



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL
Facultad de Ciencias Naturales y Matemáticas

“Estudio analítico para la predicción de las tasas de exceso de muertes por COVID-19 en una ciudad, implementando técnicas de Machine Learning”

TESIS DE GRADO

Previo la obtención del Título de:
INGENIERA EN ESTADÍSTICA INFORMÁTICA
INGENIERA EN ESTADÍSTICA

Presentado por:
Ambar Elizabeth Borbor Cedeño
Alexia Mayling Pareja Proaño

GUAYAQUIL - ECUADOR

Año: 2022

DEDICATORIA

A mis padres, a ellos les debo todo lo que soy. A mis hermanas, por su apoyo y amor constante.

Ambar Elizabeth Borbor Cedeño

A mis padres, Vilma y Alex por su incondicional amor y apoyo.

A mi tío Geovanny y a mi abuelo Pepe, quienes hubieran celebrado este logro como propio.

Alexia Mayling Pareja Proaño

AGRADECIMIENTOS

Agradezco a Dios por brindarme la capacidad y sabiduría para culminar esta etapa.

A Roberto, Alba y hermanas, no existen palabras que puedan dimensionar mi gratitud y amor hacia ustedes.

A mis perritas, gatitas y Avery, han sido mi escape de momentos de estrés.

A mis amigos y compañeros de ESPOL, su compañerismo me ayudó a llegar aquí.

A Alexia, la mejor compañera, tu amistad, dedicación y profesionalismo fueron clave para culminar este proceso favorablemente.

Ambar Elizabeth Borbor Cedeño

A mis padres, Alex y Vilma, a mi hermano Alecksey, a mis abuelitas, Martha y Piedad, quienes me inspiran a ser mejor cada día.

A toda mi familia, que siempre estuvo pendiente de mis pasos y presta a brindarme su apoyo.

A Alaska y a Álstor por estar a mi lado cada noche de desvelo.

A Marlon por acompañarme en todos los momentos de este camino.

A mis amigos y compañeros de ESPOL, por quienes recordaré con gran alegría esta etapa.

A Ambar por su gran amistad, compañía y dedicación, sin quien este paso final no hubiera sido posible.

Alexia Mayling Pareja Proaño

A la ESPOL y docentes que han formado parte de nuestra formación académica, por compartir sus conocimientos y valores a lo largo de la carrera. A la M.Sc. Heydi Roa, su guía y apoyo fueron indispensables para poder terminar con éxito el proyecto. A la PhD Sandra García y al PhD Francisco Vera por sus valiosos aportes en el desarrollo de este estudio.

Ambar y Alexia

DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, nos corresponden conforme al reglamento de propiedad intelectual de la institución. Nosotras, Ambar Elizabeth Borbor Cedeño y Alexia Mayling Pareja Proaño damos nuestro consentimiento para que la ESPOI realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”

Ambar Elizabeth Borbor Cedeño

Alexia Mayling Pareja Proaño

EVALUADORES



**PhD Sandra Lorena
García Bustos**

PROFESOR DE LA MATERIA

**M.Sc. Heydi Mariana
Roa López**

PROFESOR TUTOR

RESUMEN

El cálculo de las tasas de excesos de muertes en una comunidad es de gran importancia para la salud pública porque además de permitir a las autoridades ver desde varias aristas esta problemática, también ayuda a que las autoridades respectivas prioricen recursos que pueden resguardar la vida de las personas. Adicionalmente al utilizar variables no observables, como el índice de vulnerabilidad y las tasas de morbilidad para modelar estos excesos de muerte, entran en consideración factores que permitirán visualizar las zonas que necesitarán atención prioritaria en estados de calamidades naturales o sanitarias. En este trabajo en particular se modelan las tasas de excesos de muertes de los sectores censales del cantón Durán durante el año 2020. Este período de tiempo contempla la afectación por el COVID-19, debido a esto, las morbilidades a considerar serán aquellas agravantes de la infección provocada por este virus. Las tasas obtenidas se obtienen implementando modelos de Machine Learning: regresión espacial ponderada con distribución Poisson y modelos de mezcla con distribución Poisson. Considerando los resultados obtenidos por la técnica de K-Folds Cross Validation se seleccionó el modelo de mezcla con distribución Poisson y $k = 5$, ya que fue el que obtuvo el menor promedio de las Sumas Cuadráticas del Error: 12.03, esto permite la determinación, a través de escenarios futuros, de las zonas que deben ser atendidas en primera instancia.

Palabras Clave: Exceso de muertes, Poisson, vulnerabilidad, morbilidad, machine learning.

ABSTRACT

The calculation of death excess rates in a community is of great importance for public health because it allows the respective authorities to prioritize the resources that can protect people's lives. Additionally, when using unobservable variables, such as the vulnerability index and morbidity rates to model these excesses, factors that will allow us to visualize the areas that will need priority attention in states of natural or sanitary calamities come into consideration. In this case, the deaths excess rates of the census sectors of the Duran canton during the year 2020 will be modeled. This period contemplates the affectation by COVID-19, due to this, the morbidities to be considered will be those aggravating the infection caused by this virus. The rates obtained were modeled by implementing Machine Learning models: weighted spatial regression with Poisson distribution and mixing models with Poisson distribution. Considering the results obtained by the K-Folds Cross Validation technique, the mixture model with Poisson distribution and $k = 5$ was selected, because it presented the lowest average of the Square Sums of the Error: 12.03, this allows us to determine in future scenarios the areas that should be served in the first instance.

Keywords: *Death excess, vulnerability, morbidity, machine learning*

ÍNDICE GENERAL

CAPÍTULO 1	15
1. Introducción	15
1.1. Descripción del problema	16
1.2. Justificación del problema	17
1.3. Objetivos	19
1.3.1. Objetivo general.....	19
1.3.2. Objetivos específicos	19
1.4. Marco teórico.....	19
CAPÍTULO 2	30
2. Metodología	30
2.1. Área de estudio	30
2.2. Fuente de los datos	31
2.3. Correlación espacial.....	33
2.4. Definición de variables	34
2.5. Aplicación de modelos	34
2.6. Selección del mejor modelo	36
2.7. Limitantes.....	36
CAPÍTULO 3	37
3. Resultados.....	37
3.1. Sector es urbanos donde se registran casos de COVID-19	38
3.2. Serie de tiempo SARIMA para la estimación de exceso de muertes para el 2020	39
3.3. Índice Global de Morán para las variables independientes del modelo	42
3.4. Índice de LISA para las variables independientes del modelo	43
3.5. Mapas de correlación espacial mediante clusterización de LISA	44

3.6.	Mapas de las variables independientes.....	46
3.7.	Prueba de bondad de ajuste para la variable de exceso de muertes	47
3.8.	Modelo de regresión espacial ponderada con distribución Poisson ..	48
3.9.	Modelo de mezcla Poisson.....	49
3.10.	Elección del mejor modelo	49
3.11.	Presentación de tasas.....	51
3.12.	Mapa de calor del conjunto de prueba para sus valores de exceso de muertes versus el exceso de muertes estimado	52
CAPÍTULO 4		55
4.	Conclusiones y Recomendaciones	55
4.1.	Conclusiones.....	55
4.2.	Recomendaciones.....	56
5.	Bibliografía.....	57
6.	Anexos.....	59

ABREVIATURAS

AI	Artificial Intelligence
CV	Cross Validation
DL	Deep Learning
EM	Expectation Maximization
LISA	Local Indicator of Spatial Association
ML	Machine Learning
SC	Sector Censal
SCE	Suma Cuadrática del Error
MREP	Modelo de Regresión Espacial Ponderado
GAD	Gobierno Autónomo Descentralizado de Durán

SIMBOLOGÍA

Z_i	Desviación de una variable en una unidad espacial con respecto a su media para el Índice Global de Morán.
W_{ij}	Ponderación geográfica entre las unidades espaciales i y j para el Índice Global de Morán.
λ	Media y varianza de la distribución Poisson
$\theta^{(k)}$	Estimación de la iteración previa del algoritmo EM (Esperanza Maximización).
$Q(\theta \theta^k)$	Esperanza condicional del algoritmo EM.
$\theta^{(k+1)}$	Función del logaritmo que maximiza la esperanza del algoritmo EM.
w_j	Proporción del componente j en el modelo de mezcla.
$pr(y)$	Función densidad de la probabilidad de la variable de respuesta y de longitud N del modelo de mezcla.
$p(y \theta_j)$	Distribución cualquiera con variable y y n parámetros en el modelo de mezcla.

ÍNDICE DE FIGURAS

Ilustración 1: Casos de COVID-19 en sectores urbanos de Durán	38
Ilustración 2: Frecuencias de casos de COVID-19.....	39
Ilustración 3: Pronósticos de muertes esperadas del modelo SARIMA (1,1,1) (1,0,0) [12].....	40
Ilustración 4: Comportamiento temporal de las defunciones registradas del 2016 hasta el 2019 VS el pronóstico del año 2020.....	40
Ilustración 5: Muertes mensuales registradas Vs muertes estimadas para el 2020.....	41
Ilustración 6: Excesos de muertes calculados para el 2020.....	41
Ilustración 7: Análisis LISA del índice de vulnerabilidad	45
Ilustración 8: Variables de morbilidad y población por SC	46
Ilustración 9: Índice de vulnerabilidad de cada SC.....	47
Ilustración 10: Mapa de exceso de muertes estimadas.....	52
Ilustración 11: Acercamiento de mapa de exceso de muertes estimadas.....	53
Ilustración 12: Mapa de exceso de muertes registradas	53
Ilustración 13: Acercamiento de mapa de exceso de muertes registradas	54

ÍNDICE DE TABLAS

Tabla 1: Frecuencia de casos a nivel cantonal de las enfermedades agravantes de COVID-19	32
Tabla 2: Variables independientes utilizadas	34
Tabla 3: Criterios de información para el modelo SARIMA (1,1,1) (1,0,0) [12].....	39
Tabla 4: Índice Global de Morán para las variables del modelo.....	42
Tabla 5: SC con autocorrelación espacial según LISA VS SC sin autocorrelación espacial según LISA	44
Tabla 6: Prueba de Kolmogórov-Smirnov	48
Tabla 7: Métricas del MREP.....	48
Tabla 8: Métricas del modelo de mezcla Poisson	49
Tabla 9: Media de la SCE para los modelos	50
Tabla 10: Estimaciones de excesos de muertes VS exceso de muertes registrados por cada 10,000 habitantes.....	51
Tabla 11: Codificación de las enfermedades agravantes del COVID-19 según el CIE-10.....	59

CAPÍTULO 1

1. Introducción

Uno de los factores más relevantes de la salud pública a analizar en un país es la tasa de mortalidad, que es utilizada con la finalidad de llegar a un razonamiento sobre las distintas causas de defunción de una población.

Y así, indirectamente llegar a determinar las causas que influyen en esta tasa a los distintos niveles demográficos, como por ejemplo las decisiones políticas y competencias de los gobiernos locales, económicas, sociales o sanitarias, con el propósito de que las autoridades de dichas localidades cuenten con información objetiva para considerar en la planificación y gestión de mecanismos de prevención y alerta temprana que promuevan el bienestar de la población.

La pandemia del COVID-19 causada por la rápida propagación mundial del virus SARS-COV-2 declarada en el 11 de marzo del año 2020 por el doctor Tedros Adhanom Ghebreyesus, director de la OMS (Organización Mundial de la Salud), trajo consigo un impacto directo a la economía de Ecuador y del mundo en general, permitiendo visualizar las consecuencias de las vulnerabilidades sociales, demográficas e inclusive de salud, que no han sido consideradas para el monitoreo y seguimiento de las gestiones públicas en diversas zonas del país.

Según el Instituto Nacional de Estadística y Censos (INEC) hasta el 2020 se registró un incremento del 55% en la tasa esperada de mortalidad en el país con respecto al año 2019, siendo la mayoría de este incremento (57.92%) proveniente de casos confirmados o sospechosos de COVID-19. Considerando las eventualidades suscitadas por la pandemia, es de notar como ciertos sectores presentaron mayores obstáculos al momento de ser atendidos, debido a diversos factores como su índice de vulnerabilidad y morbilidad, así como una posible influencia espacial de los sectores aledaños, estos en consecuencia generan una mayor predisposición al incremento de la tasa de mortalidad.

Sin embargo, se pueden cuestionar las cifras proporcionadas por las entidades gubernamentales, como consecuencia de la falta de preparación ante situaciones de tipo emergente a nivel nacional. Considerando esto, en el presente trabajo de investigación se plantea la modelación de la tasa de excesos de muerte a partir de

las variables de vulnerabilidad, morbilidad y los factores de corrección correspondientes implementando técnicas de Machine Learning, con propósito de estimar la tasa de exceso de mortalidad durante el año 2020 en los sectores censales del cantón Durán, provincia del Guayas, que le permita a las autoridades municipales gestionar sus recursos dentro de las competencias pertinentes. Y con ello, lograr mitigar el impacto de los escenarios posteriores a los eventos de riesgo de la población.

1.1. Descripción del problema

El cálculo de la tasa de exceso de mortalidad regularmente se define como la cantidad de muertes registradas en una región en un determinado período de tiempo menos la cantidad esperada. Esta tasa resulta de gran interés cuando ocurren eventualidades de tipo social, económico o sanitario que desencadenen en una crisis.

Durante la pandemia por el COVID-19, diferentes regiones del mundo tuvieron dificultades al momento de estimarla, como producto de la falta de un sistema apropiado de registro de muertes, un mal etiquetado en los motivos de fallecimientos debido a la limitada cobertura de atención a las personas que presentaban los síntomas propios de la enfermedad, ya que sus muertes eran etiquetadas como neumonías atípicas e inclusive como una incidencia en la mortalidad de otras enfermedades que resultaron agravadas por la paralización de varios servicios médicos.

La vulnerabilidad es un concepto que hace referencia a aquellos grupos o sectores que por determinadas condiciones sociales o demográficas tienden a correr mayores riesgos en diversas situaciones que puedan surgir, impidiéndoles ser parte del desarrollo social que podría mejorar sus condiciones de bienestar. Diversas instituciones gubernamentales buscan investigar estas variables para que al momento de enfrentar crisis se puedan identificar aquellos sectores que requerirán atención inmediata y focalizada, con el objetivo de amortiguar las consecuencias de dichos eventos.

La Real Academia Española define a la morbilidad como la porción de una población que en un momento determinado padece alguna enfermedad. En Ecuador el levantamiento de datos sobre la morbilidad es llevado a cabo por el INEC, siendo

su última actualización publicada en el 2020, donde se presenta un registro de esta variable por sexo, rango etario, región geográfica y causas.

La influencia o autocorrelación espacial mide el grado en el que una variable geográfica está correlacionada consigo misma en dos puntos o zonas diferentes del área de estudio. En otras palabras, capta cómo varía esta variable a través del espacio geográfico y capta el grado en que los elementos locales pueden verse afectados por sus vecinos, con la finalidad de determinar patrones espaciales y describir su comportamiento (Siabato & Guzmán-Manrique, 2018).

Las tasas de excesos de muertes brindan un acercamiento a la cantidad de fallecimientos adicionales que surgen como producto de crisis sociales o sanitarias (Ritchie, et al., 2020).

Múltiples estudios buscan modelar la tasa de exceso de muertes utilizando únicamente técnicas de series de tiempo y segmentaciones por grupos etarios y sexo, sin embargo, las variables de vulnerabilidad y morbilidad que serán aplicadas en este estudio deben ser consideradas debido al impacto que tienen en el desarrollo de un sector y como este se recupera de las crisis sanitarias.

1.2. Justificación del problema

Ecuador no ha tenido una política pública sólida de vigilancia y control epidemiológico y, por lo tanto, cuando empezó la emergencia sanitaria emitida por la Organización Mundial de la Salud (OMS) a causa de COVID-19 existió una falta de capacidad de respuesta temprana y suficiente ante la situación (Velasco & Hurtado, 2020). Esto dejó una brecha abierta a pensamientos sobre cómo sería el próximo accionar político público ante nuevas situaciones de emergencia similares.

Las cifras de contagio anunciadas por el Ministerio de Salud Pública (MSP) junto con el Instituto Nacional de Investigación en Salud Pública (INSPI) son utilizadas para medir el impacto y alcance del COVID-19 en el país a través de las tasas de mortalidad y de exceso de muertes, al ser comparadas con años anteriores. El limitado poder de reacción efectiva inmediata ante la pandemia puso al descubierto problemas de diversa índole del sistema sanitario, entre ellos, la dificultad de contar con sistemas que permitan el registro de información de manera vinculada e integrada con todos los entes involucrados, por lo que la disponibilidad de los mismos fue otra

de las dificultades, pues al no contar con información oportuna y de primera mano los proyectos de investigación orientados a brindar soporte a la toma de decisiones de las autoridades y soluciones se vieron retrasados y en muchos casos inhabilitados. Sin embargo, a pesar de estas dificultades es preciso el desarrollo de análisis enfocados a la estimación del exceso de muertes como se propone en este trabajo, para que a través de los resultados de estas estimaciones se logre proponer medidas que vayan a mitigar los diversos efectos que generan los hechos mencionados anteriormente

Las tasas de mortalidad son de gran utilidad para el análisis de una población en el campo social y económico (Vera, et al., 2018). Las tasas de excesos de mortalidad arrojan indicios de cómo incidieron factores como la estructura poblacional, elementos ambientales y las dimensiones socioeconómicas durante la pandemia.

La información sobre mortalidad se convirtió en una herramienta fundamental para caracterizar el estado de salud de la población y diseñar políticas basadas en las necesidades de salud. Las medidas de salud utilizadas más frecuentemente reflejan la enfermedad y la mortalidad, incluso esta última más que la salud. Para la mayor parte de los países la información obtenida de los análisis de mortalidad les permite tomar decisiones y desarrollar estrategias para el bienestar del país (Organización Panamericana de la Salud, 2017).

Los modelos estadísticos son ampliamente utilizados debido a los beneficios que ofrecen, y gracias al Machine Learning, pueden ser entrenados y evaluados para asegurar que el modelo seleccionado explica de la mejor manera posible el comportamiento de las variables en conjunto (Amitava, et al., 2021). En este estudio se utilizarán variables de interés para las entidades gubernamentales, puesto que son variables que describen por ejemplo la vulnerabilidad, correlación espacial y morbilidad de la población, las mismas que son significativas para la determinación de la tasa de exceso de muertes de una zona específica.

Al obtener un modelo adecuado para las variables a utilizar se puede asegurar que las estimaciones son muy cercanas a la realidad socioeconómica de una zona. Y, así mismo, el modelo será capaz de ajustar con la mayor precisión posible la tasa de exceso de muertes. Esto quiere decir, bajo un contexto de pandemia o emergencia sanitaria similar a futuro, se puede tener una aproximación de la tasa de exceso de

muerres de una ciudad en Ecuador. Y finalmente, gracias a esa información se pueden generar mejores directrices por parte de las entidades gubernamentales y tomar medidas bajo un criterio técnico de prioridad para evitar que esta tasa tenga comportamientos no deseados durante un nuevo evento de emergencia que atraviese el país.

1.3. Objetivos

1.3.1. Objetivo general

Diseñar un modelo estadístico utilizando variables de vulnerabilidad, morbilidad e influencia espacial de los sectores censales del cantón Durán, provincia del Guayas, mediante la implementación de algoritmos de Machine Learning, para estimar las tasas de excesos de muerte por la pandemia del COVID-19.

1.3.2. Objetivos específicos

- Establecer la existencia de correlación espacial global y local en los sectores censales de Durán mediante el Índice Global de Morán y análisis LISA.
- Construir varios modelos estadísticos para la estimación de la tasa de exceso de muertes, implementando técnicas de Machine Learning.
- Evaluar los modelos generados aplicando la técnica estadística de validación cruzada, para la posterior selección del modelo que tiene las mejores estimaciones de tasas de excesos de muertes.

1.4. Marco teórico

1.4.1. Ciudad intermedia

La organización de Ciudades y Gobiernos Locales Unidos define a las ciudades intermedias desde dos factores: el volumen poblacional que debe ser entre 50 mil y 1 millón de habitantes y su influencia a nivel industrial y comercial tanto dentro como por fuera de sus límites geográficos.

1.4.2. Sector censal

Según el Instituto Vasco de Estadística un sector censal es una sectorización del término municipal caracterizada por estar preferentemente definida mediante límites fácilmente identificables, tales como accidentes naturales del terreno, construcciones de carácter permanente y viales. Además, por tener un tamaño entre 1.000 y 2.500 residentes, excepto en el caso que el municipio completo tenga una población total

inferior. Por último, se caracteriza por estar constituido por un promedio de 80 viviendas (INEC, 2022).

1.4.3. Exceso de muertes

El exceso de mortalidad es un indicador sociodemográfico que representa la diferencia entre las defunciones registradas para una comunidad en un período determinado y las esperadas. Es importante la determinación de esta tasa porque permite a las autoridades tomar acciones que resguarden la vida de los habitantes de dichos espacios (Yanover, Parolini, Rodríguez, Blanco, & Toledo, 2021).

1.4.4. Índices de vulnerabilidad

La vulnerabilidad social es un término que hace referencia a los factores económicos, sociales y demográficos que interfieren en la capacidad de respuesta ante alguna eventualidad que se catalogue como desastre (sanitario, social o natural) de distintos sectores o grupos de una comunidad, impidiendo que estos puedan reponerse y aumentando la gravedad de las condiciones de habitabilidad lo cual predispone el aumento de las defunciones. (Barry E. Flanagan, Gregory, Hallisey, Heitgerd, & Lewis, 2011). El desarrollo de índices que permitan identificar la vulnerabilidad de una comunidad es importante porque facilitan la planificación de las gestiones públicas antes las distintas calamidades que pueden azotar un colectivo.

1.4.5. Tasas de morbilidad

El Diccionario Médico Ilustrado de Dorland define a la morbilidad como la incidencia o prevalencia de una enfermedad o varias en una población determinada, en adición la Real Academia de la Lengua Española señala que dicha población se puede definir en términos de espacios geográficos y períodos temporales. De este indicador se pueden obtener las tasas de morbilidad, las cuáles se presentan como la cantidad de casos para cada cierta cantidad de habitantes establecida según el interés de quiénes lleven a cabo el análisis.

1.4.6. Autocorrelación espacial

Cuando las variables que son objeto de análisis cuentan con coordenadas que permiten identificar su geolocalización en uno de los varios sistemas geodésicos de coordenadas, como: WG S84, ED50, ETRS89, NAD83, PSAD56, SIRGAS, entre otros, se pueden aplicar técnicas de análisis que estudien su comportamiento considerando la geolocalización como un factor influyente. Es así como el término

autocorrelación espacial es utilizado para definir la variación sistemática a la que se someten estas variables al delimitar el espacio físico en que se desarrollan (Haining, 2002). Esta medida es positiva cuando las distintas observaciones guardan similitud en sus valores, mientras que en el escenario de que estos difieran notablemente, la autocorrelación espacial será negativa, es a través de esta interpretación que podemos citar la Primera Ley de la Geografía de Tobler: “Todos los lugares están relacionados, pero los lugares cercanos están más relacionados que los lugares lejanos.” (Ramírez & Falcón, 2015)

La identificación de la presencia de la autocorrelación espacial es importante porque implica la existencia de factores importantes que se distribuyen en la superficie de estudio que deberán ser considerados para entender el principal fenómeno de estudio.

Existen diversas técnicas estadísticas que nos permiten obtener mediciones de la autocorrelación espacial, siendo el Índice Global de Morán uno de los más utilizados.

1.4.7. Índice Global de Morán

El índice global de Morán fue presentado por primera vez en el año 1950 por Patrick Alfred Pierce Moran. Este es uno de los métodos más comunes y utilizados para determinar de forma global en un espacio delimitado el cumplimiento o no cumplimiento del principio de Tobler descrito anteriormente. Este índice se formula de la siguiente manera:

$$S_o = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{i,j} z_i z_j}{\sum_{i=1}^n z_i^2};$$

$$S_o = \sum_{i=1}^n \sum_{j=1}^n \omega_{i,j}$$

En donde:

n es el número total de unidades espaciales que están siendo estudiadas, S_o es una constante para las unidades antes mencionadas, Z_i es la desviación de una

variable en una unidad espacial con respecto a su media, W_{ij} representa una ponderación geográfica entre las unidades espaciales i y j .

Los valores de este indicador pueden variar entre -1 y 1 , un valor cercano a 0 implica que la distribución espacial de los valores de la variable es aleatoria, la interpretación de este índice puede ser evaluado con una prueba de hipótesis, en la que se plantea:

H₀: La distribución espacial de los valores de la variable analizada es aleatoria

Vs.

H₁: La distribución espacial de los valores de la variable analizada no es aleatoria

1.4.8. Índice Local de Asociación Espacial

Muchas veces se busca explorar a mayor detalle la influencia del factor espacial en las unidades de investigación, es por ello se han desarrollado técnicas que permiten obtener los índices de autocorrelación espacial a nivel local, siendo el más divulgado el propuesto por Luc Anselin en 1995, quien, a partir del Índice Global de Morán, presenta el Índice Local de Asociación Espacial, LISA (Local Index of Spatial Association) por sus siglas en inglés.

Este índice es implementado cuando se necesita identificar las unidades espaciales cuyos valores sean significativos en los diferentes indicadores estadísticos, permitiendo identificar puntos espaciales atípicos.

1.4.9. Modelo Multiplicativo ARIMA (p, d, q) * SARIMA (P, D, Q)

Los modelos multiplicativos de ARIMA (p, d, q) * SARIMA (P, D, Q) fueron sugeridos por Box y Jenkins con el fin de capturar la influencia de ambos modelos en una serie (Chávez Quisbert, 1997). Estos modelos consideran la existencia de una dependencia estacional para todos los períodos, hay ocasiones en que esta consideración no se cumple, pero esto no evita el empleo de esta metodología en la descripción de fenómenos temporales varios (González Casimiro, 2009).

1.4.10. Distribución Poisson

La distribución de Poisson fue derivada por SIMEON DENIS POISSON, quien en 1837 publicó un trabajo de Investigación en el que se presentaba una nueva distribución de probabilidad aplicado al ámbito penal (King, 1988). Poisson encontró

que cuando el tamaño de una muestra es grande y la probabilidad de ocurrencia de un evento es pequeña, el valor esperado $\mu = np$ tiende a ser una constante.

La función densidad de la probabilidad de variable de Poisson es:

$$f(y) = \frac{(\lambda)^y e^{-\lambda}}{y!} \quad y = 0, 1, \dots$$

donde s es el tamaño del intervalo, la media y la varianza de esta variable son ambas iguales a $\mu = \sigma^2 = \lambda$ (probabilidad de una ocurrencia en la unidad de medida).

Un modelo de regresión para una variable de Poisson es un modelo que permite estudiar si dicha variable depende, o no, de otra u otras variables. Si una variable de Poisson de parámetro λ es independiente de otra variable X , se cumple $\lambda = \frac{\lambda}{X}$, por consiguiente, un modelo de regresión es una función de λ en X que permite consolidar la relación de dependencia a través del coeficiente de X , y así como en los modelos lineal y logístico, es fácilmente generalizable a más variables independientes.

1.4.11. Variables Poisson

Una variable es de tipo Poisson cuando se representa el número de eventos que ocurren en un intervalo temporal o espacial de tamaño dado (s), cumpliendo las siguientes condiciones (Abraira & Pérez de Vargas, 1996): Una variable es de tipo Poisson cuando se representa a el número de eventos que ocurren en un intervalo temporal o espacial de tamaño dado (s), cumpliendo las siguientes condiciones (Abraira & Pérez de Vargas, 1996):

- El número de eventos que ocurren en el intervalo es independiente del número de los que ocurren fuera del mismo.
- Existe un intervalo lo suficientemente pequeño, de tamaño h , para el que la probabilidad de que en el mismo ocurra un sólo evento es proporcional al tamaño del intervalo, es decir es λh , siendo por tanto λ (constante) la probabilidad de que ocurra un evento en un intervalo de tamaño unidad.
- La probabilidad de que en cualquier intervalo de tamaño h ocurran dos o más eventos, es prácticamente 0.

1.4.12. Algoritmo Esperanza Maximización (EM)

El algoritmo EM es un método para la obtención de soluciones para las ecuaciones de verosimilitud en el caso que no sea posible para ellas obtener una forma analítica (Hernández Malagón, 2020). El nombre del algoritmo surge de los dos pasos principales a aplicar del mismo, donde el primer paso (E) consiste en el cálculo del valor esperado o mayormente conocido como esperanza matemática. Seguido, se aplica el paso M que consiste en aplicar el método de máxima verosimilitud para modelos probabilísticos. El algoritmo EM es un método iterativo cuya finalidad es calcular el estimador de máxima verosimilitud (ML) de un parámetro θ de una distribución de probabilidad paramétrica.

Una de las ventajas principales de aplicar algoritmo EM es que es posible aplicarlo con base de datos con registros perdidos debido a que el algoritmo trata de asociar un problema de datos incompletos con otro problema con registros completos gracias a las relaciones entre la verosimilitud de estos dos problemas mencionados (Casas & Burgo Alcalá Nalvaiz, 2017). Sin embargo, una de las desventajas es que debido a las iteraciones la carga computacional puede ser alta y como resultado obtendríamos una convergencia muy lenta.

Según Casas y Burgos (2017) este algoritmo consiste en repetir dos pasos:

1. Paso E (Esperanza):

Este paso consiste en “rellenar” los datos que faltan, en otras palabras, consiste en la creación de nuevo problema con datos completos para realizar la comparación de la verosimilitud.

Dada la estimación de la iteración previa $\theta^{(k)}$, calcule la esperanza condicional:

$$Q(\theta|\theta^k) = E\{\ln f(X|\theta)|Y, \theta^k\}$$

2. Paso M (Maximización):

Consiste en encontrar los parámetros que maximizan la función del logaritmo de la verosimilitud calculada en el paso anterior.

La función de probabilidad se maximiza bajo el supuesto de que se conocen los datos faltantes. Escoja un $\theta^{(k+1)}$ para ser algún valor de $\theta \in \Theta$ que maximiza $Q(\theta|\theta^k)$.

$$\theta^{(k+1)} = \arg \max_{\theta \in \Theta} Q(\theta, \theta^{(k)})$$

3. Repetir los dos pasos con iteraciones hasta llegar a una convergencia.

1.4.13. Modelos de mezcla

Un modelo de mezcla consiste en una distribución de densidad de probabilidad que es una combinación o mezcla convexa de varias distribuciones dando paso a un proceso de modelamiento estadístico de gran flexibilidad capaz de encontrar aplicaciones en multitud de fenómenos en disciplinas heterogéneas (Salas, 2006).

Son modelos utilizados para el ajuste de datos de tipo heterogéneo que han pasado por situaciones experimentales, donde usar una sola distribución no es factible al momento de modelar la variable de respuesta. Básicamente se interpreta que los datos provienen de dos o más subpoblaciones.

En términos matemáticos, si la variable aleatoria discreta Y es una mezcla de k componentes o dimensiones, entonces la densidad de probabilidad de Y , $pr(y)$, es la suma de las distintas distribuciones con distintos pesos w_i de los componentes. El modelo de mezcla se representa del siguiente modo (Salas, 2006):

$$pr(y) = \sum_{j=1}^k w_j p(y|\theta_j)$$

$$0 \leq w_j \leq 1 \ (\forall j); \sum_{j=1}^k w_j = 1$$

Donde w_j es la proporción del componente j en la mezcla y $pr(y)$ es la función densidad de la probabilidad de la variable de respuesta y de longitud N . $p(y|\theta_j)$ representa una distribución cualquiera con variable y y n parámetros que denotamos genéricamente como $\theta = [\theta_1, \dots, \theta_n]$.

1.4.14. Modelos de mezcla: distribución Poisson

Un modelo de mezcla con distribución Poisson consiste básicamente en un modelo de mezcla cuyas componentes de la variable de respuesta Y del modelo se

consideran pertenecientes a suma de subpoblaciones Poisson con distintos parámetros λ . Un modelo Poisson asume que el valor esperado y la varianza de la variable de respuesta son iguales (λ) y los valores de la variable de respuesta no pueden ser negativas ni estar expresadas con decimales.

1.4.15. Modelo de regresión espacial ponderada Poisson

El modelo de regresión espacial ponderada (MREP) es un método local de regresión lineal que se utiliza para modelar las relaciones (correlación) que varían espacialmente, en otras palabras, permite realizar un análisis espacial de las relaciones entre las variables independientes que los modelos globales no permiten estudiar (Molinero, et.al., 2021). El MREP posibilita tres tipos de modelos de regresión: continuo, binario y recuento. En términos de estadísticos, estos tipos de regresión se conocen como gaussiano, logística y Poisson, respectivamente. El Tipo de modelo para análisis se debería elegir dependiendo de cómo se mide o comporta la variable dependiente (Y), así como del rango de valores de las observaciones que contiene.

Una ventaja significativa del modelo MREP es que aplica una ponderación geográfica a las entidades utilizadas en cada una de las ecuaciones de regresión local. A las entidades que están más alejadas del punto de regresión se les da menos peso y, por tanto, tienen menos influencia en los resultados de regresión para la entidad de destino; las entidades más cercanas tienen más peso en la ecuación de regresión. Los pesos se determinan utilizando un kernel, que es una función de disminución de la distancia que determina la rapidez con la que los pesos disminuyen a medida que aumentan las distancias (López, et.al., 2019).

Un modelo de regresión espacial ponderada de tipo Poisson es aquel que tiene una variable Y discreta y representa el número de ocurrencias de un evento, por ejemplo, recuento de accidentes. Los modelos de recuento también se pueden utilizar si Y representa una tasa y el denominador de esta es un valor fijo como, por ejemplo, número de gente con neumonía por cada 10.000 habitantes. El modelo Poisson (recuento) asume que el valor esperado y la varianza de la variable de respuesta son iguales (λ) y las observaciones no pueden ser negativas ni estar expresadas con decimales (Pineda & Rivas, 2016). La fórmula del modelo de regresión espacial ponderada de tipo Poisson es como sigue:

$$y_i \sim \text{Poisson}[N_i \exp(\beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{k,i})]$$

Donde,

$\beta_0(u_i, v_i)$ es el parámetro específico del intercepto para la locación i .

$\beta_k(u_i, v_i)$ es el coeficiente de la variable independiente x_k en la locación i y N_i es la variable de compensación en la i -ésima ubicación.

(u_i, v_i) es la coordenada $x \sim y$ en la i -ésima ubicación.

1.4.16. Machine Learning

Machine Learning (ML) es una rama en evolución de los algoritmos computacionales que están diseñados para emular la inteligencia humana aprendiendo del entorno circundante. Son considerados el impulso en la nueva era conocida como Big data (El Naga & Murphy, 2015).

Machine Learning es el proceso de descubrir automáticamente patrones y tendencias en los datos que van más allá del simple análisis. Esta técnica hace uso de algoritmos matemáticos sofisticados para segmentar los datos y predecir la probabilidad de eventos futuros basados en eventos pasados, que no se pueden abordar mediante técnicas simples de consulta e informes (Ghatak, 2017).

Las técnicas basadas en ML se han aplicado con éxito en diversos campos que van desde el reconocimiento de patrones, la visión por computadora, la ingeniería de naves espaciales, las finanzas, el entretenimiento y la biología computacional hasta las aplicaciones biomédicas y médicas. El aprendizaje automático o Machine Learning se utiliza en búsquedas web, filtros de spam, sistemas de recomendación, calificación crediticia, detección de fraudes, negociación de acciones, diseño de medicamentos y muchas otras aplicaciones (Nichols, et. al. , 2019).

Ghatak en el 2017 señala que el aprendizaje automático está relacionado con la Artificial Intelligence (AI) y el Deep Learning (DL), cuyas definiciones para entender sus diferencias son:

Inteligencia artificial (IA) es el término más amplio que se aplica a cualquier técnica que permita a las computadoras imitar la inteligencia humana utilizando lógica, reglas

del tipo "si-entonces", árboles de decisión y aprendizaje automático (incluido el aprendizaje profundo). Por otro lado, el aprendizaje automático es un subconjunto de la inteligencia artificial que incluye técnicas estadísticas abstrusas que permiten a las máquinas mejorar en las tareas con la experiencia obtenida al ejecutarlas. Cuya finalidad es obtener la mejor aproximación posible de una función que defina nuestra variable de respuesta objetivo. Y finalmente, Deep Learning es una versión escalable de aprendizaje automático, intenta expandir el posible rango de funciones estimadas. Por ejemplo, si el ML puede aprender 1000 modelos, el DL nos permite aprender 10000 modelos. Aunque ambos tienen espacios infinitos, el Deep Learning tiene un espacio viable más grande debido a las matemáticas, al exponer las redes neuronales de múltiples capas a grandes cantidades de datos.

1.4.17. Validación cruzada

Para conjuntos de datos grandes, la muestra original de datos puede dividirse en un conjunto de entrenamiento en el que entrenar el modelo, un conjunto de validación en el que validar nuestros modelos y un conjunto de pruebas para valorar nuestro modelo entrenado. Sin embargo, cuando no tenemos grandes muestras de datos, la validación cruzada es particularmente útil. La validación cruzada es un método estadístico de evaluación y comparación de algoritmos de aprendizaje mediante la división de datos en dos conjuntos: uno utilizado para entrenar un modelo y el otro utilizado para validar el modelo (Lantz., 2013).

La validación cruzada (CV) nos permite seleccionar un modelo y estimar el error en el modelo. Cuando seleccionamos un modelo usando CV, lo hacemos seleccionando un modelo de una gama de otros modelos, entrenados en un conjunto de datos en particular y / o seleccionando los hiperparámetros de un modelo en particular.

1.4.18. K-Folds Cross Validation

La retención repetida es la base de una técnica conocida como validación cruzada de k veces (K-Folds Cross Validation), que se ha convertido en la técnica estándar de la industria para estimar el rendimiento del modelo. Pero en lugar de tomar muestras aleatorias repetidas que potencialmente podrían usar el mismo registro más de una vez, el K-Folds Cross Validation divide aleatoriamente los datos en k particiones aleatorias completamente separadas llamadas pliegues (folds). (Lantz., 2013)

Un caso extremo de K-Folds CV es el método de dejar uno fuera (leave-one-out), que realiza K-Folds CV usando un pliegue para cada uno de los ejemplos de datos. Esto asegura que se utilice la mayor cantidad de datos para entrenar el modelo. Aunque esto puede parecer útil, es muy costoso computacionalmente por lo que rara vez se usa en la práctica.

1.4.19. Criterio de Información de Akaike

El Criterio de Información de Akaike es una medida que nos brinda un criterio de selección del mejor modelo dentro de un conjunto de modelos. A diferencia de otros criterios el AIC previene el sobreajuste mediante una función que considera el número de parámetros estimados como un mecanismo de sanción. Con esta técnica se selecciona el modelo que representa una menor pérdida de información con respecto al proceso generador de los datos (Burnhm & Anderson, 2004).

1.4.20. Criterio de Información de Bayesiano

El Criterio de Información Bayesiano (BIC) desarrollado por Gideon E. Schwarz en 1978 es un criterio utilizado en la selección del mejor modelo de un conjunto de modelos generados. Se prefieren siempre aquellos modelos que presenten el BIC más bajo. La aplicación de esta medida permite evitar problemas de sobreajuste (Jayanta K., Mohan, & Tapas, 2006).

CAPÍTULO 2

2. Metodología

En este capítulo se describe el proceso de recolección, tratamiento y evaluación de las variables que se utilizaron en el proceso de modelación de la tasa de excesos de muertes en los sectores censales del cantón Durán.

Para la estimación de este indicador demográfico se aplicaron modelos de mezcla a través del algoritmo EM, que empleó modelos probabilísticos Poisson, que permitieron modelar los datos y se seleccionó al de menor Suma Cuadrática del Error.

2.1. Área de estudio

Los datos utilizados en el modelo corresponden a sectores censales (SC) del cantón Durán de la provincia del Guayas. El cantón Durán tiene tres parroquias urbanas: Eloy Alfaro, Divino Niño y El Recreo. Según datos oficiales del INEC, la proyección poblacional para el 2020 es aproximadamente 315.724 habitantes resultando en una densidad de 3912.53 habitantes/km², convirtiéndola en la sexta ciudad más poblada del país. Se encuentra en la región Litoral Ecuatoriano a la derecha del río Guayas y de la isla Santay. Así mismo es considerada como la segunda urbe más grande y poblada de la provincia del Guayas.

El cantón Durán limita con la ciudad de Guayaquil, una de las ciudades más afectadas por el COVID-19 en Latinoamérica durante el año 2020. Varios de los sectores censales de Durán cuentan con una menor densidad poblacional debido a que sus territorios tienen fines industriales. Es uno de los más importantes centros administrativos, económicos, financieros y comerciales del Ecuador. Las actividades principales de la ciudad son el comercio, y la industria. Por otro lado, Durán forma parte del área metropolitana de Guayaquil, pues su actividad económica, social y comercial está fuertemente ligada a esta urbe, por lo que es considerada una ciudad intermedia.

2.2. Fuente de los datos

El procesamiento de datos consistió en varias etapas para obtener las variables observables y no observables de interés para el presente proyecto que se detallan a continuación.

2.2.1. Casos de contagio de COVID-19

El registro de casos de contagio por COVID-19 en los sectores censales de Durán durante el año 2020 proviene de los registros del Ministerio de Salud Pública del Ecuador (MSP).

2.2.2. Índice de vulnerabilidad

El índice de vulnerabilidad fue levantado de variables que describen la situación socioeconómica de los distintos sectores censales que se analizarán, las variables que se consideran capturan las siguientes dimensiones: capacidad adaptativa, sensibilidad y exposición. Los datos de las variables provienen de fuentes primarias como el censo poblacional realizado en el 2010, proyecciones de población, hogares y viviendas al 2020, medidas del Gobierno Autónomo Descentralizado de Durán (GAD) y de los registros del Ministerio de Salud Pública, como consta en el Reporte Técnico del Análisis de Riesgo a COVID-19 con enfoque multi-amenaza (Borbor, et. al. , 2022).

2.2.3. Estimación de las tasas de excesos de muertes por sector censal en el cantón Durán en el 2020

Para modelar la tasa de exceso de muertes del 2020 se trabajó con la cantidad de defunciones reales desde el período 2016 al 2019, registradas en la base de datos del Registro Civil, y con el total de defunciones esperadas que corresponden a las defunciones reales del año 2020. La diferencia de estos dos valores representa el exceso de muertes para el periodo de estudio. Se aplicó el modelo de serie temporal SARIMA con un periodo mensual de defunciones. Las métricas de evaluación del modelo utilizadas fueron Criterio de Información de Akaike (AIC), el Criterio de Información Bayesiano (BIC) y Sigma al cuadrado.

Los datos reales sobre la cantidad de defunciones fueron tomados de los registros de acceso público del INEC con un total de 781 casos registrados para el cantón Durán. Debido a que los datos se encontraban a nivel censal se procedió a distribuir el exceso de muertes estimado según el peso de los casos de COVID-19 confirmados

en el sector censal por pruebas de Reacción en Cadena de la Polimerasa (PCR) con respecto al total de contagios registrados en el cantón.

2.2.4. Estimación de las tasas morbilidad de las enfermedades agravantes del COVID-19

Para estimar las tasas de morbilidad por sector censal se utilizó el registro de egresos hospitalarios del cantón más reciente publicado por el INEC, que a la fecha de redacción del presente proyecto es del 2019.

El proceso de recolección de los datos se presenta en 2 etapas:

- Primera etapa: Identificación de códigos de las enfermedades agravantes del COVID-19 en el CIE-10, clasificándolas en 5 categorías: diabetes y elevaciones de glucosa, enfermedades cardíacas, hipertensivas, pulmonares, cancerígenas y tumores benignos. Para mayor detalle sobre las codificaciones y diagnósticos considerados dirigirse a la *Tabla 11: Codificación de las enfermedades agravantes del COVID-19 según el CIE-10* en la sección Anexos.

Tipo de enfermedad	Frecuencia de casos en Durán
Diabetes y elevaciones de glucosa	34
Cardíacas	18
Pulmonares	64
Hipertensivas	12
Cancerígenas y tumores benignos	53

Tabla 1: Frecuencia de casos a nivel cantonal de las enfermedades agravantes de COVID-19

- Segunda etapa: Consideración del tamaño poblacional de los sectores censales como el peso para el cuál serán ponderadas las frecuencias de casos a nivel cantonal de las enfermedades definidas anteriormente.

Tasa de morbilidad paara el SC

$$= \frac{\text{Total poblacional en el SC}}{\text{Total poblacional en Durán}} * \text{Frecuencia de casos en Durán}$$

Una vez completadas estas etapas se procedió a realizar transformaciones en los valores de las tasas de morbilidad, multiplicándolas por una constante igual a 100, de tal forma que los valores se lean personas con alguna de las morbilidades por cada centenar de habitantes.

2.3. Correlación espacial

La existencia de correlación espacial ayuda a entender el comportamiento de las variables utilizadas en los modelos a lo largo de los sectores censales de Durán. Debido a que al existir correlación espacial se logra entender el proceso de clasificación o clusterización de las variables de interés. Primero se busca encontrar la existencia de correlación global mediante el Índice Global de Morán, aplicándolo a todas las variables que conformaron los modelos testeados en este estudio. La hipótesis nula de la prueba del Índice Global de Morán indica que la variable analizada está distribuida de manera aleatoria y, por otro lado, la hipótesis alterna que los datos analizados siguen una distribución no aleatoria. El valor estadístico obtenido oscila entre -1 y 1, siendo que mientras más cercano esté el valor del estadístico a 1, mayor es el grado de correlación global. El 1 significa una autocorrelación positiva perfecta, es decir, máxima concentración; mientras que el 1 negativo indica autocorrelación negativa perfecta o completa dispersión. El cero significa un patrón espacial completamente aleatorio.

Seguidamente se calculó el Índice Local de Asociación Espacial (LISA) aplicado a todas las variables, con el fin de realizar un análisis más detallado de la influencia del factor espacial en las variables de interés. Dado que es un índice local, éste brinda información relevante sobre la agrupación espacial de valores similares alrededor de un punto específico. La ventaja principal en este estudio es poder determinar la significancia estadística de los conglomerados generados entre las variables. La hipótesis nula de la prueba del análisis LISA indica que no existe asociación espacial local y, por otro lado, la hipótesis alterna sí valida la existencia de dicha correlación. Así mismo las métricas de decisión se basan en el estadístico calculado donde se busca encontrar valores similares al Índice Global de Morán que oscila entre -1 y 1,

siendo que mientras más cercano esté el valor del estadístico a 1, mayor es el grado de correlación global. El 1 significa una autocorrelación positiva perfecta, es decir, máxima concentración; mientras que el 1 negativo indica autocorrelación negativa perfecta o completa dispersión. El cero significa un patrón espacial completamente aleatorio. Es de resaltar que la prueba de hipótesis es ejecutada sobre cada uno de los 531 sectores censales por cada variable independiente de interés. El fin de este proceso es identificar los sectores que se ven afectados por factores espaciales.

2.4. Definición de variables

Las variables independientes utilizadas para modelar el exceso de muertes son:

Población total	Tamaño poblacional por SC.
Índice de Vulnerabilidad	Indicador de vulnerabilidad del SC.
Tasa de Morbilidad: Diabetes	Tasa de morbilidad de diabetes por SC.
Tasa de Morbilidad: Cardiacas	Tasa de morbilidad por enfermedades cardiacas del SC.
Tasa de Morbilidad: Hipertensos	Tasa de morbilidad de hipertensos por SC.
Tasa de Morbilidad: Pulmonar	Tasa de morbilidad por enfermedades pulmonares del SC.
Tasa de Morbilidad: Cáncer	Tasa de morbilidad por enfermedades de cáncer del SC.

Tabla 2: Variables independientes utilizadas

2.5. Aplicación de modelos

Previo a las pruebas de modelación se consideraron en la base de datos únicamente los casos de los sectores censales que presentaron un exceso de muertes igual a al menos 1. Se usaron dos técnicas de modelación con tipo de distribución Poisson, debido a la naturaleza de la variable de respuesta: la primera fue un modelo de regresión espacial ponderada y la segunda un conjunto de modelos de mezcla, que varían en la cantidad de clústers que se consideran, en este caso fue desde $k = 2$ hasta $k = 6$.

2.5.1. Modelo de regresión espacial ponderada con distribución Poisson

La técnica de modelación de regresión espacial ponderada es una versión local de los Modelos Lineales Generalizados, en este caso se asumió que la variable de respuesta sigue una distribución Poisson por su naturaleza. Dentro de los parámetros que considera el modelo se consideró un kernel Bi-cuadrático. Al considerar la ubicación geográfica de los casos, se tuvieron que utilizar las coordenadas de cada sector, las cuales estaban en formato de coordenadas longitud y latitud. Una vez adecuados los parámetros se procedió con el proceso de modelación.

2.5.2. Modelo de mezcla: distribución Poisson

Se aplicó un modelo de mezcla con distintas distribuciones Poisson en base a los hallazgos encontrados en el análisis LISA donde se dio un punto de partida sobre cuántas clases o clústeres se deben usar en el modelo de mezcla. Según el índice de LISA se debió partir con un valor de $k = 4$, es decir, 4 clústeres y esto reafirmó la validez de usar un modelo de mezclas debido a que la variable de respuesta sigue varias distribuciones Poisson con distintos valores de λ . No obstante, se probaron varias combinaciones posibles de modelos de mezcla partiendo de un modelo con $k = 2$ clases hasta un modelo con $k = 6$ clases o clústeres.

La única modificación por resaltar con respecto a la formulación de las variables en el modelo consiste en que se aplicó un «offset» de la variable de Tamaño de la población debido a que es una de las variaciones más habituales que consiste en la inclusión de un término adicional, denominado «variable de exposición», «multiplicador de tasa», o simplemente «offset», en el caso de que sea necesario tomar en consideración que el recuento para cada una de las observaciones se basa en períodos de tiempo o espacio, o bien en tamaños poblacionales distintos. Lo cual si es válido para el estudio debido a que el tamaño poblacional por sector censal varía entre 64 y 707.

Finalmente, luego de fijar todos los parámetros del modelo y de probar distintas dimensiones de clusterización se obtuvo que el mejor valor de k para el modelo de mezcla Poisson es con un $k = 5$. Es decir, un clúster adicional al sugerido por el análisis LISA aplicado al índice de vulnerabilidad.

2.6. Selección del mejor modelo

Una vez establecidos los modelos, se implementó la técnica de K-Folds Cross Validation, con $k = 10$ para el total de 176 casos que se consideraron. La aplicación de esta técnica garantiza la inexistencia de subajuste o sobreajuste, debido a que se forma el modelo con el 90% de los datos, se lo prueba en el 10%, y se repite este paso 10 veces, hasta que cada subconjunto haya sido utilizado como datos de prueba. De cada una de las iteraciones se obtuvo una Suma Cuadrática de los Errores (SCE), y al promediarlas se tiene la Media de la Suma Cuadrática de los Errores, considerada como el indicador para evaluar el mejor modelo, en conjunto con el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC).

2.7. Limitantes

En el proceso de recolección y modelación de los datos se debieron considerar algunas limitantes. Primero, no existe un registro oficial de las muertes previstas por año para los cantones del país por lo que se tuvieron que implementar modelaciones temporales con los datos históricos de defunciones. Segundo, no existe un registro de defunciones por sector censal ni de egresos hospitalarios, esto implicó la aplicación de técnicas de ponderación. Finalmente, al momento de ponderar el exceso de muertes para los sectores censales se consideraron los casos de COVID-19 como pesos, sin embargo, es de considerar que el conteo presentado por esta variable considera únicamente a los casos que fueron diagnosticados como positivos por las pruebas PCR, misma que no era accesible para la mayoría de la población del cantón, provocando la existencia de casos positivos confirmados no registradas por las instituciones públicas, este mismo hecho causa la presencia de cero contagios en 355 de los 531 sectores censales, y consecuentemente al tener cero contagios se reportaba un exceso de muertes del mismo valor. Los sectores censales que presentaban esta condición no fueron considerados como datos aplicables a los modelos implementados debido a que no representaban exceso de muertes.

CAPÍTULO 3

3. Resultados

Se obtuvieron hallazgos de interés para el actual estudio, los cuales se detallan en esta sección. Primero, se muestran las gráficas descriptivas de Durán con la finalidad de entrar en contexto de la situación por la que pasaba el cantón en el 2020. Debido a que las observaciones son por sector censal se procedió a realizar mapas de frecuencias y de calor con la finalidad de entender a nivel espacial los casos de COVID-19 registrados en los sectores urbanos. De la misma manera, se presentan los mapas de las variables de morbilidad las cuales son de enfermedades que agravan la enfermedad causada por COVID-19 y así se logró visualizar el comportamiento frecuentista de las mismas a lo largo de Durán. Segundo, se muestran los resultados obtenidos del análisis de autocorrelación espacial realizados con la finalidad de comprobar con técnicas Geoestadísticas la influencia espacial en los factores de interés. Finalmente, se muestran las métricas obtenidas de los modelos aplicados, así como los resultados obtenidos del método de validación cruzado junto con la conclusión sobre qué modelo es el más apropiado para modelar las tasas de excesos de muerte.

3.1. Sectores urbanos donde se registran casos de COVID-19

Se graficó a nivel de sectores urbanos los casos de COVID-19 registrados por el MSP en el mapa de Durán.

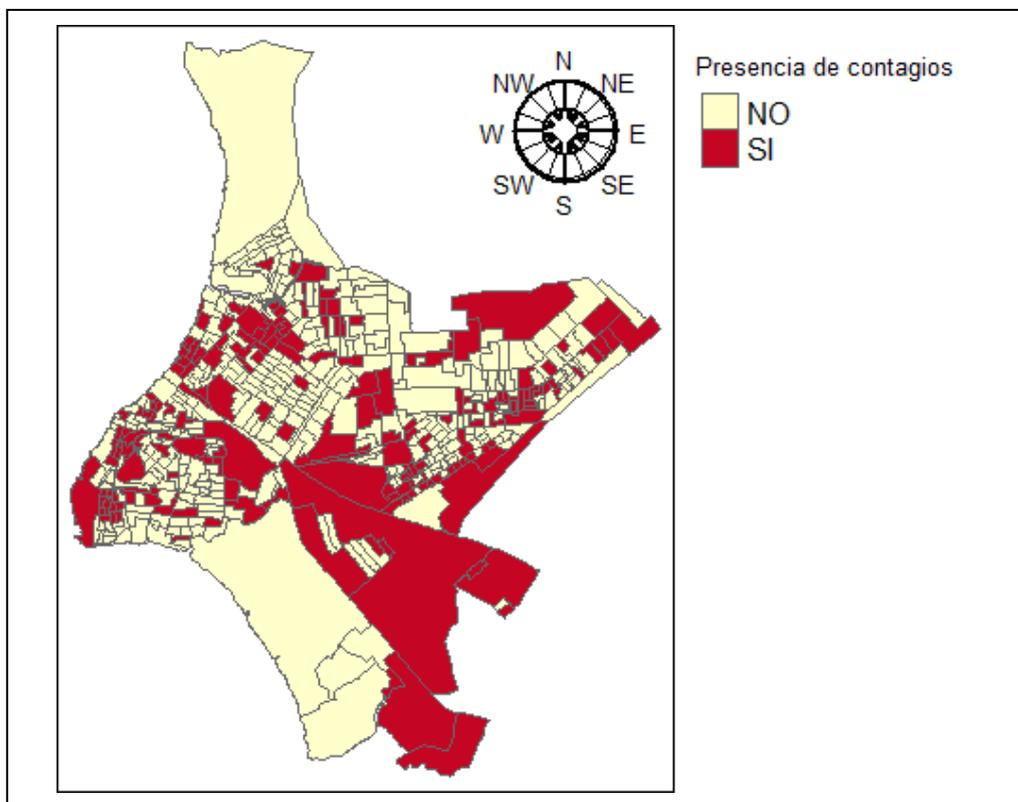


Ilustración 1: Casos de COVID-19 en sectores urbanos de Durán

Existieron 176 sectores censales donde se registraron casos y 355 zonas donde no se registraron casos de COVID-19 para el año 2020.

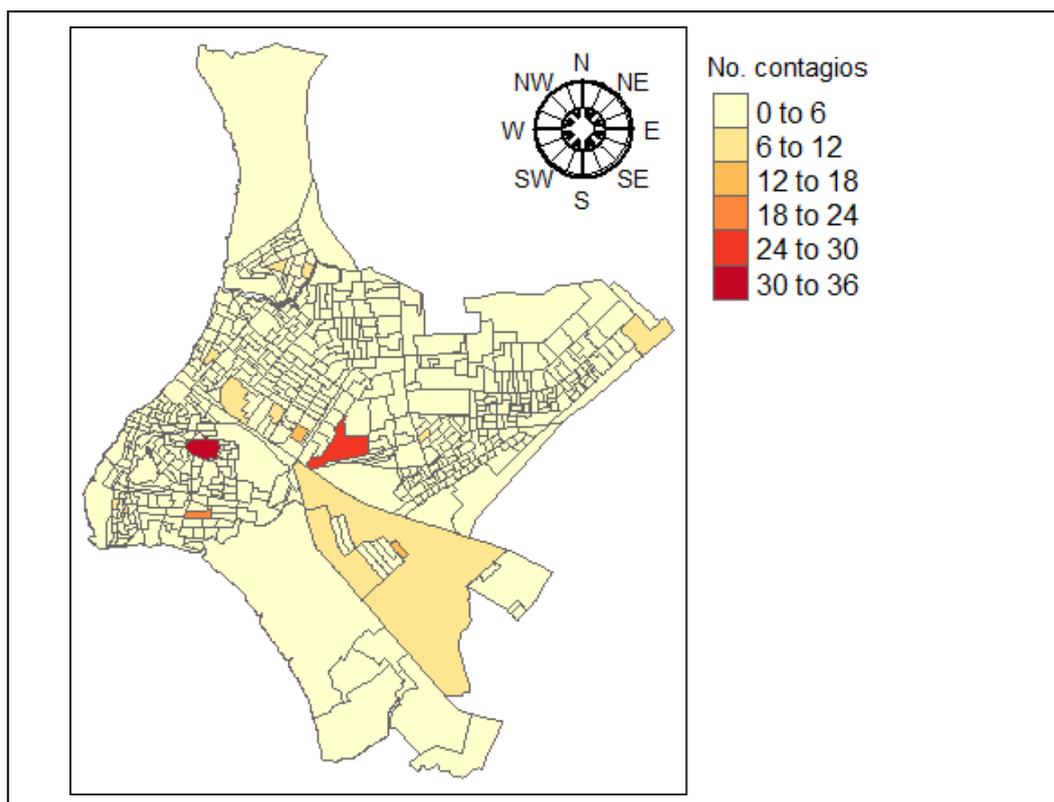


Ilustración 2: Frecuencias de casos de COVID-19

3.2. Serie de tiempo SARIMA para la estimación de exceso de muertes para el 2020

Se aplicó el modelo de serie temporal SARIMA (1,1,1) (1,0,0) [12] con orden $p = 1$ y $q = 1$, con una diferenciación para el componente estacionario y para el componente estacional se tiene orden $P = 1$ y $Q = 0$, sin diferenciación, y con un periodo $s = 12$ correspondiendo a una frecuencia mensual de defunciones. Las métricas de evaluación del modelo obtenidas son las que se presentan a continuación.

Sigma al cuadrado (σ^2)	AIC	BIC
30.497	305.191	312.592

Tabla 3: Criterios de información para el modelo SARIMA (1,1,1) (1,0,0) [12]

Cumpliendo los respectivos supuestos teóricos se obtuvo el pronóstico de defunciones esperadas para el cantón Durán del año 2020, con un total de 124 muertes. El gráfico respectivo de los pronósticos mensuales obtenidos para las defunciones es como sigue:

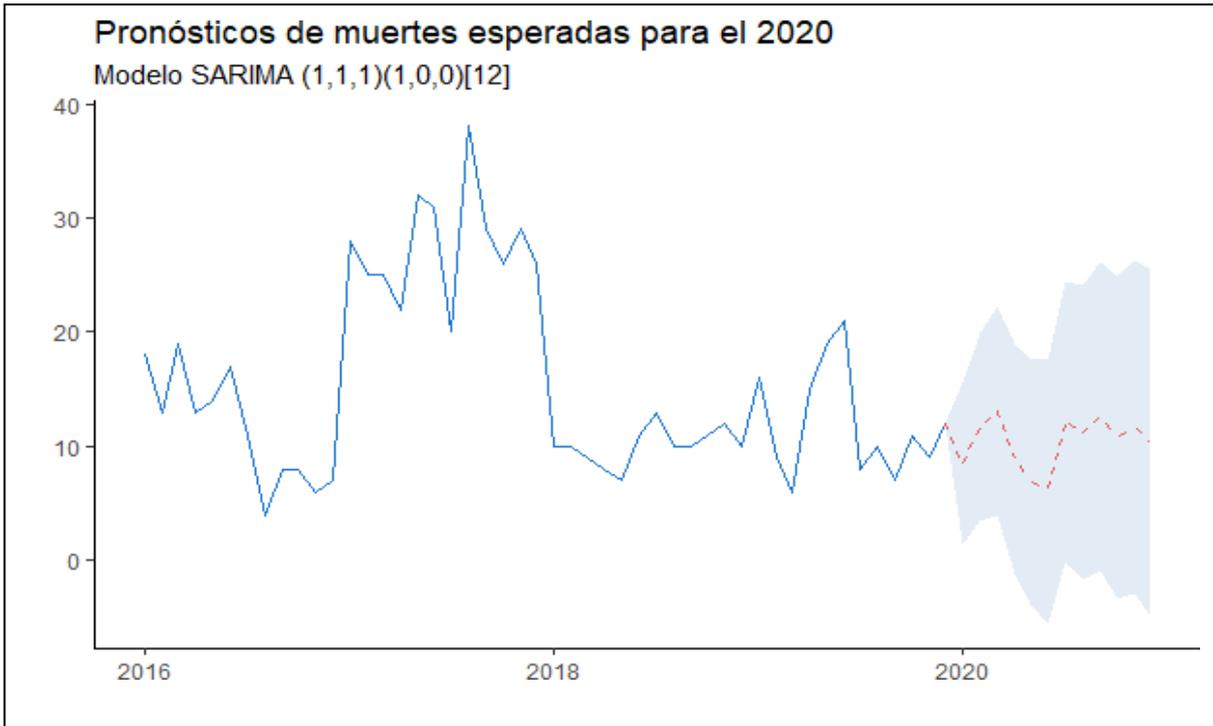


Ilustración 3: Pronósticos de muertes esperadas del modelo SARIMA (1,1,1) (1,0,0) [12]

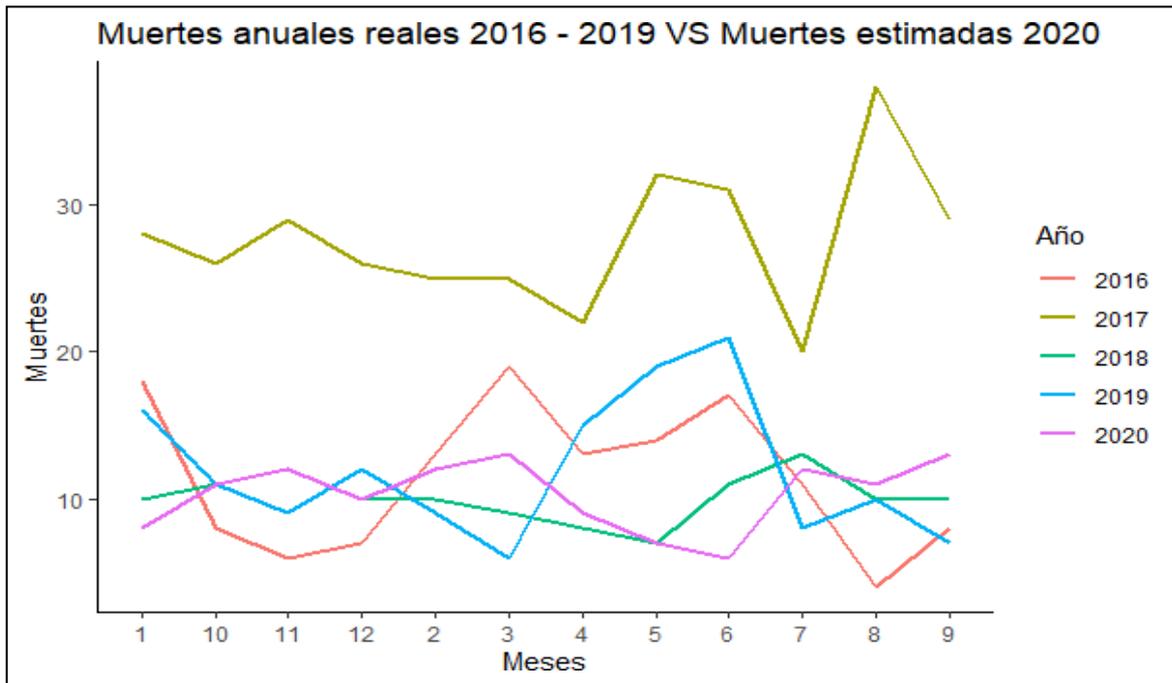


Ilustración 4: Comportamiento temporal de las defunciones registradas del 2016 hasta el 2019 VS el pronóstico del año 2020

Se visualiza que los pronósticos para el año 2020 (color morado) de la serie de tiempo desde el año 2016 al 2019 siguen el comportamiento de defunciones registradas años anteriores, mostrando un buen ajuste de la serie SARIMA. Adicionalmente, se presenta la gráfica comparativa de las series temporales de las

muertes reales registradas y las muertes estimadas para el año 2020. Y, por último, se muestra el gráfico de área del exceso de muertes calculado para el año de análisis con finalidad de entender visualmente su comportamiento.

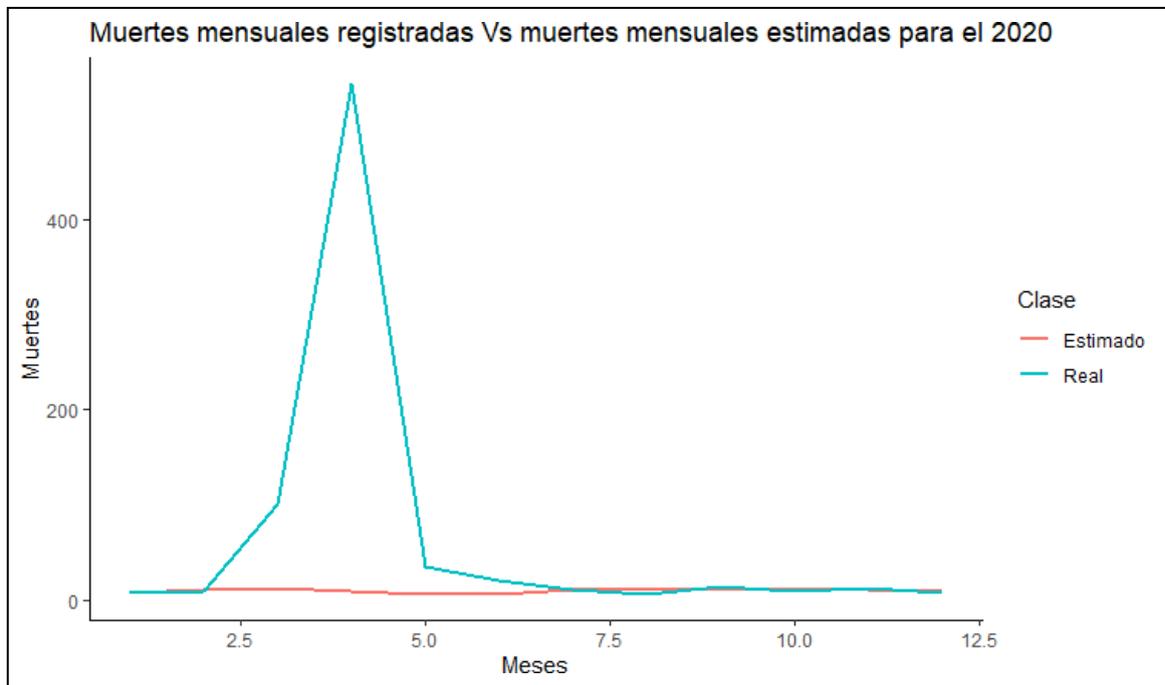


Ilustración 5: Muertes mensuales registradas Vs muertes estimadas para el 2020

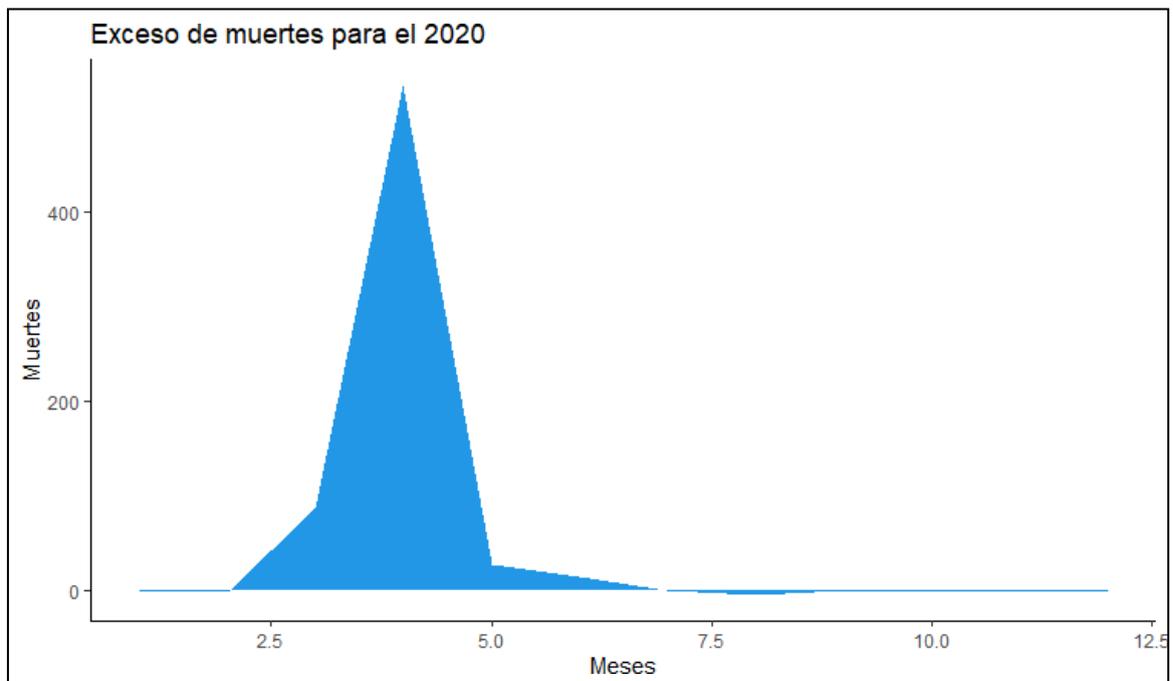


Ilustración 6: Excesos de muertes calculados para el 2020

Se observa que el exceso de muertes calculado captura el evento ocurrido en el mes de febrero hasta el mes de mayo cuando ocurrió la primera ola de contagios bajo el contexto de estudio sigue el comportamiento real registrado para el año de análisis.

3.3. Índice Global de Morán para las variables independientes del modelo

Se detallan los Índices Globales de Morán para todas las variables independientes utilizadas y para la variable dependiente en los modelos probados en este estudio. Así mismo, se observan los valores p del contraste de hipótesis planteado acerca de la existencia de autocorrelación espacial global. Se concluyó en base al valor p de la prueba y se tomó en cuenta el valor del estadístico obtenido donde los valores oscilan entre -1 y 1, siendo que mientras más cercano esté el valor del estadístico a 1, mayor es el grado de correlación global. El 1 significa una autocorrelación positiva perfecta, es decir, máxima concentración; mientras que el -1 negativo indica autocorrelación negativa perfecta o completa dispersión. El cero significa un patrón espacial completamente aleatorio.

Variable	Índice Global de Morán	Valor p
Exceso de muertes	0.00722	0.3511
Índice de Vulnerabilidad	0.27489	< 2.2e-16
Tasa de Morbilidad: Diabetes	0.33191	< 2.2e-16
Tasa de Morbilidad: Cardiacas	0.27039	< 2.2e-16
Tasa de Morbilidad: Hipertensos	0.26574	< 2.2e-16
Tasa de Morbilidad: Pulmonar	0.33042	< 2.2e-16
Tasa de Morbilidad: Cáncer y tumores benignos	0.33491	< 2.2e-16

Tabla 4: Índice Global de Morán para las variables del modelo

Dados que los valores p obtenidos en las pruebas de hipótesis realizadas son menores a 0.05, concluimos que para todas las variables independiente existe la suficiente evidencia estadística para rechazar la hipótesis nula con un nivel de confianza al 95%, en favor de la alterna, la cual indica que la distribución espacial de los valores de las variables no es aleatoria. Sin embargo, el valor p obtenido para la prueba de hipótesis para la variable dependiente es mayor a 0.05, por lo tanto, existe suficiente evidencia estadística para no rechazar la hipótesis nula, indicando que la distribución espacial del exceso de muertes no es aleatoria.

3.4. Índice de LISA para las variables independientes del modelo

Se realizó adicionalmente el cálculo del Índice Local de Asociación Espacial (LISA) y la prueba de hipótesis correspondiente a cada uno de los sectores censales por cada variable. En la siguiente tabla presentamos un resumen de la cantidad de sectores censales que basándose en la prueba de hipótesis planteada por LISA cuentan con autocorrelación local espacial y los que no.

Variable	Número de SC con autocorrelación local espacial	Número de SC sin autocorrelación local espacial
Índice de Vulnerabilidad	253	278
Tasa de Morbilidad: Diabetes	126	405
Tasa de Morbilidad: Cardiacas	99	432
Tasa de Morbilidad: Hipertensos	91	440
Tasa de Morbilidad: Pulmonar	104	427

Tasa de Morbilidad: Cáncer y tumores benignos	109	422
---	-----	-----

Tabla 5: SC con autocorrelación espacial según LISA VS SC sin autocorrelación espacial según LISA

La existencia de conglomerados permite identificar la existencia de influencia del factor espacial en algunas variables, en este caso la que más resalta es aquella que representa al índice de vulnerabilidad. Para 253 sectores censales se obtuvo un valor p menor a 0.05, por lo que se rechaza la hipótesis nula con un nivel de confianza al 95% en favor de la alterna la cual indica que estos sectores censales en efecto presentaron autocorrelación espacial local.

3.5. Mapas de correlación espacial mediante clusterización de LISA

Una de las aplicaciones de mayor ventaja fueron las visualizaciones del mapa de Durán donde se observaron los comportamientos de las variables con los respectivos clústeres y sus niveles calculados por el análisis LISA. Se realizó una gráfica del mapa de Durán con la variable no observable de índice de vulnerabilidad por sector censal y se obtuvieron mediante el análisis de LISA 3 clústeres respectivos para los índices de vulnerabilidad.

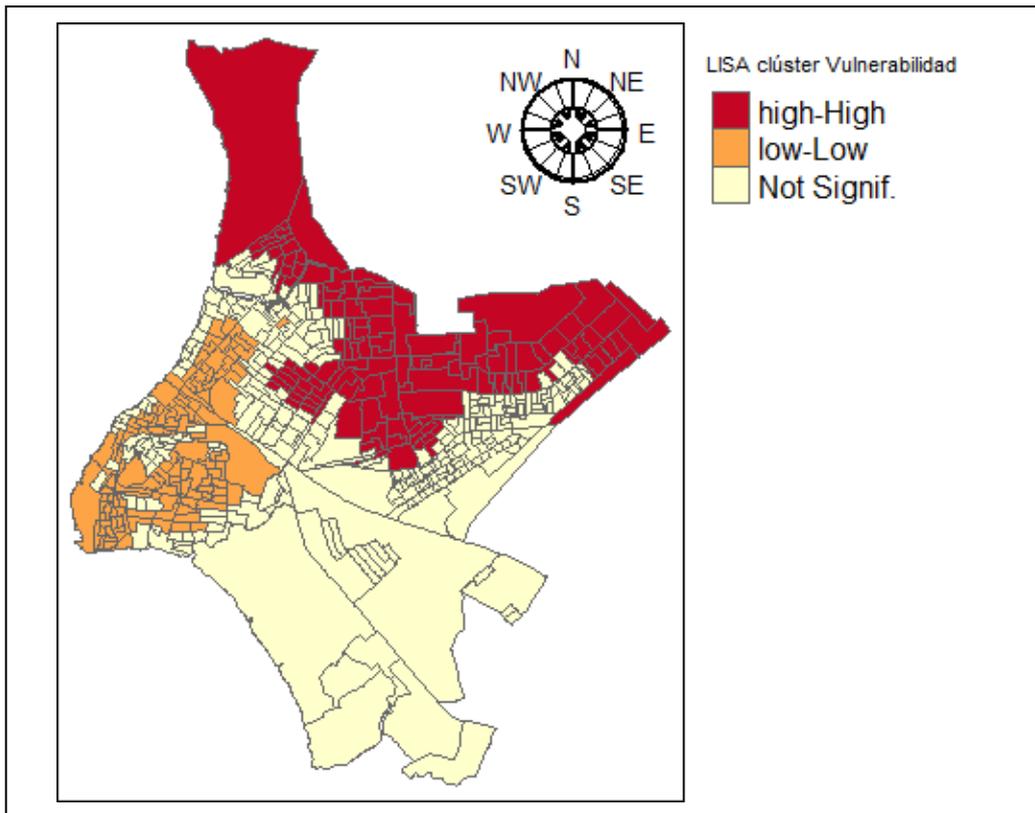


Ilustración 7: Análisis LISA del índice de vulnerabilidad

Según el análisis de correlación espacial de LISA existen 3 clasificaciones de índice de vulnerabilidad en Durán. Para el norte de Durán se observa que existen índice de vulnerabilidad “high-high”, es decir, que hay un índice alto de vulnerabilidad para esas zonas. Así mismo se observa que existen índices de vulnerabilidad de tipo “low-low” al oeste de Durán. Por último, para la zona sur y central existen índices de vulnerabilidad no significantes.

3.6. Mapas de las variables independientes

Para entender de una manera visual el comportamiento de las variables independientes se graficaron las mismas en el mapa de Durán, se muestra el número de muertes por distintas enfermedades (cáncer, cardíacas, hipertensión, pulmonar, diabetes). Las unidades se presentan son de muerte por cada cien habitantes en el SC. Adicionalmente, se muestra el número de habitantes por sector censal y por último se muestra el índice de vulnerabilidad de cada SC ubicados en los mapas de calor.

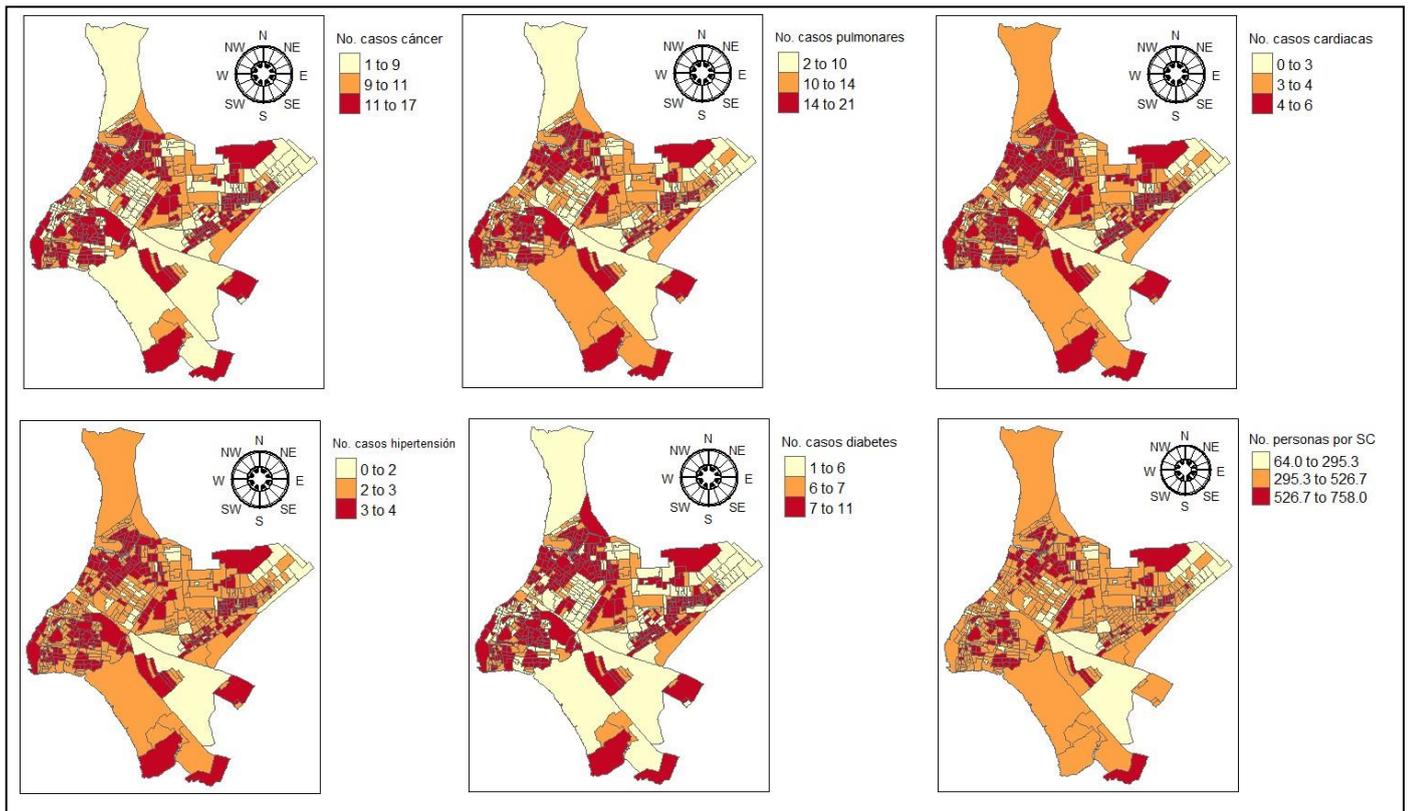


Ilustración 8: Variables de morbilidad y población por SC

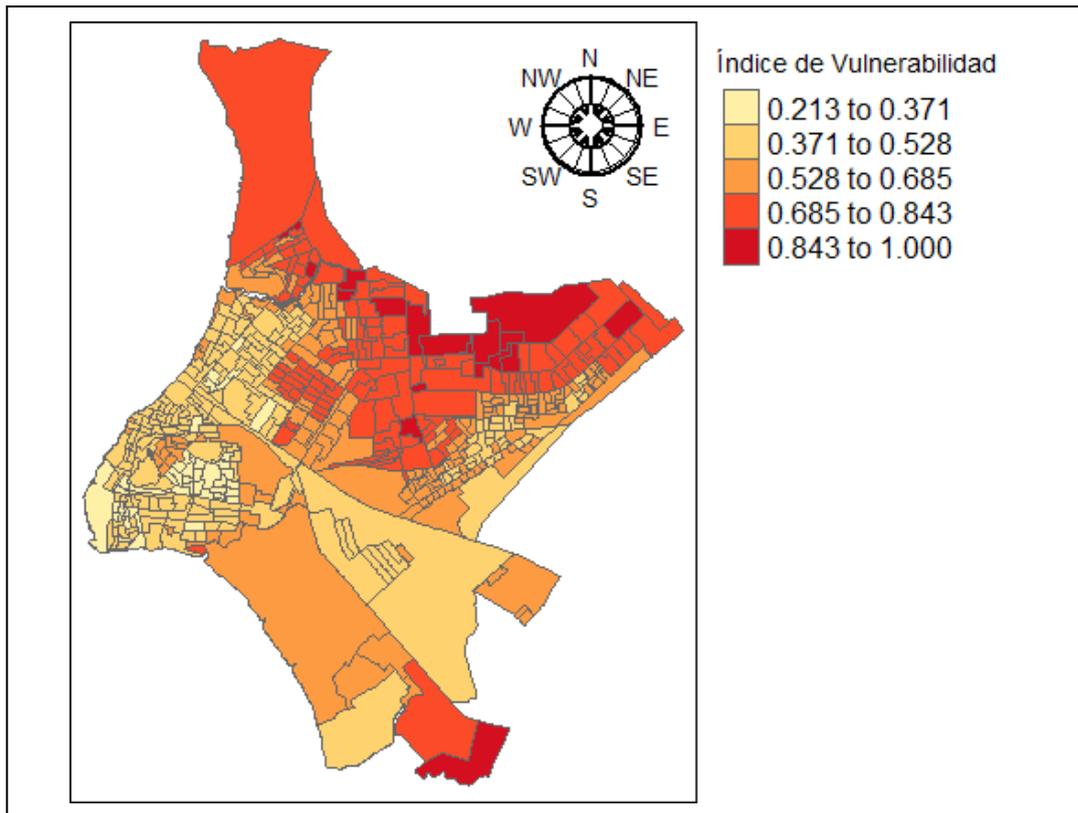


Ilustración 9: Índice de vulnerabilidad de cada SC

Se observa la incidencia de más casos para algunos SC, ya sea por dos tipos de enfermedades o por algunas enfermedades. Así mismo, se visualiza que en algunas de estas zonas de estudio existe un mayor número de habitantes, lo cual explica el motivo de que se observen más casos. Finalmente, en el caso de índice de vulnerabilidad se observan ciertas similitudes con el mapa que fue graficado con clusterización de LISA.

3.7. Prueba de bondad de ajuste para la variable de exceso de muertes

Se aplicó una prueba de bondad de ajuste a través de una prueba de Kolmogórov-Smirnov para probar con estadística inferencial que la variable de respuesta sigue una distribución Poisson. Primero, se le aplicó a la variable tomando en cuenta todas las observaciones, es decir, considerando los sectores censales que representaban un exceso de muertes igual a cero. Seguidamente, se realizó el mismo proceso para la variable de exceso de muertes sin ceros con fin de validar que provienen de una distribución Poisson.

H_0 : Los datos no pertenecen a una población Poisson(λ)

Vs.

H_a : Los datos pertenecen a una población Poisson(λ)

Prueba de bondad de ajuste Kolmogórov-Smirnov		
	Estadístico D	Valor p
Exceso de muertes con ceros	0.53461	< 2.2e-16
Exceso de muertes sin ceros	0.86939	< 2.2e-16

Tabla 6: Prueba de Kolmogórov-Smirnov

Una vez realizadas las pruebas de bondad de ajuste y dado que los valores p son menores a 0.05, en ambos casos existe la suficiente evidencia estadística para rechazar la hipótesis nula con un nivel de confianza al 95% en favor de la hipótesis alterna, que plantea que la variable de exceso de muertes sigue una distribución Poisson.

3.8. Modelo de regresión espacial ponderada con distribución Poisson

A continuación, se muestran los criterios de diagnóstico calculados para el modelo de regresión espacial ponderada Poisson considerando un kernel de Bi-cuadrático:

SCE	AIC	Pseudo R ²	Desviación
33.5297	241.5795	0.64931	165.1056

Tabla 7: Métricas del MREP

Adicionalmente, se muestra la fórmula respectiva del modelo con los coeficientes obtenidos en las iteraciones.

Exceso de muertes

$$\begin{aligned}
 &= 1.7202249 - 0.01694948 * \text{offset}(\text{population size}) + 0.1648961 \\
 &* \text{diabetes} - 0.9055762 * \text{cardiacas} - 0.3005532 * \text{hipertensos} \\
 &- 0.1466612 * \text{pulmonar} + 0.4220909 * \text{cáncer y tumores benignos} \\
 &- 0.2763454 * \text{vulnerabilidad}
 \end{aligned}$$

3.9. Modelo de mezcla Poisson

Luego de fijar todos los parámetros del modelo y de probar distintas dimensiones de clusterización se obtuvo que el mejor valor de k para el modelo de mezcla Poisson es con un k = 5. Es decir, un clúster adicional al sugerido por el análisis LISA aplicado al índice de vulnerabilidad. Para el modelo probabilístico de mezcla implementado se obtuvo que el modelo que dio un menor error cuadrático es:

$$pr(y) = \sum_{j=1}^k w_j p(y|\theta_j)$$

$$pr(y) = 0.2278 + 0.0535 + 0.1258 + 0.3319 + 0.2609$$

$$pr(y) = 0.9999$$

AIC	BIC	SCE
594.0799	704.6949	10.37378

Tabla 8: Métricas del modelo de mezcla Poisson

El modelo de mezcla presentado considera un k = 5, implicando que existen 5 conglomerados diferentes y un modelo Poisson para cada uno de ellos. La suma cuadrática del error es igual aproximadamente a 10.37, por otro lado, al ser calculado sobre un conjunto de prueba no guarda riesgos de sobreajuste.

3.10. Elección del mejor modelo

Una vez establecidos los mejores modelos por cada técnica, se procede a elegir el mejor modelo utilizando la técnica de K-Folds Cross Validation con k = 10:

Iteración	Suma Cuadrática del Error	
	Modelo de regresión espacial ponderada con distribución Poisson	Modelo de mezcla Poisson con $k = 5$
1	36.14352	13.467709
2	126.32791	16.062543
3	197.67411	15.266793
4	923.89316	18.377903
5	200.29317	9.990055
6	2394.52559	9.017666
7	158.47873	9.973332
8	225.59053	8.963395
9	74.67729	8.667107
10	33.52968	10.546429
Promedio de la Suma Cuadrática de los Errores	437.1134	12.0333

Tabla 9: Media de la SCE para los modelos

En la tabla 9 podemos ver que el Promedio de la Suma Cuadrática de los Errores para el Modelo de regresión espacial ponderada con distribución Poisson es aproximadamente 36.32 veces mayor al obtenido por el Modelo de mezcla Poisson con $k = 5$, basándonos en este indicador podemos decir que el mejor modelo es el último.

3.11. Presentación de tasas

Las proyecciones respectivas se han calculado usando la data de prueba, por lo tanto, se obtuvo un total de 14 proyecciones. Las unidades de cada estimación están por cada 10.000 habitantes, por ejemplo, se estima un exceso de muertes de 2 para el SC 090750043002.

Sector censal	Excesos de muertes estimado	Exceso de muertes registrado	Índice de vulnerabilidad
090750043002	2	2	0.4960086
090750025003	7	7	0.5569814
090750041007	1	1	0.3985278
090750039006	1	1	0.7635484
090750038001	1	15	0.4457174
090750027004	8	2	0.5557982
090750013004	9	9	0.5098463
090750027001	0	1	0.6246334
090750008007	2	2	0.5119666
090750010005	9	5	0.5234557
090750025006	11	1	0.5563333
090750018009	4	4	0.7827109
090750010010	1	7	0.6536921
090750018003	1	4	0.7326642
TOTAL	57	59	

Tabla 10: Estimaciones de excesos de muertes VS exceso de muertes registrados por cada 10,000 habitantes

El exceso de muertes estimado para SC con índice de vulnerabilidad alto no presenta mayor número en comparación al exceso de muertes reales para el año 2020. Sin embargo, en el caso de exceso de muertes reales donde se observa mayor número sí tiene mayor índice de vulnerabilidad, indicando la incidencia de la situación vulnerable de los SC en el exceso de muertes.

3.12. Mapa de calor del conjunto de prueba para sus valores de exceso de muertes versus el exceso de muertes estimado

Se localizaron geográficamente en un mapa de calor el exceso de muertes para los sectores censales del conjunto de prueba con finalidad de comparación visual.

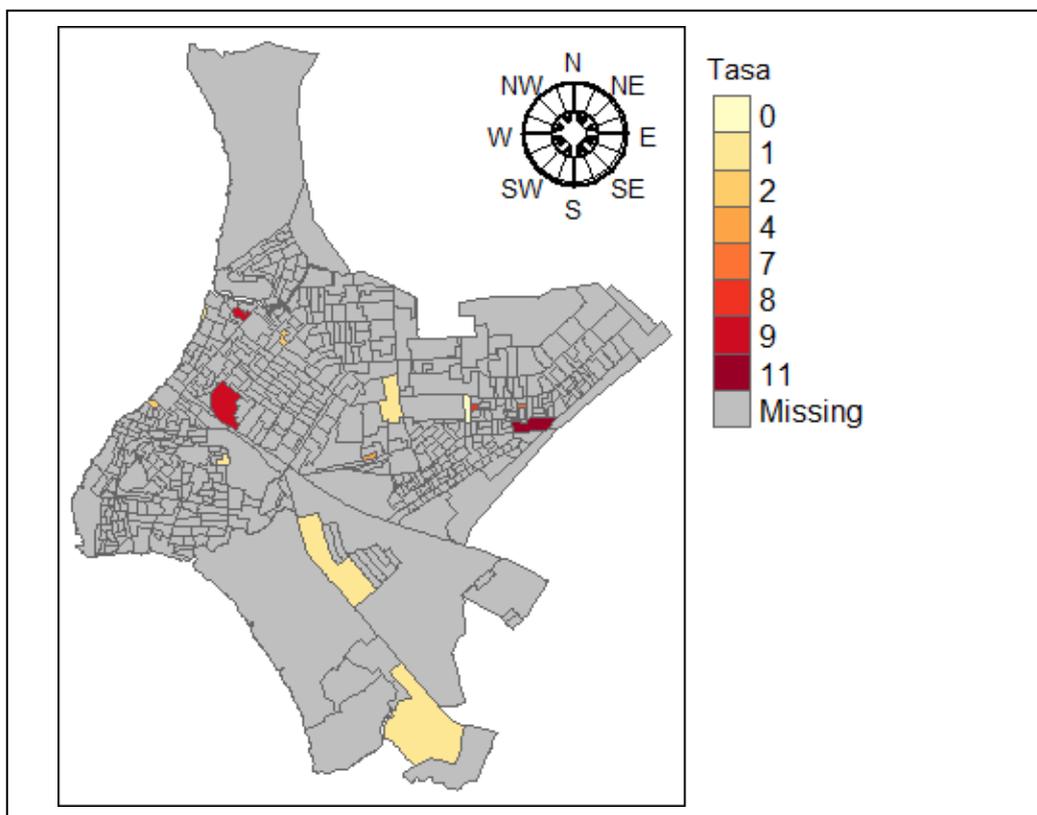


Ilustración 10: Mapa de exceso de muertes estimadas

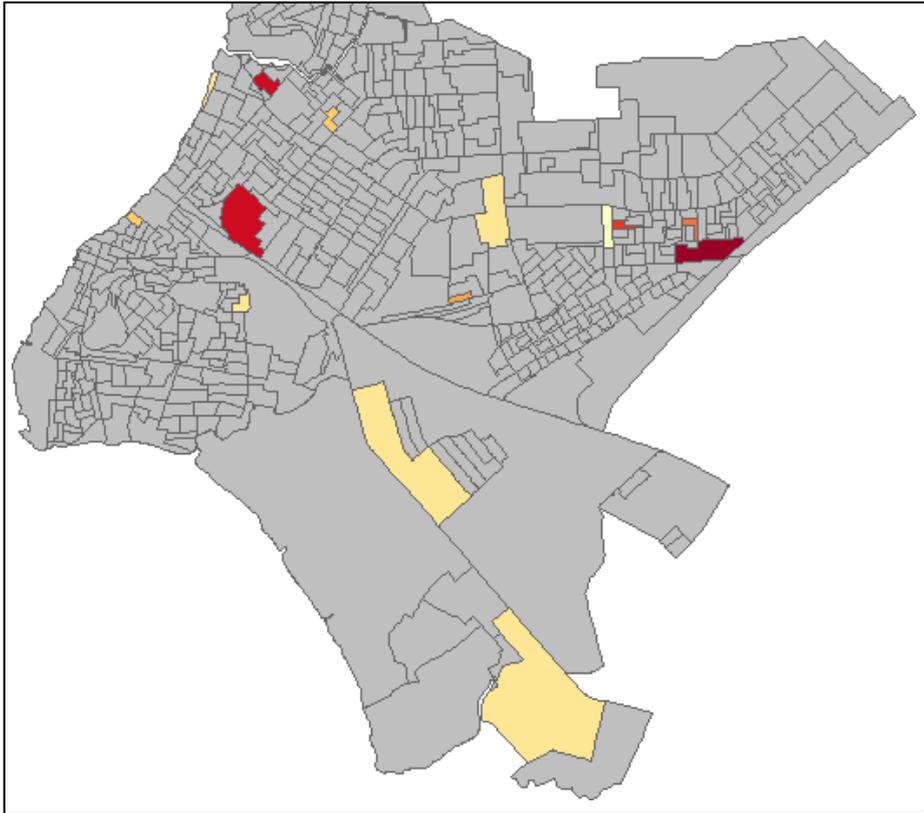


Ilustración 11: Acercamiento de mapa de exceso de muertes estimadas

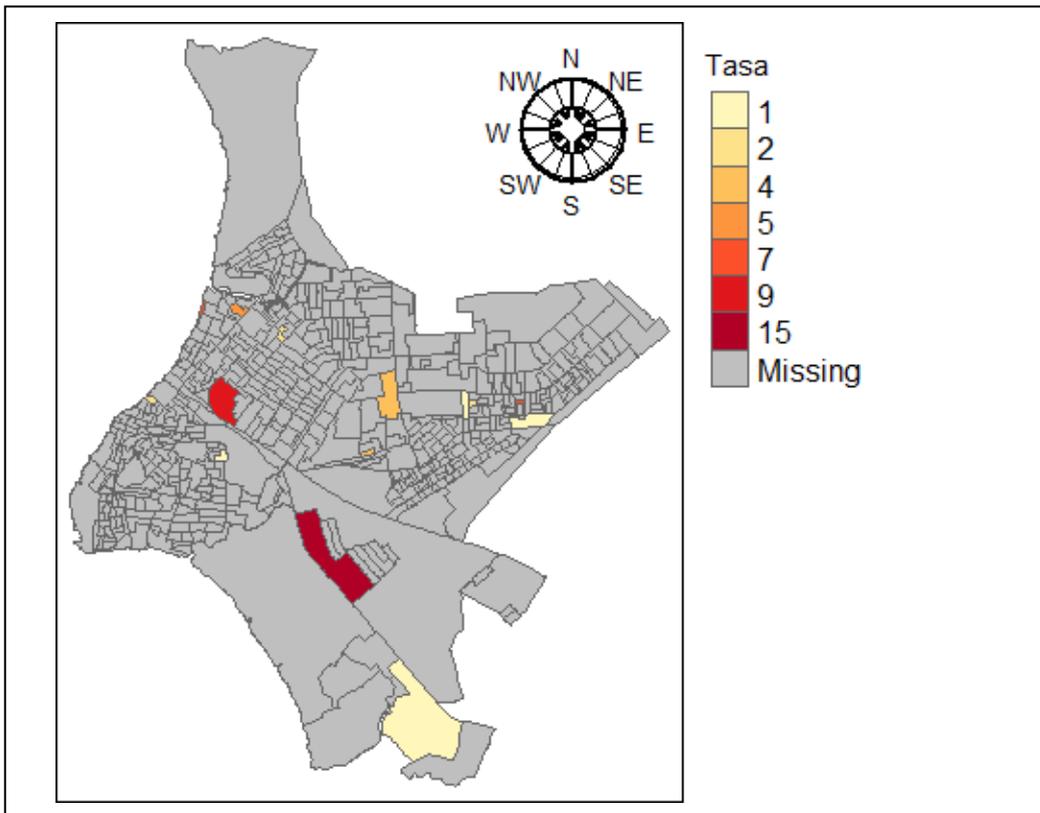


Ilustración 12: Mapa de exceso de muertes registradas

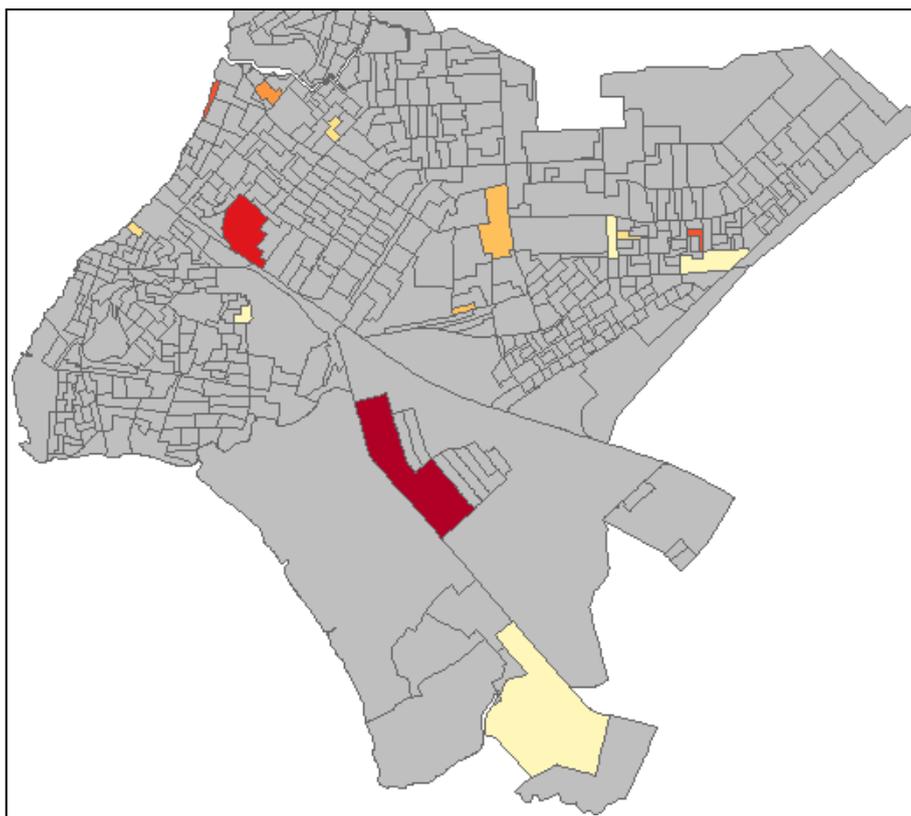


Ilustración 13: Acercamiento de mapa de exceso de muertes registradas

En los mapas de calor presentados podemos observar el comportamiento de los sectores censales que fueron utilizados en el conjunto de datos de prueba, aquellos fueron considerados en el conjunto de entrenamiento se identifican como “Missing” en la leyenda. Podemos notar un comportamiento similar en ambos que permitirá a las autoridades competentes visualizar las zonas que podrían resultar más afectadas por las eventualidades de riesgo.

CAPÍTULO 4

4. Conclusiones y Recomendaciones

4.1. Conclusiones

- 4.1.1. El modelo probabilístico de mezcla con distribución Poisson y $k = 5$ fue el modelo seleccionado para la estimación requerida debido a que logró cumplir con los objetivos de análisis planteados en el presente proyecto, esto es, estimar de manera más cercana y precisa los excesos de muertes en los sectores censales, capturando la incidencia de dimensiones como la vulnerabilidad y morbilidad en cada una de estas zonas, permitiendo medir a través de estas variables independientes la afectación que pueden llegar a tener en caso de presentarse eventos adversos y emergencias de distinta naturaleza.
- 4.1.2. El estudio de las conductas espaciales de las variables independientes incluido en este trabajo fue de un aporte importante porque permitió preseleccionar modelos que tomen a esta característica en consideración. En este caso particular, el Índice Global de Morán indicó que no existe autocorrelación global con respecto a los excesos de muertes de los sectores censales, sin embargo, las variables empleadas como independientes sí cumplen con la premisa del índice, implicando la influencia de las distribuciones espaciales de estas en el comportamiento de la variable de respuesta.
- 4.1.3. Posterior a la aplicación del Índice Global de Morán, el análisis LISA presentó la existencia de conglomerados en las distintas variables independientes. Si bien no todos los sectores censales se ven afectados por sus vecinos, el identificar aquellos que sí lo hacen brinda una perspectiva más amplia a las autoridades sobre como intervenirlos considerando su autocorrelación espacial local, por ejemplo, si se realizan gestiones que buscan mitigar la vulnerabilidad de los sectores se debería considerar importantes a aquellos que funcionan como conglomerados debido a que los beneficios se amplificarían a los sectores vecinos del intervenido y se generaría un impacto mayor en términos de reducción de daños.
- 4.1.4. El modelo seleccionado fue diseñado a nivel de sectores censales de Durán, pero siguiendo la metodología presentada se puede aplicar en otras

escalas geográficas, obteniendo resultados confiables, por ejemplo, utilizando las ciudades del país como unidades de observación.

4.1.5. La correcta estimación de los excesos de muertes en los sectores censales posibilita una mejoría en la definición de sectores prioritarios y distribución de recursos y esfuerzos ante una emergencia que amenace con el bienestar de las comunidades, dichas emergencias incluyen desastres naturales, olas de calor, inundaciones o de tipo sanitarias como en el caso estudiado en el presente proyecto de investigación.

4.2. Recomendaciones

- La falta de datos con respecto a los contagios de COVID-19 provocó una pérdida significativa de información, esto debido a que solo se consideran las pruebas de tipo PCR, se podría extender esta definición y tomar en cuenta a las pruebas de antígenos, las cuales son más accesibles para las comunidades, tanto en el sentido económico como en los puntos de toma de muestra.
- A nivel espacial se deben identificar los sectores censales que sirven como zonas industriales de distintos rubros en el cantón Durán, debido a que estos tienen una densidad poblacional baja y pueden llegar a sesgar, inflar o variar significativamente los estadísticos sensibles a valores atípicos. Se recomienda, omitir dichos sectores o implementar técnicas robustas.

5. Bibliografía

- Abraira, V., & Pérez de Vargas, A. (1996). *Métodos Multivariantes en Bioestadística*. Centro de Estudios Ramón Areces.
- Amitava Banerjee, S. C., Kotecha, D., Freitag, D. F., Denaxas, S., & Hemingway, H. (2021). Machine learning for subtype definition and risk prediction in heart failure, acute coronary syndromes and atrial fibrillation: systematic review of validity and clinical utility. *BMC medicine*, 19(1), 1-14.
- Barry E. Flanagan, B., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*.
- Borbor, M., Roa, H., & Ronquillo, Y. (2022). *Reporte técnico del Análisis de Riesgo a COVID-19 con enfoque multi-amenaza*. Centro Internacional del Pacífico para la Reducción del Riesgo de Desastres.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*.
- Casas, B. U., & Burgo Alcalá Nalvaiz, J. T. (2017). *Algoritmo EM. Aplicaciones y extensiones*. Zaragoza: Facultad de Ciencias Universidad Zaragoza.
- Chávez Quisbert, N. (1997). MODELOS ARIMA. *Revista Ciencia y Cultura*.
- Ciudades y Gobiernos Locales Unidos. (2021). *Ciudades y Gobiernos Locales Unidos*. Obtenido de Ciudades intermedias: <https://www.uclg.org/es/agenda/ciudades-intermedias>
- Dorland. (2020). *Dorland's Illustrated Medical Dictionary, 33rd Edition*. Philadelphia: ELSEVIER.
- El Naga, I., & Murphy, M. J. (2015). *What is machine learning? In machine learning in radiation oncology*. Springer, Cham.
- Ghatak, A. (2017). *Machine Learning with R*. Singapore: Springer Nature Singapore Pte Ltd.
- González Casimiro, M. P. (2009). *Análisis de series temporales: Modelos Arima*. Bilbao: Sarriko-On.
- Haining, R. (2002). Spatial Autocorrelation. *International Encyclopedia of the Social & Behavioral Sciences*.
- Hernández Malagón, O. (2020). *Sistema para la obtención de una mezcla de normales mediante el algoritmo maximización de la esperanza (EM)*.
- INEC. (2020). *Instituto Nacional de Estadística y Censos*. Obtenido de Estadísticas Vitales: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Defunciones_Generales_2020/2021-06-10_Principales_resultados_EDG_2020_final.pdf

- INEC. (2022). *Instituto Nacional de Estadísticas y Censos*. Obtenido de <https://www.ecuadorencifras.gob.ec/estadisticas/>
- Jayanta K., G., Mohan, D., & Tapas, S. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer New York.
- King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science*, 838-863.
- Lantz., B. (2013). *Machine Learning with R*. Packt Publishing.
- López, H. N., Aceros, B. M., & Luzardo, B. M. (2019). Análisis de los hurtos en Colombia durante el año 2017 mediante los modelos de regresión lineal múltiple y la regresión ponderada geográficamente. *Revista Criminalidad*, 61(3), 141-163.
- Ministerio de salud pública, M. (2020). *Ministerio de salud pública*. Obtenido de <https://www.salud.gob.ec/coronavirus-covid-19/>
- Molinero, P. R., Aguilera, B. F., & Gómez, D. M. (2021). Regresión Logística Geográficamente Ponderada para identificar los factores explicativos de la distribución de usos de suelo en escenarios futuros de crecimiento urbano. *Boletín de la Asociación de Geógrafos Españoles*(88).
- Nichols, J. A., Chan, H. W., & Baker, M. A. (2019). Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical reviews*, 11(1), 111-118.
- OMS. (2020). *Organización mundial de la salud*. Obtenido de <https://www.who.int/es/news/item/27-04-2020-who-timeline---covid-19>
- Organización Panamericana de la Salud, O. (2017). Lineamientos básicos para el análisis de la mortalidad. *Organización Panamericana de la Salud*.
- Pineda, S. C., & Rivas, C. L. (2016). Regresión geográficamente ponderada: aplicación a datos de suicidio en la Región del Bío-Bío, Concepción, Chile. *Medwave*, 66-73.
- Ramírez, L., & Falcón, V. (2015). AUTOCORRELACIÓN ESPACIAL: ANALOGÍAS Y DIFERENCIAS ENTRE EL INDICE DE MORAN Y EL INDICE GETIS Y ORD. APLICACIONES CON INDICADORES DE ACCESO AL AGUA EN EL NORTE ARGENTINO. San Luis: ResearchGate.
- Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., . . . Roser, M. (2020). Coronavirus Pandemic (COVID-19). *Our World In Data*.
- Salas, G. D. (2006). *Modelos de mezcla de distribuciones alpha-estables. Aplicación a micromatrices de expresión genética*. Granada: Editorial de la Universidad de Granada.

Velasco, A., & Hurtado, C. X. (2020). La pandemia en Ecuador. *Desigualdades, impactos y desafíos*, 4-5. Obtenido de <http://library.fes.de/pdf-files/bueros/quito/16860.pdf>

Vera, F., Solórzano, M., Gina, O., García, B. S., & Cevallos, S. (2018). Tablas de mortalidad de Ecuador continental mediante un análisis de supervivencia. *Papeles de población*, 24(97), 63-83.

Yanover, M., Parolini, S., Rodríguez, S. d., Blanco, L., & Toledo, L. (2021). *Exceso de mortalidad en el contexto de la pandemia Covid-19. Provincia de Córdoba - Argentina*. Organización Panamericana de la Salud.

6. Anexos

Tabla 11: Codificación de las enfermedades agravantes del COVID-19 según el CIE-10

Diabetes mellitus y elevaciones de glucosa	E10	Diabetes mellitus insulino dependiente
	E11	Diabetes mellitus no insulino dependiente
	E13	Otras diabetes mellitus especificadas
	R73	Nivel elevado de glucosa en la sangre
Enfermedades cardíacas	T82	Complicaciones de dispositivos protésicos, implantes e injertos cardiovasculares
	I70	Aterosclerosis
	I21	Infarto agudo de miocardio
	I50	Insuficiencia cardíaca
	I46	Paro cardíaco
	I47	Taquicardia paroxística
Enfermedades hipertensivas	I10	Hipertensión esencial (primaria)
	I11	Enfermedad cardíaca hipertensiva

Enfermedades cancerígenas y tumores benignos	C15	Tumor maligno del esófago
	C16	Tumor maligno del estómago
	C22	Tumor maligno del hígado y de las vías biliares intrahepáticas
	C24	Tumor maligno de otras partes y de las no especificadas de las vías biliares
	C25	Tumor maligno del páncreas
	C38	Tumor maligno del corazón, del mediastino y de la pleura
	C53	Tumor maligno del cuello del útero
	C61	Tumor maligno de la próstata
	C95	Leucemia de células de tipo no especificado
	D17	Tumores benignos lipomatosos
	D21	Otros tumores benignos del tejido conjuntivo y de los tejidos blandos
	D23	Otros tumores benignos de la piel
	D24	Tumor benigno de la mama
	D27	Tumor benigno del ovario
	D29	Tumor benigno de los órganos genitales masculinos
D37	Tumor de comportamiento incierto o desconocido de la cavidad bucal y de los órganos digestivos	

	D39	Tumor de comportamiento incierto o desconocido de los órganos genitales femeninos
	D40	Tumor de comportamiento incierto o desconocido de los órganos genitales masculinos
	D43	Tumor de comportamiento incierto o desconocido del encéfalo y del sistema nervioso central
	D48	Tumor de comportamiento incierto o desconocido de otros sitios y de los no especificados
	Z51	Atención paliativa
Enfermedades pulmonares	J00	Rinofaringitis aguda
	J02	Faringitis aguda
	J03	Amigdalitis aguda
	J15	Neumonía bacteriana, no clasificada en otra parte
	J18	Neumonía, organismo no especificado
	J20	Bronquitis aguda
	J22	Infección aguda no especificada de las vías respiratorias inferiores
	J36	Absceso periamigdalino
	J40	Bronquitis, no especificada como aguda o crónica

	J44	Otras enfermedades pulmonares obstructivas crónicas
	J45	Asma
	J69	Neumonitis debida a sólidos y líquidos
	J80	Síndrome de dificultad respiratoria del adulto
	J84	Otras enfermedades pulmonares intersticiales
	J90	Derrame pleural no clasificado en otra parte
	J93	Neumotórax
	J96	Insuficiencia respiratoria, no clasificada en otra parte
	R04	Hemorragias de las vías respiratorias