las mismas. Los modelos fueron trabajados al 10 % de significancia con el fin de incluir el mayor número de factores posibles.

Para el año 2004 el modelo incorporó las variables asistencia escolar (ESC), la calidad de vida (CVD), los hogares que no consumen las 3 comidas diarias (NUT) , la cobertura de vacunación y la temperatura ambiente (TEM), siendo algunas significativas incluso a un nivel de significancia del 5 %. EL R-cuadrado del modelo es de 0.63, que equivale al nivel de aceptación predictiva moderado.

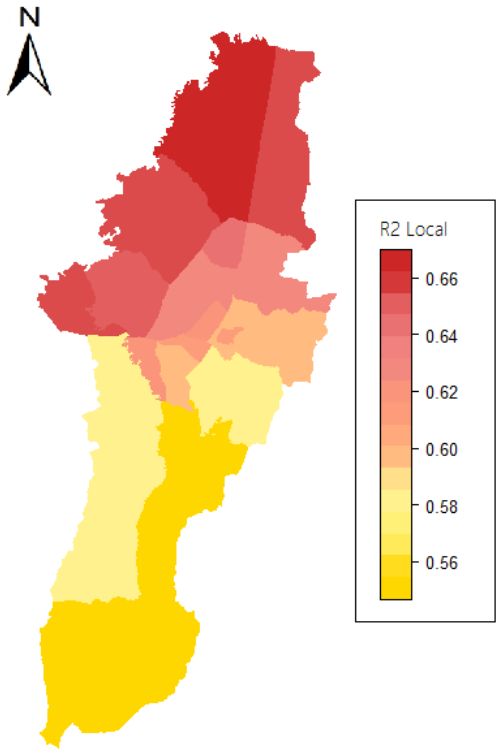
Tabla 35: Estimación regresión geográficamente ponderada: 2004

| Variables | Min | Media | Max |
|---|---|---|---|
| Intercepto | -5.382 | -6.877 | -8.563 |
| ESC | -0.073 | -0.060 | -0.043 |
| CVD | -0.099 | -0.083 | -0.069 |
| NUT | 0.056 | 0.067 | 0.075 |
| VAC | -0.010 | -0.008 | -0.007 |
| TEM | 0.064 | 0.072 | 0.081 |
| R-cuadrado | | 0.6295 | |
| R-cuadrado Ajustado | | 0.3019 | |

*Fuente: Elaboración propia en R*

La distribución del R-cuadrado local a nivel localidad, varía entre 0.56 y 0.66, siendo mayormente explicadas las localidades de la parte norte de Bogotá (Engativá, Suba, entre otras) con valores superiores a 0.60. En la zona sur se puede apreciar que el ajuste es menor; por ejemplo, la localidad de Usme presenta el menor R2 con un valor de 0.56. La capacidad explicativa de las localidades en el centro de Bogotá oscila entre 0.60 y 0.63.

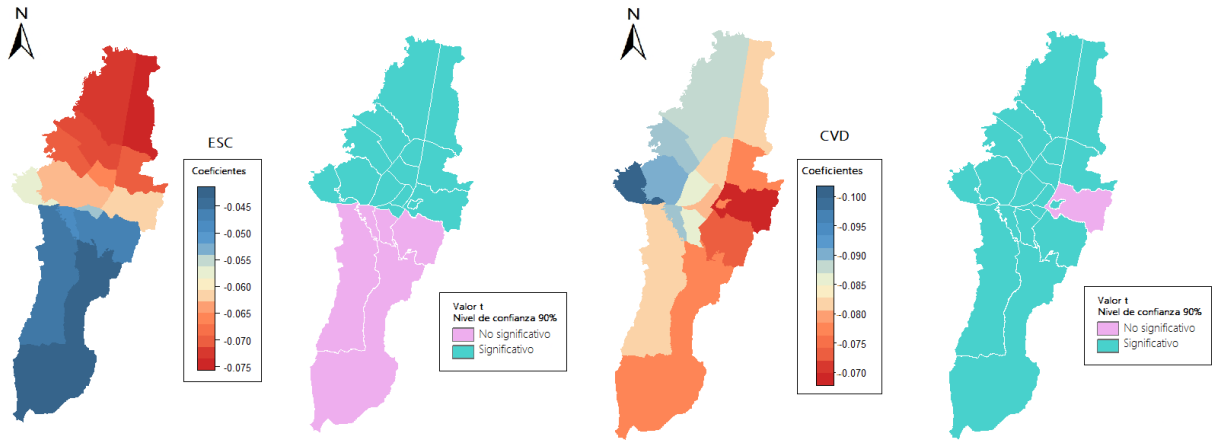
Figura 15: Distribución espacial de los R-cuadrados locales: GWR 2004




*Fuente: Elaboración propia*

De la figura 16 a la figura 18 se muestra el análisis de variabilidad espacial de los coeficientes locales de las variables explicativas para el año 2004. Por cada variable se muestra el valor de los coeficientes junto al mapa de valores t estadísticamente significativos y no significativos para las localidades de Bogotá. La asistencia escolar (Figura 16), muestra que solo los valores concentrados en las localidad del norte y centro de la ciudad son significativos con valores negativos altos, indicando que en éstas localidades es de suma importancia la asistencia escolar con el fin de mitigar el impacto en la SMR. En cuanto a la calidad de vida, su impacto es significativo en todas las localidades de Bogotá excepto en Santa Fe; los coeficientes oscilan entre -0.099 y -0.069 siendo Usme, San Cristóbal y Chapinero las localidades que se verían en menor proporción beneficiadas con la disminución de la SMR si la calidad de vida mejora para sus habitantes. La localidades de Fontibón, Kennedy, Bosa y Tunjuelito reducirían significativamente la mortalidad por neumonía si la calidad de vida aumenta.

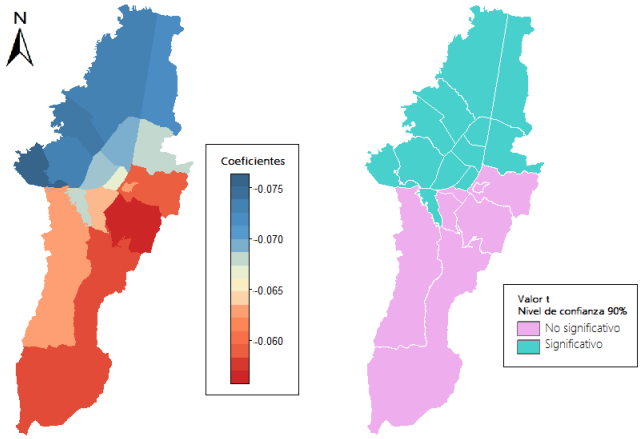
Figura 16: Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2004: Asistencia Escolar y Calidad de Vida.




*Fuente: Elaboración propia*

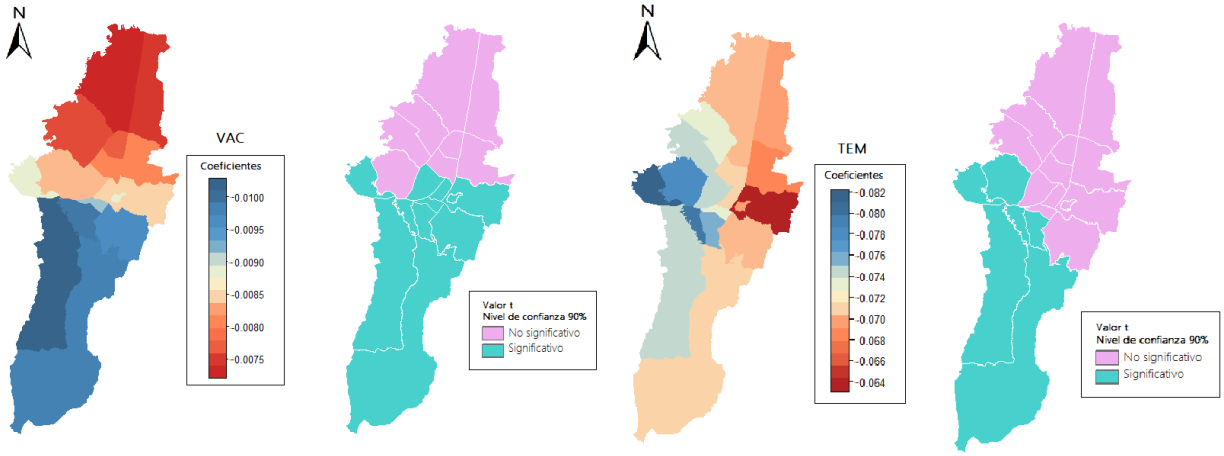
Es también evidente (figura 17) que aquellos hogares en donde no se consumió ninguna de las tres comidas diarias afecta significativamente las localidades al norte y centro de la Ciudad. Bosa sería la localidad en la que disminuiría en mayor proporción la SMR si en los hogares de esta localidad las condiciones nutricionales mejoraran. En Los Mártires, aunque también se obtendría un efecto positivo disminuyendo la SMR, el efecto no sería tan importante como en las demás localidades. Para el sur de la ciudad es inconcluso afirmar el efecto local de la variable, pues a un 10 % de significancia ésta no incide en las localidades de esta sección de la ciudad.

Figura 17: Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2004: Hogares donde no se consumió ninguna de las tres comidas diarias




*Fuente: Elaboración propia*

Figura 18: Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2004: Cobertura de vacunación con anti Haemophilus influenzae tipo B y Temperatura Ambiente.




*Fuente: Elaboración propia*

Para las localidades de Bosa, Ciudad Bolívar, Usme, Tunjuelito , Rafael Uribe Uribe y Kennedy la cobertura de vacunación y la temperatura ambiente inciden significativamente en las perturbaciones a la SMR. Si la cobertura en la vacunación aumentara y la temperatura disminuyera, las tasas de mortalidad por la neumonía bajarían. Por otro lado, el centro de la ciudad solo se vería afectado por la cobertura de vacunación en la misma forma que lo hacen las localidades mencionadas previamente. Las localidades del norte de Bogotá, como Usaquén y Suba no se verían afectadas

localmente ni por la temperatura ambiente o la cobertura de vacunación (Figura 18).

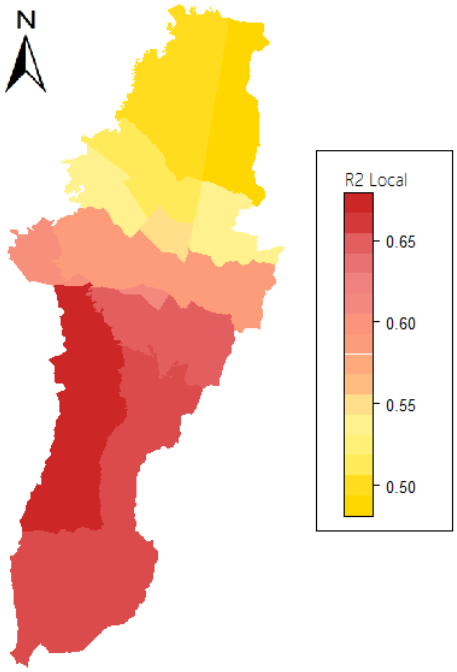
Tabla 36: Estimación Regresión geográficamente ponderada: 2007

| Variables | Min | Media | Max |
|---|---|---|---|
| Intercepto | 3.2343 | 3.7029 | 4.1258 |
| ESC | -0.1089 | -0.0966 | -0.0817 |
| DEP | - 0.0120 | -0.0105 | -0.0091 |
| R-cuadrado | | 0.5497 | |
| R-cuadrado Ajustado | | 0.3727 | |

*Fuente: Elaboración propia en R*

En el año 2007 solo dos variables fueron significativas para explicar la SMR: la asistencia escolar (ESC) y el índice de prácticas deportivas (DEP) siendo ambas significativas al 10 %. La explicación de la variabilidad de la GWR fue del 54 %, lo que indica una explicación moderada. Existe un mayor nivel de explicación local de la SMR en la localidad de Ciudad Bolivar en un 67 % aproximadamente. En las localidades del norte de Bogotá (Engativá, Suba, Usaquén) el ajuste es más pobre con un 50 %.

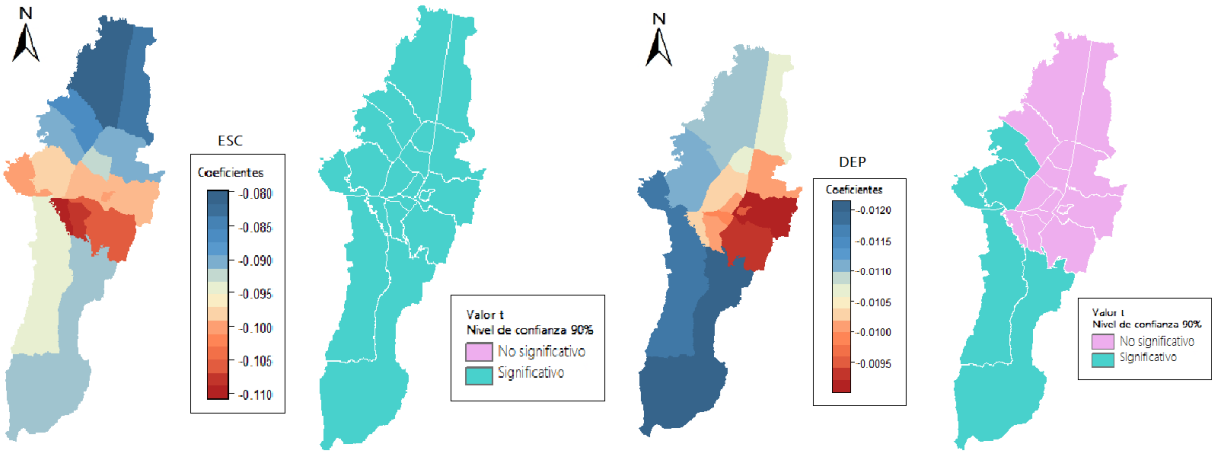
Figura 19: Distribución espacial de los R-cuadrados locales: GWR 2007




*Fuente: Elaboración propia*

En la figura 20 los coeficientes de la variable asistencia escolar son estadísticamente significativos en todo el territorio bogotano. Se observa que la relación de la SMR y la asistencia escolar es negativa en toda el área de estudio, a mayores niveles de asistencia se espera que disminuya la SMR. De la misma forma, se puede observar la variable índice de prácticas deportivas que solo afecta la región sur de Bogotá (Kennedy, Bosa, Usme y Ciudad Bolívar) y a Fontibón indicando que entre menor índice sea reportado es probable que se presenten mayores victimas a causa de la neumonía.

Figura 20: Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2007: Asistencia Escolar e índice de prácticas deportivas.




*Fuente: Elaboración propia*

En el 2011 se encuentra que la SMR se ve afectada por la calidad de vida (CVD) y el número de centros asistenciales presentes en una localidad; el efecto de ambos factores es inverso a la variable respuesta. El R-cuadrado del modelo es de 0.54 indicando que el ajuste aunque no es muy bueno, es moderado. Los coeficientes de la CVD oscilan entre -0.04 y -0.05 y los del número de IPS de -0.0035 a -0.0028.

Tabla 37: Estimación Regresión geográficamente ponderada: 2011

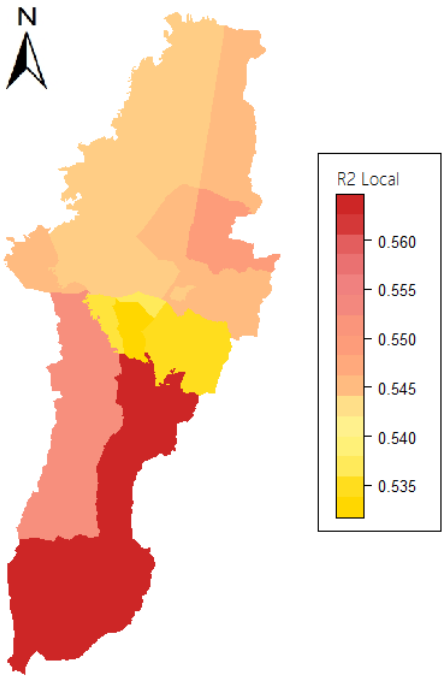
| Variables | Min | Media | Max |
|---|---|---|---|
| Intercepto | -4.2568 | -3.442 | -2.8834 |
| CVD | -0.0579 | -0.0495 | -0.0437 |
| IPSE | - 0.0035 | -0.0031 | -0.0028 |
| R-cuadrado | | 0.5575 | |
| R-cuadrado Ajustado | | 0.4175 | |

*Fuente: Elaboración propia en R*

En la figura 21 se detalla la distribución espacial de los R-cuadrado locales para la zona de estudio. Los R-cuadrado enseñan un ajuste más o menos uniforme a

través de Bogotá distando el mejor ajuste del peor en solo 3 puntos porcentuales aproximadamente. La localidad que presenta mayor bondad en el ajuste es Usme con u valor de 0.56 en tanto la que presenta el índice más bajo es su vecina, San Cristóbal con 0.53. Existe una clara aleatoriedad en el ajuste del R-cuadrado en las 19 localidades.

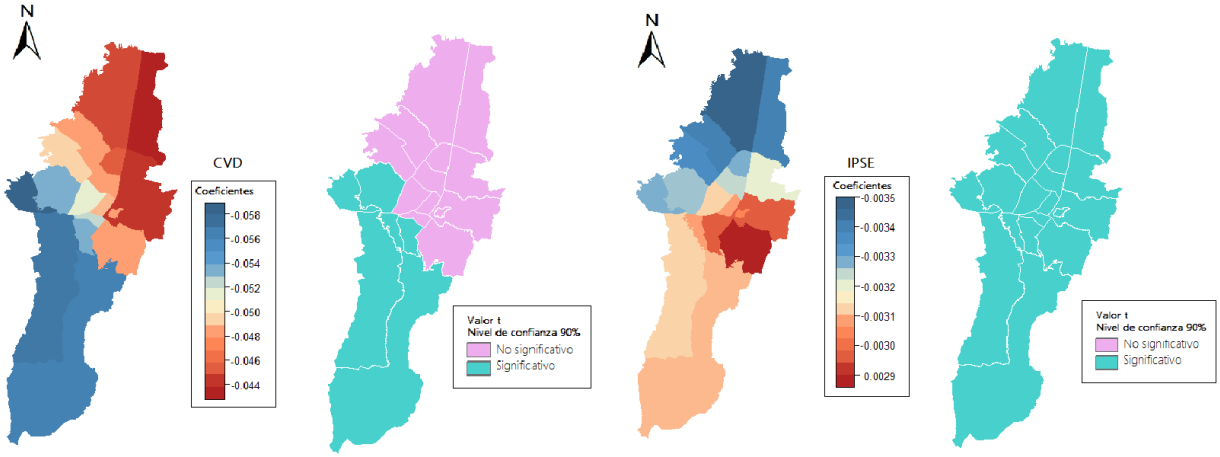
Figura 21: Distribución espacial de los R-cuadrados locales: GWR 2011




*Fuente: Elaboración propia*

La afectación por la calidad de vida es significativa solo en las localidades sur de Bogotá: Usme, Ciudad Bolívar, Bosa, Kennedy, Tunjuelito y Rafael Uribe Uribe. En estas localidades si se evidencia una mejoría en la calidad de vida, la SMR podría disminuir en éstas mismas. En las demás localidades el impacto de la calidad de vida no es estadísticamente significativo, lo que podría indicar que la SMR de esta año podría explicarse a través de otros factores. El número de centros asistenciales incide inversamente en la SMR de todo Bogotá, a menor número de IPS en una localidad la mortalidad por neumonía podría aumentar para este año.

Figura 22: Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2011: Calidad de Vida y número de IPS.




*Fuente: Elaboración propia*

El 2014 es el año con mayor número de variables incidentes en la SMR; se reportan el número de centros asistenciales por localidad (IPSE), la calidad de la vivienda (CVV), la cobertura de vacunación (VAC), la temperatura ambiente (TEM), la contaminación por PM10 y los hogares que no comieron las tres comidas diarias (NUT). Existen relaciones proporcionales e inversamente proporcionales con la mortalidad por neumonía y se aprecia un nivel de explicación alto (93 %) de la variabilidad del modelo.

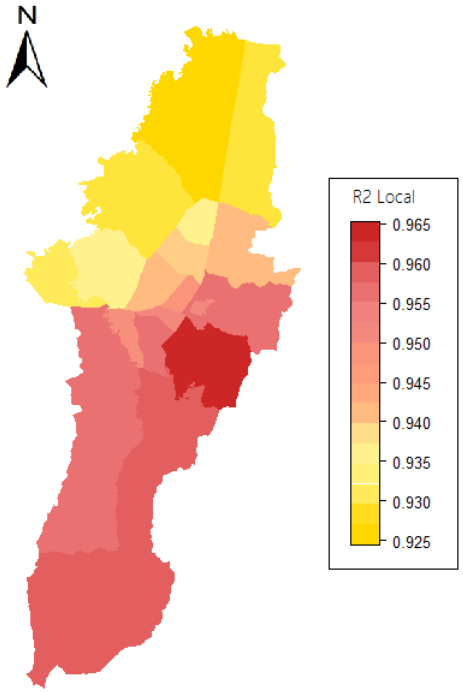
Tabla 38: Estimación regresión geográficamente ponderada: 2014

| Variables | Min | Media | Max |
|---|---|---|---|
| Intercepto | 15.8496 | 16.2748 | 16.7761 |
| IPSE | -0.0004 | -0.0004 | -0.0003 |
| CVV | -1.2261 | -1.1759 | -1.0925 |
| VAC | -0.0356 | -0.0318 | -0.0263 |
| TEM | 0.1399 | 0.2176 | 0.2581 |
| CPM | 0.0085 | 0.0094 | 0.0139 |
| NUT | 0.1099 | 0.1201 | 0.1268 |
| R-cuadrado | | 0.9366 | |
| R-cuadrado Ajustado | | 0.8617 | |

*Fuente: Elaboración propia en R*

El ajuste local muestra que si bien todas las localidades tienen valores altos, la localidad con el mejor ajuste es San Cristóbal, seguida de Usme. De las localidades cuyo ajuste a pesar de ser bueno no supera el umbral del 96.5 % se destacan Usaquén, Suba y Engativá. Es notorio que el mejor ajuste, como se presento en el año 2007 sucedió en la región sur de Bogotá.

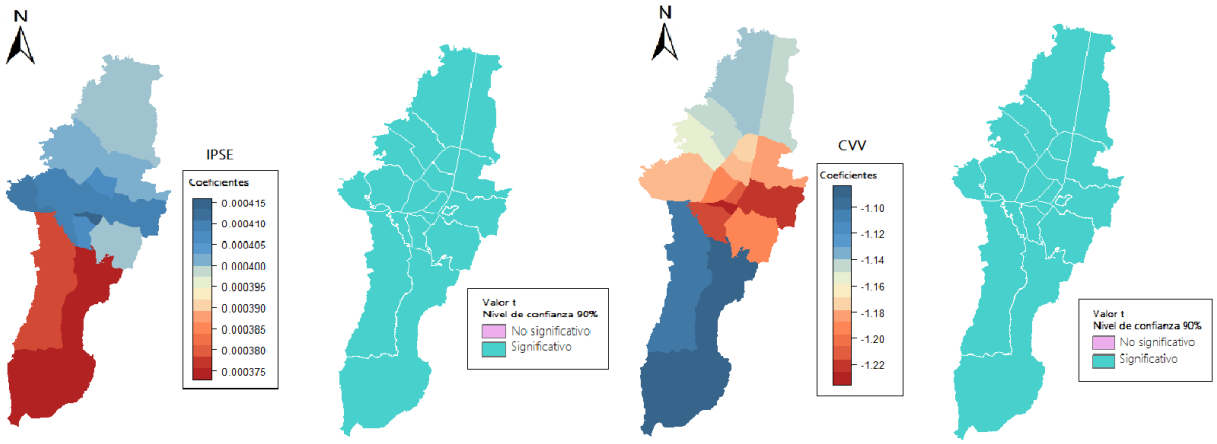
Figura 23: Distribución espacial de los R-cuadrados locales: GWR 2014




*Fuente: Elaboración propia*

Tanto el número de centros asistenciales, la calidad de la vivienda y la cobertura de vacunación son estadísticamente significativos en las 19 localidades que componen el estudio y su impacto es inverso a la variable respuesta. En el caso del número de centros asistenciales se espera que a mayor número de éstos la SMR aminore; este cambio sería más visible en localidades como Usme y Ciudad Bolívar que en Bosa o Tunjuelito. La calidad de la vivienda debería mejorar si se desea que la tasa de mortalidad por neumonía disminuya en Bogotá; Santa Fe y la Candelaria serían las localidades más beneficiadas en la caída de la SMR con mejores condiciones de vivienda. Si se habla de la cobertura de vacunación, se espera que su impacto sea mayor en el centro de la ciudad (La Candelaria, Santa Fe, Los Mártires, Antonio Nariño) que en el sur y norte de la misma. A menores coberturas de vacunación se apreciaría un aumento en la mortalidad de la enfermedad.

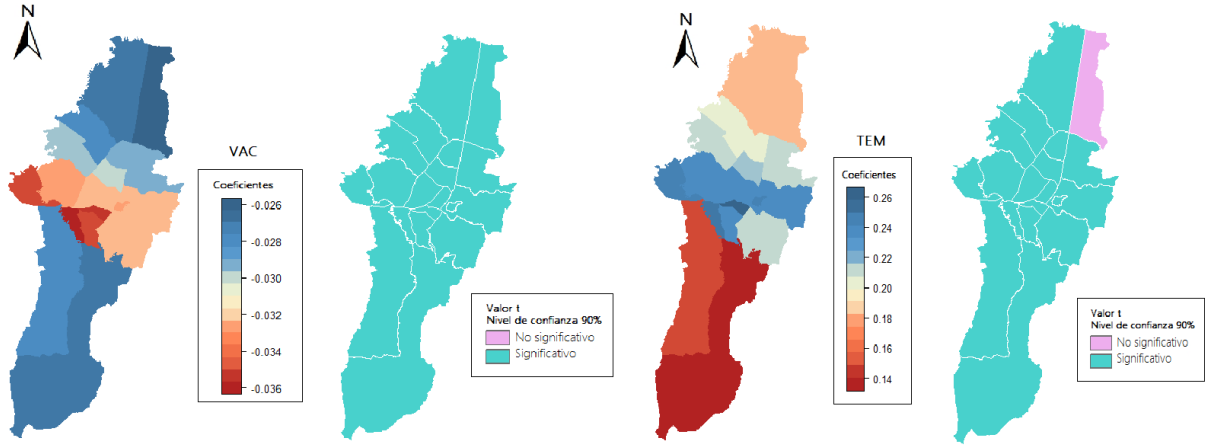
Figura 24: Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2014: Número de IPS y calidad de la vivienda.




*Fuente: Elaboración propia*

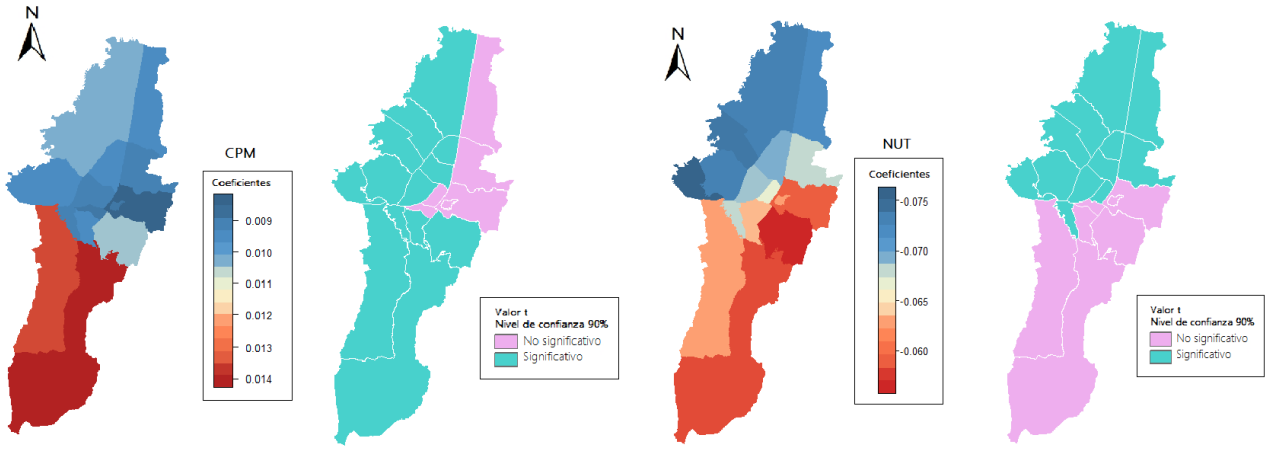
La temperatura ambiente presenta significancia estadística en casi todas las localidades de Bogotá excepto en Usaquén. A Mayores temperaturas mayor será la tasa de mortalidad. Localidades como Usme y Ciudad Bolívar se ven mayormente afectadas que las demás pertenecientes a la ciudad. Existe un comportamiento uniforme en los coeficientes de la TEM en el centro de la ciudad.

Figura 25: Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2014: Cobertura de vacunación y Temperatura Ambiente.




*Fuente: Elaboración propia*

Figura 26: Distribución espacial de los coeficientes locales y P-valores (t student) modelo GWR 2014: Contaminación por PM10 y Hogares que no consumieron las tres comidas diarias.




*Fuente: Elaboración propia*

La contaminación por PM10 es significativa al 10 % en las localidades de Bogotá obviando Usaquén, Chapinero, Santa Fe, La Candelaria, Los Mártires y Antonio Nariño. La concentración en aumento del PM10 en Usme y Ciudad Bolívar conduce a mayores muertos por neumonía en estas localidades. De la misma manera, sucede en las demás localidades aunque en una proporción más pequeña. La afectación de los hogares que no consumen las tres comidas diarias se manifiesta solo en el norte y parte del centro de la ciudad. Suba, Engativá, Fontibón y Bosa se ven afectados en mayor medida por esta variable y se estima que si en los hogares de las localidades surge la posibilidad de comer tres veces al día la SMR disminuirá.

# 7. DISCUSIÓN

Según los resultados se encontraron un número de variables asociadas a la tasa de muerte estandarizada por la neumonía, las cuales desempeñan un rol importante y significativo, no solo en el proceso de adquisición, desarrollo y mortalidad de la enfermedad, sino también en su propagación a lo largo de las localidades de Bogotá. De esta forma, los cambios en el ambiente, las dinámicas socioeconómicas, los hábitos de la población y las variables sanitarias, sumado a la creciente población y las débiles políticas entorno al control epidémico, deben ser consideradas como base para entender en qué medida y en qué dirección estimulan la enfermedad respiratoria.

A través de los años 2004, 2007, 2011 y 2014 se pusieron a prueba las mismas variables sociales, ambientales, conductuales y sanitarias, pero no todas participaron activamente en la determinación de la SMR para sus correspondientes años. Un claro ejemplo fue la cobertura de acueducto (ACU) que, basada en el marco conceptual, debía ser significativa a la hora de explicar la SMR y aún así, no se presentó en ningún año debido a que esta variable presenta una cobertura de casi el 100 % en la ciudad. Según el plan maestro de acueducto y alcantarillado desarrollado en el 2006 (Empresa de Acueducto y Alcantarillado de Bogotá, 2006) se esperaba una cobertura en el servicio del 100 % para todo el territorio bogotano para el año 2015 y aunque no se alcanzó el 100 %, el Observatorio Ambiental de Bogotá reportó un valor del 99,44 % al 99,97 % en el periodo de 10 años. En resumen, la cobertura del servicio en las localidades de Bogotá oscilo entre el 95 % y el 100 % , teniendo una variación menor al 2 % de localidad a localidad lo que explicaría que la cobertura no presentará significancia para describir la SMR. Paralelamente el índice de desempleo (IDD) y la pobreza por NBI (NBI) no fueron factores explicativos de la SMR no obstante se encontró correlación espacial en el análisis exploratorio. Se dedujo que el índice de desempleo y la pobreza por NBI no fueron factores determinantes debido a que presentaban alta correlación entre sí pero poca con respecto a la tasa; aunque la NBI se basa en cuantificar el bienestar y acceso a los recursos privados y colectivos de un individuo, indirectamente mide el estado adquisitivo que alude a los ingresos de un particular, y por ende, a la existencia de un empleo. Si bien se esperaba que el nivel de pobreza por NBI afectara los resultados de la SMR como sucedió en estudios en otros países (Jasti et al., 2008), la medición de esta condición en Colombia incluyendo el distrito capital tiene varias limitaciones en considerar la pobreza en personas con una necesidad básica insatisfecha aunque existan altos niveles de satisfacción en otras necesidades; además de tener varios componentes que dependen de características físicas del entorno y no de los niveles de vida (Departamento Nacional De Planeación, 2001).

El impacto de la asistencia escolar (ESC) es evidente en los años 2004, 2007 y 2014 (Véase 6.3. Modelo autorregresivo espacial). En estos años la relación con la SMR fue inversa, aumentos en la asistencia se traducen en la reducción de la tasa de mortalidad. Tanto el plan sectorial de Educación en Bogotá del 2004-2008 como del

2010-2014 desarrollaron políticas públicas que pretendían aumentar los niveles de asistencia escolar y mitigar el aumento en la deserción escolar. Sin embargo, con base en los datos obtenidos por los censos realizados en Bogotá la asistencia escolar presentó bajas a través de los años afectando en mayor proporción las localidades del sur de la ciudad (Bosa, Kennedy, Usme y Ciudad Bolívar) que fueron el foco de atención por los planes de educación; ni los subsidios condicionados ni la ampliación de la oferta educativa fueron suficientes para incrementar el número de niños, jovenes y adultos en los centros de educación. La Secretaría Distrital de Salud (SDS) estima que más de un 28.5 % de las muertes a causa de la neumonía son atribuidas a pacientes que por diferentes razones, no asisten a entidades educativas o cuyo nivel educativo es inferior al de los pacientes que sobreviven la enfermedad. La Vacuna del Haemophilus Influenza tipo B incidió en la variable respuesta para el 2007, 2011 y 2014. Pese a que el esquema de vacunación obligatorio de los últimos 15 años presentado por la SDS había hecho énfasis en la obligatoriedad de vacunar contra la Haemophilis inflienzae causante de enfermedades bacterianas de alto riesgo como la meningitis y la neumonía, no fue hasta el 2004, después de un brote de neumonía y bronquitis, que Bogotá decidió desarrollar más de 5 campañas anuales de vacunación gratuita contra la influenza y el neumococo. Se estima que gracias a estas medidas el número de infectados menores de 5 años disminuyó en un 47 % del año 2005 al 2016 y la mortalidad para los demás grupos etarios se redujo en un 28 % (Observatorio Ambiental de Bogotá, 2018). Lo previamente expuesto aclararía la poca suficiencia de la variable en el año 2004 y su fuerte incidencia en los años siguientes.

El número de centros asistenciales (IPSE) es una variable explicativa importante para los años 2007 y 2014. De acuerdo al Registro Especial de Prestadores de Servicios de Salud el número de centros asistenciales pasó de 921 en el 2007 a 2109 en el 2014 en el territorio Bogotano. Sin embargo, este aumento fue desproporcional respecto a las localidades y la población de las mismas en Bogotá; Chapinero contenía el 20 % de los centros asistenciales en comparación con localidades como Los Mártires o ciudad Bolívar con tan solo el 0.5 %. Esta disparidad se vio reflejada en la SMR, en la cual las localidades con menos centros asistenciales (Mártires, Ciudad Bolívar, San Cristóbal y Tunjuelito) reportaron índices más altos de mortalidad.

De los años estudiados el 2004 presentó la temperatura ambiente media más alta en Bogotá con un valor de 14.8 grados centígrados (Observatorio Ambiental de Bogotá, 2018). La importancia de este fenómeno se manifestó en el modelamiento espacial de la SMR para el 2004, en donde la temperatura representó significancia a la hora de explicar la SMR y se idéntico que a mayores temperaturas las localidades tendían a presentar mayores muertos por neumonía. Por otro lado, en el 2007 se encontró, según el índice de prácticas deportivas realizado por la encuesta de calidad de vida 2007, que más del 60 % de la población mayor a 13 años tenía tendencia al sedentarismo y se veía afectada en mayor medida por enfermedades de tipo contagioso. En el análisis realizado en este estudio parte de la SMR del 2007 es explicada por el índice de prácticas deportivas; localidades sedentarias suelen tener mayores tasas de mortalidad asociadas a la neumonía. Para el 2011 la calidad de vida es una variable

significativa para explicar la SMR. Algunos estudios sobre desigualdades sociales en salud han mostrado, en ambos sexos y para todas las edades, cómo las clases sociales más desfavorecidas y las personas o áreas geográficas con calidades de vida más pobres tienen peores indicadores de salud que la población de las clases sociales más privilegiadas o las personas o zonas geográficas con mayor riqueza, aspectos muy atingentes a la situación del SMR en este año; las localidades que presentaron menores índices en la calidad de vida se ven más afectadas por la neumonía.

Expuesto lo anterior, se afirma que el estudio y conocimiento de los aspectos no solo epidemiológicos sino también factores sociales, ambientales, conductuales y de asistencia médica espacializados afectan a la neumonía en Bogotá, pueden enriquecer y ser la base para posibles programas de control y prevención de la enfermedad así como un marco conceptual para análisis futuros. Igualmente, la importancia de la dimensión espacial en el proceso de identificación de los múltiples factores de riesgo y su distribución en el territorio. Se espera que el presente análisis sea el pilar de futuras investigaciones y que las entidades públicas y/o privadas que buscan entender las perturbaciones de la neumonía en función de variables no propiamente epidemiológicas hagan uso del mismo en aras de luchar y erradicar la enfermedad respiratoria en la capital colombiana.

# 8. CONCLUSIONES Y RECOMENDACIONES

La unión de los sistemas de información geográfica y la estadística para analizar bases de datos georeferenciadas con temas de salud pública permite estudiar las relaciones y variaciones geográficas de tasas de morbilidad y mortalidad de enfermedades en conjunto con los factores sociales, ambientales y conductuales que coadyudan en su propagación. Los estudios epidemiológicos espaciales se verán cada vez con mayor frecuencia debido al atractivo visual que poseen, la importancia que adquieren las tecnologías geográficas y el afán del hombre por convivir en ambientes más saludables. Es fundamental que estos estudios estén acompañados no sólo de principios epidemiológicos básicos, sino también de los factores externos que determinan a la enfermedad, con el fin de generar soluciones prontas a los problemas de salubridad a partir de una metodología estadística bien definida, teniendo en cuenta la calidad de los datos debido a que pueden presentar confusión y sesgo.

Por otro lado, al estudiar el comportamiento espacial de la mortalidad de la neumonía, en el análisis exploratorio se evidenció que en los últimos años las localidades donde se difundió en mayor medida la tasa de mortalidad fueron Santa Fe, Chapinero, San Cristobal, regiones en las que se recomienda realizar programas en los que se priorice la disminución de las necesidades básicas insatisfechas y se ofrezcan empleos para la población que está inactiva con el fin de que se aumenten los indicadores de calidad de vivienda y de hogares que puedan tener por lo menos tres comidas diarias. También, se deben homogeneizar las regiones con instituciones prestadoras de salud para que hayan más equipamientos que soporten la cantidad de población que tiene la ciudad. La contaminación es uno de los mayores retos debido a que en todo el territorio bogotano existen diferentes fuentes de contaminación, y tratar una causa específica, con el propósito de observar su incidencia en la mortalidad por la enfermedad, se propondría para estudios futuros. Al ser la temperatura un factor fluctuante por las características climáticas de la ciudad, evitar la morbilidad y mortalidad de la neumonía a causa de ésta es responsabilidad de la población tomando medidas mínimas para cuidar su salud. La combinación de estas variables permitirá tasas más bajas en las ubicaciones mencionadas y por lo tanto, en regiones vecinas como lo mostró este trabajo.

Se recomienda a las entidades del estado, en especial a la SDS, reparar en los factores previamente analizados no sólo con el fin de reducir su impacto en la mortalidad de la neumonía, sino también el mismo en diversas enfermedades y/o fenómenos que afligen a la población Bogotana. También se sugiere la participación activa de éstas entidades en los censos y la recopilación de la información de la población y el entorno, con el fin de enriquecer estudios similares y fortalecer las conclusiones y procedimientos a la que es sometida la información.

# A. Anexo: Datos para el cálculo de la SMR por neumonía

Tabla 39: Datos para hallar SMR 2004-2007

| Año | 2004 | | | 2007 | | |
|---|---|---|---|---|---|---|
| Localidad | Total | Población | SMR | Total | Población | SMR |
| Usaquen | 50 | 451959 | 0.928 | 69 | 479403 | 1.124 |
| Chapinero | 26 | 139231 | 1.566 | 19 | 147685 | 1.005 |
| Santa Fe | 23 | 129107 | 1.494 | 19 | 136947 | 1.084 |
| San Cristobal | 71 | 500396 | 1.190 | 59 | 530781 | 0.868 |
| Usme | 34 | 259023 | 1.101 | 34 | 274751 | 0.967 |
| Tunjuelito | 33 | 231047 | 1.198 | 34 | 245076 | 1.084 |
| Bosa | 43 | 538357 | 0.670 | 55 | 571047 | 0.752 |
| Kennedy | 87 | 920233 | 0.793 | 94 | 976111 | 0.752 |
| Fontibón | 32 | 291431 | 0.921 | 45 | 309128 | 1.137 |
| Barrios Unidos | 30 | 183091 | 1.374 | 35 | 194208 | 1.408 |
| Teusaquillo | 19 | 159666 | 0.998 | 21 | 169362 | 0.968 |
| Los Mártires | 20 | 103253 | 1.625 | 23 | 110583 | 1.625 |
| Engativá | 89 | 816070 | 0.915 | 99 | 865623 | 0.893 |
| Suba | 86 | 825011 | 0.874 | 122 | 875108 | 1.089 |
| Antonio Nariño | 17 | 106676 | 1.337 | 17 | 113153 | 1.174 |
| Puente Aranda | 51 | 295981 | 1.446 | 55 | 313954 | 1.368 |
| Candelaria | 3 | 27552 | 0.913 | 5 | 29225 | 1.336 |
| Rafael Uribe Uribe | 65 | 433352 | 1.258 | 62 | 459666 | 1.054 |
| Ciudad Bolivar | 59 | 617491 | 0.801 | 87 | 654987 | 1.037 |

*Fuente: Elaboración propia en R*

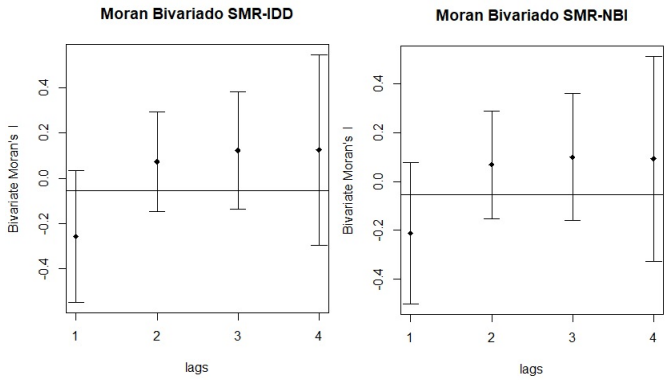
Tabla 40: Datos para hallar SMR 2011-2014

| Año | 2011 | | | 2014 | | |
|---|---|---|---|---|---|---|
| Localidad | Total | Población | SMR | Total | Población | SMR |
| Usaquen | 75 | 474773 | 1.451 | 79 | 472908 | 1.562 |
| Chapinero | 28 | 133778 | 1.922 | 36 | 126951 | 2.652 |
| Santa Fe | 15 | 109993 | 1.252 | 17 | 96534 | 1.647 |
| San Cristobal | 64 | 409799 | 1.434 | 55 | 396383 | 1.298 |
| Usme | 33 | 382876 | 0.791 | 32 | 337152 | 0.887 |
| Tunjuelito | 17 | 201843 | 0.773 | 26 | 189522 | 1.283 |
| Bosa | 45 | 583056 | 0.708 | 47 | 709039 | 0.620 |
| Kennedy | 83 | 1019949 | 0.747 | 107 | 1187315 | 0.843 |
| Fontibón | 29 | 345909 | 0.770 | 21 | 403519 | 0.486 |
| Barrios Unidos | 30 | 233781 | 1.178 | 27 | 263883 | 0.957 |
| Teusaquillo | 19 | 146583 | 1.190 | 28 | 140767 | 1.860 |
| Los Mártires | 17 | 97926 | 1.594 | 79 | 873243 | 0.846 |
| Engativá | 89 | 843722 | 0.968 | 125 | 1250734 | 0.934 |
| Suba | 97 | 1069114 | 0.833 | 13 | 94130 | 1.292 |
| Antonio Nariño | 22 | 108307 | 1.865 | 7 | 109277 | 0.599 |
| Puente Aranda | 35 | 258441 | 1.243 | 35 | 225220 | 1.453 |
| Candelaria | 5 | 24144 | 1.902 | 7 | 22633 | 2.893 |
| Rafael Uribe Uribe | 40 | 377615 | 0.973 | 51 | 353761 | 1.348 |
| Ciudad Bolivar | 70 | 639937 | 1.004 | 61 | 719701 | 0.792 |

*Fuente: Elaboración propia en R*

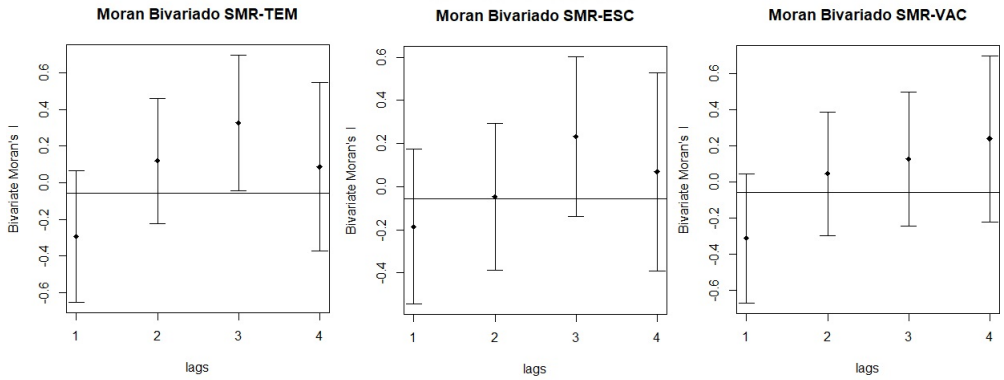
# B. Anexo: Correlogramas bivariados

Figura 27: Moran bivariado 2004



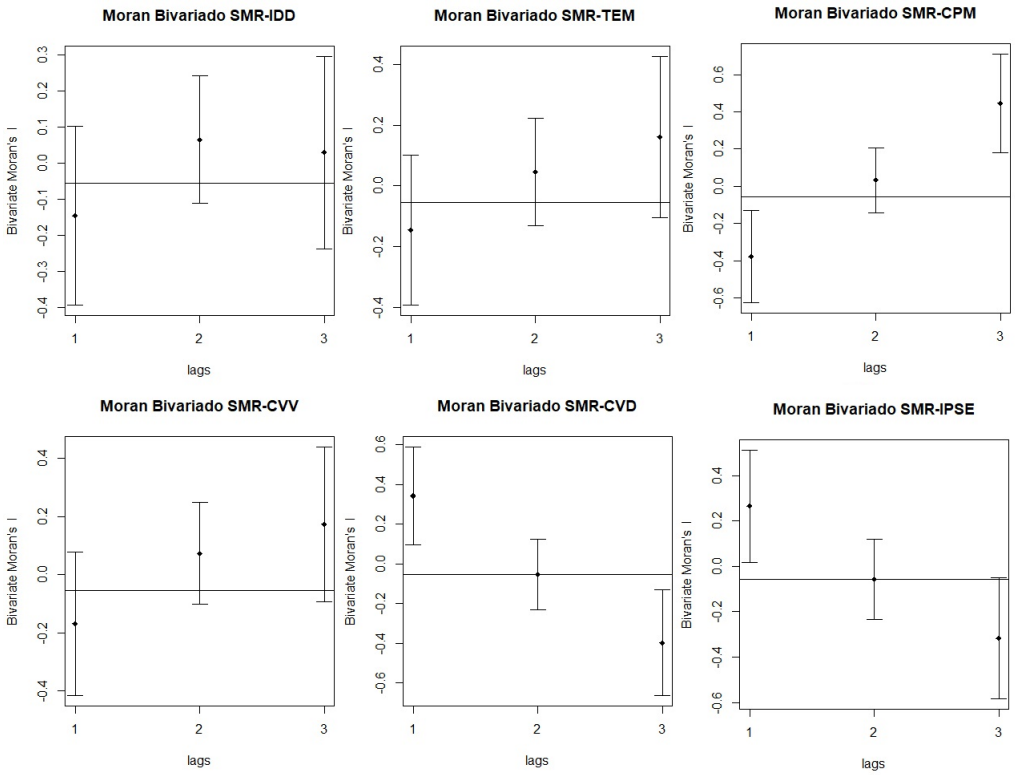

*Fuente: Elaboración propia*

Figura 28: Moran bivariado 2007



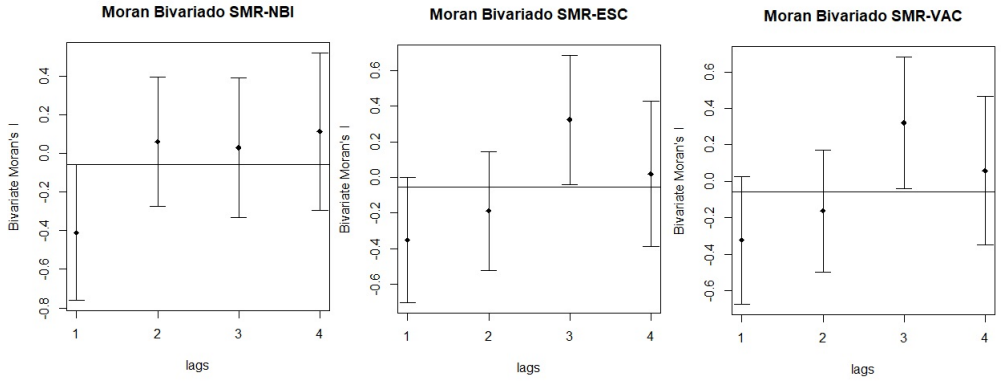

*Fuente: Elaboración propia*

Figura 29: Moran bivariado 2011




*Fuente: Elaboración propia*

Figura 30: Moran bivariado 2014




*Fuente: Elaboración propia*

# BIBLIOGRAFÍA


Acevedo, I., y Velásquez, E. (2008). Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. *Ecos de Economía*, *12*(27).

Alcaldía Mayor de Bogotá. (2018). *Ciudad: Ubicación.* Descargado 2018-01-12, de `https://http://www.bogota.gov.co/`

Ancot, L., Paelinck, J., Klaassen, L., Molle, W., Albegov, M., Andersson, A., y Snickars, F. (1982). Topics in regional development modelling. *M. Albegov, Å. Andersson and F. Snickars (eds, pp. 341-359), Regional Development Modelling in Theory and Practice. Amsterdam: North Holland*.

Anselin, L. (1988a). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, *20*(1), 1-17.

Anselin, L. (1988b). Spatial econometrics: Methods and models. En (1st ed.). Bogotá D.C., Colombia: Springer Netherlands.

Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical analysis*, *27*(2), 93–115.

Aroca, P., y Bosch, M. (2000). Crecimiento, convergencia y espacio en las regiones chilenas: 1960-1998. *Estudios de economía*, *27*(2).

Baik, I., Curhan, G. C., Rimm, E. B., Bendich, A., Willett, W. C., y Fawzi, W. W. (2000). A prospective study of age and lifestyle factors in relation to community-acquired pneumonia in us men and women. *Archives of Internal Medicine*, *160*, 3082–3088.

Baronio, A., Vianco, A., y Rabanal, C. (2012). Una introducción a la econometría espacial: dependencia y heterogeneidad. *Catedra de econometría*.

Bohorquez, M. (2009). Estadística espacial. *Universidad Nacional de Colombia, Departamento de Estadística*.

Box, G., y Cox, R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *26*, 211–252.

Brunsdon, C., Fotheringham, A. S., y Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, *28*(4), 281-298.

Buzzoa, A. R., Roberts, C., García Mollinedo, L., Mould, J., Luna Casas, G., y Sobrino Soldevilla, J. M. (2013). Morbidity and mortality of pneumonia in adults in six latin american countries. *International Journal of Infectious Diseases*, *17*(9), 673-677.

Cameron, D., y Jones, I. G. (1983). John snow, the broad street pump and modern epidemiology. *International journal of epidemiology*, *12*(4), 393-396.

Canavos, G. (1992). *Probabilidad y estadística; aplicaciones y métodos.* Iztapalapa, México: Mac Graw Hill.

Chasco Yrigoyen, C. (2003). Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales. En (1ra ed.). Madrid, España: Consejería de Economía e Innovación Tecnológica.

Cliff, A., Hagget, P., y Ord, J. (1986). Spatial aspects of influenza epidemics.

London: Pion.

Conley, T. R., y Ligon, E. A. (2002). Economic distance, spillovers, and cross country comparisons. *Journal of Economic Growth*, *7*, 157-187.

Cressie, N. N. A. (1993). Statistics for spatial data. En (Revised Edition ed.). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Dall'erba, S., y Llamosas-Rosas, I. (2006). Spatial policy for growth and equity. En M. Fischer y P. Nijkamp (Eds.), *Handbook of regional science* (p. 334-370). Berlin, Heidelberg: Springer.

Davis, R., Dougherty, E., McArthur, C., Huang, Q. S., y Baker, M. G. (2010). Influence of socioeconomic status on community-acquired pneumonia outcomes in elderly patients requiring hospitalization: a multicenter observational study. *BMC Public Health*, *10*(421).

Departamento Nacional De Planeación. (2001). Coyuntura económica e indicadores sociales. Bogotá, Colombia.

Desai, G., Sucher, K., y Chen, M. (2004). Impact of malnutrition in patients with pneumonia on length of stay. *Journal of the Academy of Nutrition and Dietetics*, *97*(9), A50.

Dray, S., Legendre, P., y Peres-Neto, P. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (pcnm). *Ecological Modelling*, *196*, 483-493.

Empresa de Acueducto y Alcantarillado de Bogotá. (2006). Plan maestro de acueducto y alcantarillado. Bogotá, Colombia.

European Respiratory Society. (2016). *Forum of international respiratory societies calls for global effort on world pneumonia day.* Descargado 2017-09-12, de `https://www.ersnet.org/the-society/news/forum-of-international-respiratory`

Evans, R. G., y Stoddart, G. L. (1990). Producing health, consuming health care. *Social Science and Medicine*, *31*, 1347–1363.

Farr, B. M., Bartlett, C. L. R., y Miller, D. L. (2000). Risk factors for community-acquired pneumonia diagnosed upon hospital admission. *Respiratory Medicine*, *94*, , 954–963.

Fine, M., Auble, T., Yealy, D., Hanusa, B., Weissfeld, L., y Singer, D. (1997). A prediction rule to identify low-risk patients with community-acquired pneumonia. *Journal of Medicine*, *336*, 243–250.

Florax, R., y Folmer, H. (1992). Specification and estimation of spatial linear regression models: Monte carlo evaluation of pre-test estimators. *Regional Science and Urban Economics*, *22*, 404-432.

Flory, J. H., Joffe, M., Fishman, N. O., Edeistein, P., y Metlay, J. (2009). Socioeconomic risk factors for bacteraemic pneumococcal pneumonia in adults. *Epidemiology and Infection*, *137*(5), 717-726.

Foody, G. M. (2003). Geographical weighting as a further refinement to regression modeling: an example focused on the ndvi-rainfall relationship. *Remote Sensing of Environment*, *88*, 283-293.

Fotheringham, A. S., Charlton, M., y Brunsdon, C. (1998). Geographically weigh-

ted regression: a natural evolution of the expansion method for spatial data analysis. *Environmental Planning A*, *30*(11), 1905-1927.

Galeano, C., Parra, F., Santa Guzmán, L., y Rodríguez, L. (2017). Modelos de regresión espacial para describir el comportamiento del cáncer gástrico en colombia para el periodo 2005-2012. *UD y la geomática*(11), 27–38.

Hennessy, T., Ritter, T., Holman, R., Bruden, D., Yorita, K., Bulkow, L., ... Smith, J. (2008). The relationship between in-home water service and the risk of respiratory tract, skin, and gastrointestinal tract infections among rural alaska natives. *CAm J Public Health*, *98*(11). doi: 10.2105/AJPH.2007.115618

Hyeyoung, L., y Sujit K., G. (2009). Performance of information criteria for spatial models. *J Stat Comput Simul*, *79*(1), 93–106.

Izquierdo, C., Oviedo, M., Ruiz, L., Sintes, X., Vera, I., Nebot, M., ... Domínguez, A. (2016). Cold, dry air is associated with influenza and pneumonia mortality in auckland, new zealand. *Influenza Other Respir Viruse*, *10*(4), 310-313.

Jasti, H., Mortesen, E., Obrosky, D., Kapoor, W., y Fine, M. (2008). Causes and risk factors for rehospitalization of patients hospitalized with community-acquired pneumonia. *Clinical Infectious Diseases*, *46*(4), 550-556.

Kaplan, V., Clermont, G., Griffin, M. F., Kasal, J., Watson, R. S., y Linde-Zwirble, W. T. (2003). Pneumonia: Still the old man's friend? *Archives of Internal Medicine*, *163*, 317–323.

Kramer, W., y Sonnberger, H. (1986). The linear regression model under test. En (1st ed.). Heidelberg, Alemania: Physica-Verlag Heidelberg.

Last, L. A. (1995). A dictionary of epidemiology. En (3rd ed.). Oxford: Oxford University Press.

LeSage, J. P., y Pace, R. K. (2010). Spatial Econometric Models. *Handbook of Applied Spatial Analysis*, 355–376.

Loeb, M. (2003). Acquired pneumonia in older people: The need for a broader perspective. *Journal of the American Geriatrics Society*, *51*, 539–543.

Loeb, M. (2004). Use of a broader determinants of health model for community-acquired pneumonia in seniors. *Clinical infectious diseases*, *38*(9), 1293–1297.

Martori, J. C., Hoberg, K., y Madariaga, R. (2008). La incorporación del espacio en los métodos estadísticos: autocorrelación espacial y segregación. *Actas del X Coloquio Internacional de Geocrítica*.

McMillen, D. P. (1996). One hundred fifty years of land values in chicago: A nonparametric approach. *Journal of Urban Economics*, *40*, 100-124.

McMillen, D. P. (2003). Spatial autocorrelation or model misspecification? *International Regional Science Review*, *26*, 208-217.

Moreno, R., y Vayá, E. (2000). Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial. En (UB 44 Manuelas ed.). Barcelona, España: Edicions Universitat Barcelona.

Moreno, R., y Vayá, E. (2002). Econometría espacial: nuevas técnicas para el análisis regional. una aplicación a las regiones europeas. *Investigaciones Regionales*, *1*, 83-106.

Observatorio Ambiental de Bogotá. (2018). *Indicadores de ciudad.* Descar-

gado 2018-03-28, de http://oab2.ambientebogota.gov.co/es/todos-los-indicadores

Ortiz, C., Rúa-Uribe, G., Suarez, C., y Mafla, M. A. (2013). Distribución espacial de casos e incidencia de dengue: análisis de la situación para medellín-colombian. *Revista Facultad Nacional de Salud Pública*, *31*(4).

Pace, K., y LeSage, J. P. (2004). Spatial autoregressive local estimation. *Spatial econometrics and spatial statistics*, 31–51.

Pace, K., y LeSage, J. P. (2006). Interpreting spatial econometric models. En M. Fischer y P. Nijkamp (Eds.), *Handbook of regional science* (p. 1535-1552). Berlin, Heidelberg: Springer.

Pamplona, P., Jaramillo, A., y Alba, A. (2015). Análisis epidemiológico espacial de la incidencia de tuberculosis en el departamento de antioquia periodo 2009 - 2013. *Ingenierías USB Medellín*.

Perdomo, C. (2005). Cómo enseñar autocorrelación espacial. *Economía Sociedad y Territorio*.

Quah, S. R. (2017). *International encyclopedia of public health (second edition)* (n.º 978-0-12-803708-9). Elsevier Inc.

Rogerson, P. A. (2001). Statistical Methods for Geography. En (1st ed.). London, UK: SAGE Publications.

Schabenberger, O., y Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. CRC press.

Secretaria Distrital de Salud de Bogotá. (2004). Guías para la prevención, control y vigilancia epidemiológica de infecciones intrahospitalarias: Neumonia nosocomial. En (1ra ed.). Boston, Massachusetts: Esfera Editores Ltda.

Serrano, R. M., y Valcarce, E. V. (2000). *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial* (Vol. 44). Edicions Universitat Barcelona.

Skinner, K. R., Montgomery, D. C., Runger, G. C., Fowler, J. W., McCarville, D. R., Rhoads, T. R., y Stanley, J. D. (2002). Multivariate statistical methods for modeling and analysis of wafer probe test data. *IEEE transactions on semiconductor manufacturing*, *15*(4), 523–530.

Taboada, L. B., Leal, A. L., Caicedo, M. P., Camargo, C. B., y Roa, J. H. (2015). Etiología de la neumonía adquirida en la comunidad en un hospital de cuarto nivel en bogotá: estudio descriptivo de un registro institucional durante los años 2007 a 2012. *Infectio*, *19*(1), 393-396.

Vincent, J.-L. H., Jesse, B., y Vincent, J. B. (2012). *Encyclopedia of intensive care medicine* (n.º 616.028). SpringerLink (Online service).

Waller, L. A., y Gotway, C. A. (2004). Applied spatial statistics for public health data. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Wood, E., Sallar, A. M., Schechter, M. T., y Hogg, R. S. (1999). Social inequalities in male mortality amenable to medical intervention in british columbia. *Social Science and Medicine*, *48*, , 1751–1758.

**Mediante las siguientes firmas certificamos que el proyecto aquí desarrollado es ORIGINAL y para efectos legales toda la información es verídica y puede ser verificada en cualquier momento por la Universidad Distrital Francisco José de Caldas y/o entidades externas.**

| | |
|---|---|
| David Enrique Payares García | Norgith Bibiana Quintero Alonso |