



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA DE ESTADÍSTICA

**“ANÁLISIS DE SERIES TEMPORALES DE CASOS
CONFIRMADOS Y FALLECIDOS POR COVID – 19 DEL
ECUADOR”**

TRABAJO DE TITULACIÓN:

TIPO: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

AUTORA: MARITZA JANETH CABRERA VALLADOLID

DIRECTOR: ING. HÉCTOR SALOMÓN MULLO GUAMINGA

Riobamba – Ecuador

2021

©2021, Maritza Janeth Cabrera Valladolid

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Yo, Maritza Janeth Cabrera Valladolid, declaro que el presente trabajo de titulación es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autora asumo la responsabilidad legal y académica de los contenidos de este trabajo de titulación, el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 12 de agosto de 2021



Maritza Janeth Cabrera Valladolid

110476995-3

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA DE ESTADÍSTICA

El Tribunal del trabajo de titulación certifica que: El trabajo de titulación: Tipo: Investigación, **ANÁLISIS DE SERIES TEMPORALES DE CASOS CONFIRMADOS Y FALLECIDOS POR COVID – 19 DEL ECUADOR** realizado por la señorita: **MARITZA JANETH CABRERA VALLADOLID**, ha sido minuciosamente revisado por los Miembros del Tribunal del trabajo de titulación, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

	FIRMA	FECHA
Dra. Jaqueline Elizabeth Balseca PRESIDENTE DEL TRIBUNAL	_____	2021-08-12 _____
Ing. Héctor Salomón Mullo DIRECTOR/A DEL TRABAJO DE TITULACION	_____	2021-08-12 _____
Ing. Johanna Enith Aguilar MIEMBRO DE TRIBUNAL	_____	2021-08-12 _____

DEDICATORIA

Primordialmente, a Dios por haberme dado la sabiduría y la fortaleza para guiarme correctamente durante mi vida y en el desarrollo de este proyecto de investigación; a mis padres Agustín Cabrera y Fanny Valladolid por darme la vida, brindarme su tiempo, amor, apoyo incondicional y motivarme a continuar por el camino del conocimiento en el transcurso de mi educación; a mis hermanos Cristobal Agustín y Danton Gregorio por regalarme su apoyo emocional y a mi hermana Elizabeth del Cisne por ser mi compañera de toda la vida siempre presente en mis derrotas y en mis victorias en el transcurso de mi preparación profesional; a mis tíos, tías, primos, primas, a mi familia y seres queridos que constituyen el pilar fundamental en mi vida; a mis inolvidables compañeros de aula y docentes por brindarme sus conocimientos, cariño, respeto, comprensión y apoyo, haciendo que mi formación profesional sea más llevadera lejos de mi familia.

Maritza Cabrera

AGRADECIMIENTO

Agradezco infinitamente a Dios, dueño de la vida y de toda la existencia, por bendecirme todos los días dándome más de lo que pueda merecer, otorgándome la vida, sabiduría, inteligencia y la alegría de celebrar este gran logro de mi vida junto a mis seres queridos.

A mis apreciados padres, que con su demostración de ejemplo y trabajo me han mostrado el sendero correcto para continuar fortaleciendo mis conocimientos, enseñándome con sus sabios consejos a ser una mujer perseverante y fuerte ante las adversidades; a mis hermanos y hermana por su apoyo incondicional para alcanzar las metas propuestas de mi vida. A toda mi familia, quienes me apoyaron desinteresadamente en todo momento para que se haga realidad este anhelo.

Así mismo mi gratitud con la Escuela Superior Politécnica de Chimborazo, a la Facultad de Ciencias y fundamentalmente a la Carrera de Estadística por abrirme las puertas de la oportunidad para formarme académicamente y ser capaz de contribuir al desarrollo del país y el buen vivir de la sociedad; a todos los docentes que fueron parte de mi desarrollo profesional, transmitiéndome sus conocimientos, experiencias y vivencias con respeto, cariño y profesionalismo.

De manera especial quiero expresar un profundo agradecimiento a mi estimado maestro quien dirigió este trabajo investigativo, Ing. Héctor Mullo por brindarme su tiempo, su visión crítica sobre muchos aspectos en el desarrollo de esta investigación, además de su valiosa colaboración y rectitud profesional; a la Ing. Johanna Aguilar miembro de Tribunal por manifestarme su ayuda absoluta, su aprecio y constancia profesional. Que Dios continúe bendiciendo su vida y su trabajo con su don de docencia para que sigan transmitiendo sus conocimientos con sabiduría, responsabilidad y entusiasmo, mil gracias.

Son muchas las personas que han contribuido en mi formación académica y profesional, me complazco en expresarles mi gratitud por su amistad, cariño, respeto, consideración, consejos, ánimos, compañía y apoyo en los momentos que más los he necesitado, en cualquier lugar en donde se encuentren quiero darles las gracias por formar parte de mi vida, que dios les bendiga.

Maritza Cabrera

ÍNDICE DE CONTENIDOS

ÍNDICE DE CONTENIDOS	vi
ÍNDICE DE TABLAS.....	ix
ÍNDICE DE GRÁFICOS.....	x
ÍNDICE DE FIGURAS.....	xi
ÍNDICE DE ANEXOS	xii
RESUMEN	xiii
ABSTRACT.....	xiv

INTRODUCCIÓN	1
--------------------	---

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL.....	5
1.1. Antecedentes.....	5
1.2. Planteamiento del Problema	8
1.2.1. <i>Enunciado del problema</i>	8
1.2.2. <i>Formulación</i>	9
1.3. Justificación.....	9
1.4. Objetivos.....	10
1.4.1. <i>Objetivo General</i>	10
1.4.2. <i>Objetivos Específicos</i>	11
1.5. Bases teóricas	11
1.5.1. <i>Virus</i>	11
1.5.2. <i>Pandemia</i>	11
1.5.3. <i>Fases de la pandemia con relación a los procedimientos preventivos</i>	11
1.5.4. <i>SRAS-CoV2 y COVID-19</i>	12
1.5.5. <i>COVID-19 como Amenaza</i>	12
1.5.6. <i>Mortalidad</i>	13
1.5.7. <i>Transmisión</i>	13
1.5.8. <i>Sintomatología y enfermedad</i>	14
1.5.9. <i>Prevención</i>	14
1.6. Bases conceptuales	15
1.6.1. <i>Series Temporales</i>	15
1.6.1.1. <i>Objetivo de la serie de tiempos</i>	15

1.6.1.2.	<i>Componentes de la serie de tiempo</i>	15
1.6.1.3.	<i>Autocorrelación ACF Y PACF</i>	16
1.6.1.4.	<i>Series de tiempo estacionarias</i>	17
1.6.1.5.	<i>Series no estacionarias</i>	18
1.6.1.6.	<i>Raíz Unitaria</i>	18
1.6.1.7.	<i>Test Augmented Dickey Fuller (ADF)</i>	19
1.6.1.8.	<i>Grado de integración</i>	20
1.6.1.9.	<i>Ruido Blanco</i>	20
1.6.2.	<i>Metodología de Box Jenkins</i>	21
1.6.2.1.	<i>Pasos de la metodología Box y Jenkins</i>	21
1.6.2.2.	<i>Criterios de selección del modelo</i>	22
1.6.2.3.	<i>Medidas del Error</i>	22
1.6.3.	<i>ARIMA - Modelo autorregresivo integrado de media móvil</i>	23
1.6.3.1.	<i>Etapas del modelo ARIMA</i>	24
1.6.3.2.	<i>Expresión general del modelo ARIMA</i>	24
1.6.3.3.	<i>Modelo autorregresivo de orden p:AR(p)</i>	25
1.6.3.4.	<i>Modelo móvil de orden q: MA(q)</i>	25
1.6.3.5.	<i>Modelo autorregresivo de medias móviles de orden (p, q): ARMA (p, q)</i>	26
1.6.3.6.	<i>Modelo autorregresivo integrado de medias móviles de orden p, d, q :</i> ARIMA (p, d, q)	26
1.6.3.7.	<i>Modelo autorregresivo integrado de medias móviles estacional de período estacional</i> S	26
1.6.4.	<i>Bootstrap</i>	27
1.6.4.1.	<i>Bootstrapping en series de tiempo</i>	28
1.6.4.2.	<i>Bootstrap de bloques</i>	28
1.6.4.3.	<i>Estimación Bootstrap del Error Típico</i>	29
1.6.4.4.	<i>Intervalo de Confianza para predicción de series Bootstrap</i>	29

CAPÍTULO II

2.	MARCO METODOLÓGICO	31
2.1.	Tipo de Investigación	31
2.2.	Diseño de la investigación No Experimental	31
2.2.1.	Localización del estudio	31
2.2.2.	Población de estudio	32
2.2.3.	Tamaño de la muestra	32
2.2.4.	Método de muestreo	32

2.2.5.	<i>Técnicas de recolección de datos</i>	33
2.2.6.	<i>Identificación de variables</i>	33
2.2.7.	<i>Modelo Estadístico</i>	33
2.3.	Variables en estudio	33
2.3.1.	<i>Operacionalización de Variables</i>	33

CAPÍTULO III

3.	RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS	34
3.1.	Análisis exploratorio	34
3.2.	Modelación ARIMA	37
3.2.1.	<i>Correlogramas</i>	37
3.2.2.	<i>Test de Dickey Fuller aumentada</i>	38
3.2.3.	<i>Diferenciación</i>	39
3.2.4.	<i>Identificación de los modelos</i>	41
3.2.5.	<i>Estimación y selección del modelo</i>	42
3.2.6.	<i>Validación del modelo</i>	47
3.2.7.	<i>Pronósticos</i>	49
3.3.	Modelación Bootstrap	53
3.3.1.	<i>Análisis del modelo</i>	53
3.3.2.	<i>Pronóstico</i>	54

	CONCLUSIONES	59
--	---------------------------	----

	RECOMENDACIONES	61
--	------------------------------	----

BIBLIOGRAFÍA

ANEXOS

ÍNDICE DE TABLAS

Tabla 1-2: Operacionalización de Variables.....	33
Tabla 1-3: Estadística descriptiva de las variables de casos confirmados y fallecidos por COVID-19 presentados a diario.....	35
Tabla 2-3: Test de Dickey Fuller (Estacionariedad) Casos de COVID-19 en Ecuador.....	39
Tabla 3-3: Criterios de información y evaluación para los posibles modelos ARIMA de la serie de Casos confirmados por COVID-19.....	43
Tabla 4-3: Criterios de información y evaluación para los posibles modelos ARIMA de la serie de Casos Fallecidos por COVID-19	45
Tabla 5-3: Coeficientes e intervalos de confianza de los modelos seleccionados.....	47
Tabla 6-3: Comprobación de supuestos de los modelos de casos confirmados y fallecidos por COVID-19	48
Tabla 7-3: Pronóstico del modelo para las series de casos confirmados y fallecidos por COVID-19	49
Tabla 8-3: Medidas de adecuación de los pronósticos metodología ARIMA	51
Tabla 9-3: Estadística Bootstrap por bloques de la serie de casos confirmados y fallecidos por COVID-19	54
Tabla 10-3: Pronóstico Bootstrap en base a los modelos ARIMA para las series de casos confirmados y fallecidos por COVID-19	54
Tabla 11-3: Medidas de adecuación de los pronósticos metodología Bootstrap.....	56

ÍNDICE DE GRÁFICOS

Gráfico 1-3: Porcentaje de casos confirmados y fallecidos	35
Gráfico 2-3: Evolución serial de casos confirmados por COVID-19	36
Gráfico 3-3: Evolución serial de casos fallecidos por COVID-19.....	36
Gráfico 4-3: Correlogramas ACF y PACF de la serie casos confirmados por COVID-19	37
Gráfico 5-3: Correlogramas ACF y PACF de la serie casos fallecidos por COVID-19.....	37
Gráfico 6-3: Evolución de la primera diferencia serial de Casos Confirmados por COVID-19	39
Gráfico 7-3: Evolución de la segunda diferencia serial de Casos Confirmados por COVID-19	40
Gráfico 8-3: Evolución de la primera diferencia serial de Casos Fallecidos por COVID-19...40	
Gráfico 9-3: Correlogramas ACF y PACF de casos Confirmados por COVID-19	41
Gráfico 10-3: Correlogramas ACF y PACF de casos fallecidos por COVID-19.....	42
Gráfico 11-3: Correlogramas de los modelos ARIMA propuestos para la serie de casos Confirmados por COVID-19	43
Gráfico 12-3: Correlogramas de los modelos ARIMA propuestos para la serie de casos fallecidos por COVID-19	46
Gráfico 13-3: Validación del modelo ARIMA(0,2,2) de casos confirmados por COVID-19....	47
Gráfico 14-3: Validación del modelo ARIMA(1,1,1) with drift de casos fallecidos por COVID-19	48
Gráfico 15-3: Pronóstico ARIMA(0,2,2) para los meses de noviembre y diciembre de casos confirmados por COVID-19.....	52
Gráfico 16-3: Pronóstico ARIMA(1,1,1) with drift para los meses de noviembre y diciembre de casos fallecidos por COVID-19.....	52
Gráfico 17-3: Pronóstico Bootstrap ARIMA(0,2,2) para los meses de noviembre y diciembre de casos confirmados por COVID-19	57
Gráfico 18-3: Pronóstico Bootstrap ARIMA(1,1,1) with drift para los meses de noviembre y diciembre de casos fallecidos por COVID-19	58

ÍNDICE DE FIGURAS

Figura 1-2: Mapa Ecuador	32
---------------------------------------	----

ÍNDICE DE ANEXOS

ANEXO A: Código en R (versión 4.0.3) del análisis exploratorio de las series de tiempo

ANEXO B: Código en R (versión 4.0.3) metodología de Box-Jenkins

ANEXO C: Código en R (versión 4.0.3) metodología Bootstrapping

RESUMEN

El presente trabajo tuvo como objetivo modelar curvas de casos confirmados y personas fallecidas de la enfermedad COVID-19, representadas por el número acumulado de casos en función del tiempo, Ecuador 2020. La investigación fue de naturaleza cuantitativa, de objeto aplicada con un nivel de profundización explicativo, de tipo inferencial inductivo con corte longitudinal. El análisis se llevó a cabo mediante datos de 168192 casos confirmados y 8357 muertes comprobadas de las 24 provincias del Ecuador, desde el 13 de marzo hasta el 31 de octubre del 2020, información que es reportada por Ministerio de Salud Pública del Ecuador en su página web de datos abiertos. Los principales resultados del análisis exploratorio de las series de tiempo fue la identificación de tendencia creciente, ausencia de estacionalidad y estacionariedad, se obtuvo los modelos seriales en base a criterios de información y medidas de adecuación para las series de casos confirmados ARIMA(0,2,2) y casos fallecidos ARIMA(1,1,1) con constante, los residuos de los dos modelos cumplen con el supuesto de estacionariedad y ruido blanco pero no cumplen con el supuesto de normalidad, la metodología Bootstrap se basó en el método estadístico de predicción ARIMA, el desempeño de las metodologías se evaluó comparando los valores reales y el pronóstico de los meses de noviembre y diciembre. Finalmente se determinó que la modelación ARIMA es adecuada para pronosticar casos confirmados y la modelación Bootstrap para los casos fallecidos, se verifica que al ser series de tiempo con diferentes características y según el desempeño de los modelos, no se puede optar solamente por un método de modelación. Se recomienda contrastar la existencia de pronósticos más precisos de casos confirmados y fallecidos por medio de distintas técnicas estadísticas como Holt Winter, árboles de decisión, redes neuronales artificiales entre otros, idóneos para trabajar con el número acumulado de casos.

Palabras clave: <ESTADÍSTICA>, <COVID-19>, <SERIES DE TIEMPO>, <MODELACIÓN>, <PRONÓSTICO>, <METODOLOGÍA BOX-JENKINS>, <ARIMA>, <BOOTSTRAP>, <ECUADOR (PAIS)>



0717-DBRAI-UPT-2021

ABSTRACT

The objective of this work was to model curves of confirmed cases and deceased people of the COVID-19 disease, represented by the accumulated number of cases as a function of time, Ecuador 2020. The research was quantitative in nature, object applied with a level of explanatory deepening, inductive inferential type with longitudinal cut. The analysis was performed using data from 168,192 confirmed cases and 8,357 verified deaths from the 24 provinces of Ecuador, from March 13 to October 31, 2020, information that is reported by the Ministry of Public Health of Ecuador on its open data web page. The main results of the exploratory analysis of the time series was the identification of increasing trend, absence of seasonality and stationarity, serial models were obtained based on information criteria and measures of adequacy for the series of confirmed cases ARIMA (0.2, 2) and death cases ARIMA (1,1,1) with drift, the residuals of the two models comply with the assumption of stationarity and white noise but do not fulfill the assumption of normality, the Bootstrap methodology was based on the statistical method of ARIMA prediction, the performance of the methodologies was evaluated by comparing the actual values and the forecast for the months of November and December. Finally, it was determined that the ARIMA modeling is adequate to forecast confirmed cases and the Bootstrap modeling for deceased cases, it is verified that it is not possible to choose only one method of modeling due to the performance of the models and considering that they are time series with different characteristics. It is recommended to contrast the existence of more accurate forecasts of confirmed and deceased cases by means of different statistical techniques such as: Holt Winter, decision trees, artificial neural networks, among others, suitable for working with the accumulated number of cases.

Keywords: <STATISTICS>, <COVID-19>, <TIME SERIES>, <MODELING>, <FORECAST>, <BOX-JENKINS METHODOLOGY>, <ARIMA>, <BOOTSTRAP>, <ECUADOR (COUNTRY)>

INTRODUCCIÓN

Este último año se ha visto afectado por una pandemia que no da tregua, causando más de 2 millones de personas contagiadas y la muerte de cientos de miles a nivel mundial, principalmente en personas de la tercera edad y aquellas que tienen factores de riesgo que los hacen más vulnerables y en pocos casos en personas jóvenes; la pandemia se da a partir de una angustiosa enfermedad denominada como COVID-19, inducida por un virus poco conocido, el SARS-CoV-2 (Pérez Velázquez, 2020, p. 1).

En diciembre de 2019, Wuhan, provincia de Hubei, China, se convirtió en el centro de atención no solo local sino a nivel mundial por un brote de neumonía de causa no conocida. Al 23 de enero de 2020, se habían detectado en China un total de 835 casos con infección por 2019-nCoV llamado inicialmente el cual proveniente de los Betacoronavirus del grupo 2B, de estos casos confirmados por laboratorio en China 25 han fallecido y el 93% permanece hospitalizados. Como la secuencia genética de este virus es de al menos 70% similar al SARS-CoV, permitió el rápido desarrollo de pruebas de diagnóstico de RT-PCR en tiempo real, la OMS lo denominó a este coronavirus como SARS-CoV-2 infección del COVID-19. Los casos presentados de COVID-19 en Wuhan no se limita a este lugar, sino que su expansión llegó a otros continentes, estableciendo que su transmisión es de persona a persona y, por lo tanto, se incrementa el riesgo de una propagación mucho más amplia de la enfermedad (Wang et al.,2020: p. 2).

El Director General de la Organización Mundial de la Salud (OMS), el doctor Tedros Adhanom Ghebreyesus, declara la epidemia de COVID-19 como una emergencia de salud pública de preocupación internacional el 30 de enero de 2020, para el 11 de marzo de 2020 la caracteriza como una pandemia, esto significa que la epidemia se ha extendido por varios países, continentes y todo el mundo, y que afecta a un gran número de personas, causando una emergencia sanitaria de gran magnitud; la OMS está trabajando con asociados de todo el mundo para ayudar a coordinar el desarrollo de una vacuna, la misma que debe ser segura y eficaz, además que debería obtener la aprobación de los organismos nacionales de reglamentación, tendría que fabricarse con arreglo a normas precisas y distribuirse (Organización Panamericana de la Salud, 2020).

Las grandes potencias al contar con un aforo económico considerable les permiten el desarrollo en ciencia y tecnología, por tanto, tienen la capacidad de desplegar estrategias biotecnológicas muy distintas para mitigar el grave daño que está dejando la pandemia, hasta a finales del año ya se encuentran realizando ensayos clínicos de antivirales que parece estar dando resultados favorables para contrarrestar el COVID-19. En el mes de agosto del 2020, Rusia se ha convertido

ya en el primer país del mundo en aprobar una vacuna contra el síndrome respiratorio agudo severo coronavirus 2 o SARS-CoV-2, esta vacuna se llama Sputnik V y fue desarrollada por el Centro Nacional de Epidemiología y Microbiología de Gamaleya - Moscú, se basa en dos vectores de adenovirus (virus que infecta las membranas de las vías respiratorias), el presidente Vladimir Putin aprobó a Sputnik V y mencionó que la vacuna funciona con bastante eficacia, ayuda a desarrollar una inmunidad fuerte y ha pasado por todas las pruebas necesarias, esta declaración fue prematura debido a que en el momento de la aprobación la vacuna no había comenzado los ensayos de fase 3, y tampoco se había publicado los resultados sobre los ensayos de la fase 2; la hija del presidente Vladimir Putin ha sido la primera en recibir esta vacuna y se ha programado el inicio de ensayo clínicos a gran escala (Burki, 2020, p. e85). La Sputnik V antes de culminar el año se encuentra compitiendo con vacunas creadas por China y grandes laboratorios occidentales, a pesar de la desconfianza y la crítica por la rapidez de sus ensayos y falta de transparencia, esta es la primera vacuna en recibir la aprobación para uso de emergencia y está por concluir la última fase de sus ensayos clínicos, la Sputnik V tiene una eficacia del 92%, según los datos preliminares divulgados por el Instituto Gamaleya (Sahuquillo, 2020).

La mayoría de los casos en América Latina son procedentes de Europa, particularmente de Italia; la OMS, ha determinado que la gran mayoría de los países de las Américas empiezan a presentar transmisión local y contención del virus, al mismo tiempo se evidencia la necesidad de mayor trabajo multidisciplinar para la prevención de enfermedades infecciosas, dado que existe una falta de infraestructura de salud para enfrentar desafíos de tales dimensiones (Rodríguez Morales et al., 2020: p. 3).

Según (Acosta, 2020, p. 2-3) en su investigación sobre la Capacidad de respuesta frente a la pandemia de COVID-19 en América Latina y el Caribe, la situación de los países de América Latina y el Caribe es particularmente preocupante debido a las características y condiciones de los sistemas de salud y la escasez de políticas de cuidado para las personas mayores, así como la situación macroeconómica y la pobreza que impera en gran parte de los países de la región, la pandemia se extendió lentamente en este territorio, por tanto, los gobiernos lograron en parte sobrellevar esta crisis adoptando medidas enérgicas tempranas que evitaron el avance de la pandemia, en comparación con los países de Asia y Europa; así mismo, Acosta toma en consideración el brote de COVID-19 durante los primeros 90 días en cada país de la región, tres países Costa Rica, Cuba y Uruguay no se encontraban en fase de transmisión comunitaria a los 90 días de iniciada la pandemia, en este estudio se presenta una tasa de incidencia acumulada promedio de 1268 casos por millón de habitantes en la región, determinando la mayor tasa de incidencia acumulada en Perú, seguido de Chile, Panamá, Ecuador y Brasil, con baja letalidad; por otro lado, la mayor tasa de letalidad se identificó en México, seguido de Ecuador y Brasil con tasas de incidencia

intermedias y bajos niveles de testeos; la mayor Tasa Cruda de Mortalidad se registró en Ecuador, seguido de Perú y Brasil y en promedio para toda la región presenta una tasa de 43 muertes por millón de habitantes, cabe señalar que la Tasa Cruda de Mortalidad se incrementa a mayor velocidad en México, Ecuador, Perú, Chile, República Dominicana, Colombia y Brasil, verificando así que estos países tienen mayores problemas en la contención de la pandemia, dado a factores económicos, sociales y políticos.

Ecuador ha sido golpeado duramente por la pandemia, debido a que no cuenta con los recursos necesarios e indispensables para controlarla, lo cual ha ocasionado una crisis política, económica, social y sanitaria. Ecuador se ha encontrado entre los tres países de Sudamérica con más casos de COVID-19, el primer boletín de mayo reporta un total de 61.262 casos, de los cuales 20.937 están confirmados, 17.535 son sospechosos, y 22.790 son casos descartados. La cantidad de muertes hasta esa fecha, a causa de la pandemia, ya suma 1.564, además de tener cifras alarmantes de personal médico contagiados y fallecidos con la agravante del no reconocimiento del contagio por COVID-19 como, accidente de trabajo, ni como enfermedad profesional, vulnerando aún más el servicio hospitalario para la atención de la población confirmada de COVID-19 (Peraza de Aparicio, 2020, p. 1).

En Ecuador, amparados en el llamado Plan Nacional de Desarrollo 2017-2021 y tras declaraciones del presidente Lenin Moreno mencionó, que “Ecuador es un país de derechos, con estabilidad económica, trabajo y empleo dignos, el derecho a la salud y educación y servicios públicos de calidad, así como la reducción de la desigualdad”, sin embargo, el gobierno adopta crueles medidas contra la población viéndose afectados los más desfavorecidos tanto niños, adultos mayores, personas con capacidades diferentes y principalmente la clase pobre y media del Ecuador, priorizando el pago de la deuda externa en medio de la pandemia, además de hacer pactos con el Fondo Monetario Internacional, así mismo, originaban acciones para maquillar la dura realidad económica, social y política, el país ofreció una imagen deteriorada ante el mundo en los meses de mayor azote del virus, por la corrupción y la incapacidad de tomar las riendas ante una realidad casi surrealista: por un lado, el escándalo sobre la compra pública de mascarillas y kits alimenticios a sobreprecio, despidos de personal médico, falta de infraestructura hospitalaria y principalmente la inconciencia ciudadana al tomar medidas de prevención contra el COVID-19 de manera adecuada y obligatoria (Molina Prendes y Mejias Herrera, 2020: pp. 4–5).

La presencia de esta enfermedad en el Ecuador ha dejado muchas muertes y contagios, además de graves secuelas tanto económicos, humanitarios, sociales, políticos, educativos, laborales y sanitarios, por lo que, es necesario desarrollar investigaciones que ayuden a contrarrestar esta crisis para lograr una buena toma de decisiones sobre el COVID-19 más adelante. La interrogante

central de esta investigación es ¿Cuáles son los modelos de series de tiempo que caracterizan las curvas de casos confirmados y fallecidos por COVID-19, representadas por el número acumulado de casos en función del tiempo registrados por el Ministerio de Salud Pública del Ecuador?

Este estudio pretende modelar curvas de series temporales en casos confirmados y fallecidos por COVID-19, representadas por el número acumulado de casos en función del tiempo, del Ecuador en el año 2020. El trabajo de investigación se considera de tipo cuantitativo y no experimental debido a que la información es obtenida de registros diarios presentados en infografías por el Ministerio de Salud Pública del Ecuador en su página web y contienen valores de casos confirmados y fallecidos por COVID 19; según el objetivo es aplicada dado que la investigación se centra en el campo de la salud principalmente en el sistema sanitario por la propagación del COVID-19, mediante la modelación de curvas de casos confirmados y personas fallecidas registradas a nivel nacional y determinar el método de predicción serial que presente características buenas con la aplicación de los modelos ARIMA.

Este trabajo está dividido en tres capítulos los mismos que se describen a continuación:

En el **Capítulo I** reúne información documental para el desarrollo del diseño metodológico de la investigación. Se detallan los antecedentes, en el cual se describe investigaciones ya desarrolladas similares al estudio planteado tanto metodológico y aplicativo. Se presenta la problemática a analizar a través de preguntas que permitan una mejor comprensión sobre la solución que se debe dar. Se detalla el justificativo y se enuncian los objetivos que se pretende alcanzar con la investigación, además de las bases teorías respecto al COVID-19 y conceptuales respecto a las técnicas que se va a aplicar dentro del estudio.

En el **Capítulo II** revisa los procesos a realizar en la investigación, muestra la ubicación en donde se realiza el estudio, y define cada uno de los ítems referente a: tipo de investigación, población de estudio, tamaño de la muestra, método de muestreo, técnicas de recolección de la información, identificación de variables y operacionalización de variable.

En el **Capítulo III** aplica de forma sistemática que el nuevo hallazgo o resultado sea consistente con conocimiento científico, se analiza los resultados obtenidos luego de haber aplicado las técnicas de modelación de las curvas de series de tiempo de casos confirmados y fallecidos de COVID-19 en el Ecuador, seleccionando el modelo que mejor ajuste las predicciones, analizando los gráficos entre datos reales y datos predichos para determinar qué tan bueno es el ajuste de cada técnica.

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL

1.1. Antecedentes

Los modelos de series de tiempo son utilizados para modelar curvas, determinar predicciones y pronósticos de datos tomados en el tiempo, estos modelos han sido aplicados en varias investigaciones previas de carácter epidemiológicos.

(Maleki et al., 2020: p. 1) Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models (Modelado y pronóstico de la tasa de propagación y muerte del coronavirus (COVID-19) en el mundo utilizando modelos de series de tiempo), la propagación del COVID-19 se está dando de forma acelerada a nivel global causando afectaciones graves, por lo que es necesario ejecutar planes y políticas estrictas para ayudar a contrarrestarlas, para ello se debe predecir y pronosticar los casos confirmados futuros utilizando modelos de series de tiempo; en esta investigación se introduce un modelo mejorado de series de tiempo aplicando distribuciones TP - SMN basados en la mezcla de escalas de dos piezas normal, sin antes identificar que la serie de tiempo clásica se basa en la simetría de la distribución del error, pero en casos reales esta simetría de los términos de error no es satisfactoria, por tanto, con los modelos propuestos que incluyen modelos de series de tiempo simétrico, asimétrico con cola ligera y asimétrico con cola pesada no gaussiana se determinó que se ajustaron inicialmente a los datos históricos de COVID-19, después se seleccionó la serie de tiempo que mejor se ajustaba a un conjunto de datos y se aplicaron los modelos seleccionados para pronosticar casos confirmados, todos los juicios demuestran que los modelos planteados son más razonables que el modelo de serie de tiempo gaussiano ordinario.

(Ye y Yang, 2020: p. 19) Analysis and prediction of confirmed COVID-19 cases in China with uncertain time series (Análisis y predicción de casos confirmados de COVID-19 en China con series de tiempo inciertas), es importante estudiar la tendencia del número acumulado de casos confirmados por COVID-19 en China, el objetivo de este estudio fue implementar un modelo de serie de tiempo incierto para analizar y predecir la evolución de los casos confirmados de COVID-19, para ello se utilizó el análisis de series de tiempo clásico tomando número acumulado de casos confirmados de COVID-19 excluyendo los casos importados del 13 de febrero al 23 de marzo de 2020 para predecir los casos confirmados acumulativos en el futuro, los datos y el modelo pasaron el ruido blanco y la prueba de estacionariedad, pero los resultados no fueron favorables, con este

resultado se empleó el análisis de series temporales inciertas incluido el análisis de residuos, estimaciones mínimas, prueba de hipótesis incierta, pronósticos e intervalo de confianza; se determinó que el análisis de series de tiempo inciertas es más apropiado que el análisis de series de tiempo clásico debido a que la prueba de hipótesis incierta y la capacidad de identificar automáticamente los valores atípicos redujeron la varianza estimada del término de perturbación a un valor aceptable.

(Yang et al., 2020: pp. 1–3) Research on COVID-19 based on ARIMA model—Taking Hubei, China as an example to see the epidemic in Italy (Investigación sobre COVID-19 basada en el modelo ARIMA Δ —Tomando Hubei, China como ejemplo para ver la epidemia en Italia), el modelo de Box-Jenkins es el modelo de predicción de series de tiempo más común, en este trabajo utilizaron datos periodos completos para modelar y validar el ARIMA, se derivaron del número diario de nuevos casos confirmados y nuevas muertes de casos de COVID-19 en Hubei, China. Los participantes del estudio aplicaron los modelos ARIMA basados en los nuevos casos y nuevas muertes en Hubei a Italia para observar su situación epidémica en los próximos 10 días, los modelos se ajustan bien y son más adecuados para la predicción a corto plazo, pero con la propagación de la epidemia la secuencia de datos cambia constantemente haciendo que el modelo cambie, siendo así, no se tiene claro el tiempo, la causa y la tendencia de la situación epidemiológica, de lo anterior se consolida que los resultados predictivos deben ser de acuerdo a la situación real.

(Saba et al.: 2021) Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types (Técnicas de aprendizaje automático para detectar y predecir el total diario de casos de infecciones y muertes por COVID-19 bajo diferentes tipos de bloqueo), este trabajo muestra la necesidad de pronosticar los casos de infección y las muertes confirmadas acumuladas diariamente en 3 diferentes tipos de encierros parcial, colectivo y completo, con el fin de seleccionar las estrategias de encierro más adecuadas y controlar la intensidad de esta pandemia, para reducir la carga en los hospitales; el estudio se basa en las restricciones parciales introducidas en Bulgaria, Grecia y Rusia, restricciones colectivas en Islandia, Holanda y Suecia, y las restricciones totales en Hubei (China), Irán e India, mediante la aplicación de series de tiempo y modelos de Machine learning, entre estos, bosques aleatorios, K vecinos más cercanos, SVM, árboles de decisión (DT), regresión polinomial, Holt Winter, ARIMA y SARIMA, la precisión y eficacia de los modelos se valoraron con los criterios de rendimiento; se aplicó un ensacado o agregación Bootstrap para elegir varios bosques aleatorios y encontrar la mejor ruta de solución al problema, el ensacado con la selección tiene como propósito, reducir la correlación entre arboles de decisión sin minimizar bruscamente la varianza, también utiliza el método de arranque que extrae al azar varias muestras de los datos

originales. Para cada país se determinó 3 modelos óptimos por lo que se torna imposible recomendar un enfoque único para modelar y pronosticar, el modelo óptimo para cada conjunto de datos permitió pronosticar casos con 10 días de anticipación y obtuvo resultados muy cercanos a los valores reales. Además, se identificó que la estrategia de encierro colectivo es la mejor política para controlar los casos confirmados y las muertes por COVID-19.

(Huamani et al. 2020) Casos confirmados y mortalidad por COVID-19 en Sudamérica: un análisis comparativo por millón de habitantes, este trabajo tiene como objetivo comparar equitativamente y evaluar la propagación del COVID-19 y las medidas adoptadas por el gobierno de los países de Sudamérica, los datos de confirmaciones y defunciones por COVID-19 fueron tomados de Our World in Data corroborados con información oficial de cada país a partir del primer reporte hasta el 29 de junio del 2020, determinando los pronósticos en países con mayor presencia de contagio por millón de habitantes, mediante modelos ARIMA y siguiendo la metodología propuesta por Box y Jenkins, Brasil presenta una tendencia creciente en el número de casos confirmados, Perú y Ecuador muestran superficialmente la misma tendencia siendo estos países los más afectados por el número elevado de contagios, Ecuador es el país que tiene mayor número de muertes por millón de habitantes y el cuarto país con más casos confirmados por millón de habitantes. Se realizó pronósticos a corto plazo para el número de contagio por millón de habitantes de los países de Bolivia, Brasil, Chile, Ecuador y Perú, según el estadístico de Ljung y Box, los residuales de cada modelo no presentan autocorrelaciones, es decir los residuales presentan un comportamiento de ruido blanco, con valores significativos altos, de esta manera no rechazando las hipótesis nula de independencia de los residuales, evidenciando la adecuación óptima de los modelos, Ecuador con el modelo $ARIMA(0,1,2)$ espera disminuir el número de contagios y posiblemente mantener una tendencia constante en las últimas dos semanas al 17 de julio.

(Hallo, Rojas y Hallo, 2020) Perspective from Ecuador, the Second Country with More Confirmed Cases of Coronavirus Disease 2019 in South America: A Review (Perspectiva de Ecuador, el segundo país con más casos confirmados de enfermedad por coronavirus 2019 en América del Sur: una revisión), la respuesta de Ecuador al nuevo virus COVID-19 comenzó el 26 de febrero de 2020 al confirmar su primer caso y a un mes después de que comenzara el brote en China. El 23 de marzo de 2020, el Ministerio de Salud de Ecuador informó sobre 789 pacientes confirmados con COVID-19 y 18 muertes relacionadas con el virus, la tasa de mortalidad fue similar a la de otros países y el número de pacientes confirmados aumentó, al igual que el número de personas en vigilancia epidemiológica y cuarentena.

(Ruiz y León, 2020, p. 5) Evolución de la enfermedad por coronavirus (COVID-19) en Ecuador, tiene como finalidad demostrar cómo se desarrolla la infección por COVID-19 en Ecuador a partir de la información proporcionada por el gobierno central con el fin de implementar oportunamente medidas que aplanen la curva creciente en el país, ha sido necesario realizar estos tipos de investigación observacional en este caso se basa en datos demográficos y estadísticos de estudios previos específicamente de reportes del 13 al 31 de marzo del 2020, los mismos que fueron extraídos de bases de datos del Sistema Nacional de Gestión de Riesgos y Emergencias (SINGRE), además de una revisión en bases de datos PubMed, OMS, Organización de las Naciones Unidas (ONU) e Instituto Nacional de Estadísticas y Censos(INEC) de los años 2013, 2019 y 2020; se determinó que el Ecuador tiene la tasa más alta de personas con COVID-19 en América del Sur con 13.15 por 100000 habitantes y supera el promedio mundial de 9.63. La tasa de letalidad en Ecuador de 3.40% se acerca al promedio mundial de 4.80%. Las provincias de Guayas, Galápagos, Cañar y Sucumbíos tienen las tasas de morbilidad por COVID-19 más altas en el país y superan la media mundial.

En este trabajo se pretende, modelar curvas de series temporales en casos confirmados y fallecidos por COVID-19 representadas por el número acumulado de casos en función del tiempo, Ecuador 2020, para lo cual se requiere determinar el modelo de las series temporales aplicando modelos ARIMA.

1.2. Planteamiento del Problema

1.2.1. Enunciado del problema

Las series de tiempo forman parte significativa de la modelación estadística, son muy utilizadas para modelar, predecir y pronosticar problemas prácticos reales, los cuales están plasmados en el tiempo y por lo tanto ordenados cronológicamente, en epidemiología las series de tiempo son usadas a corto plazo principalmente para modelar curvas epidemiológicas, obtener predicciones y pronósticos que sirven para planificar los recursos y tomar decisiones en función de lo que se espere en el futuro, teniendo una visión clara más no precisa del comportamiento de un evento.

La Organización Mundial de la Salud (OMS) declaró en “Emergencia de salud pública de interés internacional” al nuevo coronavirus llamado SARS-CoV2 y denomina a esta enfermedad como COVID-19 el 30 de enero de 2020, este virus era desconocido hasta que estallara el brote en Wuhan – China en diciembre de 2019, las principales afecciones son fiebre, síntomas respiratorios como tos y dificultad para respirar, en casos graves puede causar neumonía, síndrome respiratorio agudo severo, insuficiencia renal e incluso la muerte, se transmite de persona a persona,

expandiéndose en muchos países del mundo provocando crisis internacional sanitaria, humanitaria, social y económica.

El primer caso confirmado en el continente americano es en Estados Unidos el 20 de enero del 2020, y Brasil informa del primer caso el 26 de febrero del 2020, desde entonces el COVID – 19 se ha propagado por 54 países de América, según la Organización Panamericana de la Salud. Dando cumplimiento con el reglamento Sanitario Internacional, el 29 de febrero del 2020 Ecuador notifica su primer caso importado de COVID-19 a la Organización Panamericana de la Salud y Organización Mundial de Salud (OPS/OMS) por el Ministerio de Salud Pública (MSP) de Ecuador. Los registros de casos confirmados y fallecidos por COVID-19 son llevados por el Ministerio de Salud Pública y el COE Nacional.

Es muy importante la aplicación teórica de series de tiempo en casos confirmados y fallecidos de COVID-19 en el Ecuador, ya que permite realizar un seguimiento epidemiológico de los datos en el tiempo brindando la facilidad para la detección temprana de cambios en los patrones de ocurrencia que se pueden presentar en un futuro inmediato y con esto poder tomar decisiones y acciones.

1.2.2. Formulación

El proyecto busca dar respuesta a la siguiente pregunta:

¿Cuáles serían los modelos de series de tiempo que mejor se ajusten y caractericen las curvas de casos confirmados y fallecidos por COVID-19, representadas por el número acumulado de casos en función del tiempo registrados por el Ministerio de Salud Pública del Ecuador?

1.3. Justificación

El análisis de series de tiempo permite observar el comportamiento del experimento u observaciones dependientes ordenadas a través del tiempo. Box y Jenkins en 1970, desarrollaron una estructura metodológica el cual permite identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable que juega un papel fundamental es el tiempo, ARIMA (Autoregressive Integrated Moving Average – Modelos de promedio móvil autorregresivo integrado) es un tipo de modelo que produce pronósticos con base en una síntesis de los patrones históricos en los datos: los modelos (Melo Morín y Santana Esparza, 2016: p. 23); el modelo ARIMA se deriva de sus tres componentes AR (Autorregresivo), I (Integrado) y MA (Media Móvil). El método ARIMA se ha utilizado en salud pública para modelar el comportamiento de eventos relacionados con la salud, la primera utilización de estos modelos en el ámbito sanitario se debe

a Keewan Choi del Buró de Epidemiología del Centro para el Control de Enfermedades (CDC) y Stephen Tacker de la Universidad de Emory, ambos en Atlanta, EE.UU., quienes propusieron este método para pronosticar el número de muertes esperadas por influenza y neumonía (Choi y Thacker, 1981; Saldaña Valqui, 2016, p. 1-2). Bootsprtat permite generar nuevas series de tiempo similares a las series de tiempo observadas a partir de los residuos, para ello se debe transformar la serie de tiempo en Box-Cox y usando el método STL (Seasonal and Trend decomposition using Loess) se descompone en componentes de estacionalidad, tendencia y residuo usando Loess, cabe mencionar que Loess es un método para estimar relaciones no lineales, a partir de esto se obtendrá versiones barajadas del componente residual para obtener así la serie residual reemuestreada (Hyndman y Athanasopoulos, 2018).

El enfoque de la investigación es cuantitativo ya que se manipula cifras de casos confirmados y fallecidos a causa del COVID-19, en base a la fuente de información documental para el uso de reportes digitales proporcionado por el Ministerio de Salud Pública del Ecuador, además se emplea el método analítico para determinar la tendencia, el comportamiento estacional y predicciones mediante la aplicación de métodos ARIMA.

En función a la investigación realizada se pretende dar resultados estadísticos basados en el análisis de series temporales del método ARIMA, el cual permitirá identificar la situación con la que se enfrentará el país a corto plazo, permitiendo al gobierno, Ministerio de Salud Pública y a la población en general tomar las medidas pertinentes ante la propagación del COVID-19, ya que esta epidemia es la causante de muchos problemas tanto a nivel económico, social, humano y principalmente sanitario, por tanto, es fundamental predecir y pronosticar los casos confirmados y fallecidos de COVID-19 en Ecuador para tener indicios de lo que se espera en el futuro, y así considerar implementar planes y políticas especiales que lleven a mejorar la situación del país mediante la contención, estabilización y erradicación de contagios y muertes por COVID-19.

1.4. Objetivos

1.4.1. *Objetivo General*

Modelar curvas de casos confirmados y personas fallecidas de la enfermedad COVID-19, representadas por el número acumulado de casos en función del tiempo, Ecuador 2020.

1.4.2. *Objetivos Específicos*

- Modelar la serie temporal de casos confirmandos y fallecidos por COVID – 19 mediante modelos ARIMA;
- Modelar la serie temporal de casos confirmandos y fallecidos por COVID – 19 mediante la metodología Bootstrap;
- Determinar si la metodología ARIMA presenta mejores predicciones en comparación con la metodología Bootstrap.

1.5. Bases teóricas

1.5.1. *Virus*

DIEC (Diccionario de la lengua catalana del Instituto de Estudios Catalanes) define los virus con el Individuo de un grupo de agentes infecciosos submicroscópicos, históricamente unidos por el hecho de no poder ser retenidos por filtros bacterianos, parásitos endocelulares obligados de plantas, animales y bacterias, que consta de un núcleo central constituido por ácido desoxirribonucleico o por ácido ribonucleico y rodeado por una cubierta de proteína o cápside. DIEC, los virus son “parásitos endocelulares obligados de plantas, animales y bacterias”, es decir, están obligados a vivir dentro de una célula (Martínez, 2016, p. 2).

1.5.2. *Pandemia*

Las pandemias son grandes brotes de alguna enfermedad que afectan a varios países y plantean graves riesgos sanitarios, sociales y económicos. Un agente patógeno que circula rápidamente y se propaga por todo el planeta puede matar a decenas de millones de personas, perturbar las economías y desestabilizar la seguridad de los países. El cambio climático, la urbanización y la falta de agua y saneamiento son factores que pueden acelerar la propagación de los brotes de enfermedades y contribuir a que adquieran proporciones catastróficas (Banco Mundial, 2020). Las pandemias constituyen uno de los doce factores de riesgo mundial de que ocurra un daño inexorable para toda la humanidad (Santana López et al., 2019: p. 481).

1.5.3. *Fases de la pandemia con relación a los procedimientos preventivos*

Son alrededor de 5 y 6 fases de procedimiento preventivo como se detalla a continuación (Romero, 2020):

FASE 1: Empieza cuando se advierte la presencia del virus a los otros países los cuales deberán establecerse hospitales de atención y un protocolo sanitario en los puntos de ingreso al territorio, como fronteras y aeropuertos.

FASE 2: Es la fase de contención y comienza cuando se tiene identificadas a las personas que traen el virus desde el extranjero (casos importados). Se aplican protocolos de aislamiento, distanciamiento social, lavado de manos.

FASE 3: Es la etapa de contagio comunitario, el virus circula por todo el país. El virus ya está en la comunidad y se considera que todos los pacientes están potencialmente infectados con SARS-CoV-2 y cualquier paciente sintomático se considera COVID-19 hasta que se demuestre lo contrario.

FASE 4: En esta etapa el virus se dispersa en la comunidad: es casi imposible trazar un mapa de contagio y el número de casos se incrementan. Se debe evitar que la curva se acelere para ello se restringe la movilidad en el país.

FASE 5: Estas se dan cuando las medidas anteriores no han sido aplicadas con éxito y el número de casos salen de control. Se llega al pico más alto de la pandemia y las medidas siguen siendo extremas se prohíbe la movilidad.

1.5.4. SRAS-CoV2 y COVID-19

El SARS fue causado por un coronavirus CoV e identificado en 2003 de acuerdo con la Organización Mundial de la Salud (OMS). El coronavirus CoV es un virus animal de murciélagos que infectó a humanos en la provincia de Guangdong, China (Lai et al. 2020). Cuando se descubrió el SARS-CoV en 2003, el análisis filogenético del genoma del SARS-CoV mostró que ocupaba una posición única en el Betacoronavirus, que posteriormente se colocó en el subgénero Sarbecovirus. Después de la epidemia de SARS, se descubrió un número sin precedentes de CoV nuevos (Luk et al., 2019, p. 22). Es el nuevo coronavirus humano que ahora se llama coronavirus 2 del síndrome respiratorio agudo severo (SARS-CoV-2) anteriormente llamado HCoV-19 surgió en Wuhan, China, a fines de 2019 y ahora está causando una pandemia (Organización Mundial de la Salud, 2020; citado en van Doremalen et al., 2020: p. 1). El coronavirus pertenece a la subfamilia *Orthocoronavirinae* en la familia de *Coronaviridae* en el orden *Nidovirales*, que causa principalmente infecciones en el tracto respiratorio y gastrointestinal (Huang et al., 2020, p. 497).

1.5.5. COVID-19 como Amenaza

Hay dos razones por las que COVID-19 es una amenaza. Primero, puede matar a adultos sanos además de personas mayores con problemas de salud existentes. Los datos hasta ahora sugieren que el virus tiene un riesgo de letalidad de alrededor del 1%; Esta tasa la haría mucho más severa

que la influenza estacional típica, ubicándola en algún lugar entre la pandemia de influenza de 1957 (0.6%) y la pandemia de influenza de 1918 (2%) (Novel, 2020; citado en Gates, 2020, p. 1). En segundo lugar, COVID-19 se transmite de manera bastante eficiente. La persona infectada promedio propaga la enfermedad a otras dos o tres, una tasa de aumento exponencial. También existe una fuerte evidencia de que puede ser transmitida por personas que están levemente enfermas o incluso presintomáticas (Hoehl et al., 2020; citado en Gates, 2020, p. 1). Eso significa que COVID-19 será mucho más difícil de contener que el síndrome respiratorio del Medio Oriente o el síndrome respiratorio agudo severo (SRAS), que se propagó de manera mucho menos eficiente y solo por personas sintomáticas. De hecho, COVID-19 ya ha causado 10 veces más casos que el SARS en una cuarta parte del tiempo (Gates, 2020, p. 1).

1.5.6. Mortalidad

Es la desaparición permanente de todo signo de vida, cualquiera que sea el tiempo transcurrido desde el nacimiento con vida (cesación post-natal de las funciones vitales sin posibilidad de resucitar) (Organización Panamericana de la Salud, 2017). La lamentable cantidad de muertes causada por el COVID-19 está en función de dos indicadores: el nivel de mortalidad o probabilidad de morir por esta enfermedad, y la cantidad de población de una sociedad o área particular siendo determinante, para esto último, la estructura de edad (Manzano y Russo, 2020, p. 2).

1.5.7. Transmisión

Se transmite de persona a persona a través de gotas o contacto directo, y se ha estimado que la infección tiene un período de incubación medio de 6.4 días con un rango de 2.1 días a 11.1 días y según las observaciones de los datos del brote temprano en China continental del 10 al 24 de enero de 2020 se estimó que el número de reproducción básico medio (R_0) oscilaba entre 2.24 a 3.58 al 95% de confianza (Lai et al., 2020: p. 2). Generalmente los coronavirus se replican primordialmente en las células epiteliales del tracto respiratorio inferior y en menor medida en las células de las vías respiratorias superiores; es por eso por lo que la transmisión ocurre principalmente de pacientes con enfermedad reconocida y no de pacientes con signos leves e inespecíficos, es decir, que se cree que la propagación ocurre solo después de que se presentan signos de enfermedad del tracto respiratorio inferior. Sin embargo, los pacientes con infección por 2019-nCoV detectados en estado severo o fatal tienen una mayor probabilidad de transmitir este virus, ya que eliminan una mayor cantidad de partículas infectivas en comparación con los pacientes que presentan la infección de forma leve o asintomática (Palacios Cruz et al., 2020: p. 57).

1.5.8. Sintomatología y enfermedad

Las manifestaciones clínicas de COVID-19 son variables, van desde una enfermedad leve asintomática, hasta enfermedad severa con falla respiratoria, disfunción multiorgánica, choque séptico y muerte. La fiebre, tos seca y la fatiga son síntomas comunes. La diarrea y otros síntomas digestivos como malestar y dispepsia han sido informados en menos del 5% de los pacientes. Sin embargo, otros autores encontraron que las náuseas, el dolor abdominal y la diarrea pueden presentarse en el 50% de los pacientes, e incluso preceder a los síntomas respiratorios (Otero et al. 2020: p. 66). Luego de haber adquirido el virus, existe un periodo de 5-6 días (aunque puede variar desde 2 hasta 14 días) en el cual no aparece ningún síntoma. No se sabe con certeza, pero es probable que los infectados asintomáticos puedan contagiar durante este periodo. Los síntomas, cuando aparecen, duran aproximadamente dos semanas en casos leves, pero en casos severos pueden durar entre 3-6 semanas. La mayoría (80-85%) de los pacientes que se infectan con este virus no presentan síntomas o estos son leves, similar a un resfrío común; un 10-15% presentan neumonía que requiere hospitalización y el 5% restante corresponden a las formas más graves, que requieren ventiladores mecánicos y Unidades de Cuidados Intensivos (UCI). Es en este grupo de formas graves donde se presenta la mayor mortalidad (50%) según (Huang et al., 2020; citado en Salas Asencios et al., 2020: p. 12). Cuando existen síntomas, estos corresponden a los de una gripe común, con fiebre no elevada que dura 3-4 días, raras veces se presenta secreción nasal. Los síntomas más frecuentes son los respiratorios; al inicio hay fiebre (se presenta en más del 90% de casos), seguida de tos seca (70%). También son frecuentes al inicio de los síntomas los dolores musculares (mialgias), dolor de cabeza, sensación de fatiga o cansancio (40%) y síntomas digestivos, como vómitos o diarreas. Menos frecuente parece ser el dolor de garganta. Luego de alrededor de una semana del inicio de la enfermedad, si el paciente se agrava, puede presentarse dificultad para respirar y llevarlo hasta la muerte. El monitoreo y comunicación con el equipo médico es fundamental, para detectar tempranamente la dificultad respiratoria (Huang et al., 2020; citado en Salas Asencios et al., 2020: p. 12).

1.5.9. Prevención

La prevención está orientada a reducir la transmisibilidad del agente y por tanto el daño potencial que genera la enfermedad. Las medidas más importantes son la higiene de manos (lavado de manos o uso de alcohol-gel) y la higiene respiratoria (cubrirse la boca al toser o estornudar con el brazo o usando un papel). El uso de la mascarilla quirúrgica es para pacientes infectados y personas no infectadas sin síntomas, mientras que el uso de la mascarilla N-95 es para uso del personal de salud que atiende a los pacientes, personas que cuidan a un paciente afectado o están en contacto cercano con ellas. La bioseguridad del personal de salud, incluido el personal de

apoyo (limpieza, alimentación), es irrestricta. El uso del equipo de protección personal (EPP): guantes estériles, protectores oculares y faciales, mandilones descartables, etc., es mandatorio en el personal que atiende los pacientes (Villegas Chiroque, 2020, p. 4).

1.6. Bases conceptuales

1.6.1. *Series Temporales*

Una serie de tiempo (o serie cronológica) es “una secuencia de datos empíricos ordenados en función del tiempo; esto es, que puedan graficarse contra el tiempo”, por lo que es aplicable a un conjunto de datos registrados periódicamente, ya sea un registro diario, semanal, semestral, mensual o anual, entre otros. Los modelos de serie de tiempo pueden identificarse con la aplicación de la Metodología Box-Jenkins, mediante la cual se calculan las funciones de autocorrelación simple (FAS) y de autocorrelación parcial (FAP). Es sustancial saber que, cuando se trabaja con series de tiempo, una de las preguntas más importantes que se debe hacer el investigador sobre esta es: ¿cuál es el proceso generador de datos, del que proviene la muestra estudiada? (Contreras Juárez et al., 2016: p. 389).

1.6.1.1. *Objetivo de la serie de tiempos*

Según Chatfield (2013, pp. 7-9), son cuatro los objetivos para los cuales se analiza una serie:

- **Descripción:** Cuando se tiene la serie, el primer paso en el análisis es graficar los datos y obtener medidas descriptivas simples de las propiedades principales de la serie.
- **Explicación:** Cuando las observaciones son tomadas sobre 2 o más variables, es posible usar la variación en una serie para explicar la variación en otras series.
- **Predicción:** Dada una serie de tiempo se puede predecir valores futuros de la serie. Este es el objetivo más frecuente del análisis de series de tiempo.
- **Control:** Cuando una serie de tiempo se genera por mediciones de calidad de un proceso, el objetivo de análisis puede ser el control del proceso (Chatfield, 2013, pp. 7-9; citado en Zavala Guevara, 2017b, p. 31).

1.6.1.2. *Componentes de la serie de tiempo*

Para describir una serie de tiempo se deben identificar los siguientes componentes:

Tendencia o componente tendencial: Una serie de tiempo tiene tendencia cuando por largos periodos los valores crecen o decrecen, también puede definirse como cambios en la media (Zavala

Guevara, 2017b, p. 32). Es el resultado de factores que afectan a largo plazo a los datos, los cuales generan un patrón gradual y consistente de las variaciones de la misma serie (Páez, González y Duarte, 2016: p. 81). Indica la dirección hacia la cual se dirige la serie de tiempo, característica que lo convierte en el componente más importante. Puede ser creciente, decreciente, constante, lineal, curvilínea, entre otras; se llama también tendencia a largo plazo (León Álvarez et al., 2016: p. 374).

Componentes cíclicas y estacionales: tanto las variaciones cíclicas como las estacionales, pueden entenderse como el movimiento ondulante de los datos por encima y por debajo de la línea de tendencia, son procesos que ocurren de manera más espaciada dentro del intervalo de observación, mientras que las componentes estacionales ocurren con menor separación en el tiempo (Páez, González y Duarte, 2016: p. 82). El ciclo indica las variaciones que ocurren en una serie de tiempo en períodos más prolongados. Cuando la métrica es en años son variaciones mayores de un año, comúnmente de 2 a 10 años. La serie sube y baja suavemente a manera de onda siguiendo la tendencia; mientras que, la estacionalidad indica las variaciones que ocurren a corto plazo en una serie de tiempo con respecto a la línea de tendencia general. Ocurre en períodos fijos como días, semanas, meses, trimestres o años (León Álvarez et al., 2016: p. 375).

Componente irregular o aleatorios: que se relaciona con datos muy desviados de la tendencia. En este caso, es necesario tomar decisiones en cuanto al manejo que se le da a estos valores anormales, pues incluirlos en los cálculos añadiría una gran cantidad de ruido a la descripción de estos. Por tal razón, suele estimarse un valor, para reemplazar el dato anormal (Páez, González y Duarte, 2016: p. 82).

1.6.1.3. Autocorrelación ACF Y PACF

La autocorrelación se define como que los errores del modelo no son independientes entre sí, esto quiere decir que la $E(\mu_i, \mu_j) \neq 0$ cuando $i \neq j$, por ejemplo, en el estado de ánimo estar bien a lo largo de una semana, si esta triste, enojado, melancólico eso ya vendría a ser un error, entonces los cambios imprevistos en el estado de ánimo están relacionados o nos; que los errores están autocorrelacionados cuando existen cambios imprevistos de un tiempo a otro.

El ACF significa función de autocorrelación y el PACF para función de autocorrelación parcial. La autocorrelación calcula y traza las autocorrelaciones de una serie de tiempo. La autocorrelación es la correlación entre las observaciones de una serie de tiempo separadas por k unidades de tiempo, mientras que el ACF es el gráfico utilizado para mostrar la correlación entre los puntos, hasta la unidad de retraso inclusive. En ACF, el coeficiente de correlación está en el eje x , mientras que se muestra el número de rezagos en el eje y . De manera similar, las autocorrelaciones parciales miden la fuerza de la relación con otros términos que se tienen en

cuenta. En este caso, otros términos son los rezagos en el modelo que están interviniendo. Una autocorrelación parcial es una combinación de la relación entre una observación en una serie de tiempo con las observaciones que se excluyen en la fase de inicio con las relaciones de las observaciones intermedias (Khan y Gupta, 2020: p. 13).

1.6.1.4. Series de tiempo estacionarias

Una serie de tiempo estacionaria es **estable a lo largo del tiempo**, es decir, su media y su varianza son constantes en el tiempo: los valores de la serie de tiempo tienden a oscilar alrededor de una media constante y la variabilidad respecto a la media también permanece constante en el tiempo. Una serie de tiempo no estacionaria se caracteriza porque **la tendencia o variabilidad cambian en el tiempo** (Nossa, Angulo y Torres, 2017: p. 4).

Es importante aclarar que un modelo ARIMA como estrategia de predicción solo tendrá sentido si las características observadas en la serie permanecen en el tiempo. Lo anterior define el concepto de un modelo Estacionario, que se fundamenta en dos requisitos (León Álvarez et al., 2016: p. 377):

1. Que la media y la varianza de los datos sean constantes en el tiempo \bar{y} ,
2. Que la estructura de la covarianza (el grado de variación conjunta de dos variables aleatorias) entre dos períodos diferentes de tiempo dependa solamente de la distancia o rezago entre estos dos períodos y no del tiempo en el que se ha calculado dicha covarianza (Asteriou y Hall, 2011; citado en León Álvarez et al., 2016: p. 377).

Una serie de tiempo estacionaria es aquella cuyas propiedades no dependen del momento en que se observa la serie, cabe indicar que una serie de tiempo estacionaria no tendrá patrones predecibles a largo plazo; las series de tiempo con tendencias, o con estacionalidad, no son estacionarias; la tendencia y la estacionalidad afectarán el valor de la serie de tiempo en diferentes momentos. Por otro lado, una serie de ruido blanco es estacionaria; no importa cuando la observe, debería verse muy similar en cualquier momento. Los gráficos revelarán que la serie es aproximadamente horizontal con varianza constante, la estacionariedad se puede verificar mediante la autocorrelación y pruebas de hipótesis, para que una serie sea estacionaria su media, varianza y covarianza deben ser constante, no tiene tendencia y es independientes del tiempo (Hyndman y Athanasopoulos, 2018: p. 291). Para que sea estacionaria una serie debe cumplir con las siguientes propiedades:

- La esperanza de y_t es constante a lo largo del tiempo, no importa que estén rezagados a periodos k , entonces $E(y_t) = E(y_{t-1})$.
- La varianza de y_t es constante a lo largo de toda la serie $Var(y_t) = \sigma^2$

- La covarianza entre dos valores diferentes será la misma siempre y cuando estén separadas en k periodos $Cov(y_t, y_{t-k}) = Cov(y_{t-i}, y_{t-i-k})$.

Es estacionaria en medias, debido a que la serie fluctúa alrededor de un valor fijo siendo esta su media y los intervalos a los cuales fluctúa es la varianza, es estacionaria en tendencia si el incremento o disminución fluctúa alrededor de una línea de tendencia y cuya varianza también permanece constante. Una serie estacionaria se formula de la siguiente manera:

$y_t = \rho y_{t-1} + \varepsilon_t$, $\rho < 1$ debido a que si $\rho = 1$ esta serie ya no sería estacionaria

$$y_t = \rho(\rho y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$y_t = \rho^2 y_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t$$

$$y_t = \rho^2 (y_{t-3} + \varepsilon_{t-2}) + \rho \varepsilon_{t-1} + \varepsilon_t$$

$$y_t = \rho^3 y_{t-3} + \rho^2 \varepsilon_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t$$

⋮
⋮
⋮

$$y_t = \rho^{k+1} y_{t-(k+1)} + \sum \rho^k \varepsilon_{t-k}$$

Como $\rho < 1$ se esperaría que $\rho^{k+1} \rightarrow 0$ cuando $k \rightarrow \infty$ entonces, $y_t = \sum \rho^k \varepsilon_{t-k}$; cabe recordar que las series estacionarias tienen memoria a corto plazo, mientras que las series no estacionarias son las que tienen memoria de largo plazo.

1.6.1.5. Series no estacionarias

Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante. Pueden permanecer durante periodos largos con valores por encima o por debajo de la media, debido a la presencia de una tendencia determinística, al problema de raíces unitarias o a la presencia de una varianza heterocedástica. Las series de tiempo no estacionarias pueden transformarse en series de tiempo estacionarias aplicando los correspondientes procedimientos. Las series de tiempo pueden ser no estacionarias en media y no estacionarias en varianza (Nossa, Angulo y Torres, 2017: p. 4).

1.6.1.6. Raíz Unitaria

Decir que una serie posee raíz unitaria, equivale a decir que la serie es no estacionaria. “El número de raíces unitarias equivale al número de veces que se tiene que diferenciar una serie para hacerla estacionaria. Así, se dice que una serie $I(1)$ tiene una raíz unitaria y que una serie $I(d)$ tiene d raíces unitarias” (Brugger, 2010; citado en Guachalá Coba, 2017, p. 17).

Se realiza tests de raíces unitarias, para analizar las propiedades de la estacionariedad, si la serie o proceso tiene raíz unitaria, la serie no es estacionaria y los estimadores MCO no tienen distribución normal. Una raíz unitaria, es una tendencia estocástica en la serie temporal. Algunas veces se le llama “paseo aleatorio con deriva”. Por tanto, si la serie tiene una raíz unitaria, ésta presenta un patrón sistemático que es impredecible. Entonces, una serie temporal es estacionaria si un cambio en el tiempo no cambia la forma de la distribución; y las raíces unitarias son una causa de no estacionariedad. Existen dos procedimientos comunes para remover la tendencia de una serie temporal: Las primeras diferencias suelen ser apropiadas para series temporales integradas de orden uno $-I(1)-$; y la regresión tendencial es apropiada para series con estacionariedad tendencial $I(0)$. Se comprueba con el test de Dickey-Fuller Aumentado y el test de Phillips-Perron (López, 2018).

1.6.1.7. Test Augmented Dickey Fuller (ADF)

Es el test de Dickey-Fuller aumentada, la cual elimina la autocorrelación e indica si una serie es estacionaria o no (Marul y López, 2018): determina la estacionariedad de las variables mediante la prueba de raíz unitaria ADF, esta prueba incluye la tendencia y la intersección para verificar la estacionariedad de cada variable, elimina la autocorrelación e indica la estacionariedad de la serie (Rehman et al., 2019: p. 53).

Se propone un análisis que determine la estacionariedad de las variables mediante el uso de la prueba Dickey-Fuller aumentada (DFA) para evitar obtener resultados adulterados en regresiones con series de tiempo no estacionarias. Stock y Watson (2012) mencionan que el contraste DFA o ADF para una raíz unitaria autorregresiva contrasta la hipótesis nula $H_0: \delta = 0$ frente a la hipótesis alternativa $H_1: \delta < 0$, en la siguiente regresión (Quinde Rosales, Víctor et al., 2020: p. 62):

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + \varepsilon_t$$

Interpretado de otra forma las hipótesis representan:

H_0 : La variable es no estacionaria; tiene raíz unitaria

H_1 : La variable es estacionaria; no tiene raíz unitaria

El test DFA ayuda a detectar la autocorrelación, no posee una distribución estándar, por lo cual no puede utilizar los valores críticos habituales de la distribución normal para su contraste de raíz unitaria, en su lugar debe utilizarse un conjunto particular de valores críticos con base en la distribución del estadístico DFA con hipótesis nula H_0 ; la hipótesis alternativa de estacionariedad

implica que en la ecuación antes descrita sea $\delta < 0$, el contraste es unilateral. Por lo tanto, el valor del estadístico debe ser menor (más negativo) que los valores críticos unilaterales para rechazar H_0 la hipótesis nula (Quinde Rosales, Victor Xavier et al., 2020: p. 62). La hipótesis nula para ambas pruebas es que los datos no son estacionarios. Queremos RECHAZAR la hipótesis nula para esta prueba, por lo que queremos un valor de p menor que 5% (Holmes, Scheuerell y Ward, 2019: p. 126).

1.6.1.8. Grado de integración

Si la serie de tiempo no muestra estacionariedad, existe una forma de hacerla estacionaria a través del cálculo de las diferencias entre observaciones consecutivas, esto se conoce como diferenciación o procesos estocásticos no estacionarios homogéneos de orden de integración o número de diferenciaciones d , la cual puede ayudar a estabilizar la media de una serie temporal al eliminar los cambios en el nivel de una serie temporal y, por lo tanto, eliminar (o reducir) la tendencia y la estacionalidad (Hyndman y Athanasopoulos, 2018: pp. 291-292). Cuando la serie en estudio no es estacionaria, se la diferencia un número determinado de veces, para transformarla en estacionaria, antes de esto ya se podría aplicar la metodología de los modelos *ARMA*. “Cuando la diferenciación no es suficiente, es necesario transformar la serie tomando logaritmos neperianos y posteriormente realizar la diferenciación. En la práctica se puede presumir que en todas las series económicas se puede eliminar la tendencia mediante la diferenciación que en general para las series económicas es suficiente con una o dos diferencias” (Bórras, 2004; citado en Guachalá Coba, 2017, p. 16) .

La primera diferencia está definida como $\Delta y_t = y_t - y_{t-1} = \varepsilon_t$ un ruido blanco, haciendo que la serie sea estacionaria, a una segunda diferencia se la define como $\Delta^2 X_t = X_t - 2X_{t-1} + X_{t-2}$, a partir de la cual se define la expresión general del modelo *ARIMA(p,d,q)* como sigue (Guachalá Coba, 2017, p. 16):

$$\Delta^d X_t = \phi_1 \Delta^d X_{t-1} + \dots + \phi_p \Delta^d X_{t-p} + a_t + \dots + \phi_q a_{t-q}$$

donde $\Delta^d X_t$, expresa que sobre la serie original X_t , se han aplicado d diferencias.

1.6.1.9. Ruido Blanco

Se dice que una serie de tiempo (o proceso) es puramente aleatorio (o de ruido blanco) si tiene una media de cero y varianza constante y no está seriamente correlacionado (Zavala Guevara, 2017a, p. 32). El ruido blanco es una señal aleatoria (proceso estocástico) que se caracteriza por el hecho de que sus valores de señal en dos tiempos diferentes no guardan correlación estadística, es decir, en el eje del tiempo la señal toma valores sin ninguna relación unos con otros. El término ruido

blanco surge de un análisis de frecuencia del modelo muestra que, en analogía con la luz blanca, todas las frecuencias entran por igual (Fournies, 2015; citado en Ayala Jacinto y Samillán Vázquez, 2019: p. 34). Por tanto, un proceso estacionario a_t es un ruido blanco si: $E(a_t) = 0$, $Var(a_t) = \sigma^2$, $r_k = 0$ y $a_t \sim N(0, \sigma)$.

Es importante tener en cuenta que un ruido blanco es un ejemplo de proceso estocástico estacionario de segundo orden, las autocorrelaciones simples y parciales de un procedimiento de ruido blanco son igual a cero excepto en el rezago 0 debido que es igual a uno y decrece rápidamente a cero.

1.6.2. Metodología de Box Jenkins

Box y Jenkins (1976) publican *Time Series Analysis: Forecasting and Control* imponiendo una nueva herramienta de pronóstico, esta herramienta, conocida como metodología de Box-Jenkins (BJ), hoy es identificada como metodología ARIMA. El método BJ utiliza una estrategia iterativa para construir los modelos, que consiste en seleccionar un modelo inicial mediante la estimación de los coeficientes del modelo y el análisis de los residuales (Muñoz Santiago et al., 2017: p. 135).

1.6.2.1. Pasos de la metodología Box y Jenkins

Para la metodología se hace uso del modelo ARIMA estacional de Box y Jenkins (1976), donde los pasos básicos de esta metodología consisten en lo siguiente (Blanco Laurente, 2018, pp. 12-13):

- **Análisis preliminar:** realizar un análisis preliminar a los datos de tal modo que sean un proceso estocástico estacionario.
- **Identificación de un modelo tentativo:** especificar el orden (p, d, q) del modelo ARIMA, para ello las funciones de autocorrelación ayudan para el reconocimiento del modelo.
- **Estimación del modelo:** el siguiente paso es la estimación del modelo ARIMA identificado en el paso anterior. La estimación puede ser realizada por el método de mínimos cuadrados o máxima verosimilitud.
- **Diagnóstico de resultados:** se realiza la revisión del modelo usando test para los parámetros y residuos.
- **Proyección:** si el modelo ganador del paso anterior es el adecuado, entonces el modelo puede ser usado para la representación y proyección.

1.6.2.2. Criterios de selección del modelo

Para la estimación del modelo estacionario y parsimonioso que tenga un buen ajuste se emplean criterios de selección de modelos tales como el criterio de información de Akaike (AIC), criterio de información de Schwarz (SC) y/o el Final Prediction Error (FPE) los cuales permiten hacer comparaciones entre distintas especificaciones y de esta manera seleccionar el mejor modelo (Rojas, 2019, p. 20).

Akaike (1976) estudió de manera concisa un criterio de información (AIC) que se utiliza como una estimación de parámetro para verificar el desempeño del Modelo que da historia para el desarrollo de pruebas de teoría estadística en análisis de series de tiempo y señalando que el procedimiento de prueba de teoría no se define efectivamente como la técnica para la identificación de modelos estadísticos, el AIC no es más que la estimación del criterio de teórico de mínima información. Para el modelo $p > 0$, $q > 0$, mediante los ACF y PACF es difícil analizar la selección de orden, para lo cual se introduce el enfoque mediante la minimización de los Criterios de información de Akaike (AICc) corregidos. Si el cálculo AIC se basó en la fórmula matemática algorítmica (Akaike, 1976, p. 5-9; citado en Jain y Mallick, 2017: pp. 2-5):

$$AIC = -2\text{Log } L + 2m$$

$$AICc = AIC + \frac{2(p + q + K + 1)(p + Q + k + 2)}{n - p - K}$$

Donde $m = p + q + K + 1$ es el número de parámetros en el modelo y L es la función de verosimilitud. El mejor modelo es el que tiene el valor AIC más bajo. De este criterio aparece la formulación matemática correspondiente a los criterios de información bayesiano BIC:

$$BIC = AIC + \log (T) (p + q + P - 1)$$

Siendo el criterio que permite la selección final de los órdenes p y q , existiendo la probabilidad basada en la estimación de máxima verosimilitud.

1.6.2.3. Medidas del Error

Estas medidas permiten determinar que tan bien se ajusta un modelo a los valores reales y para determinar el rendimiento de los pronósticos, el error de pronóstico es la diferencia entre el valor predicho o pronosticado y el valor real $e_t = y_t - \hat{y}_t$, se considera 3 criterios para evaluar el desempeño del pronóstico el error absoluto medio (MAE), el error cuadrático medio (RMSE) y el error porcentual absoluto medio (MAPE), las cuales indican:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| = \text{mean}(e_t)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100$$

Donde

y_i : Valores reales obtenidos en el estudio

\hat{y}_i : Valores predichos o pronosticados por el modelo

n : Número de observaciones

El desempeño del pronóstico es mejor cuando el valor es menor, y si los resultados no son consistentes entre las tres medias Makridakis (1993) siguiere analizar la medida MAPE, el cual es considerado como indicador básico del modelo y lo toma como punto de referencia con relativa estabilidad que otros dos criterios.

1.6.3. *ARIMA - Modelo autorregresivo integrado de media móvil*

Se deriva de sus componentes: AR (Autorregresivo), I (Integrado) y MA (Medias móviles). Es importante aclarar que un modelo ARIMA como estrategia de predicción solo tendrá sentido si las características observadas en la serie permanecen en el tiempo. Lo anterior define el concepto de un modelo Estacionario, que se fundamenta en dos requisitos (León Álvarez et al., 2016: p. 377).

1. Que la media y la varianza de los datos sean constantes en el tiempo, es decir, $E(X_t) = \mu$ y la $Var(X_t) = \sigma^2$.
2. Que la estructura de la covarianza entre dos períodos diferentes de tiempo dependa solamente de la distancia o rezago entre estos dos períodos y no del tiempo en el que se ha calculado dicha covarianza. Es decir, la correlación entre X_t y X_{t-i} depende exclusivamente del número de demoras que las separan $\rho(t, i) = r_i$.

Decimos que un proceso o modelo es estacionario en sentido estricto si las funciones de distribución conjuntas como la media, las varianzas, las covarianzas y las funciones de distribución “completas”, son constantes o invariantes con respecto al desplazamiento en el tiempo (León Álvarez et al., 2016: p. 377).

1.6.3.1. Etapas del modelo ARIMA

El procedimiento original de la metodología ARIMA involucra un proceso iterativo de distintas etapas y se basa en dos grandes principios: el principio de parsimonia, que consiste en seleccionar siempre el modelo más sencillo que sea suficientemente representativo de los datos y el principio del mejoramiento iterativo, que consiste en partir de un modelo sencillo y factible, al cual se harán mejoramientos sucesivos, hasta llegar a un modelo satisfactorio, las fases a seguir son (Rosas Rojas et al., 2017: p. 3-4)

- i) **Análisis de Estacionalidad y Estacionariedad:** Se debe verificar que la serie sea estacionaria e intentar eliminar la variación estacional, si esta existe.
- ii) **Identificación del Modelo Tentativo:** Se utilizan los correlogramas simple y parcial, para determinar un patrón de las diferentes autocorrelaciones y para evaluar si existe algún tipo de variación estacional, no eliminada por las diferencias estacionales.
- iii) **Estimación de Parámetros del Modelo:** Para cada modelo tentativo se deberán estimar los parámetros. Esto se puede realizar mediante el método de mínimos cuadrados ordinarios, la ecuación de Yule-Walker o el método de máxima verosimilitud.
- iv) **Verificación del Modelo:** Se deben calcular los estadísticos clásicos para medir el ajuste de los datos utilizando medidas de error y/o criterios de información (AIC o BIC), la significancia estadística de los coeficientes estimados, pruebas de raíz unitaria sobre los residuos del modelo, también se debe verificar que no exista autocorrelación.
- v) **Pronóstico:** Los pronósticos serán construidos utilizando el método de iteración adelantada. Es importante señalar que la calidad del pronóstico declina a medida que éste se prolonga.
- vi) **Validación del pronóstico:** Se debe efectuar una prueba Diebold-Mariano, en esta prueba se busca una función objetivo que no sea cuadrática.

1.6.3.2. Expresión general del modelo ARIMA

La metodología Box-Jenkins se refiere a una serie de procedimientos para identificar, ajustar y verificar los modelos Auto Regresivos (AR(p)), Integrados (I(d)) y de Promedios Móviles (MA(q)), más conocido por sus siglas en inglés ARIMA, con los datos de serie de tiempo, el modelo se expresa de la siguiente manera (Echegaray Munenaka, 2017: p. 25):

$$ARIMA(p, d, q) \times (P, D, Q)_S$$

Donde:

p : representa el orden del proceso autorregresivo no estacional

d : es el número de diferencias utilizadas para eliminar la tendencia

q : representa el orden del proceso de medias móviles no estacional

P : representa el orden del proceso autorregresivo estacional

D : es el número de diferencias utilizadas para eliminar el factor estacional

Q : representa el orden del proceso de medias móviles estacional

S : representa el periodo del factor estacional

1.6.3.3. Modelo autorregresivo de orden p : $AR(p)$

Se basa en la idea de que el valor actual de la serie X_t , puede explicarse en función de p valores pasados $X_{t-1}, X_{t-2}, \dots, X_{t-p}$, donde p , determina el número de rezagos necesarios para pronosticar un valor actual (Paredes Rodríguez, 2017, p. 26).

$$AR(p) = ARIMA(p, 0, 0) \times (0, 0, 0)_0$$

Diremos que $\{X_t\}$ es un proceso autorregresivo $AR(p)$ si $\phi_p(B) = a_t + \varepsilon$, en el cual $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p$. El modelo lineal de $AR(p)$ puede ser:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t + \varepsilon = \sum_{i=1}^p \phi_i X_{t-i} + a_t + \varepsilon$$

Donde $\{a_t\}$ es un proceso de ruido blanco independiente de X_t . El $AR(p)$ caracteriza porque los primeros p coeficientes de la función de autocorrelación parcial son no nulos, el resto cero y la función de autocorrelación simple en general presenta muchos coeficientes no nulos que decrecen con el retardo en una mezcla de exponenciales y sinusoidales (Echegaray Munenaka, 2017: p. 26).

1.6.3.4. Modelo móvil de orden q : $MA(q)$

En el caso donde solo un número finito de los pesos no es cero, tenemos lo que es llamado proceso de promedio móvil, estos son indicadores que allanan o suavizan, en mayor o menor medida, de tal manera que eliminan determinadas fluctuaciones, sean a corto, medio o largo plazo (Cryer y Chan, 2008; citados en Ayala Jacinto y Samillán Vásquez, 2019: p. 37).

$$MA(q) = ARIMA(0, 0, q) \times (0, 0, 0)_0$$

Diremos que $\{X_t\}$ es un proceso de media móvil $MA(q)$ si $X_t = \theta_q(B) a_t + \varepsilon$, en el cual $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_q B^q$. El modelo lineal de $MA(q)$ puede ser:

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} - a_{t-p} + \varepsilon = a_t + \sum_{i=1}^q \theta_i a_{t-i} + \varepsilon$$

Donde $\{a_t\}$ es un proceso de ruido blanco independiente de X_t y $\theta_1, \theta_2, \dots, \theta_q$ son los parámetros del modelo siendo los “ponderadores” de la combinación lineal. El $MA(q)$ se caracteriza porque los primeros q coeficientes de la función de autocorrelación simple son no nulos, el resto cero y la función de autocorrelación parcial presenta varios coeficientes no nulos que decrecen con el retardo en una mezcla de exponenciales y sinusoidales (Echegaray Munenaka, 2017: p. 27).

1.6.3.5. Modelo autorregresivo de medias móviles de orden (p, q) : **ARMA (p, q)**

Se caracteriza porque los primeros q coeficientes de la función de autocorrelación simple son no nulos, el resto cero y los primeros p coeficientes de la función de autocorrelación parcial son no nulos y el resto cero (Echegaray Munenaka, 2017, p. 28).

$$ARMA(p, q) = ARIMA(p, 0, q) \times (0,0,0)_0$$

Diremos que $\{X_t\}$ es un proceso $ARMA(p, q)$ si $\phi_p(B)X_t = \theta_q(B)a_t + \varepsilon$. De lo anterior se deduce que un proceso $ARMA(p, q)$ es una mezcla de un proceso $AR(p)$ y un $MA(q)$, expresada como: $X_t = AR(p) + MA(q)$

1.6.3.6. Modelo autorregresivo integrado de medias móviles de orden (p, d, q) : **ARIMA (p, d, q)**

Este se define como $ARIMA(p, d, q) = ARIMA(p, d, q) \times (0,0,0)_0$. Diremos que $\{X_t\}$ es un proceso $ARIMA(p, d, q)$ si $\phi_p(B)\nabla^d X_t = \theta_q(B)a_t + \varepsilon$, donde $\{a_t\}$ es un proceso de ruido blanco independiente de X_t . Tomando $w_t = \nabla^d X_t$, tenemos $\phi_p(B)w_t = \theta_q(B)a_t + \varepsilon$. El proceso $ARIMA(p, d, q)$ para X_t , es equivalente al proceso $ARMA(p, q)$ para w_t . La modelación se hará sobre el proceso diferenciado w_t , considerando la función inversa de w obtendremos la serie original $X_t = f(w_t)^{-1}$ (Echegaray Munenaka, 2017, p. 29).

La forma general del modelo ARIMA se muestra en la siguiente Ecuación:

$$\nabla^d X_t = \mu + \phi_1 \nabla^d X_{t-1} + \phi_2 \nabla^d X_{t-2} + \dots + \phi_p \nabla^d X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Si no se da alguna diferencia se anula ∇^d de la ecuación.

1.6.3.7. Modelo autorregresivo integrado de medias móviles estacional de período estacional **S**. Si existía una estacionalidad evidente en los datos, utilizamos el modelo estacional ARIMA ($SARIMA(p, d, q) \times (P, D, Q)_s$). Las letras p, q , y d corresponden al significado del orden de

autorregresión, grado de diferencia, y el orden de media móvil, respectivamente. P , D , Q y S significa autoregresión estacional, integración estacional, promedio móvil estacional y duración del período estacional (He y Tao, 2018: p. 62). Diremos que $\{X_t\}$ es un proceso Autorregresivo integrado de medias móviles de periodo S si $\Phi_P(B^S) \phi_p(B) \nabla^d \nabla_{S^d} X_t = \Theta_Q(B^S) \theta_q(B) a_t + \varepsilon$, donde $\{a_t\}$ es un proceso de ruido blanco independiente de X_t ; ∇ es el operador de diferencia no estacional, ∇_S es el operador de diferencia estacional, d es el orden de diferenciación regular y D es el orden de diferenciación estacional. Tomando $w_t = \nabla^d \nabla_{S^d} X_t$, tenemos $\Phi_P(B^S) \phi_p(B) w_t = \Theta_Q(B^S) \theta_q(B) a_t + \varepsilon$. El proceso $ARIMA(p, d, q) \times (P, D, Q)_S$ para X_t es equivalente al proceso ARMA estacional w_t . Las funciones de autocorrelación simple y parcial de un proceso ARMA estacional son una mezcla de sus componente regulares y estacionales (Echegaray Munenaka, 2017: p. 30).

1.6.4. *Bootstrap*

El método bootstrap propuesto por Efron y Tibshirani (1986), es una herramienta poderosa para estimar la distribución muestral de un estadístico dado. La idea básica detrás del Bootstrap es que la inferencia sobre una población a partir de datos de muestra se pueda modelar volviendo a muestrear los datos de muestra y realizando inferencias sobre una muestra a partir de datos remuestreados. Como la población es desconocida, se desconoce el verdadero error en un estadístico de muestra respecto a su valor poblacional. En remuestras Bootstrap, la “población” es de hecho la muestra, y esto es conocido; por lo tanto, la calidad de la inferencia de la muestra “verdadera” de los datos remuestreados es medible (Efron y Tibshirani, 1986; citados en Larco Álvarez, 2018, p. 32).

Sea X una variable aleatoria que sigue una distribución desconocida F , sea $\mathbf{X}=(X_1, \dots, X_n)$ una muestra aleatoria simple de X , y sea $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Se introduce \mathbf{X} para indicar que son muestras aleatorias simples. El método bootstrap plantea como solución, aproximar el estadístico sustituyendo una distribución desconocida por una estimación de F . En función de qué estimación tomemos de F , podemos diferenciar entre dos tipos de Bootstrap (Muñoz 2019, p. 30):

- **Bootstrap no paramétrico:** Parte de que no conocemos el modelo que sigue la muestra aleatoria simple, y estima F usando la función de distribución empírica F_n .
- **Bootstrap paramétrico:** en este caso sabemos el modelo F_θ que sigue la muestra, pero no sabemos sus parámetros $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^k$. Se basa en obtener una estimación $\hat{\theta}$ de los parámetros y estimar F_θ mediante \hat{F}_θ .

Los métodos Bootstrap son robustos debido a que no asumen ninguna distribución particular de la población y han sido aplicados a diversos problemas de muestreo, sustituyendo la complejidad del análisis teórico por el poder de cálculo de los ordenadores (Efron y Tibshirani, 1994; citado en Orosco Gavilán, 2017, p. 24).

1.6.4.1. Bootstrapping en series de tiempo

Bootstrapping es una técnica bien conocida que se utiliza para estimar las propiedades de una estadística. Fue desarrollado por Bradley Efron (1979). Los casos de uso más comunes incluyen la estimación de variaciones y / o intervalos de confianza (Gonzalez, 2018). En series de tiempo se hace Bootstrap sobre los residuos para simular valores futuros de una serie usando un modelo. De manera más general, podemos generar nuevas series de tiempo que son similares a nuestras series observadas, esto se ejecuta mediante la transformación Box-Cox la cual incluye tanto logaritmos como transformaciones de potencia, y la descomposición de componentes estacionales, de tendencia y residuos a través de la metodología STL (Descomposición estacional y de tendencias usando Loess), este último es un método versátil y robusto para descomponer series de tiempo, mientras que Loess es un método para estimar relaciones no lineales. El método STL fue desarrollado por Cleveland, Cleveland, McRae y Terpenning (Hyndman y Athanasopoulos, 2018).

1.6.4.2. Bootstrap de bloques

El bootstrap de bloques intenta replicar la correlación volviendo a muestrear bloques de datos en lugar de valores individuales. Puede usarse con datos correlacionados en el tiempo (es decir, series de tiempo) pero también se puede usar con datos correlacionados en el espacio o entre grupos (Gonzalez, 2018). Este método fue propuesto por Liu y Tang (1996), el cual consiste en remuestrear bloques de observaciones consecutivas en el tiempo, de ahí resulta una estructura de dependencia de las observaciones originales se preserva dentro de cada bloque (Larco Álvarez, 2018, p. 35).

Künsch (1989) y los investigadores Liu y Singh (1992), propusieron de forma independiente el llamado bootstrap por bloques (moving blocks bootstrap o MBB). El método procede del siguiente modo (Cao Abad y Fernández Casal, 2020: p. 136):

1. Fijar un entero positivo, b , el tamaño del bloque, y tomar k igual al menor entero mayor o igual que $\frac{n}{b}$.
2. Definir los bloques (o submuestras):

$$B_{i,b} = (X_i, X_{i+1}, \dots, X_{i+b-1})Bi$$
, o simplemente B_i , para $i = 1, 2, \dots, q$ ($q = n - b + 1$).

3. Arrojar k observaciones (bloques), $\xi_1, \xi_2, \dots, \xi_k$, con distribución equiprobable sobre el conjunto de posibles bloques: $\{B_1, B_2, \dots, B_q\}$. Cada ξ_i es un vector b -dimensional $(\xi_{i,1}, \xi_{i,2}, \dots, \xi_{i,b})$.
4. Definir \vec{X}^* como el vector formado por las n primeras componentes de $(\xi_{1,1}, \xi_{1,2}, \dots, \xi_{1,b}, \xi_{2,1}, \xi_{2,2}, \dots, \xi_{2,b}, \dots, \xi_{k,1}, \xi_{k,2}, \dots, \xi_{k,b})$.

Hay que tener presente que si $b = 1$, entonces $k = n$ y con ello se obtiene un Bootstrap ordinario, pero si $b = n$, entonces $k = 1$ esto hará que tengamos un Bootstrap degenerado debido a que todas las réplicas coinciden con la muestra o serie original.

La selección de b es importante antes de aplicar Bootstrap de bloques, según Hall, Horowitz y Jing (1995) considera este problema en el contexto de la estimación bootstrap del sesgo y la varianza. Obtienen una expresión asintótica para el error cuadrático medio $n^{-2}(C_1 b^{-2} + C_2 n^{-1} b)$, donde C_1 y C_2 son constantes desconocidas que dependen del problema de estimación del que se trate. A partir de esto se concluye que el tamaño óptimo del bloque (en el sentido del error cuadrático medio) es de orden $n^{1/3}$ (Cao Abad y Fernández Casal 2020, p. 137).

1.6.4.3. Estimación Bootstrap del Error Típico

El bootstrap fue introducido por Efron y Tibshirani (1994) como un método basado en cálculos intensivos mediante un ordenador para estimar el error muestral de un estadístico. Explícitamente, la estimación bootstrap del error de un estadístico es como sigue (Orosco Gavilán, 2017, pp. 25-26):

- Se extrae B muestras Bootstrap independientes en función de la distribución empírica \hat{F} .
- Se obtiene $\hat{\theta}_b^*$, computarizando el estadístico de interés en cada una de las B muestras.
- Se estima el error muestral $\hat{\theta}$ mediante la desviación estándar de la función de distribución obtenida de los $B\hat{\theta}_i^*$, es decir a través de $\hat{F}^*(\hat{\theta}^*)$.

$$\widehat{se}_B = \sqrt{\frac{\sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}^*(\cdot))^2}{(B-1)}}, \text{ donde } \hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}_b^*}{B} \text{ (Error estándar)}$$

1.6.4.4. Intervalo de Confianza para predicción de series Bootstrap

Existen al menos cuatro fuentes de incertidumbre en la predicción mediante modelos de series de tiempo: el término de error aleatorio, las estimaciones de los parámetros, la elección del modelo para los datos históricos, y la continuación del proceso de generación de datos históricos en el futuro. Cuando se genera intervalos de predicción en general se tiene en cuenta la primera

incertidumbre acerca del error aleatorio el resto va apareciendo, lo primero que se debe hacer es simular series similares a las originales, para cada serie se ajusta un modelo de espacios de estados ETS (Error tendencia y estacionalidad) también se puede considerar como suavizado exponencial, las posibilidades para cada componente son:

Error = {A, M}, con A: aditivo y M: multiplicativo

Tendencia = {N, A, A_{re}}, con N: ninguno, A: aditivo y A_{re}: aditivo amortiguado; y,

Estacional = {N, A, M}, con N: ninguno, A: aditivo y M: multiplicativo.

Se simula las rutas de muestra del modelo ETS seleccionado, estas las rutas permitirán la incertidumbre del modelo y la incertidumbre de los parámetros, así como la incertidumbre asociada con el término de error aleatorio, a partir de esto se toma las medias y los cuantiles de estas rutas de muestra simuladas para formar pronósticos puntuales e intervalos de predicción (Hyndman y Athanasopoulos 2018).

Miranda (2003), describe algunos métodos frecuentemente utilizados a través de los cuales se pueden construir intervalos de confianza bootstrap y se hablara de dos (Orosco Gavilán, 2017, p. 27):

- **El método de aproximación normal** utiliza la misma estructura de los procedimientos paramétricos de la construcción de intervalos de confianza. Si es posible asumir que el estadístico se distribuye según la curva normal pero el cálculo del error típico resulta analíticamente difícil o no existe fórmula para su cálculo, entonces podemos emplear la distribución muestral bootstrap para estimar el error típico o insertarlo en la correspondiente expresión del intervalo de confianza paramétrico. Así el intervalo de confianza bootstrap será de la forma:

$$\left[\hat{\theta} - \widehat{se}_R Z_{1-\frac{\alpha}{2}} ; \hat{\theta} + \widehat{se}_R Z_{1-\frac{\alpha}{2}} \right] \text{ donde } \widehat{se}_R = \sqrt{\frac{\sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}^*(\cdot))^2}{(B-1)}}$$

- **El método del percentil** hace uso literal de la idea básica del bootstrap, es decir $\hat{F}^*(\hat{\theta}^*)$ se aproxima a $F(\hat{\theta})$. La idea es muy simple: un intervalo con un nivel de confianza $1 - \alpha$ incluye todos los valores de $\hat{\theta}^*$ entre los percentiles $\frac{\alpha}{2}$ y $(1 - \frac{\alpha}{2})$ de la distribución de $\hat{F}^*(\hat{\theta}^*)$. En este caso, el intervalo esta dado de la forma $[\hat{F}^{-1}(\alpha/2) ; \hat{F}^{-1}(1 - \alpha/2)]$, donde $\hat{F}^{-1}(\alpha/2)$ representa el percentil de $\alpha/2$ de la distribución del estimador $\hat{\theta}^*$, es decir un valor tal que la probabilidad acumulada hasta dicho valor sea $\alpha/2$, y $\hat{F}^{-1}(1 - \alpha/2)$ representa el percentil del $1 - \alpha/2$. El método del percentil conserva la esencia no-paramétrica del enfoque bootstrap y libera al usuario de las asunciones de la estadística paramétrica.

CAPÍTULO II

2. MARCO METODOLÓGICO

2.1. Tipo de Investigación

El tipo de investigación empleado para este estudio según:

- La naturaleza de los datos es cuantitativa, en vista que se trabaja con el número acumulado de casos confirmados y fallecidos de COVID-19.
- El objetivo es de tipo aplicada, dado que la investigación se centra en abordar el problema originado en el campo de la salud, principalmente en el sistema sanitario, a través de la aplicación de modelos de series de tiempo.
- El nivel de profundización en el objeto de estudio es explicativo, ya que se busca determinar los métodos de predicción a partir de casos confirmados y fallecidos de COVID-19.
- El tipo de inferencia inductiva, puesto que se pretende determinar la metodología de predicción serial que presente características suficientemente buenas que representen los datos históricos.
- El periodo temporal longitudinal, ya que se aprecia la evolución epidemiológica con el seguimiento a los casos confirmados y fallecidos que se reporta diariamente por el Ministerio de Salud, el cual permite estudiar la tendencia y cambios producidos en el tiempo, y así también describe las predicciones de la enfermedad COVID-19 (Hernández Sampieri et al., 2014) (Patten and Newhart, 2018).

2.2. Diseño de la investigación No Experimental

El estudio es no experimental, puesto que en la investigación se puede observar los fenómenos tal cual como se dan en su ambiente natural, no se permite la manipulación de variables ni se puede influir sobre ellas dado que ya sucedieron, al igual que sus efectos, es decir, se observa situaciones ya existentes más no provocadas por el investigador. Debido a que la recolección de los datos se efectuó en un periodo específico del tiempo (Hernández Sampieri et al., 2014).

2.2.1. Localización del estudio

El presente estudio se realizó en todo el país, Ecuador se localiza en la costa noroccidental de América del Sur, abarca la selva del Amazonas, las zonas altas andinas y las islas Galápagos;

limita por el sur y por el este con Perú, con Colombia por el norte, y al oeste con el océano pacífico.



Figura 1-2: Mapa Ecuador

Fuente: https://satellites.pro/mapa_de_Ecuador#-2.026555,-78.365479,7

2.2.2. Población de estudio

El estudio se realizó a partir de la proyección referencial poblacional de 17'510.643 habitantes para el año 2020 según el INEC y la Secretaría Nacional de Planificación y Desarrollo, en base al censo de Población y Vivienda aplicado en el 2010.

2.2.3. Tamaño de la muestra

La muestra se representó por el número acumulado de casos confirmados y fallecidos de COVID-19 de las 24 provincias del Ecuador, desde el 13 de marzo hasta el 31 de octubre del 2020 registrado en infografías proporcionados por el Ministerio de Salud Pública del Ecuador, según el último registro se tienen una muestra de 168192 confirmados y 8357 muertes por COVID-19 detectados con la prueba PCR (Reacción en Cadena de la Polimerasa).

2.2.4. Método de muestreo

Se aplica el muestreo no probabilístico por conveniencia, el cual permite que el investigador tome en consideración la muestra de análisis a partir de los datos existentes de casos confirmados y fallecidos por COVID-19 a nivel nacional.

2.2.5. Técnicas de recolección de datos

La información se obtiene del Ministerio de Salud Pública del Ecuador en su página web de datos abiertos los cuales se encuentran registrados en infografías (<https://www.gestionderiesgos.gob.ec/informes-de-situacion-covid-19-desde-el-13-de-marzo-del-2020/>), su tabulación se desarrolla en una hoja de cálculo de Microsoft Excel.

2.2.6. Identificación de variables

Para el presente estudio se propone trabajar con las variables epidemiológicas de casos confirmados y fallecidos de COVID-19, debido a la importancia y necesidad de conocer y determinar el comportamiento de dichas variables.

2.2.7. Modelo Estadístico

Modelación de series temporales por el método ARIMA y Bootstrap, mediante el software estadístico R.

2.3. Variables en estudio

2.3.1. Operacionalización de Variables

Tabla 1-2: Operacionalización de Variables

VARIABLE	DESCRIPCIÓN	TIPO	ESCALA	UNIDAD DE MEDIDA
Tiempo	Variable independiente, es una magnitud de carácter físico, periodo de que una cosa pasa de un estado a otro.	Cuantitativa discreta	Razón	Días
Confirmados	Variable dependiente, representa en número la magnitud de casos confirmados de COVID-19.	Cuantitativa discreta	Razón	Número de casos confirmados
Fallecidos	Variable dependiente, representa en número la magnitud de individuos fallecidos de COVID-19.	Cuantitativa discreta	Razón	Número de individuos fallecidos

Fuente: Ministerio de Salud Pública (Informes de Situación e Infografías – COVID 19 – desde el 29 de febrero del 2020)

Realizado por: Cabrera, Maritza, 2021

CAPÍTULO III

3. RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS

La información obtenida para este estudio fue extraída de infografías presentadas diariamente por el Ministerio de Salud Pública del Ecuador, en el cual se reportó el número de casos confirmados y fallecidos por COVID-19 a partir del 13 de marzo hasta el 31 de octubre del 2020, los datos fueron analizados en el software estadístico R. Se trabajó en el análisis de la serie de casos confirmados y la serie de casos fallecidos a causa de COVID-19, a través del modelado de curvas empleando la metodología ARIMA.

3.1. Análisis exploratorio

Hasta el 31 de octubre de 2020 en Ecuador se registró un acumulado de 168192 casos confirmados y 12670 casos de personas fallecidas por COVID-19, esta enfermedad causada por el virus SARS-CoV-2 fue reportada por el Ministerio de Salud el 13 de marzo con 23 casos confirmados y un individuo fallecido. La información ha sido registrada de forma diaria especialmente de los casos probados con la prueba PCR, los días 25 y 26 de abril y el 5 de mayo no reportó información el Ministerio de Salud debido a que existieron pruebas de COVID-19 que se encontraban en proceso, por la cual esta información no estaba registrada convirtiéndose en NA (datos perdidos o no respondidos), al tratarse de valores acumulados se asignó el número de casos presentes de la fecha anterior para que no se pierda la continuidad, no se puede hacer una imputación ni con media, mediana o moda debido a que transcurre en el tiempo y la información es acumulativa.

La matriz de datos contiene la fecha, casos confirmados y fallecidos de COVID-19, cabe señalar que en algunas fechas los incrementos y decrementos no se han mantenido de forma creciente puesto que se han producido cambios considerables en el acumulados según los reportes del 3 al 12 de abril con un acumulado diario de 20937, 22981, 22981, 20483, 21361, 19881, 20134, 20622, 20572 y 21481, el cual indica que estos días no ha mantenido en una tendencia creciente.

Del total de casos confirmados por COVID-19 se estableció que el 4.97% han fallecido, y el 95.03% se encuentran aún con la enfermedad o se han recuperado, pero fueron confirmados con COVID-19, por lo que se los mantiene en nuestros datos por requerimiento de este estudio, ya que se está trabajando netamente con el acumulado de estos casos.

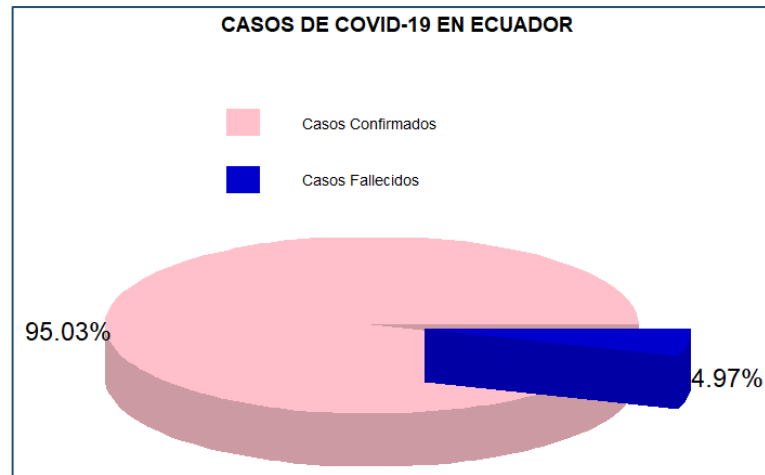


Gráfico 1-3: Porcentaje de casos confirmados y fallecidos

Realizado por: Cabrera, Maritza, 2021

En la tabla 1-3 muestra que en promedio se presentaron 722 casos confirmados y 36 muertes diarias a causa del COVID-19, alcanzando un máximo de 3821 y 410 casos confirmados y fallecidos respectivamente en un día, las dos variables siguen una distribución asimétrica positivas de tipo leptocúrtica alargándose a valores mayores que la media, claramente indica que no siguen una distribución normal.

Tabla 1-3: Estadística descriptiva de las variables de casos confirmados y fallecidos por COVID-19 presentados a diario

Variables	1° Qu	Mediana	Media	3° Qu	Desviación Estándar	Asimetría	Curtosis	Max
Casos Confirmados	301	660	722	976	546.22	1.32	3.73	3821
Casos Fallecidos	14	28	36	45	44.45	4.75	29.71	410

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

En la Gráfico 2-3 se presenta la evolución de la serie de tiempo de casos confirmados de COVID-19 a partir del 13 de marzo hasta el 31 de octubre, se puede evidenciar que no hay presencia de estacionalidad dado que no existe patrones repetidos a lo largo de todas las series, además se observó que los valores individuales no se desplazan alrededor de 65717 casos confirmados en promedio y muestra una tendencia creciente, el número de casos confirmados por el virus ha tenido un incremento diario, a partir de estos hallazgos se puede determinar que la serie temporal no presenta estacionariedad.

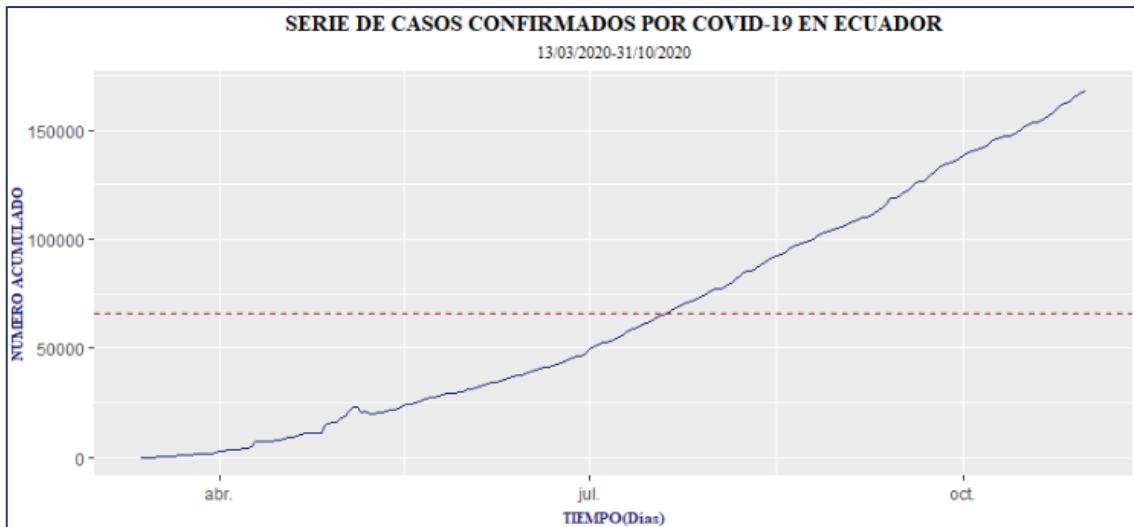


Gráfico 2-3: Evolución serial de casos confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

El Gráfico 3-3 muestra la evolución de la serie de tiempo de casos de individuos fallecidos por COVID-19, en el cual se logró observar que la serie mantiene una tendencia creciente, los valores individuales no fluctúan alrededor de la media que está representada por 4389 muertes y no existe patrones repetidos a lo largo de toda la serie, por tanto, se consideró que la serie temporal no es estacionaria.

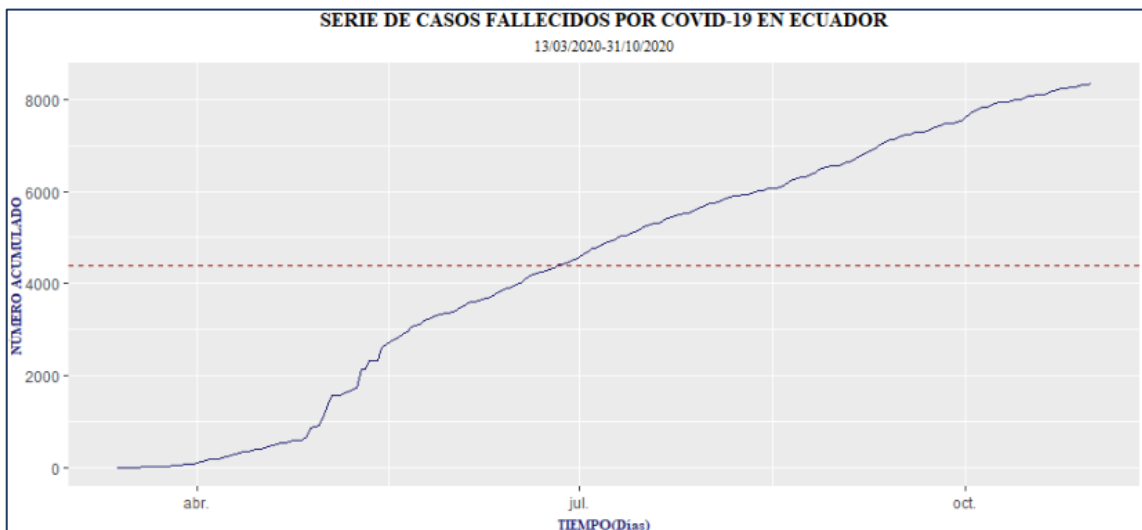


Gráfico 3-3: Evolución serial de casos fallecidos por COVID-19

Realizado por: Cabrera, Maritza, 2021

3.2. Modelación ARIMA

3.2.1. Correlogramas

Mediante en la función ACF de autocorrelación y PACF autocorrelación parcial expuesta en el Gráfico 4-3, un comportamiento de Autocorrelación decreciente, tienden a disminuir lentamente en el tiempo de forma aritmética en las correlaciones estimadas, lo que indica que no existe estacionariedad en la serie de tiempo de casos confirmados por COVID-19 en Ecuador; además se observó que los datos son dependientes positivos debido a que están por encima de la recta, por tanto, los rezagos son significativos, al igual que el primer rezago de la función de autocorrelación parcial (PACF), el resto de rezagos se encuentra dentro del intervalo de confianza y se aproximan a cero.

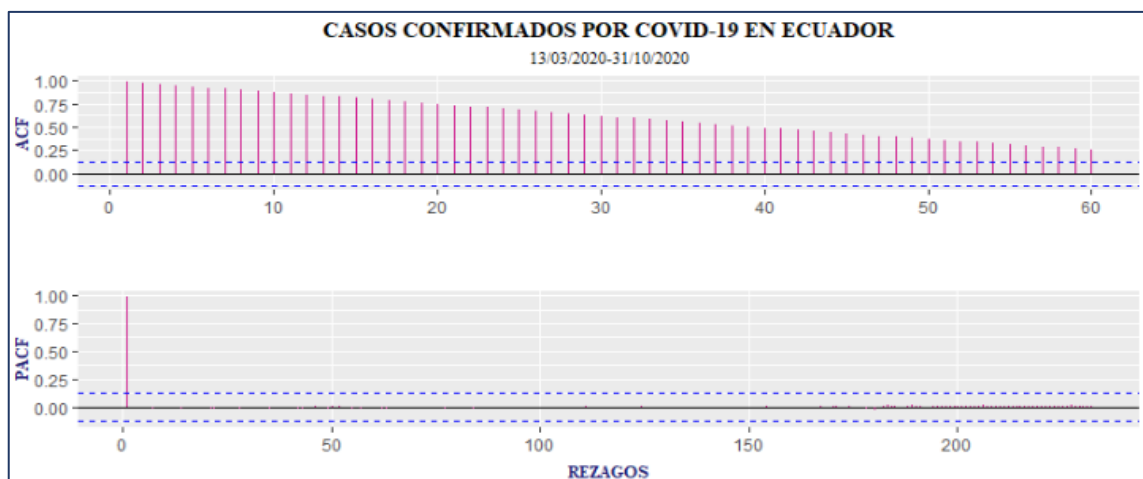


Gráfico 4-3: Correlogramas ACF y PACF de la serie casos confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

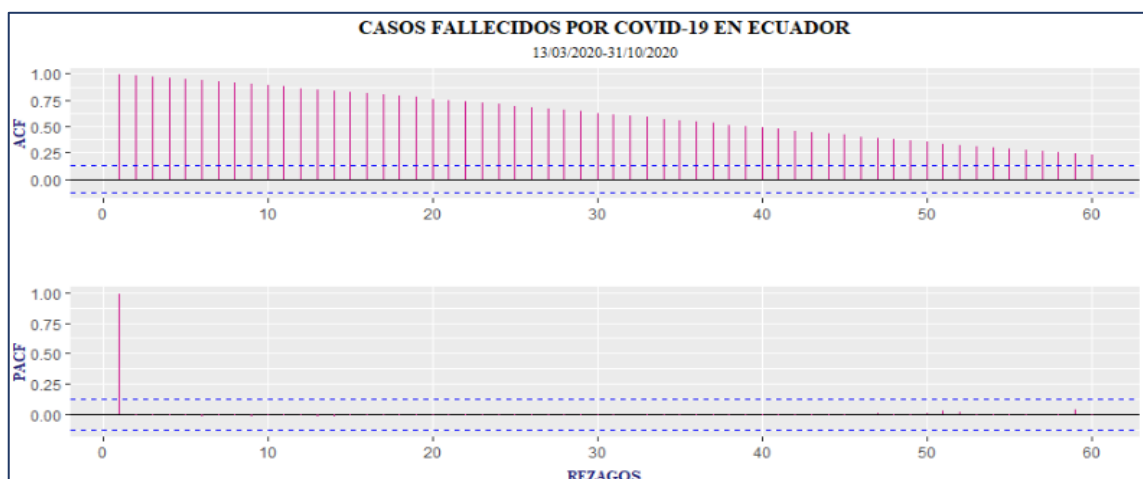


Gráfico 5-3: Correlogramas ACF y PACF de la serie casos fallecidos por COVID-19

Realizado por: Cabrera, Maritza, 2021

Se determina que no existe estacionariedad en la serie de casos fallecidos por COVID-19 pero si una clara tendencia, según la función de autocorrelación del Gráfico 5-3 se produce un lento decrecimiento aritmético en las autocorrelaciones estimadas los primeros retrasos son significativamente diferentes de cero, a medida que va decreciendo de forma gradual se acerca al intervalo de confianza. Corroborando que efectivamente los datos presentan tendencia, pero no estacionariedad, mientras las correlaciones sean cercanas a 1 da indicios para sospechar que existe raíz unitaria, por lo que se debe probar mediante una prueba de hipótesis para determinar la estacionalidad. Además, la función de autocorrelación parcial (PACF) muestra que los retardos se aproximan a cero a excepción del primer rezago, es decir, significativamente diferente de cero, debido a que se encuentra fuera de las bandas que indican el intervalo de confianza.

Las dos series presentaron tendencia, problema de raíz unitaria por el comportamiento de la serie, y no estacionariedad, según la autocorrelación y correlación parcial existe autocorrelación tanto en media móvil como en el componente autorregresivo. Se dice que una serie presenta al menos una raíz unitaria cuando la serie no es estacionaria en media o no es integrada de orden cero $I(0)$, la raíz unitaria es la causante de problemas en la inferencia estadística en los modelos de series de tiempo, debido a que la correlación es espuria, es decir, la apariencia existente en la relación de causalidad entre las variables cuando esta no existe en realidad; y esto sesga la significancia de las variables, en series de tiempo hace creer que hay un buen ajuste y relación entre las variables cuando en realidad no la hay, indicando así que no existe estacionariedad, para corregir el problema de autocorrelación se debe aplicar la prueba de Dickey Fuller aumentada debido a que le estamos agregando los rezagos de la variable dependiente.

3.2.2. Test de Dickey Fuller aumentada

La presencia de estacionariedad en las series de tiempo fue validada por la prueba de raíz unitaria de Dickey Fuller aumentada, planteando las siguientes hipótesis:

H_0 : Los casos confirmados diarios y_t tienen raíz unitaria, la serie no estacionaria

H_1 : Los casos confirmados diarios y_t no tienen raíz unitaria, la serie es estacionaria

H_0 : Los casos de fallecidos diarios y_t tienen raíz unitaria, la serie no estacionaria

H_1 : Los casos de fallecidos diarios y_t no tienen raíz unitaria, la serie es estacionaria

La Tabla 2-3 muestra el test de Dickey Fuller aumentada con un orden de retraso de 6 y un valor de p de 0.74 y 0.93, indica que no existe evidencia suficiente para decir que las series de tiempo

de casos confirmados y casos fallecidos por COVID-19 sean estacionarias a un 5% de significancia, es decir, las series no presentan estacionariedad y por ende tiene raíz unitaria.

Tabla 2-3: Test de Dickey Fuller (Estacionariedad)

Casos de COVID-19 en Ecuador

Variable	Dickey-Fuller	Valor p
Casos Confirmados	-1,6069	0,7411
Casos Fallecidos	-1,0377	0,9307

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

Se da por demostrado que las series de tiempo no son estacionarias, cabe recalcar que para aplicar el modelo ARIMA estas series deben ser estacionarias, por ello se debió encontrar la estacionariedad de las series, para lo cual se realizó la diferenciación, esto nos ayudó a estabilizar la media de las series temporales y así reducir o eliminar la tendencia y la estacionalidad.

3.2.3. Diferenciación

Se obtuvo el número de diferenciaciones que se necesita para llevar a cabo la estacionariedad de las series temporales, a través de los comandos **ndiffs** que estima el número de primeras diferencias regulares necesarias y **nsdiffs** se encarga de estimar el número de diferencias estacionales. Los respectivos cálculos muestran que la serie de casos confirmados por COVID-19 necesita dos diferenciaciones regulares y ninguna diferenciación estacional; por otro lado, la serie de tiempo que representa los casos de individuos fallecidos por COVID-19 es necesario una diferenciación regular y ninguna diferenciación estacional.

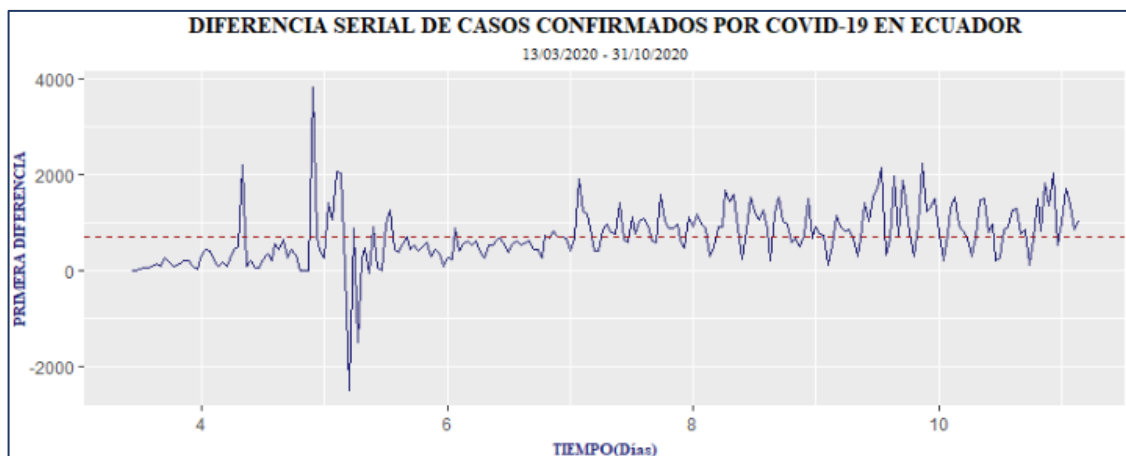


Gráfico 6-3: Evolución de la primera diferencia serial de Casos Confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

El Gráfico 6-3 presenta la serie temporal de casos confirmados por COVID-19 denotada como y_t y la primera diferencia $\nabla y_t = y_t - y_{t-1}$, para $t = 1, 2, \dots, 233$, la misma que no muestra signos de tendencia, pero si la presencia de picos altos lo que indica que la varianza no es constante además de que no existe patrones de estacionalidad según conteos mensuales teniendo como referencia 30 días hábiles; se logra observar rasgos de fluctuación alrededor de la media para concretar esta información se realizó una segunda diferencia.

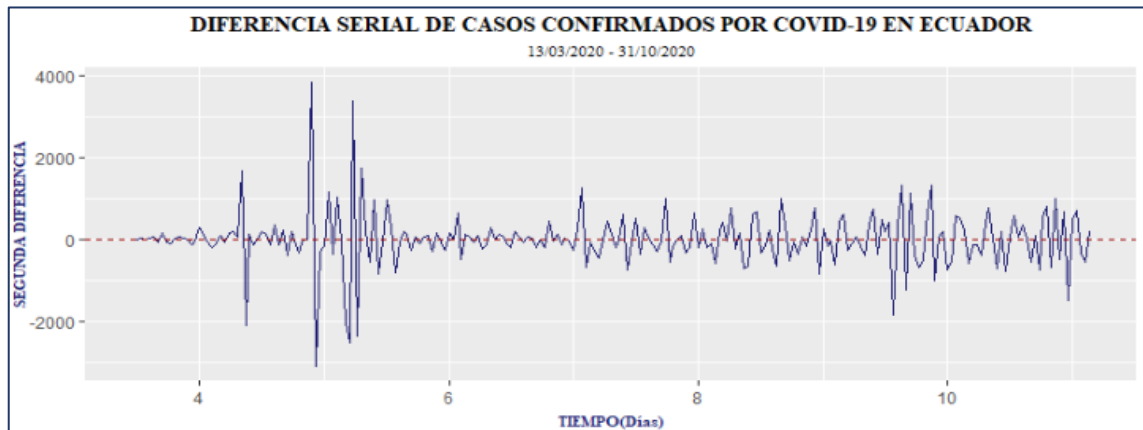


Gráfico 7-3: Evolución de la segunda diferencia serial de Casos Confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

En el Grafico 7-3 ya se pudo evidenciar con una segunda diferencia que la media es constante los datos se mueven alrededor de esta, aunque existe picos altos antes de llegar al mes de mayo hasta a mediados del mes pero para el resto se mantiene considerando la existencia de varianza constante lo que indica estacionariedad, además si tomamos como referencia a los meses de 30 días no se logra visualizar una variación periódica para decir que existe estacionalidad y mucho menos la presencia de tendencia.

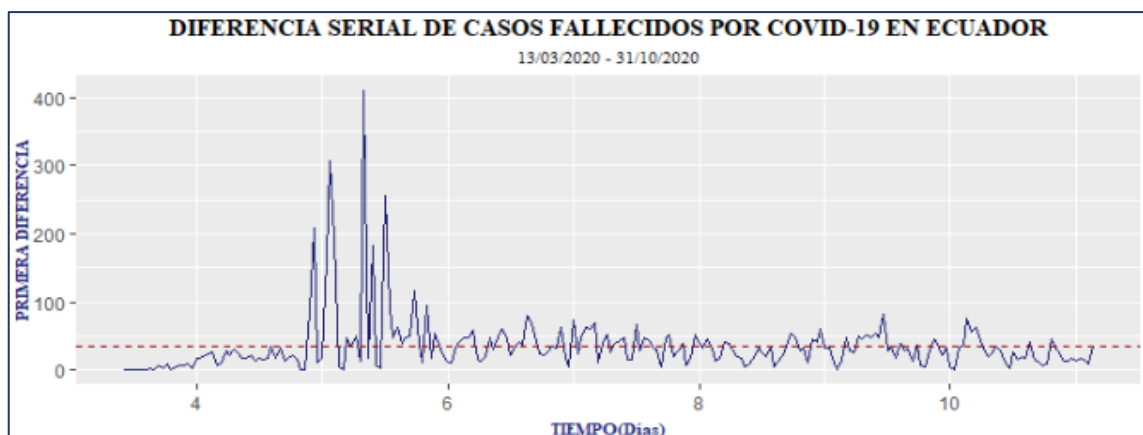


Gráfico 8-3: Evolución de la primera diferencia serial de Casos Fallecidos por COVID-19

Realizado por: Cabrera, Maritza, 2021

El Gráfico 8-3 muestra la evolución de la serie de tiempo de casos de personas fallecidas por COVID-19, luego de realizar la primera diferenciación, se determinó que la serie no muestra signos de tendencia, pero si la presencia de picos altos no estacionales, los datos fluctúan alrededor de la media por lo que se considera a la serie estacionaria en media.

Se verificó la estacionariedad con la prueba de Dickey Fuller aumentada, en la cual se obtuvo un valor de p de 0.01 para la serie de casos confirmados por COVID-19, mientras que para la serie de casos fallecidos con la diferencia devolvió un valor de p de 0.02, existe suficiente evidencia para aseverar que las series de tiempo de casos confirmados y fallecidos por COVID-19 en Ecuador son estacionarios con 2 y 1 diferencia respectivamente.

3.2.4. Identificación de los modelos

Las series y_t del número acumulado de casos confirmados por COVID-19 en Ecuador, requirió de 2 diferencias regulares para que sea estacionaria, se planteó un modelo ARIMA no estacional de parámetros $(p, 2, q)$, en el Gráfico 9-3 muestra desfases en el primer rezago de la autocorrelación simple y en la autocorrelación parcial inicia con 6 desfases en los primeros rezagos, cabe indicar que en la autocorrelación simple se observa que los rezagos 1, 7, 9, 10, 16, 26 y 35 son significativos y en el resto tienden a anularse, lo que significa que están incorrelados a partir de que se encuentren dentro del intervalo de confianza, por otro lado, en el gráfico de Autocorrelación Parcial en total se observa 12 rezagos significativos los primeros 6, 8,9,14,16,20 y 34 al igual que la autocorrelación simple estos también tienden a cero.

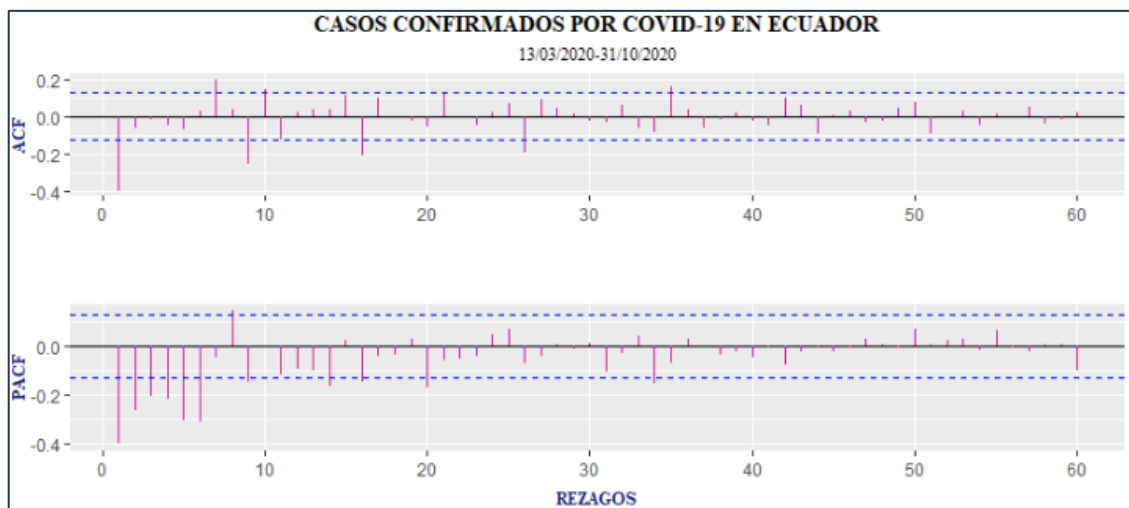


Gráfico 9-3: Correlogramas ACF y PACF de casos Confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

Es importante tener en claro que al realizar el análisis de Autocorrelación Simple determinamos el número de medias móviles y con la Autocorrelación Parcial el número de autorregresivos; gráficamente se establece un $MA(1)$ debido a que el primer retraso es significativo y las autocorrelaciones parciales presentan un decrecimiento fluctuando a cero mientras avanza el tiempo, por lo que tomando en cuenta el número de diferencias el modelo en una primera instancia se lo definió como $ARIMA(0,2,1)$, junto con las siguientes variaciones que incluyen un $ARIMA(0,2,2)$, $ARIMA(1,2,0)$, $ARIMA(2,2,0)$ y $ARIMA(1,2,1)$.

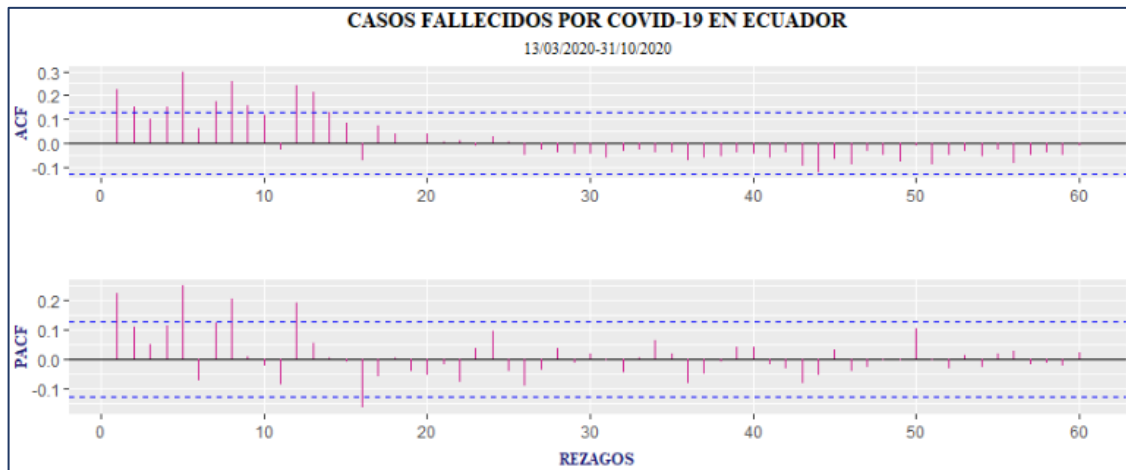


Gráfico 10-3: Correlogramas ACF y PACF de casos fallecidos por COVID-19

Realizado por: Cabrera, Maritza, 2021

En cuanto al análisis de la serie de tiempo del número acumulado de casos fallecidos por COVID-19 en Ecuador, en el Gráfico 10-3 se pudo visualizar las autocorrelaciones simples y parciales las mismas que tienen desfase en el primer rezago, en la gráfica de autocorrelación los rezagos significativos son 1, 2, 4, 5, 7, 8, 9, 12 y 13 el resto de los rezagos decaen a cero, mientras los rezagos significativos de la autocorrelación parcial son 1, 5, 8, 12 y 16 el resto de rezagos tiende a cero, es probable definir un modelo $MA(1)$ y $AR(1)$ debido a que presenta características de media móvil y autorregresivas, defendiendo un modelo $ARIMA(1,1,1)$.

3.2.5. Estimación y selección del modelo

Se analiza el Criterio de Información de Akaike AIC, AICc (corregido) y el criterio de información bayesiano (BIC) de Schwarz son parsimoniosos debido a que explica mucho con pocas variables exógenas o independiente, está indicado por el valor mínimo que representa el mejor ajuste.

Tabla 3-3: Criterios de información y evaluación para los posibles modelos ARIMA de la serie de Casos confirmados por COVID-19

MODELO	Criterios de información			Criterios de Evaluación		
	AIC	AICc	BIC	RMSE	MAE	MASE
ARIMA(0,2,1)	3571,45	3571,50	3578,33	541,33	328,69	0,433
ARIMA(0,2,2)	3567,83	3567,93	3578,15	534,83	321,09	0,423
ARIMA(1,2,0)	3639,04	3639,09	3645,92	629,19	391,75	0,516
ARIMA(2,2,0)	3624,33	3624,44	3634,66	606,65	375,78	0,495
ARIMA(1,2,1)	3567,99	3568,09	3578,32	535,03	322,39	0,424

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

En la Tabla 3-3 se puede apreciar que los mejores ajustes se centran en aquellos que tienen componente de medias móviles, el resultado sugiere un modelo ARIMA(0,2,2); los criterios en esta condición garantizan que es la mejor opción en la reproducción de la serie de datos para la variable estimada, por otro lado se presentó en este modelo un error cuadrático medio (RMSE), error absoluto medio (MAE) y error escalado absoluto medio (MASE) mínimos en comparación al resto de modelos, este modelo permitirá reducir en mejores predicciones.

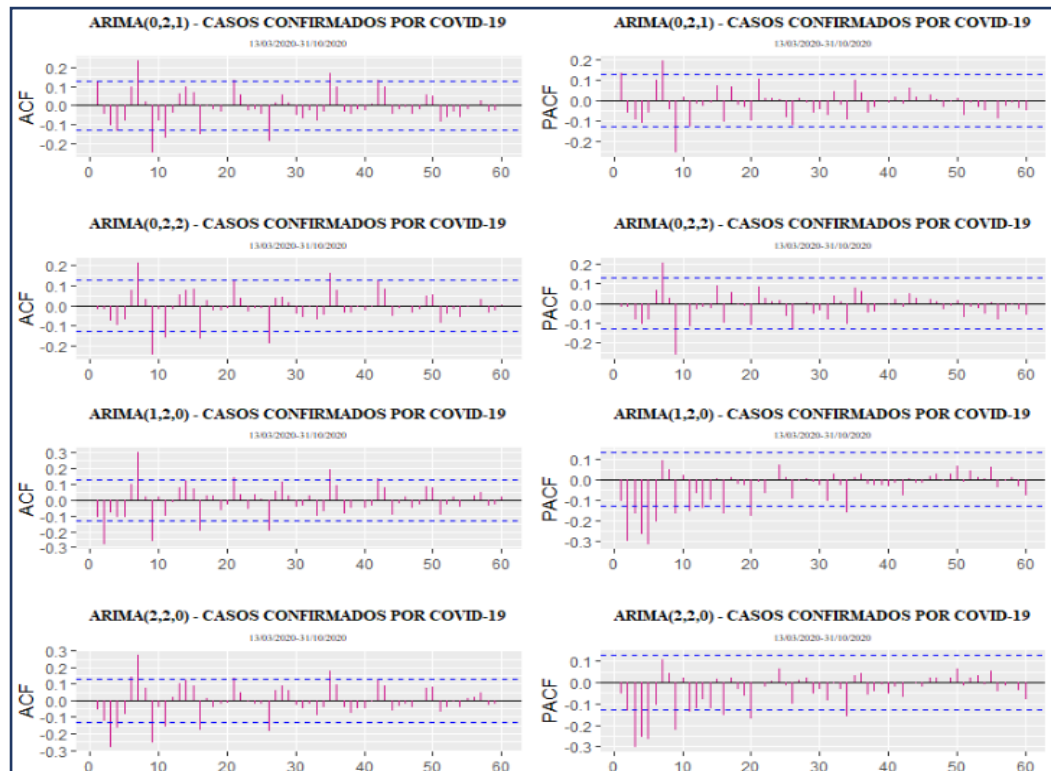


Gráfico 11-3: Correlogramas de los modelos ARIMA propuestos para la serie de casos Confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

Se analizaron los correlogramas de los modelos propuestos de la serie de casos confirmados por COVID-19 en el Gráfico 11-3, se logró identificar el número mínimo de rezagos significativos en los modelos ARIMA(0,2,2) y ARIMA(1,2,1), y la aproximación a cero se da de forma acelerada

Antes de proceder a la validación se generó un modelo mediante el comando `auto.arima()` el cual proporciona automáticamente el mejor modelo, en este ARIMA(0,2,2) es un modelo de suavizado exponencial lineal general, para la diferenciación de segundo orden se calculó a partir de $\Delta^2 X_t = X_t - 2X_{t-1} + X_{t-2}$ es igual a una función lineal de los dos últimos errores del pronóstico:

$$\begin{aligned}\Delta^2 X_t &= -\theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t \\ X_t - 2X_{t-1} + X_{t-2} &= -\theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t \\ X_t &= 2X_{t-1} - X_{t-2} - \theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t\end{aligned}$$

Cabe señalar que a es un proceso de ruido blanco independiente de X_t y los parámetros θ_1 y θ_2 son los coeficientes de MA(1) y MA(2) respectivamente, y a partir del cual se obtiene el modelo ARIMA(0,2,2) para el número acumulado diario de casos confirmados por COVID-19 en Ecuador.

$$X_t = 2X_{t-1} - X_{t-2} + 0.7868 e_{t-1} + 0.1527 e_{t-2} + e_t$$

En lo que respecta a la serie del número acumulado de casos fallecidos por COVID-19 en Ecuador, gráficamente se identificó que existen características autorregresivas y de media móvil presentándose desfase en los primeros rezagos, por tanto, se podría definir un modelo ARIMA(1,1,1), se probó también con los modelos ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(1,1,2) y ARIMA(0,1,2), para quedar claros que modelo aplicar.

El resultado de la Tabla 4-3 propone un modelo ARIMA(1,1,1) a través de los criterios de información y de evaluación se observa que el modelo sin tomar en consideración el modelo ARIMA(1,1,1) with drif (con deriva) presenta un AIC, AICc y BIC bajos por tanto se lo considera adecuado para pronosticar fallecimientos por COVID-19 y aproximarse a los datos reales, verificado también en el error cuadrático medio (RMSE), el error absoluto medio (MAE) y el error escalado absoluto medio (MASE) son bajos corroborando lo que indica el criterio de información.

Al generar un modelo automático con la función `auto.arima()` se obtiene un ARIMA(1,1,1) with drift (con deriva), lo que permite dar direccionalidad a la serie, esto ocurre en modelos con paseo aleatorio.

Tabla 4-3: Criterios de información y evaluación para los posibles modelos ARIMA de la serie de Casos Fallecidos por COVID-19

MODELO	Criterios de información			Criterios de Evaluación		
	AIC	AICc	BIC	RMSE	MAE	MASE
ARIMA(0,1,1)	2493,28	2493,33	2500,17	51,59	21,53	0,81
ARIMA(0,1,2)	2469,35	2469,46	2479,69	48,77	21,38	0,72
ARIMA(1,1,0)	2462,89	2462,95	2469,79	48,31	23,24	0,67
ARIMA(2,1,0)	2444,73	2444,84	2455,07	46,23	22,70	0,63
ARIMA(1,1,1)	2401,66	2401,76	2412,00	42,06	21,38	0,58
ARIMA(1,1,1) with drift	2399,89	2400,07	2413,68	41,81	21,32	0,54

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

La deriva o constante se incluye si la serie tiene una tendencia promedio distinta de cero como lo muestra el Gráfico 3-3; partiendo de que la serie de casos fallecidos por COVID-19 tiene tendencia positiva con media distinta de cero y se aplicó un orden de diferenciación se puede considerar la inclusión de una constante, cabe mencionar que la constante se incluye en modelos que no tienen orden de diferenciación y modelos que tienen un orden de diferenciación, normalmente en modelos de dos órdenes de diferenciación no incluye un término constante. Al incluir la constante en el modelo ARIMA(1,1,1) se observa que se minimizan considerablemente los criterios de información y evaluación, a partir de este análisis se consideró que el modelo adecuado para esta serie de tiempo es ARIMA(1,1,1). La diferencia no estacional de primer orden se calculó a partir de $\Delta X_t = X_t - X_{t-1}$, el modelo seleccionado contiene características autorregresivas y de medias móviles es mixto y se define como:

$$\begin{aligned}\Delta X_t &= \mu + \varphi_1 \Delta X_{t-1} + e_t + \theta_1 e_{t-1} \\ X_t - X_{t-1} &= \mu + \varphi_1 (X_{t-1} - X_{t-2}) + e_t + \theta_1 e_{t-1} \\ X_t &= \mu + X_{t-1} + \varphi_1 X_{t-1} - \varphi_1 X_{t-2} + \theta_1 e_{t-1} + e_t\end{aligned}$$

Teniendo en cuenta que μ es la constante del modelo, se presenta el modelo ARIMA(1,1,1) with drift (con deriva) para el número acumulado diario de casos fallecidos por COVID-19 en Ecuador:

$$X_t = 33.4104 + X_{t-1} + 0.9505X_{t-1} - 0.9505X_{t-2} - 0.8390 e_{t-1} + e_t$$

Como se puede observar en el Gráfico 12-3 las autocorrelaciones simples y parciales que presentan menores rezagos significativos pertenecen a los modelos ARIMA(1,1,1) y ARIMA(1,1,1) with drift (con deriva), la única diferencia entre estos dos modelos es la inclusión de una constante, por lo que se toma en cuenta debido a que presenta mejores características según los criterios de información y evaluación.

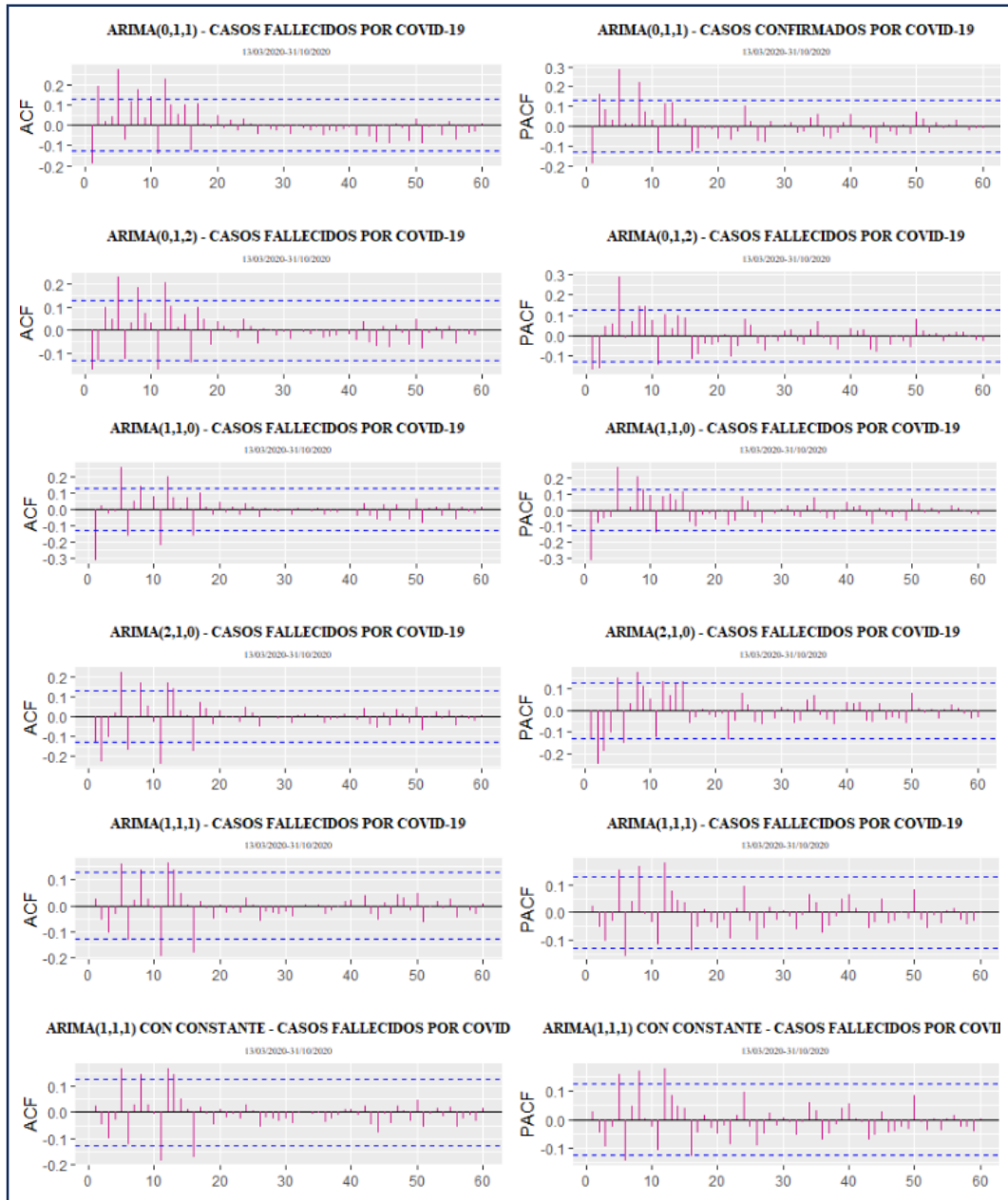


Gráfico 12-3: Correlogramas de los modelos ARIMA propuestos para la serie de casos fallecidos por COVID-19

Realizado por: Cabrera, Maritza, 2021

3.2.6. Validación del modelo

Es necesario contrastar si los parámetros del modelo para determinar si son adecuados y mantenerlos en el modelo caso contrario si alguno de los parámetros no fuera significativo debe ser separado del modelo, el estadístico de contraste está basado en la distribución t de Student y probando la siguiente hipótesis nula $\text{Coeficiente} = 0$ se dice que no significativo.

Tabla 5-3: Coeficientes e intervalos de confianza de los modelos seleccionados

SERIE	MODELO	Coeficientes	Estimación	Error estándar	Valor p	Li	Ls
Casos	ARIMA(0,2,2)	MA(1)	-0,787	0,064	2,00E-16	-0,913	-0,661
Confirmados		MA(2)	-0,153	0,063	0,0154	-0,276	-0,029
Casos	ARIMA(1,1,1) with drift	AR(1)	0,951	0,032	2,20E-16	0,887	1,014
Fallecidos		MA(1)	-0,839	0,051	2,20E-16	-0,939	-0,739
		μ	33,410	8,669	0,0001	16,420	50,401

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

En la Tabla 5-3 es posible observar que los coeficientes tanto para el modelo ARIMA(0,2,2) donde MA(1) y MA(2) son estadísticamente significativos a un 99% y 95% de confianza al igual que el modelo ARIMA(1,1,1) with drift (con deriva) donde los coeficientes AR(1), MA(1) y μ al 99% de confianza son significativos, se puede considerar que los modelos son adecuados, (-0.913; -0.661) y (-0.276; -0.029) son los intervalos de confianza del modelo ARIMA(0,2,2) y los intervalos que corresponde al modelo ARIMA(1,1,1) with drift (con deriva) cuenta con los intervalos de confianza (0.887; 1.014), (-0.939; -0.739) y (16.420; 50.401).

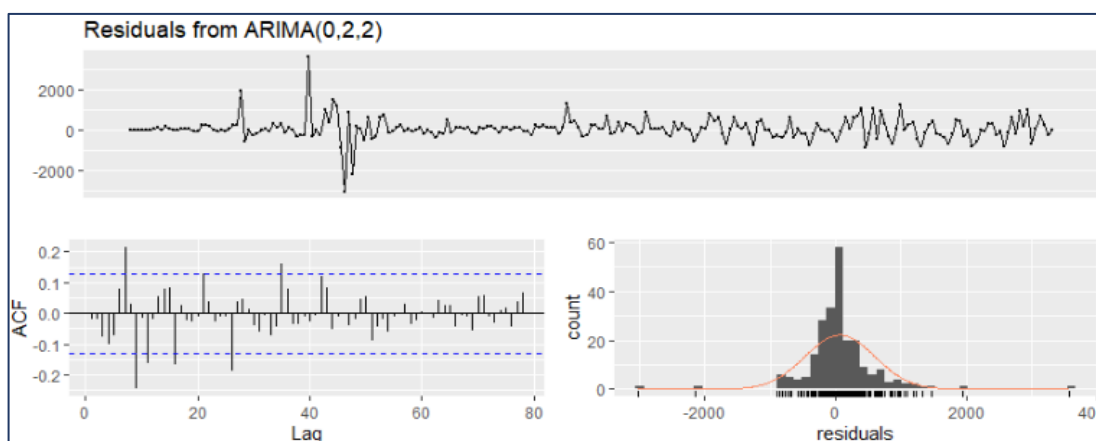


Gráfico 13-3: Validación del modelo ARIMA(0,2,2) de casos confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

En los Gráficos 13-3 se observa que los residuos del modelo ARIMA(0,2,2) son de media 0 y en líneas generales no son demasiado inestables, por tanto, tiene varianza constante y se determina que son estacionarios; además se observa en la Autocorrelación que existen algunos rezagos significativos ya que están fuera de la banda de confianza, esto da indicios de que el modelo no es adecuado para predicciones, existen los p valores de los rezagos mayores que el nivel de significancia por lo que no se puede afirmar si son ruido blanco o no, necesario un análisis más a profundidad.

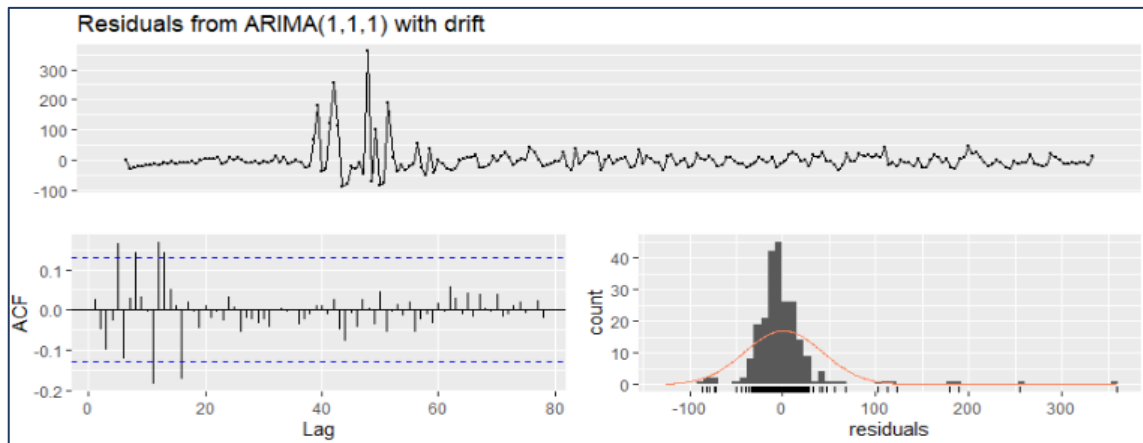


Gráfico 14-3: Validación del modelo ARIMA(1,1,1) with drift de casos fallecidos por COVID-19

Realizado por: Cabrera, Maritza, 2021

La Grafica 14-3 muestra la validación del modelo ARIMA (1,1,1) with drift (con deriva) correspondiente al número acumulado de casos fallecidos por COVID-19, en el cual se observa posible presencia de estacionariedad en los residuos, pero al parecer carece de ruido blanco debido a que existen rezagos significativos.

Tabla 6-3: Comprobación de supuestos de los modelos de casos confirmados y fallecidos por COVID-19

MODELO	p - valor de los Test		
	Estacionariedad	Independencia	Normalidad
ARIMA(0,2,2)	0,01	0.777	1,49E-12
ARIMA(1,1,1) with drift	0,01	0.694	2,20E-16

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

Al realizar el análisis de los residuos que muestra la tabla 6-3 se verificó que los modelos ARIMA de las series de casos confirmados y casos fallecidos por COVID-19 con la prueba de Dickey

Fuller aumentada cumplen con el supuesto de estacionariedad con un p valor de 0.01; el test de Jarque y Bera permitió determinar normalidad en los residuos, por lo que al rechazar la hipótesis nula los residuos de los dos modelos no siguen una distribución normal, finalmente se identificó la presencia de independencia estadística los residuos de los modelos, esto se verificó con la prueba de Ljung Box al contrastar la autocorrelación en los residuales ε_k para cada rezago, por tanto se considera series de ruido blanco con un valor p menor que el 5% del nivel de significancia.

Cabe señalar que se analizó las autocorrelaciones ya que ayudan a identificar la existencia de ruido blanco para validar el modelo, si no existe coeficientes de correlación significativos se afirma que los residuos son independientes y por tanto son Ruido Blanco.

3.2.7. Pronósticos

Los pronósticos tanto para el modelo ARIMA(0,2,2) y ARIMA(1,1,1) with drift (con deriva) se realizaron para los 2 meses siguientes noviembre y diciembre, que corresponde a 61 días, se presentó los valores reales, los pronósticos y los intervalos de confianza de acuerdo con el modelo para la serie de casos confirmados como la serie de casos fallecidos por COVID-19.

Tabla 7-3: Pronóstico del modelo para las series de casos confirmados y fallecidos por COVID-19

Fecha	Casos Confirmados				Casos Fallecidos			
	Reales	Pronóstico	LI	LS	Reales	Pronóstico	LI	LS
01/11/2020	169194	169257	168200	170315	8371	8381	8299	8464
02/11/2020	169562	170318	168656	171981	8380	8406	8282	8529
03/11/2020	170110	171380	169240	173519	8386	8431	8272	8590
04/11/2020	171433	172441	169878	175004	8394	8456	8263	8650
05/11/2020	171783	173502	170545	176459	8420	8482	8257	8708
06/11/2020	172508	174563	171230	177896	8449	8509	8251	8766
07/11/2020	173486	175624	171926	179323	8492	8535	8246	8825
08/11/2020	174907	176685	172629	180742	8507	8562	8241	8883
09/11/2020	175269	177747	173336	182157	8516	8589	8237	8941
10/11/2020	175711	178808	174046	183570	8525	8617	8234	9000
11/11/2020	176630	179869	174757	184981	8592	8645	8231	9059
12/11/2020	177513	180930	175469	186391	8614	8673	8228	9118
13/11/2020	178674	181991	176180	187803	8642	8701	8226	9177
14/11/2020	179627	183053	176890	189215	8658	8730	8225	9236

15/11/2020	180295	184114	177598	190629	8669	8759	8223	9295
16/11/2020	180676	185175	178305	192045	8674	8788	8222	9354
17/11/2020	181104	186236	179009	193463	8681	8817	8222	9413
18/11/2020	182250	187297	179711	194883	8708	8847	8221	9472
19/11/2020	183246	188358	180411	196306	8729	8877	8222	9531
20/11/2020	183840	189420	181107	197732	8750	8906	8222	9591
21/11/2020	184876	190481	181801	199160	8787	8937	8223	9650
22/11/2020	185643	191542	182492	200592	8804	8967	8225	9709
23/11/2020	185944	192603	183180	202027	8825	8997	8227	9768
24/11/2020	186436	193664	183864	203464	8861	9028	8229	9827
25/11/2020	187230	194726	184546	204905	8877	9058	8231	9886
26/11/2020	188138	195787	185224	206349	8891	9089	8234	9944
27/11/2020	189534	196848	185899	207797	8927	9120	8237	10003
28/11/2020	190909	197909	186571	209247	8938	9151	8241	10061
29/11/2020	192117	198970	187239	210701	8989	9182	8245	10120
30/11/2020	192685	200031	187904	212159	9024	9214	8249	10178
01/12/2020	193673	201093	188566	213620	9052	9245	8254	10236
02/12/2020	194876	202154	189224	215084	9103	9277	8259	10294
03/12/2020	195884	203215	189879	216551	9129	9308	8265	10352
04/12/2020	196482	204276	190531	218022	9200	9340	8270	10409
05/12/2020	197391	205337	191179	219496	9252	9372	8276	10467
06/12/2020	197998	206398	191824	220973	9268	9403	8283	10524
07/12/2020	198244	207460	192465	222454	9269	9435	8290	10581
08/12/2020	198752	208521	193104	223938	9278	9467	8297	10638
09/12/2020	199228	209582	193739	225425	9295	9499	8304	10695
10/12/2020	200379	210643	194370	226916	9320	9532	8312	10752
11/12/2020	200765	211704	194999	228410	9323	9564	8319	10808
12/12/2020	201524	212766	195624	229908	9336	9596	8328	10864
13/12/2020	202110	213827	196245	231408	9337	9628	8336	10921
14/12/2020	202180	214888	196864	232912	9337	9661	8345	10977
15/12/2020	202356	215949	197479	234419	9357	9693	8354	11032
16/12/2020	203461	217010	198091	235930	9373	9726	8363	11088
17/12/2020	204249	218071	198700	237443	9386	9758	8373	11143
18/12/2020	205003	219133	199305	238960	9396	9791	8383	11199
19/12/2020	205920	220194	199908	240480	9400	9823	8393	11254
20/12/2020	206257	221255	200507	242003	9400	9856	8403	11309

21/12/2020	206329	222316	201103	243530	9400	9889	8414	11364
22/12/2020	206364	223377	201696	245059	9400	9921	8425	11418
23/12/2020	207084	224439	202285	246592	9413	9954	8436	11473
24/12/2020	208010	225500	202872	248128	9427	9987	8447	11527
25/12/2020	208828	226561	203455	249666	9433	10020	8458	11581
26/12/2020	209274	227622	204036	251208	9439	10053	8470	11635
27/12/2020	209355	228683	204613	252753	9439	10085	8482	11689
28/12/2020	209758	229744	205187	254302	9441	10118	8494	11742
29/12/2020	210326	230806	205759	255853	9447	10151	8507	11796
30/12/2020	211512	231867	206327	257407	9469	10184	8519	11849
31/12/2020	212512	232928	206892	258964	9473	10217	8532	11902

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

Para determinar la precisión de los pronósticos se analiza las medidas de adecuación, la tabla 8-3 muestra un error porcentual promedio absoluto del pronóstico respecto a los datos reales (MAPE) es de 4.51% para la serie de casos confirmados y 2.72% para casos fallecidos, con un error cuadrático medio (RMSE) de 10721.54 y 321.36, y una desviación absoluta media de 9003.79 para casos confirmados y 251.45 para casos fallecidos aplicando la metodología ARIMA, comparando los dos modelos el que presenta mejor pronóstico es el ARIMA(1,1,1) with drift (con deriva) correspondiente a los casos fallecidos.

Tabla 8-3: Medidas de adecuación de los pronósticos metodología ARIMA

SERIE	MODELO	MAD	RMSE	MAPE
Casos Confirmados	ARIMA(0,2,2)	9003,79	10721,54	4,51%
Casos Fallecidos	ARIMA(1,1,1) with drift	251,45	321,36	2,72%

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

Se observa los pronósticos y los valores reales de los meses de noviembre y diciembre mediante el modelo ARIMA(0,2,2) presentada en el gráfico 15-3 tiene una tendencia creciente, los valores pronosticados de casos confirmados por COVID-19 al inicio del mes de noviembre se aproximaron a la curva Regist. Nov-Dic de datos reales de los meses pronosticados y en el transcurso del tiempo se ha incrementado, a final del año la aproximación del pronóstico no

explico correctamente a los datos reales existiendo una diferencia de 20416 casos, esto se corroboró en la Tabla 7-3.

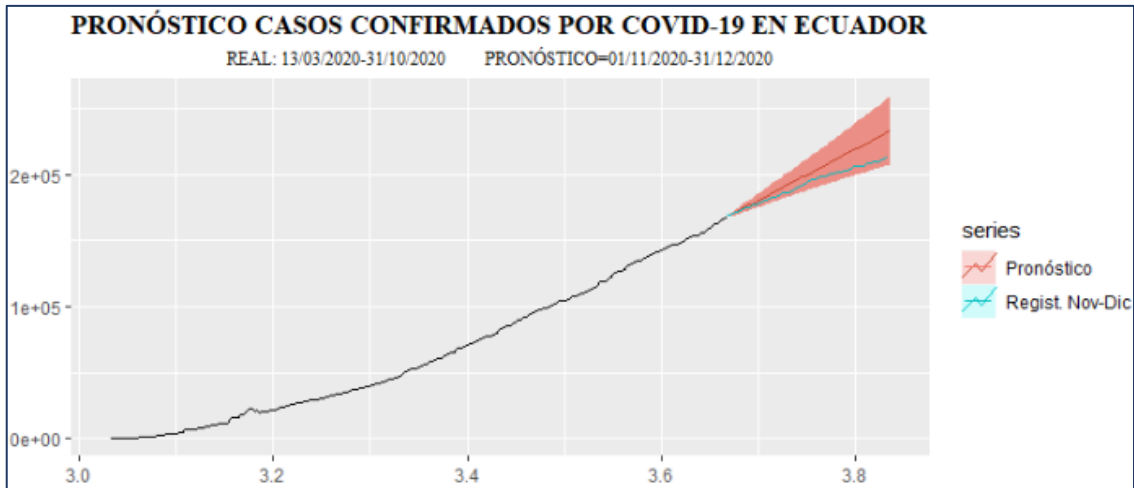


Gráfico 15-3: Pronóstico ARIMA(0,2,2) para los meses de noviembre y diciembre de casos confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

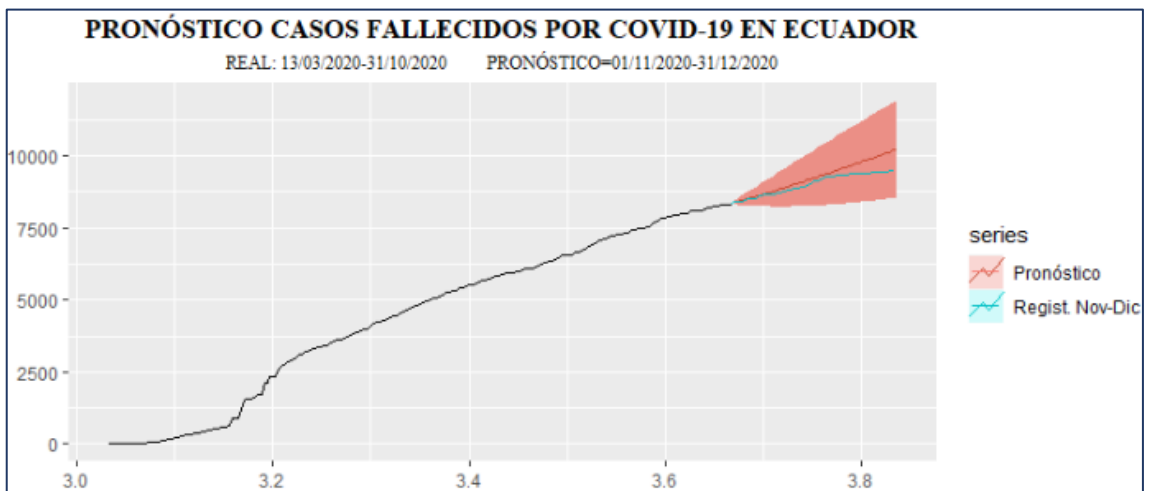


Gráfico 16-3: Pronóstico ARIMA(1,1,1) with drift para los meses de noviembre y diciembre de casos fallecidos por COVID-19

Realizado por: Cabrera, Maritza, 2021

En el Gráfico 16-3 se puede observar los pronósticos y la información real de los casos fallecidos por COVID-19, se determina en la curva de Regist. Nov-Dic en comparación con la curva de Pronóstico de casos confirmados, mostrando una buena aproximación, los pronósticos siguen una tendencia positiva al igual que los datos originales a mediados del mes de diciembre los pronósticos empiezan a alejarse de los originales, para el 31 de diciembre se determinó 9473

muerdes confirmadas y con el pronóstico del modelo $ARIMA(1,1,1)$ with drift se obtuvo 10217 muerdes con una diferencia de 744 fallecidos.

3.3. Modelación Bootstrap

Debido a que los residuos de las series de tiempo lineales presentan autocorrelación es necesario aplicar Bootstrap, al muestrear los datos originales se produce una distribución única estadística, la más utilizada es la media, para construirla se debe calcular en cada réplica de un conjunto de replicas. Para mejorar las predicciones es muy útil agregar un Bootstrap (ensacado) evitando el sobreajuste en el modelo de la serie de tiempo y obtener estabilidad.

Se aplicó un **bootstrapping por bloques** para series de tiempo, esta consiste en volver a muestrear en bloques consecutivamente valores de los datos originales y alinearlos a una muestra Bootstrap, empleando los comandos de R que permiten realizar un remuestreo por bloques, `bld.mbb.bootstrap()` es el generador de versiones o repeticiones bootstrap de una serie de tiempo basado en Box-Cox y Loess, este comando acepta la serie de tiempo original, número de versiones del arranque, y el tamaño del bloque; el comando `tsboot()` implementa métodos de remuestreo para series de tiempo, si el bootstrap es estacionario de debe generar un bloque de tipo geométrico, `tsboot` incluye la serie de tiempo original, el método estadístico que se emplea en la serie de tiempo en este caso el modelo ARIMA, el número de réplicas de arranque bootstrapping y el tipo de simulación y la longitud promedio del bloque.

Cabe indicar que `tsboot()` en realidad no realiza remuestreo, se debe crear una función que devuelva la serie de tiempo de arranque, `tsboot` entrega como salida las estadísticas Bootstrap. Para la modelación de las curvas seriales se utilizó una semilla (2021), réplicas de 100, 16 bloques y 1000 simulaciones teniendo en cuenta el método estadístico de predicción ARIMA para la serie de tiempo de casos confirmados y casos fallecidos por COVID-19.

3.3.1. Análisis del modelo

La Tabla 9-3 muestra el análisis del booststrap en bloques, con una longitud de bloque fijo de 16, y 100 réplicas de arranque, tanto para la serie de casos confirmados y casos fallecidos los intervalos de confianza se generaron a un 95%, para la serie de casos confirmados fueron examinados los coeficientes MA(1) y MA(2) obteniendo una media estimada de -1.0148 y 0.00148 las mismas que se encuentran en los intervalos de confianza (-1.0167; -0.9819) y (-0.0104; 0.0214) respectivamente, y se presenta el sesgo entre los coeficientes del modelo original y la media de los coeficientes generados de las 100 réplicas, coeficiente MA(2) muestra un

coeficiente menor de 0.17; mientras que para los casos fallecidos su medias son 0.0831, -0.1040 y -0.1096 los mismos que se encuentra dentro del intervalo de confianza percentil (-0.030; 0.917), (-1.00; 0.002) y (-7.660; 6.620), obteniendo sesgos -0.87, 0.74 y -33.52 para los coeficientes de este modelo.

Tabla 9-3: Estadística Bootstrap por bloques de la serie de casos confirmados y fallecidos por COVID-19

Serie	Estadísticas Bootstrap por Bloques						
	Coefficientes	Sesgo	Error Estándar	Media	LI	LS	
Casos confirmados	MA(1)	-0,7868	-0,2280	0,0233	-1,0148	-1,068	-0,998
	MA(2)	-0,1527	0,1675	0,0233	0,0148	-0,003	0,068
	AR(1)	0,9505	-0,8675	0,2687	0,0831	-0,030	0,917
Casos fallecidos	MA(1)	-0,8390	0,7349	0,2898	-0,1040	-1,000	0,002
	μ	33,410	-33,5201	3,4568	-0,1096	-7,660	6,620

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

Los intervalos de confianza de la modelación Bootstrap brindan mejor presión en los pronósticos debido que son más angostos, se determina también que el sesgo no es nulo por tanto los coeficientes no son insesgados o centrados.

3.3.2. Pronóstico

Tabla 10-3: Pronóstico Bootstrap en base a los modelos ARIMA para las series de casos confirmados y fallecidos por COVID-19

Fecha	Bootstrap - Casos Confirmados				Bootstrap - Casos Fallecidos			
	Reales	Pronóstico	LI	LS	Reales	Pronóstico	LI	LS
01/11/2020	169194	169402	168208	170380	8371	8372	8287	8458
02/11/2020	169562	170536	169199	171990	8380	8398	8298	8483
03/11/2020	170110	171712	169967	173593	8386	8423	8289	8572
04/11/2020	171433	172896	170668	175122	8394	8451	8305	8630
05/11/2020	171783	173976	170943	176415	8420	8473	8262	8692
06/11/2020	172508	175114	171649	178712	8449	8500	8236	8789
07/11/2020	173486	176247	173151	180340	8492	8524	8249	8845
08/11/2020	174907	177391	173688	181725	8507	8552	8266	8955
09/11/2020	175269	178554	174655	183023	8516	8579	8240	8975

10/11/2020	175711	179635	175476	184075	8525	8602	8238	9023
11/11/2020	176630	180780	176882	185317	8592	8632	8240	9064
12/11/2020	177513	181960	177718	187516	8614	8661	8242	9180
13/11/2020	178674	183054	178312	189275	8642	8685	8211	9273
14/11/2020	179627	184277	178853	191021	8658	8720	8190	9381
15/11/2020	180295	185386	179420	192183	8669	8746	8161	9454
16/11/2020	180676	186566	180211	193595	8674	8776	8140	9517
17/11/2020	181104	187707	181395	195379	8681	8805	8117	9623
18/11/2020	182250	188844	182606	196831	8708	8834	8110	9695
19/11/2020	183246	189971	183143	197943	8729	8863	8116	9720
20/11/2020	183840	191092	184351	199959	8750	8889	8134	9794
21/11/2020	184876	192126	184723	201239	8787	8909	8204	9833
22/11/2020	185643	193233	185890	203158	8804	8938	8132	9915
23/11/2020	185944	194286	186569	204314	8825	8960	8173	9975
24/11/2020	186436	195338	186408	205649	8861	8981	8158	10018
25/11/2020	187230	196376	187394	206794	8877	9004	8146	10047
26/11/2020	188138	197434	187133	208335	8891	9024	8138	10110
27/11/2020	189534	198481	188243	209655	8927	9044	8131	10152
28/11/2020	190909	199637	188758	211497	8938	9074	8117	10183
29/11/2020	192117	200785	190275	213135	8989	9099	8076	10252
30/11/2020	192685	201841	190137	214231	9024	9121	8023	10295
01/12/2020	193673	202953	191251	215577	9052	9146	8008	10349
02/12/2020	194876	204009	191055	217150	9103	9167	7994	10389
03/12/2020	195884	205163	191310	218732	9129	9196	7992	10473
04/12/2020	196482	206319	191463	219532	9200	9224	7972	10541
05/12/2020	197391	207462	192565	220968	9252	9250	7941	10589
06/12/2020	197998	208535	193365	222769	9268	9274	7898	10623
07/12/2020	198244	209665	194815	224537	9269	9300	7845	10683
08/12/2020	198752	210721	195963	226231	9278	9320	7774	10748
09/12/2020	199228	211839	196709	227962	9295	9347	7714	10792
10/12/2020	200379	212981	197113	229273	9320	9373	7626	10899
11/12/2020	200765	214140	197899	230359	9323	9400	7569	10968
12/12/2020	201524	215231	199021	232130	9336	9424	7528	11015
13/12/2020	202110	216337	199480	233569	9337	9449	7418	11081
14/12/2020	202180	217426	199627	235770	9337	9473	7361	11160
15/12/2020	202356	218483	200239	237256	9357	9494	7276	11238

16/12/2020	203461	219655	201705	239183	9373	9524	7158	11314
17/12/2020	204249	220756	203241	240624	9386	9549	7088	11438
18/12/2020	205003	221827	203420	241603	9396	9572	6999	11570
19/12/2020	205920	222901	203026	243263	9400	9595	6929	11639
20/12/2020	206257	224076	203804	245106	9400	9624	6788	11714
21/12/2020	206329	225165	204399	246665	9400	9647	6696	11727
22/12/2020	206364	226350	204937	249088	9400	9678	6612	11725
23/12/2020	207084	227550	205019	250194	9413	9707	6497	11800
24/12/2020	208010	228678	205341	250914	9427	9734	6400	11861
25/12/2020	208828	229856	205292	253406	9433	9766	6276	11896
26/12/2020	209274	230888	205706	253852	9439	9785	6177	11991
27/12/2020	209355	232025	206657	255028	9439	9813	6139	12087
28/12/2020	209758	233095	207385	256316	9441	9834	6123	12173
29/12/2020	210326	234264	208564	258290	9447	9865	6121	12227
30/12/2020	211512	235405	209826	259789	9469	9891	6146	12239
31/12/2020	212512	236506	210999	262391	9473	9914	6192	12306

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

Los pronósticos presentados en la Tabla 10-3 tienen tendencia creciente, la serie de tiempo de casos fallecidos con el método Bootstrap tomando el modelo de casos fallecidos ARIMA(1,1,1) with drift (con deriva) como punto de partida, presenta predicciones más aproximadas a los datos reales que los pronósticos Bootstrap de los casos confirmados, se proyecta hasta el 31 de diciembre un total de 9914 muertes y 236506 casos confirmados por COVID-19; a partir de esto se puede decir que para los casos confirmados los pronósticos por Bootstrap están más alejados de los datos reales, mientras que para los casos fallecidos con el modelo Bootstrap los valores pronosticados más se aproximan a la realidad.

Tabla 11-3: Medidas de adecuación de los pronósticos metodología Bootstrap

SERIE	MODELO	MAD	RMSE	MAPE
Casos Confirmados	Bootstrap ARIMA(0,2,2)	10942,36	12838,47	5,49%
Casos Fallecidos	Bootstrap ARIMA(1,1,1) with drift	136,10	177,34	1,48%

Fuente: Software estadístico R

Realizado por: Cabrera, Maritza, 2021

La tabla 11-3 muestra las medidas de adecuación los cuales permitirá determinar la precisión de los pronósticos al aplicar la metodología Bootstrap, el 5.49% y 1.48% representa el error porcentual promedio absoluto (MAPE) de la serie de casos confirmados y casos fallecidos por COVID-19 correspondientemente, cuentan con un error cuadrático medio (RMSE) de 12838.47 y 177.34, y una desviación absoluta media de 10942.36 y 136.10 respectivamente para cada serie.

A partir del análisis de las medidas de adecuación del pronóstico, se logró identificar un error absoluto porcentual medio de 4.511 y 5.493 para la serie de casos confirmados con el modelo ARIMA y metodología Bootstrap correspondientemente, definiendo que se produce pronósticos más acertados para la serie de tiempo de casos confirmados por COVID-19 con el modelo ARIMA(0,2,2); mientras ocurre todo lo contrario para la serie de casos fallecidos, con errores absolutos porcentuales promedio de 2.721 y 1.481 perteneciente al método ARIMA y Bootstrap respectivamente, se determina que el menor error promedio absoluto le corresponde a los pronósticos que se aproxima de mejor manera a los datos observables, siendo así, el modelo preciso para pronosticar el número de casos confirmados para los meses de noviembre y diciembre es ARIMA(0,2,2) y para casos fallecidos el modelo Bootstrap.

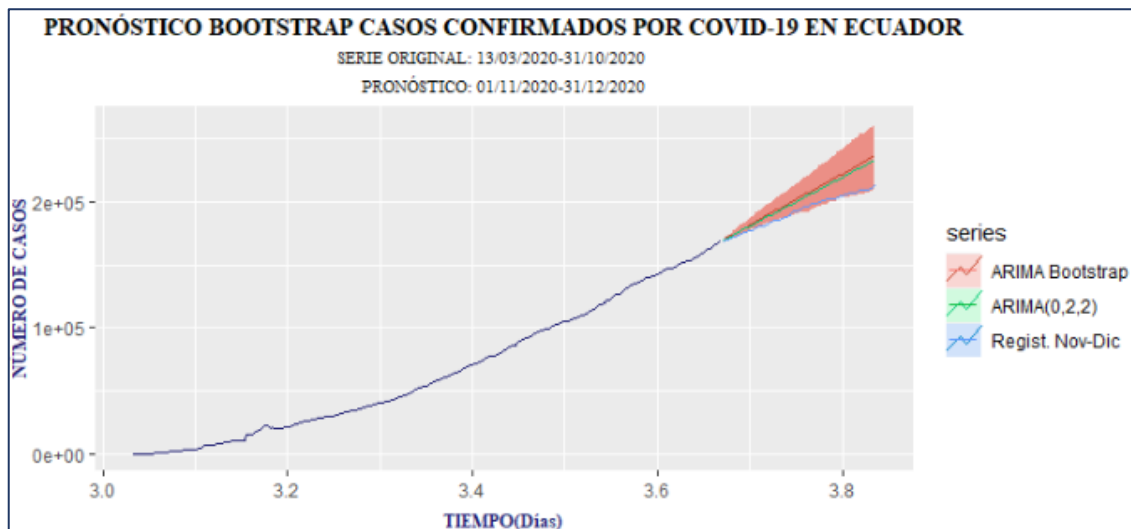


Gráfico 17-3: Pronóstico Bootstrap ARIMA(0,2,2) para los meses de noviembre y diciembre de casos confirmados por COVID-19

Realizado por: Cabrera, Maritza, 2021

Se puede observar en la Gráfico 17-3 que la curva del modelo bootstrap se encuentra más alejado de la curva Regist. Nov-Dic correspondientes a los registros diarios de los meses de noviembre y diciembre mientras que la curva del modelo ARIMA(0,2,2) se acopla de mejor manera a curva la Regist. Nov-Dic, cabe señalar que intervalo de confianza si cubre a los valores reales a un 95%, pero no se considera adecuado para pronosticar los casos confirmados por COVID-19.

Claramente se observa el Grafico 18-3 que no existe gran diferencia entre las curvas de los modelos ARIMA y bootstrap para la serie de casos fallecidos, determinando que la curva más precisa es de obtenida mediante el metodolo bootstrap debido a que la curva de su pronóstico esta más aproximada a la curva Regist. Nov-Dic y los intervalos de confianza a un 95% son más estrechos, por lo tanto brinda mejores pronósticos que el modelo ARIMA(1,1,1) with drift (con deriva).

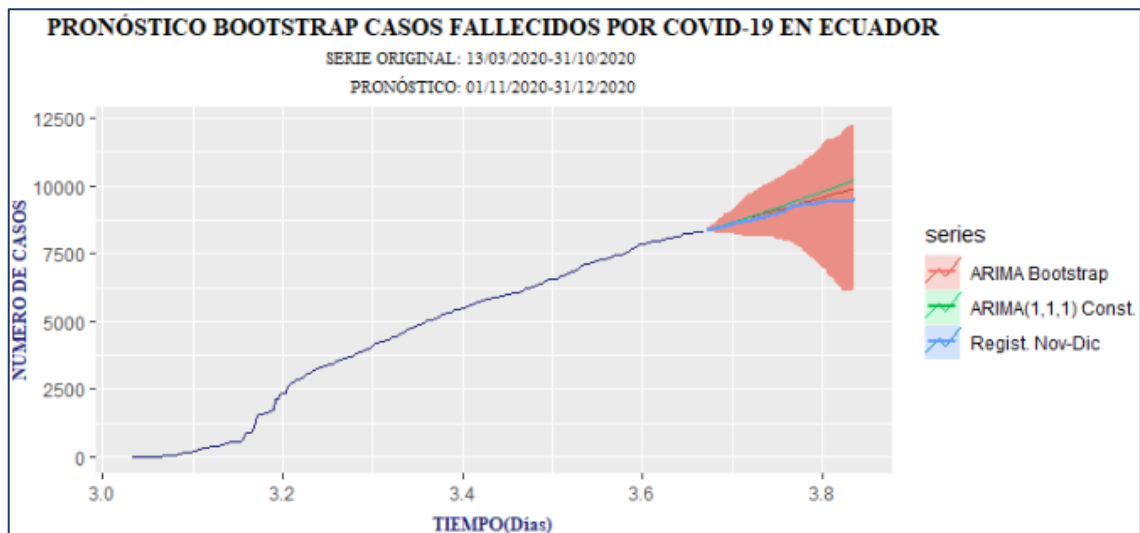


Gráfico 18-3: Pronóstico Bootstrap ARIMA(1,1,1) with drift para los meses de noviembre y diciembre de casos fallecidos por COVID-19

Realizado por: Cabrera, Maritza, 2021

Se logró identificar que los pronósticos más adecuados y precisos que mejor explican la curva Regist. Nov-Dic correspondiente a los datos reales de casos presentados en los meses de noviembre y diciembre, para las dos series de tiempo no están representados por el mismo modelo, el método ARIMA con un $p = 0$, $d = 2$ y $q = 2$ para la serie de casos confirmados por COVID-19 es más apropiado para ejecutar predicciones entre el 01 de noviembre y 31 de diciembre de 2020, y para la serie de casos fallecidos con la metodología Bootstrap presenta pronósticos más precisos que el modelo ARIMA con parámetros $p = 1$, $d = 1$, $q = 1$ y la inclusión de una constante.

CONCLUSIONES

En este estudio, se modela las curvas de las series temporales de casos confirmados y fallecidos por COVID-19 en Ecuador, con la aplicación de la modelación ARIMA y Bootstrap, tomando la información diaria acumulada proporcionada por el Ministerio de Salud Pública a partir del 13 de marzo hasta el 31 de octubre de 2020, específicamente los casos comprobados con la prueba PCR, revelando las siguientes conclusiones:

- Las series de tiempo correspondientes a los casos confirmados y a los casos fallecidos presentan tendencia creciente no estacionaria con respecto a la media y no se exteriorizan patrones estacionariedad, esto se comprueba con el test Dickey Fuller aumentada la cual muestra que es necesario realizar diferencias de segundo orden para casos confirmados y primer orden para casos fallecidos, ambas de forma regular; con el análisis de los criterios de información y adecuación de varios modelos, se determinó que el modelo con mayor ajuste que describe el comportamiento diario de casos confirmados es ARIMA(0,2,2) con un AIC de 3567.83 y un error cuadrático medio de 534.83, y para la serie de casos fallecidos es ARIMA(1,1,1) with drift (con deriva) presentando un AIC de 2399.89 y 41.81 de error cuadrático medio, se obtuvo los pronósticos de los meses de noviembre y diciembre con el modelo ARIMA para los casos confirmados el cual esta errado en un 4.51%, y para los casos fallecidos sus pronósticos están errando en 2.72%, estos modelos ARIMA seleccionados a pesar de presentar coeficientes significativos no son del todo confiables debido a que no cumplen con los supuestos de normalidad.
- Mediante la aplicación de modelos Bootstrap en 16 bloques, 100 réplicas de arranque, 1000 simulaciones y la semilla 2021 generados a partir de los modelos ARIMA con Bootstrap de casos confirmados y casos fallecidos, se determina que al aplicar Bootstrap al modelo ARIMA(0,2,2) en el pronóstico se obtiene un error porcentual absoluto medio de 5.49%, una desviación media absoluta de 10942.36 y un error cuadrático medio de 12838.47, para el Bootstrap del modelo ARIMA(1,1,1) with drift para la serie de casos fallecidos por COVID-19 la desviación absoluta media es de 136.10, el error cuadrático medio de 177.34 y un 1.48% de errar en el pronóstico.
- Se determina que el modelo ARIMA(0,2,2) explica de mejor manera la evolución de casos confirmados por COVID-19 teniendo en cuenta el 4.51% de errar en el pronóstico frente a un 5.49% de error porcentual absoluto medio con el método Bootstrap, mientras que para los casos fallecidos por COVID-19 el método Bootstrap es el más apropiado para pronosticar

debido a que presenta un error de 1.48% por debajo del 2.72% presentado en el modelo ARIMA(1,1,1) with drift, siendo estos modelos eficientes al obtener menor error de pronóstico. En consecuencia, mediante el modelo ARIMA se obtiene mejores pronósticos para la serie de casos confirmados, y para los casos fallecidos por COVID-19 el método Bootstrap es más adecuado para pronosticar estos casos para los meses de noviembre y diciembre.

Los hallazgos encontrados en esta investigación guardan relación con el estudio realizado por Tanzila Saba, Ibrahim Abunadi, Mirza Naveed y Amjad Rehman (2021), en el que sustenta que las técnicas de modelación óptimas para predicción dependen de los criterios de rendimiento MAPE, MAE y RMSE, el modelo ARIMA devuelve pronósticos precisos de casos confirmados y muertes por COVID-19 en los países Bulgaria, Grecia, Rusia, Irán, India y Holanda, y solamente en casos confirmados de Islandia y Suecia mientras que las muertes se ven mejor pronosticados por modelos de Machine Learning (aprendizaje automático), estos modelos fueron adecuados para pronosticar los casos en Hubei (China), cabe indicar que los árboles de decisión incluyó el ensacado o agregación Bootstrap. Por otro lado, el trabajo antes mencionado no emplea la modelación Bootstrap de series de tiempo, pero permite identificar que de ningún modo se puede recomendar un enfoque único para modelar y pronosticar, debido a las características del territorio en estudio haciendo que la evolución de los casos confirmados y muertes de COVID-19 difieran, por lo que las técnicas de modelación óptimas en muchos de los casos no son las mismas para cada serie, sin embargo, permiten obtener pronósticos acertados.

RECOMENDACIONES

- Aplicar la metodología Bootstrap se recomienda generar muestras grandes y una semilla, además considerar diferentes tamaños de réplicas y bloques para evaluar los hallazgos, esto permitirá no perder la información, mantener estático el resultado generado y delimitar resultados apropiados.
- En posteriores análisis se debe tener presente el cumplimiento de todos los supuestos en la aplicación de modelos ARIMA para evaluar la incidencia que tiene sobre la modelación Bootstrap, debido a que juegan un papel sustancial dentro de la modelación de series temporales para determinar el modelo que permitirá la obtención de pronósticos precisos.
- Analizar las series de casos confirmados y fallecidos por COVID-19 en el Ecuador mediante la aplicación de otras técnicas como el modelo de Holt Winter y trasladar al estudio en modelos de árboles de decisión, redes neuronales artificiales, modelo logarítmico Sigmoidal y modelo Suspected-Infected-Recovered (Sospechoso-Infectado-Recuperado SIR) aptos para trabajar con el número acumulado de casos, que fusionados con la metodología Bootstrap se puede obtener hallazgos interesantes para contrastar el desempeño de los modelos.

BIBLIOGRAFÍA

- ACOSTA, L.D.**, 2020. Capacidad de respuesta frente a la pandemia de COVID-19 en América Latina y el Caribe. *Revista Panamericana de Salud Pública* [en línea], vol. 44, pp. 1. [Consulta: 16 noviembre 2020]. ISSN 1020-4989, 1680-5348. DOI 10.26633/RPSP.2020.109. Disponible en: <https://iris.paho.org/handle/10665.2/52654>.
<https://iris.paho.org/bitstream/handle/10665.2/52654/v44e1092020.pdf?sequence=1&isAllowed=y>
- AKAIKE, H.**, 1976. An information criterion (AIC). *Mathematical Sciences* 14, pp. 5-9.
- ASTERIOU, D. y HALL, S.G.**, 2011. ARIMA models and the Box–Jenkins methodology. *Applied Econometrics*, vol. 2, no. 2, pp. 265-286.
- AYALA JACINTO, A.R. y SAMILLÁN VÁSQUEZ, C.J.**, 2019. Metodología de Box Jenkins vs Redes Neuronales Artificiales para construir un modelo de pronóstico del precio de compra de cierre mensual de las acciones del Banco de Crédito del Perú en la Bolsa de Valores de Lima, abril de 2005 - febrero de 2018. En: Accepted: 2019-10-02T21:01:10Z, *Universidad Nacional Pedro Ruiz Gallo* [en línea], [Consulta: 11 junio 2020]. Disponible en: <http://repositorio.unprg.edu.pe/handle/UNPRG/5322>.
- BANCO MUNDIAL**, 2020. Pandemias: Panorama general. *World Bank* [en línea]. [Consulta: 31 mayo 2020]. Disponible en: <https://www.bancomundial.org/es/topic/pandemics>.
- BLANCO LAURENTE, L.F.**, 2018. Proyección de la producción de papa en Puno: Una aplicación de la metodología de Box-Jenkins. *Semestre Económico* [en línea], vol. 7, no. 1, pp. 7-24. [Consulta: 11 junio 2020]. ISSN 2523-0840. Disponible en: <http://revistas.unap.edu.pe/seconomico/index.php/SECONOMICO/article/view/160>.
- BOX, G.E. y JENKINS, G.M.**, 1976. *Time Series Analysis: Forecasting and Control*. S.l.: Holden-Day. ISBN 978-0-8162-1104-3.
- BRUGGER, S.**, 2010. Capital Especulativo y Crisis Bursátil en América Latina, Contagio, Crecimiento y Convergencia (1993-2005) [tesis doctoral]. *Facultad de Economía, UNAM*,
- BURKI, T.K.**, 2020. The Russian vaccine for COVID-19. *The Lancet Respiratory Medicine* [en línea], vol. 8, no. 11, pp. e85-e86. [Consulta: 12 diciembre 2020]. ISSN 2213-2600, 2213-2619. DOI 10.1016/S2213-2600(20)30402-1. Disponible en: [https://www.thelancet.com/journals/lanres/article/PIIS2213-2600\(20\)30402-1/abstract](https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(20)30402-1/abstract).
- BÓRRAS, B.**, 2004. Econometría empresarial II. *Universidad Autónoma de Madrid (UAM)* [en línea]. [Consulta: 20 enero 2021]. Disponible en: <https://www.doccity.com/es/econometria-empresarial-ii-modelos-lineales-sin-estacionalidad/3124592/>.
- CAO ABAD, R. y FERNÁNDEZ CASAL, R.**, 2020. *Técnicas de Remuestreo* [en línea]. S.l.: s.n. [Consulta: 10 enero 2021]. Disponible en: https://rubenfcasal.github.io/book_remuestreo/.

- CHATFIELD, C.**, 2013. *The Analysis of Time Series: Theory and Practice* [en línea]. S.l.: Springer. ISBN 978-1-4899-2925-9. Disponible en: [https://books.google.es/books?hl=es&lr=&id=u1D5BwAAQBAJ&oi=fnd&pg=PR12&dq=The+analysis+of+time+series+theory+and+practices+Chatfield+\(1978\)&ots=LAUnkCYTP-&sig=lv0Y8URwsY13eY-f8noZ2Q7Pruw#v=onepage&q&f=false](https://books.google.es/books?hl=es&lr=&id=u1D5BwAAQBAJ&oi=fnd&pg=PR12&dq=The+analysis+of+time+series+theory+and+practices+Chatfield+(1978)&ots=LAUnkCYTP-&sig=lv0Y8URwsY13eY-f8noZ2Q7Pruw#v=onepage&q&f=false).
- CHOI, K. y THACKER, S.B.**, 1981. An evaluation of influenza mortality surveillance, 1962–1979: I. Time series forecasts of expected pneumonia and influenza deaths. *American journal of epidemiology*, vol. 113, no. 3, pp. 215-226.
- CONTRERAS JUÁREZ, A., et al.**, 2016. Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos. *Estudios Gerenciales* [en línea], vol. 32, no. 141, pp. 387-396. [Consulta: 2 junio 2020]. ISSN 0123-5923. DOI 10.1016/j.estger.2016.11.002. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0123592316300754>.
- CRYER, J.D. y CHAN, K.-S.**, 2008. *Time Series Analysis. New York: Springer Science* [en línea]. 2. New York: Springer-Verlag. [Consulta: 20 enero 2021]. Springer Texts in Statistics. ISBN 978-0-387-75958-6. Disponible en: <https://www.springer.com/gp/book/9780387759586>.
- ECHEGARAY MUNENAKA, V.C.**, 2017. Pronóstico de demanda utilizando la metodología de Box-Jenkins. En: Accepted: 2018-04-12T21:05:08Z, *Universidad Nacional Jorge Basadre Grohmann* [en línea], pp. 6-50. [Consulta: 11 junio 2020]. Disponible en: <http://repositorio.unjbg.edu.pe/handle/UNJBG/1532>.
- EFRON, B.**, 1979. Computers and the theory of statistics: thinking the unthinkable. *SIAM review*, vol. 21, no. 4, pp. 460-480. ISSN 0036-1445.
- EFRON, B. y TIBSHIRANI, R.**, 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pp. 54-75. ISSN 0883-4237.
- EFRON, B. y TIBSHIRANI, R.**, 1994. An Introduction to the Bootstrap. New York, NY: Chapman & Hall. *CRC Monographs on Statistics & Applied Probability*,
- FOURNIES, A.**, 2015. *Modelos ARIMA y Método Box and Jenkins*. DOI 10.13140/RG.2.1.2907.0883
- GATES, B.**, 2020. Responding to Covid-19 ¿A Once-in-a-Century Pandemic? *New England Journal of Medicine* [en línea], vol. 382, no. 18, pp. 1677-1679. [Consulta: 31 mayo 2020]. ISSN 0028-4793. DOI 10.1056/NEJMp2003762. Disponible en: <https://doi.org/10.1056/NEJMp2003762>.
- GONZALEZ, O.**, 2018. Bootstrapping de datos de series de tiempo. *Quantdare* [en línea]. [Consulta: 10 enero 2021]. Disponible en: <https://quantdare.com/bootstrapping-time-series-data/>.

- GUACHALÁ COBA, B.L.**, 2017. Pronóstico de ventas de la Farmacéutica Sanofi usando series temporales. [en línea], pp. 1-96. Disponible en: <http://www.dspace.uce.edu.ec/bitstream/25000/13350/1/T-UCE-0011-336.pdf>.
- HALLO, A., et al.**, 2020. Perspective from Ecuador, the Second Country with More Confirmed Cases of Coronavirus Disease 2019 in South America. *Cureus* [en línea], vol. 12, no. 3. [Consulta: 4 junio 2020]. ISSN 2168-8184. DOI 10.7759/cureus.7452. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7187996/>.
- HE, Z. y TAO, H.**, 2018. Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. *International Journal of Infectious Diseases* [en línea], vol. 74, pp. 61-70. [Consulta: 2 junio 2020]. ISSN 1201-9712. DOI 10.1016/j.ijid.2018.07.003. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1201971218344618>.
- HOEHL, S., et al.**, 2020. Evidence of SARS-CoV-2 infection in returning travelers from Wuhan, China. *New England Journal of Medicine*, vol. 382, no. 13, pp. 1278-1280.
- HOLMES, E., et al.**, 2019. Applied time series analysis for fisheries and environmental data. URL <https://nwfsc-timeseries.github.io/atsa-labs>. NOAA Fisheries, Northwest Fisheries Science Center [en línea], vol. 2725. Disponible en: <https://nwfsc-timeseries.github.io/atsa-labs>.
- HUAMANI, N.B., et al.**, 2020. Casos confirmados y mortalidad por COVID-19 en Sudamérica. *PURIQ* [en línea], vol. 2, no. 3, pp. 307-327. [Consulta: 3 febrero 2021]. ISSN 2707-3602. DOI 10.37073/puriq.2.3.95. Disponible en: <https://www.revistas.unah.edu.pe/index.php/puriq/article/view/95>.
- HUANG, C., et al.**, 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* [en línea], vol. 395, no. 10223, pp. 497-506. [Consulta: 13 diciembre 2020]. ISSN 0140-6736. DOI 10.1016/S0140-6736(20)30183-5. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0140673620301835>.
- HYNDMAN, R.J. y ATHANASOPOULOS, G.**, 2018. *Forecasting: principles and practice* [en línea]. S.l.: OTexts. ISBN 978-0-9875071-1-2. Disponible en: <https://Otexts.com/fpp2/>.
- JAIN, G. y MALLICK, B.**, 2017. A Study of Time Series Models ARIMA and ETS. [en línea]. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. [Consulta: 9 enero 2021]. ID 2898968. Disponible en: <https://papers.ssrn.com/abstract=2898968>.
- KHAN, F.M. y GUPTA, R.**, 2020. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience* [en línea], vol. 1, no. 1, pp. 12-18. [Consulta: 22 diciembre 2020]. ISSN 2666-4496. DOI 10.1016/j.jnlssr.2020.06.007. Disponible en: <http://www.sciencedirect.com/science/article/pii/S2666449620300074>.
- KÜNSCH, H.R.**, 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, vol. 17, no. 3, pp. 1217-1241. ISSN 0090-5364.

- CHIH, L., et al.**, 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19). *International Journal of Antimicrobial Agents* [en línea], vol. 55, no. 3, pp. 105924. [Consulta: 31 mayo 2020]. ISSN 0924-8579. DOI 10.1016/j.ijantimicag.2020.105924. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0924857920300674>.
- LARCO ÁLVAREZ, D.M.**, 2018. Diseño de gráficos de control no paramétricos para datos autocorrelacionados e independientes utilizando técnicas de remuestreo. En: Accepted: 2018-12-06T16:48:36Z [en línea], [Consulta: 11 junio 2020]. Disponible en: <http://bibdigital.epn.edu.ec/handle/15000/19891>.
- LEÓN ÁLVAREZ, A.L., et al.**, 2016. Ronda clínica y epidemiológica. Series de tiempo. *Iatreia* [en línea], vol. 29, pp. 373-381. ISSN 0121-0793. Disponible en: http://www.scielo.org.co/scielo.php?pid=S0121-07932016000300373&script=sci_abstract&tlng=es#:~:text=El%20an%C3%A1lisis%20de%20series%20de,que%20se%20relacionan%20entre%20s%C3%AD.
- LIU, R.Y. y SINGH, K.**, 1992. Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*, vol. 225, pp. 248.
- LIU, R.Y. y TANG, J.**, 1996. Control charts for dependent and independent measurements based on bootstrap methods. *Journal of the American Statistical Association*, vol. 91, no. 436, pp. 1694-1700. ISSN 0162-1459.
- LUK, H.K.H., et al.**, 2019. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infection, Genetics and Evolution* [en línea], vol. 71, pp. 21-30. [Consulta: 31 mayo 2020]. ISSN 1567-1348. DOI 10.1016/j.meegid.2019.03.001. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1567134819300310>.
- MAKRIDAKIS, S.**, 1993. Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting* [en línea], vol. 9, no. 4, pp. 527-529. [Consulta: 26 enero 2021]. ISSN 0169-2070. Disponible en: https://econpapers.repec.org/article/eeeintfor/v_3a9_3ay_3a1993_3ai_3a4_3ap_3a527-529.htm.
- MALEKI, M., et al.**, 2020. Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models. *Chaos, Solitons & Fractals* [en línea], vol. 140, pp. 110151. [Consulta: 12 diciembre 2020]. ISSN 0960-0779. DOI 10.1016/j.chaos.2020.110151. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0960077920305476>.
- MANZANO, F.A. y RUSSO, G.**, 2020. ¿Cómo se mide la mortalidad? Tratando de ayudar en la crisis por el coronavirus (COVID-19). [en línea], DOI 10.13140/RG.2.2.15022.46402. Disponible en: https://www.researchgate.net/publication/340104902_Como_se_mide_la_mortalidad_Tratando_de_ayudar_en_la_crisis_por_el_coronavirus_COVID-19.

- MARTÍNEZ, H.**, 2016. Virus: Una petita introducció al món dels virus. *Biologia on-line. Revista de divulgació de la Facultat de Biologia* [en línea], vol. 5, no. 2. [Consulta: 4 junio 2020]. ISSN 2339-5745. Disponible en: https://revistes.ub.edu/index.php/b_on/article/view/16214.
- LÓPEZ, C.**, 2018. 🏆 Series temporales con R → Estacionariedad y raíces unitarias. *FinanzasZone* [en línea]. [Consulta: 20 enero 2021]. Disponible en: <https://finanzaszone.com/analisis-y-prediccion-de-series-temporales-con-r-ii-estacionariedad-y-raices-unitarias/>.
- MELO MORÍN, J. y SANTANA ESPARZA, G.**, 2016. Minado de series de tiempo utilizando la metodología ARIMA. *Revista de Investigación y Desarrollo*, vol. 2, no. 5, pp. 21-31.
- MIRANDA MOLES, A.**, 2003. El Método de Remuestreo y su Aplicación en la Investigación Biomédica, pp. 149.
- MOLINA PRENDES, N. y MEJIAS HERRERA, M.L.**, 2020. Impacto social de la COVID-19 en Brasil y Ecuador. *EDUMECENTRO* [en línea], vol. 12, no. 3, pp. 277-283. [Consulta: 16 noviembre 2020]. ISSN 2077-2874. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2077-28742020000300277&lng=es&nrm=iso&tlng=pt.
- MUÑOZ, R.**, 2019. *Métodos de remuestreo: Jackknife y Bootstrap* [en línea]. Murcia: Universidad de Murcia facultad de Matemáticas. Disponible en: https://www.um.es/documents/118351/14493552/Jord%C3%A1+Mu%C3%B1oz+TF_48742912+%28publicar%29.pdf/41a71924-1bfd-4617-a28d-9c9709a17582.
- MUÑOZ SANTIAGO, A. et al.**, 2017. Pronóstico del precio de la energía en Colombia utilizando modelos ARIMA con IGARCH. En: ISSNE 2145-454x, *Revista de Economía del Rosario* [en línea], vol. 20, no. 1, pp. 127-161. ISSN 0123-5362. Disponible en: https://revistas.urosario.edu.co/xml/5095/509554391005/509554391005visor_jats.pdf.
- NOSSA, L.P., et al.**, 2017. Estimación espectral de series de tiempo de absorbancia uv-vis para el monitoreo de calidad de aguas. *Ingeniería* [en línea], vol. 22, no. 2, pp. 3. [Consulta: 2 junio 2020]. ISSN 0121-750X. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=5973271>.
- NOVEL, C.P.E.R.E.**, 2020. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi* [en línea], vol. 41, no. 2, pp. 145. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32064853/>.
- ORGANIZACIÓN MUNDIAL DE LA SALUD**, 2020. Coronavirus disease (COVID-2019) situation reports. *World Health Organization* [en línea]. Disponible en: <https://www.who.int/emergencias/diseases/novel-coronavirus-2019/situation-reports/>.

ORGANIZACIÓN PANAMERICANA DE LA SALUD, 2017. *Lineamientos básicos para el análisis de la mortalidad* [en línea]. Washington DC: s.n. ISBN 978-92-75-31981-9. Disponible en: <https://iris.paho.org/handle/10665.2/34492>.

ORGANIZACIÓN PANAMERICANA DE LA SALUD y O.M.S., 2020. La OMS caracteriza a COVID-19 como una pandemia - OPS/OMS | Organización Panamericana de la Salud. [en línea]. [Consulta: 16 noviembre 2020]. Disponible en: <http://www.paho.org/es/noticias/11-3-2020-oms-caracteriza-covid-19-como-pandemia>.

OROSCO GAVILÁN, C., 2017. Metodología Bootstrap en el análisis serial de los ingresos mensuales. Hotel El Centro, Ayacucho 2011 - 2014. En: Accepted: 2018-08-25T14:36:27Z, *Universidad Nacional de San Cristóbal de Huamanga* [en línea], pp. 81. [Consulta: 11 junio 2020]. Disponible en: <http://repositorio.unsch.edu.pe/handle/UNSCH/1776>.

OTERO, W., et al., 2020. Procedimientos endoscópicos y pandemia COVID-19. *Revista Colombiana de Gastroenterología*, vol. 35, no. 1, pp. 65-75.

PÁEZ, E., et al., 2016. Aplicación de series de tiempo en la realización de pronósticos de producción. *Fuentes: El reventón energético* [en línea], vol. 14, no. 1, pp. 79-88. [Consulta: 2 junio 2020]. ISSN 1657-6527. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=6371161>.

PALACIOS CRUZ, M., et al., 2020. COVID-19, una emergencia de salud pública mundial. *Revista clínica española* [en línea], vol. 221, no. 1, pp. 55-61. ISSN 1578-1860. DOI 10.1016/j.rce.2020.03.001. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32204922>. PubMed: 32204922

PAREDES RODRÍGUEZ, N.P., 2017. Pronóstico del caudal máximo del río Huaura por Modelos ARIMA y redes neuronales. En: Accepted: 2018-08-29T17:32:11Z, *Universidad Nacional de Trujillo* [en línea], pp. 7-81. [Consulta: 11 junio 2020]. Disponible en: <http://dspace.unitru.edu.pe/handle/UNITRU/10637>.

PERAZA DE APARICIO, C.X., 2020. Salud laboral frente a la pandemia del COVID-19 en Ecuador. *MediSur* [en línea], vol. 18, no. 3, pp. 507-511. [Consulta: 16 noviembre 2020]. ISSN 1727-897X. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1727-897X2020000300507&lng=es&nrm=iso&tlng=pt.

PÉREZ VELÁZQUEZ, L.C., 2020. La COVID-19: reto para la ciencia mundial. *Anales de la Academia de Ciencias de Cuba*, vol. 10, no. 2, pp. 763. ISSN 2304-0106.

QUINDE ROSALES, V., et al., 2020. Relación entre el crecimiento y el desarrollo económico: caso Ecuador. *Revista Universidad y Sociedad* [en línea], vol. 12, no. 3, pp. 60-66. [Consulta: 20 enero 2021]. ISSN 2218-3620. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2218-36202020000300060&lng=es&nrm=iso&tlng=pt.

QUINDE ROSALES, V., et al., 2020. Análisis de cointegración entre el crecimiento económico y deterioro medio ambiental: un análisis empírico del desarrollo sostenible de América Latina y el Caribe. *Revista de Economía del Caribe* [en línea], vol. 0, no. 24, pp. 8-25-25. [Consulta: 22 diciembre 2020]. ISSN 2145-9363. DOI 10.14482/rec.v0i24.11297. Disponible en: <http://rcientificas.uninorte.edu.co/index.php/economia/article/view/11297>.

REHMAN, A., et al., 2019. Economic perspectives of cotton crop in Pakistan: A time series analysis (1970–2015) (Part 1). *Journal of the Saudi Society of Agricultural Sciences* [en línea], vol. 18, no. 1, pp. 49-54. [Consulta: 21 diciembre 2020]. ISSN 1658-077X. DOI 10.1016/j.jssas.2016.12.005. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1658077X1630162X>.

RODRÍGUEZ MORALES, A.J., et al., 2020. Preparación y control de la enfermedad por coronavirus 2019 en América Latina. *Acta Médica Peruana* [en línea], vol. 37, no. 1, pp. 3-7. [Consulta: 16 noviembre 2020]. ISSN 1728-5917. DOI 10.35663/amp.2020.371.909. Disponible en: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1728-59172020000100003&lang=es.

ROJAS, V.C., 2019. Un análisis de series de tiempo mediante modelos SARIMAX para la proyección de demanda de carga en el puerto del Callao. *Revista de análisis económico y financiero* [en línea], vol. 1, no. 3, pp. 17. [Consulta: 9 enero 2021]. ISSN 2617-9989. Disponible en: <https://www.aulavirtualusmp.pe/ojs/index.php/raef/article/view/1694>.

ROMERO, M.R., 2020. Guía de buenas prácticas en Odontología para Uruguay durante la pandemia Covid-19. *Odontostomatología* [en línea], vol. 22, no. especial, pp. 25-37. Disponible en: http://www.scielo.edu.uy/scielo.php?script=sci_arttext&pid=S1688-93392020000200025.

ROSAS ROJAS, E., et al., 2017. Pronóstico de las remesas familiares de México 2017, aplicación empírica mediante modelos ARIMA y alisado exponencial. En: Accepted: 2017-11-13T19:06:39Z [en línea], pp. 20. ISSN 2463-0527. Disponible en: <http://ri.uaemex.mx/handle/20.500.11799/67644>.

RUIZ, G.P. y LEÓN, A.C., 2020. Evolución de la enfermedad por coronavirus (COVID-19) en Ecuador. *La Ciencia al Servicio de la Salud* [en línea], vol. 11, no. 1, pp. 5-15. [Consulta: 4 junio 2020]. ISSN 1390-874X. Disponible en: <http://revistas.espech.edu.ec/index.php/cssn/article/view/441>.

SABA, T., et al., 2021. Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types. *Microscopy Research and Technique* [en línea], vol. n/a, no. n/a. [Consulta: 9 febrero 2021]. ISSN 1097-0029. DOI <https://doi.org/10.1002/jemt.23702>. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jemt.23702>.

SAHUQUILLO, M.R., 2020. Rusia pone en órbita mundial su vacuna Sputnik V contra el coronavirus. *EL PAÍS* [en línea]. [Consulta: 12 diciembre 2020]. Disponible en:

<https://elpais.com/sociedad/2020-12-11/rusia-pone-en-orbita-mundial-su-vacuna-sputnik-v-contras-el-coronavirus.html>.

SALAS ASENCIOS, R., et al., 2020. Coronavirus COVID-19: conociendo al causante de la pandemia. *The Biologist* [en línea], vol. 18, no. 1, pp. 9-27. [Consulta: 31 mayo 2020]. ISSN 1994-9073. DOI 10.24039/rtb2020181442. Disponible en: <http://revistas.unfv.edu.pe/index.php/rtb/article/view/442>.

SALDAÑA VALQUI, E.J., 2016. *Modelo predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad de pacientes hospitalizados* [en línea]. Trujillo – Perú: Universidad Privada Antenor Orrego - UPAO. [Consulta: 19 enero 2021]. Disponible en: <http://repositorio.upao.edu.pe/handle/upaorep/1927>.

SANTANA LÓPEZ, B.N., et al., 2019. Creencias y actitudes de trabajadores sanitarios y estudiantes de enfermería de una región de España ante una pandemia de gripe. *Revista Peruana de Medicina Experimental y Salud Pública* [en línea], vol. 36, pp. 481-486. [Consulta: 31 mayo 2020]. ISSN 1726-4634, 1726-4634, 1726-4642. DOI 10.17843/rpmesp.2019.363.4371. Disponible en: <https://www.scielosp.org/article/rpmesp/2019.v36n3/481-486/es/>.

STOCK, J.H., et al., 2012. *Introducción a la Econometría* [en línea]. 3era. Madrid - España: Pearson Educación, S.A. ISBN 978-84-8322-967-5. Disponible en: <https://danielmorochoruiz.files.wordpress.com/2018/05/0000017.pdf>.

VAN DOREMALEN, N., et al., 2020. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine* [en línea], vol. 382, no. 16, pp. 1564-1567. [Consulta: 31 mayo 2020]. ISSN 0028-4793. DOI 10.1056/NEJMc2004973. Disponible en: <https://doi.org/10.1056/NEJMc2004973>.

VILLEGAS CHIROQUE, M., 2020. Pandemia de COVID-19: pelea o huye. *Revista Experiencia en Medicina del Hospital Regional Lambayeque*, vol. 6, no. 1.

WANG, C., et al., 2020. A novel coronavirus outbreak of global health concern. *The Lancet* [en línea], vol. 395, no. 10223, pp. 470-473. [Consulta: 16 noviembre 2020]. ISSN 01406736. DOI 10.1016/S0140-6736(20)30185-9. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S0140673620301859>.

YANG, Q., et al., 2020. Research on COVID-19 based on ARIMA model—Taking Hubei, China as an example to see the epidemic in Italy. *Journal of Infection and Public Health* [en línea], vol. 13, no. 10, pp. 1415-1418. [Consulta: 12 diciembre 2020]. ISSN 1876-0341. DOI 10.1016/j.jiph.2020.06.019. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1876034120305293>.

YE, T. y YANG, X., 2020. Analysis and prediction of confirmed COVID-19 cases in China with uncertain time series. *Fuzzy Optimization and Decision Making* [en línea], [Consulta: 12 diciembre 2020]. ISSN 1573-2908. DOI 10.1007/s10700-020-09339-4. Disponible en: <https://doi.org/10.1007/s10700-020-09339-4>.

ZAVALA GUEVARA, J.A., 2017a. *Pronóstico de la exportación pesquera por redes neuronales y modelos ARIMA* [en línea]. Trujillo - Perú: Universidad Nacional de Trujillo. [Consulta: 11 junio 2020]. Disponible en: <http://dspace.unitru.edu.pe/handle/UNITRU/10400>.

ZAVALA GUEVARA, J.A., 2017b. *Pronóstico de la exportación pesquera por redes neuronales y modelos ARIMA* [en línea]. Trujillo - Perú: Universidad Nacional de Trujillo. Disponible en: <http://dspace.unitru.edu.pe/handle/UNITRU/10400>.

ANEXOS

ANEXO A: Código en R (versión 4.0.3) del análisis exploratorio de las series de tiempo

Librerías utilizadas

```
library(broman)
```

```
library(extrafont)
```

```
library(modeest)
```

```
library(psych)
```

```
library(ggplot2)
```

```
library(gridExtra)
```

```
library(reshape2)
```

```
library(tseries)
```

```
library(lmtest)
```

```
library(ggfortify)
```

```
library(prophet)
```

```
library(tidyverse)
```

```
library(caret)
```

```
library(knitr)
```

```
library(purrr)
```

```
library(forecast)
```

```
library(plotrix)
```

Lectura de datos

```
setwd("C:/Users/... ") # seleccionar el directorio
```

```
datos<- read.csv("CasosDia.csv",sep = ";")
```

```
attach(datos)
```

```
str(datos)
```

Estadísticas descriptivas

```
summary(datos$Confirmados)
```

```
sd(datos$Confirmados)
```

```
summary(datos$Fallecidos)
```

```
sd(datos$Fallecidos)
```

Proporciones de casos de COVID-19

```
NC <- sum(datos$Confirmados)
```

```
NF <- sum(datos$Fallecidos)
```

```
prNF <- NF/NC*100;prNF
```

```
prNC <- (NC-NF)/NC*100;prNC
```

```
proporciones <- c(prNC, prNF) # creamos un vector con proporciones
```

Diagrama de pastel

```
etiquetas <- c("95.03%", "4.97%") # vector con etiquetas
```

```
pie3D(proporciones,labels=etiquetas, border = F, explode=.1, main="CASOS DE COVID-19 EN  
ECUADOR", col = c("Pink","Blue3"))
```

```
legend("topright", c("Casos Confirmados", "Casos Fallecidos"), cex = 0.9, fill = c("Pink",  
"Blue3"), border = F,box.col = F)
```

Series de tiempo

```
SerieC <- ts(PCR_Confirmados, start = c(3,13), frequency = 365)
```

```
SerieF <- ts(PCR_Fallecidos, start = c(3,13), frequency = 365)
```

Grafico de la serie de tiempo

```
mediaC <- mean(SerieC)
```

```
gc <- ggplot(datos, aes(Fecha, PCR_Confirmados)) + geom_line(lwd=0.7, color="Midnight  
Blue") + geom_hline(aes(yintercept=mediaC), colour="#990000", linetype="dashed") + ggtitle  
("SERIE DE CASOS CONFIRMADOS POR COVID-19 EN ECUADOR", subtitle =  
"13/03/2020-31/10/2020") + theme(plot.title = element_text(family = "Times New Roman",  
size=12, hjust = 0.5, face="bold", lineheight=1), plot.subtitle= element_text(family = "Times  
New Roman", size=9,hjust = 0.5, face="plain",lineheight=1)) + scale_x_date("TIEMPO(Días)")  
+ ylab("NÚMERO ACUMULADO") + xlab("") + theme(axis.title.x = element_text(family =  
"Times New Roman", face="bold", vjust=-0.5, colour="Midnight Blue", size=8)) +  
theme(axis.title.y = element_text(family = "Times New Roman", face="bold", vjust=0.5,  
colour="Midnight Blue", size=8))
```

```
mediaF <- mean(SerieF)
```

```
gf <- ggplot(datos, aes(Fecha, PCR_Fallecidos)) + geom_line(size = 0.7, color="Midnight  
Blue") + geom_hline(aes(yintercept=mediaF), colour="#990000", linetype="dashed") + ggtitle  
("SERIE DE CASOS FALLECIDOS POR COVID-19 EN ECUADOR", subtitle = "13/03/2020-  
31/10/2020") + theme(plot.title = element_text(family = "Times New Roman", size=12, hjust =  
0.5, face="bold", lineheight=1), plot.subtitle= element_text(family = "Times New  
Roman",size=9, hjust = 0.5, face="plain", lineheight=1)) + scale_x_date("TIEMPO(Días)") +  
ylab("CASOS FALLECIDOS") + xlab("") + theme(axis.title.x = element_text(family = "Times  
New Roman", face="bold", vjust=-0.5,colour="Midnight Blue", size=8)) + theme(axis.title.y =  
element_text(family = "Times New Roman", face="bold", vjust= 0.5, colour="Midnight Blue",  
size=8))
```

```
grid.arrange(gc, gf)
```

ANEXO B: Código en R (versión 4.0.3) metodología de Box-Jenkins

Correlogramas de las series de tiempo

```
acfC <- ggAcf(SerieC, plot = TRUE, lwd=.7,col="Violet Red", lag.max = 60) + ggtitle ("CASOS
CONFIRMADOS POR COVID-19 EN ECUADOR", subtitle = "13/03/2020-31/10/2020") +
theme(plot.title = element_text(family = "Times New Roman", size=12, hjust = 0.5, face="bold",
lineheight=1), plot.subtitle= element_text(family = "Times New Roman",size=9, hjust =
0.5,face="plain",lineheight=1)) + ylab("ACF") + xlab("") + theme(axis.title.x =
element_text(family = "Times New Roman", face="bold", vjust=-0.5, colour="Midnight Blue",
size=9)) + theme(axis.title.y = element_text(family = "Times New Roman", face="bold",
vjust=0.5,colour="Midnight Blue", size=9))
pacfC <- ggPacf(SerieC, plot = TRUE, lwd=.7, col="Violet Red", lag.max = 60)+ ggtitle (" ") +
ylab("PACF") + xlab("REZAGOS") + theme(axis.title.x = element_text(family = "Times New
Roman", face="bold", vjust=-0.5, colour="Midnight Blue", size=9)) + theme(axis.title.y =
element_text(family = "Times New Roman", face="bold", vjust=0.5, colour="Midnight Blue",
size=9))
grid.arrange(acfC, pacfC)
```

```
acfF <- ggAcf(SerieF, plot = TRUE, lwd=0.7,col="Violet Red", lag.max = 60) + ggtitle ("CASOS
FALLECIDOS POR COVID-19 EN ECUADOR", subtitle = "13/03/2020-31/10/2020") +
theme(plot.title = element_text(family = "Times New Roman", size=12, hjust = 0.5, face="bold",
lineheight=1), plot.subtitle= element_text(family = "Times New Roman",size=9, hjust =
0.5,face="plain",lineheight=1)) + ylab("ACF") + xlab("") + theme(axis.title.x =
element_text(family = "Times New Roman", face="bold", vjust=-0.5, colour="Midnight Blue",
size=9)) + theme(axis.title.y = element_text(family = "Times New Roman", face="bold",
vjust=0.5,colour="Midnight Blue", size=9))
```

```
pacfF <- ggPacf(SerieF, plot = TRUE, lwd=.7, col="Violet Red", lag.max = 60)+ ggtitle (" ") +
ylab("PACF") + xlab("REZAGOS") + theme(axis.title.x = element_text(family = "Times New
Roman", face="bold", vjust=-0.5, colour="Midnight Blue", size=9)) + theme(axis.title.y =
element_text(family = "Times New Roman", face="bold", vjust=0.5, colour="Midnight Blue",
size=9))
grid.arrange(acfF,pacfF)
```

Estacionariedad (Test Augmented Dickey-Fuller)

```
# Ho: Raiz unitaria (No estacionario)
```

```
# p < 0.05 -> Ho se rechaza - Es estacionario
```



```
adf.test(SerieC, alternative = "stationary")
adf.test(SerieF, alternative = "stationary")
```

Diferenciación

```
SerieC_diff1 <- diff(SerieC, differences = 1)
ndiffs(SerieC_diff1)
SerieC_diff2 <- diff(SerieC, differences = 2)
ndiffs(SerieC_diff2)
SerieF_diff1 <- diff(SerieF, differences = 1)
ndiffs(SerieF_diff1)
```

Identificación de los modelos

```
acfC_diff <- ggAcf(SerieC_diff2, plot = TRUE, lwd=0.7, col="Violet Red", lag.max = 60) +
ggtitle ("CASOS CONFIRMADOS POR COVID-19 EN ECUADOR", subtitle = "13/03/2020-
31/10/2020") + theme(plot.title = element_text(family = "Times New Roman", size=12, hjust =
0.5, face = "bold", lineheight=1), plot.subtitle= element_text(family = "Times New Roman",
size=9, hjust = 0.5, face="plain", lineheight=1)) + ylab("ACF") + xlab("") + theme(axis.title.x =
element_text(family = "Times New Roman", face="bold", vjust=-0.5, colour="Midnight Blue",
size=9)) + theme(axis.title.y = element_text(family = "Times New Roman", face="bold",
vjust=0.5, colour="Midnight Blue", size=9))
pacfC_diff <- ggPacf(SerieC_diff2, plot = TRUE, lwd=0.7, col="Violet Red", lag.max = 60) +
ggtitle (" ") + ylab("PACF") + xlab("REZAGOS") + theme(axis.title.x = element_text(family =
"Times New Roman", face="bold", vjust=-0.5, colour="Midnight Blue", size=9)) +
theme(axis.title.y = element_text(family = "Times New Roman", face="bold", vjust=0.5,
colour="Midnight Blue", size=9))
grid.arrange(acfC_diff,pacfC_diff)
```

```
acfF_diff <- ggAcf(SerieF_diff1, plot = TRUE, lwd=0.7, col="Violet Red", lag.max = 60) +
ggtitle ("CASOS FALLECIDOS POR COVID-19 EN ECUADOR", subtitle = "13/03/2020-
31/10/2020") + theme(plot.title = element_text(family = "Times New Roman", size=12, hjust =
0.5, face = "bold", lineheight=1), plot.subtitle= element_text(family = "Times New Roman",
size=9, hjust = 0.5, face="plain", lineheight=1)) + ylab("ACF") + xlab("") + theme(axis.title.x =
element_text(family = "Times New Roman", face="bold", vjust=-0.5, colour="Midnight Blue",
size=9)) + theme(axis.title.y = element_text(family = "Times New Roman", face="bold",
vjust=0.5, colour="Midnight Blue", size=9))
pacfF_diff <- ggPacf(SerieF_diff1, plot = TRUE, lwd=0.7, col="Violet Red", lag.max = 60) +
ggtitle (" ") + ylab("PACF") + xlab("REZAGOS") + theme(axis.title.x = element_text(family =
```

```
"Times New Roman", face="bold", vjust=-0.5, colour="Midnight Blue", size=9)) +  
theme(axis.title.y = element_text(family = "Times New Roman", face="bold", vjust=0.5,  
colour="Midnight Blue", size=9))  
grid.arrange(acfF_diff,pacfF_diff)
```

Estimación y selección del modelo

```
Arima(SerieC, c(0,2,1))
```

```
Arima(SerieC, c(0,2,2))
```

```
Arima(SerieC, c(1,2,0))
```

```
Arima(SerieC, c(2,2,0))
```

```
Arima(SerieC, c(1,2,1))
```

```
auto.arima(SerieF, trace = T)
```

```
Arima(SerieF, c(0,1,1))
```

```
Arima(SerieF, c(0,1,2))
```

```
Arima(SerieF, c(1,1,0))
```

```
Arima(SerieF, c(2,1,0))
```

```
Arima(SerieF, c(1,1,1))
```

```
auto.arima(SerieC, trace = T)
```

```
arima6 <- Arima(SerieF, c(1,1,1), include.drift = T)
```

Validación del modelo

```
arimaC <- auto.arima(SerieC)
```

```
arimaF <- auto.arima(SerieF)
```

1. Coeficientes significativos

```
coeftest(arimaC)
```

```
confint(arimaC)
```

```
checkresiduals(arimaC)
```

```
coeftest(arimaF)
```

```
confint(arimaF)
```

```
checkresiduals(arimaF)
```

2. Estacionariedad de los RESIDUOS

```
adf.test(arimaC$residuals, alternative = "stationary")
```

```
adf.test(arimaF$residuals, alternative = "stationary")
```

```
## 3. Supuesto de Ruido Blanco
```

```
Box.test(arimaC$residuals,type="Lj")
```

```
Box.test(arimaF$residuals, type="Ljung-Box")
```

```
## 4. Normalidad de los residuos
```

```
qqnorm(arimaC$residuals, main="Residuos ARIMA (0,2,2)", sub="Gráfico Q para evaluar normalidad")
```

```
qqline(arimaC$residuals)
```

```
shapiro.test(arimaC$residuals)
```

```
jarque.bera.test(arimaC$residuals)
```

```
qqnorm(arimaF$residuals, main="Residuos ARIMA(0,1,0)", sub="Gráfico Q para evaluar normalidad")
```

```
qqline(arimaF$residuals)
```

```
shapiro.test(arimaF$residuals)
```

```
jarque.bera.test(arimaF$residuals)
```

```
# Pronósticos
```

```
prediccionC <- forecast(arimaC, h=61, level = 95);prediccionC
```

```
prediccionF <- forecast(arimaF, h=61, level = 95);prediccionF
```

```
## Datos de los meses de noviembre y diciembre
```

```
NovDicC<- ts(datosAct$PCR_Confirmados, start = 3.672, frequency = 365);NovDicC
```

```
NovDicF <- ts(datosAct$PCR_Fallecidos, start = 3.672, frequency = 365)
```

```
## Gráfico
```

```
autoplot(SerieC)+ggtitle ("PRONÓSTICO CASOS CONFIRMADOS POR COVID-19 EN ECUADOR") + labs(subtitle = "REAL: 13/03/2020-31/10/2020 PRONÓSTICO=01/11/2020-31/12/2020") + theme(plot.title = element_text(family="Times New Roman",hjust = 0.5,face="bold",lineheight=1))+ theme (plot.subtitle = element_text(family="Times New Roman",size=9,hjust = 0.5, face="plain",lineheight=1)) + autolayer(prediccionC, series="Pronóstico") + autolayer(NovDicC, colour=T, series="Regist. Nov-Dic")
```

```
autoplot(SerieF)+ggtitle ("PRONÓSTICO CASOS FALLECIDOS POR COVID-19 EN ECUADOR") + labs(subtitle = "REAL: 13/03/2020-31/10/2020 PRONÓSTICO=01/11/2020-31/12/2020") + theme(plot.title = element_text(family="Times New Roman",hjust = 0.5,face="bold",lineheight=1)) + theme (plot.subtitle = element_text(family="Times New Roman",size=9,hjust = 0.5, face="plain",lineheight=1)) + autolayer(prediccionF, series="Pronóstico") + autolayer(NovDicF, colour= T, series="Regist. Nov-Dic")
```

ANEXO C: Código en R (versión 4.0.3) metodología Bootstrapping

Estadísticas Bootstrap

```
bootAC <- function(tsb) { #statistic for time - series bootstrap
  fit <- arima(tsb, c(0,2,2))
  return(coef(fit))
}
```

```
bootAF <- function(tsb) { #statistic for time - series bootstrap
  fit <- Arima(tsb, c(1,1,1),include.drift = T) #fit with AR(1)
  return(coef(fit))
}
```

```
bootAC(SerieC)
```

```
bootAF(SerieF)
```

```
require(boot)
```

```
set.seed(2021)
```

```
bootC <- tsboot(SerieC, bootAC, R =100, l =16, sim = "fixed", n.sim = 1000); bootC
```

```
mean(bootC$t[,1])
```

```
mean(bootC$t[,2])
```

```
boot.ci(bootC, index =1)
```

```
boot.ci(bootC, index =2)
```

```
set.seed(2021)
```

```
bootF <- tsboot(SerieF, bootAF, R =100, l =16, sim = "fixed", n.sim = 1000); bootF
```

```
summary(bootF)
```

```
mean(bootF$t[,1])
```

```
mean(bootF$t[,2])
```

```
mean(bootF$t[,3])
```

```
boot.ci(bootF, index =1)
```

```
boot.ci(bootF, index =2)
```

```
boot.ci(bootF, index =3)
```

Función para generar los bootstrapping

```
sim_BootPro <- function(data, R = 100, nsim=1000, bloque = 16, h, ARIMA=auto.arima,
level=95){
```

```
  set.seed(2021)
```

```
  sim <- bld.mbb.bootstrap(data, num=R, block_size = bloque) # Genera versiones bootstrap de
  una serie de tiempo usando el bootstrap de descomposición basado en Box-Cox y Loess.
```

```

h <- as.integer(h)
future <- matrix(0, nrow=nsim, ncol=h)
future <- sim %>% map(function(x){
simulate(ARIMA(x),nsim=h)}) %>%
unlist() %>% matrix(ncol = h, nrow = nsim, byrow = TRUE)
start <- tsp(data)[2]+1/365
bootstrap <- structure(list(
mean = future %>% colMeans() %>% ts(start = start, frequency = 365),
lower = future %>% as.data.frame() %>%
map_dbl(quantile, prob = 0.025) %>%
ts(start = start,frequency = 365),
upper = future %>% as.data.frame() %>%
map_dbl(quantile, prob = 0.975) %>%
ts(start = start,frequency = 365),
level=level),
class="forecast")
assign("bootstrap", bootstrap, envir = .GlobalEnv)
bootstrap
}

```

Pronósticos

```

sim_BootC <- sim_BootPro(SerieC, h=61)
sim_BootF <- sim_BootPro(SerieF, h=61)
arima_C <- SerieC %>% auto.arima() %>% forecast(h=h,level=95)
arima_F <- SerieF %>% auto.arima() %>% forecast(h=h,level=95)

```

Grafico de pronostico

```

autoplot(SerieC, colour ="Midnight Blue") + xlab("TIEMPO(Días)") + ylab("NÚMERO DE
CASOS") + autolayer(sim_BootC, series="ARIMA Bootstrap", PI=T) + autolayer(arima_C,
series="ARIMA(0,2,2)", PI=F) + autolayer(NovDicC, series="Regist. Nov-Dic", lwd=0.6) +
ggtitle ("PRONÓSTICO BOOTSTRAP CASOS CONFIRMADOS POR COVID-19 EN
ECUADOR") + labs(subtitle = "SERIE ORIGINAL: 13/03/2020-31/10/2020
\nPRONÓSTICO: 01/11/2020-31/12/2020") + theme(plot.title = element_text(family="Times
New Roman",size=12, hjust = 0.5,face="bold",lineheight=1)) + theme(plot.subtitle =
element_text(family="Times New Roman",size=8,hjust = 0.5, face="plain",lineheight=1.5)) +
theme(axis.title.x = element_text(family = "Times New Roman", face="bold", vjust=-0.5,

```

```
colour="Midnight Blue", size=9)) + theme(axis.title.y = element_text(family = "Times New Roman", face="bold", vjust=0.5, colour="Midnight Blue", size=9))
```

```
autoplot(SerieF, colour ="Midnight Blue") + xlab("TIEMPO(Días)") + ylab("NÚMERO DE CASOS") + autolayer(sim_BootF, series="ARIMA Bootstrap", PI=T) + autolayer(arima_F, series="ARIMA(1,1,1) Const.", PI=F) + autolayer(NovDicF, series="Regist. Nov-Dic", lwd=1) + ggtitle ("PRONÓSTICO BOOTSTRAP CASOS FALLECIDOS POR COVID-19 EN ECUADOR") + labs(subtitle = "SERIE ORIGINAL: 13/03/2020-31/10/2020 \nPRONÓSTICO: 01/11/2020-31/12/2020") + theme(plot.title = element_text(family="Times New Roman",size=12, hjust = 0.5,face="bold",lineheight=1)) + theme(plot.subtitle = element_text(family="Times New Roman",size=8,hjust = 0.5, face="plain",lineheight=1.5)) + theme(axis.title.x = element_text(family = "Times New Roman", face="bold", vjust=-0.5, colour="Midnight Blue", size=9)) + theme(axis.title.y = element_text(family = "Times New Roman", face="bold", vjust=0.5, colour="Midnight Blue", size=9))
```



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

**DIRECCIÓN DE BIBLIOTECAS Y RECURSOS DEL APRENDIZAJE
UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y DOCUMENTAL**

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 25 / 08 / 2021

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: <i>Maritza Janeth Cabrera Valladolid</i>
INFORMACIÓN INSTITUCIONAL
Facultad: <i>Ciencias</i>
Carrera: <i>Estadística</i>
Título a optar: <i>Ingeniera en Estadística Informática</i>
f. Analista de Biblioteca responsable: <i>Lic. Luis Caminos Vargas Mgs.</i>



0717-DBRAI-UPT-2021