

**UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE ECONOMÍA Y EMPRESA**

**CARRERA DE NEGOCIOS INTERNACIONALES**

**TEMA:**

**Aplicación de modelo ARIMA para el pronóstico de exportación de flores del  
Ecuador.**

**AUTORES:**

**Li Ye, Kevin Kinjang**

**Paz y Miño Robles, Zhukov Jafet**

**Trabajo de integración curricular previo a la obtención del título de**

**Licenciado en Negocios Internacionales**

**TUTOR:**

**Ing. Carrera Buri, Félix Miguel, Mgs.**

**Guayaquil, Ecuador**

**04 de septiembre del 2023**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL  
**FACULTAD DE ECONOMÍA Y EMPRESA**  
**CARRERA NEGOCIOS INTERNACIONALES**

**CERTIFICACIÓN**

Certificamos que el presente trabajo de integración curricular fue realizado en su totalidad por **Li Ye, Kevin Kinjang y Paz y Miño Robles, Zhukov Jafet** como requerimiento para la obtención del título de **Licenciados en Negocios Internacionales**.

**TUTORES**

f. 

Ing. Carrera Buri, Félix Miguel, Mgs.

**DIRECTOR DE LA CARRERA**

f. \_\_\_\_\_

Ing. Hurtado Cevallos, Gabriela Elizabeth Mgs.

**Guayaquil, a los 04 del mes de septiembre del año 2023**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA**  
**CARRERA NEGOCIOS INTERNACIONALES**

**DECLARACIÓN DE RESPONSABILIDAD**

**Li Ye, Kevin Kinjang y Paz y Miño Robles, Zhukov Jafet**

**DECLARO QUE:**

El Trabajo de Integración Curricular, **Aplicación de modelo ARIMA para el pronóstico de exportación de flores del Ecuador** previo a la obtención del título de **Licenciados en Negocios Internacionales**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Integración Curricular referido.

**Guayaquil, a los 04 del mes de septiembre del año 2023**

**LOS AUTORES:**

f.   
\_\_\_\_\_

**Li Ye, Kevin Kinjang**

f.   
\_\_\_\_\_

**Paz y Miño Robles, Zhukov Jafet**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA  
CARRERA NEGOCIOS INTERNACIONALES

### AUTORIZACIÓN

**Nosotros, Kevin Kinjang Li Ye y Zhukov Jafet Paz y Miño Robles**

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Integración Curricular, **Aplicación de modelo ARIMA para el pronóstico de exportación de flores del Ecuador**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

**Guayaquil, a los 04 del mes de septiembre del año 2023**

**LOS AUTORES:**

f. \_\_\_\_\_

**Li Ye, Kevin Kinjang**

f. \_\_\_\_\_

**Paz y Miño Robles, Zhukov Jafet**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA  
CARRERA NEGOCIOS INTERNACIONALES

### REPORTE URKUND

**URKUND**

Documento	<a href="#">Aplicación de modelo ARIMA para el pronóstico de exportación de flores del ecuador.docx</a> (D173166783)
Presentado	2023-08-27 13:43 (-05:00)
Presentado por	kevin.li@cu.ucsg.edu.ec
Recibido	felix.carrera01.ucsg@analysis.orkund.com
Mensaje	TESIS AVANCE FINAL LI YE Y PAZ Y MIÑO <a href="#">Mostrar el mensaje completo</a> 0% de estas 49 páginas, se componen de texto presente en 0 fuentes.

URKUND interface footer with navigation icons: search, zoom, quote, cursor, up, left, right.

...

Ing. Carrera Buri, Félix Miguel

## **AGRADECIMIENTO**

No encuentro palabras suficientes para expresar la gratitud que siento por mis padres y por todo lo que han hecho por mí. En los momentos más difíciles, siempre estuvieron allí para darme una mano. Su dedicación, amor y apoyo incondicional han sido la base sobre los cuales he construido mi camino y han dado forma a la persona que soy hoy.

Además, no puedo dejar pasar la oportunidad de expresar mi más profundo agradecimiento a Eugenia y Jorge asimismo a sus padres. Los consejos y experiencias que me han brindado me han guiado e inspirado a lo largo de este viaje.

Quiero agradecer a Vivian y a sus padres por la compañía que me brindaron en los primeros semestres de universidad. Me han ayudado a sentirme menos solo cuando llegue a esta ciudad y sus consejos me han facilitado en la nueva etapa de mi vida.

Asimismo, quiero agradecer a mi compañero de tesis Jafet por su colaboración y apoyo durante todo el proceso de este trabajo. Enfrentamos desafíos y celebramos avances. Su habilidad para colaborar, debatir y trabajar en equipo fue un factor clave en el éxito de este proyecto.

Agradezco a todos mis amigos que conocí en la Universidad, por todos los buenos momentos y las alegrías compartida. Su amistad fue importante y han dejado una marca perdurable en mi experiencia universitaria.

Por último, quiero agradecer a todos los profesores por su dedicación y orientación a lo largo de mi trayectoria universitaria. Cada clase fue una oportunidad para aprender y crecer. Y adicionalmente, deseo expresar mi más sincero agradecimiento a mi tutor de tesis al Ing. Carrera Buri Félix Miguel su orientación ha sido fundamental para el desarrollo de este trabajo.

Kevin Kinjang Li Ye

## **DEDICATORIA**

A lo largo de mi camino académico, mis padres han sido mi fuente constante de apoyo, inspiración y amor. Por ello quiero dedicarles esta tesis como un reflejo de la gratitud que siento por haberme brindado las herramientas y oportunidades para crecer como estudiante y como persona.

Kevin Kinjang Li Ye

## **AGRADECIMIENTO**

En este capítulo de mi padre y madre son dos personas especiales que han estado a mi lado desde momentos inmemorables, gracias a sus consejos y su apoyo moral y espiritual he podido salir adelante, le agradezco infinitamente a mi padre Walther Zhukov Paz y Miño por darme las pautas, los consejos, herramientas y aptitudes necesarias para poder lograr cumplir con mis objetivos, y a mi madre Lucrecia Robles, le agradezco por su amor incondicional, por su sabiduría espiritual y por enseñarme a sobrellevar todos los problemas de la vida y ser resiliente respecto a las adversidades que esta misma nos presenta.

Agradezco haber conocido a diversas personas a lo largo de este camino. A Genesis Mendieta, una persona que me mostró siempre su apoyo y comprensión, gracias por tu cariño, por tus consejos y por apoyarme siempre a ser mejor persona. Mis amigos Camilo, Rafael, Fernando y en especial mi compañero de tesis Kevin Li son personas que siempre estarán presentes en mi vida y los momentos que pasamos juntos serán longevos, les agradezco por hacer que mi vida universitaria haya sido más placentera.

Por último, pero no menos importante a mis profesores, educadores que me han compartido sus conocimientos y me han convertido en un mejor profesional y un agradecimiento en especial a Félix Carrera, que además de haber sido mi maestro y tutor de tesis, fue el responsable que hizo que inclinara y fascinará por el mundo de la ciencia de datos

Zhukov Jafet Paz y Miño Robles

## **DEDICATORIA**

Gracias al sacrificio de mi padre y de mi madre me encuentro donde estoy, por lo que este proyecto y a su vez la finalización de una etapa de mi vida va dedicado a mis padres Walther Zhukov Paz y Miño y Lucrecia Robles, espero esto pueda reflejar lo agradecido que estoy con todo lo que han hecho por mí.

Adicional quiero dedicar esta tesis a mi hermana Fiorella Fiore, a quien amo mucho y a mi difunto hermano Walther Andrés Paz y Miño Robles, nunca pensé en adelantarte en esta carrera llamada vida así que esta tesis va por ti hermano mayor, en tu nombre culmino esta etapa de mi vida, por lo que no pudiste terminar en vida lo hago yo por ti.

Zhukov Jafet Paz y Miño Robles



**UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL  
FACULTAD DE ECONOMÍA Y EMPRESA  
CARRERA DE NEGOCIOS INTERNACIONALES**

**TRIBUNAL DE SUSTENTACIÓN**

f. \_\_\_\_\_

**Ing. Hurtado Cevallos, Gabriela Elizabeth Mgs.**

**DIRECTORA DE CARRERA**

f. \_\_\_\_\_

**(Apellidos, Nombres completos)**

**COORDINADOR DEL ÁREA**

f. \_\_\_\_\_

**(Apellidos, Nombres completos)**

**OPONENTE**



**UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL  
FACULTAD DE ECONOMÍA Y EMPRESA  
CARRERA DE NEGOCIOS INTERNACIONALES**

**CALIFICACIÓN**

---

**Li Ye Kevin Kinjang**

---

**Paz y Miño Robles Zhukov Jafet**

## Índice General

Introducción .....	2
Problemática.....	6
Justificación.....	9
Alcance.....	10
Objetivos .....	11
Objetivos generales .....	11
Objetivos específicos.....	11
Capítulo II .....	12
Marco Teórico .....	12
Sector Florícola .....	12
Series de temporales .....	13
Series temporales univariantes .....	18
Componentes de una serie de tiempo .....	18
Tendencia Secular.....	18
Variación Cíclica .....	19
Variación Estacional .....	19
Movimientos Irregulares.....	19
Tipos de Series de Tiempo .....	19
Aditiva .....	19
Multiplicativa.....	19
Mixta.....	20
Estacionaria .....	21
Serie de tiempo estacionaria .....	21
No Estacionaria .....	21
Media Constante y Varianza Constante .....	22
Media Constante y Varianza Creciente .....	23

Media no Constante y Varianza no Constante .....	23
Media no Constante y Varianza Creciente .....	24
Características de las Series de Tiempo .....	25
Ajustes .....	25
Ajuste por Calendario .....	26
Ajuste Poblacional .....	26
Ajuste de Inflación.....	26
Ajuste Logarítmico o Matemático .....	26
Rezagos o Retardo .....	27
Clasificación .....	27
Estacionarias .....	27
No estacionarias .....	27
Modelo ARIMA .....	27
Modelo autorregresivo (AR).....	31
Modelo de medias móviles (MA) .....	31
Modelo autorregresivo de medias móviles.....	31
Marco Conceptual .....	34
Series de tiempo .....	34
Componentes de una serie de tiempo .....	34
Tendencia.....	34
Estacional.....	34
Ciclo.....	34
Irregular .....	34
ARIMA.....	35
Modelo Autorregresivo (AR).....	35
Modelo Integrado (I).....	35
Modelo Media Móvil (MA).....	35

Modelo Autorregresivo y de Medias Móviles (ARMA) .....	36
Marco Legal .....	37
Capitulo III.....	38
Metodología .....	38
Población y Muestra .....	38
Normalización .....	39
Metodología de Box y Jenkins .....	40
La metodología Box-Jenkins consta de tres etapas principales: .....	40
Identificación y selección del modelo .....	40
Estimación de los parámetros .....	41
Validación del modelo obtenido .....	41
Ruido Blanco .....	43
Proceso Autoregresivo .....	44
Media Móvil .....	45
Estacionariedad .....	46
Estacionalidad .....	47
Raíz Unitaria.....	47
Forecast .....	49
Análisis de resultados.....	51
READXL.....	51
XTS .....	52
DPLYR .....	52
TIDYVERSE.....	54
LUBRIDATE .....	54
TSERIES .....	58
ASTSA .....	58
FORECAST.....	59

FOREIGN.....	59
TIMSAC.....	59
VARs.....	59
MFILTER.....	59
DYNLM.....	60
NLME.....	60
PROPHET.....	61
GRIDEXTRA.....	61
GGPUBR.....	61
GGSTATSPLOT.....	62
Líneas de Códigos.....	63
Conclusión de análisis de datos.....	78
Conclusión.....	80
Recomendaciones.....	81
Referencias.....	83

## Índice tablas

Tabla 1. Coeficientes de TM.....	67
Tabla 2. Datos entrenamiento de TM.....	67
Tabla 3. Ljung-Box Test de TM .....	70
Tabla 4. Coeficientes de FOB .....	73
Tabla 5. Datos entrenamiento de FOB.....	74
Tabla 6. Ljung-Box Test de FOB.....	76

## Índice Figuras

Figura 1. Serie de tiempo estacionaria .....	21
Figura 2. Serie de Tiempo no Estacionaria .....	22
Figura 3. Rendimientos diarios de un activo financiero .....	22
Figura 4. Precio de un activo financiero .....	23
Figura 5. Tasa de desempleo .....	23
Figura 6. PIB de un país en desarrollo .....	24
Figura 7. Metodología de Box Jenkins para el análisis de series de tiempo .....	42
Figura 8. Ruido Blanco .....	44
Figura 9. Distribución de valores mensuales del año 1990 a 2023 de las exportaciones de flores del Ecuador. ....	64
Figura 10. Evolución del peso neto en Toneladas y FOB de las exportaciones de flores del Ecuador del año 1990 a 2023. ....	65
Figura 11. Residuos de ARIMA.....	68
Figura 12. Función de autocorrelación simple de los residuos (ACF).....	69
Figura 13. Histograma de los residuos .....	70
Figura 14. Proyección del peso neto de las exportaciones de flores del Ecuador.....	71
Figura 15. Residuos de ARIMA.....	74
Figura 16. Función de autocorrelación simple de los residuos (ACF).....	75
Figura 17. Histograma de los residuos .....	76
Figura 18. Proyección de FOB de las exportaciones de flores del Ecuador. ....	78

## **Resumen**

El mundo está siempre en constante cambio y esto influye en la forma de vivir el día a día de las personas, los avances tecnológicos de la reciente década han representado un paso significativo en diversos ámbitos en los cuales destacan diversas herramientas y conocimientos como inteligencia de negocios para una mejor toma de decisiones. El presente estudio tiene como objetivo explicar y evidenciar cómo el modelo estadístico autorregresivo integrado de media móvil se utiliza para analizar y pronosticar series temporales. Las variables utilizadas para llevar a cabo el pronóstico fueron peso neto y FOB, de la cual se aplicó el modelo ARIMA (1, 0, 2) (0, 1, 1) para la variable peso neto y el modelo ARIMA (2, 0, 1) (0, 1, 2) para la variable FOB. Para alcanzar el objetivo, se empleó el programa Rstudio, el cual contribuyó a gestionar eficazmente desde la preparación de los datos hasta la ejecución de los análisis estadísticos. Como objeto se fijó las exportaciones del sector florícola del Ecuador en el periodo 1990-2023, siendo las exportaciones de flores un indicador económico relevante para el país. El estudio recopila datos históricos de las exportaciones de flores y realiza un análisis exploratorio para comprender las tendencias y patrones presentes en los datos. Los resultados obtenidos revelaron información valiosa para las empresas que operan en el sector agrícola.

**Palabras clave:** Inteligencia de Negocios, Modelo Autorregresivo Integrado (AR), Modelo de Medias Móviles(MA), Modelo Autorregresivo Integrado de Medias Móviles (ARIMA), Series de Tiempo, R-Studio, Modelos predictivos, Estacionariedad.

## **Abstract**

The world is always changing, and this influences the way people live their daily lives. The technological advances of the recent decade have represented a significant step forward in various fields in which different tools and knowledge stand out, such as business intelligence for better decision making. The present study aims to explain and demonstrate how the autoregressive integrated moving average statistical model is used to analyze and forecast time series. The variables used to carry out the forecast were net weight and FOB, from which the ARIMA (1, 0, 2) (0, 1, 1) model was applied for the net weight variable and the ARIMA (2, 0, 1) (0, 1, 2) model for the FOB variable. To achieve the objective, the Rstudio program was used, which helped to manage efficiently from the preparation of the data to the execution of the statistical analyses. The objective was the exports of the Ecuadorian flower sector for the period 1990-2023 since flower exports are a relevant economic indicator for the country. The study compiles historical data on flower exports and performs an exploratory analysis to understand the trends and patterns present in the data. The results obtained revealed valuable information for companies operating in the agricultural sector.

**Keywords:** Business Intelligence, Integrated Autoregressive Model (AR), Moving Average Model (MA), Autoregressive Integrated Moving Average Model (ARIMA), Time Series, R-Studio, Predictive models, Stationarity.

## Résumé

Les progrès technologiques de la dernière décennie ont représenté une avancée significative dans divers domaines dans lesquels différents outils et connaissances se distinguent, tels que l'intelligence économique pour une meilleure prise de décision. Cette étude vise à expliquer et à démontrer comment le modèle statistique de la moyenne mobile intégrée autorégressive est utilisé pour analyser et prévoir les séries temporelles. Les variables utilisées pour effectuer les prévisions sont le poids net et la FAB, à partir desquelles le modèle ARIMA (1, 0, 2) (0, 1, 1) a été appliqué pour la variable poids net et le modèle ARIMA (2, 0, 1) (0, 1, 2) pour la variable FAB. Pour atteindre l'objectif, le programme Rstudio a été utilisé, ce qui a permis de gérer efficacement la préparation des données jusqu'à l'exécution des analyses statistiques. L'objet de l'étude était les exportations du secteur floricole équatorien pour la période 1990-2023, les exportations de fleurs étant un indicateur économique pertinent pour le pays. L'étude compile des données historiques sur les exportations de fleurs et effectue une analyse exploratoire pour comprendre les tendances et les modèles présents dans les données. Les résultats obtenus révèlent des informations précieuses pour les entreprises opérant dans le secteur agricole.

**Mots-clés :** Intelligence économique, modèle autorégressif intégré (AR), modèle de moyenne mobile (MA), modèle autorégressif intégré de moyenne mobile (ARIMA), séries chronologiques, R-Studio, modèles prédictifs, stationnarité.

## Introducción

El impacto de la globalización a lo largo de los años ha resultado beneficioso para diversos países, pues ésta ha permitido la expansión de la inversión y el comercio internacional en diversas naciones (Parkin et al., 2007, p. 6). La importancia del comercio exterior y del comercio internacional radica en que estos permiten impulsar el crecimiento económico de un país y, a través de un correcto manejo de éste alcanzar el desarrollo económico (Abril et al., 2022, como se citó en Acosta Palomeque, Pazmiño Arroyo, & Cerda Prado, 2018).

Además, la globalización ha transformado la manera en la que las organizaciones empresariales y naciones se desenvuelven y a su vez ha permitido y facilitado el comercio internacional con el uso de las tecnologías de la información como softwares de ciencia de datos dentro de su sistema empresarial u organizacional, como lo es la inteligencia artificial que sigue evolucionando con rapidez y desempeñará un papel esencial en nuestra vida cotidiana (Zhao H.,2018).

La introducción de la Inteligencia de los Negocios en las áreas económica y comercial ha beneficiado a muchos países para poder darle un mejor entendimiento a la situación comercial mediante un análisis. Con el avance tecnológico este se ha ido aplicando para poder manejar los grandes datos. Asimismo, la inteligencia de los negocios tiene un impacto positivo y significativo en el intercambio de conocimientos, la innovación organizacional y la obtención de una ventaja competitiva (Rosa et al., 2016).

La inteligencia de los negocios puede definirse como: El conjunto de habilidades, tecnologías, aplicaciones y buenas prácticas utilizadas para ayudar a una organización determinada a obtener mayor conocimiento de su contexto comercial, imprescindible para tomar mejores decisiones encaminadas a la adquisición de ventajas competitivas (Anabel et al.,2015, como se citó en Rotaeché, 2009; Castaño, 2012 y Pourmojib et al., 2013). También “se entiende como un proceso analítico respaldado por tecnología que reúne y transforma datos fragmentados de empresas y mercados en información o conocimiento sobre objetivos, oportunidades y posiciones de una organización” (Bernhard, 2015, pag.1163).

De igual manera, la inteligencia de los negocios ayuda a que las empresas puedan captar las diferentes oportunidades que se pueden mostrar mediante el uso de esta herramienta, a su vez, mejora la eficiencia operativa y ayuda a aumentar la productividad en las operaciones comerciales. Por lo que, “la inteligencia de negocios actúa como un factor estratégico para una empresa u organización, generando una ventaja competitiva que proporciona información privilegiada para responder a los problemas de negocio” (Murillo y Cáceres, 2013, p. 122)

En Estados Unidos (EE.UU), Canadá y los países que conforman la Unión Europea (EU) han sabido aprovechar el apogeo tanto la inteligencia artificial como la ciencia de datos e inteligencia de negocios, en los diversos sectores económicos y políticos. Estados Unidos cuenta con 40% de empresas de inteligencia artificial además de más de una docena de agencias.

Por consiguiente, las naciones en vía de desarrollo deben de estar a la vanguardia de los avances de las tecnologías de la información como lo son los programas de inteligencia artificial a fin de poder analizar de manera más precisa las distintas situaciones de la vida cotidiana y poder implementar y aplicar estrategias eficientes a partir de los datos en busca de resultados óptimos.

La exportación es una de las actividades económicas más importantes que beneficia a muchos países, igualmente, es también fundamental para muchas empresas que buscan oportunidades de expandir sus negocios y la presencia en el mercado global, la cual aumenta la base de sus clientes y diversifica las fuentes de sus ingresos.

En el Ecuador, se halla una economía diversificada en la cual se encuentran distintos productos que se exportan para el beneficio y el desarrollo económico del país. Los principales productos que son exportados según el Banco Central del Ecuador (2021), el petróleo es uno de los principales productos que se exportan en el Ecuador, por el otro lado, los productos no petroleros también son una fuente importante de ingresos de exportación para el país. Los productos no petroleros que más se destacan en la exportación son el banano, atún, camarón, cacao y flores naturales en la cual estas representan el 74% de las exportaciones no petroleras según el Ministerio de Comercio Exterior (2018).

Ecuador es un país enriquecido en diversos ámbitos, según el Banco Central del Ecuador (BCE), su principal fuente de ingreso es el sector petrolero, durante el año 2022 el valor total de las exportaciones de la nación fue de 32.658 millones USD de los cuales 11.587 corresponden al sector petrolero, no obstante, durante los meses de enero a febrero del 2023 las exportaciones de dicho sector han sufrido un declive del -20.8% en contraste del periodo enero-febrero 2022. Sin embargo, se presenta una variación positiva de 1.9% en las exportaciones no petroleras en los primeros dos meses del 2023 en contraste a las exportaciones de enero a febrero del año 2022, además las exportaciones de flores naturales muestran un incremento del 1% a lo largo del primer bimestre del periodo.

En el ámbito económico global, como señala Toledo (2017) “el rol de las exportaciones para promover el desarrollo económico ha sido objeto de estudio al menos desde los años cincuenta” (p. 79). La participación del país en el comercio internacional es un factor de mucha importancia para la economía ecuatoriana ya que depende en gran medida las exportaciones porque es un pilar fundamental en el crecimiento económico del país para lograr el desarrollo económico y social siendo una de las metas principales de todo país (Chamba et al., 2021). Uno de los períodos claves del Estado Ecuatoriano en la interacción de estos factores para el crecimiento económico fue cuando dinamizó su comercio Internacional con el auge cacaotero permitiendo a la economía expandir sus ingresos y fortaleciendo el sector laboral (Ayala, 2014)

Holanda evidentemente es el principal país productor y exportador de flores, con una producción valorada en 2000 millones de dólares americanos, seguido por Colombia, con valor de exportaciones de 600 millones de dólares americanos. También hay países que sobresalen como productores y exportadores, tal es el caso de Ecuador, México, Brasil, Costa Rica, Kenia, Tailandia, Taiwán, entre otros (van Uffelen y Groot, 2005).

Arbeláez Torres, G (1993) indica que en Colombia a partir del año 1965 el sector de floricultura, con fin de exportación, ha sido de gran crecimiento. Sin embargo, no fue hasta hace décadas luego del 1965 que comenzó a desarrollarse mejor y a tener mayor relevancia en la economía del país en el 1992 ocupando el tercer puesto en exportaciones agropecuarias (Asocolflores,1992). Las especies florales que

destacan en el país son el cultivo de rosa con el 33,5 %, hortensias 20,5 %, crisantemo 12%, clavel 11.6%, Alstroemeria 4,9%, 17,5% otros (Ceniflores,2022). Además, el Centro de Innovación de Floricultura Colombiana informa que “el país exporta a más de 100 países y es el segundo mayor exportador del mundo después de Holanda”. Por otro lado, México destina el 20% de su producción florícola para exportación, siendo la rosa la especie que más destaca de las demás.

La floricultura es una de las actividades más dinámicas y modernas desarrolladas a nivel mundial por su aporte a la generación de empleo, desarrollo de sectores rurales y los ingresos que genera en el comercio exterior (Morocho et al., 2021). El sector florícola en el Ecuador tiene una trayectoria aproximada de 38 años, su primer cultivo de las flores fueron los crisantemos y claveles (Coraima et al., 2021, p. 76). La ubicación geográfica del Ecuador ha sido ideal para el cultivo de flores y esto ha conllevado a un crecimiento significativo en las últimas décadas. De igual forma, el sector florícola es uno de los más fundamentales del país, puesto que “en el mundo, las flores ecuatorianas son conocidas como una de las mejores, por ser un producto que cumple los estándares exigidos de calidad en el exterior.” (Coraima et al., 2021, p. 80). La industria florícola del Ecuador se enfoca en la producción de flores de alta calidad y con gran variedad de tamaños y colores para poder competir en los mercados internacionales.

El sector florícola del Ecuador es uno de los principales pilares económicos del país y forma parte de una importante actividad comercial y económica en la balanza comercial. A lo largo del año 2022 dicho sector tuvo una participación en las exportaciones no petroleras del 4.5% (Banco Central del Ecuador, 2023). De acuerdo con Coraima et al (2021) “Existen variedades de flores ecuatorianas; en cuales las principales flores exportadas en el mundo son las siguientes, rosas, claveles, gypsophilas, claveles y alstroemerias; debido que son consideradas como un producto de calidad en los mercados internacionales por ser hermosas y mejores en el mundo.” (pág. 76). Mientras que los principales destinos de exportaciones de flores ecuatorianas son los Estados Unidos, Kazajistán, Rusia, Canadá, Unión Europea y entre otros.

## **Problemática**

Cuando se habla de América Latina en el campo de la tecnología siempre se le conoce como una parte del continente que menos desarrollo tiene ya sea porque existe una diferencia en el acceso y el uso de la tecnología, tanto si, las comunidades rurales y lugares de bajos recursos carecen de internet y de infraestructura adecuada para utilizar las oportunidades tecnológicas (Olarte, 2013). Por ello, la parte sur de América no tiene altas ventajas competitivas en ciertos sectores empresariales como en países más desarrollados.

Además de ello, está marcada la desigualdad económica ya que forma parte de un obstáculo fundamental en el desarrollo tecnológico y educacional. Ya sea por la dificultad en el acceso a la educación de calidad ya que para muchos países de América Latina no se encuentran capacitados para poder preparar a las personas con habilidades tecnológicas, estadísticas y matemáticas, por lo tanto, existe una capacitación deficiente la cual permite que se pueda manejar la metodología adecuada en la estadística, además, existe una carencia en las oportunidades en los campos laborales que estén relacionadas altamente profesional en el campo de la estadística en los países latinoamericanos y en Ecuador y esto conlleva a que los países latinoamericano no tengan capital humano suficiente con experiencia en ese campo laboral.

La ciencia, tecnología e innovación es uno de los retos que tienen mayormente los países latinoamericanos para poder estar en un entorno cada vez más globalizado y competitivo (Barragán, 2013). Según Olarte (2013) América Latina es un lugar geopolíticamente contingente y se le ha conocido como un lugar premoderno, del tercer mundo y subdesarrollado, por lo tanto, siempre se le ubica en un lugar de subordinación frente a otros mundos. Por ello, es un tema que se debe priorizar en los países latinoamericanos para poder estar a la par en el desarrollo tecnológico e innovación con los países desarrollados.

En el campo científico los países latinoamericanos y el caribe han tenido varios avances en los últimos 20 años, sin embargo, existe poca inversión en investigación, desarrollo e innovación, frente a los países desarrollados (León et al., 2020). Debido a la enorme cantidad de datos generados por la sociedad hoy en día, conceptos como Big Data, minería de datos, analíticas de datos y open data se abren paso herramientas que

permiten su tratamiento (Lasso, 2022). Sin embargo, estos tipos de conceptos han llegado a ser utilizados de manera ineficaz ya que América Latina no lo han utilizado de la manera adecuada por la falta de conocimiento sobre la importancia de las estadísticas (Veloso et al., 2022). Y por esta comprensión limitada de los conceptos de estadísticas, al ser una disciplina con un amplios conceptos y técnicas, pueden llegar a ser un factor clave en la cual las empresas no tienen mucha confiabilidad en aplicar profundamente estos tipos de herramientas estadísticas ya que pueden llegar a ser ineficaz y nada productivo.

De acuerdo con Escobar Sailema (2019) las empresas ecuatorianas no entienden la importancia de la implementación de distintos modelos estadísticos, lo que conlleva a una aplicación errónea de los distintos métodos de estimación, que ha sido el motivo principal de las malas decisiones empresariales además de una mala planificación de la producción que se refleja en una rentabilidad baja.

Por lo que, las empresas limitan el presupuesto lo cual afecta la capacidad de llevar a cabo la utilización de esas herramientas. Por consiguiente, esto llega a afectar en varios ámbitos de la investigación estadística, uno de los problemas que enfrentan estos países es la falta de calidad y confiabilidad de los datos, ya que la deficiencia en la recolección y procesamiento de datos deterioran el proceso y perjudican en el modelo estadístico que se va a implementar y llega a ser bastante ineficiente.

Según Jacho Guashca (2020) el estudio y aplicación de la estadística permite obtener resultados confiables y de esa manera tomar decisiones eficientes. Además, se presentan conceptos que ayudan a dar un mejor entendimiento del contenido y las pautas necesarias para el seguimiento de la investigación, que brinde resultados eficientes y confiables en la actualidad.

Ecuador cuenta con problemas de falta de información actualizada en las distintas bases de datos del país como el INEC, SUPERCIAS, SRI, e incluso las bases de datos públicas no se encuentran ni estandarizadas ni interconectadas, además de tener un sistema altamente burocrático, este suceso no permite obtener resultados confiables y causa un retroceso en la obtención de datos que se refleja en modelos estadísticos incompletos y poco fiables y por lo tanto en malas decisiones en la aplicación de diversas estrategias.

Por lo tanto, los problemas actuales en la estadística en América Latina, influye negativamente en la insuficiencia de recopilación de datos para el sector de exportación de flores, ya que estos datos deben ser precisos y actualizados. La falta de importancia al sector estadístico es un error que persiste hasta el día de hoy en el Ecuador. Es por ello por lo que se observa ver distintas empresas realizar malas decisiones y estrategias, puesto que una buena aplicación de la estadística, manejada por un profesional, permitiría a organizaciones, instituciones e incluso gobiernos poder tener un mejor criterio frente a situaciones en las que se deben de analizar incontables datos y por consiguiente aplicar las bases de la estadística para poder detectar problemas a futuro, mediante predicciones, regresiones etc. Asimismo, es relevante resaltar que, en los países latinoamericanos, el costo de implementación de estas nuevas tecnologías estadísticas no llega a ser bastante accesible para muchas empresas, sin embargo, muy pocas empresas que están más desarrolladas están empezando a aplicar estas herramientas y esto por consiguiente crea un desbalance en la competencia en los negocios con la implementación de estas herramientas.

El impulso fuerte en el uso de computadoras en el INEC empezó débilmente en los años 90 (Villacis, 2021). Por lo que podemos comprender que el Ecuador ha empezado a usar las estadísticas recientemente en ese año, y esto puede deber a una falta de gestión en las instituciones de estadísticas y gubernamentales del Ecuador. Además de ello, es fundamental recalcar que en el Ecuador se evidencia una carencia en la aplicación de conocimientos estadísticos en la exportación de flores.

Asimismo, en el Ecuador, aunque se recopilen datos y realicen análisis estadísticos, esto puede haber problemas por la falta de conocimiento para utilizar los datos en las tomas de decisiones estratégicas empresariales en las competencias de las exportaciones de flores. Esto puede ser causado por problemas en el acceso a información comparativa de otros mercados de exportación de flores y esto puede afectar a diversas empresas para poder determinar mediante modelos estadísticos la estrategia ideal para poder tomar la adecuada decisión.

## **Justificación**

La exportación en el Ecuador es una de las actividades comerciales que más aportan al desarrollo económico del país, siendo los productos no petroleros uno de los más importantes en la comercialización internacional. Puesto que, ayuda al país generando nuevos empleos y mejorando el crecimiento de la economía. Además de ello, es de recalcar que las flores naturales del Ecuador son una de las más reconocidas a nivel mundial por la calidad y la variedad que ofrecen. Por lo tanto, es importante que el sector florícola del Ecuador deba ser estudiada y analizada para identificar las oportunidades de crecimiento en lo que respecta a la exportación.

La predicción de las exportaciones de flores es primordial para poder establecer y administrar de manera apropiada la producción y comercialización de las flores ecuatorianas. Una predicción adecuada favorece a los productores y exportadores a tomar una apropiada decisión respecto a las cantidades convenientes en la comercialización internacional. Lo cual mejora significativamente la producción y productividad en el sector florícola.

Además de ello, es importante que la predicción de las exportaciones de flores se realice con un modelo adecuado en el proceso de pronóstico y debe ser apropiado para el objetivo particular del usuario del pronóstico (Witt et al., 2003; Danese y Kalcsmidt, 2011). Por lo tanto, para esta investigación el modelo ARIMA es el modelo adecuado para estos datos, debido a que el modelo en general proporciona proyecciones más precisas y es más confiable que otras técnicas estadísticas comunes (Lai y Dzombak, 2020).

Por consiguiente, este estudio tiene como objetivo contribuir, mediante un correcto análisis predictivo del modelo ARIMA al sector florícola, proporcionando resultados pronosticados y evaluados los cuáles permitirán a los productores de poder brindar una gestión adecuada y tomar una mejor decisión para la futura producción de flores naturales ecuatoriana, lo que puede fortalecer la productividad y la eficiencia de las empresas.

## **Alcance**

En la actualidad, las organizaciones públicas del Ecuador poseen grandes dimensiones de datos adquiridos por diferentes empresas en el sector florícola. Los datos obtenidos para la presente investigación se recopilaron del Banco Central del Ecuador, se usará una variable la cual es la cantidad en miles de millones de dólares de la exportación de flores desde el año 1990 hasta marzo del año 2023. Por consiguiente, los datos serán procesados mediante software de la Inteligencia de los Negocios “R Studio”, en la que se facilitará considerablemente el proceso para realizar el análisis predictivo con el uso de la tecnología en los negocios. El análisis del modelo ARIMA se centra solamente en la predicción de valores futuras que en esta ocasión es la variable de flores exportadas, por lo que, no se utilizará variables externas que afecten al resultado del pronóstico. Finalmente, este estudio está dirigida principalmente para las empresas del sector florícola del Ecuador, con la finalidad de dar a conocer esta herramienta de la inteligencia de los negocios y mostrar la importancia que esta trae para poder administrar y analizar de una manera adecuada a la producción de las flores naturales.

## **Objetivos**

### **Objetivos generales**

Analizar mediante el modelo ARIMA la predicción de la exportación de flores del Ecuador en el periodo 1990 - 2023.

### **Objetivos específicos**

- Examinar el marco teórico y sus conceptos mediante la revisión literaria acerca del tema de investigación.
- Comprender la metodología del modelo autorregresivo integrado de media móvil para la aplicación de un modelo predictivo.
- Analizar y describir los resultados del modelo de predicción ARIMA

## **Capítulo II**

### **Marco Teórico**

#### **Sector Florícola**

La industria de exportación de flores en Ecuador desempeña un papel importante en la economía del país. Ecuador es conocido por exportar diversos productos, como petróleo, plátanos, camarones y flores (Veuthey & Gerber, 2012). Las flores son uno de los productos de exportación dominantes en términos de valor (Veuthey & Gerber, 2012). Además, ha contribuido a diversificar la economía de Ecuador. Mientras que el país es principalmente agrario y exporta productos como plátanos y camarones, la industria de las flores ha proporcionado una fuente alternativa de ingresos y de ingresos de exportación (González-Andrade et al., 2010). Esta diversificación ayuda a reducir la dependencia del país de un único producto básico y contribuye a la estabilidad económica general (Castillo, 2017). El país se ha convertido en el tercer exportador mundial de flores (Prado & Vanel, 2020). La mayoría de las flores que se consumen en Norteamérica se importan de países como Colombia y Ecuador (Chandler & Sánchez, 2012). La industria florícola ecuatoriana se caracteriza por ciclos de vida cortos de los productos, una gran variedad de productos, una demanda volátil y cambiante, y procesos de entrega largos e inflexibles (Mañay et al., 2022).

El crecimiento de la industria de exportación de flores en Ecuador ha tenido impactos positivos. La industria ha contribuido al crecimiento económico del país y ha proporcionado una importante fuente de ingresos y empleo (Mañay et al., 2022). La industria de exportación de flores también se ha considerado una explotación sostenible de los recursos naturales de Ecuador (Guaita-Pradas et al., 2023).

La industria de exportación de flores de Ecuador se enfrenta a la competencia de otros países, en particular Colombia y Kenia. Colombia es uno de los principales exportadores de flores cortadas a Norteamérica (Chandler y Sánchez, 2012). Kenia también es un factor importante en el mercado mundial de la flor cortada, ya que una gran parte de las flores cortadas que se venden en Alemania se importan de Kenia (Rombach et al., 2021). La competencia de estos países pone de manifiesto la necesidad de que la industria floral ecuatoriana mantenga su posición competitiva en el mercado mundial (Guaita-Pradas et al., 2023). Por lo que, esta competencia en el

mercado obliga a crear estrategias para mejorar en cuanto la toma de decisiones, ventaja competitiva, optimización de recursos y entre otros. Y una de las estrategias es el uso de predicciones en el mercado de flores en el Ecuador. Para comprender y predecir las tendencias y patrones de la exportación de flores en Ecuador, es crucial utilizar el análisis de series temporales. Los datos de series temporales brindan una perspectiva útil de la dinámica de la industria, lo que permite identificar los factores y relaciones clave que influyen en el rendimiento de las exportaciones.

La estadística es esencial para analizar y comprender los datos de series temporales. Ofrece un marco para comprender las relaciones, los patrones y las tendencias que se encuentran en los datos. La estadística se ha utilizado en diversas tareas en los análisis temporales, como identificar la tendencia, seleccionar modelos, comprobar hipótesis y evaluar la calidad.

Una aplicación común de la estadística en el análisis de series temporales es en la detección de tendencias, esto es importante para comprender el comportamiento a largo plazo y hacer pronósticos, además se han utilizado varios métodos estadísticos como el análisis de regresión lineal para identificar las tendencias en los datos de la serie en estudio (Forkel et al, 2013). Asimismo, la comprobación de hipótesis es otra aplicación de la estadística en el análisis de la misma, estas pruebas permiten detectar cambios o impactos significativos en los datos de la serie (Scharwachter y Muller, 2020). La estadística también se utiliza para evaluar la calidad de los métodos de análisis de series temporales. Diversas medidas estadísticas, como el error cuadrático medio (MSE), el error cuadrático medio radical (RMSE) y los coeficientes de correlación, se utilizan para evaluar la precisión y el rendimiento de los modelos de previsión (Y. et al., 2021). La previsión es una tarea fundamental en el análisis de este tipo de series y la estadística ofrece las herramientas y técnicas para realizar predicciones precisas. Los modelos estadísticos, como los modelos de medias móviles integradas autorregresiva (ARIMA) y las redes neuronales han sido utilizado ampliamente para los pronósticos en las series temporales (Chen y Liu, 2020).

### **Series de temporales**

El análisis de series temporales se remonta a los inicios de la investigación científica. Los científicos llevan mucho tiempo recopilando y analizando conjuntos de datos para comprender mejor la dinámica de diversos sistemas (Fulcher et al. 2013).

El análisis de éstas se ha utilizado ampliamente para estimar tendencias futuras y explicar patrones históricos en diversos sectores. Puede pensarse que tuvo inicios en 1664 cuando Sir Isaac Newton descompuso una señal luminosa en sus componentes a diferentes frecuencias haciendo pasar la señal por un prisma de vidrio. Una imagen multicolor fue lanzada sobre una pared opuesta. El llamó a esto el espectro, y corresponde a la operación de filtrado. Newton no llevó a cabo un análisis cuantitativo de su serie de tiempo (Abril J., 2011).

Según Abril (2011) Sir W. Herschel realizó un análisis cuantitativo usando termómetros en 1800. Éste midió la energía promedio en varias bandas de frecuencia del espectro de la luz solar emplazando termómetros a lo largo de ese espectro.

“Los fundamentos matemáticos del análisis de las series de tiempo se empezaron a dar a mediados del siglo XIX cuando Gouy representó a la luz blanca como una serie de Fourier. Posteriormente Lord Rayleigh reemplazó la serie por una integral. En 1881 S. P. Langley refinó el experimento de Herschel considerablemente al medir la energía de la luz con un bolómetro espectral (un aparato que él inventó y que hace uso de la corriente eléctrica generada en un cable por radiación incidental)”. Un avance sustancial en el análisis de la serie en estudio correspondiente a las señales luminosas ocurrió en 1891 cuando A. A. Michelson inventó el interferómetro. Este aparato permitía la medición del valor promedio.

Para valores grandes de  $n$  y valores determinados del rezago (lag)  $u$ , donde  $n$  es la cantidad de observaciones disponibles de la señal luminosa bajo estudio  $y(t)$ , la cual se la considera como una función continua del tiempo  $t$ . Esto permitía la estimación de la función de auto covarianzas de la señal.

Los investigadores también han desarrollado métodos para el descubrimiento de patrones en datos de series temporales. Estos métodos implican emparejar subsecuencias de una serie con otra y analizar los patrones estadísticos de los resultados (zhu et al., 2018). Los análisis temporales han evolucionado para abordar los desafíos y las demandas al analizar y predecir patrones en datos dependientes del tiempo. La compresión y aplicación, del análisis de este tipo de series, en varios dominios han mejorado significativamente gracias al desarrollo de algoritmos y técnicas avanzadas.

Las series de tiempo se aplican en el ámbito profesional en todas sus extensiones ya que se han realizado varios métodos para tener mejor precisión en

cuanto al modelo de predicción. Uno de los objetivos principales es desarrollar estadísticamente y explicar el comportamiento de las variables a través de los datos históricos (Correa, 2004). Son una herramienta esencial para poder planificar datos futuros ya que es el conocimiento de una variable a través del tiempo para, a partir de este conocimiento, poder realizar predicciones (Napoleón, 2013). Los datos de una serie se caracterizan y distinguen de otros datos puesto que, como ya mencionado con anterioridad, el análisis del mismo permite observar cómo las variables cambian en el tiempo, por lo tanto se determina que el factor tiempo es crucial, en otros términos, el tiempo es una variable esencial porque revela cómo se ajustan los datos en los puntos de los datos, así como los resultados finales (Tableau, 2023).

De acuerdo a C.W.J. Granger y Paul (1986) denotan en su libro *Forecasting Economic Time Serie* señala que en diversos ámbitos profesionales se observa la aplicación de esta técnica de predicción como por ejemplo, en el sector económico con datos mensuales de empleo; en sociología y sus reportes de cifras de criminalidad; en el campo de la meteorología se distingue los reportes de precipitaciones, temperatura, velocidad del viento; en el ámbito profesional de la medicina observamos la aplicación de esta herramienta en electrocardiogramas y electroencefalogramas; en sistemas físicos vibratorios como la subida de un coche por la superficie ; además están presente también en las ramas de la sismología, oceanografía y geomorfología y cabe destacar también el área de la astrología y los estudios de brillo estelar y actividad solar. El análisis de dicha serie se utiliza a menudo en el ámbito financiero para predecir los precios de las acciones. Pandey (2021) destacó la importancia de su análisis para comprender las tendencias de los precios de las acciones y elaborar previsiones basadas en datos históricos. Además, en otros campos como la medicina, los estudios de series temporales han sido cruciales para evaluar los peligros para la salud relacionados con la exposición a corto plazo a las partículas (PM) y a la contaminación atmosférica, además estos estudios tienen una larga historia en los campos de la epidemiología de las PM y la aplicación de análisis de series temporales y han conducido al establecimiento de normas reguladoras (Bell et al., 2004). Por otra parte, también existe análisis en el campo de la evaluación del comportamiento sísmico, tras un seísmo importante, se necesitan series temporales de movimientos del terreno en lugares no instrumentados para llevar a cabo un análisis no lineal de la historia de respuesta específica del lugar (Tamhidi et al., 2022). El análisis de series temporales

también se ha utilizado para anticipar la demanda de agua. Alnsour y Ijam (2023) desarrollaron un modelo de previsión de la demanda de agua en cascada basado en el análisis de series temporales. Basándose en los factores de la serie temporal, utilizaron datos anteriores para controlar los cambios y anticipar la futura demanda de agua.

Además, Villavicencio (2010) nos indica otras áreas y subáreas en la que resalta la implementación de esta herramienta.

### **Economía y Marketing**

- Pronóstico del empleo y desempleo.
- Evolución del índice de precios de algún producto en específico.
- Beneficios netos mensuales de cierta entidad bancaria.
- Índices del precio del petróleo, por ejemplo

### **Demografía**

- Número de habitantes por año.
- Tasa de mortalidad infantil por año.

### **Medioambiente**

- Evolución horaria de niveles de óxido de azufre y de niveles de óxido de nitrógeno en una ciudad durante una serie de años.
- Lluvia recogida diariamente en una localidad.
- Temperatura media mensual.
- Medición diaria del contenido en residuos tóxicos en un río.

Por consiguiente en base a las diversas definiciones recopilada por diversos científicos, se define a una serie temporal,  $(\{y_t\}, t=1, 2, \dots, n)$ , como el conjunto de observaciones o medidas realizadas secuencialmente en intervalos predeterminados y de igual, o aproximadamente igual, duración. La característica específica de una serie temporal es, por tanto, que las observaciones están ordenadas en el tiempo (Beltrán Barco, 2003). Las series de tiempo sirven para explicar el valor que toma en un momento de tiempo determinado la variable analizada (Sarmiento et al., 2013). El enfoque Box- Jenkins, considera que la serie de tiempo es una realización de un proceso estocástico (Ramírez, 2021). Las series de tiempo por sí solas se reducen a una simple organización temporal de muestras que permiten la descripción de algunos parámetros básicos de un proceso estocástico (Sarmiento et al., 2013).

El conjunto de observaciones de una serie temporal puede ser recopilada en periodos de tiempo tales como: segundo, minutos, horas, días, semanas, meses, bimestres, trimestres, semestres, o en el transcurso de un año, preferiblemente en intervalos equivalentes. A pesar de que la fluctuación de la serie puede ser observada de manera gráfica, distinguir las individualidades cada una puede contener de la misma puede llegar a ser enrevesada. Los procedimientos de muestreo y recolección de datos son críticos para determinar la calidad de un estudio y la susceptibilidad de los resultados (Gibbs et. al, 2007), una cantidad extensa de datos permite asegurar la consistencia y la fiabilidad de la serie temporal. Según Parra (2020) la calidad predictiva dependerá del tipo de serie con la que el analista trabaje, es decir si ésta sigue algún tipo de patrón de comportamiento con efectos fijos medios (conocida como serie determinista) ciertamente se obtendrá predicciones no tan confiables con una grado de error bajo, No obstante, si la serie no presenta ningún patrón de comportamiento específico (denominada serie aleatoria, no determinista o estocástica), seguramente nuestras predicciones carecerán de validez por completo.(p.29)

Box G. et. al. (2016) señala en su obra *Time Series Analysis: Forecasting and Control*, qué existen dos modelos matemáticos como ya mencionados con anterioridad, uno es el determinista y el otro lo denomina modelo estocástico. Se afirma que el modelo determinista es aquel que sigue un patrón de comportamiento y se fijan condiciones para que al ejecutar el experimento se determine el resultado. Es posible predecir con mayor fiabilidad gracias que se puede derivar un modelo matemático basado en leyes de física lo cual hace posible determinar el valor de una cantidad dependiente del tiempo para casi exactamente cualquier momento en el tiempo. (Box G. et al., 2016, p.6)

### **Proceso Estocástico**

Es una sucesión de variables aleatorias  $\{y_t\}$ , donde  $t = -\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty$  o  $\{Y_t\} = -\infty$ , un proceso estocástico es un conjunto de variables aleatorias correspondientes a distintos intervalos de tiempo. Consecuentemente, una serie temporal puede considerarse como una realización muestral de la  $n$  variables aleatorias que forman su proceso estocástico generador o la función de distribución conjunta que la genera (Bustamante R., 2014).

Se hace referencia a los modelos estocásticos en términos de probabilidad debido a que existen variables aleatorias que no siguen un patrón y resulta imposible de predecir con exactitud, por lo tanto, los datos históricos no conceden formular un modelo para determinar un resultado preciso del experimento en cuestión Mendoza (2002). Sin embargo, es posible derivar un modelo para calcular la probabilidad de que un valor futuro se sitúe dentro de dos límites específicos Box G, et al. (2016).

### **Series temporales univariantes**

Una serie univariante es aquella que analiza el comportamiento de la tendencia y la variabilidad de la variable observada en relación al tiempo. La notación matemática sería;  $y_1, y_2, \dots, y_N$ ;  $(y_t)_{t=1}^N$ ;  $(y_t : t = 1, \dots, N)$ , donde  $y_t$  es la observación n°  $t$  ( $1 \leq t \leq N$ ) de la serie y  $N$  es el número de observaciones de que consta la serie completa (el tamaño o la longitud de la serie). Las  $N$  observaciones  $y_1, y_2, \dots, y_N$  pueden recogerse en un vector columna  $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$  de orden  $N \times 1$ .

### **Componentes de una serie de tiempo**

El objeto del análisis de una serie temporal es el conocimiento de un patrón de comportamiento para prever una acción futura ya que son un conjunto de observaciones ordenadas en el tiempo que muestra la evolución de un fenómeno o variable a lo largo de él. Lo habitual es identificar cuatro componentes que son: tendencia, estacionalidad, ciclos y residuos (Nicasio, 2018).

#### **Tendencia Secular**

Se denomina el principal componente de la serie de tiempo, la misma determina el crecimiento o decrecimiento de la serie a largo plazo. Este es el tipo de tendencia que perdura durante un periodo muy largo. Los datos sobre precios, exportaciones e importaciones, por ejemplo, reflejan tendencias obviamente crecientes a lo largo del tiempo (Scribd, 2017).

Para dar con el hallazgo de esta, es crucial disponer de una serie de ciclos completos además de ser extensa, para que ésta no se vea influida por la fase del ciclo en que finaliza la serie, por lo que, a veces, resulta difícil separar ambos componentes (Parra, 2023).

### **Variación Cíclica**

Son oscilaciones alrededor de la tendencia producidas por periodos alternativos de prosperidad y depresión (Hernández, 2015, p.5). Este componente se presenta en series que son afectadas por fenómenos físicos o económicos que ocurren con una periodicidad variable (Hernández et al., 2008).

### **Variación Estacional**

Es lo opuesto a la tendencia y a la fluctuación cíclica, la variación estacional o la estacionalidad nos indica un comportamiento periódico y predecible, describe acontecimientos que se producen a intervalos específicos y regulares en el transcurso de un año y se repite año tras año. Esto se produce cuando los puntos de datos cercanos en el tiempo tienden a estar relacionados. (Tableau, 2023)

### **Movimientos Irregulares**

Se denomina movimientos irregulares a aquellos que presentan oscilaciones en la serie formada por hechos u acontecimientos que individualmente no son importante pero que al combinarse pueden ser significativos, por lo tanto, estas no siguen ningún patrón y resultan impredecibles estos por ejemplo son catástrofes, pandemias etc (Hanke & Wichern, 2010, pág. 166).

### **Tipos de Series de Tiempo**

#### **Aditiva**

Se estructura gracias a la suma de la tendencia, estacionalidad, variación cíclica regular, variación cíclica irregular, ruido. y se expresa de la siguiente manera:

$$x_t = T_t + E_t + C_t + R_t$$

#### **Multiplicativa**

Se compone mediante la multiplicación de la tendencia, estacionalidad, variación cíclica regular, variación cíclica irregular, ruido. y se expresa de la siguiente manera:

$$x_t = T_t * E_t * C_t * R_t$$

### **Mixta**

Se componen sumando y multiplicando la Tendencia, estacionalidad, variación cíclica regular, variación cíclica irregular, ruido.

$$x_t = T_t * E_t * C_t + R_t$$

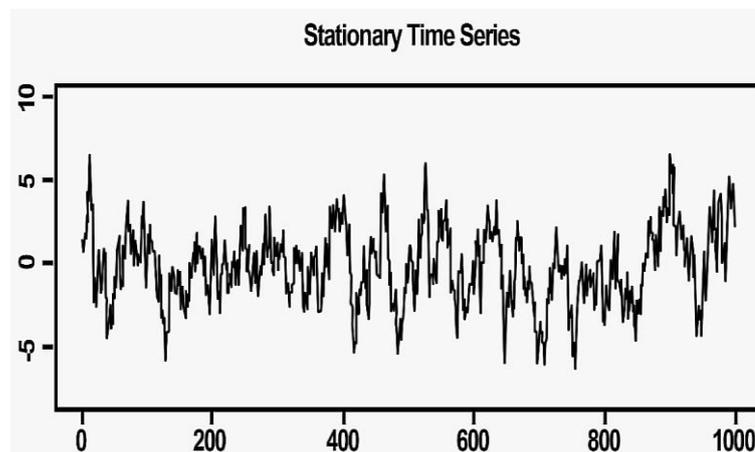
## Estacionaria

La media y la varianza de una serie estacionaria no cambian con el tiempo, y tampoco siguen una tendencia.

- Media constante
- Varianza constante
- La función de autocorrelación disminuye cuando aumenta  $k$

**Figura 1.**

### Serie de tiempo estacionaria



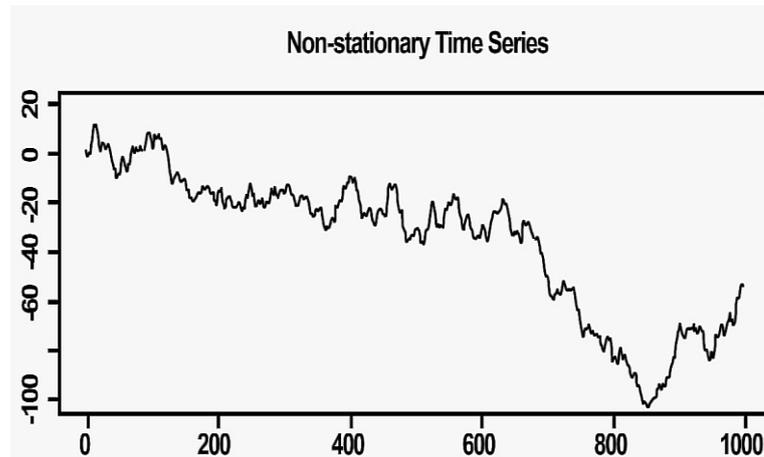
*Nota:* Adaptada de Serie de tiempo estacionaria, por O’reilly, 2023, O’reilly, (<https://www.oreilly.com/library/view/hands-on-machine-learning/9781788992282/15c9cc40-bea2-4b75-902f-2e9739fec4ae.xhtml>).

## No Estacionaria

- La media no es única
- Varía durante el período analizado.
- La variancia no es constante
- La función de autocorrelación decae lentamente en el tiempo.

**Figura 2.**

**Serie de Tiempo no Estacionaria**

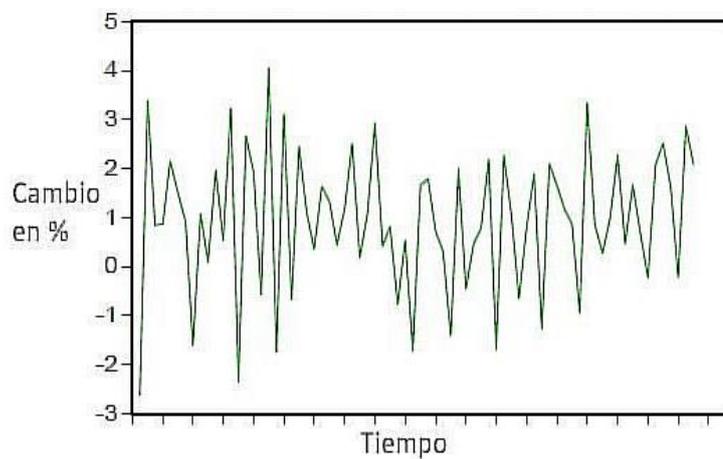


*Nota:* Adaptada de Serie de tiempo no estacionaria, de O’reilly, 2023, O’reilly, (<https://www.oreilly.com/library/view/hands-on-machine-learning/9781788992282/15c9cc40-bea2-4b75-902f-2e9739fec4ae.xhtml>)

**Media Constante y Varianza Constante**

**Figura 3.**

**Rendimientos diarios de un activo financiero**

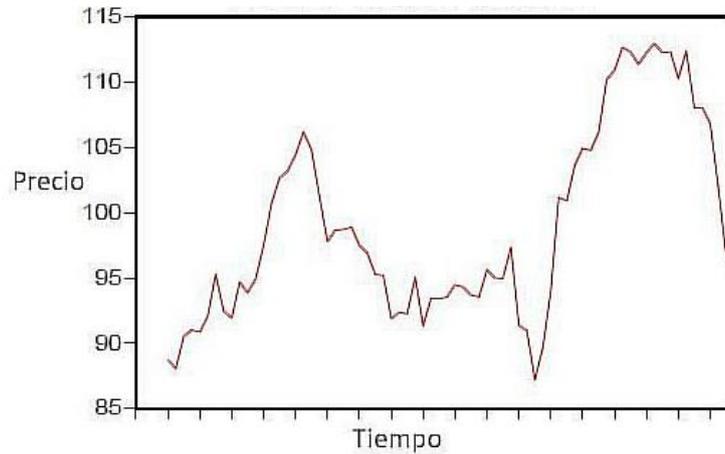


*Nota.* El gráfico anterior representa el cambio porcentual diario de un activo financiero sin mucha volatilidad Adaptado de Tipos de series temporales por José Francisco López, 2018, Economipedia (<https://economipedia.com/definiciones/tipos-de-series-temporales.html>).

## Media Constante y Varianza Creciente

Figura 4.

### Precio de un activo financiero

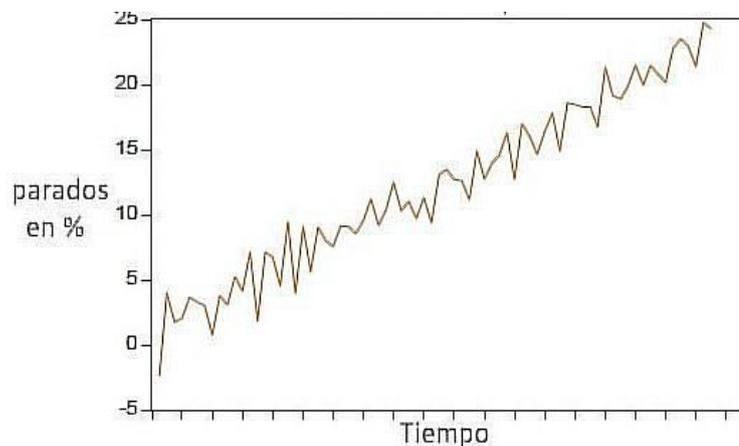


*Nota.* Las acciones de algunas empresas que presentan volatilidad, las materias primas o incluso el mercado de divisas (FOREX) presentan este tipo de comportamiento. Adaptado de Tipos de series temporales por José Francisco López, 2018, Economipedia (<https://economipedia.com/definiciones/tipos-de-series-temporales.html>).

## Media no Constante y Varianza no Constante

Figura 5.

### Tasa de desempleo

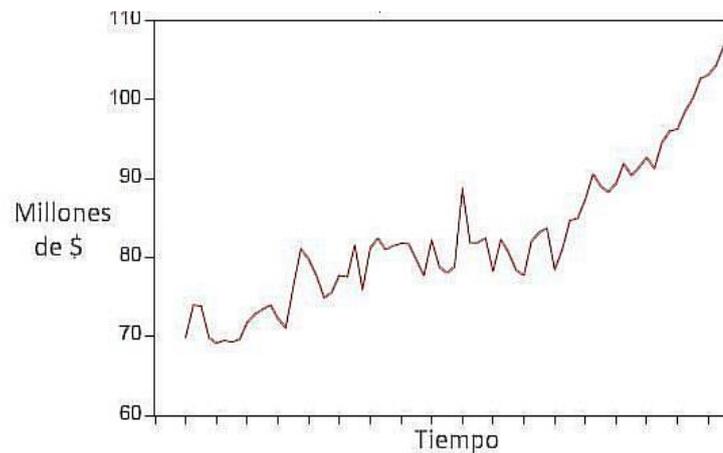


*Nota.* Se observa la evolución de la tasa de desempleo en una época de fuerte recesión económica que dura varios años. La varianza es constante o acotada, es decir, se mantiene en niveles muy parecidos en todo el periodo pero la media se mantiene creciente. Adaptado de Tipos de series temporales por José Francisco Lopéz, 2018, Economipedia, (<https://economipedia.com/definiciones/tipos-de-series-temporales.html>).

### Media no Constante y Varianza Creciente

**Figura 6.**

#### PIB de un país en desarrollo



*Nota.* El PIB de un país es un indicador económico que refleja el valor monetario de todos los bienes y servicios finales producidos por un territorio en un determinado periodo de tiempo. Se utiliza para medir la riqueza que genera un país (Morena y Sevilla, 2021).. Adaptado de Tipos de series temporales por José Francisco Lopéz, 2018, Economipedia (<https://economipedia.com/definiciones/tipos-de-series-temporales.html>).

Los modelos de pronósticos más usados que consideran la relación del comportamiento de variables en el tiempo se destacan tres: i) los de regresión lineal simple y múltiple ii) las series de tiempos y la técnica de Box y Jenkins (1970). (Amaris et al., 2017)

Tableau (2023) indica que los modelos de análisis de series temporales incluyen:

*Clasificación:* Identifica y asigna categorías a los datos.

*Ajuste de curvas:* Traza los datos a lo largo de una curva para estudiar las relaciones de las variables dentro de los datos.

*Análisis descriptivo:* Identifica patrones en los datos de la serie como tendencias, ciclos o variaciones estacionales.

*Análisis explicativo:* Intenta comprender los datos y sus relaciones, así como la relación causa-efecto.

*Análisis exploratorio:* destaca las principales características de los datos de las series temporales, normalmente en formato visual.

*Predicción:* Predice datos futuros éste se basa en tendencias históricas, utiliza los datos históricos como modelo para los datos futuros, prediciendo los escenarios que podrían darse a lo largo de los futuros puntos de la trama.

*Análisis de intervención:* Estudia cómo un acontecimiento puede cambiar los datos.

*Segmentación:* Divide los datos en segmentos para mostrar las propiedades subyacentes de la información de origen.

### **Características de las Series de Tiempo**

- La mayoría de las series tienen una tendencia. Su valor medio cambia con el tiempo, estas se denominan series no estacionarias.
- Algunas describen meandros, es decir, suben y bajan sin ninguna tendencia obvia o tendencia a revertir hacia algún punto.
- En ciertas se presentan Shocks persistentes por lo que las alteraciones repentinas en estas series tardan mucho tiempo en desaparecer.
- Algunas series se mueven conjuntamente, es decir tienen co- movimientos positivos, como por ejemplo las diferentes tasas de interés.
- La volatilidad de algunas series varía en el tiempo y esto produce que puedan ser más variables en un año que en otro.

### **Ajustes**

Los ajustes de una serie temporal se utilizan para poder interpretar de manera más sencilla la serie y así mismo modelar de una manera menos compleja además se enfoca en simplificar los patrones de datos históricos eliminando fuentes de variación

conocidas Tejado(201. Existen cuatro tipos de ajustes, por calendario, por poblaciones, por inflación y por transformaciones matemáticas es decir logaritmos Tajada (2022). Este proceso tiene en cuenta la estructura interna, como la autocorrelación, la tendencia o las variaciones estacionales.

### **Ajuste por Calendario**

Parte de la variación observada en los datos estacionales puede deberse a efectos simples del calendario. Es mucho más fácil eliminar la variación antes de hacer cualquier análisis posterior.

### **Ajuste Poblacional**

Cualquier serie de tiempo que se vea afectada por cambios en la población se puede ajustar para obtener datos per cápita, esto es:

$$Y_t/Pob_t$$

### **Ajuste de Inflación**

Cuando los datos se ven afectados por el valor del dinero, es mejor ajustarlos antes de modelar e interpretar. Si  $P_t$  es el índice de precios e  $Y_t$  es el precio de la variable observada, entonces  $(Y_t)^A = (Y_t/P_t) \times 100$  es el precio de la variable observado a valores en pesos del año base.

### **Ajuste Logarítmico o Matemático**

Si los datos muestran una variación que aumenta o disminuye con el nivel de la serie, entonces una transformación puede ser útil.

Por ejemplo, una transformación logarítmica:  $y_t = \log Y_t$  o  $y_t = \ln Y_t$  Los logaritmos son útiles porque son interpretables:

- Los cambios en un valor logarítmico son cambios relativos (o porcentuales) en la escala original.
- Si se usa la base logarítmica 10, entonces un aumento de 1 en la escala logarítmica corresponde a una multiplicación de 10 en la escala original.
- Si se usa la base e, entonces los aumentos son exponenciales.
- Si algún valor de la serie original es cero o negativo, los logaritmos no son posibles.

## Rezagos o Retardo

El operador de retardo se representa con el símbolo B (*a veces L, del inglés Backshift o Lag operator*) y se define de acuerdo con que:

$$BX_t \equiv X_{t-1}, B^d X_t \equiv X_{t-d} \quad (d \geq 1 \text{ entero})$$

donde  $X_t$  es una variable (real o aleatoria) referida a un momento  $t$  determinado Mauricio (2007). Un rezago es simplemente un desplazamiento en el tiempo de una serie de tiempo; en otras palabras, se utiliza el valor de la serie en un momento anterior para predecir el valor de la misma en un momento posterior (Romaní, 2014).

El primer rezago de una serie de tiempo es  $Y_t$  es  $Y_{t-1}$ .

El rezago  $j$ -ésimo de una serie de tiempo es  $Y_t$  es  $Y_{t-j}$ .

Una forma conveniente de representar rezagos es usando el operador de rezagos:  $Y_{t-j} = L^j Y_t$

## Clasificación

### Estacionarias

Normalmente, una serie temporal estacionaria puede describirse eficazmente mediante su media, varianza y función de autocorrelación, o bien mediante su media, varianza y función de densidad espectral (Box G. et. al, 2016, p.21). Esta a su vez es determinada por ser constante a lo largo del tiempo debido a que su media y varianza son asimismo constante.

### No estacionarias

En la quinta edición de *Time Series Analysis: Forecasting and Control* Box, G. et al. (2016) señala qué en ciertas subáreas de industrias, como las cifras de ventas o el valor de las acciones en la bolsa de valores, presentan una particularidad y esta es un comportamiento no estacionario el cuál no varía alrededor de una media fija.

## Modelo ARIMA

Una clase importante de modelos con datos longitudinales es la familia de los modelos autorregresivos de media móvil, conocido como ARIMA, generalmente aplicados para describir tendencias y generar predicciones a partir de valores pasados

de las series (Broz y Virgo, 2014). Esta herramienta estadística se utiliza para el análisis y pronóstico de series temporales y se utilizan ampliamente en diversas áreas, desde la economía y las finanzas hasta la meteorología y la epidemiología, permitiendo realizar pronósticos precisos y tomar decisiones informadas basadas en los patrones históricos. Chafloque et al. (2019) menciona que el modelo ARIMA destaca por ser simple y práctico para modelar y pronosticar las series de tiempo ya que solo requieren observaciones históricas de una variable, por lo que es menos costoso en la recopilación de datos (Conde, 2013). Asimismo, es considerado para muchos académicos para realizar pronósticos o predicciones en series de tiempo (Chafloque et al., 2019).

Las diferentes versiones de los modelos ARIMA, propuestos por Box y Jenkins (1970) para identificar, estimar y diagnosticar modelos dinámicos de series temporales, se han aplicado en más de dos tercios de los estudios posteriores al año 2000 que utilizaron técnicas de previsión con series de tiempo (Conde, 2013). Altamirano (2022) explica que este modelo se ha descrito como una alternativa para los modelos econométricos durante los años de 1960 debido que algunos economistas expresaron cierta preocupación sobre la fiabilidad predictiva de los modelos a gran escala, son prácticamente lo más exactos para la predicción disponibles hasta el momento y son especialmente adecuados para las series de periodicidad cortas (Countin, 2007). “Estas destacan porque con un número reducido de parámetros permiten explicar la estructura de correlación que domina una serie” (Cedeño y Carpio, 2022, p. 31). “En una competencia para pronosticar con el mejor modelo, los enfoques econométricos para el pronóstico se enfatizan cuando se utilizan datos anuales, mientras que los modelos de series temporales suelen mostrar ventaja para frecuencias más altas por ejemplo los datos mensuales” (Kumar y Sharma, 2016, p. 108). Además, “se analizan sobre una serie estacionaria y se necesitan como mínimo 50 datos para que su predicción sea confiable” (Box et. al, 2015, como se citó Rojas et. al, 2022, p. 655). Por esta razón, este modelo tiene una larga historia de confiabilidad en la predicción y análisis de datos ya que este tipo de series de tiempo es bastante flexible (Altamirano, 2022).

El ARIMA utiliza las series de tiempo para poder pronosticar un valor en el futuro, estos modelos son más apropiados para predicciones a largo plazo que para corto plazo (Rojas et al., 2020; Broz y Virgo, 2014). Este modelo predictivo es bastante

funcional ya que según Evans (2003) para lograr mejores predicciones, la mayoría de los economistas y pronosticadores, han recurrido al uso de métodos estadísticos que no involucran teoría económica como los modelos de series de tiempo univariados. Este enfoque ha surgido por una variedad de razones, que reflejan tanto las limitaciones de la teoría económica como la necesidad de un análisis más empírico y preciso de los datos económicos.

Según Goh y Law (2002), diversos estudios han demostrado cómo los modelos ARIMA y sus diferentes variantes obtienen buenos resultados en los pronósticos y superan en la mayoría de los casos a otros métodos de series de tiempo, como el estudio de Farooqi (2014), Ramlan (2021), Ghosh (2017), Upadhyay (2013) en el que se utilizó el ARIMA para pronosticar las exportaciones e importaciones de bienes, arrojando buenos resultados en la predicción, asimismo Allen (1994) reporta que la predicción de producción y precios con modelos econométricos arroja pobres resultados, sin embargo, encontró que las diferentes opciones de modelos de series de tiempo autorregresivos integrados de promedios móviles (ARIMA) lograban mejores resultados en cuanto precisión (López, 2013).

Sin embargo, es importante recalcar que el modelo ARIMA requiere un largo número de observaciones para poder determinar el mejor modelo para las series de datos (Altamirano et al., 2022). Ya que este explica el comportamiento de una serie temporal a partir de las observaciones pasadas de la propia serie y a partir de los errores pasados de previsión (Conde, 2013), y esto con la finalidad de obtener resultados confiables debido a la dependencia temporal que se utiliza en el modelo. Es útil también en situaciones en las que los datos de series temporales muestran fluctuaciones periódicas de estacionalidad que se repiten con casi la misma intensidad cada año (Box y Jenkins, 1976), ya que el modelo se adapta bien a las series temporales con estacionalidad ya que puede capturar y modelar los patrones estacionales. Una de la suposición fundamental utilizada para poder construir este modelo es que la serie temporal debe ser considerada lineal y tenga una distribución normal (Ghosh, 2017) y asimismo el modelo asume que existen valores para cada periodo de tiempo, esto no es común en muchas series de tiempo, lo que sugiere que se deben calcular los datos faltantes para considerar el modelo ARIMA (Ramírez, 2021). Amaris (2017) además nos indica que los pronósticos mediante este modelo requieren objetivos que sean orientados a identificar la consistencia, veracidad y la calidad de la información

recopilada para un mejor análisis, ya que pueden presentar inconvenientes si las series involucradas han experimentado cambios estructurales en su trayectoria (Broz y Virgo, 2014).

El modelo ARIMA (p, d, q) donde p, se define como el orden del proceso autorregresivo; d, el número de diferencias que son necesarias para que un proceso sea estacionario; y q, el orden del proceso de medias móviles (Ramírez, 2021). Y se representa mediante la siguiente expresión matemática:

$$y_t = a + \phi y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (1)$$

En la cual,

- **Y<sub>t</sub>**, es el valor de la serie de tiempo en el instante de tiempo “t”
- **a**, es una constante
- **φ**, son los coeficientes autorregresivos
- **θ**, son los coeficientes de la media móvil
- **ε(t)** es un término de error estocástico de ruido blanco "t".

Donde φ corresponde al coeficiente autorregresivo a determinar, θ coeficiente de media móvil a determinar, ε término de error y Y<sub>t-p</sub> es el registro normalizado de la serie a modelar (Amaris et al., 2017).

Existe casos especiales muy importantes en el modelo ARIMA, en la cual los siguientes modelos representan casos no estacionarios que son muy comunes en las prácticas:

1. El proceso (0,1,1):

$$\begin{aligned} y_t &= a_t - \theta_1 a_{t-1} \\ &= (1 - \theta_1 B) a_t \end{aligned}$$

2. El proceso (0,2,2):

$$\begin{aligned} y_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \\ &= (1 - \theta_1 B - \theta_2 B^2) a_t \end{aligned}$$

3. El proceso (1,1,1):

$$y_t - \phi_1 y_{t-1} = a_t - \theta_1 a_{t-1}$$

### Modelo autorregresivo (AR)

Un modelo es autorregresivo si predice el futuro con los valores del pasado (Altamirano et al., 2022). Considera que el valor de la serie estacionaria en el tiempo presente  $t$  depende de todos los valores pasados que ha tomado la serie, ponderados por un factor de peso que mide la influencia de ese valor pasado en el valor presente; y de una perturbación aleatoria presente (Hernandez et al., 2008).

Una de las características que puede tener una serie de tiempo es que los valores de la variable de estudio dependen de sus valores anteriores. En términos matemáticos, un modelo autorregresivo (AR) calcula el valor de la variable  $y$  en el instante  $t$ . Donde  $a$  y  $\phi$  son constantes propias de cada serie y  $\epsilon_t$  es el ruido blanco (Montes et al., 2016).

$$y_t = a + \phi y_{t-1} + \epsilon_t$$

Este proceso es estacionario si  $|\phi| < 1$ ; cuando  $\phi = 1$  el proceso se llama caminata aleatoria Colonescu (2016).

### Modelo de medias móviles (MA)

Los datos de una serie de tiempo dependen del ruido blanco que se encuentre asociado a cada una de las observaciones que constituyen la serie. En este caso, no sólo se incluye un término que resume los ruidos de la totalidad de la serie ( $\epsilon_t$ ) sino que en cada observación ( $y_i$ ) se deberá agregar el ruido correspondiente ( $\epsilon_i$ ) (Montes et al., 2016). Cuando sólo las últimas perturbaciones pasadas afectan significativamente el valor presente de la serie, el modelo se denomina Media Móvil de orden  $q$  notado MA ( $q$ ) (Hernández et al., 2008). Un modelo de medias móviles puede obtenerse a partir de un modelo autorregresivo sin más que realizar sucesivas sustituciones (Arce y Mahía, s.f).

$$y_t = a + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

### Modelo autorregresivo de medias móviles

Estos dos modelos básicos para series estacionarias se combinan para producir los modelos ARMA ( $p, q$ ). En general las series de tiempo no son estacionarias, pero por medio de transformaciones de varianza y de diferencias pueden ser transformadas en estacionarias (Hernández et al., 2008). El análisis de los modelos AR y de los

modelos MA resulta sencillo, pues simplemente se está considerando una componente de la serie de estudio donde p y q representan, respectivamente, el orden de autorregresión y el orden de media móvil. (Montes et al., 2016). Sin embargo, un modelo ARMA solo tiene utilidad en la estimación cuando  $Y_t$  es estacionaria, significando que la media, varianza y covarianza de la serie sean constantes a lo largo del tiempo (Sánchez et al., 2013).

La expresión general de un modelo ARMA se presenta en la siguiente ecuación:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

que puede ponerse de la forma:

$$Y_t - \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

es decir,  $Y_t(1 - \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p) = a_t(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$

El proceso ARMA (p, q) es estacionario si lo es su componente autorregresiva, y es invertible si lo es su componente de medias móviles (Fernández, s.f).

- Un modelo ARMA (p,q) es estacionario si las raíces del polinomio definido por  $(1 - \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p)$  caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación:

$$x^p - \phi_1 x^{p-1} - \phi_2 x^{p-2} - \dots - \phi_{p-1} x - \phi_p = 0 \text{ sean todas inferiores a uno en módulo.}$$

- Un modelo ARMA (p,q) es invertible si las raíces del polinomio en B definido mediante  $(1 - \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$  caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación:

$$x^q - \theta_1 x^{q-1} - \theta_2 x^{q-2} - \dots - \theta_{q-1} x - \theta_q = 0 \text{ sean todas inferiores a uno en módulo.}$$

Un modelo ARMA puede ajustarse a cualquier patrón de datos. Sin embargo, los valores de p y q se deben de especificar, por ejemplo:

ARMA (1,0)

$$Y_t = \phi_1 Y_{t-1} + a_t$$

ARMA (2,0)

$$y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t$$

ARMA (0,1)

$$Y_t = a_t - \theta_1 a_{t-1}$$

ARMA (0,2)

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

ARMA (1,1)

$$Y_t = \phi_1 Y_{t-1} + a_t - \theta_1 a_{t-1}$$

## **Marco Conceptual**

### **Series de tiempo**

Las series de tiempo son secuencias de muestras de una variable tomadas típicamente en instantes sucesivos y espaciados uniformemente. Estas series son herramientas estadísticas utilizadas para explicar (y en algunos casos predecir) el valor que toma en un momento de tiempo determinado la variable analizada (López y Martínez, 2013).

### **Componentes de una serie de tiempo**

#### **Tendencia**

Una serie de tiempo con tendencia es aquella que contiene un componente de largo plazo que representa el crecimiento o declinación de la serie a través de un período amplio (Murillos, 2003).

#### **Estacional**

Se define como estacional una serie de tiempo con un patrón de cambio a sí mismo año tras año. Por lo regular, el desarrollo de una técnica de pronóstico estacional comprende la selección de un método multiplicativo o uno de adición y estimar después índices estacionales a partir de la historia de la serie (Murillos, 2003).

#### **Ciclo**

El efecto cíclico se define como la fluctuación en forma de onda alrededor de la tendencia. Los patrones cíclicos tienden a repetirse en los datos cada dos, tres o más años. Es difícil establecer un modelo para estos patrones cíclicos, ya que no son estables (Murillos, 2003).

#### **Irregular**

El componente irregular de la serie de tiempo es el factor residual, es decir, “todo lo que sobra” y toma en consideración las desviaciones de los valores reales de

la serie analizada en comparación con los esperados; es el elemento aleatorio (Murillos, 2003).

## **ARIMA**

El modelo ARIMA permite describir un valor como una función lineal de datos anteriores y errores debidos al azar, además, puede incluir un componente cíclico o estacional. Es decir, debe contener todos los elementos necesarios para describir el fenómeno. Box y Jenkins recomiendan como mínimo 50 observaciones en la serie temporal (Fernández, s.f).

$$y_t = a + \phi y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

### **Modelo Autorregresivo (AR)**

Un modelo se define como autorregresivo si la variable interna de un periodo determinado es posible explicarla por las observaciones, de ella misma, de los periodos anteriores (de su propio pasado), añadiendo, además, un término de error (Nicasio et al., 2018).

$$y_t = a + \phi y_{t-1} + \epsilon_t$$

### **Modelo Integrado (I)**

Se refiere al estado de la variable, es decir, si se va a trabajar sobre el valor sin modificación sobre su primera o segunda diferencia, entendiendo la primera diferencia simplemente como la primera variación de la serie en estudio (Fernández et al., 2019).

### **Modelo Media Móvil (MA)**

Un modelo de los denominados de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Estos modelos se denotan normalmente con las siglas MA (Fernández et al., 2019)

$$y_t = a + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

### **Modelo Autorregresivo y de Medias Móviles (ARMA)**

A la combinación formada por procesos autorregresivos y de medias móviles se conoce como proceso ARMA. Llamado también proceso mixto, si este contiene p términos autorregresivos y q términos de medias móviles, se dice que es de orden ARMA (p,q) (Chávez, 1997).

$$y_t = a + \phi y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$$

## **Marco Legal**

En el ámbito legal se han generado varios planes alineados a una normativa legal entre los cuales se destacan: El Plan Nacional de Telecomunicaciones y Tecnologías de Información del Ecuador 2016-2021. El Convenio Marco de Cooperación Interinstitucional firmado con los diferentes entes como el Ministerio de Telecomunicaciones y de La Sociedad de la Información, la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación.

Los Planes Nacionales para el Desarrollo (2013-2017). La Ley Orgánica de Telecomunicaciones. El Código Orgánico de La Economía Social de los Conocimientos, Creatividad e Innovación. Así como los artículos que contemplan el desarrollo de la investigación científica y de la innovación tecnológica, de las TIC y diversas áreas del conocimiento que están en la Constitución de la República del Ecuador de los cuales se puede destacar los (Art. 281, Art. 385, Art. 387, Art. 423, Art. 388, Art. 313) que permiten alinearse y ser partícipes de la adopción de estas tecnologías emergentes para el cambio y adopción de una cultura de transformación digital para un mejor desarrollo a nivel empresarial en el ámbito local con proyección internacional, lo que se convierte en un impacto positivo.

## Capítulo III

### Metodología

#### Población y Muestra

En el campo del análisis de series temporales, el modelo ARIMA se ha establecido como una herramienta efectiva para comprender patrones y realizar predicciones futuras, el modelo utiliza información pasada como observaciones, tendencias y estacionalidad para predecir valores en el futuro. A diferencia de otros tipos de investigaciones, esta no requiere el concepto de población y muestra, dado que ARIMA es un modelo univariante, que solo tiene en cuenta los valores pasados de la variable que se predice, no incorpora datos de población ni de muestra (Beard et al., 2019). En lugar de utilizar parámetros externos como el tamaño de la población o de la muestra, los modelos ARIMA pretenden capturar los patrones temporales y la autocorrelación dentro de una única serie temporal (Liu et al., 2011).

Otra razón por la que los modelos ARIMA pueden no utilizarse población y muestras es que no son adecuados para captar relaciones no lineales en los datos, los modelos ARIMA asumen la linealidad y pueden no ser capaces de pronosticar con precisión datos de series temporales que presentan patrones no lineales o tienen factores no lineales. En tales casos, el supuesto de linealidad de los modelos ARIMA puede limitar su capacidad para proporcionar previsiones precisas (Xu et al., 2022).

Los valores históricos de la variable que se proyecta sirven de entrada principal para los modelos ARIMA. Con el fin de encontrar patrones, tendencias y estacionalidad que puedan utilizarse para predecir valores futuros, estos modelos analizan los valores históricos de las series temporales.

Aunque la población y el tamaño de la muestra no se incorporan a los modelos ARIMA, estas características pueden influir indirectamente en la precisión de las previsiones. Por ejemplo, una mayor población o tamaño de la muestra puede ofrecer más datos, lo que podría mejorar la capacidad del modelo para identificar tendencias y generar previsiones precisas. Un menor tamaño de la población o de la muestra suele

asociarse a proyecciones menos precisas, ya que hay una escasez de datos (Wolfe et al., 2021).

## **Normalización**

La normalización de una base de datos es un principio fundamental en el diseño de bases de datos relacionales. Consiste en organizar los datos de una base de datos para eliminar redundancias y mejorar la calidad y el rendimiento de los datos. El proceso de normalización consiste en dividir una base de datos en varias tablas y establecer relaciones entre ellas. El objetivo es minimizar la duplicación de datos y garantizar que cada dato se almacena en un único lugar (Albarak y Bahsoon, 2018).

Existen varias ventajas de la normalización de las bases de datos. En primer lugar, mejora la integridad de los datos al reducir su redundancia, ya que los datos redundantes pueden provocar incoherencias y anomalías en la base de datos. Al eliminar la redundancia mediante la normalización, se mantiene la integridad de los datos y se minimiza el riesgo de incoherencias (Albarak y Bahsoon, 2018).

En segundo lugar, la normalización mejora la coherencia y precisión de los datos. Cuando los datos se almacenan en varios lugares, resulta difícil actualizarlos y mantener su coherencia en toda la base de datos. La normalización garantiza que cada dato se almacena en un único lugar, lo que facilita su actualización y precisión (Albarak y Bahsoon, 2018).

En tercer lugar, la normalización mejora la recuperación de datos y el rendimiento de las consultas. Al descomponer una base de datos en varias tablas y establecer relaciones entre ellas, las consultas pueden ejecutarse con mayor eficacia. La normalización reduce la necesidad de complejas uniones y mejora el rendimiento general de la base de datos (Albarak y Bahsoon, 2018).

Además, la normalización facilita el mantenimiento y la escalabilidad de la base de datos. A medida que aumenta el tamaño y la complejidad de una base de datos, resulta más difícil mantenerla y actualizarla. La normalización simplifica la estructura de la base de datos y facilita su gestión y ampliación a medida que evolucionan las necesidades de la organización (Albarak y Bahsoon, 2018).

## **Metodología de Box y 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 (Muñoz et al, 2017). Medina (2011) señala que propusieron esta metodología rigurosa con el objetivo de identificar, estimar y diagnosticar modelos dinámicos para datos de series temporales. Este método de predicción se basa en el análisis de las propiedades probabilísticas o estocásticas de las series de tiempo económicas (Gujarati, 2010).

La metodología Box Jenkins se utiliza para el proceso óptimo de creación de modelos en ARIMA, lo que ha hecho que ARIMA sea popular (Ghosh, 2017). En la cual consiste en estimar modelos autorregresivos de media móvil (ARIMA) con el fin de encontrar el mejor ajuste para datos longitudinales (Broz y Virgo (2014). Es un método ampliamente utilizado para el análisis y pronóstico de series temporales y se ha convertido en una herramienta fundamental en el campo de las series temporales.

Desarrollar un modelo basado en series de tiempo no es algo para nada fácil, sin embargo, Box y Jenkins lograron crear una metodología bastante práctica para estimar los modelos de series de tiempo (Hernandez et al., 2008). El método Box-Jenkins 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 et al., 2017). Es importante retomar la necesidad en la metodología Box-Jenkins, pues resulta imprescindible utilizar una serie estacionaria o, en su defecto, su transformación diferenciada (Oliva de con et al., 2016).

### **La metodología Box-Jenkins consta de tres etapas principales:**

#### **Identificación y selección del modelo**

La primera es la identificación y selección del modelo la cual se basa a partir del análisis de los diagramas de función de autocorrelación ACF y autocorrelación parcial PACF, asegurando de que el comportamiento de la variable sea estacionario y se identifique también un componente de estacionalidad (Guzman, 2019). Si la secuencia de las observaciones no demuestra estacionariedad, debe convertirse en una

serie estacionaria mediante una diferenciación apropiada para hacer las series bajo consideración estacionaria (Ghosh, 2017). Es importante destacar el hecho de que este procedimiento está sesgado hacia el uso del juicio personal porque no existen reglas claras sobre cómo decidir sobre los componentes AR Y MA apropiados. Por lo tanto, la experiencia juega un papel fundamental en este sentido (Rosales, 2022).

### **Estimación de los parámetros**

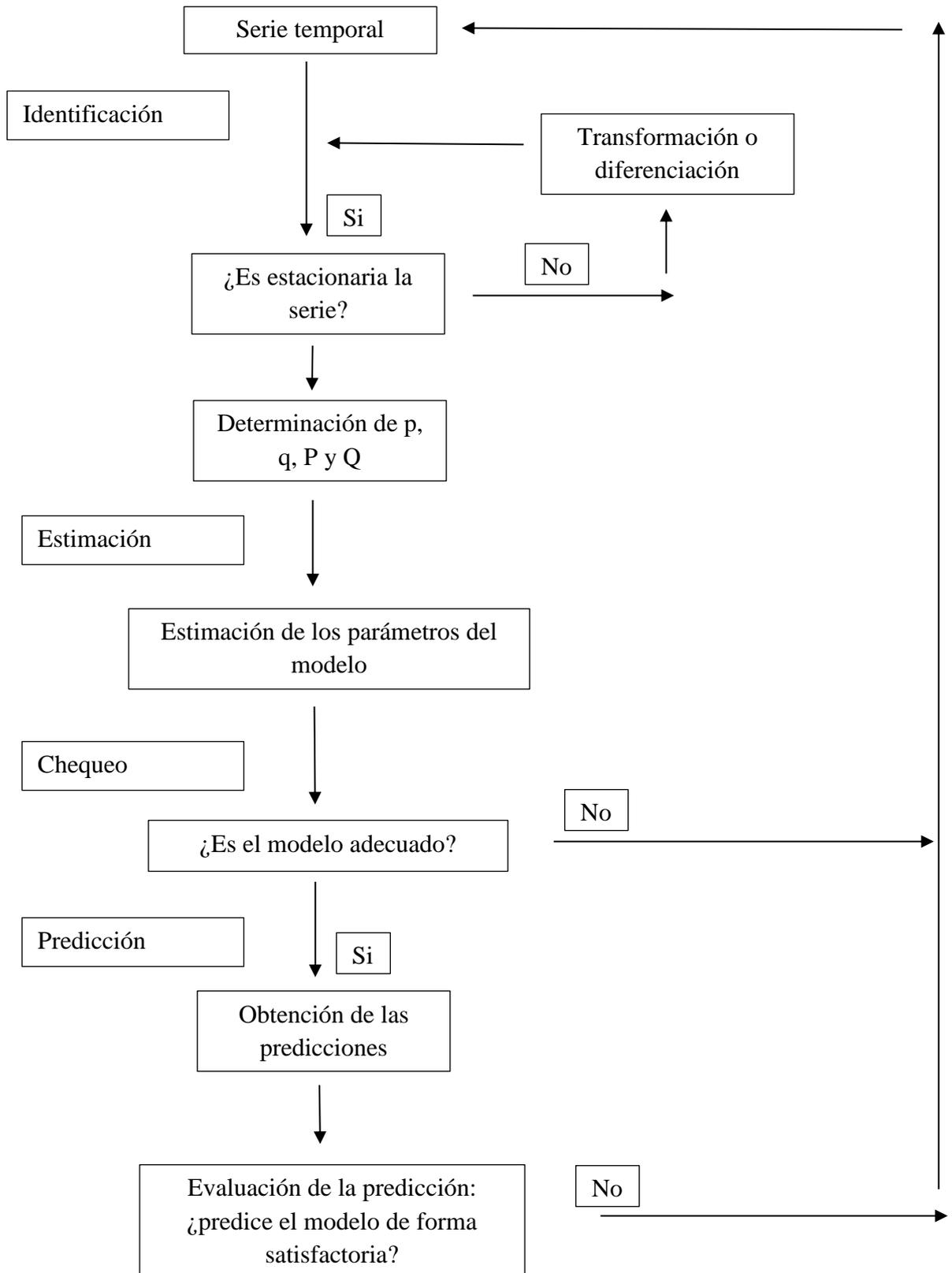
Una segunda fase consistirá en la estimación de los parámetros a través de la aplicación de algoritmos para la obtención de coeficientes del modelo lineal que de un mejor ajuste del modelo a las condiciones reales de la variable. Esto se puede hacer mediante estimación de máxima verosimilitud o por mínimos cuadrados no lineales (Guzman, 2019).

### **Validación del modelo obtenido**

Y, por último, una tercera fase que consiste en la validación y ajuste del modelo a través de pruebas de independencia de los residuos, cuyas medias y varianzas deben ser constante en el tiempo, o especialmente manejar una distribución normal (Guzman, 2019). La selección del modelo construido se puede probar a través de un diagnóstico analítico (Gosh, 2017). La verificación de diagnóstico generalmente se realiza generando el conjunto de residuos y probando si satisfacen las características de un proceso de ruido blanco (Rosales, 2022).

**Figura 7.**

**Metodología de Box Jenkins para el análisis de series de tiempo**



*Nota.* La figura muestra los pasos para realizar el pronóstico del modelo autorregresivo integrado media móvil. Fuente: Gorina et al., 2011

## **Ruido Blanco**

El proceso denominado ruido blanco que es una secuencia de variables aleatorias de media cero, varianzas constante y covarianzas nulas. El ruido blanco en un pronóstico se refiere a una señal aleatoria que tiene una densidad espectral de potencia constante en todas las frecuencias (Ade et al., 2019). Es un tipo de proceso estocástico que se caracteriza por variables aleatorias independientes y no correlacionadas con media cero y varianzas constante (Fryz & Scherbak, 2023). En el pronóstico de series temporales, el ruido blanco representa el componente aleatorio e impredecible de los datos que el modelo no puede explicar ni predecir.

Igualmente, se utiliza a menudo como punto de referencia para evaluar el rendimiento de los modelos de previsión (Kurov et al., 2018). Si un modelo de pronóstico no puede superar la precisión de un pronóstico de ruido blanco, indica que el modelo no está capturando ningún patrón o relación significativa en los datos.

En algunos casos, se puede añadir ruido blanco a un modelo de previsión para tener en cuenta las fluctuaciones aleatorias e impredecibles de los datos (Li y Guo, 2021). Esto puede ayudar a mejorar la capacidad del modelo para captar la variabilidad y la incertidumbre inherentes a las series temporales.

El ruido blanco en las previsiones es importante por varias razones. Esta sirve de referencia para evaluar el rendimiento de los modelos de previsión. Si un modelo de previsión no puede superar la precisión de una previsión con ruido blanco, indica que el modelo no está captando ninguna pauta o relación significativa en los datos (Lawrence et al., 2006).

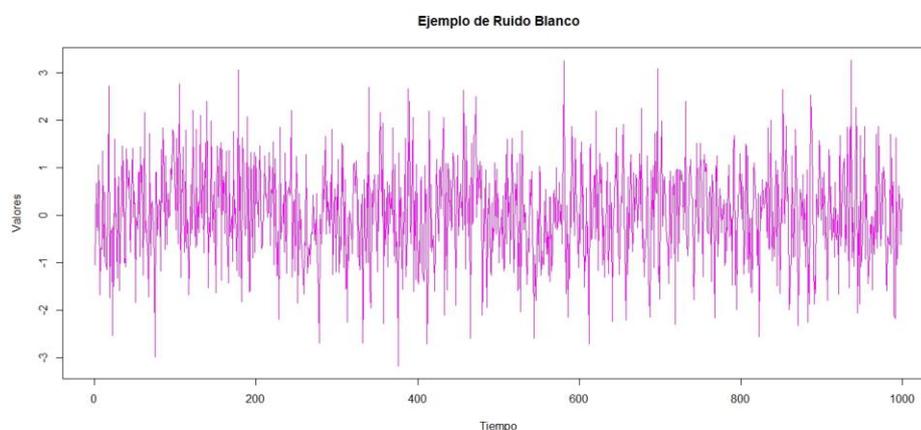
Además, el ruido blanco representa el componente aleatorio e impredecible de los datos que el modelo no puede explicar ni predecir. La incorporación del ruido blanco a los modelos de previsión ayuda a tener en cuenta la variabilidad y la incertidumbre inherentes a las series temporales (Ragozin y Pletenkova, 2022). Esto puede dar lugar a predicciones más sólidas y precisas, especialmente en situaciones en las que los datos presentan fluctuaciones aleatorias o ruido (Angwin et al., 2017).

También se utiliza en diversas técnicas y modelos estadísticos para mejorar los resultados de las predicciones. Por ejemplo, en el campo de la econometría, a menudo se asume que los residuos generados por los pronosticadores siguen un comportamiento de ruido blanco (Chlebus et al., 2021). Este supuesto ayuda a garantizar que los errores de previsión no estén sesgados y presenten características aleatorias (Chlebus et al., 2021).

Por último, el ruido blanco es relevante en el contexto de la incertidumbre del modelo y la estimación de parámetros. Los modelos estocásticos que incorporan ruido blanco pueden representar mejor las incertidumbres y fluctuaciones del sistema modelado. Al considerar el ruido blanco como una fuente de incertidumbre, permite realizar previsiones más realistas y fiables (Arnold et al., 2013).

### **Figura 8.**

#### **Ruido Blanco**



*Nota.* La gráfica representa un ejemplo del ruido blanco. Tomado de Legorreta (2015). (<https://dlegorreta.wordpress.com/tag/ruido-blanco/>)

#### **Proceso Autoregresivo**

En esta etapa, el proceso autorregresivo se implementa para poder determinar si la serie estudia requiere de una diferenciación o integración, esto indica las veces que se requiere para poder transformar la serie en estacionaria. Definimos un modelo como autorregresivo si la variable endógena de un período  $t$  es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiendo, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios

con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda  $Y_t$  puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error (innovación)(Arce & Bahía, 2001, p5).

Los procesos autorregresivos presentan función de autocorrelación parcial abreviada como ACFP, con un número finito de valores distinto de cero. Un proceso  $AR(p)$  tiene los primeros  $p$  términos de la función de autocorrelación parcial distintos de cero y los demás son nulos. Esta afirmación es muy fuerte, y en la práctica se considera que una muestra dada proviene de un proceso autorregresivo de orden  $p$  si los términos de la función de autocorrelación parcial son casi cero a partir del que ocupa el lugar  $p$  (Casimiro, 2009).

Los modelos autorregresivos se abrevian con la palabra AR tras la que se indica el orden del modelo:  $AR(1)$ ,  $AR(2)$ ,.....etc. El orden del modelo expresa el número de observaciones retrasadas de la serie temporal analizada que intervienen en la ecuación.

### **Media Móvil**

El proceso de media móvil es un componente del modelo ARIMA (Autoregressive Integrated Moving Average). En el modelo ARIMA, el proceso de media móvil (MA) es uno de los tres subprocesos, junto con la autorregresivo (AR) y la diferenciación, que forman una serie temporal estacionaria Chadsuthi et al. (2015).

El proceso MA en ARIMA implica el uso de los términos de error pasados o residuales para predecir el valor actual de la serie temporal. Asume que el valor actual de la serie es una combinación lineal de los términos de error pasados y un término constante. El proceso MA ayuda a capturar las dependencias a corto plazo y las fluctuaciones aleatorias de los datos.

Es especialmente útil para modelizar y predecir series temporales de datos que presentan fluctuaciones aleatorias e impredecibles. Ayuda a tener en cuenta el ruido o el componente aleatorio de los datos que no pueden explicar los componentes autorregresivos o de tendencia.

La estimación y selección del orden adecuado del proceso MA en el modelo ARIMA puede determinarse utilizando diversas técnicas estadísticas, como el Criterio de Información de Akaike (AIC) o el Criterio de Información Bayesiano (BIC). Estos criterios ayudan a identificar el orden óptimo que minimiza los errores residuales y proporciona el mejor ajuste a los datos.

### **Estacionariedad**

Para poder implementar un modelo de pronóstico ARIMA se debe determinar si la serie estudiada es adecuada para el modelo mencionado. En el análisis de series temporales el objetivo es utilizar la teoría de procesos estocásticos para determinar qué proceso estocástico ha sido capaz de generar la serie temporal bajo estudio con el objetivo de caracterizar el comportamiento de la serie y que permita predecir el futuro. Si se quieren conseguir métodos de predicción consistentes, no se puede utilizar cualquier tipo de proceso estocástico, sino que es necesario que la estructura probabilística del mismo sea estable en el tiempo. La filosofía que subyace en la teoría de la predicción es siempre la misma: se aprende de las regularidades del comportamiento pasado de la serie y se proyectan hacia el futuro. Por lo tanto, es preciso que los procesos estocásticos generadores de las series temporales tengan algún tipo de estabilidad. Si, por el contrario, en cada momento de tiempo presentan un comportamiento diferente e inestable, no se pueden utilizar para predecir. A estas condiciones que se les impone a los procesos estocásticos para que sean estables para predecir, se les conoce como estacionariedad.

El concepto de estacionariedad se puede caracterizar bien en términos de la función de distribución o de los momentos del proceso, la función de distribución de un proceso estocástico incluye todas las funciones de distribución para cualquier subconjunto finito de variables aleatorias del proceso y los momentos del proceso hace referencia a el conjunto de las medias de todas las variables aleatorias del proceso, el conjunto de la varianza de todas las variables aleatorias y la covarianza entre todo par de variables aleatorias del proceso en estudio. En el primer caso, se hablará de estacionariedad en sentido estricto y, en el segundo, de estacionariedad de segundo orden o en covarianza (González, 2009)

Se puede calcular la sucesión de medias y varianzas por años, si se obtienen variaciones significativas crecientes y decrecientes a lo largo de los años, indica que no hay estacionariedad. Este resultado conduce a tomar logaritmos y diferenciar la serie original con el objetivo de atenuar la falta de estacionariedad en media y varianza.

Una serie será estacionaria en varianza cuando pueda mantenerse el supuesto de que existe una única varianza para toda la serie temporal, es decir, cuando la variabilidad de la serie en torno a su media se mantenga constante a lo largo del tiempo. Si la serie no es estacionaria en varianza, se utilizan las transformaciones estabilizadoras de varianza, es decir, las transformaciones Box-Cox (González, 2009).

Las transformaciones Box-Cox incluye una familia infinita de funciones: raíz cuadrada, inversa, la transformación más utilizada en la práctica económica es la logarítmica. Los instrumentos que se utilizan para analizar la estacionariedad en varianza de una serie son el gráfico de la serie original y de las transformaciones correspondientes. Si la serie no es estacionaria en media se puede lograr la estacionariedad transformándola, tomando diferencias. Así, si la serie no es estacionaria en media, se tomarán de sucesivas diferencias de orden 1 sobre la serie hasta obtener una serie estacionaria.

## **Estacionalidad**

Si la serie temporal presenta una tendencia, lo primero que debe de hacerse es convertirla en estacionaria mediante una diferenciación de orden  $d$ . Una vez diferenciada la serie, una buena estrategia consiste en comparar los correlogramas de la función de autocorrelación (ACF) y la función de autocorrelación parcial (ACFP), proceso que suele ofrecer una orientación para la formulación del modelo orientativo.

## **Raíz Unitaria**

La prueba Dickey-Fuller es una prueba estadística utilizada en el modelado ARIMA para determinar la estacionariedad de una serie temporal (Wang et al., 2018). La estacionariedad es un supuesto crucial para los modelos ARIMA, ya que están diseñados para trabajar con datos de series temporales estacionarias. Una serie temporal estacionaria es aquella en la que la media, la varianza y la autocovarianza no cambian con el tiempo. Y busca una raíz unitaria, un signo de no estacionariedad, en

la serie temporal. La serie temporal tiene una raíz unitaria, lo que indica que no es estacionaria, y ésta es la hipótesis nula de la prueba. Si el estadístico de la prueba es significativamente distinto de cero y el valor p está por debajo de un determinado umbral (normalmente 0,05), se rechaza la hipótesis nula, lo que indica que la serie temporal es estacionaria.

La prueba Dickey-Fuller aumentada (ADF) es una extensión de la prueba Dickey-Fuller original que tiene en cuenta factores adicionales como la tendencia y la estacionalidad en las series temporales (Wang et al., 2018). La prueba ADF se utiliza comúnmente en la modelización ARIMA para comprobar la estacionariedad y determinar el orden apropiado de diferenciación (d) necesario para que la serie temporal sea estacionaria.

En el modelado ARIMA, la prueba Dickey-Fuller se aplica normalmente a la serie temporal original, y si se encuentra que no es estacionaria, se realiza la diferenciación hasta que se alcanza la estacionariedad (Wang et al., 2018). El orden de diferenciación (d) es un parámetro importante en los modelos ARIMA y se determina en función de los resultados de la prueba Dickey-Fuller. Para determinar si una serie de tiempo tiene una raíz unitaria, se utilizan pruebas estadísticas como la prueba de Dickey-Fuller o la prueba de Phillips-Perron.

Estas pruebas evalúan si la serie de tiempo es estacionaria o si tiene una raíz unitaria presente. Si se rechaza la hipótesis nula de la presencia de una raíz unitaria, se puede modelar la serie de tiempo utilizando métodos apropiados para datos estacionarios. Si la presencia de una raíz unitaria no se rechaza, se pueden aplicar técnicas de diferenciación para transformar la serie en una forma estacionaria antes de modelar (Villegas, 2020).

ARIMA es una abreviatura que representa AutoRegressive Incorporated Moving Normal. Es una especulación de la más directa Normal Móvil Auto-Regresiva e incluye el pensamiento de combinación. Esta abreviatura es inconfundible, capturando las partes clave del propio modelo. Rápidamente, lo son:

**AR(p): Autoregresión.** Un modelo que utiliza la conexión necesaria entre la percepción y un cierto número de percepciones flojas.

**I(d): Integrado.** La utilización de la diferenciación de las percepciones crudas (por ejemplo, restando una percepción de la percepción en el paso de tiempo pasado) para hacer que la disposición temporal sea estacionaria.

**MA(q).** Hace referencia a las medias móviles, estas se utilizan en el modelo ARIMA para suavizar los datos de series de tiempo y reducir el ruido en los datos.

Cada una de estas partes se indica inequívocamente en el modelo como un parámetro. Se utiliza la documentación estándar de ARIMA(p,d,q) donde los parámetros se sustituyen por calidades de números enteros para mostrar rápidamente el modelo ARIMA particular que se está utilizando.

Los parámetros del modelo ARIMA se caracterizan como seguidos:

p: El número de percepciones de holgura incorporadas en el modelo, también llamado petición de holgura.

d: Las ocasiones en que las percepciones crudas se diferencian se llaman también el nivel de diferencia.

q: El tamaño de la ventana de movimiento normal, adicionalmente llamado la solicitud de movimiento normal.

Se elabora un modelo de recaída recta que incluye el número y el tipo de términos predefinidos, y la información se establece por un nivel de diferenciación para hacerla estacionaria, por ejemplo, para expulsar patrones y estructuras regulares que influyen negativamente en el modelo de recaída.

Se puede utilizar una estimación de 0 para un parámetro, lo que demuestra que no se utiliza ese componente del modelo. Por lo tanto, el modelo ARIMA puede ser diseñado para reproducir la capacidad de un modelo ARMA, e incluso un modelo básico AR, I o Mama.

## **Forecast**

La base del forecast en el modelo ARIMA es la aplicación de la información de valores pasados y presente observado para predecir los valores futuros, el pronóstico es la última fase en un modelo de predicción ARIMA. Se selecciona el

modelo ARIMA que minimiza la medida de error (Ayala, 2021) la cuál utiliza los patrones que han generado las variables pasadas, para poder así generar el *forecast*, en este paso se debe de establecer para cuantos periodos se va a pronosticar y de igual manera los periodos/ intervalos de tiempos deben de ser los mismos de la serie original. Se utiliza un algoritmo que busca el mejor modelo ARIMA para los datos de la serie de tiempo, basándose en diferentes criterios de selección de modelos basándose en diferentes criterios de selección de modelos, como el criterio de información de Akaike (AIC) y ajusta diferentes modelos ARIMA con diferentes combinaciones de  $p$ ,  $d$  y  $q$ , y selecciona el modelo que minimiza la medida de error seleccionada para un pronóstico con mayor confiabilidad (Amarís et al., 2017).

## Análisis de resultados

### READXL

readxl es un paquete diseñado para hacer una sola tarea: importar hojas de Excel a R. Esto hace que sea un paquete ligero y eficiente, a cambio de no contar con funciones avanzadas.

#### *excel\_format*

Determinar si los archivos son xls o xlsx (o de la familia xlsx)

#### *excel\_sheets*

La lista de nombres de hojas es especialmente útil cuando se desea iterar sobre todas las hojas de un libro de trabajo.

#### *readxl\_example*

readxl viene con algunos archivos de ejemplo en su directorio inst/extdata. Esta función facilita el acceso a los mismos.

#### *readxl\_progress*

Por defecto, readxl muestra un spinner de progreso a menos que una de las siguientes opciones sea TRUE:

El spinner se desactiva explícitamente configurando `options(readxl.show_progress = FALSE)`.

El código se ejecuta en una sesión no interactiva (`interactive()` es FALSE).

El código es ejecutado por knitr / rmarkdown.

El código se ejecuta en un chunk de RStudio notebook. readxl utiliza el paquete progress under-the-hood y por lo tanto también es sensible a cualquier opción que consulte.

#### *read\_excel*

Leer archivos xls y xlsx

`read_excel()` llama a `excel_format()` para determinar si la ruta es xls o xlsx, basándose en la extensión del archivo y en el propio archivo, en ese orden.

## **XTS**

`xts` es un paquete de R que ofrece una serie de funcionalidades para trabajar con datos indexados en el tiempo. `xts` amplía `zoo`, otro popular paquete para el análisis de series temporales.

### *extensible times series*

Convierta fácilmente una de las muchas clases de series temporales (y no temporales) de R en un verdadero objeto basado en el tiempo que hereda todos los métodos de `zoo`, al tiempo que permite nuevas herramientas basadas en el tiempo cuando sea apropiado.

Además, se pueden utilizar `xts` para crear nuevos objetos que pueden contener atributos arbitrarios nombrados durante la creación como pares nombre=valor.

## **DPLYR**

De acuerdo con Pubs by RStudio (2023) el paquete `dplyr` fue desarrollado por Hadley Wickham de RStudio y es una versión optimizada de su paquete `plyr`. El paquete `dplyr` no proporciona ninguna nueva funcionalidad a R per se, en el sentido que todo aquello que podemos hacer con `dplyr` lo podríamos hacer con la sintaxis básica de R.

El paquete `dplyr` entrega una serie de funciones muy útiles para la manipulación de data frames, permitiendo simplificar las acciones sobre este tipo de datos. Las funciones de `Dplyr` están construidas en C++ y son por tanto bastante más rápidas. Este proporciona una serie de verbos o funciones que mejoran el acceso y manipulación de datos. Los principales son:

- Arrange
- Filter

- Group
- Mutate
- Rename
- Select
- Summarize

### *Transmute*

Realiza modificaciones manteniendo solamente las variables creadas y omitiendo las restantes. Con esta gramática podemos comunicar mediante nuestro código que es lo que estamos haciendo en los data frames a otras personas (asumiendo que conozcan la gramática). Esto es muy útil, ya que proporciona una abstracción que anteriormente no existía. Por último, cabe destacar que las funciones del paquete dplyr son muy rápidas, puesto que están implementadas con el lenguaje C + +. Todas las funciones en este paquete tienen en común una serie de argumentos. En particular.

- El primer argumento es el data frame
- Los otros argumentos describen qué hacer con el data frame especificado en el primer argumento, podemos referirnos a las columnas en el data frame directamente sin utilizar el operador \$, es decir sólo con el nombre de la columna/variable.
- El valor de retorno es un nueva data frame.
- Los data frames deben estar bien organizados/estructurados, es decir debe existir una observación por columna y, cada columna representa una variable, medida o característica de esa observación. Para ello, es muy útil el uso del paquete tidy.

Cuando trabajes con datos debes:

- Averiguar qué quieres hacer.
- Describir esas tareas en forma de programa informático.
- Ejecutar el programa.

El paquete dplyr facilita y agiliza estos pasos:

- Al limitar sus opciones, le ayuda a pensar en sus retos de manipulación de datos.

- Proporciona "verbos" sencillos, funciones que corresponden a las tareas de manipulación de datos más comunes, para ayudarle a traducir sus pensamientos en código.
- Utiliza backends eficientes, para que pases menos tiempo esperando al ordenador.

## **TIDYVERSE**

A alto nivel, el tidyverse es un lenguaje para resolver retos de ciencia de datos con código R. Su principal objetivo es facilitar la conversación entre un humano y un ordenador sobre los datos. De forma menos abstracta, el tidyverse es una colección de paquetes de R que comparten una filosofía de diseño de alto nivel y una gramática y estructuras de datos de bajo nivel, de modo que aprender un paquete facilita el aprendizaje del siguiente. El tidyverse engloba las tareas repetidas que constituyen el núcleo de todo proyecto de ciencia de datos: importación, ordenación, manipulación, visualización y programación de datos. Tidyverse permite facilitar el trabajo estadístico y la generación de trabajos reproducibles. Está compuesto de los siguientes paquetes:

- Readr
- Ggplot2
- Purr
- Tibble
- Dplyr
- Tidyr
- Stringr
- Forecast

## **LUBRIDATE**

Lubridate es un paquete de R que facilita el trabajo con fechas y horas. Conseguir que R acepte que tus datos contienen las fechas y horas que crees que

contienen puede ser complicado. Lubridate lo simplifica. Identifique el orden en que aparecen el año, el mes y el día en sus fechas.

### *Análisis de fechas*

Las funciones de análisis sintáctico de Lubridate leen cadenas en R como objetos fecha-hora POSIXct. Los usuarios deben elegir la función cuyo nombre modele el orden en que aparecen los elementos año ('y'), mes ('m') y día ('d') en la cadena que se va a analizar: dmy(), myd(), ymd(), ydm(), dym(), mdy(), ymd\_hms()). Parse\_date\_time() proporciona un analizador sintáctico muy flexible y fácil de usar.

Lubridate también puede analizar fechas parciales de cadenas en objetos Period con las funciones hm(), hms() y ms().

Lubridate incorpora un analizador POSIX muy rápido. La mayoría de los formatos de strptime() y varias extensiones están soportadas para las localizaciones inglesas. Ver parse\_date\_time() para más detalles.

### *Manipulación de fechas*

Lubridate distingue entre momentos en el tiempo (conocidos como instantes()) y lapsos de tiempo (conocidos como lapsos de tiempo, véase Timespan). Los intervalos de tiempo se dividen a su vez en objetos Duration, Period e Interval.

### *Instantes*

Los instantes son momentos específicos del tiempo. Date, POSIXct y POSIXlt son las tres clases de objetos que Base R reconoce como instantes. is.Date() comprueba si un objeto hereda de la clase Date. is.POSIXt() comprueba si un objeto hereda de las clases POSIXlt o POSIXct. is.instant() comprueba si un objeto hereda de cualquiera de las tres clases.

now() devuelve la hora actual del sistema como un objeto POSIXct. today() devuelve la fecha actual del sistema. Por conveniencia, 1970-01-01 00:00:00 se guarda en origen. Este es el instante a partir del cual se calculan los tiempos POSIXct. Pruebe unclass(now()) para ver la estructura numérica que subyace a los objetos POSIXct.

Cada objeto POSIXct se guarda como el número de segundos que ocurrieron después de 1970-01-01 00:00:00.

Conceptualmente, los instantes son una combinación de medidas en diferentes unidades (es decir, años, meses, días, etc.). Los valores individuales de estas unidades pueden extraerse de un instante y establecerse con las funciones accesorias `second()`, `minute()`, `hour()`, `day()`, `yday()`, `mday()`, `wday()`, `week()`, `month()`, `year()`, `tz()` y `dst()`. Nota: las funciones accesorias se denominan según la forma singular de un elemento. No deben confundirse con las funciones auxiliares de periodo que tienen como nombre la forma plural de las unidades (por ejemplo, `seconds()`).

### ***Husos horarios***

Lubridate proporciona dos funciones de ayuda para trabajar con zonas horarias. `with_tz()` cambia la zona horaria en la que se muestra un instante. La hora del reloj mostrada para el instante cambia, pero el momento de tiempo descrito sigue siendo el mismo. `force_tz()` cambia sólo el elemento de zona horaria de un instante. La hora del reloj mostrada sigue siendo la misma, pero el instante resultante describe un nuevo momento de tiempo.

### ***Lapso de tiempo***

Un intervalo de tiempo es un periodo de tiempo que puede o no estar relacionado con un instante concreto. Por ejemplo, tres meses es un lapso de tiempo. También lo es una hora y media. Base R utiliza objetos de la clase `difftime` para registrar intervalos de tiempo. Sin embargo, las personas no siempre son coherentes en la forma en que esperan que se comporte el tiempo. A veces el paso del tiempo es una progresión monótona de instantes que debería ser tan matemáticamente fiable como la recta numérica. En otras ocasiones, el tiempo debe seguir convenciones y reglas complejas para que las horas de reloj que vemos reflejen lo que esperamos observar en términos de luz diurna, estación del año y congruencia con el reloj atómico. Para navegar mejor por los matices del tiempo, lubridate crea tres clases de `timespan` adicionales, cada una con su propio comportamiento específico y coherente: Intervalo, Período y Duración.

`is.difftime()` comprueba si un objeto hereda de la clase `difftime`. `is.timespan()` comprueba si un objeto hereda de cualquiera de las cuatro clases `timespan`.

### ***Duraciones***

Las duraciones miden la cantidad exacta de tiempo que transcurre entre dos instantes. Esto puede dar lugar a resultados inesperados en relación con la hora del reloj si en el intervalo se produce un segundo bisiesto, un año bisiesto o un cambio en el horario de verano (DST).

Las funciones para trabajar con duraciones incluyen `is.duration()`, `as.duration()` y `duration()`. `dseconds()`, `dminutes()`, `dhours()`, `ddays()`, `dweeks()` y `dyears()` longitudes convenientes.

### ***Periodos***

Los periodos miden el cambio en el tiempo del reloj que se produce entre dos instantes. Los periodos proporcionan predicciones sólidas de la hora del reloj en presencia de segundos bisiestos, años bisiestos y cambios en el horario de verano.

Las funciones para trabajar con periodos incluyen `is.period()`, `as.period()` y `period()`. `seconds()`, `minutes()`, `hours()`, `days()`, `weeks()`, `months()` and `years()` crean rápidamente periodos de longitudes convenientes.

### ***Intervalos***

Los intervalos son espacios de tiempo que comienzan en un instante específico y terminan en un instante específico. Los intervalos conservan información completa sobre un intervalo de tiempo. Son la única forma fiable de convertir entre periodos y duraciones.

Las funciones para trabajar con intervalos son `is.interval()`, `as.interval()`, `interval()`, `int_shift()`, `int_flip()`, `int_aligns()`, `int_overlaps()` y `%within%`. Los intervalos también pueden manipularse con `intersect`, `union` y `setdiff()`.

### ***Varios***

decimal\_date() convierte un instante en un decimal de su año. leap\_year() comprueba si un instante ocurre durante un año bisiesto. pretty\_dates() proporciona un método para hacer saltos bonitos para fechas-horas. Lakers es un conjunto de datos que contiene información sobre la temporada de baloncesto 2008-2009 de Los Angeles Lakers.

## **TSERIES**

Análisis de series temporales y finanzas computacionales.

Calcula la prueba Dickey-Fuller aumentada para la nulidad de que  $x$  tiene una raíz unitaria.

Se utiliza la ecuación de regresión general que incorpora una constante y una tendencia lineal y el  $\rho$  para un coeficiente autorregresivo de primer orden igual a uno. El número de retardos utilizado en la regresión es  $k$ . El valor por defecto de  $\text{trunc}((\text{longitud}(x)-1)^{1/3})$  corresponde al límite superior sugerido sobre la tasa a la que el número de rezagos,  $k$ , debe hacerse crecer con el tamaño de la muestra para la configuración general ARMA( $p,q$ ). Obsérvese que para  $k$  igual a cero se calcula la prueba Dickey-Fuller estándar. Los valores  $p$  se interpolan a partir de la Tabla 4.2, p. 103 de Banerjee et al. (1993). Si la estadística calculada está fuera de la tabla de valores críticos, se genera un mensaje de advertencia.

## **ASTSA**

Análisis estadístico aplicado de series temporales. Contiene conjuntos de datos y secuencias de comandos para analizar series temporales tanto en el dominio de la frecuencia  $t$  como en el del tiempo, incluido el modelado del espacio de estados. Este permite al usuario realizar gráficas de autocorrelación parcial así como simple, de función de autocovarianza, generar un AR(1) y AR(2), además permite generar las ruido blanco para ser interpretados por el analista. Esta librería facilita evaluar el modelo de la serie otorgando gráficos estadísticos como; *p values for Ljung-Box, ACF of residuales and Standardized Residuals* lo que conlleva a una mejor interpretación de la distribución de los datos y si estos son dependientes o no uno del otro.

## **FORECAST**

El paquete de forecast consta de métodos y herramientas de visualización y análisis de previsiones de series temporales univariantes, incluido el alisamiento exponencial mediante modelos de espacio de estados y la modelización automática ARIMA. Realiza previsiones automáticas mediante modelos de suavización exponencial, ARIMA, el método Theta, splines cúbicos, así como otros métodos de previsión habituales a través de la validación de las series temporales, se podrá determinar la probabilidad de que un acierto en mayor o menor magnitud. Se utiliza las librerías, *tidyverse*, *lubridate*, *forecast* y el paquete *forecast* se emplea en conjunto de la función *auto.arima*.

## **FOREIGN**

Lectura y escritura de datos almacenados por algunas versiones de 'Epi Info', 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', y para la lectura y escritura de algunos archivos 'dBase'. Además de la importación, el paquete "foreign" también permite exportar datos desde R a estos formatos para compartirlos o trabajar con ellos en otros programas.

## **TIMSAC**

Este paquete proporciona funciones para el análisis estadístico, la predicción y el control de series temporales.

## **VARS**

Estimación, selección de lag, pruebas de diagnóstico, previsión, análisis de causalidad, descomposición de la varianza del error de previsión y funciones de respuesta al impulso de modelos VAR y estimación de modelos SVAR y SVEC. Se usa en econometría, finanzas además en series de tiempo.

## **MFILTER**

El paquete mFilter implementa varios filtros de series temporales útiles para suavizar y extraer componentes de tendencia y cíclicos de una serie temporal. Las

rutinas se utilizan habitualmente en economía y finanzas, aunque también deberían ser de interés en otras áreas. Actualmente, el paquete incluye los filtros Christiano-Fitzgerald, Baxter-King, Hodrick-Prescott, Butterworth y de regresión trigonométrica.

## **DYNLM**

El paquete "DYNLM" (Dynamic Linear Models) de RStudio es una herramienta para ajustar y analizar modelos lineales dinámicos. Los modelos lineales dinámicos son una clase de modelos estadísticos que se utilizan para analizar series temporales en las que las observaciones están correlacionadas en el tiempo y pueden influir entre sí.

Permite modelar y predecir series temporales especificando las relaciones lineales entre las variables de interés a lo largo del tiempo, teniendo en cuenta la autocorrelación y la dependencia entre las observaciones sucesivas.

## **NLME**

El paquete "nlme" (Non-Linear Mixed Effects) es una importante herramienta para el análisis de datos longitudinales y correlacionados. Permite la modificación de modelos mixtos no lineales, lo que le permite manejar situaciones en las que los efectos varían tanto entre sujetos como a lo largo del tiempo. Esto es particularmente útil en estudios en los que se recopilan mediciones repetidas en el tiempo de los mismos individuos, y donde es crucial considerar la estructura de correlación inherente en los datos.

El paquete "nlme" también proporciona métodos para estimar parámetros, inferencia estadística, diagnóstico de ajuste de modelos y visualización de resultados, además de permitir la especificación de efectos fijos y aleatorios en los modelos. Los estudios médicos para evaluar la eficacia de los tratamientos a lo largo del tiempo, las investigaciones en ciencias sociales para comprender cómo cambian las actitudes o los comportamientos, y el análisis en biología para modelar el crecimiento de organismos a lo largo de múltiples mediciones son algunos ejemplos comunes de aplicación.

## **PROPHET**

Prophet es un procedimiento de predicción que se puede usar tanto en R como en Python. Es rápido y proporciona previsiones totalmente automatizadas que los científicos de datos y analistas pueden ajustar manualmente. Se pueden hacer pronósticos y predicciones de series temporales con facilidad y precisión con él. Este paquete modela las tendencias y los efectos estacionales en los datos de series temporales utilizando un enfoque aditivo. Puede administrar datos que incluyen cambios en la tendencia, efectos de días festivos y otros eventos especiales, además de factores estacionales. El "prophet" es más beneficioso que algunos otros métodos de predicción porque puede lidiar con datos ausentes y faltantes.

## **GRIDEXTRA**

Proporciona una serie de funciones a nivel de usuario para trabajar con gráficos de "cuadrícula", en particular para organizar múltiples gráficos de cuadrícula en una página y dibujar tablas. Esto permitirá observar de manera gráfica todos los aspectos de la serie de estudio a tratar y cómo cambian de acuerdo con los ajustes que se le implementen.

## **GGPUBR**

El paquete "ggpubr" de RStudio es una herramienta que permite mejorar y ajustar las visualizaciones creadas con el paquete "ggplot2." Debido a su flexibilidad y capacidad para generar visualizaciones de alta calidad, "ggplot2" es ampliamente utilizado para crear gráficos en R. Sin embargo, en ocasiones se requiere un mayor control y una presentación más sofisticada de los resultados gráficos.

El paquete "ggpubr" se utiliza para agregar capas adicionales, temas, etiquetas, estadísticas de resumen y otras características a las trazas generadas con "ggplot2." Además, puede crear visualizaciones comunes como gráficos de barras, gráficos de cajas, gráficos de dispersión, gráficos de líneas y más con detalles adicionales para mejorar la legibilidad y la presentación.

## **GGSTATSPLOT**

El paquete `{ggstatsplot}` del lenguaje de programación R proporciona una sintaxis de una sola línea para enriquecer las visualizaciones basadas en `{ggplot2}` con los resultados del análisis estadístico-incrustados en la propia visualización. De este modo, el paquete ayuda a los investigadores a adoptar un flujo de trabajo de exploración e información de datos riguroso, fiable y sólido.

## Líneas de Códigos

### Flores <- Read\_excel

La función Read Excel se utiliza para leer y cargar datos de hojas de cálculos de Excel (.xlsx), y se nombró con el objeto “Flores”.

### Flores1 <- Flores[, -c(1)]

En el lenguaje R existe la posibilidad de almacenar información en un “objeto”, con la finalidad de realizar procedimientos estadísticos y plantear gráficas sin ser necesario repetir toda una línea de código, de tal manera se le crea el objeto “Flores1” y se le asigna los datos de “Flores” mediante el uso del operador “<-”. Por consiguiente “Flores1” es una nueva base de datos, luego de que se procede a extraer un elemento del vector usando los corchetes de tal manera “fecha” fue eliminada de la base mediante la función Flores[, -c(1)], donde 1 representa la primera columna de la base Flores1.

### Series <- ts(Flores1, start = c(1990,1), end = c(2023,5), frequency = 12)

Se convierte el objeto “Series” a serie temporal con la función “ts()”. La construcción de objetos de la clase *ts* requiere que los datos de partida estén distribuidos regularmente en la escala temporal es decir que haya siempre un dato por día, o dos datos por mes (Rpubs, 2023).

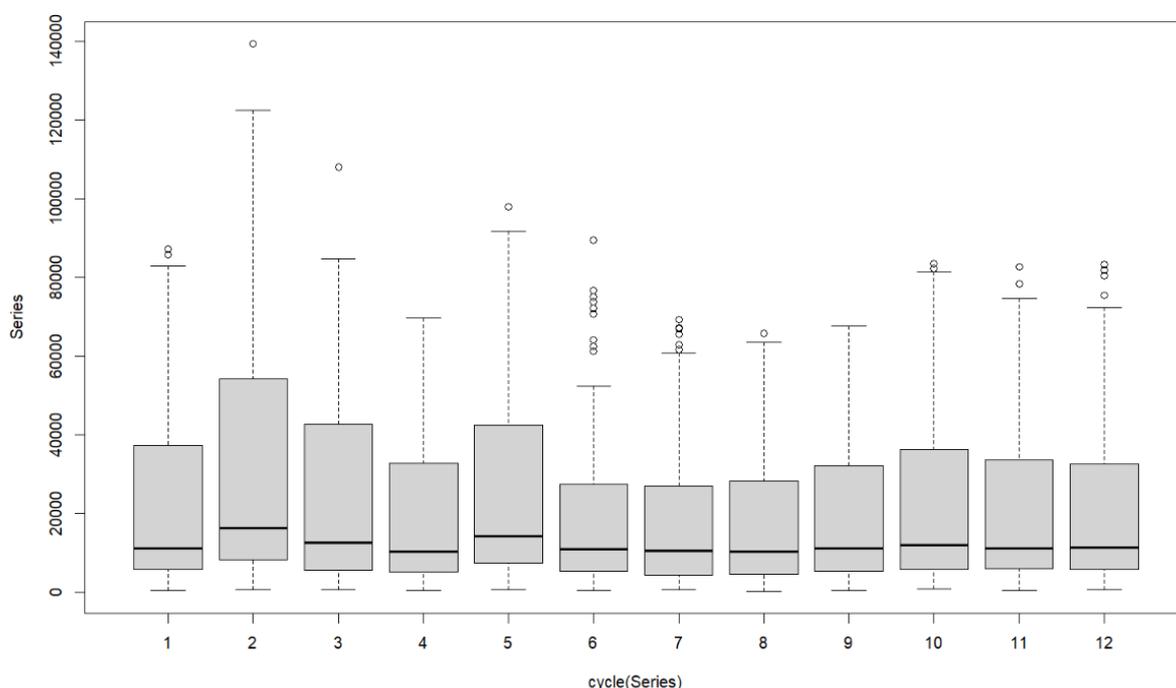
*Start* indica desde qué año comienza la serie de tiempo y la función “c(1990,1)” se encarga de crear un vector numérico. Llamamos esta función y le damos como argumento los elementos que deseamos combinar en un vector, separados por comas. El 1 indica en que momento o secuencia inicia por otro lado *end* indica en qué año finaliza la misma c(2023,5) donde el 5 indica en que secuencia de ese año termina. El argumento frequency = 12 indica la frecuencia o periodicidad que habrá a lo largo de cada año, en este caso se utiliza 12 por los meses que contiene un año.

### Plot1 <- boxplot(Series~cycle(Series))

En la función del objeto Plot1, en esta línea se está creando un gráfico de caja “BoxPlot” para visualizar la distribución de los valores en función del ciclo de la serie temporal. Esta consiste en una caja que abarca el rango intercuartílico, con una línea en medio que muestra la mediana. Los “bigotes” se extienden desde la caja para señalar la dispersión de la mayoría de los datos y los puntos fuera de los bigotes pueden indicar valores atípicos. Este gráfico refleja una visión general de la simetría de la distribución de los datos; si la mediana no está en el centro del diagrama de caja se sobreentiende que la distribución no es simétrica.

**Figura 9.**

**Distribución de valores mensuales del año 1990 a 2023 de las exportaciones de flores del Ecuador.**



Un diagrama de caja es un método estandarizado para representar gráficamente una serie de datos numéricos a través de sus cuartiles. De esta manera, se muestran a simple vista la mediana y los cuartiles de los datos y cada cuartil representa el 0.25 de la unidad, y también pueden representarse sus valores atípicos. En la figura 8 se observa que a lo largo de los años las exportaciones mensuales de flores han mantenido una mediana próxima a los \$10.000 USD y esta tiende a ser constante durante los meses de julio a diciembre y en los primeros 5 meses del año existe una varianza notable y además se denota valores extremadamente atípicos en el mes de febrero.

```
plot2 <-plot(Series, ylab="Precio", main="Toneladas y precio FOB de Flores")
```

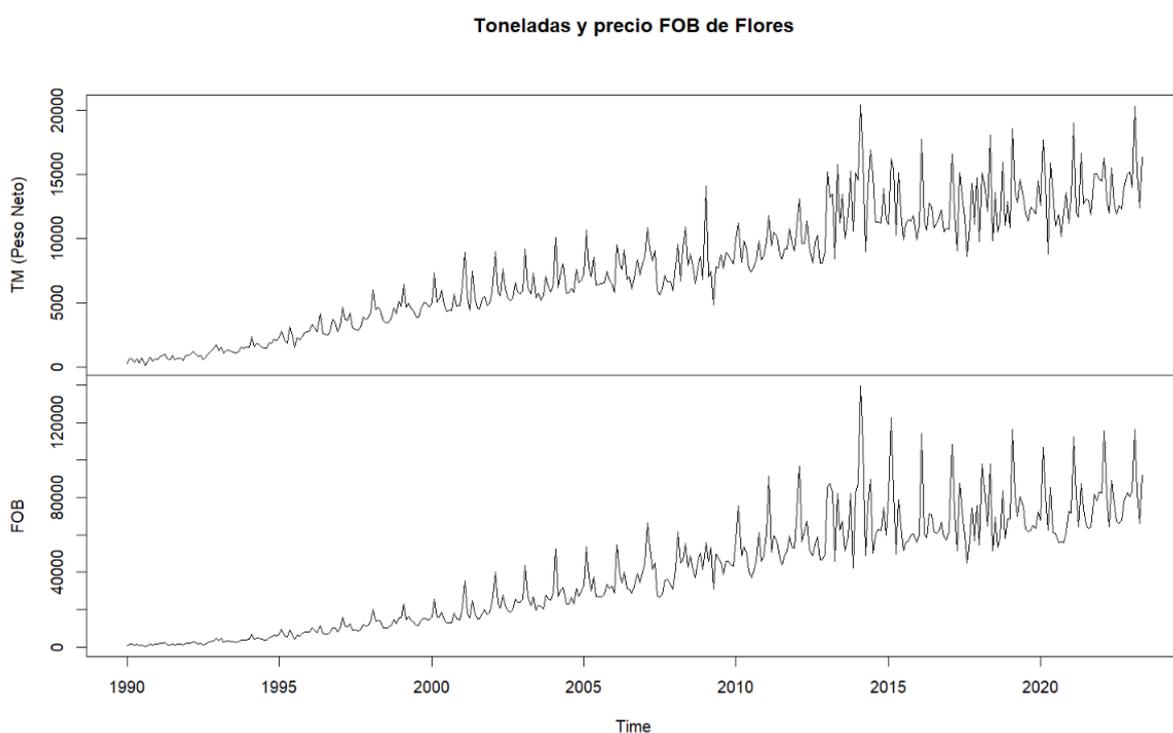
En la función del objeto Plot2, está siendo utilizada para crear un gráfico de línea simple utilizando el objeto de la serie temporal “Series”.

En Ylab = “Precio”, esta línea establece la etiqueta “precio” en el eje Y del gráfico.

En main = “Toneladas y precio FOB de Flores”, esta línea establece la etiqueta “Toneladas y precio FOB de Flores” en el eje X del gráfico.

### Figura 10.

### Evolución del peso neto en Toneladas y FOB de las exportaciones de flores del Ecuador del año 1990 a 2023.



Este gráfico ilustra la evolución de las exportaciones de flores del Ecuador en términos de valor FOB (Free On Board), y peso neto a lo largo del periodo comprendido entre 1990 y 2023. El valor FOB representa el valor de las mercancías en el punto de embarque, excluyendo los costos de transporte y seguro. Además, es importante destacar que el crecimiento en el peso neto de las exportaciones demuestra la capacidad de la industria para abastecer la demanda creciente en los mercados

internacionales. Por lo que, la tendencia a lo largo de estos años es claramente ascendente, evidenciando el crecimiento constante de la industria de flores en el país.

**par(mfrow= c(2,2))**

Este parámetro sirve para dividir la ventana gráfica en forma de matriz para almacenar en cada celda un gráfico diferente. La forma para modificar este parámetro es `mfrow=c(nf, nc)`, donde **nf es el número de filas** y **nc el número de columnas** en las cuales se va a dividir la ventana gráfica (Github, 2023)

**modelo <- auto.arima(Series[,c(1)], seasonal=T, stepwise=F, approximation=T)**

La función `auto.arima()` se utiliza para ajustar automáticamente un modelo ARIMA. La especificación “auto” significa que la función buscará automáticamente el mejor modelo de ARIMA.

en `Series[,c(1)]` está seleccionando la primera columna de la serie temporal del objeto “Series”.

`Seasonal = T`, al establecer “T” de “TRUE” significa que establece que el modelo ARIMA ajustado debe incluir componentes estacionales.

`Stepwise = F`, el `stepwise` controla si el ajuste del modelo se realiza en un proceso paso a paso. Al establecer F de “False”, se estaría evitando el enfoque.

`approximation=T`, al establecer “T” de “TRUE” esto indica que se utiliza una aproximación al ajustar el modelo ARIMA.

**summary(modelo)**

La función “summary” muestra un resumen general sobre las variables de la data frame (mínimo, máximo, media, mediana, primer y tercer cuartil). Además, muestra la frecuencia de cada categoría.

**Tabla 1.****Coefficientes de TM**

Coeficientes					
	AR1	MA1	MA2	SMA1	DRIFT
	0.9347	-0.9034	0.1231	-0.7711	36.3222
S.E	0.0283	0.0586	0.0535	0.0387	4.3845
	Sigma 2=	Log			
	1560711	Likelihood = -			
		3328.4			
	AIC =	AICc =	BIC =		
	6668.8	6669.02	6692.58		

**Tabla 2.****Datos entrenamiento de TM**

Training							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
	-	1222.516	785.4344	-	10.6407	0.71812	0.004911
	2.6798			2.4981			

**ResidualTM <- checkresiduals(modelo)**

La función `checkresiduals` toma el modelo y realiza pruebas y visualizaciones de los residuos para verificar si cumplen con ciertas propiedades deseables. En esta prueba se validan estadísticos de Ljung-Box, Ruido Blanco, función de autocorrelación parcial (ACP) y simple (ACF).

```
checkresiduals(object, lag, test, plot = TRUE, ...)
```

`object`

Un modelo de serie temporal, un objeto de previsión o una serie temporal (se supone que son residuos).

## Lag

Número de rezagos a utilizar en la prueba de Ljung-Box o Breusch-Godfrey. Si falta, se fija en  $\min(10, n/5)$  para datos no estacionales, y  $\min(2m, n/5)$  para datos estacionales, donde  $n$  es la longitud de la serie, y  $m$  es el período estacional de los datos. Además, debe ser al menos  $df+3$ , donde  $df$  son los grados de libertad del modelo. De este modo se garantiza que haya al menos 3 grados de libertad en la prueba de chi-cuadrado.

## Test

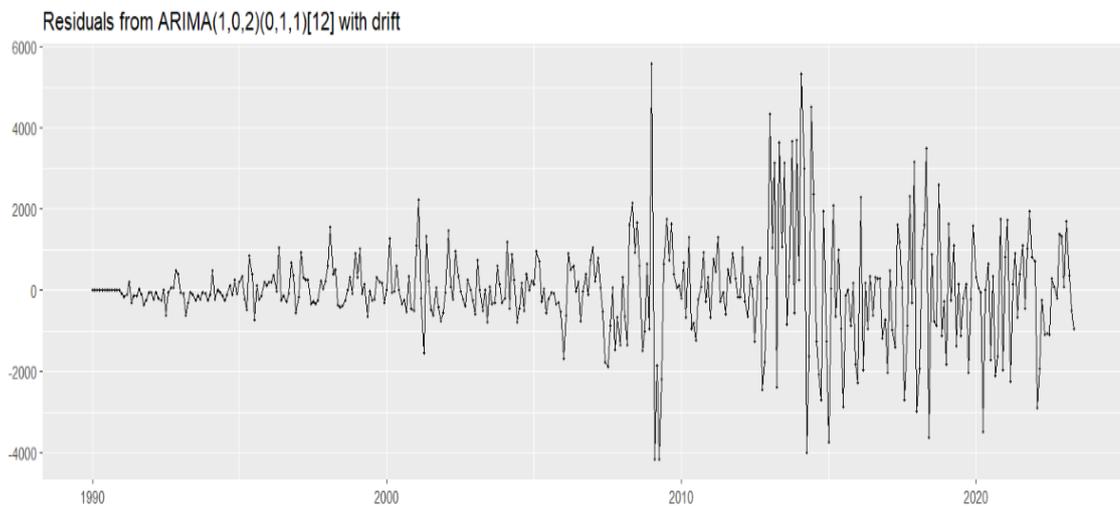
Prueba para utilizar la correlación serial. Por defecto, si el objeto es de clase `lm`, entonces `test="BG"`. En caso contrario, `test="LB"`. Si se establece `test=FALSE`, no se imprimirán los resultados de la prueba.

## Plot

Si es `TRUE`, producirá el gráfico.

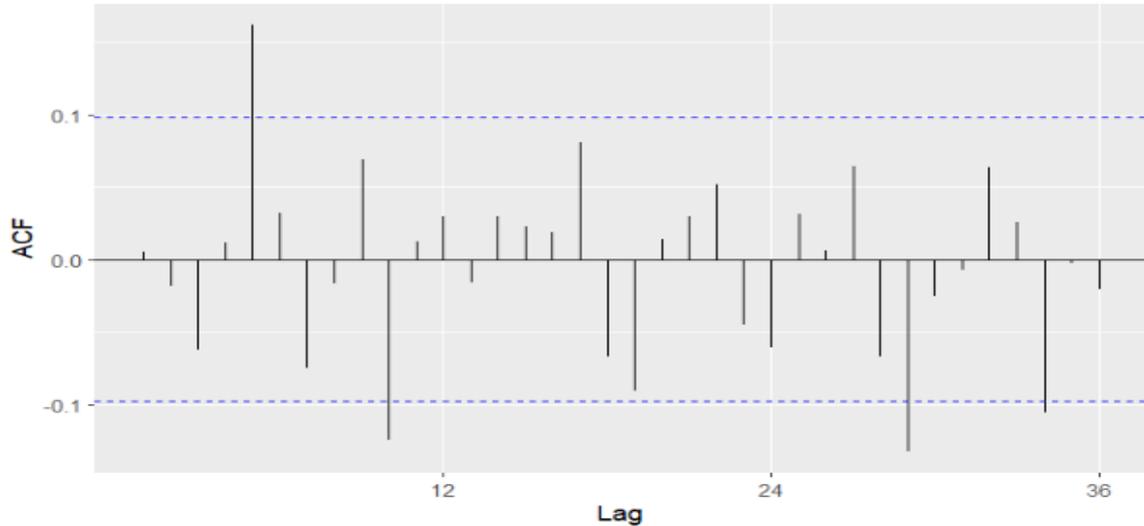
### Figura 11.

#### Residuos de ARIMA



**Figura 12.**

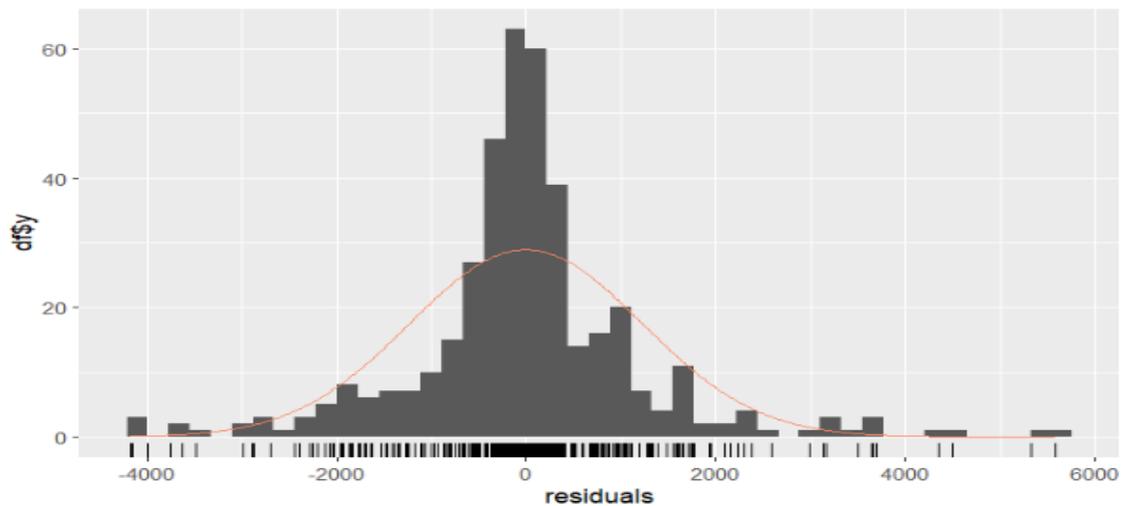
**Función de autocorrelación simple de los residuos (ACF)**



Función de autocorrelación de los residuos: La función de autocorrelación de los residuos se utiliza para identificar patrones de autocorrelación en los residuos. Si no hay correlaciones significativas en la función de autocorrelación de los residuos, se puede inferir que el modelo cumple con el supuesto de que los residuos son independientes. Para los valores que superan los rangos del ACF en el modelo, indica una correlación significativa en ese punto de tiempo y puede ser un indicador de que se necesitan ajustes adicionales en el modelo, además se observa 4 valores que están fuera de los rangos establecidos, desfase con correlación significativa positiva en el desfase 5 seguida de una onda que alterna entre positiva y negativa y con correlación negativa en el desfase 10,29 y 34. La interpretación de los resultados del ACF debe hacerse en conjunto con otros estadísticos y gráficos para obtener una evaluación completa del modelo ARIMA.

**Figura 13.**

**Histograma de los residuos**



Valores p de los estadísticos de chi-cuadrada de Ljung-Box: Los valores p de los estadísticos de chi-cuadrada de Ljung-Box se utilizan para evaluar la autocorrelación de los residuos. Si todos los valores p son mayores que 0.05, se puede concluir que los residuos son independientes y no hay autocorrelación significativa.

**Tabla 3.**

**Ljung-Box Test de TM**

---

Ljung-Box Test		
Data: Residuals from		
ARIMA (1,0,2)(0,1,1)		
with drift		
Q = 37.203	Df = 20	p-value = 0.01106
Model df: 4	Total lag used: 24	

---

**ForecastTM <- forecast(modelo, h=36)**

La función “Forecast” retorna predicciones e información acerca de modelos ARIMA univariantes, esta toma los datos del objeto “modelo” y “h” que es el número de periodos a predecir. En este estudio se estableció generar un pronóstico para los próximos 36 meses en la serie temporal.

## ForecastTM

La función del objeto “ForecastTM” se visualizará mediante una tabla de los datos de peso pronosticados para los siguientes 36 meses.

## ForecastTMPlot <- autoplot(ForecastTM)

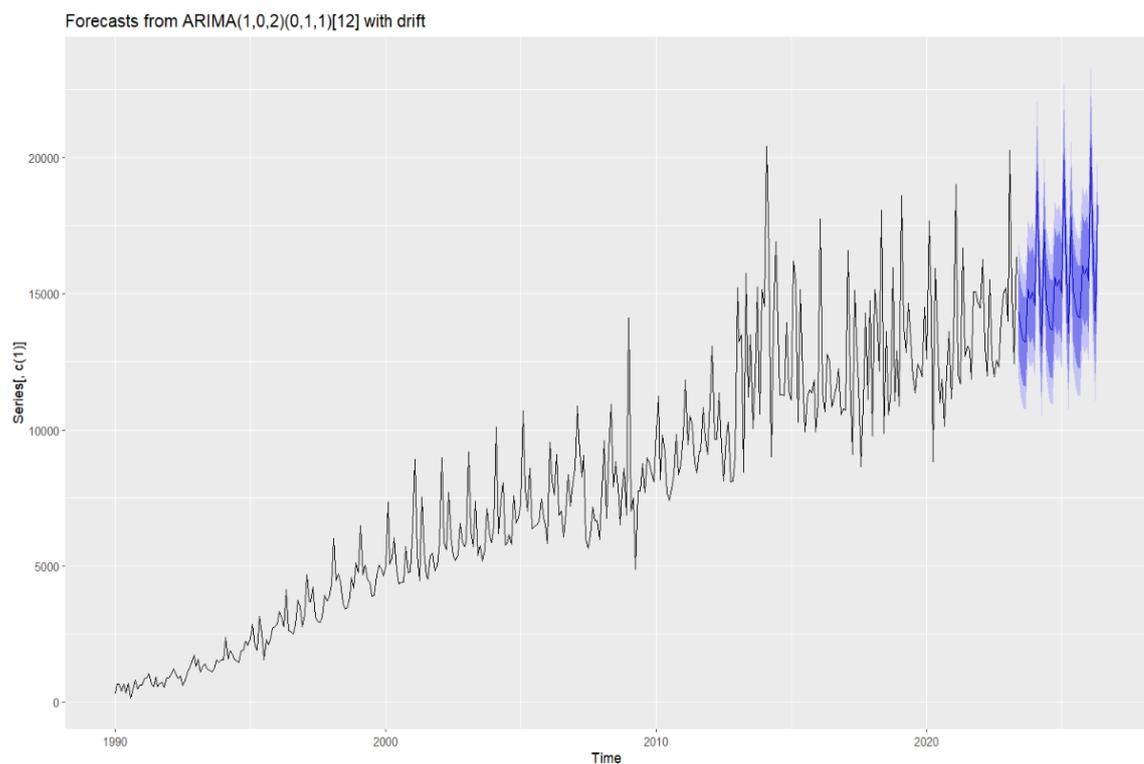
La función “autoplot” toma el objeto “ForecastTM” y crea un gráfico a partir del objeto para visualizar los pronósticos.

## ForecastTMPlot

La función del objeto “ForecastTMPlot” se visualizará la gráfica con los pronósticos creado en la función anterior

## Figura 14.

### Proyección del peso neto de las exportaciones de flores del Ecuador.



A lo largo de las décadas pasadas, las exportaciones de flores ecuatorianas han demostrado un crecimiento constante en términos de peso neto, lo que indica la capacidad de la industria para satisfacer la demanda internacional de manera constante y en aumento. La proyección sugiere que esta tendencia de crecimiento en el peso neto de las exportaciones se mantendrá. (Valor en toneladas)

**modelo2 <- auto.arima(Series[,c(2)], seasonal=T, stepwise=F, approximation=T)**

La función `auto.arima()` se utiliza para ajustar automáticamente un modelo ARIMA. La especificación “auto” significa que la función buscará automáticamente el mejor modelo de ARIMA.

en `Series[,c(2)]` está seleccionando la segunda columna de la serie temporal del objeto “Series”.

`Seasonal = T`, al establecer “T” de “TRUE” significa que establece que el modelo ARIMA ajustado debe incluir componentes estacionales.

`Stepwise = F`, el `stepwise` controla si el ajuste del modelo se realiza en un proceso paso a paso. Al establecer F de “False”, se estaría evitando el enfoque.

`approximation=T`, al establecer “T” de “TRUE” esto indica que se utiliza una aproximación al ajustar el modelo ARIMA.

**summary(modelo2)**

La función “summary” muestra un resumen general sobre las variables de la data frame (mínimo, máximo, media, mediana, primer y tercer cuartil). Además, muestra la frecuencia de cada categoría.

**Tabla 4.**

**Coefficientes de FOB**

Coefficientes						
	AR1	AR2	MA1	SMA1	SMA2	DRIFT
	0.6292	0.2246	-0.6001	-	-0.157	208.98
				0.4950		
S.E.	0.1197	0.0681	0.1146	0.0555	0.051	29.3957
	Sigma 2=		Log			
	47751178		Likelihood=			
			-3990.91			
	AIC=	AICc=	BIC=			
	7995.81	7996.11	8023.56			

**Tabla 5.**

**Datos entrenamiento de FOB**

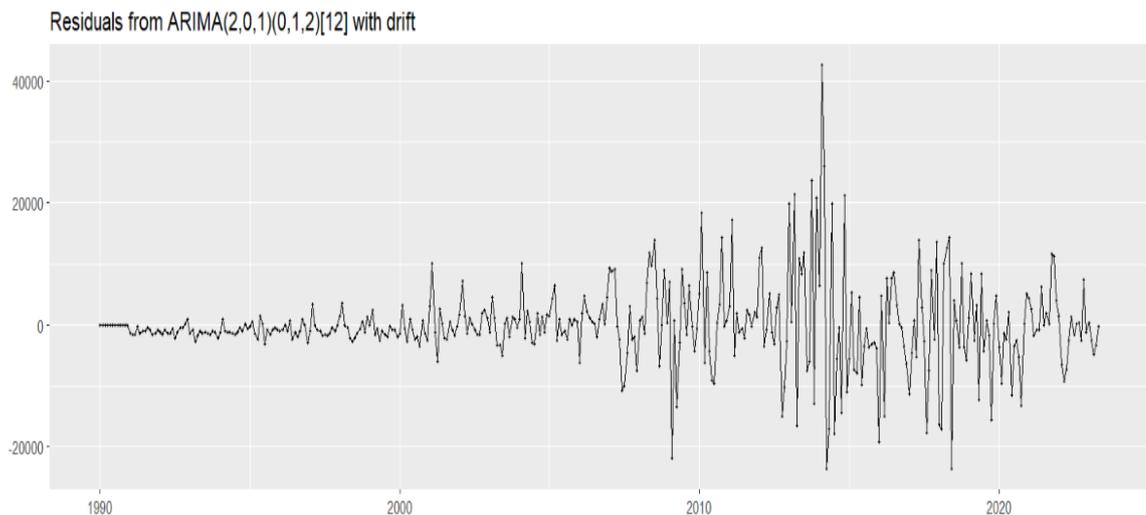
Training							
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	
-	6753.53	4101.393	-	14.7794	0.7454	0.007267	
59.1532			7.7273				

**ResidualFOB <- checkresiduals(modelo2)**

La función checkresiduals toma el modelo y realiza pruebas y visualizaciones de los residuos para verificar si cumplen con ciertas propiedades deseables.

**Figura 15.**

**Residuos de ARIMA**

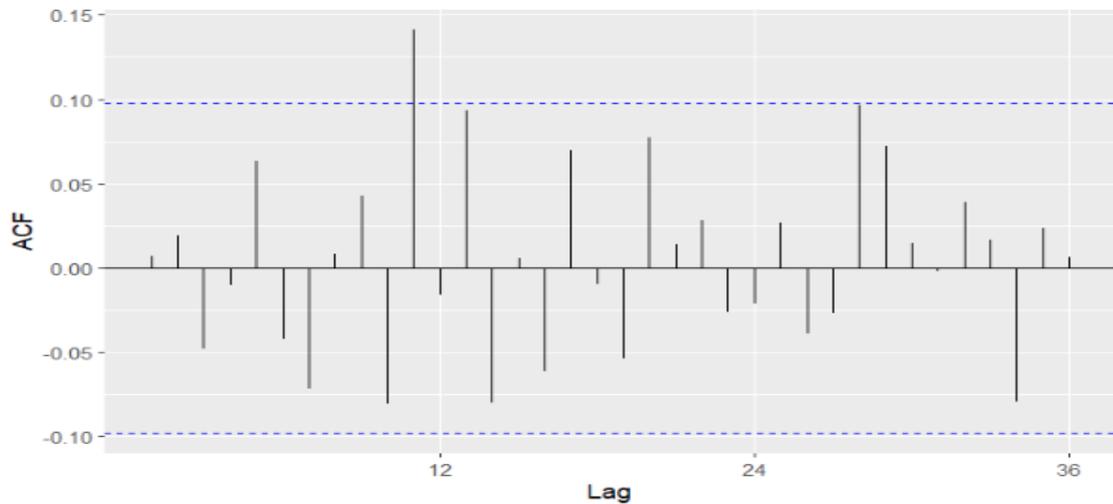


Utilice la gráfica de residuos vs. orden para verificar el supuesto de que los residuos son independientes entre sí. Los residuos independientes no muestran tendencias ni patrones cuando se muestran en orden cronológico. Los patrones en los puntos podrían indicar que los residuos que están cercanos entre sí podrían estar correlacionados y, por lo tanto, podrían no ser independientes. Lo ideal es que los

residuos que se muestran en la gráfica se ubiquen aleatoriamente alrededor de la línea central:

**Figura 16.**

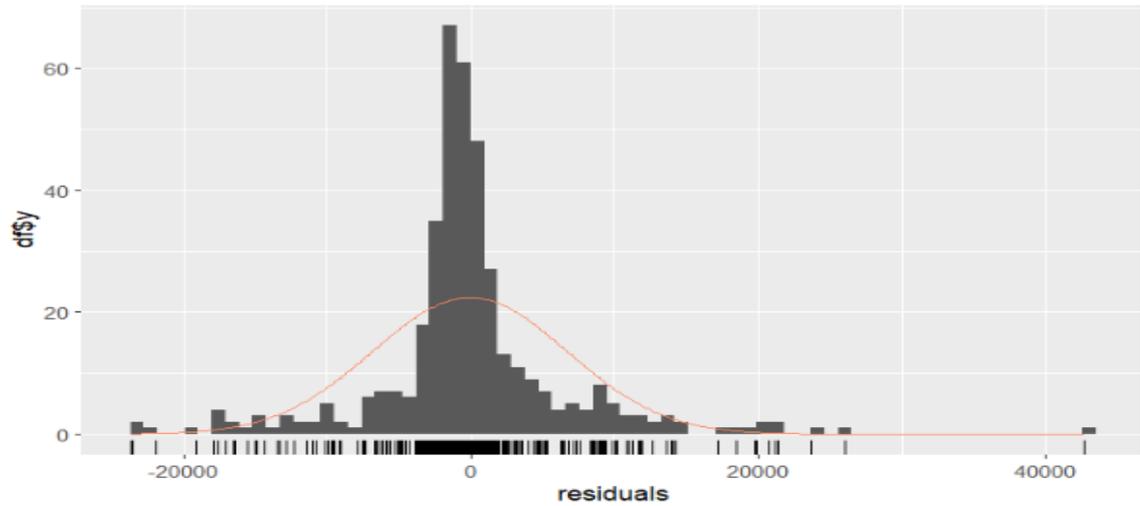
**Función de autocorrelación simple de los residuos (ACF)**



La función de autocorrelación de los residuos se utiliza para identificar patrones de autocorrelación en los residuos. Si no hay correlaciones significativas en la función de autocorrelación de los residuos, se puede inferir que el modelo cumple con el supuesto de que los residuos son independientes. Para los valores que superan los rangos del ACF en el modelo, indica una correlación significativa en ese punto de tiempo y puede ser un indicador de que se necesitan ajustes adicionales en el modelo. En el modelo se observa únicamente un valor fuera del rango del ACF en el desfase 11. La interpretación de los resultados del ACF debe hacerse en conjunto con otros estadísticos y gráficos para obtener una evaluación completa del modelo ARIMA.

**Figura 17.**

**Histograma de los residuos**



El histograma de residuos es una herramienta esencial en el análisis de regresión y modelado estadístico que nos permite evaluar si los residuos del modelo se distribuyen de manera aproximada a una distribución normal alrededor de cero. En este caso, el histograma de residuos en términos de FOB muestra claramente que los residuos están muy cercanos a cero, lo que sugiere un buen ajuste del modelo a los datos observados.

**Tabla 6.**

**Ljung-Box Test de FOB**

---

Ljung-Box Test		
Data: Residuals from		
ARIMA(2,0,1)(0,1,2) with		
drift		
Q = 32.084	Df = 19	p-value = 0.03058
Model df: 5	Total lag used: 24	

---

**ForecastFOB <- forecast(modelo2, h=36)**

La función “Forecast” que toma los datos del “modelo2” para generar pronósticos para los próximos 36 meses en la serie temporal.

**ForecastFOB**

La función del objeto “ForecastFOB” se visualizará mediante una tabla de los datos de exportación de flores pronosticados para los siguientes 36 meses.

**ForecastFOBPlot <- autoplot(ForecastFOB)**

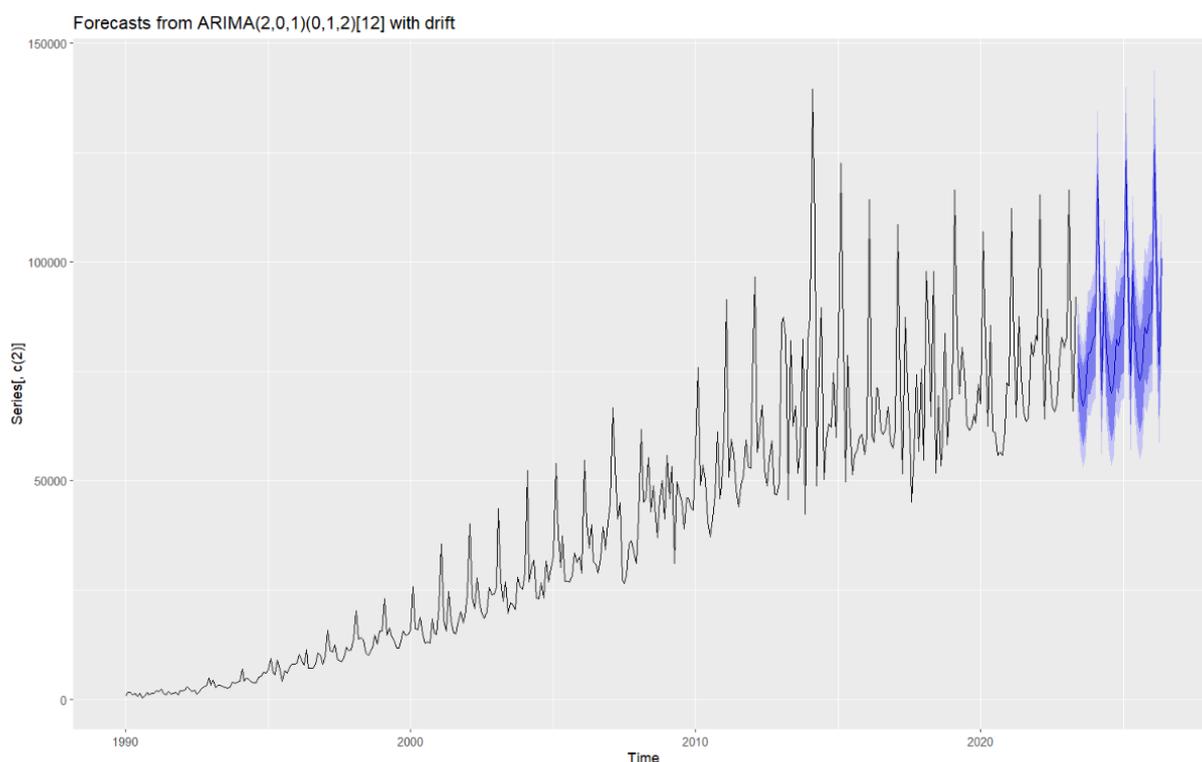
La función “autoplot” toma el objeto el objeto “ForecastFOB” y crea un gráfico para visualizar los pronósticos.

**ForecastFOBPlot**

La función del objeto “ForecastFOBPlot” se visualizará la gráfica con los pronósticos creado en la función anterior.

**Figura 18.**

**Proyección de FOB de las exportaciones de flores del Ecuador.**



El gráfico de proyección de las exportaciones de flores del Ecuador en términos de valor FOB (Free On Board) desde 1990 hasta 2023 presenta una tendencia creciente a lo largo del periodo. Es importante reconocer que las proyecciones están sujetas a incertidumbres y cambios en el entorno económico y comercial. Sin embargo, en función de los datos históricos y la dirección general de la industria de flores en el Ecuador, esta proyección sugiere que la tendencia creciente en las exportaciones de flores se mantendrá en los próximos tres años. (Valor en \$)

**Conclusión de análisis de datos**

Se fijaron como objetivos principales de la investigación la implementación del modelo ARIMA y la evaluación de su capacidad para predecir las exportaciones de flores en el Ecuador. Para abordar estos objetivos, se siguió un enfoque sistemático que involucró la recopilación y preparación de datos, la selección y ajuste del modelo ARIMA más adecuado, y la interpretación de los resultados obtenidos. Se utilizaron datos históricos obtenidos del Banco Central del Ecuador de exportaciones de flores

en Ecuador, abarcando un período suficientemente amplio (Enero 1990- Mayo 2023) con el propósito de capturar variaciones estacionales y tendencias a largo plazo. Estos datos fueron sometidos a un proceso de preprocesamiento que incluyó la identificación y corrección de valores atípicos, la transformación de datos no estacionarios y la división en conjuntos de entrenamiento y prueba.

El modelo ARIMA fue implementado a través de pasos de identificación, estimación y diagnóstico. Se realizaron análisis de autocorrelación y autocorrelación parcial para determinar los órdenes adecuados para el modelo ARIMA. Luego, el modelo fue ajustado a los datos de entrenamiento y validado utilizando los datos de prueba. Los resultados revelaron una capacidad prometedora del modelo ARIMA para predecir las exportaciones de flores en el Ecuador, mostrando un ajuste razonable a los datos observados y una capacidad para capturar patrones estacionales.

Se observa que las exportaciones de flores han tenido y sigue teniendo una tendencia creciente a lo largo de los años, además se refleja en la gráfica (1) que sigue un patrón de varianza constante en sus límites inferiores y superiores que crece de igual manera. El enfoque de esta investigación es notable en la aplicación de métodos estadísticos avanzados en la predicción económica, específicamente mediante la utilización del modelo ARIMA. La elección de este modelo refleja una comprensión profunda de las necesidades de predicción en un mercado tan dinámico como el de las exportaciones de flores.

Gracias a la implementación del software y del lenguaje `r` resalta que las exportaciones alcanzan el límite superior en los meses de febrero de cada año a lo largo de la serie de tiempo y se observa esto en los valores atípicos que son extremadamente lejanos a la mediana del gráfico (). Los procesos llevados a cabo ARMA permitió la elaboración del modelo autorregresivo.... con el cual se aplica en el programa y este nos arroja 2 modelos de forecast; uno para el peso neto en Toneladas pronosticadas a exportar,  $(1,0,2)(0,1,1)$  y otro modelo para el valor en dólares americanos pronosticado  $(2,0,1)(0,1,2)$ ; además se enfatiza la comparativa de los residuos, funciones de autocorrelación y los estadísticos de Ljung-Box para la elección y evaluación del modelo estadístico a implementar para el pronóstico final

Después de analizar los datos, se encontró que la exportación de flores en los futuros años seguirá un patrón creciente. Además, se concluye que el modelo ARIMA es adecuado para predecir la exportación de flores en Ecuador y que los resultados obtenidos son útiles para la toma de decisiones en el sector de la floricultura.

El modelo ARIMA se presenta como una herramienta efectiva para predecir las exportaciones de flores. Su capacidad para identificar patrones estacionales y tendencias en los datos temporales lo convierte en una herramienta útil en la industria de las flores. Al capturar las fluctuaciones estacionales en la demanda y los patrones de crecimiento a lo largo del tiempo, el modelo autorregresivo integrado de medias móviles otorga información valiosa para la gestión y planificación de las exportaciones de flores.

Es importante considerar que dicho modelo puede enfrentar desafíos si las exportaciones de flores están influenciadas por factores externos impredecibles o eventos inusuales. En tales casos, combinar ARIMA con otros enfoques o datos adicionales podría mejorar aún más la precisión de las predicciones.

## **Conclusión**

Esta investigación explora un tema estrechamente relacionado con el contexto económico y agrícola de Ecuador explorando el uso de modelos estadísticos ARIMA para pronosticar las exportaciones de flores, uno de los sectores clave de la economía nacional. El objetivo del estudio es determinar el volumen de exportación de flores (en toneladas) y el valor monetario (FOB) en los próximos años mediante la implementación de un modelo estadístico denominado ARIMA en conjunto con el software de programación R-Studio.

Para comprender y captar mejor la investigación, se lleva a cabo un estudio detallado y una revisión de la literatura de todos los conceptos del modelo estadístico autorregresivo integrado de media móvil, al tiempo que se complementa con diversos estudios que han aplicado estas mismas técnicas en la práctica.

Mediante la revisión literaria de la investigación presente se ha alcanzado cumplir con el objetivo de comprender la metodología del modelo autorregresivo integrado de media móvil para la aplicación de un modelo predictivo, gracias a un

estudio y análisis exhaustivo se logró abordar la metodología del modelo ARIMA y sus procesos matemática y estadísticos que el misma ejecuta con la finalidad de precisar y aplicar un modelo que detecta patrones o tendencias en una serie de tiempo respecto a una base de datos históricas, esto permite que a través del factor de diferenciación de la serie a estudiar, se ajuste de no estacionaria a estacionaria con el objetivo de arrojar los valores más certeros para una predicción. En base a la teoría y la aplicación práctica de este enfoque se logró resultados importantes que aportan de gran manera al área de análisis de series de tiempo y a la predicción. La comprensión de la metodología fue clave para la implementación del modelo predictivo en este estudio ya que además se adquirió la capacidad de modelar, ajustar y evaluar los modelos de predicción.

La comprensión de los fundamentos de la metodología del modelo ARIMA en conjunto de la aplicación del programa de R-Studio fueron clave para cumplir el objetivo final de esta investigación el cuál estipula analizar y describir los resultados del modelo de predicción ARIMA. La implementación de la herramienta informática permitió la construcción del modelo predictivo de una manera rápida y eficaz, la herramienta además permite observar valores estadísticos, visualización de resultados, gráficas, histogramas, funciones de correlación, análisis de residuos que fueron claves para analizar la capacidad predictiva de los modelos elaborados. En conclusión, la combinación de herramientas de software especializadas como R-Studio y un sólido conocimiento estadístico nos permitió no solo lograr con éxito nuestros objetivos, sino también desarrollar una comprensión profunda y crítica de la utilidad y relevancia de los modelos ARIMA como herramienta predictiva para determinar las futuras exportaciones del sector florícola en el Ecuador.

### **Recomendaciones**

Las implicaciones de los resultados son significativas para los actores involucrados en el sector de exportaciones de flores en Ecuador. Las predicciones generadas por el modelo ARIMA pueden proporcionar una guía valiosa para la toma de decisiones y la planificación estratégica en un mercado caracterizado por su volatilidad.

Para las futuras investigaciones y/o estudios que deseen realizar pronósticos ya sea en la misma área o diferente, la herramienta de R-Studio resulta útil para obtener resultados óptimos, sin embargo, se debe de tener conocimientos previos tanto en estadística descriptiva como en inferencial sin dejar de lado el aprendizaje del lenguaje del programa. La implementación de estos conocimientos en el ámbito profesional permitirá mejorar la capacidad de exportación y así mismo la de producción de las empresas floricultoras del país. Se recomienda hacer énfasis en la capacitación de conocimientos estadísticos sin dejar de lado la adecuada inducción hacia las técnicas de inteligencia de negocios.

Basándose en los resultados obtenidos, se recomienda que los actores del sector de exportaciones de flores en Ecuador consideren la incorporación del modelo ARIMA como una herramienta complementaria en su proceso de toma de decisiones. Sin embargo, es importante reconocer las limitaciones inherentes al modelo ARIMA, como su sensibilidad a cambios abruptos y eventos extraordinarios que podrían no estar reflejados en los datos históricos, se debe utilizar con precaución y complementar con otras fuentes de información, especialmente en momentos de cambios económicos o situaciones excepcionales que puedan afectar las predicciones.

## Referencias

- Abril, C. M., Núñez, L. I., Briceño, J. P., & Robles, J. C. (2022). La competitividad de las exportaciones florícolas del Ecuador con Colombia hacia el mercado de los Estados Unidos. *Investigación y Desarrollo*, 7-15.
- Ade, P., Aguirre, J., Ahmed, Z., Aiola, S., Ali, A., Alonso, D., y Zhu, N. (2019). The Simons Observatory: Science Goals and Forecasts. *J. Cosmol. Astropart. Phys.*
- Aidin Tamhidi, N. M. (2022). Uncertainty quantification of ground motion time series generated at uninstrumented sites. *Sage Journals*.
- Albarak, M., Bahsoon, R. (2018). Prioritizing Technical Debt In Database Normalization Using Portfolio Theory and Data Quality Metrics.
- Alnsour, M. I. (2023). Specifying a cascade water demand forecasting model using time-series analysis: a case of Jordan. *Water Resour.*
- Altamirano, J. (2022). Modelamiento y predicción del efecto COVID-19 en el sistema laboral ecuatoriano. 1-18.
- Amaris, G., Ávila, H., & Guerrero, T. (2017). Aplicación de modelo ARIMA para el análisis de series de volúmenes anuales en el río Magdalena. *Scielo*, 88-101. <https://doi.org/10.14483/udistrital.jour.tecnura.2017.2.a07>
- Angwin, A., Wilson, W., Arnott, W., Signorini, A., Barry, R., Copland, D. (2017). White Noise Enhances New-word Learning In Healthy Adults. *Sci Rep*, 1(7).
- Antonino Parisi Fernández, L. A. (2019). Predicción de variaciones en el precio del petróleo con el modelo de optimización ARIMA, innovando con fuerza bruta operacional. *Tec Empresarial*, 53-70.
- Arnold, H., Moroz, I., Palmer, T. (2013). Stochastic Parametrizations and Model Uncertainty In The Lorenz '96 System. *Phil. Trans. R. Soc. A.*, 1991(371).
- Asmita Pandey, R. S. (2021). *International Journal for Research in Applied Science & Engineering Technology*.
- Ayala Mora, M. E. (2014). *Historia, tiempo y conocimiento del pasado: estudio sobre periodización general de la historia ecuatoriana; una interpretación interparadigmática*. Quito: Corporación Editorial Nacional.
- Ayala, R. (2021, Noviembre 30). Series Temporales, Op. 1: ARIMA. *Data Science Research Perú*.
- Barragán, A. (2014). Ciencia, tecnología e innovación en el desarrollo de México y América Latina. *Desafíos de la ciencia la tecnología y la innovación*.

- Desarrollo, educación y trabajo, tomo I, María del Carmen del Valle, Ana Mariño e Ismael Núñez. *Problemas del desarrollo*.
- Beard, E., Marsden, J., Brown, J., Tombor, I., Stapleton, J., Michie, S., ... & West, R. (2019). Understanding and Using Time Series Analyses In Addiction Research. *Addiction*, 10(114), 1866-1884.
- Bell, M., Samet, J. (2004). Time-Series Studies of Particulate Matter. *Annual Reviews*.
- Ben D. Fulcher, M. A. (2013). Highly comparative time-series analysis: the empirical structure of time series and their methods. The Royal Society.
- Bernhard Wierder, M. L. (2015). The impact of Business Intelligence on the quality of decision making - a mediation model. *Conference on ENTERprise Information Systems*, 1163-1171.
- Box, G. E. P. and Jenkins, G. 1976. Time series analysis: forecasting and control. Ed. Holden-day. San Francisco, CA. 575 p.
- Cedeño Troya, F. (2022). Modelos predictivos de sistemas de información en la gestión de abastecimientos del sector ferretero. *Revista Ciencia y Tecnología*.
- Cedillo Villavicencia, C. G. (2021). El sector florícola del Ecuador y su aporte a la Balanza Comercial Agropecuaria: período 2009 - 2020. *Revista Científica y Tecnológica UPSE*, 74-82.
- Chamba Bernal, J., Bermeo, L. (2021). Variables determinantes en el crecimiento económico del Ecuador función Cobb-Douglas 2007-2019. *Revista del Instituto Tecnológico Superior Jubones*, 109-112.
- Chandler, S., Sanchez, C. (2012). Genetic Modification; the Development Of Transgenic Ornamental Plant Varieties. *Plant Biotechnology Journal*, 8(10), 891-903.
- Chen, M., Chen, G., Lin, T. (2015). Application Of An Autoregressive Integrated Moving Average Model For Predicting Injury Mortality In Xiamen, China. *BMJ Open*, 12(5)
- Chlebus, M., Dyczko, M., Woźniak, M. (2021). Nvidia's Stock Returns Prediction Using Machine Learning Techniques For Time Series Forecasting Problem. *Central European Economic Journal*, 55(8), 44-62.
- Coutin Marie, G. (2007). Utilización de modelos ARIMA para la vigilancia de enfermedades transmisibles. *Revista Cubana de Salud Pública*.

- de Lima Veloso, A., García, L. (2022). Estado y estadística. La importancia de los sistemas oficiales de estadística para las democracias modernas. *Revista mexicana de ciencias políticas y sociales*.
- Eidizadeh, R. S. (2016). Analysing the role of business intelligence, knowledge sharing and organisational innovation on gaining competitive advantage. *Journal of Workplace Learning*, 250-267.
- Evans, (2003). Practical Business Forecasting. Malden Massachusetts. *Blackwell Publishers*.
- Farooqi, A. (2014). ARIMA Model Building and Forecasting on Imports and Exports of Pakistan. 157-168.
- Fidel de la Oliva de Con, R. J. (2016). Aproximación a la metodología Box-Jenkins para la predicción de la tasa de cambio EUR/USD. *La Habana*.
- Fryz, M., Scherbak, L. (2023). Properties Of Discrete-time Conditional Linear Cyclostationary Random Processes In the Problems Of Energy Informatics. *Sist. dosl. energ.*
- Gallego, M. J., Rodríguez, A., Mínguez, J., & F., J. P. (2018). Modelos ARIMA para la predicción del gasto conjunto de oxígeno de vuelo y otros gases en el Ejército del Aire. *Sanidad mil*, 223-229.
- Gaxiola, N. C. (2013). Análisis de la llegada de turistas internacionales a México. *Investigación administrativa*, 20-34.
- Ghosh, S. (2017). Forecasting Cotton Exports in India using the ARIMA model. *Amity Journal of Economics*, 36-52.
- Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, 499-510.
- González, M. P. (2009). *Análisis de series temporales: Modelos ARIMA*. Sarrik-on. <http://hdl.handle.net/10810/12492>
- González, M. P. (2009). *Análisis de series temporales: Modelos ARIMA*. Sarrik-on. <http://hdl.handle.net/10810/12492>
- Guaita-Pradas, I., Mañay, L., Marques-Perez, I. (2023). Competitiveness Of Ecuador's Flower Industry In the Global Market In The Period 2016–2020. *Sustainability*, 7(15)

- Guaita-Pradas, I., Mañay, L., Marques-Perez, I. (2023). Competitiveness Of Ecuador's Flower Industry In the Global Market In The Period 2016–2020. *Sustainability*, 7(15).
- Gujatari, D. N. (2010). *Econometría*. México: The McGraw-Hill.
- Guzmán, J. E. (2019). Desarrollo de un modelo de pronóstico en la metodología de Box-Jenkins para la demanda de estudiantes de pregrado de la fundación universitaria tecnológico COMFENALCO de la ciudad de cartagena de Indias desde 2002 a 2018.
- Hernández S., Ca, Pedraza M., Lf, & Escobar Diaz, A. (2008). Aplicaciones de la serie de tiempo en modelos de tráfico para una red de datos. *Scientia Et Technica*, XIV (38), 31-36.
- Huanzhou Zhu, Z. G.-T. (2018). Developing a pattern discovery method in time series data and its GPU acceleration. *Big Data Mining and Analytics*.
- Indrajeet Y. Javeri, M. T. (2021). Improving Neural Networks for Time Series Forecasting using Data Augmentation and AutoML.
- Jacho, A. D., Loaiza, I. H., & López, J. L. (2020). La importancia de la estadística para el éxito de resultados en una investigación. *Revista Imaginario Social*, 3(1). <https://doi.org/10.31876/is.v3i1.9>
- Kurov, A., Sancetta, A., Strasser, G., Wolfe, M. (2018). Price Drift Before U.S. Macroeconomic News: Private Information About Public Announcements?. *J. Financ. Quant. Anal.*, 1(54), 449-479.
- Lawrence, M., Goodwin, P., O'Connor, M., Önkal, D. (2006). Judgmental Forecasting: a Review Of Progress Over The Last 25years. *International Journal of Forecasting*, 3(22), 493-518.
- Li, S., Guo, S. (2021). Persistence and Extinction Of A Stochastic Sis Epidemic Model With Regime Switching And Lévy Jumps. *DCDS-B*, 9(26).
- Liu, Q., Liu, X., Jiang, B. et al. (2011). Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC Infect Dis* 11, 218
- Lóez Sarmiento, D. A., & Martínez Alayón, C. A. (2013). Modelado de pérdidas en una transmisión de video por medio de series de tiempo ARIMA y SARIMA. *Revista Tecnura*, 53-63.
- Lorente-Leyva, L. L., Alemany, M. M., Peluffo-Ordóñez, D. H., & Herrera-Granda, I. D. (2020). A Comparison of Machine Learning and Classical Forecasting

- Methods: A Case Study of Ecuadorian Textile Industry. *Machile Learning, Optimization, and Data Science*, 131-142.
- Luis Adrián Lasso Cardona, D. F. (2022). Aplicaciones de la Datificación y Big Data en América Latina entre el 2015 y 2019. *Revista Logos Ciencia y Tecnología*, 125-143.
- Malpartida, J. Y. (2022). Predicción de la contaminación atmosférica generada por las emisiones del CO2 en el Perú utilizando los métodos ARIMA y Redes Neuronales. *TecnoHumanismo Revista Científica*, 114-125.
- Manoj Kumar, S. S. (2016). Forecasting tourist in flow in South East Asia: A case of Singapore. *Tourism and Mangement Studies*, 107-119.
- María Raquel Chafloque Céspedes, V. M. (2018). Modelación de las dinámicas, estimación y análisis de las exportaciones mineras: Caso Peruano. *Alternativas*, 37-45.
- Matthias Forkel, N. C. (2013). Trend Change Detection in NDVI Time Series: Effects of Inter-Annual Variability and Methodology. *Remote Sens*.
- Mendoza, H, Bautista, G. (2002). Probabilidad y Estadística. Universidad Nacional de Colombia, <http://www.virtual.unal.edu.co/cursos/ciencias/2001065/>. Licencia: Creative Commons BY-NC-ND
- Michael Parkir, G. E. (2007). *Macroeconomía versión para Latinoamérica*. México: Pearson Educación.
- Michón, C. A., & Vizconde, T. A. (2011). Modelo SARIMA para la llegada mensual de visitantes extranjeros por el Aeropuerto Internacional "Jorge Chávez". *Scientia*.
- Montero Posada, A., Vega, J. (2015). Solución de inteligencia de negocio para métricas de gestión de proyectos. *Revista Cubana de Ciencias Informáticas*, 85-97.
- Müller, E. S. (2020). Two-Sample Testing for Event Impacts in Time Series. Society for Industrial and Applied Mathematics .
- Muñoz, A., Urquijo, J., Castro, A., & Lombana, J. (2017). Pronóstico del precio de la energía en Colombia utilizando modelos ARIMA con IGARCH. *Revista de Economía del Rosario*, 127-161.
- Murillo, M. J., & Cáceres, G. (2013). Business intelligence y la toma de decisiones financieras: una aproximación teórica. *Revista LOGOS CIENCIA & TECNOLOGÍA*, 119-138.

- Murillo S, J. ., Trejos, A. ., & Carvajal Olaya, P. . (2023). estudio del pronóstico de la demanda de energia electrica, utilizando modelos de series de tiempo. *Scientia Et Technica*, 3(23). Recuperado a partir de <https://revistas.utp.edu.co/index.php/revistaciencia/article/view/7379>
- Nazeeh, N., Suarez-Lopez, J. (2020). Summary Data Of Home Proximity To the Nearest Greenhouse (Floricultural) Crops And Areas Of Greenhouse Crops Around Various Distances From Homes In Agricultural Settings In Ecuador. *Data in Brief*, (31)
- P. G., A. (1994). Economic forecasting in agriculture. *International Journal of Forecasting*, 81-135.
- Quisbert, N. C. (1997). Modelos ARIMA. *Revista Ciencia Cultura*.
- Ragozin, A., Pletenkova, A. (2022). The Application Of Digital Signal Processing Technology To Improve the Accuracy Of Forecasting Time Series Data In Anomaly Detection Systems In The Observed Processes Of Automated Process Control Systems. *IS*, 4(22).
- Ramírez, L., Cristancho, S. L., & José-Alejandro, C.-L. (2021). Modelos de fluctuaciones de precios agrícolas: Estudio comparativo de frutas tropicales fresas en Colombia. *Revista de Ciencias Sociales*.
- Ramlan, M. N. (2021). Evaluating Forecast performance of malaysian goods exports for 2021-2022 with Box Jenkins Methodology and Arima model. *Focus on Research in Contemporary Economics*, 157-180.
- Raynolds, L. (2012). Fair Trade Flowers: Global Certification, Environmental Sustainability, and Labor Standards. *Rural Sociol*, 4(77), 493-519.
- Rombach, M., Dean, D., Widmar, N., Bitsch, V. (2021). The Ethically Conscious Flower Consumer: Understanding Fair Trade Cut Flower Purchase Behavior In Germany. *Sustainability*, 21(13).
- Samuel-Luis Rojas, B. R. (2020). La producción de vainilla en México: análisis y pronóstico. *Revista Mexicana de Ciencias Agrícolas*.
- Sánchez, E., Barreras, A., Pérez, C., Figueroa, F., & Angel, O. (2013). Aplicación de un modelo ARIMA para pronosticar la producción de leche de bovino en baja california, México. *Tropical and Subtropical Agroecosystems*, 315-324.
- Sierra, M. F. (2013). Introducción. Ciencia, tecnología y América Latina: perspectivas situadas. *Universitas Humanística*, 2013.

- Toledo, W. (2017). El rol de las exportaciones en el crecimiento económico: evidencia de una muestra de países de América Latina y el Caribe. *Revista de Economía*, 78-100.
- Upadhyay, V. K. (2013). Modelling and forecasting export and import of Indian wood based panel using ARIMA models. *Elixir International Journal*.
- Veuthey, S., Gerber, J. (2012). Accumulation By Dispossession In Coastal Ecuador: Shrimp Farming, Local Resistance and The Gender Structure Of Mobilizations. *Global Environmental Change*, 3(22), 611-622.
- Villacis, B. (2021). Experticia estadística en la administración pública ecuatoriana: mecanismos de emergencia y legitimación. *Revista de Ciencias Sociales*, 81-102.
- Villavicencio, J. (2010). Introducción a series de tiempo. Puerto Rico.
- Wolfe, M., Archana, A., Catoe, D., Coffman, M., Dorevich, S., Graham, K., ... & Boehm, A. (2021). Scaling Of Sars-cov-2 Rna In Settled Solids From Multiple Wastewater Treatment Plants To Compare Incidence Rates Of Laboratory-confirmed Covid-19 In Their Sewersheds. *Environ. Sci. Technol. Lett.*, 5(8), 398-404.
- Xu, X., Du, Y., Lie, J. (2022). An Improved Arima Model Based On Regularized Gaussian Basis Function and Its Application To Stock Price Forecasting. *BCPBM*, (30), 112-121.
- Yinghua Liu, H. C. (2020). Time Series Data Reconstruction Method Based on Probability Statistics and Machine Learning. *Multimedia Technology and Enhanced Learning*.



## DECLARACIÓN Y AUTORIZACIÓN

Nosotros, **Li Ye, Kevin Kinjang**, con C.C: # 0924565591 y **Paz y Miño Robles, Zhukov Jafet** con C.C: # 0926807777 autores del trabajo de integración curricular: **Aplicación de modelo ARIMA para el pronóstico de exportación de flores del Ecuador**, previo a la obtención del título de **Licenciados en Negocios Internacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de integración curricular para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

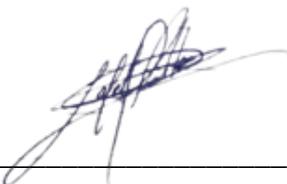
2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de integración curricular, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, **04 de septiembre de 2023**

f. 

Nombre : **Li Ye, Kevin Kinjang**

C.C: **0924565591**

f. 

Nombre: **Paz y Miño Robles, Zhukov Jafet**

C.C: **0926807777**



<b>REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA</b>			
<b>FICHA DE REGISTRO DE TRABAJO DE INTEGRACIÓN CURRICULAR</b>			
<b>TEMA Y SUBTEMA:</b>	<b>Aplicación de modelo ARIMA para el pronóstico de exportación de flores del Ecuador</b>		
<b>AUTOR(ES)</b>	<b>Li Ye Kevin Kinjang y Paz y Miño Robles Zhukov Jafet</b>		
<b>REVISOR(ES)/TUTOR(ES)</b>	<b>Carrera Buri, Félix Miguel, Mgs.</b>		
<b>INSTITUCIÓN:</b>	Universidad Católica de Santiago de Guayaquil		
<b>FACULTAD:</b>	<b>Facultad de Economía y Empresa</b>		
<b>CARRERA:</b>	<b>Negocios Internacionales</b>		
<b>TÍTULO OBTENIDO:</b>	<b>Licenciado en Negocios Internacionales</b>		
<b>FECHA DE PUBLICACIÓN:</b>	04 de septiembre de 2023	<b>No. DE PÁGINAS:</b>	89
<b>ÁREAS TEMÁTICAS:</b>	Negocios, oferta y demanda, Modelos predictivos, Estacionariedad.		
<b>PALABRAS CLAVES/ KEYWORDS:</b>	Inteligencia de Negocios, Modelo Autorregresivo Integrado (AR), Modelo de Medias Móviles (MA), Modelo Autorregresivo Integrado de Medias Móviles (ARIMA), Series de Tiempo, R-Studio,		
<b>RESUMEN/ABSTRACT</b> El mundo está siempre en constante cambio y esto influye en la forma de vivir el día a día de las personas, los avances tecnológicos de la reciente década han representado un paso significativo en diversos ámbitos en los cuales destacan diversas herramientas y conocimientos como inteligencia de negocios para una mejor toma de decisiones. El presente estudio tiene como objetivo explicar y evidenciar cómo el modelo estadístico autorregresivo integrado de media móvil se utiliza para analizar y pronosticar series temporales. Las variables utilizadas para llevar a cabo el pronóstico fueron peso neto y FOB, de la cual se aplicó el modelo ARIMA (1, 0, 2) (0, 1, 1) para la variable peso neto y el modelo ARIMA (2, 0, 1) (0, 1, 2) para la variable FOB. Para alcanzar el objetivo, se empleó el programa Rstudio, el cual contribuyó a gestionar eficazmente desde la preparación de los datos hasta la ejecución de los análisis estadísticos. Como objeto se fijó las exportaciones del sector florícola del Ecuador en el periodo 1990-2023, siendo las exportaciones de flores un indicador económico relevante para el país. El estudio recopila datos históricos de las exportaciones de flores y realiza un análisis exploratorio para comprender las tendencias y patrones presentes en los datos. Los resultados obtenidos revelaron información valiosa para las empresas que operan en el sector agrícola.			
<b>ADJUNTO PDF:</b>	<input checked="" type="checkbox"/> SI		<input type="checkbox"/> NO
<b>CONTACTO CON AUTOR/ES:</b>	<b>Li Ye Kevin Kinjang</b> <b>Teléfono: +593 960785056</b> <b>Paz y Miño Robles Zhukov Jafet</b> <b>Teléfono: +593 959567639</b>		E-mail: <a href="mailto:kevinliye18@gmail.com">kevinliye18@gmail.com</a>  E-mail: <a href="mailto:Jafet.zhukov@gmail.com">Jafet.zhukov@gmail.com</a>
<b>CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UIC):</b>	<b>Nombre:</b> Román Bermeo, Cynthia Lizbeth <b>Teléfono:</b> +593-984228698 <b>E-mail:</b> <a href="mailto:cynthia.roman@cu.ucsg.edu.ec">cynthia.roman@cu.ucsg.edu.ec</a>		
<b>SECCIÓN PARA USO DE BIBLIOTECA</b>			
<b>Nº. DE REGISTRO (en base a datos):</b>			
<b>Nº. DE CLASIFICACIÓN:</b>			
<b>DIRECCIÓN URL (tesis en la web):</b>			