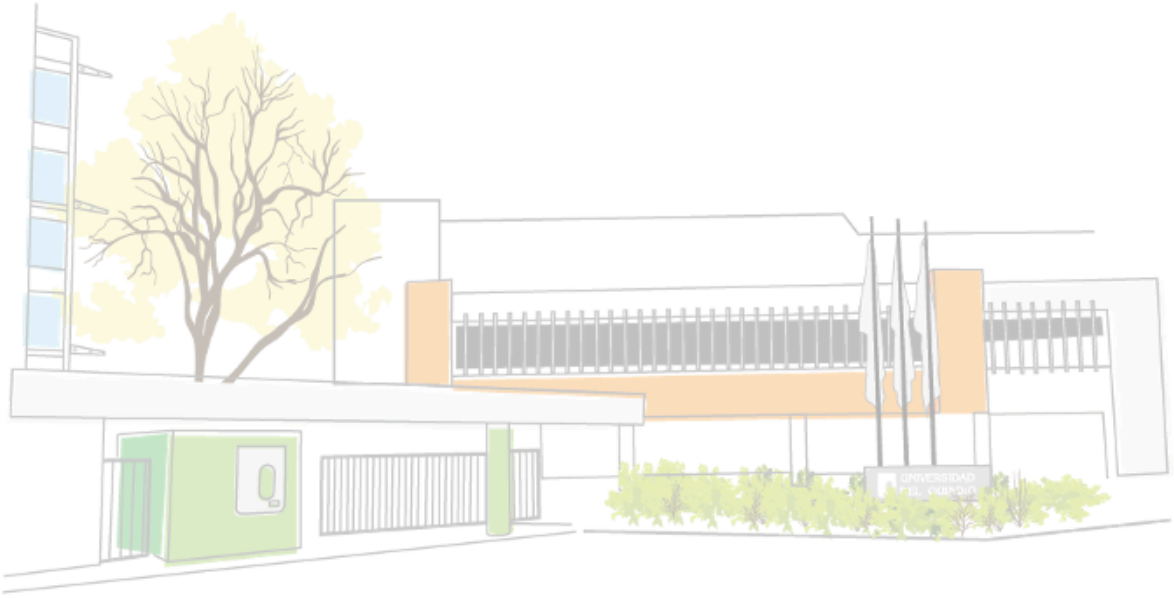
UNIVERSIDAD DEL QUINDÍO



MODELACIÓN Y ANÁLISIS ESTADÍSTICO DE LA SERIE TEMPORAL DE INFECTADOS POR SARS-COV2 EN COLOMBIA

Alexandra Sánchez Arias

Universidad del Quindío

Facultad de Ciencias de la Educación

Licenciatura en Matemáticas

Modelación y análisis estadístico de la serie temporal de infectados por sars-cov2 en colombia

Informe escrito de trabajo de Grado

Alexandra Sánchez Arias

Director: Ph.D Gladys Elena Salcedo Echeverry

Modalidad: Trabajo de investigación

Linea de investigación: Modelos para la información autocorrelacionada en el tiempo y en el espacio

Área de profundización: Estadística

Marzo 2022

Índice

Agradecimientos

Inicialmente, doy gracias infinitas a Dios, quien es la fuente de toda sabiduría, y a quien le debo mi vida y mi vocación no sólo como matemática sino como su futura esposa, además agradezco a toda mi familia, especialmente a mis padres Eliseo Sánchez Rivera y Francy Helena Arias Moreno, quienes han dado toda su vida por mí y me han formado con excelentes valores humanos. También doy gracias a mis hermanas clarisas de Montenegro, que siempre me han animado a culminar mis estudios y me han acompañado con la oración, al igual que mi amigo Robert Álvarez Sánchez, que me ayudó a no desfallecer en el proceso, y a toda mi comunidad universitaria, tanto directivos como docentes, que dejaron una gran huella no sólo en mi mente sino en mi corazón, especialmente la profe Gladys Elena Salcedo Echeverry, de la cual aprendí mucho y a quien le tengo una gran admiración y cariño.

Resumen

El presente trabajo de investigación pretende analizar estadísticamente el comportamiento de la serie temporal del número de infectados diarios por el Coronavirus en Colombia, desde el 06 de marzo del 2020 hasta el 09 de agosto del 2021, evaluando el efecto de las medidas de control de la pandemia, especialmente la vacunación, mediante la identificación de diferentes clústeres de variabilidad y modelación de cada uno de ellos por medio de algunos modelos predictivos para series de tiempo. Se concluye que el modelo más adecuado para ajustar la tendencia de la serie es el cuadrático y un SARIMA para los errores, incluyendo una estacionalidad de 14 días para la mayoría de los casos, excepto para el primero, y el quinto clúster, en cuyo caso la estacionalidad es de 7 días.

Palabras clave: Clasificación K-means, coronavirus, modelos SARIMA, SARS-Cov-2, series de tiempo.

Abstract

This work aims to statistically analyze the behavior of the time series of the number of daily infected by the Coronavirus in Colombia, from March 6, 2020 to August 9, 2021, and to evaluate the effect of the control measures of the pandemic, especially vaccination, through the identification of different clusters and then modeling each of them by means of some predictive model for time series. We conclude that in most cases the time series follows a quadratic trend plus a SARIMA model for the errors. The order of seasonality is 14 days for clusters 2, 3 and 4, and for clusters 1 and 5 is 7 days.

Keywords: K-means classification, coronavirus, ARIMA models, SARS-Cov-2, time series.

1. Introducción

A finales del 2019, la comunidad médica de Wuhan, una provincia de Hubei en China, se alarmó por el descubrimiento de un grupo de 27 casos de lo que en un principio se creyó que era neumonía como lo indica Gutiérrez, Hernández y Puche (2020) y cuya causa fue desconocida hasta que diversos estudios epidemiológicos determinaron que se debía a un nuevo tipo de coronavirus al que llamaron Síndrome Respiratorio Agudo Severo Coronavirus 2 o SARS-Cov-2 por sus siglas en inglés, y a partir de allí el mundo focalizó su atención en este acontecimiento; dicha atención se debió no sólo a su alto nivel de infectabilidad sino a las consecuencias que este virus causó en toda la humanidad, ya que su manera de esparcirse por todo el mundo fue muy veloz y cada vez más creciente, generando multitud de contrariedades económicas y sanitarias, muchas de ellas con consecuencias irreversibles o mortales. La humanidad tuvo que asumir una actitud de defensiva frente a este nuevo enemigo invisible, no obstante, para controlar tanto el virus como sus consecuencias, se debía primero estudiar su comportamiento y características que permitieran identificar los mecanismos a ejecutar, por ello, la comunidad académica a nivel mundial se enfocó en desarrollar análisis investigativos basados en el comportamiento del virus, sin embargo, se debía tener en cuenta lo que plantean Arenas, Cota y otros al aclarar que «las particularidades de las epidemias actuales exigen un replanteamiento de los modelos convencionales hacia modelos a medida» (Arenas, Cota, Gómez-Gardeñes, Gómez, Granell, Matamalas, Soriano, Steinegger, 2020, p.1). Es decir, si se quiere entender el nuevo coronavirus es necesario adaptar algunas teorías ya conocidas a las características especiales que posee dicho virus, generando nuevos modelos para describir este fenómeno.

Gran parte de teoría científica matemática, biológica, médica, etc, se dispuso a analizar los datos que iba dejando el nuevo coronavirus para intentar ponerle freno a su expansión con estrategias adaptadas a las exigencias del virus. No obstante, el proceso no era sencillo, ya que como mencionan Yang y Wang (2020) “El virus es nuevo, y aunque se especula ampliamente que los animales salvajes como murciélagos, civetas y visones son los respon-

sables del inicio de la pandemia, el origen de la infección es incierto" (p.2), así que en principio eran difíciles los estudios porque no se disponía de mucha información. Además, en un informe científico emitido por la Organización Mundial de la Salud (OMS, 2020) el SARS-Cov2 puede transmitirse por contacto directo, indirecto o cercano con personas infectadas a través de cualquier tipo de secreción que posea el virus, adicionalmente, las secreciones pueden contaminar superficies u objetos, creando fómites (superficies contaminadas) que propician la propagación, lo que indica que su peligro aumenta gracias a su fácil manera de esparcirse, es tanto así que, como informan Franco, González y otros:

> "El número de personas que cada persona infectada contagia directamente (R0) es de aproximadamente 2,5, es decir, cada persona infectada, al cabo de un mes, habrá sido responsable del contagio directo de 2,5 personas y del contagio indirecto de 244" (Franco, González-Jaramillo, González-Jaramillo, Gómez-López, Gómez-Restrepo, Palacio, 2020, p.2);

lo que resulta preocupante para la población en general, ya que el crecimiento es muy rápido e incluso incontrolable en lugares donde ocurran aglomeraciones. Ahora bien, teniendo en cuenta lo anterior, la estrategia más directa de lucha contra el nuevo coronavirus es el aislamiento social y la cuarentena obligatoria como medidas de mitigación, sin embargo, dicha estrategia era insostenible por mucho tiempo debido a las implicaciones económicas que genera, especialmente en los países más pobres. En el caso particular de Colombia, según el Departamento Administrativo Nacional de Estadística (DANE, 2020) Colombia tiene una tasa de 48,1 % de empleo informal y de 14.2 % de desempleo, lo que indica que el 62,3 % de los colombianos aptos para laborar no tienen un sustento económico que les permita quedarse encerrados en sus casas.

Por otra parte, cabe mencionar que el sistema de salud en Colombia posee diversas falencias, como lo afirma el doctor Herazo al enunciar que:

> "La salud en Colombia presenta una evaluación negativa, debido a varias situaciones, entre las cuales están: la corrupción generalizada, la injusticia e

> inequidad biológica, social, económica y política, el considerar la prestación de servicios de salud como un negocio especulativo, y la mentalidad curativa y no preventiva de empresarios, dirigentes, directivos y profesionales de la salud" (Herazo, 2010, p. 2);

por lo que resulta muy complejo brindar una amplia y buena asistencia médica en caso de alto contagio por Covid19.

Finalmente, cabe añadir que, como concluyen Gutiérrez, Hernández y Puche:

> "Diversas aproximaciones estadísticas han sido empleadas para la predicción del número de infectados en una epidemia entre los que destacan modelos de regresión lineal multivariada (Du et al., 2020; Ghosal, Sengupta, Majumder, & Sinha, 2020), redes neuronales (Al-Najjar & Al-Rousan, 2020), simulaciones basadas en el modelo SIR (Anastassopou- lou, Russo, Tsakris, & Siettos, 2020; Giordano et al., 2020) e incluso ya se han creado paquetes de software de redes neuronales profundas que pretenden detectar la enfermedad por medio del uso de imágenes de rayos X (Khan, Shah, & Bhat, 2020). Sin embargo, a pesar de la apabullante cantidad de información sobre la predicción de la enfermedad, es difícil obtener resultados confiables debido a que sólo muestran tendencias y una alta aleatoriedad" (Gutiérrez, Hernández y Puche, 2020, p. 3).

Por todo lo anterior, el desarrollo de investigaciones basadas en el análisis de los datos puede ampliar el panorama estudiado hasta el momento y brindar nuevos caminos que permitan el avance investigativo del tema en cuestión, tal como lo afirma Montesinos y Hernández:

> "Estos modelos son muy útiles porque capturan propiedades esenciales de la dispersión de una enfermedad de una forma simplificada. Además, al modificar los parámetros del modelo se pueden representar o descubrir situaciones que difícilmente se pueden obtener mediante experimentación. Por lo tanto, contri-

> buyen a prevenir futuras situaciones patológicas, determinar la prevalencia e incidencia y coadyuvar a tomar decisiones objetivas para el control o supresión de las enfermedades infecciosas" (Montesinos y Hernández, 2007, p.8).

Ahora bien, es evidente que tanto en Colombia como en todos los países afectados por esta pandemia, resulta de gran interés para los entes de control y manejo de la salud pública y para los gobiernos, el conocimiento de la dinámica de las series temporales relacionadas con el número de infecciones, recuperaciones y muertes por el virus, sus asociaciones y la identificación de distintos factores que influyen en el comportamiento de estas series, tal como lo asegura Ocaña (2004) cuando concluye que "los métodos estadísticos son herramientas de gran utilidad para la toma de decisiones en la gestión y política sanitaria". Es por ello que en el presente trabajo de investigación se tiene como objetivo analizar estadísticamente el comportamiento de la serie de infectados diarios por el Coronavirus en Colombia de acuerdo a distintos clústeres de variabilidad, y evaluar el efecto de las medidas de control de la pandemia, especialmente la vacunación mediante una metodología adecuada para el tema en cuestión, que se ejecuta por medio de algunos momentos específicos, como son: la recolección de los datos correspondientes al número de infectados diarios por COVID-19 en Colombia, la selección del número adecuado de clústeres aplicando la clasificación k-means para identificar conglomerados de distinta variabilidad de la serie y para identificar los instantes exactos donde la serie presentó cambios importantes, el ajuste de 1470 modelos predictivos para series de tiempo, la comparación de los modelos con base en diversos aspectos relevantes y la elección del modelo más adecuado por cada clúster. En ese orden de ideas, en el presente estudio investigativo, se encontró que el modelo más adecuado para ajustar la tendencia era el cuadrático, y un SARIMA que modela la autocorrelación de los residuales con una estacionalidad de 14 dias, excepto en el primer y último clúster en donde la estacionalidad es de 7 días debido a una alta transmisión y un menor control, además, al ejecutar cada uno de los modelos, se puede concluir que las combinaciones pueden ser muy amplias y variadas, sin embargo, al momento de elegir el mejor modelo no basta con tener

en cuenta únicamente el menor valor del RMSE o el MAPE, sino que también es importante tener en cuenta la significancia del p-valor, los valores arrojados en los pronósticos, la autocorrelación de los residuales, y entender que como lo expresó el gran George Edward Pelham Box «todos los modelos están mal, pero algunos son muy útiles».

2. Estado del arte

Con respecto al tema en cuestión se encontraron los siguientes antecedentes investigativos:

- Arenas et al. (2020) en su artículo titulado "A mathematical model for the spatiotemporal epidemic spreading of COVID19" tuvieron como objetivo adaptar un modelo de movilidad de metapoblación de enfoque de cadena microscópica de Markov (MMCA) para predecir la incidencia de las epidemias en una población a lo largo del tiempo y así evaluar la efectividad de las medidas de control. Para ello, utilizaron un análisis y ajuste cuantitativo del modelo MMCA introduciendo modificaciones al modelo estándar de metapoblación. Finalmente, obtuvieron como conclusión que el pico de incidencia puede producir un colapso en la capacidad sanitaria para atender a los infectados, especialmente en las unidades de cuidados intensivos.

- Grillo, Santaella, Guerrero y Bravo (2020) en su artículo titulado "Modelos matemáticos y el COVID-19", tuvieron como objetivo analizar los principales elementos usados para la construcción de los modelos a partir de patrones epidemiológicos, para lograr describir la interacción, explicar la dinámica de infección y recuperación, así como para predecir posibles escenarios que pueden presentarse con la introducción de medidas en salud pública como el distanciamiento social y cuarentenas, específicamente para el caso de la pandemia desatada por el nuevo virus SARS-Cov-2/COVID-19. Para ello, hicieron un recorrido teórico por los modelos matemáticos más usados y llegaron a la conclusión de que ante la actual pandemia desatada por la transmisión de SARS-Cov-2, la construcción de modelos matemáticos a partir de patrones epidemiológicos ha permitido describir las interacciones, explicar la dinámica de la infección y recuperación, así como predecir posibles escenarios que pueden presentarse con la introducción de medidas como el distanciamiento social y las cuarentenas. Sin embargo, existen retos importantes en la identificación de casos positivos y de

muertes relacionadas con la infección, datos que son claves en la estimación de tasas y números reproductivos.

- Yang y Wang (2020) en su artículo titulado "A mathematical model for the novel coronavirus epidemic in Wuhan, China", tuvieron como objetivo proponer un modelo matemático para investigar el brote actual de la enfermedad por coronavirus 2019 (COVID_19) en Wuhan, China, que describiera las múltiples vías de transmisión en la dinámica de la infección y enfatizara en el papel del reservorio ambiental en la transmisión y propagación de esta enfermedad. Esta investigación se llevó a cabo en 3 momentos principalmente, en primera instancia se analizó la información existente del virus, luego se presentó el modelo y los supuestos admitidos para establecer el modelo, y por último, se realizó una simulación numérica incorporando los datos de infección reportados para la ciudad de Wuhan. Los autores llegaron a la conclusión de que su modelo puede predecir la aparición de un pico epidémico, pero que si se usan tasas de transmisión el pico que se predice es poco realista y muy alto.

- Gutiérrez, Hernández y Puche (2020) en su artículo titulado "Estimación de casos de COVID-19 en países de Suramérica empleando modelos ARIMA", tuvieron como objetivo emplear modelos ARIMA (Autorregresivo Integrado de Promedio Móvil), para la estimación de nuevos contagios usando datos públicos disponibles de Venezuela y la región suramericana. Por ello, utilizaron una metodología econométrica basada en modelos dinámicos que utiliza datos de series temporales con el fin de describir la evolución temporal como una función lineal de datos previos y errores debidos al azar. El método utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Los análisis fueron realizados empleando el lenguaje de programación estadístico R (R Core Team, 2020) (versión 3.6.3), en conjunto con el entorno interactivo de desarrollo R Studio (RStudio Team, 2020). Finalmente concluyeron que la estimación para el total acumulado de casos en la región suramericana, tiene en su mayoría un claro crecimiento exponencial, además

la implementación de los modelos ARIMA es una aproximación matemática que permite hacer estimaciones de la dinámica de las enfermedades infecciosas en función del tiempo, debido a su fácil estructura y rápida aplicabilidad.

- Córdova y Santa (2021) en su artículo titulado "Aplicación del método autorregresivo integrado de medias móviles para el análisis de series de casos COVID-19 en Perú", tuvieron como objetivo estimar un modelo Autorregresivo Integrado de Medias Móviles (ARIMA) para el análisis de series de casos de COVID-19, en Perú. El método que utilizaron se basó en un análisis de series temporales univariante; los datos utilizados se refieren a la cantidad de casos nuevos acumulados de COVID-19 del 06 de marzo al 11 de junio de 2020. Para el análisis del ajuste del modelo se utilizaron los coeficientes de autocorrelación (ACF), el contraste de raíces unitarias de Dickey-Fuller Aumentado (ADF), el Criterio de Información Bayesiano Normalizado (BIC Normalizado), el error porcentual medio absoluto (MAPE) y el test de Box-Ljung. Finalmente, llegaron a la conclusión de que los resultados obtenidos con el modelo ARIMA, comparados con los datos observados, muestran un ajuste adecuado de los valores; y aunque este modelo, de fácil aplicación e interpretación, no simula el comportamiento exacto en el tiempo puede considerarse una herramienta simple e inmediata para aproximar el número de casos.

- Córdova y Santa (2021) en su artículo titulado "Precisión del pronóstico de la dinámica de propagación de la COVID-19 en Perú", tuvieron como objetivo analizar la precisión del pronóstico del modelo suavizado de Brown para predecir la propagación de la COVID-19 en Perú, entre el 6 de marzo al 30 de mayo del 2020. Para ello utilizaron el modelo suavizado de Brown que consiste en realizar dos suavizaciones exponenciales, a partir de las cuales se calcula el pronóstico: en la primera se emplean los valores observados en la serie de tiempo; y la segunda, la serie que ha sido obtenida mediante la primera atenuación. Las medidas de precisión utilizadas fueron el error medio del pronóstico (EMP), el error cuadrático medio (ECM), la desviación absoluta de la

media (DAM) y el error porcentual absoluto medio (EPMA). Para determinar si los datos se ajustan al modelo evaluado se utilizó el coeficiente de determinación (R2). Finalmente, concluyeron que el error de precisión o error porcentual absoluto medio (EPAM) fue del 9,03 %, con un coeficiente de determinación (R2) de 0,8078; lo que indica que los datos se ajustan en un 80,78 % al modelo evaluado.

- Wilches y Castillo (2020) en su artículo titulado "Aproximación matemática del modelo epidemiológico SIR para la comprensión de las medidas de contención contra la COVID-19", tuvo como objetivo desarrollar el modelo SIR y su aplicación para predecir el curso de la pandemia por Covid-19 en la ciudad de Santa Marta (Colombia), a fin de comprender la razón que subyacía a varias de las medidas de contención adoptadas por los Estados del mundo en la lucha contra la pandemia, para ello se propuso deducir matemáticamente las ecuaciones del modelo SIR, aplicarlas para simular el recorrido de la pandemia por Covid-19 en la ciudad colombiana de Santa Marta y reflexionar acerca de cómo los factores de control epidemiológicos presentes en las ecuaciones son claves para comprender y explicar las medidas de contención empleadas en la lucha contra la Covid-19. Finalmente, concluyó que el modelo epidemiológico SIR es un abordaje apropiado para comprender los mecanismos de lucha contra la pandemia, además, los resultados mostraron que el modelo B garantiza una menor cantidad de infectados totales y diarios, relacionados con una distribución de infectados más homogénea, con lo cual la probabilidad de colapso hospitalario es menor. Además, una perspectiva futura de ese trabajo sería considerar la contribución explícita de los fallecidos dentro del modelo.

- Franco et al. (2020) en su artículo titulado "Proyecciones de impacto de la pandemia COVID-19 en la población colombiana, según medidas de mitigación. Datos preliminares de modelos epidemiológicos para el período del 18 de marzo al 18 de abril de 2020", tuvieron como primer objetivo: realizar predicciones del curso de la infección en el horizonte temporal desde marzo 18 a abril 18 del 2020, según diferentes medidas

de aislamiento aplicadas. Segundo objetivo: modelar la mortalidad y la necesidad de recursos hospitalarios, estratificando por edad el escenario de contagio del 70 % de la población. Para el primer objetivo, se basaron en el número de casos confirmados en el país hasta marzo 18, 2020 (n=93). Como suposiciones para el modelo, se incluyó un índice de contagio R0=2,5 y el índice de casos reales por cada caso confirmado. Para la proporción de pacientes que necesitarían cuidados intensivos u otros cuidados intrahospitalarios, se basaron en datos aportados por el Imperial College of London. Para el segundo objetivo usaron como tasa de mortalidad por edad, datos aportados por el Instituto Superiore di Sanità en Italia. Finalmente, concluyeron que las medidas de mitigación que han sido implementadas hasta la fecha por el gobierno colombiano se fundamentan en evidencia suficiente para pensar que es posible reducir significativamente el número de casos contagiados y con esto, el número de pacientes que requerirán manejo hospitalario.

- Diaz (2020) en su artículo titulado "Precisión del pronóstico de la propagación del COVID-19 en Colombia", tuvo como objetivo presentar la precisión de un pronóstico de la dinámica de transmisión del COVID-19 en Colombia, para ello utilizó la base de datos de las personas infectadas con el Covid-19, esta información corresponde al período 6 de marzo al 14 de abril de 2020. Para su análisis de predicción manejó el método modelo de Brown, utilizando el paquete estadístico SPSS v.25. Finalmente, llegó a la conclusión de que el uso de modelación matemática se ha desarrollado en grado representativo en las últimas décadas y son de gran impulso para ilustrar escenarios eficaces de prevención y control de enfermedades infectocontagiosas.

- Diaz (2020) en su artículo titulado "Perspectiva del COVID-19 en Colombia para el año 2021", tuvo como objetivo presentar una perspectiva del contagio de personas recuperadas y fallecidas por el COVID-19 en Colombia para 2021; el método utilizado para calcular el pronóstico fue un modelo ARIMA, teniendo en cuenta el registro de la información por parte del Instituto Nacional de Salud, hasta el 6 de noviembre

2020 y considerando el punto de vista económico, social y de salud. El trabajo de investigación se realizó mediante un enfoque cuantitativo y llegó a la conclusión de que de acuerdo con la proyección del COVID-19 para 2021 en Colombia por el método ARIMA(0,1,0), se estimó que para finalizar el mes de diciembre se tendrá un número de 4'973.547 personas contagiadas, 4'784.987 personas recuperadas, y 110.159 de personas fallecidas. El autor recomienda utilizar otros modelos de pronóstico, por ejemplo, el modelo de suavizado exponencial de Brown y afirma que la utilización de modelación matemática ha progresado en grado representativo en las últimas décadas y son de gran impulso para ilustrar escenarios eficaces de prevención y control de enfermedades infectocontagiosas, esto con la finalidad de seguir monitoreando el SARS-Cov-2, y poder controlar su velocidad de propagación en Colombia.

3. Marco conceptual

En este capítulo se presenta un poco de teoría básica sobre la regresión lineal y la estimación de los parámetros, lo cual servirá de fundamento para el desarrollo de la estimación del modelo de regresión funcional.

3.1. El modelo clásico de regresión lineal

Cuando se desea conocer la relación que existe entre una variable dependiente y otra o varias explicativas o independientes, desde un enfoque estadístico, una familia de modelos muy utilizada en la práctica es aquella conformada por los modelos de regresión. El término regresión fue introducido por primera vez a finales del siglo XIX por el ilustre polímetra Sir Francis Galton en su libro "natural inheritance" en el año 1889, en donde registró la realización de estudios con el interés de predecir la estatura de los hijos a partir de la estatura de los padres y después de reunir alturas de padres e hijos concluyó que en general, padres altos tenían hijos altos y padres bajos tenían hijos bajos como si existiera un *retroceso* o *regresión* en las estaturas (Galton, 1889); desde entonces se utiliza el término regresión para relacionar variables estadísticamente.

El modelo de regresión más sencillo y aplicado en la práctica es el modelo de regresión lineal simple, el cual relaciona una variable dependiente Y con una variable independiente o explicativa X, mediante la expresión lineal

$$Y = b_0 + b_1 X + e,$$

donde b_0 y b_1 son los parámetros del modelo y e representa el error aleatorio, el cual en la práctica corresponde a factores causales no incluídos en el modelo, errores de medida, perturbaciones aleatorias, fuentes de no linealidad, entre otros.

En la práctica se dispone de valores observados de las variables X y Y para n individuos, o sea se tiene un conjunto de valores $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ a partir de los cuales se deben estimar los parámetros del modelo. Un método común de estimación es el

método de los mínimos cuadrados, el cual es muy amigable porque se apoya en técnicas del cálculo diferencial. Sin embargo, dependiendo de los supuestos probabilísticos que se hagan sobre el error aleatorio e, se pueden aplicar otros métodos de estimación como la máxima verosimilitud o métodos bayesianos por ejemplo. Un supuesto muy común y que facilita el procedimiento de estimación es que tenga una distribución Normal de media cero y varianza σ^2; esto es, $e \sim N(0, \sigma^2)$, en cuyo caso la varianza del error es otro parámetro a estimar.

El modelo de regresión se denomina *múltiple* si se relaciona la variable dependiente Y con dos o más variables explicativas $X_1, X_2, ..., X_p$ en la forma

$$Y = b_0 + b_1 X_1 + b_2 X_2 + ... + b_p X_p + e,$$

donde de nuevo $b_0, b_1, b_2, ..., b_p$ son los parámetros del modelo junto con σ^2, la varianza del error aleatorio e. En su forma simple o múltiple, el carácter del modelo de regresión pasa a ser no lineal cuando existen relaciones arbitrarias entre las variables dependientes e independientes, llevando a cabo algoritmos de estimación iterativos.

Considere ahora la situación en que se observa o se registra una característica particular Y a un individuo, en distintos instantes $t_1, t_2, \ldots, t_n$ en el tiempo, el vector de observaciones es del tipo $\{(t_1, y_1), (t_2, y_2), ..., (t_n, y_n)\}$. Un conjunto de observaciones de este tipo se conoce como una serie de tiempo, una señal o un dato funcional, y por depender del *tiempo* o por tratarse de medidas repetidas sobre un mismo individuo, pueden acarrear una autocorrelación importante la cual debe ser considerada en el modelo. Para esta situación se presentan más adelante algunos modelos lineales para series de tiempo. Sin embargo, cuando la serie temporal no es autocorrelacionada pero presenta algún tipo de tendencia dependiente de t, un modelo de regresión como los anteriores resulta apropiado.

Por otra parte, cuando se tienen observaciones seriales o registros en el tiempo de las variables X y Y para n individuos, se habla de datos funcionales, en cuyo caso el modelo de regresión toma la denominación funcional ya que tiene la forma (para el caso más simple)

$$Y(t) = b_0(t) + b_1(t)X(t) + e(t), t \in \tau \subset \mathbb{R};$$

note que en este caso los parámetros no son de tipo escalar sino que son funciones dependientes de t.

Lo que se pretende en este trabajo es comparar algunos modelos lineales y no lineales para analizar e interpretar el comportamiento de la serie temporal del número de infectados diarios y la tasa de mortalidad diaria por el coronavirus en Colombia. Para ello, inicialmente se desarrolla en detalle la teoría y estimación del modelo de regresión lineal clásico, y se presentan algunos modelos lineales clásicos para series de tiempo y otros de caracter funcional, entre los cuales se selecciona el mejor ajuste a las series de interés.Como se mencionó anteriormente, suponga que se tienen n observaciones de las variables X y Y para n individuos, es decir, se tiene el conjunto de pares ordenados

$$\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}.$$

Si dichos puntos se representan en un plano cartesiano y ellos siguen una tendencia lineal, interesa encontrar la ecuación de la recta que más se aproxima a todo el conjunto de puntos, la cual se denomina la recta de regresión tal como se muestra a continuación

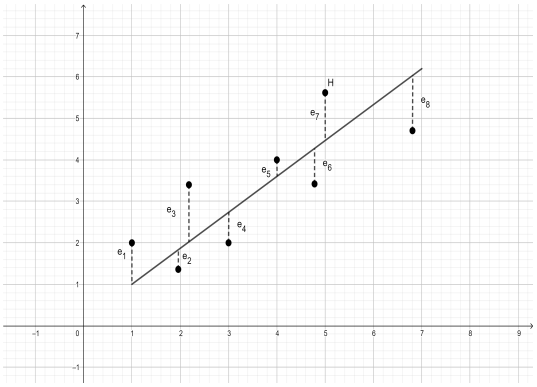


Figura 1: Recta de regresión

El modelo de regresión lineal simple está dado por la ecuación

$$Y = b_0 + b_1 X + e, \tag{1}$$

donde

- Y es la variable dependiente o variable de respuesta.
- X es la variable independiente o variable explicativa.
- b_0 y b_1 son dos valores poblacionales desconocidos pero que se estiman a través del conjunto de observaciones, y representan el intercepto y la pendiente de la recta, respectivamente.
- e es una variable aleatoria que representa el error del modelo. (Galvis, García, Hurtado y Salcedo, 2006, p.140)

La información obtenida para los n individuos aparece generalmente organizada como se muestra en el Cuadro 1

Individuo	Y	X
1	y_1	x_1
2	y_2	x_2
$\vdots$	$\vdots$	$\vdots$
n	y_n	x_n

Cuadro 1: Observaciones de las variables X y Y para n individuos.

Ahora bien, aplicando el modelo (1) para cada punto observado $(x_i, y_i),\ i = 1, 2, \ldots, n$, se genera el siguiente sistema de ecuaciones

$$\begin{aligned} y_1 &= b_0 + b_1 x_1 + e_1 \\ y_2 &= b_0 + b_1 x_2 + e_2 \\ y_3 &= b_0 + b_1 x_3 + e_3 \\ &\vdots \\ y_n &= b_0 + b_1 x_n + e_n, \end{aligned} \tag{2}$$

donde cada e_i representa la distancia del punto observado (x_i, y_i) a la recta de regresión $Y = b_0 + b_1 X$, y los n términos $e_1, e_2, \ldots, e_n$ se denominan los errores aleatorios del modelo. Un supuesto sobre estos errores es que sean no correlacionados entre sí y distribuídos normalmente con media cero y varianza constante σ^2 (Galvis, García, Hurtado y Salcedo, 2006, p.139). Un buen ajuste es aquel para el cual estos errores son mínimos, la cual es la idea del método de estimación por mínimos cuadrados y que se desarrolla a continuación.

3.2. Estimación por mínimos cuadrados ordinarios

Dado un conjunto de obervaciones (x_i, y_i) con $i = 1, 2, \ldots, n$ y considerando el modelo (1), los errores aleatorios satisfacen la ecuación

$$e_i = y_i - b_0 - b_1 x_i, \; i = 1, 2, \cdots, n.$$

Una función objetivo para estimar b_0 y b_1 es la suma del cuadrado de todos los errores

$$Q(b_0, b_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 . \tag{3}$$

Se trata entonces de encontrar expresiones para b_0 y b_1 de tal manera que (3) alcance el mínimo, por lo que se debe simplificar la expresión (3) y derivar parcialmente a Q con respecto a b_0 y b_1, generándose el sistema de ecuaciones 2×2 como se presenta a continuación.

$$
\begin{aligned}
Q(b_0, b_1) &= \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 \\
&= \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2 \\
&= \sum_{i=1}^{n} \left(y_i^2 - 2y_i(b_0 + b_1 x_i) + (b_0 + b_1 x_i)^2 \right) \\
&= \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} (y_i(b_0 + b_1 x_i)) + \sum_{i=1}^{n} \left(b_0^2 + 2b_0 b_1 x_i + b_1^2 x_i^2 \right) \\
&= \sum_{i=1}^{n} y_i^2 - 2b_0 \sum_{i=1}^{n} y_i - 2b_1 \sum_{i=1}^{n} x_i y_i + n b_0^2 + 2b_0 b_1 \sum_{i=1}^{n} x_i + b_1^2 \sum_{i=1}^{n} x_i^2.
\end{aligned}
$$

Las derivadas parciales son

$$
\begin{cases}
\frac{\partial Q}{\partial b_0} &= -2 \sum_{i=1}^{n} y_i + 2nb_0 + 2b_1 \sum_{i=1}^{n} x_i \\
\frac{\partial Q}{\partial b_1} &= -2 \sum_{i=1}^{n} x_i y_i + 2b_0 \sum_{i=1}^{n} x_i + 2b_1 \sum_{i=1}^{n} x_i^2;
\end{cases}
$$

así que igualando ambas derivadas a cero se obtiene

$$
\begin{aligned}
\frac{\partial Q}{\partial b_0} &= -2 \sum_{i=1}^{n} y_i + 2nb_0 + 2b_1 \sum_{i=1}^{n} x_i = 0 \\
b_0 &= \frac{1}{n} \left(\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i \right) \\
b_0 &= \bar{y} - b_1 \bar{x}
\end{aligned}
$$

y

$$\begin{aligned}
\frac{\partial Q}{\partial b_1} = -2\sum_{i=1}^{n} x_i y_i + 2b_0 \sum_{i=1}^{n} x_i + 2b_1 \sum_{i=1}^{n} x_i^2 &= 0 \\
-2\sum_{i=1}^{n} x_i y_i + 2\left(\frac{1}{n}\left(\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i\right)\right)\sum_{i=1}^{n} x_i + 2b_1 \sum_{i=1}^{n} x_i^2 &= 0 \\
-2\sum_{i=1}^{n} x_i y_i + 2\frac{1}{n}\left(\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i - b_1\left(\sum_{i=1}^{n} x_i\right)^2\right) + 2b_1 \sum_{i=1}^{n} x_i^2 &= 0 \\
-2\sum_{i=1}^{n} x_i y_i + 2\frac{1}{n}\left(\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i\right) + 2b_1\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right) &= 0,
\end{aligned}$$

de donde

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}.$$

Por lo tanto, los estimadores obtenidos por mínimos cuadrados ordinarios son

$$\begin{cases}
\hat{b}_0 &= \bar{y} - b_1 \bar{x} \\
\hat{b}_1 &= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}.
\end{cases}$$

(Galvis, García, Hurtado y Salcedo, 2006, p.142)

3.2.1. El enfoque matricial del procedimiento de estimación

El sistema de ecuaciones (2) se puede expresar matricialmente como

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}, \tag{4}$$

o también como

$$Y = X\beta + \varepsilon, \tag{5}$$

en donde, $Y = [y_1, y_2, \ldots, y_n]'$, $\beta = [b_o, b_1]'$, $\varepsilon = [e_1, e_2, \ldots, e_n]'$. Ahora bien, de (5) se puede despejar ε de la siguiente manera

$$\varepsilon = Y - X\beta, \tag{6}$$

y al multiplicar (6) por la izquierda por su transpuesto y realizar las operaciones pertinentes se obtiene de nuevo la suma de los cuadrados de los errores,

$$\begin{aligned} \mathbf{Q}(\beta) = \varepsilon^T\varepsilon = \sum_{i=1}^{n} e_i^2 &= (Y - X\beta)^T(Y - X\beta) \\ &= \left[Y^T - (X\beta)^T\right][Y - X\beta] \\ &= \left[Y^T - \beta^T X^T\right][Y - X\beta] \\ &= Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta \\ &= Y^TY - 2(\beta^TX^TY) + \beta^TX^TX\beta \end{aligned} \tag{7}$$

Luego, para obtener el β que mínimice (7) se deriva parcialmente a $\mathbf{Q}(\beta)$ con respecto a β y se iguala a cero, de la siguiente forma

$$\begin{aligned} \frac{\partial \mathbf{Q}(\beta)}{\partial \beta} = -2X^TY + 2X^TX\beta &= 0 \\ X^TX\beta &= X^TY, \end{aligned}$$

lo que nos conduce a la ecuación conocida como sistema de ecuaciones normales, en donde la solución existe si $det\,(X^TX) \neq 0$, en cuyo caso se puede multiplicar por la izquierda de la ecuación anterior por $(X^TX)^{-1}$,

$$\begin{aligned}(X^TX)^{-1}(X^TX)\beta &= (X^TX)^{-1}(X^TY)\\ I\beta &= (X^TX)^{-1}(X^TY)\\ \hat{\beta} &= (X^TX)^{-1}(X^TY).\end{aligned}$$

Más específicamente, $\hat{\beta} = (\hat{b}_0, \hat{b}_1)^T$.

Este procedimiento es análogo para estimar por mínimos cuadrados los parámetros del modelo de regresión múltiple, solamente cambian las especificaciones de los vectores Y, β, ε y de la matriz de diseño X (Galvis, García, Hurtado y Salcedo, 2006, p.147).

3.3. Selección y comparación de modelos

Volviendo al modelo (1) y a su representación gráfica dada en la Figura 1, queda entendido que cuando se habla del modelo de regresión lineal simple generalmente se piensa en el ajuste de una recta para la modelación de la tendencia del conjunto de observaciones. Sin embargo, no siempre dicha tendencia sigue una línea recta sino algún otro tipo de curva dependiendo de la función que representa la relación entre las variables X y Y; por ejemplo, una función cuadrática, logarítmica, exponencial, radical, hiperbólica, etc. En este caso el modelo de regresión es dado por expresiones como las siguientes, respectivamente:

$$
\begin{aligned}
Y &= b_0 + b_1 X^2 + e, \\
Y &= b_0 + b_1 ln(X) + e, \\
ln(Y) &= b_0 + b_1 X + e, \\
Y &= b_0 + b_1 \sqrt{X} + e, \\
Y &= b_0 + b_1 \frac{1}{X} + e.
\end{aligned}
$$

Observe entonces que en casos como los anteriores, la linealidad del modelo de regresión se refiere a la linealidad de sus parámetros y el procedimiento de estimación por mínimos cuadrados ordinarios es esencialmente el mismo, excepto que las observaciones de la Tabla 1 o de las fórmulas de $\hat{b}_0$ y $\hat{b}_1$ se cambian previamente por transformaciones de las observaciones de acuerdo a la función especificada en el modelo.

Existen diversos criterios para comparar distintos modelos ajustados a un conjunto de observaciones y seleccionar el mejor ajuste. La mayoría de estos criterios están basados en el hecho deseable que los errores e_i sean lo suficientemente pequeños, lo cual está directamente relacionado no solamente con el supuesto que su valor esperado sea cero sino además que su varianza σ^2 sea pequeña. Dos funciones de riesgo basadas en la variabilidad de los errores son el error cuadrático medio (MSE del inglés *mean squared error*) o el error absoluto medio (MAE del inglés *mean absolute error*), de tal forma que se considera mejor modelo aquel para el cual estas medidas de riesgo sean mínimas;

$$
\begin{aligned}
\text{MSE} &= \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n} \\
\text{MAE} &= \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}.
\end{aligned}
$$

Otra medida de selección del modelo es el error porcentual absoluto medio (MAPE del inglés *mean absolute porcentual error*) dado por la expresión

$$\text{MAPE} = \frac{100}{n} \times \sum_{i=1}^{n} |\frac{\hat{y}_i - y_i}{y_i}|.$$

(Hurtado y Salcedo, 1999, p.71).

3.4. Series de tiempo

Una serie de tiempo es un conjunto de observaciones $\{X_t,\ t \in \tau\}$ donde cada observación es registrada en un punto t sobre un conjunto dado τ, que geralmente representa el tiempo, pero también podría representar el plano, el espacio, etc. Cuando las observaciones son hechas sobre un conjunto de puntos discretos se le llama una serie discreta y es el caso más común de las series que aparecen en los problemas prácticos; si las observaciones son hechas en tiempos continuos diremos que la serie es continua.

Una definición más formal de las series de tiempo se encuentra enmarcada dentro de la teoría de los procesos estocásticos, y se menciona brevemente a continuación.

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio de probabilidad y sea τ un conjunto de índices; un ***proceso estocástico*** *es una función $X(t, \omega)$, definida en $\tau \times \Omega$, tal que para cada t fijo $X(t, \omega)$ es una variable aleatoria sobre $(\Omega, \mathcal{A}, \mathcal{P})$ y para cada ω fijo, $X(t, \omega)$ es una trayectoria, realización, función muestral del proceso o una serie de tiempo.*

De aquí en adelante se suprime la variable ω y se representará la serie $X(t, \omega)$ simplemente como X_t, asumiendo que es discreta.

Previo a la modelación de una serie de tiempo es importante evaluar la presencia de características particulares como estacionariedad, estacionalidad, antependencia, larga memoria, comportamiento caótico y no linealidades, entre otras. Algunas de estas características se discuten a continuación.

3.4.1. Estacionariedad de una serie de tiempo

La no estacionariedad de una serie de tiempo tiene que ver con cambios a través del tiempo, bien sea en las distribuciones de probabilidad conjunta de las variables aleatorias del proceso estocástico o en las estructuras de segundo orden.

El proceso $\{X_t,\, t \in \tau\}$ se dice que es **estrictamente o fuertemente estacionario** si las distribuciones de probabilidad conjuntas de $(X_{t_1},\, X_{t_2}, ...,\, X_{t_k})'$ y de $(X_{t_1+h},\, X_{t_2+h}, ...,\, X_{t_k+h})'$ son las mismas para todo $t_1,\, t_2,\, ...,\, t_k,\, h \in \tau$, o sea, son invariantes a translaciones en el tiempo. Esto es,

$$F(x_{t_1},\, x_{t_2}, ...,\, x_{t_k}) = F(x_{t_1+h},\, x_{t_2+h}, ...,\, x_{t_k+h}),\ \forall\ t_1,\, t_2, ...,\, t_k,\, h \in \tau.$$

Dado el proceso $\{X_t,\, t \in \tau\}$ se definen

- la **función media** del proceso por

$$\mu_t = E(X_t) = \int_{-\infty}^{\infty} X_t\, dF(x_t),$$

- la **función varianza** del proceso por

$$\sigma_t^2 = E(X_t - \mu_t)^2 = \int_{-\infty}^{\infty} (X_t - \mu)^2\, dF(x_t),$$

- y la **función de autocovarianza** entre las variables X_{t_1} y X_{t_2} por

$$\gamma_x(t_1, t_2) = E(X_{t_1} - \mu_1)(X_{t_2} - \mu_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (X_{t_1} - \mu_1)(X_{t_2} - \mu_2)\, dF(x_{t_1},\, x_{t_2});$$

las cuales existen si las integrales de Riemann-Stieltjes son finitas.

Cuando un proceso $\{X_t,\ t \in \tau\}$ es estrictamente estacionario con $EX_t^2 < \infty$, se tiene en particular que $F_{X_{t_k}}(x)$ permanece invariante para todo k y

$$F_{X_{t_1},\, X_{t_2}}(x_{t_1},\ x_{t_2}) = F_{X_{t_1+k},\, X_{t_2+k}}(x_{t_1+k},\ x_{t_2+k})$$

implicando que

i) $E(X_t) = \mu_t = \mu, \quad \forall\, t \in \tau$;

ii) $Var(X_t) = \sigma^2 = \sigma_x^2, \quad \forall\, t \in \tau$;

iii) $\gamma_x(t_i, t_j) = \gamma_x(|t_i - t_j|) = \gamma_x(k)$, para cualquier $t_i, t_j, k \in \tau$. Esto es, la covarianza entre las dos variables X_{t_i} y X_{t_j} sólo depende de su separación $|t_i - t_j| = k$, definiéndose así la función de autocovarianza $\gamma_x(k),\ k \in \mathbb{Z}$.

Note que $\gamma_x(0) = \sigma_x^2$ y $\gamma_x(k) = \gamma_x(-k)$, es decir, la función de autocovarianza tiene su máximo en cero y es una función simétrica alrededor de cero.

Un proceso $\{X_t,\, t \in \tau\}$ que satisface las propiedades (i), (ii) y (iii) se dice que es **débilmente estacionario** o **estacionario de segundo orden**.

Por otra parte, un proceso $\{X_t,\, t \in \tau\}$ es **Gaussiano** si y solamente si las funciones de distribución de $\{X_{t_1}, X_{t_2}, ..., X_{t_k}\}$, para cualquier $k \in \mathbb{N}$, son todas normales multivariadas.

Note que un proceso fuertemente estacionario es débilmente estacionario; sin embargo, no todo proceso débilmente estacionario es fuertemente estacionario, a menos que este sea Gaussiano.

La dependencia que una serie de tiempo tiene de su pasado, es decir su antedependencia, se describe mediante las estructuras de autocorrelación. (Hurtado y Salcedo, 1999, p.4)

3.4.2. Autocorrelación simple y parcial

La **función de autocorrelación simple** del proceso $\{X_t, t \in \tau\}$ a rezago k se define mediante la función

$$\rho_x(k) = Corr(X_t,\ X_{t+k}) = \frac{\gamma_x(k)}{\gamma_x(0)}.$$

Tres propiedades básicas de la función de autocorrelación simple son las siguientes:

i) $\rho_x(0) = 1$.

ii) $|\rho_x(k)| \leq 1, \forall k = 0, 1, 2, \ldots$

iii) $\rho_x(-k) = \rho_x(k), \ \forall k = 0, 1, 2, \ldots$

Como puede observarse las funciones de autocovarianza y autocorrelación miden la relación lineal entre las variables X_t y X_{t+k} del proceso. Otra manera de medir esta relación, pero eliminando el efecto de las $k-1$ variables intermedias, es a través de la función de autocorrelación parcial. Por la forma como está definida, se dice que la función de autocorrelación parcial muestra una correlación más pura o más neta entre las variables X_t y X_{t+k}.

Condidere los errores e_t y e_{t+k} resultantes de los siguientes modelos de regresión lineal

$$X_t = \alpha_1 X_{t+1} + \alpha_2 X_{t+2} + \ldots + \alpha_{k-1} X_{t+k-1} + e_t$$

y

$$X_{t+k} = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \ldots + \beta_{k-1} X_{t+k-1} + e_{t+k}.$$

(Hurtado y Salcedo, 1999, p.17)

La **función de autocorrelación parcial** entre X_t y X_{t+k} corresponde a la función de correlación entre los errores e_t y e_{t+k} dada por

$$\phi_x(k) = Corr(e_t, \ e_{t+k}).$$

Cuando ambas funciones de autocorrelación, la simple y la parcial, son nulas para todo $k \neq 0$ se dice que el proceso es **sin memoria** o es un ruido blanco.

Una sucesión de variables aleatorias $\{a_t, \ t \in \tau\}$ es un proceso **ruido blanco** de media cero si

i) $E(a_t) = 0, \quad \forall t \in \tau$

ii) $\gamma(k) = \begin{cases} \sigma_a^2 & \text{para } k = 0 \\ 0 & \text{para } k \neq 0, \end{cases}$

es decir, su esperanza es cero, su varianza es constante y las $\{a_t,\ t = 1, 2, ...\}$, son variables aleatorias no correlacionadas (Hurtado y Salcedo, 1999, p.20).

Otra característica importante de una serie temporal es su comportamiento lineal o no lineal. A continuación se desarrollan algunos modelos muy utilizados en la práctica.

3.4.3. Algunos modelos para series de tiempo

Existen varias familias de modelos para series de tiempo tanto de carácter lineal como no lineal; desde los mismos modelos de errores o de regresión, otros denominados estructurales, los conocidos modelos ARIMA y ARFIMA, modelos con heteroscedasticidad condicional, o modelos con parámetros cambiando en el tiempo. Enseguida se presentan brevemente algunos de estos modelos.

3.5. Modelos de errores o de regresión

Suponga que la serie temporal observada $\{X_t,\ t = 1, 2, \ldots, T\}$ se puede descomponer en la forma

$$X_t = f_t + a_t,\ t = 1, 2, \ldots, T,$$

donde el término f_t se denomina señal y a_t es el ruido. Si f_t es una función determinística, el modelo se denomina de regresión y las observaciones son no autocorrelacionadas. Los casos más simples son el modelo de media constante más ruido, $X_t = M + a_t$, o el modelo de tendencia lineal $X_t = \alpha + \beta t + a_t = T_t + a_t$, donde la función T_t describe tendencia de la serie y es lineal; esta tendencia también podría ser de tipo polinómica, exponencial, logarítmica, etc.

En el análisis clásico de las series de tiempo cuando

$$X_t = T_t + S_t + C_t + a_t,$$

se dice que la serie puede descomponerse en tres términos determinísticos denominados tendencia, estacionalidad y ciclo, más un ruido aleatorio (Morettin y Toloi, 2006). Sin

embargo, dependiendo de la **estructura** de la misma serie, resulta más realístico incluir o considerar términos con naturaleza variable.

3.6. Modelos estructurales

Los modelos estructurales o de tipo espacio de estados consideran la variabilidad de algunos de los términos determinísticos en la descomposición anterior y son muy usados en la práctica. Algunos de estos modelos son:

3.6.1. Modelo de media local

En este caso la media cambia con t de acuerdo a un paseo aleatorio. Más específicamente,

$$X_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \eta_t,$$

los ruidos $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$ y son independientes entre sí. En este modelo la estimación de μ_t es una media móvil de las observaciones anteriores con una constante de suavizamiento que es función de la razón señal a ruido, $r = \frac{\sigma_\eta^2}{\sigma_\epsilon^2}$.

3.6.2. Modelo de tendencia local

Es un modelo muy utilizado en la práctica, en el cual el nivel de la serie cambia con t y además presenta inclinación (o "drift") que también varía con t de acuerdo a un paseo aleatorio. El modelo es dado por las ecuaciones

$$\begin{aligned} Z_t &= \mu_t + \epsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \xi_t, \end{aligned}$$

donde $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$ y $\xi_t \sim N(0, \sigma_\xi^2)$ y son independientes entre sí; la representación de este modelo en la forma espacio de estados es

$$Z_t = [1 \quad 0] \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \epsilon_t \tag{8}$$

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_{t-1} \\ \xi_{t-1} \end{bmatrix}. \tag{9}$$

El pronóstico es una recta con nivel e inclinación y la fuerza con que μ_t y β_t cambian con t depende de las razones $r_1 = \frac{\sigma_\eta^2}{\sigma_\epsilon^2}$ y $r_2 = \frac{\sigma_\xi^2}{\sigma_\epsilon^2}$. (Morettin y Toloi, 2006).

3.6.3. Modelo de tendencia local y componente estacional

Se trata del modelo de tendencia local que además incorpora un término estacional que puede ser estocástico, esto es

$$Z_t = \mu_t + S_t + \epsilon_t \tag{10}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \tag{11}$$

$$\beta_t = \beta_{t-1} + \xi_t, \tag{12}$$

donde $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$ y $\xi_t \sim N(0, \sigma_\xi^2)$ son independientes. La componente estacional es tal que, si s es el período

$$S_t + S_{t-1} + S_{t-2} + \ldots + S_{t-s+1} = a_t.$$

Note que, si $\sigma_a^2 > 0$ el término estacional es estocástico, y si $\sigma_a^2 = 0$ la componente es determinística.

Sin pérdida de generalidad, la representación de este modelo en la forma espacio de estados con $s = 4$ es dada por

$$
\begin{aligned}
Z_t &= [1\ 0\ 1\ 0\ 0]X_t + \epsilon_t \\
X_t &= \begin{bmatrix} \mu_t \\ \beta_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \xi_t \\ a_t \\ 0 \\ 0 \end{bmatrix}.
\end{aligned} \tag{13}
$$

Otra familia clásica de modelos lineales para series de tiempo son los modelos ARIMA; diferente a los modelos de descomposición discutidos hasta ahora, esta familia la caracteriza las estructuras de autocorrelación de las observaciones o de perturbaciones aleatorias. Los modelos pueden ser autorregresivos, de promedios móviles o mixtos autorregresivos y de medias móviles. Mientras que los modelos estructurales se identifican a través del comportamiento y conocimiento de la serie de análisis, los modelos ARMA los caracteriza las estructuras de autocorrelación de la serie.

3.7. Modelos autorregresivos

Un modelo Autorregresivo de orden p establece una relación entre las variables del proceso del tipo

$$X_t = C + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t, \quad p \in \mathbb{Z}^+, \tag{14}$$

donde $\{a_t,\ t \in T\}$ es un ruido blanco que reune la parte aleatoria de la relación.

El modelo autorregresivo de orden p se denota en forma simplificada como $AR(p)$. Cuando el proceso es estacionario en media, si se toma la esperanza a ambos lados de (14) se tiene

$$E[X_t] = E[C] + \phi_1 E[X_{t-1}] + \phi_2 E[X_{t-2}] + \cdots + \phi_p E[X_{t-p}] + E[a_t],$$

lo cual es igual a

$$\mu = C + \phi_1 \mu + \phi_2 \mu + \cdots + \phi_p \mu,$$

luego la constante del modelo está dada por

$$C = \mu \left(1 - \sum_{i=1}^{p} \phi_i\right).$$

Reemplazando C en (14) y denotando por $\tilde{X}_t$ a $\tilde{X}_t = X_t - \mu$, se tiene la siguiente representación del modelo $AR(p)$

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \cdots + \phi_p \tilde{X}_{t-p} + a_t, \tag{15}$$

la cual corresponde a un $AR(p)$ de media cero.

Otra manera de escribir la ecuación que relaciona las variables del proceso es en términos del operador de rezagos B en la siguiente forma

$$\tilde{X}_t = \phi_1 B \tilde{X}_t + \phi_2 B^2 \tilde{X}_t + \cdots + \phi_p B^p \tilde{X}_t + a_t,$$

de donde

$$\tilde{X}_t = (\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B_t^p)\tilde{X}_t + a_t,$$

o también

$$a_t = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B_t^p)\tilde{X}_t,$$

y denotando por $\phi_p(B)$ a $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 + \cdots + \phi_p B_t^p$, se tiene la representación $\phi_p(B)\tilde{X}_t = a_t$. La ecuación $\phi_p(B) = 0$ se denomina ecuación característica del proceso, y el proceso $AR(p)$ es estacionario cuando todas las raíces complejas de esta ecuación están fuera del círculo unitario.

En particular, cuando $p = 1$, la ecuación que relaciona las variables del proceso en el modelo autorregresivo (15) toma la forma

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + a_t,$$

donde el operador lineal $\phi(B)$ está dado por $1 - \phi_1 B = \phi(B)$ y la ecuación característica $1 - \phi_1 B = 0$ tiene su raíz fuera del círculo unitario cuando $|\phi_1| < 1$, lo cual corresponde a la condición de estacionariedad para el proceso $AR(1)$.

Si el proceso también es estacionario en varianza se tiene que

$$V(\tilde{X}_t) = \phi_1^2 V(\tilde{X}_t) + V(a_t),$$

de donde

$$\sigma_x^2 = \frac{\sigma_a^2}{(1-\phi_1^2)}.$$

Luego la condición de estacionariedad, $|\phi_1| < 1$, también garantiza que σ_x^2 sea finita y positiva.

La función de autocorrelación simple del proceso $AR(1)$ tiene la forma

$$\rho_x(k) = \phi_1^k, \quad \text{para} \quad |k| \geq 1,$$

y por lo tanto cuando el proceso es estacionario, la función $\rho_x(k)$ decrece a cero en forma exponencial.

Por su parte, la función de autocorrelación parcial tiene la forma

$$\phi(k) = \begin{cases} \phi_1 & \text{si} \quad k = 1 \\ 0 & \text{si } k \neq 1. \end{cases}$$

Volviendo al proceso (15), cuando $p = 2$, la ecuación que relaciona las variables del proceso toma la forma

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + a_t, \tag{16}$$

donde el operador lineal $\phi_2(B)$ está dado por $1 - \phi_1 B - \phi_2 B^2 = \phi_2(B)$, y la ecuación característica $1 - \phi_1 B - \phi_2 B^2 = 0$ tiene sus raíces fuera del círculo unitario cuando se cumplen las tres condiciones siguientes

$$|\phi_2| < 1,$$

$$\phi_1 + \phi_2 < 1,$$

$$\phi_2 - \phi_1 < 1,$$

las cuales corresponden a las condiciones de estacionariedad del proceso $AR(2)$, y además también garantizan que la varianza de X_t sea finita, dado que

$$Var(\tilde{X}_t) = \frac{(1-\phi_2)\sigma_a^2}{(1+\phi_2)(1-\phi_1-\phi_2)(1+\phi_1-\phi_2)}.$$

Si el modelo (16) se premultiplica por X_{t-k} y se toma valor esperado se llega a la ecuación en diferencias

$$\gamma_x(k) = \phi_1\gamma_x(k-1) + \phi_2\gamma_x(k-2), \qquad |k| \geq 1.$$

Si la anterior ecuación se divide entre $\gamma_x(0)$ entonces se obtiene la ecuación en diferencias

$$\rho_x(k) = \phi_1\rho_x(k-1) + \phi_2\rho_x(k-2), \qquad |k| \geq 1,$$

cuya ecuación característica es justamente $\phi_2(B) = 0$. La función de autocorrelación simple $\rho_x(k)$, para el proceso $AR(2)$, depende de la solución de la ecuación anterior, la cual a su vez depende de las soluciones de la ecuación característica, así:

i) Si ambas raíces denotadas G_1^{-1} y G_2^{-1} son reales pero distintas entonces $\rho_x(k) = A_1G_1^k + A_2G_2^k$ con A_1 y A_2 dos constantes resultantes de aplicar las condiciones iniciales $\rho(0) = 1$ y $\rho_1 = \frac{\phi_1}{1-\phi_2}$.

ii) Si ambas raíces son reales e iguales (G^{-1}), entonces $\rho_x(k) = A_1G^k + A_2kG^k$ donde A_1 y A_2 se obtienen de las condiciones iniciales.

iii) Si las raíces son un par de complejos conjugados denotados G_1^{-1} y G_2^{-1}, entonces la función de autocorrelación tiene una forma sinusoidal dado que

$$\rho_x(k) = \frac{r^k \sin(k\omega + \alpha)}{\sin\alpha}$$

donde r es el módulo de G_1, ω es el ángulo entre G_1 y el eje horizontal.

La función de autocorrelación parcial del proceso $AR(2)$ tiene la forma

$$\phi(k) = \begin{cases} \rho_1 = \frac{\phi_1}{1-\phi_2} & \text{si } k = 1 \\ \frac{\rho_2-\rho_1^2}{1-\rho_1^2} = \phi_2 & \text{si } k = 2 \\ 0 & \text{si } k \geq 3. \end{cases}$$

En general, el proceso $AR(p)$ definido en (15), puede reescribirse en la forma

$$\phi_p(B)\tilde{X}_t = a_t,$$

y la ecuación característica $\phi_p(B) = 0$, asociada al proceso, es una ecuación polinómica de grado p y por lo tanto tiene p raíces, G_1^{-1}, G_2^{-1}, ..., G_p^{-1}, con lo cual se genera la solución general de la ecuación en diferencias , de la forma

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \cdots + A_p G_p^k, \quad |G_i| < 1,$$

donde los A_i se obtienen de las condiciones iniciales y los G_i son reales o complejos; por lo tanto la función de autocorrelación simple ρ_k es una combinación lineal de funciones sinusoidales amortiguadas, (Morettin y Toloi, 2006).

Note que la función de autocorrelación parcial caracteriza los procesos autorregresivos, ya que el número de valores no nulos de esta función corresponden a su orden.

3.8. Modelos de medias móviles

Volviendo al tipo de relación lineal que presenten las variables del proceso, esta podría depender solamente de las variables contenidas en un proceso ruido blanco, surgiendo así los modelos de Medias Móviles (MA).

Un modelo de Medias Móviles de orden q establece la siguiente relación

$$\tilde{X}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad q \in \mathbb{Z}^+, \tag{17}$$

donde $\{a_t,\ t \in T\}$ es un proceso de ruido blanco.

En términos del operador de rezagos B, la ecuación (17) puede representarse como

$$\tilde{X}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t$$

$$\tilde{X}_t = \theta_q(B) a_t,$$

donde $\theta_q(B)$ representa el polinomio característico del proceso.

Observe que $E(\tilde{X}_t) = 0$ y $Var(\tilde{X}_t) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_a^2$.

El proceso $MA(q)$ es siempre estacionario y si las raíces de la ecuación $\theta_q(B) = 0$ caen fuera del círculo unitario, se dice que el proceso es invertible.

En la ecuación (17) cuando $q = 1$ se tiene el proceso de medias móviles de orden 1, $MA(1)$

$$\tilde{X}_t = (1 - \theta_1 B)a_t,$$

así que el proceso es invertible si $|\theta_1| < 1$.

La función de autocorrelación del proceso es dada por

$$\rho_x(k) = \begin{cases} \frac{-\theta_1}{1+\theta_1^2} & \text{si} \quad k = 1 \\ 0 & \text{si} \quad k > 1, \end{cases}$$

y la función de autocorrelación parcial es dada por

$$\phi_x(k) = \frac{-\theta_1^k(1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}, \quad k \geq 1.$$

Note que en un proceso $MA(1)$ la función de autocorrelación simple se vuelve cero a partir del rezago dos, mientras que la función de autocorrelación parcial decrece en forma exponencial.

Volviendo a la ecuación (17), cuando $q = 2$ se tiene el proceso de medias móviles de orden 2, $MA(2)$

$$\tilde{X}_t = (1 - \theta_1 B - \theta_2 B^2)a_t.$$

El proceso es siempre estacionario, y su invertibilidad depende de que se satisfagan las siguientes condiciones

$$|\theta_2| < 1,$$

$$\theta_1 + \theta_2 < 1,$$

$$\theta_2 - \theta_1 < 1.$$

La función de autocorrelación del proceso $MA(2)$ es dada por

$$\rho_x(k) = \begin{cases} \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} & \text{si } k=1 \\ \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} & \text{si } k=2 \\ 0 & \text{si } k>2 \end{cases}$$

La función de autocorrelación parcial del proceso $MA(2)$ decae a cero, en forma exponencial o sinusoidal, dependiendo de los signos y magnitudes de θ_1 y θ_2, o equivalentemente, dependiendo de las raíces de la ecuación $1-\theta_1 B-\theta_2 B^2=0$. Si las raíces son todas reales, la función decae a cero en forma exponencial, pero si existe algún par de raíces complejas, entonces el comportamiento dominante es sinusoidal.

En el proceso $MA(q)$, la varianza es dada por $Var(\tilde{X}_t)=\sigma_a^2\sum_{j=0}^{q}\theta_j^2$. Por su parte, la función de autocorrelación simple tiene la forma

$$\rho_x(k) = \begin{cases} \frac{-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2} & \text{si } k=1,2,\ldots,q \\ 0 & \text{si } k>q. \end{cases}$$

Como en el proceso $MA(2)$, la función de autocorrelación parcial del proceso $MA(q)$ decae a cero, en forma exponencial o sinusoidal, dependiendo de las raíces de la ecuación característica $1-\theta_1 B-\cdots-\theta_q B^q=0$, (Morettin y Toloi, 2006).

Note que, en los procesos de medias móviles, el número de valores no nulos de la función de autocorrelación simple caracterizan su orden.

3.9. Modelos ARMA y Modelos ARIMA

Un proceso de media cero $\{\tilde{X}_t\}$ se denomina un proceso ARMA(p,q) cuando es a la vez autorregresivo de orden p y medias móviles de orden q, esto es

$$\tilde{X}_t=\phi_1\tilde{X}_{t-1}+\phi_2\tilde{X}_{t-2}+\cdots+\phi_p\tilde{X}_{t-p}-\theta_1 a_{t-1}-\theta_2 a_{t-2}-\cdots-\theta_q a_{t-q}+a_t,\quad q\in\mathbb{Z}^+, \quad (18)$$

donde $\{a_t,\ t\in T\}$ es un proceso de ruido blanco. En términos del operador de rezagos, su forma simplificada es dada por

$$\Phi_p(B)\tilde{X}_t=\Theta_q(B)a_t.$$

En la práctica q y p son generalmente valores pequeños.

Cuando en particular el proceso $\{X_t\}$ no es estacionario y presenta fuertes tendencias lineales o polinómicas, previo a la modelación requiere de una transformación en diferencias del tipo

$$Y_t = (1-B)^d \tilde{X}_t = \tilde{X}_t - \tilde{X}_{t-d},$$

la cual elimina dichas tendencias y Y_t es estacionario. Cuando esto ocurre el modelo dado por

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + a_t, \quad q \in \mathbb{Z}^+, \tag{19}$$

se denomina autorregresivo integrado de medias móviles y se denota ARIMA(p,d,q). Generalmente, $d = 1$ si la serie presenta una fuerte tendencia lineal y rara vez $d = 2$. Valores de d mayores a 2 no son usuales. En términos del operador de rezagos la forma simplificada del modelo ARIMA(p,d,q) es dada por

$$\Phi_p(B)(1-B)^d \tilde{X}_t = \Theta_q(B) a_t.$$

En cuanto a las funciones de autocorrelación simple y parcial de los procesos ARMA(p,q), sin entrar en detalles y dada su utilidad en la práctica para identificar el modelo generador del proceso, basta decir que estas resultan de combinaciones del comportamiento del caso autorregresivo y del caso de medias móviles, (Hurtado y Salcedo, 1999, p.46).

Finalmente, a_t es un ruido no correlacionado, y conviene que a_t se distribuya normalente con media cero y varianza σ_a^2. Los siguientes modelos son muy comunes en la práctica:

$$
\begin{aligned}
ARIMA(1,0,0) &: X_t = \phi_1 X_{t-1} + a_t; \\
ARIMA(0,0,1) &: X_t = \theta_1 a_{t-1} + a_t; \\
ARIMA(1,0,1) &: X_t = \phi_1 X_{t-1} + \theta_1 a_{t-1} + a_t; \\
ARIMA(1,1,0) &: (1-\phi_1 B)(1-B)X_t = a_t; \\
&\quad (1-\phi_1 B)W_t = a_t; \\
&\quad W_t = \phi_1 W_{t-1} + a_t; \\
ARIMA(1,1,1) &: W_t = \phi_1 W_{t-1} + \theta_1 a_{t-1} + a_t;
\end{aligned}
$$

donde $W_t$es la serie de tiempo de la primera diferencia de X_t (Hurtado y Salcedo, 1999, p.27).

3.10. Procesos estacionales

Un proceso estacional es aquel que tiene un comportamiento singular que se repite cada cierto número de etapas observadas en una serie de tiempo. La estacionalidad del proceso se identifica gracias a las funciones de autocorrelación simple y parcial de las observaciones realizadas en el periodo de estacionalidad. Ahora bien, para incluir el componente estacional en un modelo ARIMA como el siguiente

$$\Phi_p(B)(1-B)^d \tilde{X}_t = \Theta_q(B)a_t,$$

se debe tener en cuenta que los residuales a_t no resultan independientes, es por esto que se pueden relacionar a través de una ecuación de la forma

$$\Phi_P(B^s)(1-B^s)^D a_t = \Theta_Q(B^s)c_t,$$

donde c_t es ahora un ruido blanco. Ahora bien, si se despeja a_t en cualquiera de las ecuaciones anteriores y se reemplaza en la sobrante, se llega a

$$\Phi_P(B^s)\Phi_p(B)(1-B)^d(1-B^s)^D\tilde{X}_t = \Theta_Q(B^s)\Theta_q(B)c_t.$$

Para la anterior ecuación, se necesita ahora 7 parámetros que son: p, d, q para la parte no estacional y s, P, D, Q para la parte estacional, de modo que se construya el modelo

$$SARIMA(p,d,q) * (P,D,Q)_s,$$

(Hurtado y Salcedo, 1999, p.86).

3.11. Algoritmo K-means

Este algoritmo es uno de los más utilizados actualmente por diversos investigadores que pretenden realizar algún tipo de agrupamiento. Fue desarrollado en 1967 por James MacQueen, y su popularidad ha venido en aumento debido a que se trata de un método sencillo de clasificación no supervisada, que consiste principalmente en dividir cierto número de observaciones en k grupos o Clústeres. De manera más estructurada, para un conjunto de n observaciones $(x_1, x_2, x_3, ..., x_n)$, el algoritmo k-means se encarga de construir una división de las observaciones en k conjuntos $\{S_1, S_2, ..., S_k\}$ donde $(k \leq n)$, minimizando la suma de cuadrados de las distancias de las observaciones dentro de cada grupo S_i con respecto a su centroide μ_i; esto es, se trata de hallar el conjunto S tal que,

$$\min_S \sum_{i=1}^{k} \sum_{x_n \epsilon S_i} \|x_n - \mu_i\|^2$$

donde $S = \{S_1, S_2, ..., S_k\}$ es el conjunto de los k clústeres y generalmente μ_i es la media de los datos en cada clúster; (González y Ticona, 2019).

3.12. Gráficos de silueta

Este tipo de gráficos tiene como objetivo presentar una evaluación acerca de qué tan acertada es una clasificación realizada previamente. El valor de la silueta es una medida

que hace refrencia a cuán similar es un objeto de algún grupo con respecto al resto de los objetos pertenecientes al mismo grupo, este valor oscila entre -1 y 1, y a medida que el valor se acerca a 1, refleja que el objeto está muy bien asociado con su propio grupo y muy alejado del resto de grupos. Ahora bien, cuando este tipo de gráficas se utiliza para evaluar una clasificación k-means, lo que ocurre es lo siguiente: Para una observación cualquiera i que pertenezca a un clúster C_i, se tiene

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \epsilon C_i, i \neq j} d(i, j),$$

donde $d(i, j)$ es la distancia entre los datos i y j en el clúster C_i, y se divide por $|C_i| - 1$ porque no se incluye la distancia $d(i, i)$. Luego, se define la diferencia media del punto i a algún clúster $C \neq C_i$ como la media de la distancia desde i a todos los puntos que pertenezcan a C, así:

$$b(i) = \underset{k \neq i}{min} \frac{1}{|C_k|} \sum_{j \epsilon C_k} d(i, j).$$

En este caso, se afirma que el clúster con la diferencia mínima es el clúster vecino de i. Ahora bien, para definir una *silhouette* o un valor de un conjunto de datos i se tiene

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}, \; si \; |C_i| > 1 \; y \; s(i) = 0 \; si \; |C_i| = 1,$$

que se puede sintetizar en

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & si \; a(i) < b(i) \\ 0 & si \; a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1 & si \; a(i) > b(i) \end{cases}$$

e indica que, para que $s(i)$ se aproxime a uno $a(i) < b(i)$, ya que eso indicaría que i es muy similar a los datos del clúster al cuál pertenece y muy diferente a los datos del clúster vecino, (González y Ticona, 2019).

4. Análisis estadístico de los datos

En esta sección se aplican las correspondientes metodologías para la clusterización y modelación de la serie temporal del número de infectados diarios por SAR-COV2 en Colombia desde Marzo 6 de 2020 hasta el 9 de agosto de 2021.

4.1. Obtención de Clústeres

Inicialmente, para llevar a cabo el primer objetivo específico de la presente investigación, se consultó en la página web *https://ourworldindata.org/* el número de infectados diarios por coronavirus y el número acumulado de los infectados por coronavirus en Colombia, desde que se registró el primer caso hasta el día 09 de Agosto del año 2021, obteniendo así un total de 522 datos tanto del número de infectados diarios como del valor acumulado.

Posteriormente, se realizó la clasificación en clústeres, utilizando el algoritmo k-means, y para determinar la cantidad adecuada de clústeres se empleó el gráfico de siluetas que se observa en la figura 2, en donde se compara la clasificación de 2, 3, 4 y 5 clústeres.

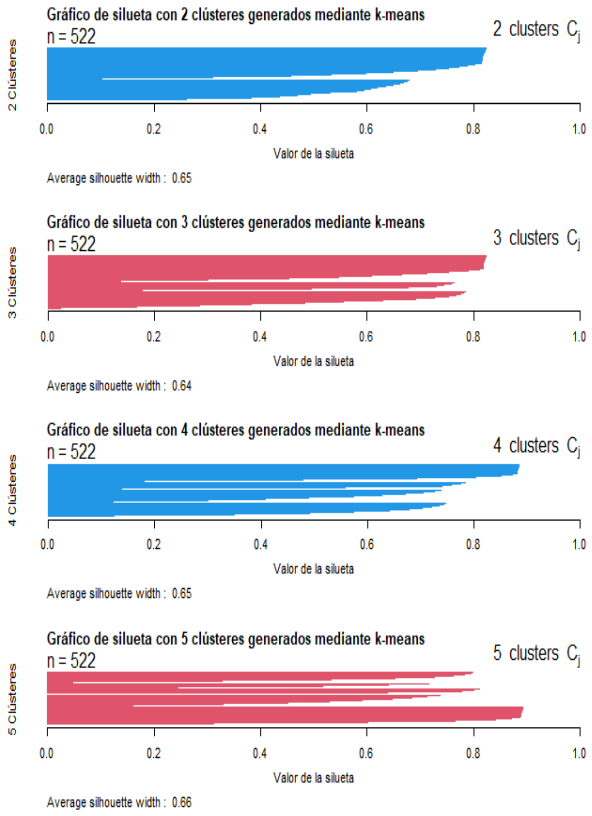


Figura 2: Gráfico de siluetas que hacen referencia a la cantidad adecuada de clústeres

Ahora bien, dado que la medida promedio silueta más alta se obtiene para 5 clústeres, se decide trabajar utilizando dicha clasificación, la cual realiza las particiones de las observaciones en los días 177, 301, 413 y 469 después de la primera observación, que corresponden a las siguientes fechas: 29 de Agosto de 2020, 31 de diciembre de 2020, 22 de abril de 2021, 17 de Junio de 2021 y que se observan en la Figura 3.

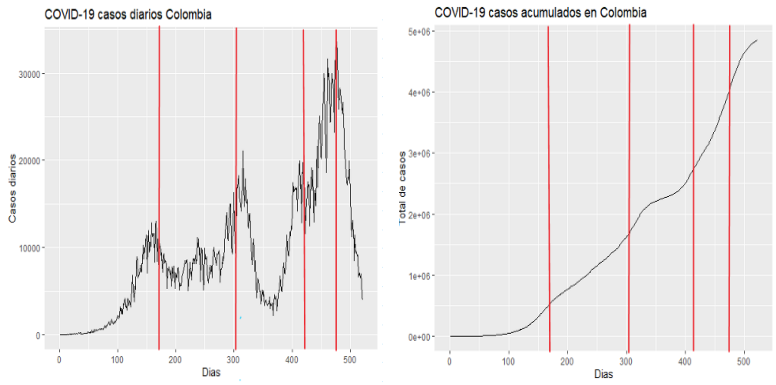


Figura 3: Clasificación de los datos en 5 clústeres

Al analizar con más detalle cada una de las fechas mencionadas anteriormente se encuentra lo siguiente:

- 29 de Agosto de 2020: En los primeros 3 meses de la pandemia, los casos no eran muy significativos debido al cuidado que cada persona tenía, a la cuarentena estricta que tuvo el país y a otros mecanismos que se utilizaron, tales como los toques de queda, el pico y cédula, el distanciamiento, entre otros. Sin embargo, una vez pasados los 100 primeros días desde que se registró el primer caso, las medidas de mitigación fueron disminuyendo, al igual que el miedo que tenían las personas, además, el gobierno pensando en la reactivación económica, realizó una actividad llamada Dia sin IVA el 19 de Junio del año 2020, en donde se registraron grandes aglomeraciones y violaciones a las medidas de mitigación, que probablemente aumentaron los casos, de igual manera, el día 01 de septiembre se levantó la cuarentena obligatoria y el contagiarse o no ahora dependía de los cuidados personales.

- 31 de diciembre de 2020: Durante el mes de diciembre, debido a temas culturales propios del país, se registró un gran aumento en el número de contagios, tal como se menciona en el Boletín de Prensa número 75 del ministerio de salud: «La realización de fiestas, celebraciones masivas, encuentros en hogares con diferentes familias,

aglomeraciones en comercios y relajamiento de las medidas en diciembre, presentó unas consecuencias graves en mortalidad e infecciones». Sin embargo, en la gráfica también se observa que luego de pasar el pico de diciembre, empieza a presentarse un descenso en los casos entre el día número 337 en adelante, que corresponde al mes de febrero de 2021, y aquí se hace necesario resaltar que en Colombia la vacunación al personal médico y priorizado inició el 17 de febrero de dicho año.

- 22 de abril de 2021: Como se mencionó anteriormente, en la medida que la vacuna se fué implementando a más personas los contagios fueron disminuyendo, sin embargo, la mentalidad de muchas personas que ya estaban vacunadas fué la de creer que la vacuna los haría inmunes, y ya estaba al alcance de muchos, por ello, las medidas de autocuidado bajaron muchísimo y la curva de casos diarios empezó a subir nuevamente entre el 20 de marzo y el 22 de abril aproximadamente.

- 17 de Junio de 2021: Entre el 22 de abril y el 17 de Junio, se presentó el pico más alto de contagios que se ha visto hasta ahora, y probablemente se deba a la cantidad de manifestaciones sociales que se presentaron en el país a partir del 28 de abril y que se extendieron durante algunos meses, provocando una gran cantidad de aglomeraciones que en la mayoría de los casos no utilizaban adecuadamente los protocolos de bioseguridad. No obstante, cabe añadir que a partir del mes de Junio, la vacunación se incrementó en gran medida, lo que generó posteriormente una caída significativa en la serie (Ver Anexo 14).

4.2. Modelación de la serie temporal en cada clúster

Una vez ejecutada la clasificación por k-means en la serie y seleccionado el número adecuado de clústeres, se procedió a modelar cada uno de los clústeres teniendo en cuenta los modelos exponencial, cuadrático, ARIMA y SARIMA, bajo todas las posibles combinaciones entre ceros y unos de sus parámetros, lo que arrojó un total de 1470 modelos, que

se evaluaron todos (uno a uno) bajo el software Statgraphics y se descartaron teniendo en cuenta la significancia estadística (Ver Anexos). Ahora bien, teniendo en cuenta lo anterior, por cada clúster se hizo un análisis particular de los modelos restantes y se eligieron los mejores con base en sus valores de RMSE, MAE y MAPE (cuando no se tenían ceros).

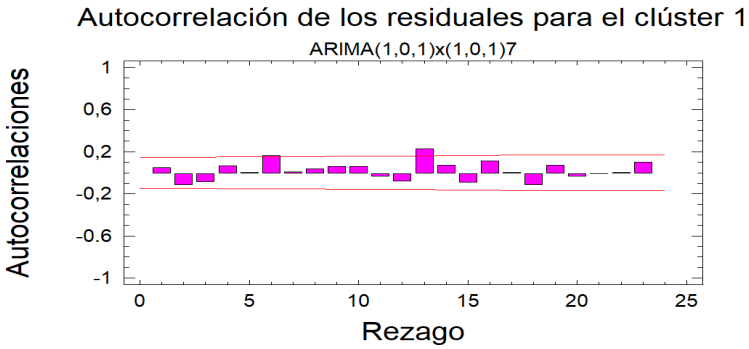
4.3. Modelación del primer clúster

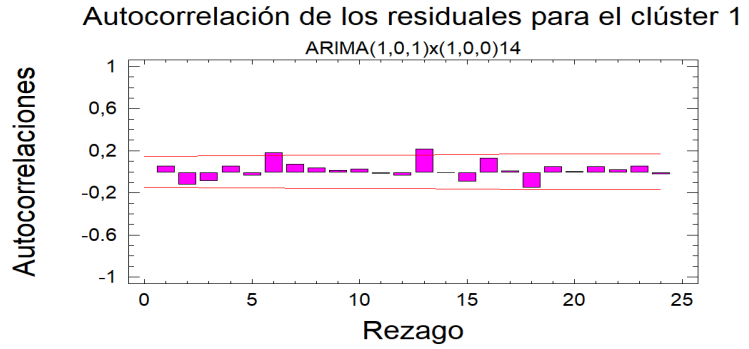
En este primer clúster se incluyen los primeros 177 datos, que corresponden al número de infectados desde el 06 de Marzo de 2020 hasta el 28 de Agosto de 2020. Para este clúster, se ajustaron 315 modelos de los cuales se descartaron 129 de ellos por su no significancia y de los 186 restantes (que aparecen en los cuadros 7, 8,9,10 y 11) se escogieron los 5 mejores que se expresan a continuación junto con sus valores de RMSE y MAE.

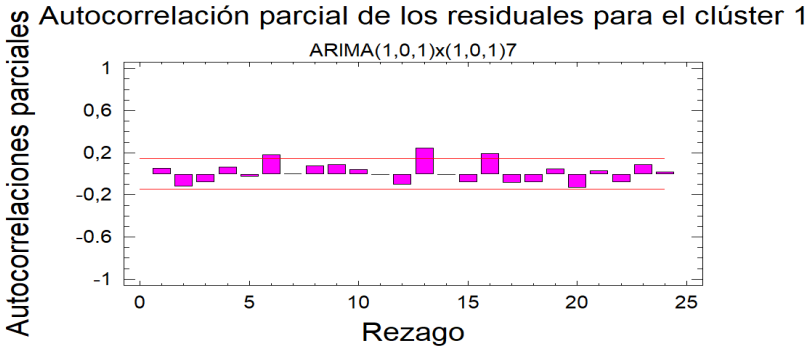
Clúster 1 ($n_1 = 177$)			
Modelo	Ecuación	RMSE	MAE
Exponencial+$e_t \sim SARIMA(1,0,1)*(1,0,1)_7$	$exp(3,85009 + 0,0374796t)+$ $\Phi_1(B^7)\Phi_1(B)\tilde{X}_t = \Theta_1(B^7)\Theta_1(B)c_t$	855,377	486,734
Cuadrático+$e_t \sim SARIMA(0,1,1)*(1,0,1)_7$	$492,11 - 40,9767t + 0,62243t^2+$ $\Phi_1(B^7)(1-B)^1\tilde{X}_t = \Theta_1(B^7)\Theta_1(B)c_t$	850,433	461,652
ARIMA $(1,0,1)$	$\Phi_1(B)\tilde{X}_t = \Theta_1(B)a_t$	826,691	432,012
SARIMA $(1,0,1)*(1,0,1)_7$	$\Phi_1(B^7)\Phi_1(B)\tilde{X}_t = \Theta_1(B^7)\Theta_1(B)c_t$	**802,057**	**429,159**
SARIMA $(1,0,1)*(1,0,0)_{14}$	$\Phi_1(B^{14})\Phi_1(B)\tilde{X}_t = \Theta_1(B)c_t$	**814,717**	**430,414**

Cuadro 2: Modelos ajustados para la serie del Clúster 1

Como se puede observar, los 2 modelos que compiten por el mejor ajuste, son: SARIMA $(1,0,1)*(1,0,1)_7$ y SARIMA $(1,0,1)*(1,0,0)_{14}$, con residuales normalmente distribuidos y no autocorrelacionados, según sus funciones de autocorrelación, las cuales aparecen a continuación:







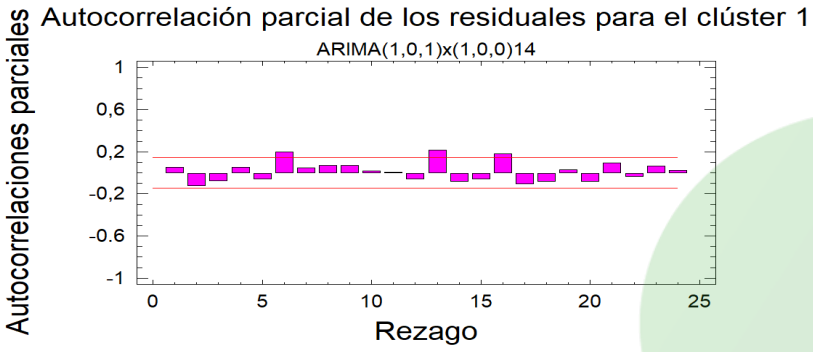


Figura 4: Fas y Fap de los residuales

Ahora bien, para evaluar el efecto de la reactivación económica y la mitigación de las medidas de control gubernamentales, se compara los datos reales y el pronóstico de 20 dias adelante según los dos mejores modelos, como se observa a continuación:

Figura 5: Comparación de los pronósticos a 20 días adelante

4.4. Modelación del segundo clúster

En este segundo clúster se incluyen 124 datos, que corresponden al número de infectados desde el 29 de agosto de 2020 hasta el 31 de diciembre de 2020. Para este clúster, se ajustaron 315 modelos de los cuales se descartaron 64 de ellos por su no significancia y de los 251 restantes (que aparecen en los cuadros 12, 13,14,15,16 y 17) se escogieron los 5 mejores los cuales se expresan a continuación junto con sus valores de RMSE, MAE y MAPE.

Clúster 2 ($n_2 = 124$)				
Modelo	Ecuación	RMSE	MAE	MAPE
Exponencial+$e_t \sim SARIMA(1,0,1)*(1,0,1)_{14}$	$exp(8,77743 + 0,00371437t)+$ $\Phi_1(B^{14})\phi_1(B)\tilde{X}_t = \Theta_1(B^{14})\theta_1(B)c_t$	**1058,48**	**783,55**	—
Cuadrático+$e_t \sim SARIMA(1,0,0)*(1,0,1)_{14}$	$7976,67 - 47,0442t + 0,65041t^2+$ $\Phi_1(B^{14})\phi_1(B)\tilde{X}_t = \Theta_1(B^{14})c_t$	**1049,26**	**786,94**	—
SARIMA$(1,0,1)*(1,0,1)_{14}$	$\Phi_1(B^{14})\phi_1(B)\tilde{X}_t = \Theta_1(B^{14})\theta_1(B)c_t$	1097,22	793,66	10,01
SARIMA$(1,1,0)*(1,0,1)_{14}$	$\Phi_1(B^{14})\phi_1(B)(1-B)^1\tilde{X}_t = \Theta_1(B^{14})\theta_1(B)c_t$	1088,81	794,12	10,15
SARIMA$(0,1,1)*(1,0,1)_{14}$	$\Phi_1(B^{14})(1-B)^1\tilde{X}_t = \Theta_1(B^{14})\theta_1(B)c_t$	1069,05	782,91	10,04

Cuadro 3: Modelos ajustados para la serie del Clúster 2

Como se puede observar, hay 2 modelos que compiten por el mejor ajuste, son: $exp(8,77743+0,00371437t) + e_t \sim SARIMA(1,0,1)*(1,0,1)_{14}$ y $7976,67 - 47,0442t + 0,65041t^2 + e_t \sim SARIMA(1,0,0)*(1,0,1)_{14}$, con residuales normales y no autocorrelacionados como es de esperar, y cuyas gráficas aparecen a continuación:

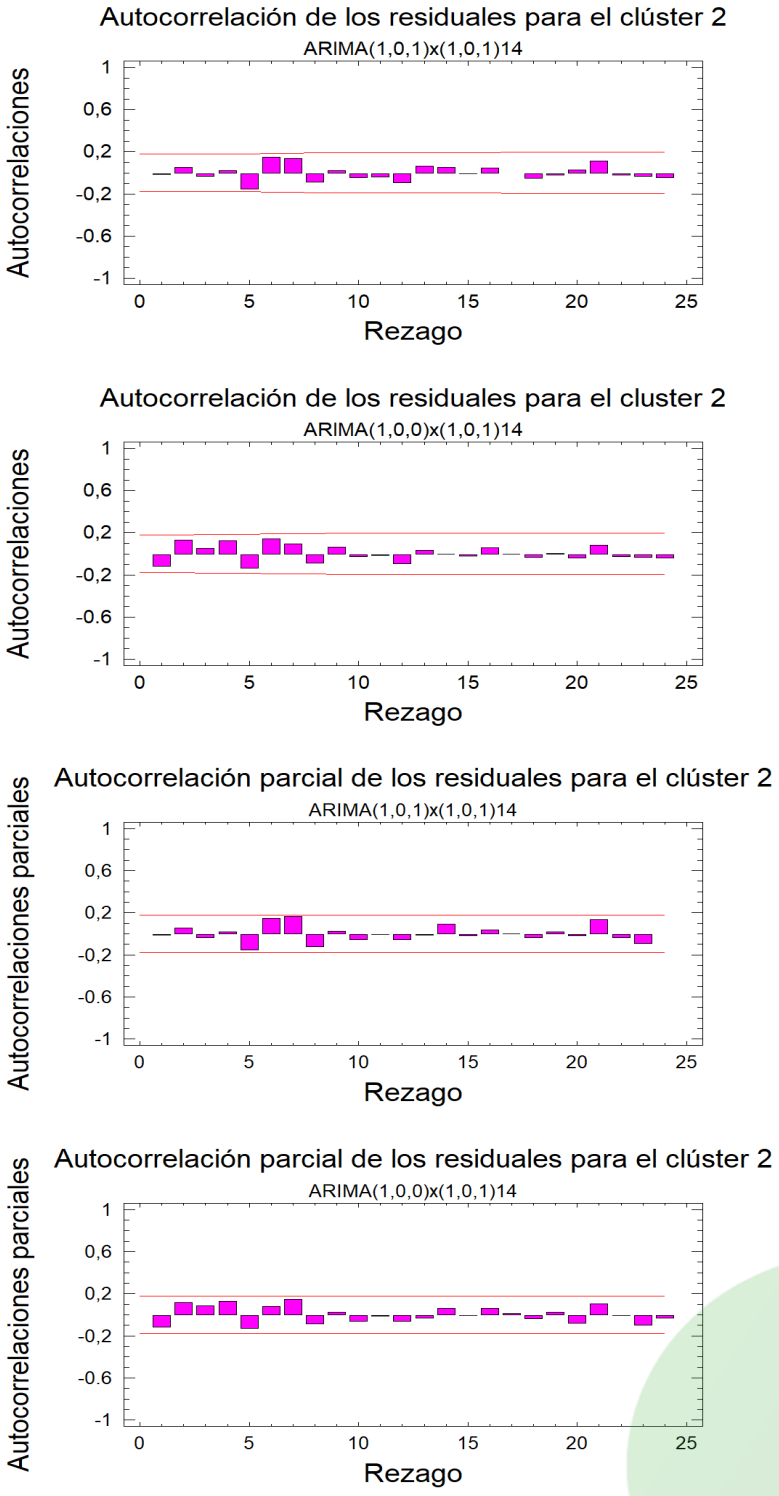


Figura 6: Fas y Fap de los residuales a 20 días adelante

Ahora bien, para evaluar el efecto de las actividades decembrinas, se compara los datos reales y los pronósticos a 20 días adelante, dando como resultado lo siguiente:

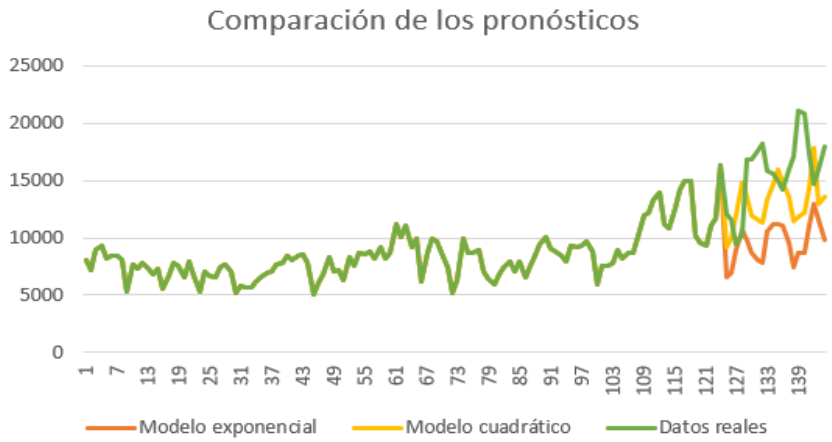


Figura 7: Comparación de los pronósticos

4.5. Modelación del tercer clúster

En este tercer clúster se incluyen 112 datos, que corresponden al número de infectados desde el 01 de enero de 2021 hasta el 22 de Abril de 2021. Para este clúster, se ajustaron 210 modelos de los cuales se descartaron 112 de ellos por su no significancia y de los 98 restantes (que aparecen en los cuadros 18, 19 y 20) se escogieron los 5 mejores que se expresan a continuación junto con sus valores de RMSE, MAE y MAPE.

Clúster 3 ($n_3 = 112$)				
Modelo	Ecuación	RMSE	MAE	MAPE
Cuadrático+$e_t \sim SARIMA(1,1,1)*(0,1,1)_{14}$	$21263,5 - 564,058t + 4,77391t^2 +$ $\Phi_1(B)(1-B)^1(1-B^{14})^1\tilde{X}_t = \Theta_1(B^{14})\Theta_1(B)c_t$	**1195,05**	**923,282**	—
Cuadrático+$e_t \sim SARIMA(0,1,0)*(0,1,1)_{14}$	$21263,5 - 564,058t + 4,77391t^2 +$ $(1-B)^1(1-B^{14})^1\tilde{X}_t = \Theta_1(B^{14})c_t$	**1196,8**	**923,037**	—
SARIMA$(1,1,1)*(0,1,1)_{14}$	$\Phi_1(B)(1-B)^1(1-B^{14})^1\tilde{X}_t = \Theta_1(B^{14})\Theta_1(B)c_t$	1241,87	897,945	11,28
SARIMA$(0,1,0)*(0,1,1)_{14}$	$(1-B)^1(1-B^{14})^1\tilde{X}_t = \Theta_1(B^{14})c_t$	1225,26	907,275	11,38
SARIMA$(1,1,1)*(0,1,1)_7$	$\Phi_1(B)(1-B)^1(1-B^7)^1\tilde{X}_t = \Theta_1(B^7)\Theta_1(B)c_t$	1271,17	929,433	11,66

Cuadro 4: Modelos ajustados a la serie del Clúster 3

Como se puede observar, los 2 modelos que compiten por el mejor ajuste, son: $21263,5 - 564,058t + 4,77391t^2 + e_t \sim SARIMA(1,1,1)*(0,1,1)_{14}$ y $21263,5 - 564,058t + 4,77391t^2 + e_t \sim SARIMA(0,1,0)*(0,1,1)_{14}$, con residuales no autocorrelacionados y normalmente distribuidos como se observa a continuación.

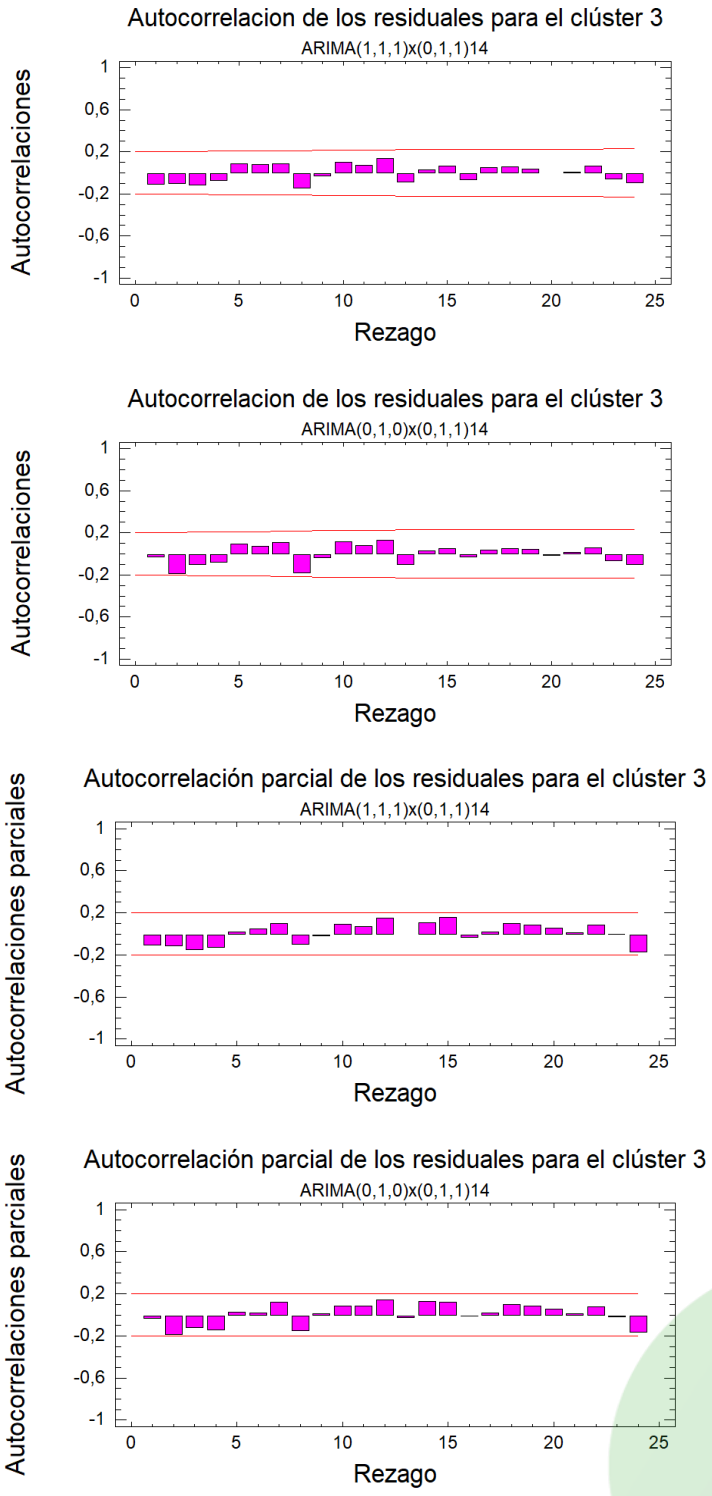


Figura 8: Fas y Fap de los residuales

Ahora bien, para evaluar el efecto de los primeros esquemas de vacunación y la pérdida de autocuidados debido a los mismos, se compara entre los datos reales y el pronóstico a 20 días adelante para los dos mejores modelos, dando como resultado lo siguiente :

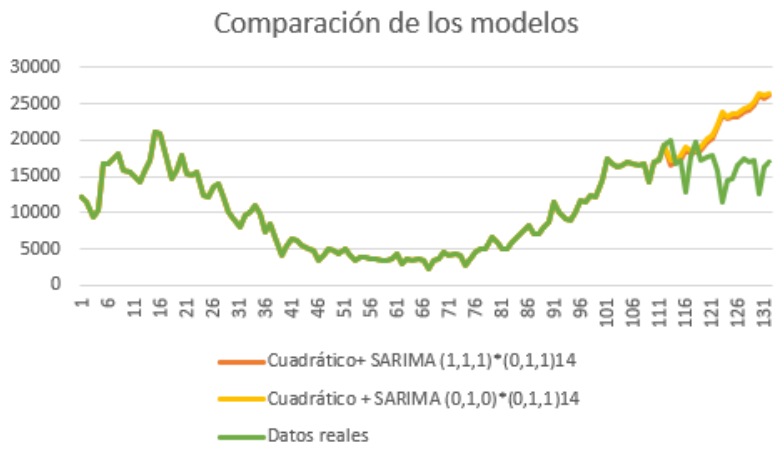


Figura 9: Comparación de los pronósticos a 20 días adelante

4.6. Modelación del cuarto clúster

En este cuarto clúster se incluyen 56 datos, que corresponden al número de infectados desde el 23 de abril del 2021 hasta el 17 de Junio del 2021. Para este clúster, se ajustaron 315 modelos de los cuales se descartaron 164 de ellos por su no significancia y de los 151 restantes (que aparecen en los cuadros 21, 22, 23 y 24) se escogieron los 5 mejores, que se expresan a continuación junto con sus valores de RMSE, MAE y MAPE.

Clúster 4 ($n_4 = 56$)				
Modelo	Ecuación	RMSE	MAE	MAPE
Exponencial+$e_t \sim SARIMA(1,0,0) * (1,0,1)_{14}$	$exp\,(9,54616 + 0,011971t)+$ $\Phi_1(B^{14})\Phi_1(B)\tilde{X}_t = \Theta_1(B^{14})c_t$	2211,79	1738,9	—
Cuadrático+$e_t \sim SARIMA(1,0,0) * (1,0,1)_{14}$	$16726,2 - 116,287t + 6,45641t^2+$ $\Phi_1(B^{14})\Phi_1(B)\tilde{X}_t = \Theta_1(B^{14})c_t$	**2114,36**	**1705,03**	—
Cuadrático+$e_t \sim SARIMA(1,1,1) * (1,0,1)_{14}$	$16726,2 - 116,287t + 6,45641t^2+$ $\Phi_1(B^{14})\Phi_1(B)(1-B)^1\tilde{X} = \Theta_1(B^{14})\Theta_1(B)c_t$	**2128,64**	**1709,32**	—
Cuadrático+$e_t \sim SARIMA(0,0,1) * (1,0,1)_{14}$	$16726,2 - 116,287t + 6,45641t^2+$ $\Phi_1(B^{14})\tilde{X}_t = \Theta_1(B^{14})\Theta_1(B)c_t$	2193,84	1775,67	—
SARIMA$(1,1,1) * (1,0,1)_{14}$	$\Phi_1(B^{14})\Phi_1(B)(1-B)^1\tilde{X}_t = \Theta_1(B^{14})\Theta_1(B)c_t$	2299,46	1748,18	9,36

Cuadro 5: Modelos ajustados a la serie del Clúster 4

Como se puede observar, los 2 modelos que compiten por el mejor ajuste, son: Cuadrático+$e_t \sim SARIMA(1,0,0) * (1,0,1)_{14}$ y Cuadrático+$e_t \sim SARIMA(1,1,1) * (1,0,1)_{14}$, con residuales normales y no autocorrelacionados como se observa en las siguientes figuras.

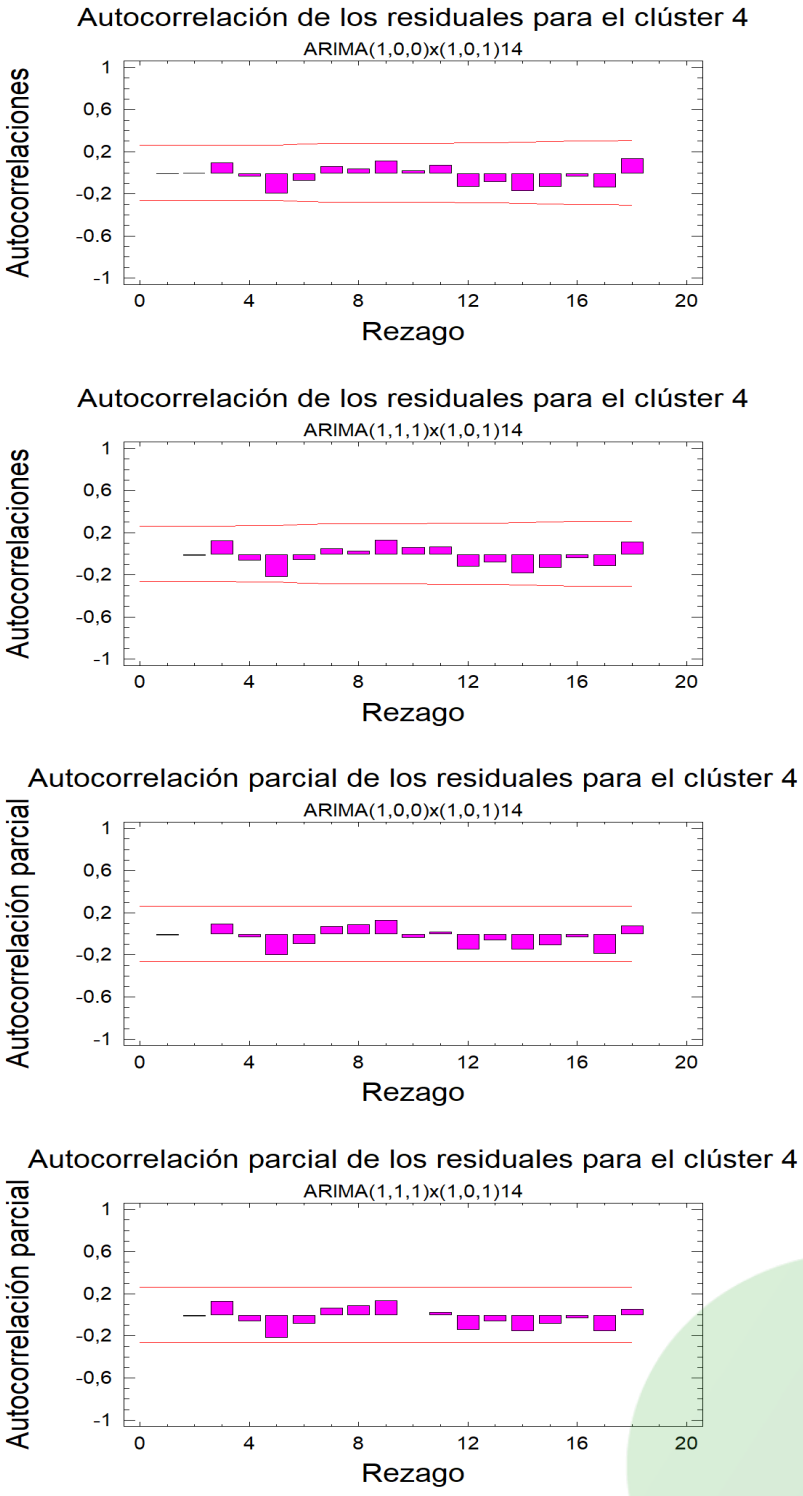


Figura 10: Fas y Fap de los residuales

Ahora bien, para evaluar el efecto de las manifestaciones sociales, se comparan los datos reales y el pronóstico a 20 días adelante a través de los dos mejores modelos, como se muestra a continuación:

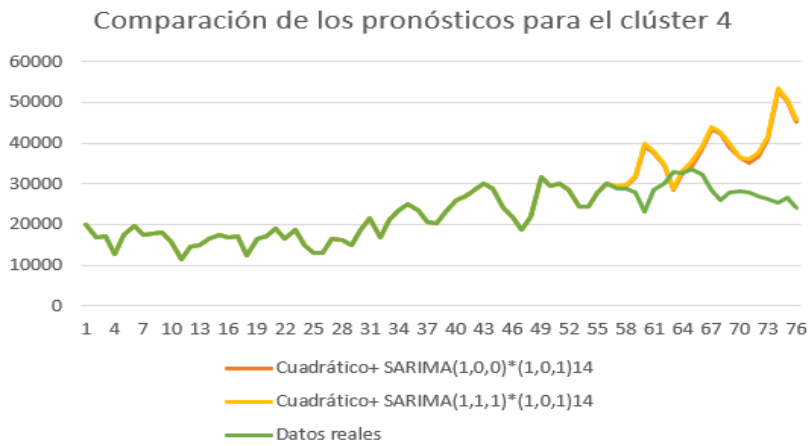


Figura 11: Comparación de los pronósticos a 20 días adelante

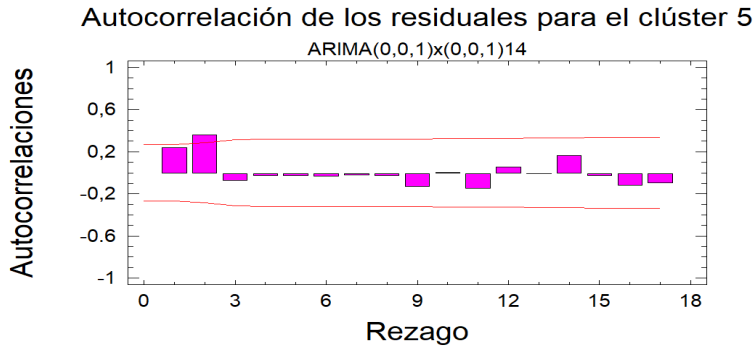
4.7. Modelación del quinto clúster

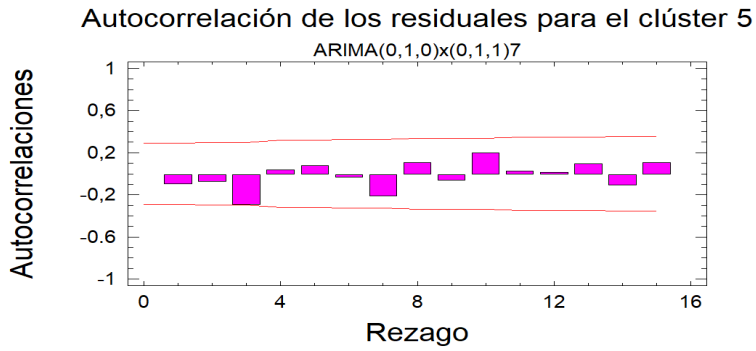
En este quinto y último clúster se incluyen 53 datos, que corresponden al número de infectados desde el 17 de junio del 2021 hasta el 09 de agosto del 2021. Para este clúster, se ajustaron 315 modelos de los cuales se descartaron 200 de ellos por su no significancia y de los 115 restantes (que aparecen en los cuadros 25, 26 y 27) se escogieron los 5 mejores que se expresan a continuación junto con sus valores de RMSE, MAE y MAPE.

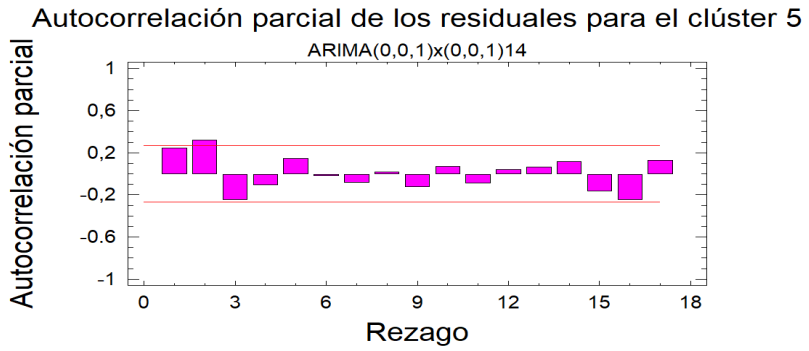
Clúster 5 ($n_5 = 53$)				
Modelo	Ecuación	RMSE	MAE	MAPE
Exponencial+$e_t \sim SARIMA(0,1,0) * (0,1,1)_7$	$exp(10,6553 - 0,0360541t)+$ $(1-B)^1(1-B^7)^1\tilde{X}_t = \Theta_1(B^7)c_t$	1476,61	1137,97	—
Cuadrático+$e_t \sim SARIMA(1,0,0) * (0,1,1)_7$	$32391,8 - 416,418t - 2,69767t^2+$ $\Phi_1(B)(1-B^7)^1\tilde{X}_t = \Theta_1(B^7)c_t$	1327,77	1015,54	—
Cuadrático+$e_t \sim SARIMA(0,0,1) * (0,0,1)_{14}$	$32391,8 - 416,418t - 2,69767t^2+$ $\tilde{X}_t = \Theta_1(B^{14})\Theta_1(B)c_t$	**1295,61**	**1008,39**	—
Cuadrático+$e_t \sim SARIMA(1,0,0) * (1,0,1)_{14}$	$32391,8 - 416,418t - 2,69767t^2+$ $\Phi_1(B^{14})\Phi_1(B)\tilde{X}_t = \Theta_1(B^{14})c_t$	1342,25	991,937	—
SARIMA$(0,1,0) * (0,1,1)_7$	$(1-B)^1(1-B^7)^1\tilde{X}_t = \Theta_1(B^7)c_t$	**1464,35**	**1087,13**	**8,31**

Cuadro 6: Modelos ajustados a la serie del Clúster 5

Como se puede observar, los mejores 2 modelos son: Cuadrático+$e_t \sim SARIMA(0,0,1)*(0,0,1)_{14}$ y SARIMA $(0,1,0) * (0,1,1)_7$, con errores normales y no autocorrrelacionados como se observa en la siguiente figura.







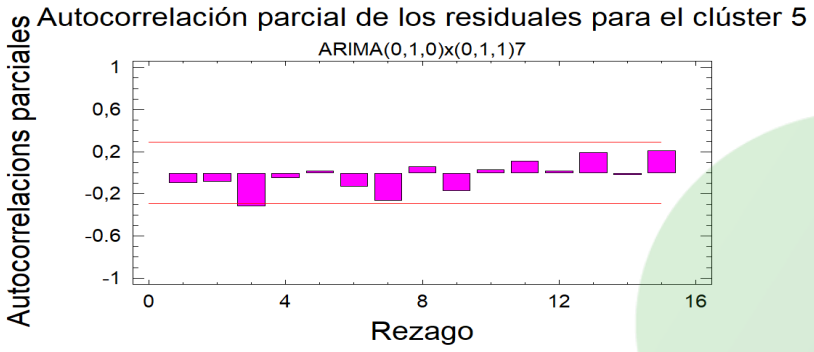


Figura 12: Fas y Fap de los residuales

Ahora bien, para evaluar el efecto de la vacunación masiva, se comparan los datos reales con los 20 pronósticos adelante de los 2 mejores modelos, cuyo resultado se observa en la siguiente Figura.

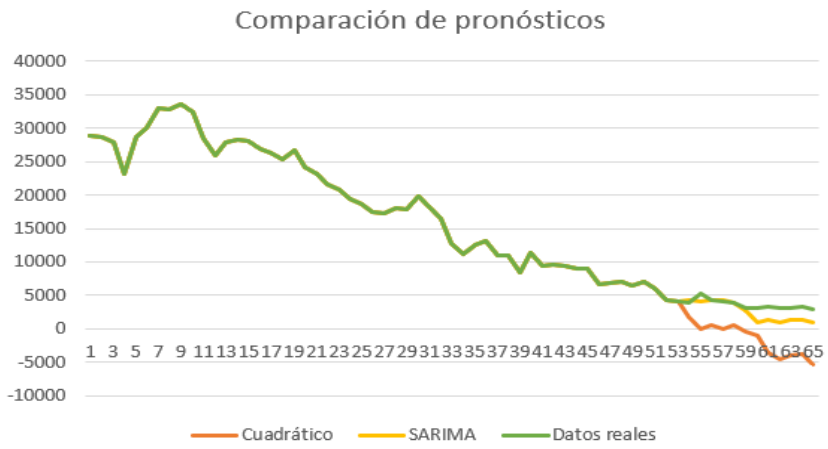


Figura 13: Comparación de los pronósticos a 20 días adelante

5. Resultados y discusiones

En términos generales, en la mayoría de los clústeres se evidenció que el modelo más adecuado para ajustar la tendencia es el cuadrático, y un SARIMA para modelar la autocorrelación de los residuales. Cabe mencionar que lo anterior no concuerda con la hipótesis muchas veces planteada a través de los medios, afirmando que el comportamiento era de tipo exponencial. Además, la estacionalidad de 7 o 14 días probablemente se encuentre relacionada con facores como las cuarentenas impuestas a cada 15 días, el período de incubación del virus y/o la espera en la obtención de los resultados de las pruebas, pues en los primeros meses las pruebas estaban centralizadas en unas pocas ciudades, y había una demora entre el envío de las muestras y la obtención de los resultados. Por otra parte, de manera particular se obtuvieron los siguientes resultados:

- En el clúster 1, el modelo SARIMA$(1,0,1)*(1,0,1)_7$ es el modelo con el menor valor de RMSE y MAE, además, en las gráficas de la autocorrelación simple y parcial de sus residuales, los valores se mantienen dentro de la banda de confianza en su mayoría, y al sobreponer los pronósticos de este modelo con los datos reales son más cercanos.

- En el clúster 2, el modelo cuadrático + SARIMA $(1,0,0)*(1,0,1)_{14}$ es el modelo con el menor valor de RMSE, además, en las gráficas de la autocorrelación simple y parcial de sus residuales, los valores se mantienen dentro del intervalo de confianza, y al sobreponer los pronósticos de este modelo compuesto junto con los datos reales resultó ser el más adecuado.

- En el clúster 3, el modelo más adecuado es el cuadrático + SARIMA $(1,1,1)*(0,1,1)_{14}$, ya que tiene el menor valor de RMSE y MAE, además, en las gráficas de la autocorrelación simple y parcial de sus residuales, los valores se mantienen dentro del intervalo de confianza, y al sobreponer los 20 pronósticos siguientes junto con los datos reales resultó ser el modelo que mejor ajusta los datos para este clúster.

- En el clúster 4, el modelo cuadrático + SARIMA $(1,0,0)*(1,0,1)_{14}$ es el modelo que presentó mejor ajuste con menor valor de RMSE y MAE, además, por las gráficas de la autocorrelación simple y parcial de sus residuales, se verifica la hipótesis de no autocorrelación residual. Por su parte, con respecto a los pronósticos el modelo sobrestima los valores reales, esto es de esperarse dado el modelo, sin embargo los datos reales reflejaron una caída importante, debido al efecto de los diferentes programas de vacunación que han demostraron su eficacia en el control del virus.

- En el clúster 5, el modelo cuadrático + SARIMA $(0,0,1)*(0,0,1)_{14}$ es el que ajusta mejor los datos. Con respecto a los pronósticos del modelo, estos van siempre en descenso debido a la tendencia que traen los datos en algunos momentos, subestimando los valores reales.

6. Conclusiones

- Para el análisis de la serie temporal del número de infectados por SARS-Cov2 en Colombia, la clasificación más adecuada es aquella en la que se utilizan 5 clústeres, ya que en el gráfico de silueta, se puede observar que la medida promedio silueta más alta (0.66) se obtiene para 5 clústeres.

- En todos los clústeres se evidenció que entre los modelos ejecutados, en general, el cuadrático es el que mejor modela la tendencia de la serie y el SARIMA el que mejor modela la autocorrelación con una estacionalidad de 14 dias, excepto en el primer clúster en donde la estacionalidad es de 7 días, lo que explica una alta transmisión y un menor control, además, al ejecutar cada uno de los modelos, se puede concluir que las combinaciones pueden ser muy amplias y variadas, sin embargo, al momento de elegir el mejor modelo no basta con basarse en el menor valor del RMSE o el MAPE, sino también es importante tener en cuenta la significancia del modelo, los valores arrojados en los pronósticos, la autocorrelación de los residuales, etc.

- Para todos los clústeres, los modelos ofrecen una buena predicción, excepto en el cuarto clúster, donde se evidencia una sobrestimación con respecto a los valores reales debido al condicionamiento que hubo en la serie principalmente por los efectos de la vacunación.

- Debido a que las cifras del número de infectados cada día varían y el virus aún está vigente, para futuras investigaciones, se propone modelar la serie con mayor cantidad de datos, variando el número de clústeres a utilizar y los parámetros de los modelos, además, se podría evaluar el efecto de otras variantes del virus.

Referencias

[1] Arenas, J. Cota, W. Gómez-Gardeñes, J. Gómez, S. Granell, S. Matamalas, J. Soriano, D. Steinegger, B. (2020). "A mathematical model for the spatiotemporal epidemic spreading of COVID19". Physical Review X, American Physical Society, 10(4): 41-55. DOI: 10.1103/PhysRevX.10.041055.

[2] Bravo, L. Grillo, E. Guerrero R. Santaella, J. (2020). "Modelos matemáticos y el COVID19". Colombia médica, Universidad del Valle, 51(2): 42-77. Recuperado de http://doi.org/10.25100/cm.v51i2.4277

[3] Castillo, M. y Wilches, J. (2020). "Aproximación matemática del modelo epidemiológico SIR para la comprensión de las medidas de contención contra la COVID19". Esp salud pública, Universidad del Magdalena, 94: 1-11. Recuperado de: https://medes.com/publication/154923.

[4] Córdova, D. y Santa, F. (2021). "Aplicación del método autorregresivo integrado de medias móviles para el análisis de series de casos de COVID19 en Perú". Facultad de medicina humana URP, Universidad Ricardo Palma, 21(1). DOI: 10.25176/RFMH.v21i1.3307.

[5] Córdova, D. y Santa, F. (2020). " Precisión del pronóstico de la dinámica de propagación de la COVID-19 en Perú". Horizonte médico, Universidad San Martín de Porres, 20(3). Recuperado de https://dx.doi.org/10.24265/horizmed.2020.v20n3.06

[6] Díaz, J. (2020). "Perspectiva del COVID-19 en Colombia para el año 2021". Repertortio de medicina y cirugía, Fundación Universitaria de ciencias de la salud, 29(1): 128-133. DOI:10.31260/RepertMedCir.01217372.1136.

[7] Díaz, J. (2020). "Precisión del pronóstico de la propagación del COVID-19 en Colombia". Repertortio de medicina y cirugía, Fundación Universitaria de ciencias de la salud, 29(1): 27-33. DOI: 10.31260/RepertMedCir.01217372.1045.

[8] Franco, O. Gómez, A. Gómez, C. González, N. González,V. Palacio, C. (2020). "Proyecciones de impacto de la pandemia COVID-19 en la población colombiana, según medidas de mitigación. Datos preliminares de modelos epidemiológicos para el periodo del 18 de marzo al 18 de abril de 2020". Salud Pública, Universidad Nacional, 22(2):1-6. DOI: https://doi.org/10.15446/rsap.V22n2.85789

[9] Galton, F. (1889). *Natural inheritance*. Londres: McMillan y Co.

[10] Galvis, D. García, M. Hurtado L. y Salcedo, G. (2006). *Estadística básica.* Armenia: Conceptos gráficos Ltda.

[11] Gonzáles, H. Ticona, U. (2019). "Clustering, mediterraneidad y comercio internacional: aplicación empírica de los algoritmos Partitioning Around Medoids y K-means". Revista latinoamericana de desarrollo económico, Universidad Católica Boliviana San Pablo, 32(1):95-129. Recuperado de: http://www.scielo.org.bo/scielo.php?pid=S207447062019000200005&script=sci_arttext

[12] Gutiérrez, E. Hernández, F. y Puche, R. (2020). "Estimación de casos de COVID-19 en Promedio Móvil". Revista Especializada en Gestión Social del Conocimiento, Observatorio Nacional de Ciencia, Tecnología e Innovación, 5(3): 11-25. Recuperado de https://docs.bvsalud.org/biblioref/2020/09/1119237/art1-gutierrez.pdf

[13] Herazo, B. (2010). "Algunos problemas de la salud en Colombia". Universitas Odontológica, Pontificia Universidad Javeriana, 29(63): 37-39. Recuperado en https://dialnet.unirioja.es/servlet/articulo?codigo=3986883

[14] Hernández, C. y Montesinos, O. (2007). "Modelos matemáticos para enfermedades infecciosas". Salud pública de Mexico, Universidad de Colima, 49(3). Recuperado de https://www.scielosp.org/pdf/spm/2007.v49n3/218-226

[15] Hurtado, L. y Salcedo, G. (1999). *Series temporales con aplicaciones a la epidemiología y a la ecología.* Armenia: Conceptos gráficos Ltda.

[16] Morettin, P. y Toloi, C. (2006). *Análise de séries temporais. Associaçao Brasileira de Estadística, projeto Fisher.* São Paulo: Edgard Blücher Ltda.

[17] Ocaña, R. (2004). "Eficacia del análisis de series temporales para la planificación sanitaria del cancer en España". Atención primaria, Sociedad Española de Medicina de Familia y Comunitaria, 34(1): 15-19. DOI: https://doi.org/10.1016/S0212-6567(04)79445-7.

[18] Yang, C. y Wang, J. (2020). "A mathematical model for the novel coronavirus epidemic in Wuhan, China". Math Biosci Eng, Departament of Health y Human Services USA, 17(3): 65-74. DOI: 10.3934/mbe.2020148.

7. Anexos

# ENVÍO	FECHA	FARMACÉUTICA	CANTIDAD
1	feb-15	Pfizer	50.310
2	20-feb	Sinovac	192.000
3	24-feb	Pfizer	50.310
4	mar-01	Pfizer (Covax)	117.000
5	mar-03	Pfizer	99.450
6	mar-06	Sinovac	958.764
7	mar-07	Sinovac	574.880
8	mar-10	Pfizer	99.450
9	mar-17	Pfizer	100.620
10	mar-20	AstraZeneca (Covax)	244.800
11	mar-20	Sinovac	774.320
12	mar-24	Pfizer	23.400
13	mar-25	Pfizer	77.220
14	31-mar	Pfizer	100.620
15	abr-03	Pfizer	280.800
16	abr-09	Pfizer	270.270
17	abr-11	Sinovac	500.000
18	abr-12	Pfizer	551.070
19	abr-21	Pfizer	549.900
20	abr-25	AstraZeneca (Covax)	912.000
21	abr-27	Sinovac	1.000.000
22	abr-29	Pfizer	549.900
23	may-01	Sinovac	1.000.000
24	may-05	Pfizer	391.950
25	may-09	Sinovac	1.000.040
26	may-12	Pfizer	391.950
27	may-16	Sinovac	500.000
28	may-17	Pfizer (Covax)	546.390
29	may-19	Pfizer	391.950
30	may-23	Sinovac	1.000.000
31	may-24	Pfizer (Covax)	546.390
32	may-26	Pfizer	394.290
33	jun-02	Pfizer	538.200
34	jun-03	AstraZeneca (Covax)	909.600
35	jun-03	Pfizer	539.370
36	jun-06	Sinovac	1.000.000
37	jun-09	Pfizer	538.200
38	jun-10	Pfizer	539.370
39	jun-13	AstraZeneca	204.000
40	jun-16	Pfizer	538.200
41	jun-17	Pfizer	539.370
42	jun-19	Astrazeneca	126.900
43	jun-23	Pfizer	539.370
44	jun-23	Janssen	480.000
45	jun-24	Pfizer	539.370
46	jun-26	Sinovac (400.000 para privados)	1.900.000
47	jun-27	Sinovac para privados	1.100.000
48	jun-29	Pfizer	540.540
49	jun-30	Pfizer	539.370
50	jul-01	Janssen (Donación EEUU)	2.500.000
51	jul-07	Pfizer	104.130
52	jul-10	Sinovac	1.000.000
53	jul-11	Astrazeneca	408.000
54	jul-11	Sinovac	1.000.000
55	jul-15	Pfizer	132.210
56	jul-17	Astrazeneca	326.400
57	jul-18	Astrazeneca	139.800
58	jul-18	Sinovac para privados	600.000
59	jul-21	Pfizer	308.880
60	jul-25	Moderna	3.500.000
61	jul-28	Pfizer	443.430
62	ago-04	Pfizer	524.160
63	ago-08	Astrazeneca	204.000
64	ago-11	Pfizer	394.290
65	ago-11	Astrazeneca	591.100
66	ago-15	Moderna	150.000
67	ago-18	Pfizer	372.060
68	ago-19	Astrazeneca	163.200
69	ago-25	Pfizer	381.420
70	ago-31	Pfizer	620.100
71	sep-02	Astrazeneca	408.000
72	sep-08	Pfizer	339.300
73	sep-08	Astrazeneca	752.400
74	sep-13	Sinovac (COVAX)	2.097.600
75	sep-13	Astrazeneca (COVAX) Donación España	957.600
76	sep-15	Pfizer	339.300
77	sep-19	Moderna	689.220
78	sep-22	Pfizer	339.300
79	sep-23	Astrazeneca	486.600
80	sep-24	Janssen	1.249.200
81	sep-25	Janssen	1.585.500
		TOTAL A 25 DE SEPTIEMBRE	48.459.104

Figura 14: Número de vacunas que han llegado a Colombia

Modelo	RMSE	MAE	MAPE
Exponencial +			
SARIMA (0,1,1)*(0,1,1)7	961	539	
SARIMA (0,1,1)*(1,0,1)7	941	512	
SARIMA (0,1,1)*(1,1,0)7	990	567	
SARIMA (0,1,1)*(0,1,0)7	1147	603	
SARIMA (0,1,1)*(1,0,0)7	954	513	
SARIMA (1,0,1)*(0,1,1)7	973	859	
SARIMA (1,0,1)*(1,0,1)7	855	486	
SARIMA (1,0,1)*(1,1,0)7	919	536	
SARIMA (1,0,1)*(0,1,0)7	1109	571	
SARIMA (1,1,0)*(0,1,1)7	974	543	
SARIMA (1,1,0)*(1,0,1)7	954	515	
SARIMA (1,1,0)*(1,1,0)7	1038	590	
SARIMA (1,1,0)*(0,1,0)7	1246	659	
SARIMA (0,1,0)*(0,1,1)7	1013	571	
SARIMA (0,1,0)*(1,1,0)7	1120	642	
SARIMA (0,1,0)*(0,1,0)7	1350	730	
SARIMA (0,0,1)*(1,1,1)7	1504	813	
SARIMA (0,0,1)*(1,1,0)7	1634	855	
SARIMA (0,0,1)*(0,1,0)7	1655	853	
SARIMA (0,0,1)*(0,0,1)7	1899	899	
SARIMA (0,0,1)*(1,0,0)7	1510	793	
SARIMA (1,0,0)*(0,1,1)7	1011	654	
SARIMA (1,0,0)*(1,1,0)7	1116	641	
SARIMA (1,0,0)*(0,1,0)7	1299	682	
SARIMA (0,1,1)*(0,1,1)14	1004	588	
SARIMA (0,1,1)*(1,1,0)14	1046	593	
SARIMA (0,1,1)*(0,1,0)14	1103	643	
SARIMA (0,1,1)*(0,0,1)14	950	517	
SARIMA (0,1,1)*(1,0,0)14	949	516	
SARIMA (1,0,1)*(0,1,1)14	897	531	
SARIMA (1,0,1)*(1,1,0)14	934	551	
SARIMA (1,0,1)*(0,1,0)14	1011	595	
SARIMA (1,0,1)*(0,0,1)14	863	480	
SARIMA (1,0,1)*(1,0,0)14	860	481	
SARIMA (1,1,0)*(0,1,1)14	1009	594	
SARIMA (1,1,0)*(1,1,0)14	1059	592	
SARIMA (1,1,0)*(0,1,0)14	1148	459	
SARIMA (1,1,0)*(1,0,0)14	957	522	
SARIMA (0,1,0)*(0,1,1)14	1047	628	
SARIMA (0,1,0)*(1,1,0)14	1131	635	
SARIMA (0,1,0)*(0,1,0)14	1220	698	
SARIMA (0,0,1)*(0,1,1)14	2056	1064	
SARIMA (0,0,1)*(1,0,1)14	1976	983	
SARIMA (0,0,1)*(1,1,0)14	1978	1106	
SARIMA (0,0,1)*(0,1,0)14	2154	1062	

Cuadro 7: Modelos ajustados al clúster 1 con sus respectivos valores de RMSE y MAE

ARIMA (1,0,0)	935	488	
SARIMA			
SARIMA (0,1,1)*(0,1,1)7	835	454	
SARIMA (0,1,1)*(1,0,1)7	817	423	
SARIMA (0,1,1)*(1,1,0)7	898	500	
SARIMA (0,1,1)*(0,1,0)7	1069	547	
SARIMA (1,0,1)*(0,1,1)7	836	449	
SARIMA (1,0,1)*(1,0,1)7	802	429	
SARIMA (1,0,1)*(1,1,0)7	895	496	
SARIMA (1,0,1)*(0,1,0)7	1062	549	
SARIMA (1,1,0)*(0,1,1)7	896	504	
SARIMA (1,1,0)*(1,1,0)7	944	551	
SARIMA (1,1,0)*(0,1,0)7	1210	626	
SARIMA (0,1,0)*(0,1,1)7	958	537	
SARIMA (0,1,0)*(1,1,0)7	1084	601	
SARIMA (0,1,0)*(0,1,0)7	1314	695	
SARIMA (0,0,1)*(0,1,1)7	1128	630	
SARIMA (0,0,1)*(1,0,1)7	1078	583	
SARIMA (0,0,1)*(1,1,0)7	1117	631	
SARIMA (0,0,1)*(0,1,0)7	1143	637	
SARIMA (0,0,1)*(0,0,1)7	2185	1224	
SARIMA (0,0,1)*(1,0,0)7	1117	603	
SARIMA (1,0,0)*(0,1,1)7	951	523	
SARIMA (1,0,0)*(1,1,0)7	1000	550	
SARIMA (1,0,0)*(0,1,0)7	1111	610	
SARIMA (0,1,1)*(0,1,1)14	859	496	
SARIMA (0,1,1)*(1,1,0)14	925	529	
SARIMA (0,1,1)*(0,1,0)14	994	563	
SARIMA (0,1,1)*(0,0,1)14	818	429	
SARIMA (0,1,1)*(1,0,0)14	816	431	
SARIMA (1,0,1)*(0,1,1)14	861	494	
SARIMA (1,0,1)*(1,1,0)14	922	523	
SARIMA (1,0,1)*(0,1,0)14	991	555	
SARIMA (1,0,1)*(0,0,1)14	815	428	
SARIMA (1,0,1)*(1,0,0)14	814	430	
SARIMA (1,1,0)*(0,1,1)14	918	564	
SARIMA (1,1,0)*(1,1,0)14	983	547	
SARIMA (1,1,0)*(0,1,0)14	1094	605	
SARIMA (0,1,0)*(0,1,1)14	986	581	
SARIMA (0,1,0)*(1,1,0)14	1080	594	
SARIMA (0,1,0)*(0,1,0)14	1177	662	
SARIMA (0,0,1)*(0,1,1)14	1304	784	
SARIMA (0,0,1)*(1,1,0)14	1288	774	
SARIMA (0,0,1)*(0,1,0)14	1339	807	
SARIMA (0,0,1)*(0,0,1)14	2088	1333	
SARIMA (0,0,1)*(1,0,0)14	1231	714	
SARIMA (1,0,0)*(0,1,1)14	981	774	

Cuadro 8: Modelos ajustados al clúster 1 con sus respectivos valores de RMSE y MAE

SARIMA (1,0,0)*(0,0,1)7	947	543	
SARIMA (1,0,0)*(1,0,0)7	931	525	
SARIMA (0,1,1)*(1,1,1)14	895	533	
SARIMA (0,1,1)*(0,1,1)14	906	527	
SARIMA (0,1,1)*(1,1,0)14	945	540	
SARIMA (0,1,1)*(0,1,0)14	1004	572	
SARIMA (0,1,1)*(0,0,1)14	865	466	
SARIMA (0,1,1)*(1,0,0)14	860	477	
SARIMA (1,0,1)*(0,1,1)14	906	528	
SARIMA (1,0,1)*(1,1,0)14	945	538	
SARIMA (1,0,1)*(0,1,0)14	1003	559	
SARIMA (1,0,1)*(0,0,1)14	865	465	
SARIMA (1,0,1)*(1,0,0)14	860	465	
SARIMA (1,1,0)*(0,1,1)14	969	547	
SARIMA (1,1,0)*(1,1,0)14	104	558	
SARIMA (1,1,0)*(0,1,0)14	1103	957	
SARIMA (1,1,0)*(0,0,1)14	945	948	
SARIMA (1,1,0)*(1,0,0)14	963	498	
SARIMA (0,1,0)*(0,1,1)14	1073	610	
SARIMA (0,1,0)*(1,0,1)14	885	632	
SARIMA (0,1,0)*(1,1,0)14	1117	624	
SARIMA (0,1,0)*(0,1,0)14	1196	661	
SARIMA (0,1,0)*(0,0,1)14	1054	575	
SARIMA (0,1,0)*(1,0,0)14	1042	564	
SARIMA (0,0,1)*(1,1,1)14	1060	665	
SARIMA (0,0,1)*(0,1,1)14	1072	681	
SARIMA (0,0,1)*(1,1,0)14	1160	696	
SARIMA (0,0,1)*(0,1,0)14	1177	687	
SARIMA (0,0,1)*(0,0,1)14	991	686	
SARIMA (0,0,1)*(1,0,0)14	1000	597	
SARIMA (1,0,0)*(0,1,1)14	981	595	
SARIMA (1,0,0)*(1,1,0)14	1035	619	
SARIMA (1,0,0)*(0,1,0)14	1085	630	
SARIMA (1,0,0)*(0,0,1)14	935	537	
SARIMA (1,0,0)*(1,0,0)14	934	540	
ARIMA (0,1,1)	906	493	
ARIMA (1,0,1)	904	494	
ARIMA (1,1,0)	993	516	
ARIMA (0,1,0)	1116	603	
ARIMA (0,0,1)	1033	632	
ARIMA (1,0,0)	981	577	
ARIMA			
ARIMA (0,1,1)	832	440	
ARIMA (1,0,1)	826	432	
ARIMA (1,1,0)	882	478	
ARIMA (0,1,0)	936	491	
ARIMA (0,0,1)	3073	1939	

Cuadro 9: Modelos ajustados al clúster 1 con sus respectivos valores de RMSE y MAE

ARIMA (1,0,0)	935	488	
SARIMA			
SARIMA (0,1,1)*(0,1,1)7	835	454	
SARIMA (0,1,1)*(1,0,1)7	817	423	
SARIMA (0,1,1)*(1,1,0)7	898	500	
SARIMA (0,1,1)*(0,1,0)7	1069	547	
SARIMA (1,0,1)*(0,1,1)7	836	449	
SARIMA (1,0,1)*(1,0,1)7	802	429	
SARIMA (1,0,1)*(1,1,0)7	895	496	
SARIMA (1,0,1)*(0,1,0)7	1062	549	
SARIMA (1,1,0)*(0,1,1)7	896	504	
SARIMA (1,1,0)*(1,1,0)7	944	551	
SARIMA (1,1,0)*(0,1,0)7	1210	626	
SARIMA (0,1,0)*(0,1,1)7	958	537	
SARIMA (0,1,0)*(1,1,0)7	1084	601	
SARIMA (0,1,0)*(0,1,0)7	1314	695	
SARIMA (0,0,1)*(0,1,1)7	1128	630	
SARIMA (0,0,1)*(1,0,1)7	1078	583	
SARIMA (0,0,1)*(1,1,0)7	1117	631	
SARIMA (0,0,1)*(0,1,0)7	1143	637	
SARIMA (0,0,1)*(0,0,1)7	2185	1224	
SARIMA (0,0,1)*(1,0,0)7	1117	603	
SARIMA (1,0,0)*(0,1,1)7	951	523	
SARIMA (1,0,0)*(1,1,0)7	1000	550	
SARIMA (1,0,0)*(0,1,0)7	1111	610	
SARIMA (0,1,1)*(0,1,1)14	859	496	
SARIMA (0,1,1)*(1,1,0)14	925	529	
SARIMA (0,1,1)*(0,1,0)14	994	563	
SARIMA (0,1,1)*(0,0,1)14	818	429	
SARIMA (0,1,1)*(1,0,0)14	816	431	
SARIMA (1,0,1)*(0,1,1)14	861	494	
SARIMA (1,0,1)*(1,1,0)14	922	523	
SARIMA (1,0,1)*(0,1,0)14	991	555	
SARIMA (1,0,1)*(0,0,1)14	815	428	
SARIMA (1,0,1)*(1,0,0)14	814	430	
SARIMA (1,1,0)*(0,1,1)14	918	564	
SARIMA (1,1,0)*(1,1,0)14	983	547	
SARIMA (1,1,0)*(0,1,0)14	1094	605	
SARIMA (0,1,0)*(0,1,1)14	986	581	
SARIMA (0,1,0)*(1,1,0)14	1080	594	
SARIMA (0,1,0)*(0,1,0)14	1177	662	
SARIMA (0,0,1)*(0,1,1)14	1304	784	
SARIMA (0,0,1)*(1,1,0)14	1288	774	
SARIMA (0,0,1)*(0,1,0)14	1339	807	
SARIMA (0,0,1)*(0,0,1)14	2088	1333	
SARIMA (0,0,1)*(1,0,0)14	1231	714	
SARIMA (1,0,0)*(0,1,1)14	981	774	

Cuadro 10: Modelos ajustados al clúster 1 con sus respectivos valores de RMSE y MAE

SARIMA (1,0,0)*(1,0,1)14	981	574	
SARIMA (1,0,0)*(0,1,0)14	1048	594	
SARIMA (1,0,0)*(0,0,1)14	1104	621	

Cuadro 11: Modelos ajustados al clúster 1 con sus respectivos valores de RMSE y MAE

Modelo	RMSE	MAE	MAPE
Exponencial+			
SARIMA (1,1,1)*(1,1,1)7	1196	872	
SARIMA (1,1,1)*(0,1,1)7	1191	874	
SARIMA (1,1,1)*(1,0,1)7	1124	821	
SARIMA (1,1,1)*(1,0,0)7	1223	920	
SARIMA (0,1,1)*(0,1,1)7	1195	835	
SARIMA (0,1,1)*(1,0,1)7	1146	813	
SARIMA (0,1,1)*(1,1,0)7	1308	953	
SARIMA (0,1,1)*(0,1,0)7	1493	1077	
SARIMA (0,1,1)*(0,0,1)7	1195	835	
SARIMA (0,1,1)*(1,0,0)7	1273	919	
SARIMA (1,0,1)*(0,1,1)7	1165	836	
SARIMA (1,0,1)*(1,0,1)7	1098	792	
SARIMA (1,0,1)*(1,1,0)7	1261	948	
SARIMA (1,0,1)*(0,1,0)7	1420	1070	
SARIMA (1,1,0)*(0,1,1)7	1205	851	
SARIMA (1,1,0)*(1,0,1)7	1165	835	
SARIMA (1,1,0)*(1,1,0)7	1316	951	
SARIMA (1,1,0)*(0,1,0)7	1502	1077	
SARIMA (1,1,0)*(0,0,1)7	1302	946	
SARIMA (1,1,0)*(1,0,0)7	1280	925	
SARIMA (0,1,0)*(0,1,1)7	1262	917	
SARIMA (0,1,0)*(1,0,1)7	1217	904	
SARIMA (0,1,0)*(1,1,0)7	1399	1030	
SARIMA (0,1,0)*(0,1,0)7	1606	1209	
SARIMA (0,1,0)*(0,0,1)7	1320	980	
SARIMA (0,1,0)*(1,0,0)7	1309	968	
SARIMA (0,0,1)*(1,1,1)7	1336	1024	
SARIMA (0,0,1)*(0,1,1)7	1353	1011	
SARIMA (0,0,1)*(1,1,0)7	1456	1156	
SARIMA (0,0,1)*(0,1,0)7	1574	1254	
SARIMA (0,0,1)*(0,0,1)7	1333	1005	
SARIMA (0,0,1)*(1,0,0)7	1326	1010	
SARIMA (1,0,0)*(0,1,1)7	1181	870	
SARIMA (1,0,0)*(1,0,1)7	1172	871	
SARIMA (1,0,0)*(1,1,0)7	1284	982	
SARIMA (1,0,0)*(0,1,0)7	1439	1124	
SARIMA (1,0,0)*(0,0,1)7	1217	902	
SARIMA (1,0,0)*(1,0,0)7	1203	897	
SARIMA (1,1,1)*(0,1,1)14	1232	885	
SARIMA (1,1,1)*(1,0,1)14	1090	802	
SARIMA (1,1,1)*(0,0,1)14	1242	898	
SARIMA (1,1,1)*(1,0,0)14	1233	900	
SARIMA (0,1,1)*(0,1,1)14	1238	862	
SARIMA (0,1,1)*(1,0,1)14	1071	784	
SARIMA (0,1,1)*(1,1,0)14	1378	1017	
SARIMA (0,1,1)*(0,1,0)14	1483	1154	
SARIMA (0,1,1)*(0,0,1)14	1286	903	
SARIMA (0,1,1)*(1,0,0)14	1265	887	

Cuadro 12: Modelos ajustados al clúster 2 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (1,0,1)*(0,1,1)14	1209	866	
SARIMA (1,0,1)*(1,0,1)14	1058	783	
SARIMA (1,0,1)*(1,1,0)14	1343	1019	
SARIMA (1,0,1)*(0,1,0)14	1439	1132	
SARIMA (1,1,0)*(0,1,1)14	1242	878	
SARIMA (1,1,0)*(1,0,1)14	1084	793	
SARIMA (1,1,0)*(1,1,0)14	1383	1026	
SARIMA (1,1,0)*(0,1,0)14	1491	1150	
SARIMA (1,1,0)*(0,0,1)14	1291	911	
SARIMA (1,1,0)*(1,0,0)14	1271	892	
SARIMA (0,1,0)*(0,1,1)14	1294	930	
SARIMA (0,1,0)*(1,0,1)14	1125	861	
SARIMA (0,1,0)*(1,1,0)14	1450	1088	
SARIMA (0,1,0)*(0,1,0)14	1588	1216	
SARIMA (0,1,0)*(0,0,1)14	1312	945	
SARIMA (0,1,0)*(1,0,0)14	1302	937	
SARIMA (0,0,1)*(1,1,1)14	1451	1065	
SARIMA (0,0,1)*(0,1,1)14	1444	1065	
SARIMA (0,0,1)*(1,0,1)14	1394	1026	
SARIMA (0,0,1)*(1,1,0)14	1643	1243	
SARIMA (0,0,1)*(0,1,0)14	1737	1322	
SARIMA (0,0,1)*(0,0,1)14	1389	1027	
SARIMA (1,0,0)*(0,1,1)14	1227	887	
SARIMA (1,0,0)*(1,0,1)14	1054	792	
SARIMA (1,0,0)*(1,1,0)14	1366	1047	
SARIMA (1,0,0)*(0,1,0)14	1475	1161	
SARIMA (1,0,0)*(0,0,1)14	1227	891	
SARIMA (1,0,0)*(1,0,0)14	1220	891	
ARIMA (1,1,1)	1272	927	
ARIMA (0,1,1)	1334	997	
ARIMA (0,1,0)	1342	996	
ARIMA (0,0,1)	1394	1047	
ARIMA (1,0,0)	1255	912	
Cuadrático +			
SARIMA (1,1,1)*(0,1,1)7	1170	865	
SARIMA (1,1,1)*(1,0,1)7	1117	823	
SARIMA (1,1,1)*(0,0,1)7	1207	892	
SARIMA (1,1,1)*(1,0,0)7	1196	886	
SARIMA (0,1,1)*(0,1,1)7	1190	832	
SARIMA (0,1,1)*(1,0,1)7	1117	800	
SARIMA (0,1,1)*(1,1,0)7	1308	952	
SARIMA (0,1,1)*(0,1,0)7	1493	1078	
SARIMA (0,1,1)*(0,0,1)7	1294	938	
SARIMA (0,1,1)*(1,0,0)7	1271	916	
SARIMA (1,0,1)*(1,0,1)7	1072	792	
SARIMA (1,0,1)*(1,1,0)7	1250	946	
SARIMA (1,1,0)*(0,1,1)7	1202	848	
SARIMA (1,1,0)*(1,0,1)7	1138	817	
SARIMA (1,1,0)*(1,1,0)7	1316	950	
SARIMA (1,1,0)*(0,1,0)7	1503	1077	

Cuadro 13: Modelos ajustados al clúster 2 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (1,1,0)*(0,0,1)7	1299	943	
SARIMA (1,1,0)*(1,0,0)7	1279	922	
SARIMA (0,1,0)*(0,1,1)7	1260	913	
SARIMA (0,1,0)*(1,0,1)7	1183	900	
SARIMA (0,1,0)*(1,1,0)7	1399	1029	
SARIMA (0,1,0)*(0,1,0)7	1606	1209	
SARIMA (0,1,0)*(0,0,1)7	1318	977	
SARIMA (0,1,0)*(1,0,0)7	1307	964	
SARIMA (0,0,1)*(0,1,1)7	1259	966	
SARIMA (0,0,1)*(1,0,1)7	1255	958	
SARIMA (0,0,1)*(1,1,0)7	1409	1113	
SARIMA (0,0,1)*(0,1,0)7	1554	1228	
SARIMA (0,0,1)*(0,0,1)7	1265	966	
SARIMA (0,0,1)*(1,0,0)7	1260	967	
SARIMA (1,0,0)*(1,1,1)7	1149	849	
SARIMA (1,0,0)*(0,1,1)7	1146	845	
SARIMA (1,0,0)*(1,0,1)7	1086	809	
SARIMA (1,0,0)*(1,1,0)7	1267	968	
SARIMA (1,0,0)*(0,1,0)7	1431	1113	
SARIMA (1,0,0)*(0,0,1)7	1187	879	
SARIMA (1,0,0)*(1,0,0)7	1176	871	
SARIMA (1,1,1)*(0,1,1)14	1219	891	
SARIMA (1,1,1)*(1,0,1)14	1070	803	
SARIMA (1,1,1)*(1,1,0)14	1369	1045	
SARIMA (1,1,1)*(0,0,1)14	1211	885	
SARIMA (1,1,1)*(1,0,0)14	1204	887	
SARIMA (0,1,1)*(0,1,1)14	1234	855	
SARIMA (0,1,1)*(1,0,1)14	1068	778	
SARIMA (0,1,1)*(1,1,0)14	1377	1012	
SARIMA (0,1,1)*(0,1,0)14	1482	1152	
SARIMA (0,1,1)*(0,0,1)14	1283	899	
SARIMA (0,1,1)*(1,0,0)14	1263	881	
SARIMA (1,0,1)*(1,1,1)14	1183	845	
SARIMA (1,0,1)*(0,1,1)14	1179	844	
SARIMA (1,0,1)*(1,0,1)14	1014	764	
SARIMA (1,0,1)*(1,1,0)14	1324	1018	
SARIMA (1,0,1)*(0,1,0)14	1426	1127	
SARIMA (1,0,1)*(0,0,1)14	1196	871	
SARIMA (1,1,0)*(1,1,1)14	1243	872	
SARIMA (1,1,0)*(0,1,1)14	1239	871	
SARIMA (1,1,0)*(1,0,1)14	1083	790	
SARIMA (1,1,0)*(1,1,0)14	1382	1022	
SARIMA (1,1,0)*(0,1,0)14	1490	1149	
SARIMA (1,1,0)*(0,0,1)14	1289	908	
SARIMA (1,1,0)*(1,0,0)14	1269	888	
SARIMA (0,1,0)*(1,1,1)14	1296	925	
SARIMA (0,1,0)*(0,1,1)14	1292	928	
SARIMA (0,1,0)*(1,0,1)14	1132	867	
SARIMA (0,1,0)*(1,1,0)14	1450	1086	
SARIMA (0,1,0)*(0,1,0)14	1588	1214	

Cuadro 14: Modelos ajustados al clúster 2 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (0,1,0)*(0,0,1)14	1310	943	
SARIMA (0,1,0)*(1,0,0)14	1300	935	
SARIMA (0,0,1)*(0,1,1)14	1323	1008	
SARIMA (0,0,1)*(1,1,0)14	1540	1200	
SARIMA (0,0,1)*(0,1,0)14	1651	1263	
SARIMA (1,0,0)*(0,1,1)14	1187	862	
SARIMA (1,0,0)*(1,0,1)14	1049	786	
SARIMA (1,0,0)*(1,1,0)14	1339	1042	
SARIMA (1,0,0)*(0,1,0)14	1452	1140	
SARIMA (1,0,0)*(0,0,1)14	1191	872	
SARIMA (1,0,0)*(1,0,0)14	1186	876	
ARIMA (1,1,1)	1237	905	
ARIMA (0,1,1)	1331	970	
ARIMA (0,1,0)	1340	993	
ARIMA (0,0,1)	1296	987	
ARIMA (1,0,0)	1216	890	
ARIMA			
ARIMA (1,1,1)	1283	951	11
ARIMA (0,1,1)	1336	983	12
ARIMA (1,0,1)	1337	977	12
ARIMA (0,1,0)	1344	999	12
ARIMA (0,0,1)	4913	4414	51
ARIMA (1,0,0)	1344	992	12
SARIMA			
SARIMA (1,1,1)*(0,1,1)7	1192	873	10
SARIMA (1,1,1)*(1,0,1)7	1122	817	10
SARIMA (1,1,1)*(0,0,1)7	1250	926	11
SARIMA (1,1,1)*(1,0,0)7	1236	917	11
SARIMA (0,1,1)*(0,1,1)7	1195	835	10
SARIMA (0,1,1)*(1,0,1)7	1120	804	10
SARIMA (0,1,1)*(1,1,0)7	1308	953	11
SARIMA (0,1,1)*(0,1,0)7	1493	1077	13
SARIMA (0,1,1)*(0,0,1)7	1298	944	11
SARIMA (0,1,1)*(1,0,0)7	1274	919	11
SARIMA (1,0,1)*(0,1,1)7	1187	843	10
SARIMA (1,0,1)*(1,0,1)7	1148	817	10
SARIMA (1,0,1)*(1,1,0)7	1266	946	11
SARIMA (1,0,1)*(0,0,1)7	1298	938	11
SARIMA (1,0,1)*(1,0,0)7	1274	912	11
SARIMA (1,1,0)*(0,1,1)7	1205	852	10
SARIMA (1,1,0)*(1,0,1)7	1148	821	10
SARIMA (1,1,0)*(1,1,0)7	1316	951	11
SARIMA (1,1,0)*(0,1,0)7	1503	1077	13
SARIMA (1,1,0)*(0,0,1)7	1303	949	11
SARIMA (1,1,0)*(1,0,0)7	1281	925	11
SARIMA (0,1,0)*(0,1,1)7	1263	918	11
SARIMA (0,1,0)*(1,0,1)7	1215	904	11
SARIMA (0,1,0)*(1,1,0)7	1399	1030	12
SARIMA (0,1,0)*(0,1,0)7	1606	1209	15
SARIMA (0,1,0)*(0,0,1)7	1321	982	12

Cuadro 15: Modelos ajustados al clúster 2 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (0,1,0)*(1,0,0)7	1310	971	12
SARIMA (0,0,1)*(0,1,1)7	1452	1122	13
SARIMA (0,0,1)*(1,0,1)7	1385	1080	13
SARIMA (0,0,1)*(1,1,0)7	1485	1162	13
SARIMA (0,0,1)*(0,1,0)7	1584	1263	15
SARIMA (0,0,1)*(0,0,1)7	3215	2715	31
SARIMA (0,0,1)*(1,0,0)7	1546	1195	14
SARIMA (1,0,0)*(0,1,1)7	1223	912	11
SARIMA (1,0,0)*(1,0,1)7	1183	881	10
SARIMA (1,0,0)*(1,1,0)7	1293	980	12
SARIMA (1,0,0)*(0,1,0)7	1443	1123	14
SARIMA (1,0,0)*(0,0,1)7	1320	977	12
SARIMA (1,0,0)*(1,0,0)7	1308	967	12
SARIMA (1,1,1)*(1,1,1)14	1239	883	10
SARIMA (1,1,1)*(0,1,1)14	1233	885	10
SARIMA (1,1,1)*(1,0,1)14	1173	848	10
SARIMA (1,1,1)*(1,1,0)14	1384	1017	12
SARIMA (1,1,1)*(0,1,0)14	1490	11590	14
SARIMA (1,1,1)*(0,0,1)14	1250	907	11
SARIMA (1,1,1)*(1,0,0)14	1239	906	11
SARIMA (0,1,1)*(0,1,1)14	1238	863	10
SARIMA (0,1,1)*(1,0,1)14	1069	782	10
SARIMA (0,1,1)*(1,1,0)14	1378	1018	12
SARIMA (0,1,1)*(0,1,0)14	1483	1155	14
SARIMA (0,1,1)*(0,0,1)14	1288	908	11
SARIMA (0,1,1)*(1,0,0)14	1266	891	11
SARIMA (1,0,1)*(0,1,1)14	1234	880	10
SARIMA (1,0,1)*(1,0,1)14	1097	793	10
SARIMA (1,0,1)*(1,1,0)14	1353	1022	12
SARIMA (1,0,1)*(0,1,0)14	1447	1132	13
SARIMA (1,0,1)*(0,0,1)14	1287	901	11
SARIMA (1,0,1)*(1,0,0)14	1266	886	10
SARIMA (1,1,0)*(0,1,1)14	1242	878	10
SARIMA (1,1,0)*(1,0,1)14	1088	794	10
SARIMA (1,1,0)*(1,1,0)14	1383	1026	12
SARIMA (1,1,0)*(0,1,0)14	1491	1151	14
SARIMA (1,1,0)*(0,0,1)14	1292	914	11
SARIMA (1,1,0)*(1,0,0)14	1272	895	11
SARIMA (0,1,0)*(0,1,1)14	1295	931	11
SARIMA (0,1,0)*(1,0,1)14	1132	867	11
SARIMA (0,1,0)*(1,1,0)14	1450	1088	13
SARIMA (0,1,0)*(0,1,0)14	1588	1216	15
SARIMA (0,1,0)*(0,0,1)14	1313	948	11
SARIMA (0,1,0)*(1,0,0)14	1303	939	11
SARIMA (0,0,1)*(0,1,1)14	1689	1241	14
SARIMA (0,0,1)*(1,0,1)14	1489	1099	13
SARIMA (0,0,1)*(1,1,0)14	1731	1289	14
SARIMA (0,0,1)*(0,1,0)14	1793	1355	15
SARIMA (0,0,1)*(0,0,1)14	3437	2883	33
SARIMA (0,0,1)*(1,0,0)14	1684	1225	14

Cuadro 16: Modelos ajustados al clúster 2 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (1,0,0)*(0,1,1)14	1271	924	11
SARIMA (1,0,0)*(1,0,1)14	1169	869	10
SARIMA (1,0,0)*(1,1,0)14	1384	1058	12
SARIMA (1,0,0)*(0,1,0)14	1488	1169	14
SARIMA (1,0,0)*(0,0,1)14	1313	945	11
SARIMA (1,0,0)*(1,0,0)14	1301	935	11

Cuadro 17: Modelos ajustados al clúster 2 con sus respectivos valores de RMSE, MAE y MAPE

Modelo	RMSE	MAE	MAPE
Cuadrático +			
SARIMA (1,1,1)*(0,1,1)7	1246	925	
SARIMA (1,1,1)*(1,0,1)7	1338	939	
SARIMA (1,1,1)*(1,1,0)7	1483	1131	
SARIMA (1,1,1)*(0,1,0)7	1648	1234	
SARIMA (1,1,1)*(0,0,1)7	1371	1004	
SARIMA (1,1,1)*(1,0,0)7	1365	993	
SARIMA (1,0,1)*(0,1,0)7	1540	1152	
SARIMA (1,0,1)*(0,0,1)7	1303	955	
SARIMA (1,0,1)*(1,0,0)7	1315	949	
SARIMA (0,1,0)*(0,1,1)7	1250	943	
SARIMA (0,1,0)*(1,0,1)7	1349	947	
SARIMA (0,1,0)*(1,1,0)7	1492	1134	
SARIMA (0,1,0)*(0,1,0)7	1671	1220	
SARIMA (0,1,0)*(0,0,1)7	1390	1019	
SARIMA (0,1,0)*(1,0,0)7	1384	1003	
SARIMA (0,0,1)*(0,1,0)7	1828	1263	
SARIMA (0,0,1)*(0,0,1)7	1470	1076	
SARIMA (0,0,1)*(1,0,0)7	1587	1130	
SARIMA (1,0,0)*(0,1,1)7	1233	921	
SARIMA (1,0,0)*(1,0,1)7	1328	922	
SARIMA (1,0,0)*(1,1,0)7	1456	1102	
SARIMA (1,0,0)*(0,1,0)7	1573	1130	
SARIMA (1,0,0)*(0,0,1)7	1325	977	
SARIMA (1,0,0)*(1,0,0)7	1333	963	
SARIMA (1,1,1)*(0,1,1)14	1195	923	
SARIMA (1,1,1)*(1,0,1)14	1362	984	
SARIMA (1,1,1)*(1,0,0)14	1394	980	
SARIMA (1,0,1)*(1,0,1)14	1282	920	
SARIMA (0,1,0)*(1,1,1)14	1313	1027	
SARIMA (0,1,0)*(0,1,1)14	1196	923	
SARIMA (0,1,0)*(1,0,1)14	1382	995	
SARIMA (0,1,0)*(1,1,0)14	1403	1058	
SARIMA (0,1,0)*(0,1,0)14	1733	1261	
SARIMA (0,1,0)*(1,0,0)14	1416	988	
SARIMA (0,0,1)*(1,1,1)14	1704	1366	
SARIMA (0,0,1)*(0,1,1)14	1637	1185	
SARIMA (0,0,1)*(1,0,1)14	1598	1158	
SARIMA (0,0,1)*(1,1,0)14	1908	1439	
SARIMA (0,0,1)*(0,1,0)14	2330	1596	
SARIMA (1,0,0)*(1,1,1)14	1280	1023	
SARIMA (1,0,0)*(0,1,1)14	1210	868	
SARIMA (1,0,0)*(1,0,1)14	1355	966	
SARIMA (1,0,0)*(1,1,0)14	1395	1044	
SARIMA (1,0,0)*(0,1,0)14	1641	1190	
SARIMA (1,0,0)*(1,0,0)14	1377	962	

Cuadro 18: Modelos ajustados al clúster 3 con sus respectivos valores de RMSE, MAE y MAPE

ARIMA (0,1,0)	1438	1043	
ARIMA (0,0,1)	1724	1218	
ARIMA (1,0,0)	1370	1001	
ARIMA			
ARIMA (0,1,0)	1445	1045	12
ARIMA (0,0,1)	5776	4927	51
ARIMA (1,0,0)	1445	1036	12
SARIMA			
SARIMA (1,1,1)*(0,1,1)7	1271	929	11
SARIMA (1,1,1)*(1,0,1)7	1337	944	11
SARIMA (1,1,1)*(1,1,0)7	1479	1129	14
SARIMA (1,1,1)*(0,1,0)7	1646	1232	14
SARIMA (1,1,1)*(0,0,1)7	1371	1004	12
SARIMA (1,1,1)*(1,0,0)7	1365	988	11
SARIMA (1,0,1)*(0,1,0)7	1571	1169	14
SARIMA (0,1,0)*(0,1,1)7	1278	954	12
SARIMA (0,1,0)*(1,0,1)7	1347	961	11
SARIMA (0,1,0)*(1,1,0)7	1487	1135	13
SARIMA (0,1,0)*(0,1,0)7	1668	1228	14
SARIMA (0,1,0)*(0,0,1)7	1389	1016	12
SARIMA (0,1,0)*(1,0,0)7	1382	1001	11
SARIMA (0,0,1)*(0,1,1)7	1874	1477	17
SARIMA (0,0,1)*(1,0,1)7	1837	1434	16
SARIMA (0,0,1)*(1,1,0)7	1867	1445	17
SARIMA (0,0,1)*(0,1,0)7	1953	1530	18
SARIMA (0,0,1)*(0,0,1)7	3619	2979	32
SARIMA (0,0,1)*(1,0,0)7	1912	1462	17
SARIMA (1,0,0)*(0,1,1)7	1277	943	11
SARIMA (1,0,0)*(1,0,1)7	1341	945	11
SARIMA (1,0,0)*(1,1,0)7	1460	1099	13
SARIMA (1,0,0)*(0,1,0)7	1596	1139	13
SARIMA (1,0,0)*(0,0,1)7	1392	1009	12
SARIMA (1,0,0)*(1,0,0)7	1379	996	11
SARIMA (1,1,1)*(1,1,1)14	1322	927	14
SARIMA (1,1,1)*(0,1,1)14	1241	897	11
SARIMA (1,1,1)*(1,0,1)14	1365	990	11
SARIMA (1,1,1)*(0,0,1)14	1404	998	12
SARIMA (1,1,1)*(1,0,0)14	1391	982	11
SARIMA (0,1,0)*(1,1,1)14	1346	1021	15
SARIMA (0,1,0)*(0,1,1)14	1225	907	11
SARIMA (0,1,0)*(1,0,1)14	1383	1008	12
SARIMA (0,1,0)*(1,1,0)14	1408	1070	16
SARIMA (0,1,0)*(0,1,0)14	1727	1277	16
SARIMA (0,1,0)*(0,0,1)14	1424	1001	12
SARIMA (0,1,0)*(1,0,0)14	1411	988	11
SARIMA (0,0,1)*(0,1,1)14	2813	2031	25
SARIMA (0,0,1)*(1,0,1)14	2633	1914	23

Cuadro 19: Modelos ajustados al clúster 3 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (0,0,1)*(1,1,0)14	2820	1998	24
SARIMA (0,0,1)*(0,1,0)14	2913	2248	28
SARIMA (0,0,1)*(0,0,1)14	3813	3036	32
SARIMA (0,0,1)*(1,0,0)14	2707	2013	24
SARIMA (1,0,0)*(1,1,1)14	1372	1037	15
SARIMA (1,0,0)*(0,1,1)14	1353	1010	13
SARIMA (1,0,0)*(1,0,1)14	1380	994	12
SARIMA (1,0,0)*(1,1,0)14	1556	1152	13
SARIMA (1,0,0)*(0,1,0)14	1851	1288	15
SARIMA (1,0,0)*(1,0,0)14	1423	1014	12

Cuadro 20: Modelos ajustados al clúster 3 con sus respectivos valores de RMSE, MAE y MAPE

Modelo	RMSE	MAE	MAPE
Exponencial +			
SARIMA (1,1,1)*(1,1,1)7	2476	1731	
SARIMA (1,1,1)*(0,1,1)7	2445	1694	
SARIMA (1,1,1)*(1,0,1)7	2298	1760	
SARIMA (1,1,1)*(1,1,0)7	2852	2182	
SARIMA (1,1,1)*(0,1,0)7	3238	2574	
SARIMA (1,1,1)*(0,0,1)7	2585	2029	
SARIMA (1,1,1)*(1,0,0)7	2556	2057	
SARIMA (0,1,1)*(1,0,1)7	2461	1835	
SARIMA (0,1,0)*(0,1,1)7	2458	1750	
SARIMA (0,1,0)*(1,0,1)7	2566	1948	
SARIMA (0,1,0)*(1,1,0)7	2973	2237	
SARIMA (0,1,0)*(0,1,0)7	3510	2818	
SARIMA (0,1,0)*(0,0,1)7	2847	2272	
SARIMA (0,1,0)*(1,0,0)7	2804	2230	
SARIMA (0,0,1)*(0,1,1)7	2483	1875	
SARIMA (0,0,1)*(1,0,1)7	2437	1911	
SARIMA (0,0,1)*(1,1,0)7	2966	2382	
SARIMA (0,0,1)*(0,1,0)7	3277	2606	
SARIMA (0,0,1)*(0,0,1)7	2625	2124	
SARIMA (0,0,1)*(1,0,0)7	2604	2122	
SARIMA (1,0,0)*(0,1,1)7	2287	1642	
SARIMA (1,0,0)*(1,0,1)7	2282	1771	
SARIMA (1,0,0)*(1,1,0)7	2745	2140	
SARIMA (1,0,0)*(0,1,0)7	3129	2495	
SARIMA (1,0,0)*(0,0,1)7	2525	2048	
SARIMA (1,0,0)*(1,0,0)7	2494	2061	
SARIMA (1,1,1)*(1,1,1)14	2633	1893	
SARIMA (1,1,1)*(0,1,1)14	2611	1950	
SARIMA (1,1,1)*(1,0,1)14	2280	1742	
SARIMA (1,1,1)*(1,1,0)14	2781	1782	
SARIMA (1,1,1)*(0,1,0)14	3336	2648	
SARIMA (1,1,1)*(1,0,0)14	2614	1979	
SARIMA (0,1,1)*(1,0,1)14	2335	1773	
SARIMA (0,1,0)*(1,1,1)14	2685	1851	
SARIMA (0,1,0)*(0,1,1)14	2677	1979	
SARIMA (0,1,0)*(1,0,1)14	2361	1860	
SARIMA (0,1,0)*(1,1,0)14	2784	1669	
SARIMA (0,1,0)*(0,1,0)14	3472	2762	
SARIMA (0,1,0)*(1,0,0)14	2786	2161	
SARIMA (0,0,1)*(1,1,1)14	2602	2042	
SARIMA (0,0,1)*(0,1,1)14	2623	1972	
SARIMA (0,0,1)*(1,0,1)14	2365	2840	
SARIMA (0,0,1)*(1,1,0)14	2873	2139	
SARIMA (0,0,1)*(0,1,0)14	3569	2930	
SARIMA (1,0,0)*(1,1,1)14	2511	1810	

Cuadro 21: Modelos ajustados al clúster 4 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (1,0,0)*(0,1,1)14	2485	2836	
SARIMA (1,0,0)*(1,0,1)14	2211	1738	
SARIMA (1,0,0)*(1,1,0)14	2688	1917	
SARIMA (1,0,0)*(0,1,0)14	3217	2570	
SARIMA (1,0,0)*(1,0,0)14	2558	1988	
ARIMA (0,1,0)	2936	2358	
ARIMA (0,0,1)	2711	2189	
ARIMA (1,0,0)	2637	2146	
Cuadrático +			
SARIMA (1,1,1)*(0,1,1)7	2468	1741	
SARIMA (1,1,1)*(1,0,1)7	2265	1781	
SARIMA (1,1,1)*(1,1,0)7	2835	2149	
SARIMA (1,1,1)*(0,1,0)7	3230	2529	
SARIMA (1,1,1)*(0,0,1)7	2515	2058	
SARIMA (1,1,1)*(1,0,0)7	2494	2053	
SARIMA (0,1,1)*(1,0,1)7	2464	1833	
SARIMA (0,1,0)*(0,1,1)7	2470	1768	
SARIMA (0,1,0)*(1,0,1)7	255	1957	
SARIMA (0,1,0)*(1,1,0)7	2974	2230	
SARIMA (0,1,0)*(0,1,0)7	3511	2817	
SARIMA (0,1,0)*(0,0,1)7	2841	2277	
SARIMA (0,1,0)*(1,0,0)7	2801	2236	
SARIMA (0,0,1)*(0,1,1)7	2368	1823	
SARIMA (0,0,1)*(1,0,1)7	2324	1870	
SARIMA (0,0,1)*(1,1,0)7	2883	2245	
SARIMA (0,0,1)*(0,1,0)7	3220	2499	
SARIMA (0,0,1)*(1,0,0)7	2515	2083	
SARIMA (1,0,0)*(0,1,1)7	2276	1680	
SARIMA (1,0,0)*(1,0,1)7	2248	1791	
SARIMA (1,0,0)*(1,1,0)7	2729	2098	
SARIMA (1,0,0)*(0,1,0)7	3113	2444	
SARIMA (1,0,0)*(0,0,1)7	2485	2052	
SARIMA (1,0,0)*(1,0,0)7	2464	2045	
SARIMA (1,1,1)*(1,1,1)14	2562	1768	
SARIMA (1,1,1)*(0,1,1)14	2635	1911	
SARIMA (1,1,1)*(1,0,1)14	2128	1709	
SARIMA (1,1,1)*(1,1,0)14	2631	1721	
SARIMA (1,1,1)*(0,1,0)14	3341	2607	
SARIMA (1,1,1)*(1,0,0)14	2556	1984	
SARIMA (0,1,1)*(1,0,1)14	2321	1770	
SARIMA (0,1,0)*(1,1,1)14	2660	1822	
SARIMA (0,1,0)*(0,1,1)14	2673	1962	
SARIMA (0,1,0)*(1,0,1)14	2352	1856	
SARIMA (0,1,0)*(1,1,0)14	2754	1660	
SARIMA (0,1,0)*(0,1,0)14	3472	2759	
SARIMA (0,1,0)*(1,0,0)14	2785	2182	
SARIMA (0,0,1)*(1,1,1)14	2323	1700	

Cuadro 22: Modelos ajustados al clúster 4 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (0,0,1)*(0,1,1)14	2512	1964	
SARIMA (0,0,1)*(1,0,1)14	2193	1775	
SARIMA (0,0,1)*(1,1,0)14	2489	1638	
SARIMA (0,0,1)*(0,1,0)14	3453	2834	
SARIMA (1,0,0)*(1,1,1)14	2389	1706	
SARIMA (1,0,0)*(0,1,1)14	2484	1831	
SARIMA (1,0,0)*(1,0,1)14	2114	1705	
SARIMA (1,0,0)*(1,1,0)14	2500	1677	
SARIMA (1,0,0)*(0,1,0)14	3010	2540	
ARIMA (0,1,0)	2927	2372	
ARIMA (0,0,1)	2579	2129	
ARIMA (1,0,0)	2572	2092	
ARIMA			
ARIMA (0,1,0)	2945	2385	12,5
ARIMA (0,0,1)	11639	10655	52,05
ARIMA (1,0,0)	2945	2342	12,3
SARIMA			
SARIMA (1,1,1)*(1,0,1)7	2381	1794	9,6
SARIMA (1,1,1)*(1,1,0)7	2854	2194	10,6
SARIMA (1,1,1)*(0,1,0)7	3245	2584	12,5
SARIMA (1,1,1)*(0,0,1)7	2734	2186	11,3
SARIMA (1,1,1)*(1,0,0)7	2676	2169	11,03
SARIMA (0,1,1)*(1,0,1)7	2465	1834	9,7
SARIMA (0,1,0)*(0,1,1)7	2455	1745	8,7
SARIMA (0,1,0)*(1,0,1)7	2564	1943	10,4
SARIMA (0,1,0)*(1,1,0)7	2973	2240	10,9
SARIMA (0,1,0)*(0,1,0)7	3510	2819	13,8
SARIMA (0,1,0)*(0,0,1)7	2851	2281	11,9
SARIMA (0,1,0)*(1,0,0)7	2805	2245	11,6
SARIMA (0,0,1)*(1,0,1)7	3063	2320	11,5
SARIMA (0,0,1)*(1,1,0)7	3335	2570	12,2
SARIMA (0,0,1)*(0,1,0)7	3454	2758	13,2
SARIMA (0,0,1)*(0,0,1)7	7938	6665	32,4
SARIMA (0,0,1)*(1,0,0)7	3227	2476	12,09
SARIMA (1,0,0)*(0,1,1)7	2417	1719	8,6
SARIMA (1,0,0)*(1,0,1)7	2526	1822	9,3
SARIMA (1,0,0)*(1,1,0)7	2817	2150	10,4
SARIMA (1,0,0)*(0,1,0)7	3186	2553	12,3
SARIMA (1,0,0)*(0,0,1)7	2853	2245	11,7
SARIMA (1,0,0)*(1,0,0)7	2799	2224	11,4
SARIMA (1,1,1)*(0,1,1)14	2608	1940	9,2
SARIMA (1,1,1)*(1,0,1)14	2299	1748	9,3
SARIMA (1,1,1)*(1,1,0)14	2812	1776	8,01
SARIMA (1,1,1)*(0,1,0)14	3344	2677	12,5
SARIMA (1,1,1)*(1,0,0)14	2705	2045	10,2
SARIMA (0,1,1)*(1,0,1)14	2335	1775	9,5
SARIMA (0,1,0)*(1,1,1)14	2694	1864	8,8

Cuadro 23: Modelos ajustados al clúster 4 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (0,1,0)*(0,1,1)14	2680	1984	9,5
SARIMA (0,1,0)*(1,0,1)14	2359	1864	10,13
SARIMA (0,1,0)*(1,1,0)14	2794	1683	7,4
SARIMA (0,1,0)*(0,1,0)14	3472	2764	13,03
SARIMA (0,1,0)*(1,0,0)14	2784	2189	11,08
SARIMA (0,0,1)*(1,0,1)14	3562	2840	13,7
SARIMA (0,0,1)*(0,1,0)14	4185	3424	15,7
SARIMA (0,0,1)*(0,0,1)14	7800	6590	32,3
SARIMA (0,0,1)*(1,0,0)14	3543	2712	12,9
SARIMA (1,0,0)*(1,1,1)14	2688	1830	8,7
SARIMA (1,0,0)*(0,1,1)14	2633	1929	9,2
SARIMA (1,0,0)*(1,0,1)14	2629	1917	9,6
SARIMA (1,0,0)*(1,1,0)14	2783	1690	7,5
SARIMA (1,0,0)*(0,1,0)14	3316	2666	12,4
SARIMA (1,0,0)*(1,0,0)14	2765	2144	10,7

Cuadro 24: Modelos ajustados al clúster 4 con sus respectivos valores de RMSE, MAE y MAPE

Modelo	RMSE	MAE	MAPE
Exponencial +			
SARIMA (1,1,1)*(1,1,1)7	1580	1199	
SARIMA (1,1,1)*(0,1,1)7	1487	1157	
SARIMA (1,1,1)*(1,1,0)7	1314	947	
SARIMA (0,1,0)*(1,1,1)7	2625	1239	
SARIMA (0,1,0)*(0,1,1)7	1476	1137	
SARIMA (0,1,0)*(1,1,0)7	1976	1517	
SARIMA (0,1,0)*(0,1,0)7	2378	1703	
SARIMA (0,0,1)*(1,1,1)7	2466	1835	
SARIMA (0,0,1)*(0,1,1)7	2118	1697	
SARIMA (0,0,1)*(0,1,0)7	3400	2373	
SARIMA (1,0,0)*(1,1,1)7	1527	1149	
SARIMA (1,0,0)*(0,1,1)7	1478	1136	
SARIMA (1,0,0)*(1,0,1)7	1767	1265	
SARIMA (1,0,0)*(1,1,0)7	1929	1486	
SARIMA (1,0,0)*(0,1,0)7	2248	1668	
SARIMA (1,1,1)*(0,1,1)14	1610	1149	
SARIMA (0,1,0)*(0,1,1)14	1520	1104	
SARIMA (0,1,0)*(1,1,0)14	2140	1395	
SARIMA (0,1,0)*(0,1,0)14	2354	1961	
SARIMA (0,0,1)*(1,1,1)14	2813	1931	
SARIMA (0,0,1)*(0,1,1)14	2077	1584	
SARIMA (0,0,1)*(1,0,1)14	2067	1413	
SARIMA (0,0,1)*(1,1,0)14	3522	2663	
SARIMA (0,0,1)*(0,1,0)14	3894	2689	
SARIMA (0,0,1)*(0,0,1)14	1890	1480	
SARIMA (0,0,1)*(1,0,0)14	2529	1996	
SARIMA (1,0,0)*(1,1,1)14	1822	1380	
SARIMA (1,0,0)*(0,1,1)14	1496	1091	
SARIMA (1,0,0)*(1,0,1)14	1677	1169	
SARIMA (1,0,0)*(1,1,0)14	2118	1344	
SARIMA (1,0,0)*(0,1,0)14	2227	1840	
ARIMA (0,1,0)	1838	1331	
Cuadrático +			
SARIMA (1,1,1)*(1,1,1)7	1445	1172	
SARIMA (1,1,1)*(1,1,0)7	1944	1459	
SARIMA (0,1,0)*(1,1,1)7	1538	1187	
SARIMA (0,1,0)*(0,1,1)7	1474	1090	
SARIMA (0,1,0)*(1,1,0)7	1927	1484	
SARIMA (0,1,0)*(0,1,0)7	2348	1695	
SARIMA (0,0,1)*(1,1,1)7	1774	1423	
SARIMA (0,0,1)*(0,1,1)7	1604	1259	
SARIMA (0,0,1)*(1,1,0)7	2265	1859	
SARIMA (0,0,1)*(0,1,0)7	2640	2054	
SARIMA (1,0,0)*(1,1,1)7	1462	1090	
SARIMA (1,0,0)*(0,1,1)7	1327	1015	

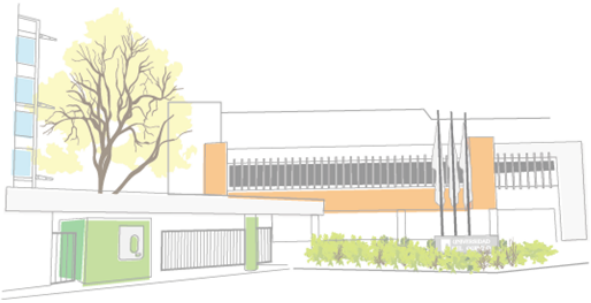
Cuadro 25: Modelos ajustados al clúster 5 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (1,0,0)*(1,1,0)7	1837	1443	
SARIMA (1,0,0)*(0,1,0)7	2184	1643	
SARIMA (1,1,1)*(0,1,1)14	1551	1137	
SARIMA (1,1,1)*(1,1,0)14	1966	1352	
SARIMA (0,1,0)*(0,1,1)14	1514	1093	
SARIMA (0,1,0)*(1,1,0)14	2053	1330	
SARIMA (0,1,0)*(0,1,0)14	2288	1884	
SARIMA (0,0,1)*(1,1,1)14	1843	1395	
SARIMA (0,0,1)*(0,1,1)14	1660	1231	
SARIMA (0,0,1)*(1,0,1)14	1468	1102	
SARIMA (0,0,1)*(1,1,0)14	2319	1778	
SARIMA (0,0,1)*(0,1,0)14	2716	2236	
SARIMA (0,0,1)*(0,0,1)14	1295	1008	
SARIMA (0,0,1)*(1,0,0)14	1720	1300	
SARIMA (1,0,0)*(1,1,1)14	1580	1125	
SARIMA (1,0,0)*(0,1,1)14	1397	922	
SARIMA (1,0,0)*(1,0,1)14	1342	991	
SARIMA (1,0,0)*(1,1,0)14	1937	1295	
SARIMA (1,0,0)*(0,1,0)14	2178	1752	
ARIMA (0,1,0)	1754	1302	
ARIMA (0,0,1)	1868	1422	
ARIMA (1,0,0)	1625	1230	
ARIMA			
ARIMA (0,1,0)	1820	1376	9,2
ARIMA (0,0,1)	10557	965	49,05
ARIMA (1,0,0)	1770	1292	8,5
SARIMA			
SARIMA (1,1,1)*(1,1,1)7	1536	1167	8,9
SARIMA (1,1,1)*(0,1,1)7	1478	1120	8,5
SARIMA (1,1,1)*(1,0,1)7	1722	1245	8,6
SARIMA (1,1,1)*(1,1,0)7	1945	1467	12,2
SARIMA (0,1,0)*(1,1,1)7	1537	1187	9,1
SARIMA (0,1,0)*(0,1,1)7	1464	1987	8,3
SARIMA (0,1,0)*(1,0,1)7	1719	1237	8,5
SARIMA (0,1,0)*(1,1,0)7	1828	1488	12,2
SARIMA (0,1,0)*(0,1,0)7	2350	1696	13,05
SARIMA (0,0,1)*(1,1,1)7	2867	2277	15,7
SARIMA (0,0,1)*(1,0,1)7	2897	2328	16,4
SARIMA (0,0,1)*(1,1,0)7	1317	2665	19,4
SARIMA (0,0,1)*(0,1,0)7	3451	2916	21,6
SARIMA (0,0,1)*(0,0,1)7	5631	4636	25,5
SARIMA (0,0,1)*(1,0,0)7	2921	2442	16,9
SARIMA (1,0,0)*(1,1,1)7	1612	1252	9,7
SARIMA (1,0,0)*(0,1,1)7	1632	1204	9,1
SARIMA (1,0,0)*(1,0,1)7	1726	1242	8,6
SARIMA (1,0,0)*(1,1,0)7	1910	1448	11,8
SARIMA (1,0,0)*(0,1,0)7	2277	1662	12,6

Cuadro 26: Modelos ajustados al clúster 5 con sus respectivos valores de RMSE, MAE y MAPE

SARIMA (1,1,1)*(1,1,1)14	1747	1319	11,02
SARIMA (1,1,1)*(0,1,1)14	1538	1130	9,4
SARIMA (1,1,1)*(1,0,1)14	1704	1241	8,2
SARIMA (1,1,1)*(1,1,0)14	1969	1339	13,5
SARIMA (1,1,1)*(0,0,1)14	1792	1345	9,3
SARIMA (1,1,1)*(1,0,0)14	1783	1354	9,4
SARIMA (0,1,0)*(1,1,1)14	1729	1323	11,07
SARIMA (0,1,0)*(0,1,1)14	1497	1083	9,09
SARIMA (0,1,0)*(1,0,1)14	1705	1202	7,8
SARIMA (0,1,0)*(1,1,0)14	2054	1322	11,3
SARIMA (0,1,0)*(0,1,0)14	2290	1891	16,4
SARIMA (0,1,0)*(0,0,1)14	1763	1336	9,3
SARIMA (0,1,0)*(1,0,0)14	1783	1336	9,2
SARIMA (0,0,1)*(1,1,1)14	3062	2437	23,4
SARIMA (0,0,1)*(0,1,1)14	3452	2860	22,6
SARIMA (0,0,1)*(1,0,1)14	3163	2563	19,4
SARIMA (0,0,1)*(1,1,0)14	3378	2359	21,3
SARIMA (0,0,1)*(0,1,0)14	5254	4461	38,4
SARIMA (0,0,1)*(0,0,1)14	5474	4029	19,7
SARIMA (0,0,1)*(1,0,0)14	4013	3334	25,9
SARIMA (1,0,0)*(1,1,1)14	1819	1201	10,7
SARIMA (1,0,0)*(0,1,1)14	1799	1322	10,9
SARIMA (1,0,0)*(1,0,1)14	1760	1281	9,07
SARIMA (1,0,0)*(1,1,0)14	2054	1294	11,07
SARIMA (1,0,0)*(0,1,0)14	2273	1845	15,9
SARIMA (1,0,0)*(1,0,0)14	1749	1285	8,9

Cuadro 27: Modelos ajustados al clúster 5 con sus respectivos valores de RMSE, MAE y MAPE



LICENCIATURA EN MATEMÁTICAS

Tel: (57) 6 735 9300 Ext 382

Carrera 15 Calle 12 Norte

Armenia, Quindío - Colombia

licenciaturaenmatematicas@uniquindio.edu.co