

# Mortalidad por neumonía en Colombia 2010 – 2019: análisis y predicción con series de tiempo

Pneumonia mortality in Colombia 2010 – 2019: analysis and prediction with time series

Author:

- Arias-Ortiz Wanderley-Augusto<sup>1\*</sup>
- Barajas-Tarazona Erik-Jesús<sup>2</sup>
- González-Veloza John-Fredy<sup>3</sup>

## Innovaciencia

ISSN: 2346-075X

E- ISSN: 2346-075X

Innovaciencia 2022; 10(1); 1-15

<http://dx.doi.org/10.15649/2346075X.2846>

### INVESTIGACIÓN CIENTÍFICA

#### Cómo citar este artículo:

Cómo citar este artículo: Arias-Ortiz W., Barajas-Tarazona E., González-Veloza J., Mortalidad por neumonía en Colombia 2010 – 2019: análisis y predicción con series de tiempo, *Innovaciencia* 2022; 10(1): 1-15, DOI: <http://dx.doi.org/10.15649/2346075X.2846>

#### Fecha de recepción:

Recibido: 10 de agosto de 2022

Aceptado: 14 de septiembre de 2022

Publicado: 01 de diciembre de 2022

#### Palabras clave:

Neumonía; Estudios de Series Temporales; Salud Pública; Mortalidad.

#### Keywords:

Pneumonia; Time Series Studies; Colombia; Public Health; Mortality.

### RESUMEN

**Introducción:** La neumonía es una patología de origen respiratorio principalmente de origen infeccioso y que representa la séptima causa de muerte en Colombia, con una alta incidencia de morbilidad y mortalidad en menores de cinco años. De cara al cierre del PDSP 2011-2022, el contexto de la pandemia por COVID-19 y la creación del nuevo plan de desarrollo y PDSP se hace necesario caracterizar la mortalidad por neumonía y diseñar un modelo que permita predecir el comportamiento de la mortalidad para orientar los esfuerzos en salud pública. **Metodología:** Se diseñó un ecológico analítico con análisis de series de tiempo y una predicción empleando la metodología y Box-Jenkins. **Resultados y conclusiones:** Se identificó que hay mayor proporción de mortalidad en adultos mayores de 65 años, en cuanto al sexo, no se identificaron diferencias en las proporciones, en cuanto al régimen de aseguramiento se identificaron diferencias estadísticas, así como una asociación estadística entre las variables de régimen de aseguramiento y sexo ( $p < 0,001$ ). Por otro lado, se logró diseñar un modelo SARI-MA  $(0,1,2) \times (1,1,1)_{12}$  con un AIC=1252.81, un RMSE=96 y un MAPE=22,0%.

### ABSTRACT

**Introduction:** pneumonia is a respiratory pathology mainly of infectious origin and represents the seventh leading cause of death in Colombia, with a high incidence of morbidity and mortality in children under five years of age. In view of the closing of the PDSP 2011-2022, the context of the covid-19 pandemic and the creation of the new development plan and PDSP, it is necessary to characterize pneumonia mortality and design a model to predict mortality behavior to guide public health efforts. **Methodology:** an analytical ecological analysis was designed with time series analysis using the methodology of Bello l & Martínez and Box-Jenkins. **Results:** it was identified that there is a higher proportion of mortality in adults over 65 years of age; as for sex, no differences were identified in the proportions; as for the insurance regime, statistical differences were identified, as well as a statistical association between the variables of insurance regime and sex ( $p < 0.001$ ). On the other hand, a SARIMA  $(0,1,2) \times (1,1,1)_{12}$  model was designed with an AIC=1252.81, an RMSE=96 and a mape=22.0%. **Conclusions:** a possible upward trend was identified from 2014 to 2018, and a decreasing trend after 2018. Pneumonia mortality in Colombia was mainly represented by the age group 65 years and older with 73.6% ( $n=42,070$ ), likewise, it is relevant to mention that there were no differences in the distribution by sex. A strong difference between insurance regimes was identified. A SARIMA $(0,1,2) \times (1,1,1)_{12}$  model was defined with a RSME of 96 and a MAPE of 22.0%.



- <sup>1\*</sup> Corresponding author. Docente, Universidad El Bosque, Facultad Escuela Colombiana de Medicina, Grupo de Investigación: Medicina Comunitaria y Salud Colectiva, [waarias@unbosque.edu.co](mailto:waarias@unbosque.edu.co)
- <sup>2</sup> Matemático, Fundación Universitaria Los Libertadores, Facultad de Ingeniería y Ciencias Básicas, Especialización en Estadística Aplicada, [ejbarajas@libertadores.edu.co](mailto:ejbarajas@libertadores.edu.co)
- <sup>3</sup> Docente, Fundación Universitaria Los Libertadores, Facultad de Ingeniería y Ciencias Básicas, Especialización en Estadística Aplicada, [jgonzalezv02@libertadores.edu.co](mailto:jgonzalezv02@libertadores.edu.co)

## INTRODUCCIÓN

La neumonía es una infección que afecta principalmente al tejido parenquimatoso pulmonar. Esta infección puede estar ocasionada por distintos patógenos, entre los cuales se destacan el *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, y recientemente el SARS-CoV-2 (1). Dentro del ámbito clínico, la neumonía se puede categorizar de acuerdo con la fuente de la infección en: hospitalaria, extrahospitalaria, asociada al ventilador y relacionada con las instalaciones de atención de la salud (2).

Según datos de la Organización Mundial de la Salud (OMS) para el 2015, el 15% de las muertes en menores de cinco años son atribuibles a la Neumonía, mortalidad que es acentuada si el niño tiene comorbilidades que comprometan su sistema inmunológico, o si está expuesto a factores como la contaminación del aire en entornos interiores, el hacinamiento y la exposición al humo de tabaco por parte de los padres (3).

### Contexto de la Mortalidad por Neumonía

De acuerdo con el observatorio de salud distrital de Bogotá, SALUDATA, para el 2016, la neumonía representaba la sexta causa de muerte en todos los grupos de edad en el distrito capital (4). Y según Global Burden of Disease 2019 en Colombia desde el 2009 al 2019 la mortalidad por infecciones respiratorias de vías inferiores tuvo un incremento del 12,4% (5), y con una carga de la enfermedad estimada de 225.558.08 Años de Vida Ajustados por Discapacidad (AVAD) IC95% [168.552.01 - 296.859.33] que la ubica en el séptimo lugar de los países con mayor AVAD en Sur América y el Caribe, superada por países como Brasil que tuvo 1.994.790.01 AVAD IC95% [1.851.232.50 - 2.127.570.61] y que supera a países como Venezuela que reportaron 153.211.88 AVAD IC95% [123.261.36 - 214.866.4] (6).

En Colombia, una de las políticas en salud pública más trascendentales es el Plan Decenal de Salud Pública (PDSP) 2012 - 2021, el cual se realizó en el contexto de la ley 1122 del 2007 que comenzó a dar relevancia a la salud pública como eje central de la salud y del Plan Nacional de Desarrollo (PND) 2010 – 2014. Uno de los frentes del PDSD es involucrar una política enmarcada en la “Equidad en Salud”, la cual se fundamenta en que “(...) *la salud se logre cuando todas las personas alcanzan su potencial de salud independientemente de sus condiciones sociales, culturales y económicas*” (7).

El PDSP cuenta con seis dimensiones que representan los componentes estructurales de la política de estado, en este sentido algunas de estas dimensiones involucran metas relacionadas con la calidad del aire, condiciones de hábitat y específicamente la mortalidad por enfermedades respiratorias. Al respecto se destacan la realización de estudios de carga ambiental de la enfermedad (dimensión Salud ambiental), reducir la mortalidad prematura por enfermedades no transmisibles, reducir el tabaquismo en personas de 18 a 69 años, lograr el cumplimiento de ambientes 100% libres de humo de tabaco y sus derivados a nivel nacional (dimensión Vida saludable y condiciones no transmisibles), implementación del programa de prevención y control de las Infecciones Asociadas a la Atención en Salud (IAAS), reducción progresiva de la mortalidad por Infección Respiratoria Aguda buscando llegar a una mortalidad por neumonía de menos de 8,9 casos por cada 100mil menores de cinco años (dimensión vida saludable y enfermedades transmisibles), sin embargo, para el informe presentado en el 2018 se evidenció una disminución del 14,2% (2012 – 2016) con 13.8 casos por 100.000 menores de cinco años (7,8).

Por otro lado, desde la formulación de la Política de Atención Integral en Salud (PAIS) y el Modelo Integral de Atención en Salud (MIAS), se formularon estrategias de atención con enfoque diferencial y se crearon Rutas Integrales de Atención en Salud (RIAS) cuyo objetivo principal es “*garantizar la atención integral en salud a las personas, familias y comunidades a partir de intervenciones de valoración integral de la salud, detección temprana, protección específica, diagnóstico, tratamiento, rehabilitación, paliación y educación para la salud, teniendo en cuenta el mejoramiento de la calidad en todo el continuo de la salud*”(9), de este modo, dichas rutas se han fundamentado en tres: Ruta de Promoción y Mantenimiento de la salud, Rutas integrales de atención para grupos de riesgo, y Rutas integrales de atención específica para eventos, lo que permite inferir un esfuerzo por parte del estado en generar política pública alrededor de patología no transmisibles y que, son susceptibles de evaluar. (10)

### **Análisis de Series de Tiempo en Salud**

Una serie de tiempo es un conjunto de datos ordenados en el tiempo que están igualmente espaciados, estas se pueden descomponer en tres elementos: la tendencia, estacionalidad y residuales. La tendencia representa los cambios (aumentos o disminución) de la serie, la estacionalidad es el periodo de tiempo en el que ocurren determinados patrones, y los residuales representan el comportamiento que no puede ser explicado por la tendencia y la estacionalidad (11).

Existen diversos modelos de series de tiempo, dentro de los cuales se encuentra el SARIMA que permite capturar la estacionalidad (S) de la serie, en la combinación de términos autorregresivos (AR), y en términos de promedios móviles (MA) con un elemento que me indica la diferencia de la serie (I) a partir del criterio del estadístico. Es usual que estos modelos se escriban como  $SARIMA(p,d,q,P,D,Q)$ , en el que el término  $(p)$  es el orden AR no estacional, la  $(d)$  es la diferenciación de orden no estacional,  $(q)$  es el término MA de orden no estacional,  $(P)$  término AR de orden estacional,  $(D)$  la diferenciación de orden estacional y  $(Q)$  el término MA de orden estacional. De este modo los modelos SARIMA tienen la forma:

#### **Ecuación 1. Forma del modelo SARIMA.**

$$\Phi_P(B^S)\phi_p(B)(1 - B^S)^D(1 - B)^dX_t = \Theta_Q(B^S)\theta_q(B)Z_t \quad (1)$$

Donde  $\Phi$  y  $\phi$  son los coeficientes autorregresivos estacional y no estacional respectivamente,  $\Theta$  y  $\theta$  son los coeficientes de media móvil estacional y no estacional a determinar,  $X_t$  es la observación en el tiempo  $t$ ,  $Z_t$  son los errores de estimación en el instante  $t$ , y  $B$  es un operador de retroceso.

Un paso importante al realizar un modelo es evaluar la precisión del pronóstico, para ello existen diferentes métricas para conocer la magnitud del error (12). Una de ellas es el Error Cuadrático Medio (RMSE) que se define como la raíz cuadrática del error medio:

#### **Ecuación 2. Error Cuadrático Medio (RMSE).**

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (2)$$

El MAPE (Error Porcentual Absoluto Medio) es otra medida de precisión, definida como la suma de los errores absolutas divididas por el valor actual ( $x_t$ ).

#### **Ecuación 3. Error Porcentual Absoluto Medio (MAPE)**

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{d_t} \quad (3)$$

De acuerdo con, la revisión bibliográfica los modelos de series temporales son ampliamente utilizados para predecir y analizar enfermedades infecciosas como es el caso del estudio realizado por Luz et al. 2008 que ajustaron un (modelo autorregresivo integrado de media móvil (SARIMA) para predecir la incidencia del dengue en Rio de Janeiro, Brasil de

1997 a 2004, aplicando la metodología Box-Jenkins, con el fin de plantear estrategias adecuadas para reducir el aumento de casos. El mejor modelo encontrado fue un  $SARIMA(2,0,0) \times (1,0,0)_{12}$ , de tal manera que los casos de dengue se puedan predecir con el número de casos ocurridos en uno, dos y doce meses antes (13).

Otros estudios parecidos fueron realizados por Reichert et al. 2018, para predecir los casos de gripe y su asociación con la gripe en Estados Unidos, 1955 - 1999. Otros estudios similares fueron hechos por Ruchiraset et al., 2018 cuyo objetivo era predecir el número de casos de personas que presentaban neumonía en la provincia de Chiang Mai, Tailandia, y su asociación con la polución del aire (PM10) con un modelo autorregresivo integrado de media móvil (SARIMA), donde los resultados mostraron que el mejor modelo que se ajustó fue un  $SARIMA(1,0,2) \times (2,0,0)$  teniendo en cuenta doce años de casos mensuales, variables climáticas y la contaminación para obtener el modelo más óptimo (14,15).

Finalmente, teniendo en cuenta el marco de las políticas de salud pública, y el contexto epidemiológico y de salud de Colombia, se hace necesario realizar un análisis frente al comportamiento de la mortalidad por neumonía teniendo en cuenta que la finalización del PDSP fue en el 2021 y es relevante generar recomendaciones y oportunidades de mejora frente a las metas propuestas. En este sentido, los objetivos de este estudio son describir el comportamiento de la mortalidad por neumonía no especificada en Colombia para el periodo 2010 – 2020 teniendo en cuenta variables sociodemográfica, y predecir el comportamiento para próximos periodos de la mortalidad por neumonía no especificada.

## MATERIALES Y MÉTODOS

Se realizó un estudio ecológico retrospectivo analítico de recolección documental, con análisis de serie de tiempo. Los datos se obtuvieron de las Estadísticas Vitales (EEVV) reportadas por el Departamento Administrativo Nacional de Estadística (DANE) y el Archivo Nacional de Datos (ANDA) de Colombia.

Se incluyeron las muertes reportadas desde enero del 2010 y hasta diciembre de 2020 que tuvieran como causa básica de muerte el código J189 “Neumonía no especificada”, se excluyeron los casos que presentaran una proporción de datos faltantes superior al 30% en las variables incluidas en el estudio.

Para el análisis descriptivo de la mortalidad se seleccionaron las variables que se muestran en la Tabla 1. Teniendo en cuenta el tipo de variable, se calcularon frecuencias absolutas y porcentajes, y asociación estadística con la prueba de Chi-Cuadrado de Pearson.

**Tabla 1. Variables incluidas en el análisis.**

Variable	Definición Operativa	Tipo de variable	Escala de medición
Sexo	Condición biológica del fallecido	Cualitativa	Nominal, politómico (Masculino, Femenino, Indeterminado)
Edad	Edad del fallecido al momento de la muerte de acuerdo con la agrupación de edades propuesta por el DANE	Cuantitativa	Ordinal, politómico (Menores de 1 año, de 1 a 4 años, de 5 a 14 años, de 15 a 44 años, de 45 a 64 años, de 65 y más años, edad desconocida)
Departamento de residencia del fallecido	Departamento de residencia del fallecido teniendo en cuenta los códigos de departamento establecidos por el DANE	Cualitativa	Nominal, politómico correspondiente a los departamentos de Colombia
Régimen de Aseguramiento	Régimen de aseguramiento en salud al que pertenecía el fallecido	Cualitativa	Nominal, politómico (Contributivo, Subsidiado, Excepción, Especial, No asegurado, Sin información)

Fuente: Construcción propia.

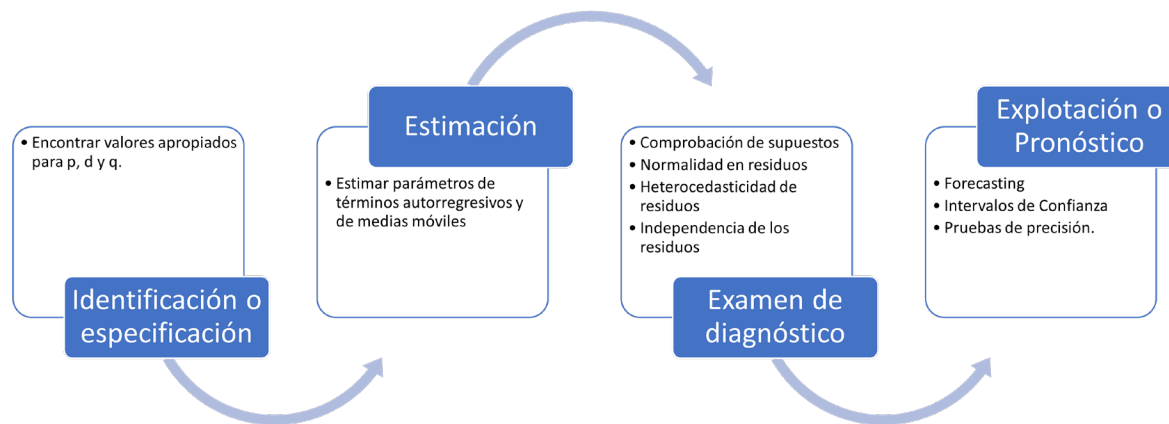
En el análisis de serie de tiempo se seleccionó como variable principal la frecuencia mensual de la mortalidad por la causa anteriormente mencionada, siguiendo la metodología propuesta por Bello L & Martínez 2002 (14).

En primer lugar, se analizó si la serie de datos correspondía a una serie no aleatoria mediante el análisis gráfico de la secuencia y la prueba de Rachas, a continuación, se realizó un análisis exploratorio de los datos en cual se identificaron valores atípicos, extremos y perdidos, y se analizó gráficamente el comportamiento de la serie (16).

En segundo lugar, teniendo en cuenta la tendencia de la serie, se procedió a aplicar una diferenciación por periodo y una diferenciación anual. Se realizó la caracterización de los componentes de la serie para determinar la estacionalidad, se analizaron los autocorrelogramas diferenciados, se empleó la prueba de Dickey-Fuller para evaluar la estacionariedad y, se aplicó un análisis gráfico para determinar la tendencia.

Posteriormente, para realizar la identificación estadística del modelo, se aplicó la metodología Box-Jenkins, en la cual se realiza un proceso iterativo que inicia con la identificación primaria del modelo, luego la estimación de parámetros, a continuación, se procede con el diagnóstico del modelo y, finalmente la explotación o pronóstico, esta última se realizó dividiendo el conjunto de datos en dos partes, una de entrenamiento que va desde 2010-01-01 hasta 2019-12-01 y la otra parte de prueba desde 2020-01-01 hasta 2020-12-01 (17). (Figura 1)

**Figura 1. Metodología Box-Jenkins.**



**Adaptado de:** Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control, 5th Edition | Wiley. 2021;

Luego teniendo en cuenta en primera instancia el autocorrelograma y autocorrelograma parcial, y a partir de esto, se diseñaron distintos modelos de diferentes órdenes SARIMA, y se determinó el mejor modelo aplicando el Criterio de Información Akaike (AIC), el error cuadrático medio (RMSE), Error porcentual absoluto medio (MAPE) y el principio de parsimonia (18). Posterior a ello, para aplicar la validación de supuestos del modelo, se realizó el análisis de heterocedasticidad, aleatoriedad y normalidad de los residuales aplicando las pruebas de Goldfeld-Quant, Ljung-Box y Shapiro-Wilk respectivamente (19,20).

El análisis descriptivo fue realizado mediante el software estadístico SPSS vr.26 y para el análisis de series de tiempo se aplicó el lenguaje de programación Python con las librerías Statsmodels, Pandas, Numpy, Matplotlib (21-25). El nivel de significancia estadística fue de 0.05.

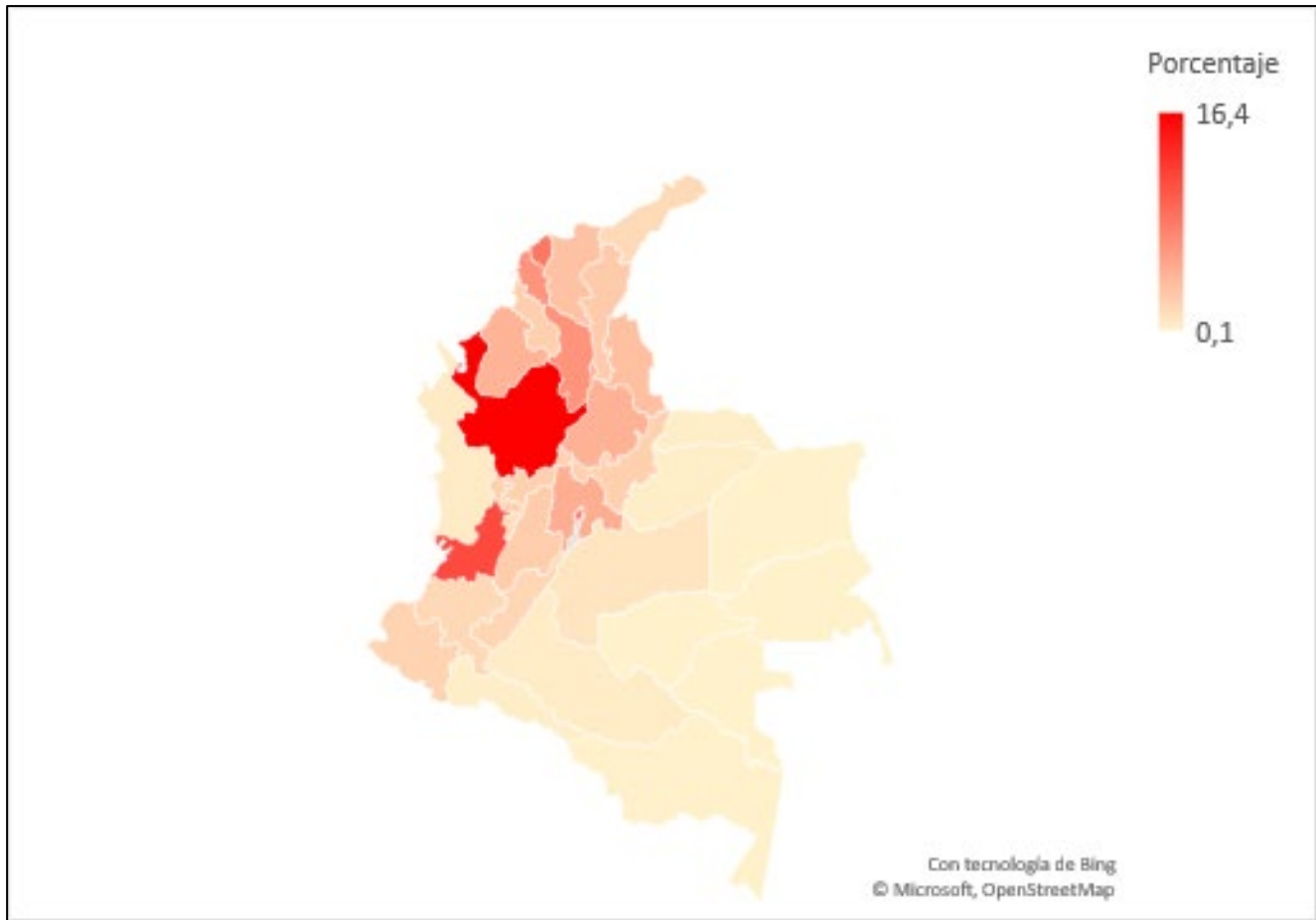
## RESULTADOS

Se incluyeron reportes de 57.133 fallecimientos desde enero de 2010 hasta diciembre de 2020 obtenidos de los reportes del ANDA y DANE.

### *Descripción de la Mortalidad por Neumonía en Colombia 2010 – 2020.*

En cuanto al comportamiento de la mortalidad, se identificó que la mayor proporción de muertes para el periodo estuvo reportada en la Región Andina, específicamente en departamentos como Antioquia con un 16.40% (n=9.385), seguido de Valle del Cauca con un 11.40% (n=6485). (Figura 2)

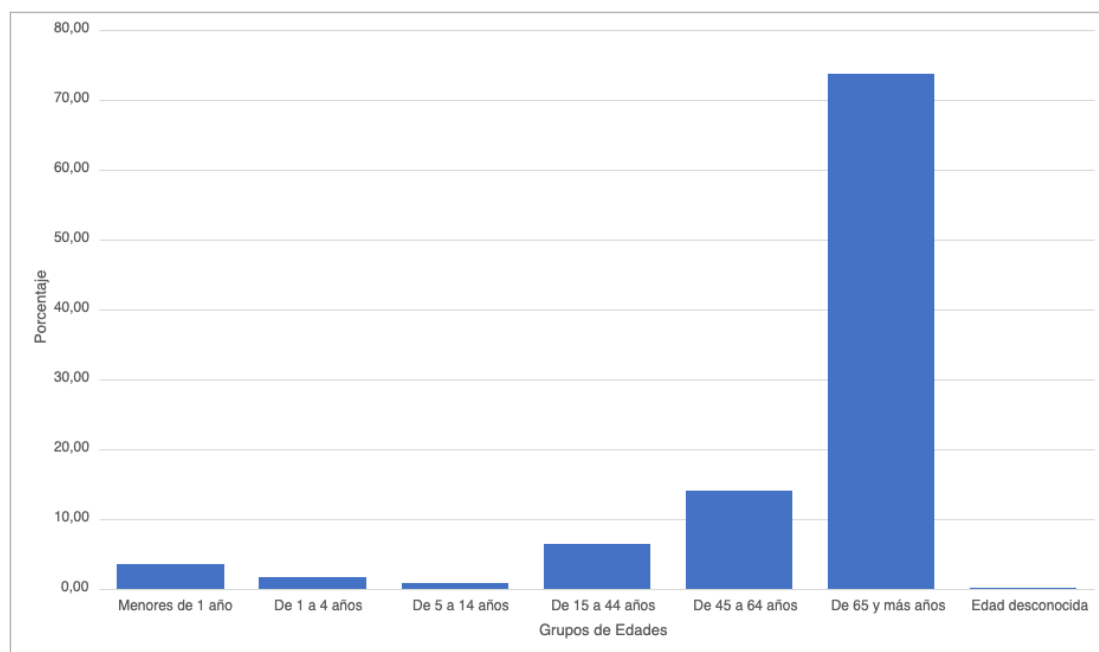
**Figura 2. Distribución geográfica de la mortalidad por neumonía en Colombia periodo 2010 – 2020.**



Fuente: Construcción propia.

En cuanto a la distribución por sexo, se identificó que la mortalidad estuvo representada de una forma similar entre hombres y mujeres 50.5% (n=28.831) y 49.5% (n=28.302) respectivamente. Por su parte en la edad, se identificó que el 73,6% (n=42.070) fue para el grupo de 65 años y más, mientras que el grupo de 5 a 14 años fue el que presentó una menor frecuencia 0,8% (n=474), también se identificaron algunos sujetos con edad desconocida (n= 23). (Figura 3)

**Figura 3. Distribución de muertes por neumonía no especificada por grupos de edad para el periodo 2019 – 2020 en Colombia.**



Fuente: Construcción propia.

Por otro lado, se encontró que el 51,58% (n=29.469) de las muertes se presentaron en el régimen subsidiado, de los cuales la distribución por sexo tuvo una representación que varió en aproximadamente un punto porcentual, en contraste con lo ocurrido en el régimen contributivo que representó un 38,77% (n=22.151) con una representación del sexo femenino del 40,81% (n=11.550) y el masculino con 36.77% (n=10.601). Al evaluar la asociación entre el régimen de seguridad social y el sexo se encontró un  $\chi^2= 234,28$   $p<0.001$ . (Tabla 2)

**Tabla 2. Distribución de muertes por neumonía no especificada por régimen de aseguramiento y sexo periodo 2019 - 2020 en Colombia.**

Régimen de seguridad social	Sexo				Total por régimen	
	Masculino		Femenino			
	n	% <sup>¥</sup>	n	% <sup>¥</sup>	n	% <sup>§</sup>
Contributivo	10601	36,77	11550	40,81	22151	38,77
Subsidiado	15021	52,10	14448	51,05	29469	51,58
Excepción	1060	3,68	890	3,14	1950	3,41
Especial	86	0,30	101	0,36	187	0,33
No asegurado	1681	5,83	1109	3,92	2790	4,88
Sin información	382	1,32	204	0,72	586	1,03
Total	28831	100	28302	100	57133	100

<sup>¥</sup>Por sexo. <sup>§</sup>Por régimen

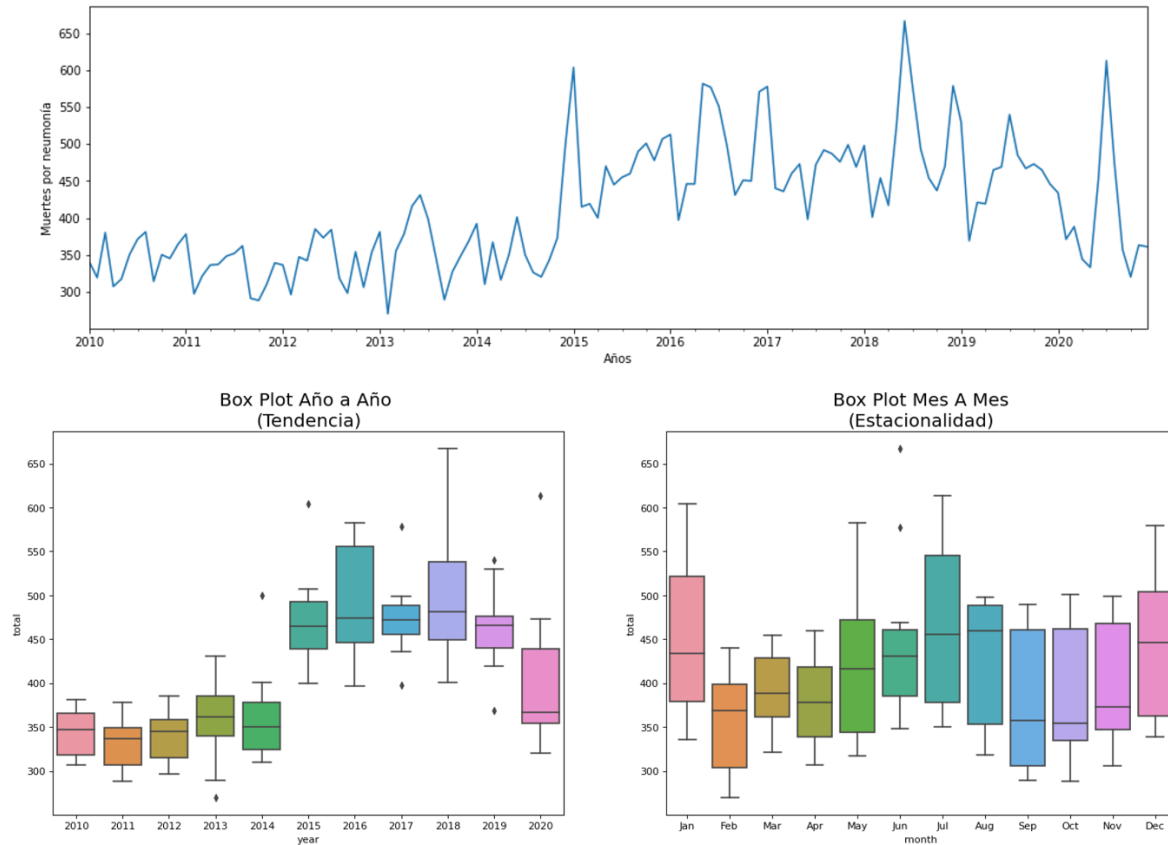
Fuente: elaboración propia.



### Caracterización de la Serie de Tiempo

Al realizar el análisis de la serie de tiempo, se identificó una posible tendencia al alza del periodo 2014 al 2019, y una decreciente posterior al 2019, sin embargo, en este último año se encontró un comportamiento atípico superior. Por otro lado, se calculó la prueba de Rachas  $Z = -8.72$   $p < 0,001$  indicando que los datos no se distribuyeron de forma aleatoria. Por otro lado, se identificaron valores atípicos en los años 2013 que correspondió a febrero ( $n=270$ ), 2014 en diciembre ( $n=500$ ), 2015 en enero ( $n=604$ ), 2017 en junio ( $n=398$ ) y enero ( $n=578$ ), 2019 en febrero ( $n=369$ ) y julio ( $n=540$ ), y 2020 en julio ( $n=613$ ), y mensualmente, se identificaron valores atípicos en el mes de junio correspondientes a los años 2016 ( $n=577$ ) y 2018 ( $n=667$ ). (Figura 4)

**Figura 4. Análisis gráfico de la serie de tiempo de mortalidad por neumonía en Colombia 2010 – 2020.**

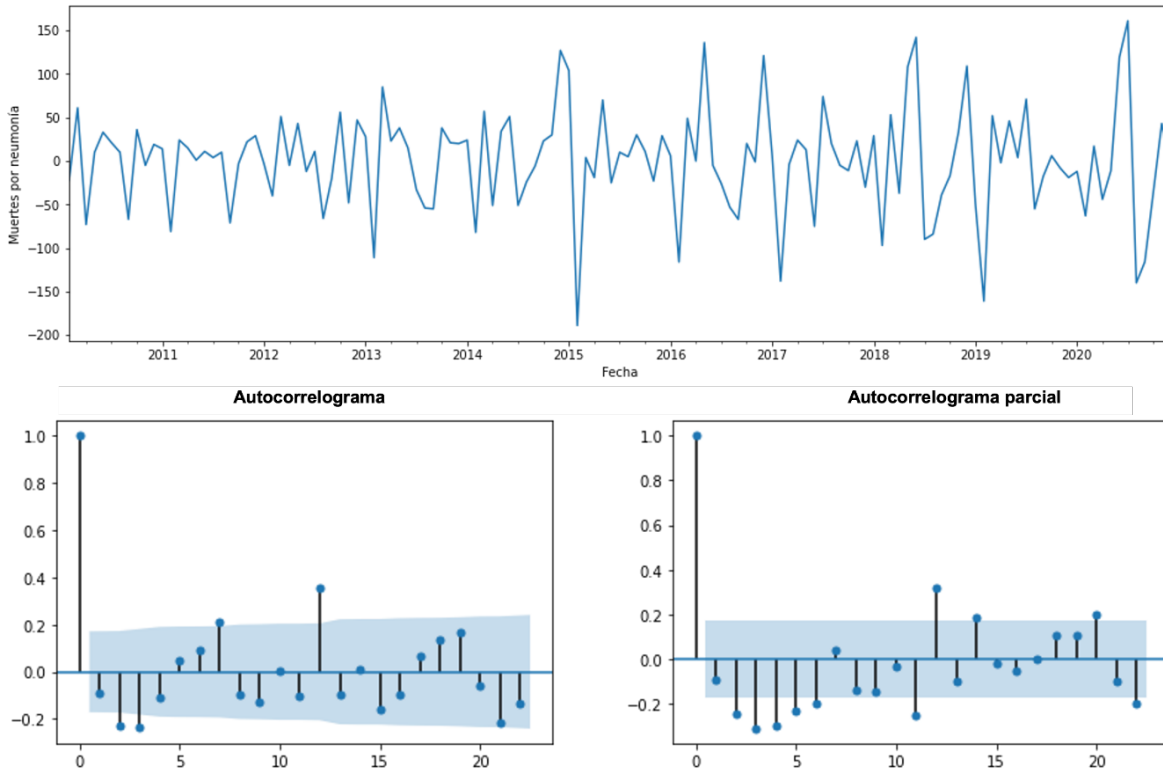


Fuente: construcción propia.

A partir de la diferenciación de la serie, se obtuvo una serie de tiempo sin tendencia y estacionaria, con una estacionalidad anual identificada a partir del autocorrelograma (ACF) y el autocorrelograma parcial (PACF) en los cuales se evidencia una significancia estadística fuerte en el rezago 12, y en cuanto a la evaluación de estacionariedad se encontró una prueba Dickey-Fuller  $Z = -3.59$   $p = 0.005$  lo que indica que la serie es estacionaria. (Figura 5)



**Figura 5. Comportamiento de la serie diferenciada y diferencia estacional de orden 12.**



Fuente: construcción propia.

Partiendo del ACF y PACF se propusieron modelos  $SARIMA(p,d,q,P,D,Q)_{12}$  con distintos órdenes en los parámetros  $p(0,1,2)$ ,  $q(0,1,2)$ ,  $P(0,1)$ ,  $Q(0,1)$  y  $d,D(1)$ , finalmente, teniendo en cuenta el criterio AIC principalmente y otras medidas de precisión como el RMSE y el MAPE. A partir de esto, se determinó que el mejor modelo con el mejor ajuste fue el  $SARIMA(0,1,2) \times (1,1)_{12}$  de acuerdo con el criterio AIC es el que presenta menor valor, (Tabla 3) el RMSE fue de 96 y el MAPE de 22,0%

**Tabla 3. Resultados de la simulación de modelos SARIMA12 de diferentes órdenes.**

Orden	AIC
SARIMA(0,1,2)x(1,1,1)	1252,81
SARIMA(1,1,2)x(1,1,1)	1254,22
SARIMA(2,1,1)x(1,1,1)	1254,37
SARIMA(0,1,2)x(1,1,0)	1255,76
SARIMA(2,1,1)x(0,1,1)	1257,24

Fuente: construcción propia.

Al analizar los supuestos del modelo seleccionado se identificó que en la prueba Goldfeld-Quandt  $F=2.47$   $p=0.01$  comprobando que el modelo es heterocedástico, Ljung-Box  $X^2=38.97$   $p=0,52$  indicando que los residuales se distribuyen de forma independiente, Shapiro-Wilk  $p=0.005$ . (Tabla 4).

**Tabla 4. Coeficientes del modelo.**

Componentes	Coeficientes	Error Estándar	Z	p	IC95%	
					Inf	Sup
MA1	-0,23	0,10	-2,28	0,02	-0,42	-0,03
MA2	-0,41	0,10	-3,89	0,00	-0,62	-0,20
ARs 1	-0,43	0,14	-2,98	0,00	-0,71	-0,14
MA s 2	-0,52	0,14	-3,61	0,00	-0,80	-0,24
Varianza	1541,98	204,23	7,55	0,00	1141,69	1942,26

Fuente: construcción propia.

**Simulación de datos.**

Una vez definido el modelo, se evalúa la magnitud de error con el conjunto de datos de prueba del año 2020. En la tabla 5 columna 2 se registran los datos reales y en la 3 los generados por el modelo. Las columnas 5 y 6 son los intervalos de confianza de la predicción. (Tabla 5)

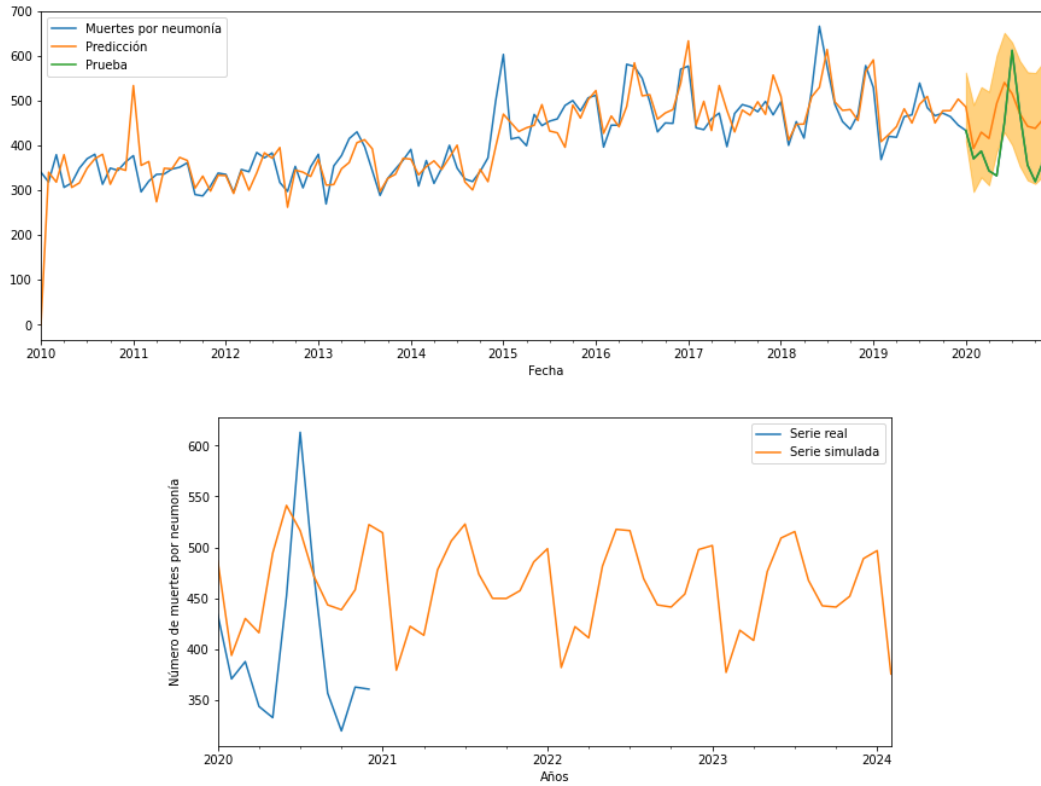
**Tabla 5. Resultados de Simulación.**

Mes	Muertes	Predicción	IC 95%	
			Límite inferior	Límite superior
2020-01-01	434	487	410	564
2020-02-01	371	394	297	491
2020-03-01	388	430	329	531
2020-04-01	344	416	312	521
2020-05-01	333	495	387	602
2020-06-01	452	541	430	653
2020-07-01	613	516	402	631
2020-08-01	473	472	354	589
2020-09-01	357	444	323	564
2020-10-01	320	439	315	563
2020-11-01	363	458	332	585
2020-12-01	361	522	393	652

Fuente: Elaboración propia

En la figura 6 se muestra dos gráficos, el primero se compone de una línea azul de las muertes registradas hasta diciembre del 2019, la línea naranja son los pronósticos realizados por el modelo, la verde representa el conjunto de prueba del año 2020 y una franja naranja que indica los intervalos de confianza para los datos de prueba. En la segunda gráfica se muestra el modelo para un intervalo de tiempo mayor, considerando la predicción hasta finales del año 2023, mostrando un patrón estacional con predicciones de valores mayores en los meses de julio y agosto, y el menor registro ocurre en el mes de febrero. (Figura 6)

**Figura 6. Gráfico de modelo de predicción de la serie de tiempo.**

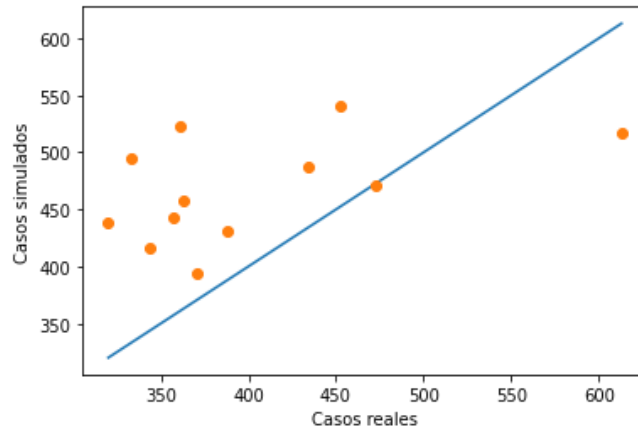


Fuente: Elaboración propia.

Para entender mejor como los resultados de la predicción se ajustan a la situación real, se realizó la figura 7. En dicha figura se evidencia la evaluación de la correspondencia de los casos reales versus los simulados, para lo cual se trazó una línea azul que representaría una juste perfecto de los datos.

Un buen indicador de que los datos simulados se ajustan bien a los casos reales es que los puntos tengan una buena aproximación a la línea azul, en este caso se puede resaltar que en los casos reales que se registran muertes menores a 450, el modelo tiende a aumentar esas muertes, es decir sobrestimando los casos. (Figura 7)

**Figura 7. Casos reales frente a simulados.**



Fuente: construcción propia.

## DISCUSIÓN

Dentro de los principales hallazgos del estudio se puede resaltar que la mortalidad por neumonía en Colombia estuvo principalmente representada por el grupo de edad de 65 años y más con el 73,6% (n=42.070), de igual forma, es relevante mencionar que no se presentaron diferencias en la distribución por sexo. En cuanto a la distribución geográfica se hace importante describir que la mayor proporción de las muertes se reportaron en el departamento de Antioquia, el cual se ha caracterizado por tener altos indicadores en la calidad del aire, lo cual puede representar uno de los principales factores que influyen en la mortalidad(26).

Al respecto el estudio realizado por Romero A en 2009 frente a factores asociados con la mortalidad de ancianos hospitalizados por neumonía en comunidad, menciona que los pacientes con 85 años y más tienen un OR=3,14 IC95% [1,11 – 8,82] representando un riesgo tres veces mayor en pacientes de esta edad comparado con el resto de las edades, lo cual coincide con los hallazgos del estudio, de igual forma, se menciona que el sexo no presenta un factor de riesgo con OR=1,00 IC95% [0,41 – 2,41] lo cual también se constata en el estudio, sin embargo, cabe aclarar que el estudio realizado por Romero A fue en el contexto Cubano, lo cual permitiría comparar que el escenario de la isla pudiera ser de cierta forma similar al colombiano, hecho que debería evaluarse en más estudios (27).

Sin embargo, en el estudio de Machado-Alba J, et al del 2013 (n=297) que está adaptado al contexto colombiano, se reportó que, a partir de un modelo de regresión logística, la variable de 65 años o más (OR<sub>A</sub>=0,55 IC [0,17 – 1,72] no contó con significancia estadística (28). Entre tanto, en el estudio de Quintero-González, et al de 2020 (n=114) también para el contexto colombiano, se identificó que una edad mayor de 90 años tenía un RR=1,62 IC95%[1,05 – 2,68] p=0,04 para letalidad por neumonía a los 30 días (29). Lo cual, contrasta con los resultados de este estudio en términos de la alta proporción de fallecidos de 64 y más años.

Por otro lado, partiendo de una perspectiva económica, se hace relevante considerar el alto costo de la neumonía a nivel hospitalario, teniendo en cuenta que es una condición que es susceptible de ser prevenida, al respecto el estudio de Calderón C y Dennis R 2014, estima que el costo promedio de la atención por neumonía por *Streptococcus pneumoniae* es de 12'264.174±14'264.174 que en mayor proporción se debe a los medicamentos (44%) y la estancia hospitalaria (35%), lo cual implica un llamado a la prevención de la enfermedad y a la promoción de la salud (30).

En relación con la prevención de la neumonía, específicamente para el caso de la Neumonía Adquirida en Comunidad, para el contexto colombiano la Asociación Colombiana de Infectología generó la Guía para la inmunización del adolescente y adulto, en la cual se describen recomendaciones relacionadas con vacunación y, por otro lado, el National Health, Lung and Blood Institute recomienda acciones como el lavado de manos, mitigar el tabaquismo (31,32).

Adicional a esto, se puede decir que una de las principales causas de hospitalización por neumonía se debe a las condiciones de calidad del aire, en el estudio de Neupane B et al 2010 se reporta que las variables de contaminación NO<sub>2</sub> presentan un OR=2,30 IC95% [1,25 – 4,21] p=0,007, y PM<sub>2,5</sub> un OR= 2,26 IC [1,20 – 4,24], lo cual indica el impacto de la calidad del aire en la neumonía (33). En este sentido, según el Informe del estado de la calidad del aire en Colombia 2020 el riesgo atribuible por contaminación de aire extramural en Colombia para Enfermedad Respiratoria Aguda es del 17,0% comparado con un global del 8,0%, y se presenta una mayor exposición a mala calidad del aire en regiones como el Urabá Antioqueño y la región andina, lo cual puede relacionarse con la mayor ocurrencia de casos de mortalidad en estas regiones de acuerdo con los hallazgos del estudio (34).

Por otro lado, se ha llegado a un modelo SARIMA que mostro el menor valor AIC que pronostica el número de casos de muertes por neumonía utilizando los errores de estimación de los casos registrados de uno, dos meses atrás y las ocurrencias de doce meses antes por la estacionalidad hallada. El modelo realiza predicciones para un año con un RMSE o una variabilidad de 95.96 que en la mayoría de los casos aumentan su valor en los casos que se registraron muertes menores a 450 y presenta un porcentaje de equivocación del 22,0%, tendiendo a sobrestimar el número de casos.

El patrón de las muertes por neumonía es similar cada año, aumenta en los meses de enero, julio y diciembre que tienen los mayores niveles de contaminación y disminuye en los otros en especial el mes de febrero se registran menos casos, esta estacionalidad es captada por el modelo(34–36).

Algunos estudios han comprobado la relación de la contaminación con muertes asociadas a la neumonía es un factor importante para predecir, como en este trabajo se realizó un análisis univariado se sugiere realizar estudios incluyendo la variable concentración de material particulado (PM) (35–37).

Una de las principales limitaciones del estudio es que el comportamiento de los residuales no cumplió con los supuestos de normalidad y homocedasticidad por la influencia de picos altos que se presentaron en los años 2014, 2018 y 2019, por consiguiente los intervalos de confianza no dan una medida fiable, como solución para este problema se planteó realizar una descomposición robusta, pero por la naturaleza de los datos no se realizó, teniendo en cuenta que para el contexto de las ciencias de la salud se prefiere no emplear métodos de sustitución o de reducción para valores atípicos.

También es relevante tener en cuenta que, el modelo que se debe ir recalibrando cada año para tener una mejor predicción del año siguiente. Otra de las limitaciones del estudio es la calidad de la información, teniendo en cuenta que al ser registrada en bases de datos puede ser susceptible de errores de digitación o de mala clasificación incurriendo en posibles sesgos de información, los cuales se intentaron controlar mediante la depuración de la base de datos, sin embargo, se sospecha de un sesgo residual.

De otro modo, cabe resaltar que no se encuentran muchos estudios en el contexto colombiano y latinoamericano que empleen modelos de series de tiempo como el SARIMA en enfermedades crónicas no transmisibles, como el caso de la neumonía, por tanto, se sugiere realizar estudios en el tema para fortalecer la comprensión de los fenómenos de salud-enfermedad en escenarios de investigación y academia.

## CONCLUSIONES

Se identificó que en términos sociodemográficos la mortalidad por neumonía se presenta en pacientes principalmente adultos, que la región antioqueña representa una mayor proporción de las muertes a nivel nacional. Por otro lado, se identificó que la proporción de muertes es similar para ambos sexos, y que esta variable, junto con el aseguramiento están asociadas significativamente.

Finalmente, se obtuvo un modelo  $SARIMA(0,1,2) \times (1,1,1)_{12}$  e con un RMSE de 96 y el MAPE de 22,0%, que tiene un buen ajuste predictivo que es útil para realizar pronósticos de mortalidad con un intervalo de confianza del 95%.

## REFERENCIAS

1. Neumonía | Diagnóstico clínico y tratamiento 2021 | AccessMedicina | McGraw Hill Medical [Internet]. [cited 2022 Apr 8]. Available from: <https://accessmedicina-mhmedical-com.ezproxy.unbosque.edu.co/content.aspx?sectionid=254925278&bookid=3002&Resultclick=2>
2. Neumonía, bronquiectasia y absceso pulmonar | Harrison. Manual de Medicina, 20e | AccessMedicina | McGraw Hill Medical [Internet]. [cited 2022 Apr 8]. Available from: <https://accessmedicina-mhmedical-com.ezproxy.unbosque.edu.co/content.aspx?sectionid=252610515&bookid=2943&Resultclick=2>
3. Neumonía [Internet]. [cited 2022 Apr 8]. Available from: <https://www.who.int/es/news-room/fact-sheets/detail/pneumonia>
4. Causas Mortalidad | SALUDATA [Internet]. [cited 2022 Apr 8]. Available from: <https://saludata.saludcapital.gov.co/osb/index.php/datos-de-salud/demografia/causasmortalidad/>
5. Abbafati C, Abbas KM, Abbasi-Kangevari M, Abd-Allah F, Abdelalim A, Abdollahi M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet [Internet]. 2020 Oct 17 [cited 2022 Apr 8];396(10258):1204–22. Available from: <http://ghdx.healthdata.org/gbd->
6. Burden of Disease - Our World in Data [Internet]. [cited 2022 Apr 8]. Available from: <https://ourworldindata.org/burden-of-disease>

7. Ministerio de Salud y Protección Social. Plan decenal de salud pública 2012-2021. Ministerio de salud y protección social [Internet]. 2012;12–538. Available from: <https://www.minsalud.gov.co/salud/publica/epidemiologia/Paginas/cargue-plan-territoriales-de-salud.aspx>
8. Ministerio de Salud y Protección Social. Informe de gestión avance del Plan Decenal de Salud Pública 2012-2021 [Internet]. 2018 [cited 2022 Jun 10]. Available from: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/PSP/informe-gestion-plan-decenal-salud-publica-2018.pdf>
9. Ministerio de Salud y Protección Social. Modelo Integral de Atención en Salud – MIAS Plan de capacitación sectorial. Ministerio De Salud Colombia [Internet]. 2016;27. Available from: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/01-pais-mias-capacitacion-sectorial.pdf>
10. Ministerio de Salud y Protección Social. Manual metodológico para la elaboración de las RIAs. Ministerio de Salud y Protección Social [Internet]. 2016;106. Available from: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/Manual-metodologico-rias.pdf>
11. Peixeiro M. Time Series Forecasting in Python. Mammig; 2021.
12. Vandeput N. Data Science for Supply Chain Forecast. 2018.
13. Luz PM, Mendes BVM, Codeço CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *Am J Trop Med Hyg* [Internet]. 2008 Dec;79(6):933–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19052308>. <https://doi.org/10.4269/ajtmh.2008.79.933>
14. Ruchiraset A, Tantrakarnapa K. Time series modeling of pneumonia admissions and its association with air pollution and climate variables in Chiang Mai Province, Thailand. *Environmental Science and Pollution Research*. 2018 Nov 26;25(33):33277–85. <https://doi.org/10.1007/s11356-018-3284-4>
15. Reichert H, Suh M, Jiang X, Movva N, Bylsma LC, Fryzek JP, et al. Mortality Associated With Respiratory Syncytial Virus, Bronchiolitis, and Influenza Among Infants in the United States: A Birth Cohort Study From 1999 to 2018. *J Infect Dis*. 2022 Aug 15;226(Supplement\_2):S246–54. <https://doi.org/10.1093/infdis/jiac127>
16. Una metodología de series de tiempo para el área de la salud; caso práctico [Internet]. [cited 2022 Jun 10]. Available from: [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0120-386X2007000200014](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-386X2007000200014)
17. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. *Time Series Analysis: Forecasting and Control*, 5th Edition | Wiley. 2021;
18. Box GEP, Jenkins GM. Some Comments on a Paper by Chatfield and Prothero and on A Review by Kendall. *J R Stat Soc Ser A*. 1973;136(3):337. <https://doi.org/10.2307/2344995>
19. Ljung GM, Box GEP. On a measure of lack of fit in time series models. *Biometrika* [Internet]. 1978 Aug 1;65(2):297–303. Available from: <https://doi.org/10.1093/biomet/65.2.297>
20. Box GEP, Pierce DA. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *J Am Stat Assoc* [Internet]. 1970 Jun 4;65(332):1509–26. Available from: <http://www.jstor.org/stable/2284333>. <https://doi.org/10.1080/01621459.1970.10481180>
21. Van R, Drafe F. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace; 2009.
22. Seabold S, Perktold J. statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference. 2010. <https://doi.org/10.25080/Majora-92bf1922-011>
23. McKinney W. Data structures for statistical computing in python. In: 9th Python in Science Conference. 2010. p. 51–6. <https://doi.org/10.25080/Majora-92bf1922-00a>
24. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020 Sep 17;585(7825):357–62. <https://doi.org/10.1038/s41586-020-2649-2>

25. Hunter JD. Matplotlib: A 2D graphics environment. *Computing in science & engineering*. 2007;9(3):90–5. <https://doi.org/10.1109/MCSE.2007.55>
26. Grisales H, Piñeros J, Nieto E. Carta de la enfermedad, costos derivados y geolocalización de la mortalidad, atribuibles a la contaminación por PM2,5 Medellín, 2010 - 2016. Medellín; 2021 May.
27. Romero Cabrera ÁJ. Factores asociados con la mortalidad de ancianos hospitalizados por neumonía adquirida en la comunidad. *Med Int Mex*. 2009;25(5):344–51.
28. Machado-Alba JE, Isaza B, Sepúlveda M. Factores asociados a mortalidad por neumonía en un hospital de tercer nivel de Colombia. *Acta Médica Colombiana*. 2013;38(4):233–8. <https://doi.org/10.36104/amc.2013.156>
29. Quintero-González DC, Ocampo JM, Reyes-Ortiz CA. Factores asociados con la letalidad por neumonía en una unidad de atención del paciente geriátrico agudo: una cohorte retrospectiva. *Biomédica*. 2020 Dec 2;40(4):734–48. <https://doi.org/10.7705/biomedica.5244>
30. Calderón C, Dennis R. Costos económicos de neumonía adquirida en comunidad, meningitis y bacteriemia por *Streptococcus pneumoniae* en una población adulta que requirió hospitalización en Bogotá, Colombia. *Biomédica*. 2013 Oct 7;34(1):92. <https://doi.org/10.7705/biomedica.v34i1.1553>
31. Gómez Muñoz JM, Gómez Rincón JC, Alí Munive A, Cano Gutierrez CA, Coral Alvarado PX, Coronell Rodríguez W, et al. Guías para la inmunización del adolescente y adulto en Colombia. Documento de actualización, 2016. *Infectio*. 2016 Oct;20(4):192–210. <https://doi.org/10.1016/j.infect.2016.08.001>
32. National Heart L and BI. Neumonía - Prevención | NHLBI, NIH [Internet]. [cited 2022 Jun 16]. Available from: <https://www.nhlbi.nih.gov/es/salud/neumonia/prevencion>
33. Neupane B, Jerrett M, Burnett R, Marrie T, Arain A, Loeb M. Long-Term Exposure to Ambient Air Pollution and Risk of Hospitalization with Community-acquired Pneumonia in Older Adults. *American Journal of Respiratory and Critical Care Medicine*. 2010;181:47–53. <https://doi.org/10.1164/rccm.200901-0160OC>
34. Ideam. Informe del estado de la calidad del aire en Colombia 2020. Bogotá, D.C.; 2021.
35. Tasci SS, Kavalci C, Kayipmaz AE. Relationship of Meteorological and Air Pollution Parameters with Pneumonia in Elderly Patients. *Emerg Med Int* [Internet]. 2018 [cited 2022 Jun 19];2018:1–9. Available from: </pmc/articles/PMC5884022/> <https://doi.org/10.1155/2018/4183203>
36. Ruchiraset A, Tantrakarnapa K. Time series modeling of pneumonia admissions and its association with air pollution and climate variables in Chiang Mai Province, Thailand. *Environmental Science and Pollution Research* [Internet]. 2018 Nov 1 [cited 2022 Jun 19];25(33):33277–85. Available from: <https://link.springer.com/article/10.1007/s11356-018-3284-4>. <https://doi.org/10.1007/s11356-018-3284-4>
37. Peabody FW. THE CARBON DIOXIDE CONTENT OF THE BLOOD IN PNEUMONIA. *J Exp Med* [Internet]. 1912 Nov 11 [cited 2022 Jun 19];16(5):701. Available from: </pmc/articles/PMC2124990/?report=abstract>. <https://doi.org/10.1084/jem.16.5.701>