

**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO**  
**FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS**  
**ESCUELA PROFESIONAL DE ESTADÍSTICA**



**TESIS**

**“Modelo estadístico para pronóstico de cantidad mensual producida de caña de azúcar. Enero 1983 – Abril 2021. Perú”**

**Para optar el Título Profesional de:**  
**Licenciada en Estadística**

**Presentado por:**

Br. Mino Ilma Rosa Maruja

Br. Pizarro Hoyos Cinthia Liseth

**Asesor:**

Dr. Rodríguez López Wilver Omero

**Lambayeque, 2024**



**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
ESCUELA PROFESIONAL DE ESTADÍSTICA**



**TESIS**

**“Modelo estadístico para pronóstico de cantidad mensual producida de  
caña de azúcar. Enero 1983 – Abril 2021. Perú”**

**Presentado por:**

.....  
Br. Mino Ilma Rosa Maruja  
Investigadora

.....  
Br. Pizarro Hoyos Cinthia Liseth  
Investigadora

**Asesor:**

.....  
Dr. Rodríguez López Wilver Omero

**Lambayeque, 2024**



**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
ESCUELA PROFESIONAL DE ESTADÍSTICA**



**TESIS**

**“Modelo estadístico para pronóstico de cantidad mensual producida de  
caña de azúcar. Enero 1983 – Abril 2021. Perú”**

**Aprobado por:**

.....  
Mg. Manuel Francisco Hurtado Sánchez  
Presidente

.....  
M.Sc. Débora Esther Mejía Pacheco  
Secretario

.....  
Mg. Luis Enrique Tuñoque Gutiérrez  
Vocal

## REPORTE TURNITIN

“Modelo estadístico para pronóstico de cantidad mensual producida de caña de azúcar. Enero 1983 – Abril 2021. Perú”

### INFORME DE ORIGINALIDAD

<b>14%</b>	<b>13%</b>	<b>2%</b>	<b>5%</b>
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

### FUENTES PRIMARIAS

<b>1</b>	<b>hdl.handle.net</b> Fuente de Internet	<b>4%</b>
<b>2</b>	<b>repositorio.unprg.edu.pe:8080</b> Fuente de Internet	<b>3%</b>
<b>3</b>	<b>www.minagri.gob.pe</b> Fuente de Internet	<b>2%</b>
<b>4</b>	<b>docplayer.es</b> Fuente de Internet	<b>1%</b>
<b>5</b>	<b>repositorio.unprg.edu.pe</b> Fuente de Internet	<b>1%</b>
<b>6</b>	<b>repositorio.unapiquitos.edu.pe</b> Fuente de Internet	<b>&lt; 1%</b>
<b>7</b>	<b>repositorio.upse.edu.ec</b> Fuente de Internet	<b>&lt; 1%</b>
<b>8</b>	<b>www.hnb.hr</b> Fuente de Internet	<b>&lt; 1%</b>
<b>9</b>	<b>tesis.unap.edu.pe</b> Fuente de Internet	

Dr. Wilber Omer Rodríguez López  
DNI: 17636159  
Asesor

		< 1 %
10	<b>Submitted to Universidad Nacional de Frontera</b> Trabajo del estudiante	< 1 %
11	<b>Submitted to Universidad Nacional Pedro Ruiz Gallo</b> Trabajo del estudiante	< 1 %
12	<b>Submitted to Universidad Cesar Vallejo</b> Trabajo del estudiante	< 1 %
13	<b>Submitted to Higher Education Commission Pakistan</b> Trabajo del estudiante	< 1 %
14	<b>Submitted to Pontificia Universidad Catolica del Peru</b> Trabajo del estudiante	< 1 %
15	<b>repositorio.ucv.edu.pe</b> Fuente de Internet	< 1 %
16	<b>repositorio.uss.edu.pe</b> Fuente de Internet	< 1 %
17	<b>www.minag.gob.pe</b> Fuente de Internet	< 1 %
18	<b>repositorio.uci.cu</b> Fuente de Internet	< 1 %



**Dr. Wiker Omero Rodríguez López**  
DNI: 17636159  
Asesor

---

Excluir citas      Activo  
Excluir bibliografía      Activo

Excluir coincidencias < 15 words



---

**Dr. Wilber Omer Rodríguez López**  
DNI: 17636159  
Asesor



## Recibo digital

Este recibo confirma que su trabajo ha sido recibido por Turnitin. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega: Cinthia Liseth Pizarro Hoyos  
Título del ejercicio: Quick Submit  
Título de la entrega: "Modelo estadístico para pronóstico de cantidad mensual pr...  
Nombre del archivo: INFORME\_ESTADISTICA\_13-10-2023\_turnitin4.docx  
Tamaño del archivo: 1.15M  
Total páginas: 43  
Total de palabras: 6,821  
Total de caracteres: 35,625  
Fecha de entrega: 17-oct.-2023 06:37p. m. (UTC-0500)  
Identificador de la entre... 2199083569



## CONSTANCIA DE VERIFICACIÓN DE ORIGINALIDAD

Yo, Dr. Rodríguez López Wilver Omero, asesor y usuario revisor del documento de tesis titulado: “Modelo estadístico para pronóstico de cantidad mensual producida de caña de azúcar. Enero 1983 – Abril 2021. Perú”

Cuyas autoras son, Br. Mino Ilma Rosa Maruja identificada con el documento de identidad N° 47450990 y Br. Pizarro Hoyos Cinthia Liseth identificada con el documento de identidad N° 47586754; declaro que la evaluación realizada por el Programa informático TURNITIN, ha arrojado un porcentaje de similitud del 14%, verificable en el Resumen de Reporte automatizado de similitudes que se acompaña.

El suscrito analizó dicho reporte y concluyó que cada una de las coincidencias detectadas dentro del porcentaje de similitud permitido no constituyen plagio y que el documento cumple con la integridad científica y con las normas para el uso de las citas y referencias establecidas en los protocolos respectivos. Se cumple con adjuntar el recibo digital a efectos de la trazabilidad respectiva del proceso.

Lambayeque, 19 de octubre de 2023.



.....  
Dr. Rodríguez López Wilver Omero  
DNI: 17636159  
Asesor

# ACTA DE SUSTENTACIÓN



**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO**  
**FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS**  
**DECANATO**  
Ciudad Universitaria - Lambayeque  
LICENCIADA - RESOLUCIÓN DEL CONSEJO DIRECTIVO N° 015 -2023-SUNEDU / CD



0021



## ACTA DE SUSTENTACIÓN N° 008.-2024.-D/FACFyM

Siendo las 09:00 ..... a.m. ..... del día miércoles 24 de enero del 2024, se reunieron los miembros del jurado evaluador de la Tesis titulada:

Modelo estadístico para pronóstico de cantidad mensual producida de caña de azúcar Enero 1983-Abril 2021 Perú

Designados por Resolución N° 839-2021-Virtual-D/FACFyM de fecha 17 de noviembre 2021 con la finalidad de evaluar y calificar la sustentación de la tesis antes mencionada, conformada por los siguientes docentes:

Mg. Manuel Francisco Hurtado Sánchez Presidente  
M.Sc. Debara Esther Mejía Pacheco Secretario  
Lic. Est. Luis Enrique Tuñogue Gutiérrez Vocal

La tesis fue asesorada por (el) (la) Dr. Lic. Wilver Omero Rodríguez López nombrado por Resolución N° 839-2021-Virtual-D/FACFyM de fecha 17 de noviembre de 2021

El Acto de Sustentación fue autorizado por Resolución N° 038-2024-D/FACFyM de fecha 15 de enero de 2024

La Tesis fue presentada y sustentada por (el) (los) Bachiller (es) Pizarro Hoya Cinthia ..... y tuvo una duración de 30 minutos.

Después de la sustentación, y absueltas las preguntas y observaciones de los miembros del jurado se procedió a la calificación respectiva, otorgándole el Calificativo de Diecisiete (17) en la escala vigesimal, mención (Buena).

Por lo que queda(n) apto(s) para obtener el Título Profesional de Licenciada en Estadística de acuerdo con la Ley Universitaria 30220 y la normatividad vigente de la Facultad de Ciencias Físicas y Matemáticas y la Universidad Nacional Pedro Ruiz Gallo.

Siendo las 9:30 ..... a.m. ..... se dio por concluido el presente acto académico, dándose conformidad al presente acto con la firma de los miembros del jurado.

CERTIFICO: Que, es copia fiel del original  
Fecha: 23/01/2024  
Manuel Francisco Hurtado Sánchez Presidente  
Debara Esther Mejía Pacheco Secretario  
Lic. Est. Luis Enrique Tuñogue Gutiérrez Vocal  
Dr. Lic. Wilver Omero Rodríguez López Asesor

## DECLARACIÓN JURADA DE ORIGINALIDAD

Mino Ilma Rosa Maruja y Pizarro Hoyos Cinthia Liseth investigadores principales, y con el apoyo del Dr. Rodríguez López Wilver Omero asesor del trabajo de investigación “Modelo estadístico para pronóstico de cantidad mensual producida de caña de azúcar. Enero 1983 – Abril 2021. Perú”, declaramos bajo juramente que este informe no es objeto de plagio, ni contiene información falsa. En caso de que se demuestre lo contrario, estamos dispuestos anular este informe y enfrentar las consecuencias administrativas que correspondan, incluso la cancelación de cualquier título o grado otorgado como resultado de este trabajo.

Lambayeque, 19 de octubre de 2023.

Nombre Investigador (es):



.....  
Br. Mino Ilma Rosa Maruja



.....  
Br. Pizarro Hoyos Cinthia Liseth

Nombre Asesor:



.....  
Dr. Rodríguez López Wilver Omero

## **DEDICATORIA**

Le dedico esta investigación a Dios, a mis padres y hermanos quienes han sido el motor y esperanza para guiarme en mi etapa profesional.

También, quiero dedicar este trabajo de investigación a mi esposo, compañero de vida por su paciencia, comprensión y apoyarme en cada momento que afronte durante mi investigación. Gracias por enseñarme a ser perseverante.

Cinthia Liseth

Le dedico esta investigación en primer lugar a Dios, a mis padres que sin ellos no hubiera logrado una meta más en mi vida profesional, gracias por estar a mi lado en esta etapa de mi pregrado, su apoyo moral y entusiasmo que me brindaron para seguir adelante en mis propósitos.

También, quiero dedicar esta investigación a mi esposo, esta persona que estuvo apoyándome en cada decisión que tomara, a esa persona que tuvo paciencia y entrega para conmigo.

A mi hijo, eres la razón que me levante cada día a esforzarme por el presente y el mañana eres mi principal motivación.

Rosa Maruja

## **AGRADECIMIENTO**

Agradezco a Dios por darme fuerzas y valor de superar cada obstáculo que se presentaron en el camino.

A nuestro asesor el Dr. Wilver Omero Rodríguez López por la colaboración como especialista en nuestra carrera profesional quien con sus conocimientos ha sabido orientarnos en esta investigación.

Cinthia Liseth y Rosa Maruja

# ÍNDICE GENERAL

DEDICATORIA.....	xi
AGRADECIMIENTO.....	xii
ÍNDICE GENERAL.....	xiii
ÍNDICE DE TABLAS.....	xv
ÍNDICE DE FIGURAS.....	xvi
RESUMEN.....	xviii
ABSTRACT.....	xix
INTRODUCCIÓN.....	1
CAPÍTULO I. DISEÑO TEÓRICO.....	3
1.1. Antecedentes.....	3
1.2. Bases teóricas.....	4
1.2.1. Serie de Tiempo.....	4
1.2.2. Modelos ARIMA (p, d, q).....	6
1.2.3. Modelo SARIMA (p, d, q) (P, D, Q).....	6
1.2.4. Metodología de Box Jenkins.....	6
1.2.5. Caña de azúcar.....	8
CAPITULO II. DISEÑO METODOLÓGICO.....	13
2.1. Diseño de contrastación de hipótesis/Procedimiento a seguir en la investigación. ..	13
2.2. Población y muestra.....	13
2.3. Técnicas e instrumentos de recolección de datos. ....	13
2.4. Técnicas estadísticas para el procesamiento de los datos. ....	13
CAPITULO III. RESULTADOS.....	15
3.1. Principales características de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 abril 2021. ....	15
3.2. Estacionariedad y estacionalidad de cantidad mensual producida de caña de azúcar en el Perú, enero 1983 abril 2021. ....	16
3.2.1. Análisis de la estacionariedad mediante Dickey – Fuller.....	18
3.2.2. Análisis de la estacionariedad mediante Dickey – Fuller en primeras diferencias. ....	19
3.3. Modelo de serie de tiempo de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 abril 2021. ....	22

3.4. Pronosticar la cantidad mensual producida de caña de azúcar en el Perú, para los años mayo 2021, diciembre 2023. ....	39
CAPITULO IV. DISCUSIÓN DE RESULTADOS .....	41
CAPITULO V. CONCLUSIONES .....	43
Recomendaciones .....	44
Bibliografía.....	45
Anexos.....	48

## ÍNDICE DE TABLAS

<b>Tabla 1:</b> Estadísticos descriptivos de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021. ....	15
<b>Tabla 2:</b> Dickey Fuller Aumentado aplicada de la serie original. ....	18
<b>Tabla 3:</b> Dickey Fuller Aumentado aplicada de la serie original en primeras diferencias..	19
<b>Tabla 4:</b> Pruebas estadísticas aplicadas a los residuales con relación si tienden a una distribución independiente y normal del modelo SARIMA 4 (0, 1, 2) (1, 1, 1).....	22
<b>Tabla 5:</b> Pruebas estadísticas aplicadas a los residuales con relación si tienden a una distribución independiente y normal del modelo SARIMA 6 (0, 1, 3) (1, 1, 1).....	25
<b>Tabla 6:</b> Pruebas estadísticas aplicadas a los residuales con relación si tienden a una distribución independiente y normal del modelo SARIMA 8 (1, 1, 1) (1, 1, 1).....	28
<b>Tabla 7:</b> Pruebas estadísticas aplicadas a los residuales con relación si tienden a una distribución independiente y normal del modelo SARIMA 16 (2, 1, 2) (1, 1, 1).....	31
<b>Tabla 8:</b> Evaluación de los modelos propuestos por medio de los errores de pronósticos.	34
<b>Tabla 9:</b> Pronóstico de cantidad mensual producida de caña de azúcar en el Perú, para los años mayo 2021, diciembre 2023 .....	39

## ÍNDICE DE FIGURAS

<b>Figura 1:</b> Representación gráfica del proceso de realización de un modelo ARIMA.....	8
<b>Figura 2:</b> Cadena productiva de la caña de azúcar en el Perú .....	9
<b>Figura 3:</b> Proceso de Comercialización del Azúcar en el Perú .....	10
<b>Figura 4:</b> Cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021. .....	16
<b>Figura 5:</b> Coeficientes de autocorrelación simple de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021. ....	17
<b>Figura 6:</b> Coeficientes de autocorrelación parcial de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021. ....	17
<b>Figura 7:</b> Cantidad mensual producida de caña de azúcar en el Perú, enero 1983 – abril 2021 en primeras diferencias.....	19
<b>Figura 8:</b> Función de autocorrelación simple de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021 en primeras diferencias.....	20
<b>Figura 9:</b> Función de autocorrelación parcial de la serie original diferenciada en un periodo tanto en la parte regular y estacional .....	21
<b>Figura 10:</b> Prueba de Ljung-Box de los residuales del modelo SARIMA 4 (0, 1, 2) (1, 1, 1).....	22
<b>Figura 11:</b> Normalidad de los datos del modelo SARIMA 4 (0, 1, 2) (1, 1, 1) por la prueba de kolmogorov Smirnov.....	23
<b>Figura 12:</b> Función de autocorrelación simple y parcial de los residuales del modelo SARIMA 4 (0, 1, 2) (1, 1, 1).....	24
<b>Figura 13:</b> Prueba de Ljung-Box de los residuales del modelo SARIMA 6 (0, 1, 3) (1, 1, 1).....	25
<b>Figura 14:</b> Normalidad de los datos del modelo SARIMA 6 (0, 1, 3) (1, 1, 1) por la prueba de kolmogorov Smirnov.....	26
<b>Figura 15:</b> Función de autocorrelación simple y parcial de los residuales del modelo SARIMA 6 (0, 1, 3) (1, 1, 1) .....	27
<b>Figura 16:</b> Prueba de Ljung-Box de los residuales del modelo SARIMA 8 (1, 1, 1) (1, 1, 1).....	28
<b>Figura 17:</b> Normalidad de los datos del modelo SARIMA 8 (1, 1, 1) (1, 1, 1) por la prueba de kolmogorov Smirnov.....	29

<b>Figura 18:</b> Función de autocorrelación simple y parcial de los residuales del modelo SARIMA 8 (1, 1, 1) (1, 1, 1).....	30
<b>Figura 19:</b> Prueba de Ljung-Box de los residuales del modelo SARIMA 16 (2, 1, 2) (1, 1, 1).....	31
<b>Figura 20:</b> Normalidad de los datos del modelo SARIMA 16 (2, 1, 2) (1, 1, 1) por la prueba de kolmogorov Smirnov.....	32
<b>Figura 21:</b> Función de autocorrelación simple y parcial de los residuales del modelo SARIMA 16 (2, 1, 2) (1, 1, 1) .....	33
<b>Figura 22:</b> Valores de la serie original y del pronóstico del modelo SARIMA 8 (1, 1, 1) (1, 1, 1).....	38

## RESUMEN

El estudio se basó en determinar un modelo estadístico para el pronóstico de la cantidad mensual de caña de azúcar producida en el Perú, enero 1983 - abril de 2021. La metodología empleada se dio bajo el enfoque cuantitativo, con diseño descriptivo longitudinal, además, la técnica fue la observación documental indirecta y la muestra se conformó por 460 observaciones de la variable. Se usó el método de Box - Jenkins, que brindó el modelo óptimo de pronóstico el SARIMA (1, 1, 1) (1, 1, 1), la misma que fue diferenciada en un período, mostró coeficientes estimados de los parámetros significativos, residuos independientes y con distribución normal, y para su elección se consideró el menor BIC (5095,52) y por el criterio de parsimonia.

Palabras claves: Modelo, Producción y SARIMA.

## **ABSTRACT**

The study was based on determining a statistical model for the forecast of the monthly amount of sugar cane produced in Peru, January 1983 - April 2021. The methodology used was based on the quantitative approach, with a longitudinal descriptive design, in addition, the The technique was indirect documentary observation and the sample was made up of 460 observations of the variable. The Box-Jenkins methodology was used, which provided the optimal forecast model SARIMA (1, 1, 1) (1, 1, 1), which was differentiated in a period, showed estimated coefficients of the significant parameters, independent residues with a normal distribution, and for their selection the lowest BIC (5095.52) was considered and by the parsimony criterion.

Keywords: Model, production and SARIMA.

## INTRODUCCIÓN

La caña de azúcar es uno de los principales cultivos en el Perú debido a las condiciones agroclimáticas presentando altos rendimientos respecto a otros países. El impacto que se tuvo por la pandemia en el sector agrícola afectó en la producción ya que se dieron medidas sanitarias para combatir el COVID – 19, existiendo inmovilización social por ello las azucareras restringieron la cantidad de trabajadores. (Ministerio de Desarrollo Agrario y Riego, [MIDAGRI], 2020)

El MIDAGRI (2020) reporta que la producción en 2020 registró 10,5 millones de toneladas, representando un 4% menos que en el 2019. Esta caída se basó en las cosechas de bajo rendimiento que se redujeron en 2,2%, lo que también tuvieron bajos rendimientos (-1,8%), dado que pasó del 2019 al 2020 de 126 a 124 toneladas por hectárea respectivamente. Además, los departamentos con mayores volúmenes de producción fueron Lambayeque (-14,9%), La Libertad (-3,1%) y Lima (-9,6%). No obstante, el MIDAGRI reporta que en el primer trimestre del 2021 se recuperó la producción a 2,3 millones de toneladas (5,6%), manteniéndose en un incremento de las áreas de cosecha (4,2%) y mejor nivel de rendimiento (1,3%).

La situación problemática por las cuales se realizó el estudio se basó en que existe una disminución de la cantidad de la caña de azúcar, un fenómeno que tiene importantes implicaciones económicas y sociales. Por otro lado, en el Perú no existen reportes que muestren los modelos de pronóstico de la producción de la caña de azúcar, siendo necesario conocer el modelo con la finalidad de analizar la variable. Esta ausencia de información confiable y predictiva dificulta la toma de decisiones informadas y la implementación de estrategias efectivas para abordar los desafíos que pueda enfrentar el sector. En este sentido el trabajo se enfocó en proponer un modelo estadístico siguiendo los pasos de Box Jenkins, la cual se formuló la siguiente pregunta: ¿Cuál es el modelo estadístico para el pronóstico de cantidad mensual producida de caña de azúcar en el Perú, de enero 1983 - abril 2021?

En la exploración de resolver la problemática anterior, se planteó como objetivo general, determinar un modelo estadístico para el pronóstico de cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021. Asimismo, los objetivos específicos fueron los siguientes: Describir las principales características de la serie; analizar la parte estacionaria y estacional de la serie; determinar el modelo de serie de tiempo de la serie y pronosticar la cantidad mensual producida de caña de azúcar en el Perú, para los años mayo

2021 - diciembre 2023.

El estudio se justifica dado que no hay trabajos que predigan la variable con el método de Box y Jenkins, en este sentido el modelo contribuirá adoptar decisiones sólidas en el sector económico y agrícola.

Es importante destacar que nuestra investigación presenta ciertas limitaciones que deben ser consideradas. Aunque se realizaron pruebas exhaustivas para verificar los supuestos subyacentes, no se llevó a cabo un proceso completo de validación del modelo, siendo una etapa crucial para asegurar la fiabilidad y la precisión de los resultados obtenidos.

Para finalizar, la tesis se conformó por: Capítulo I, se muestra el diseño teórico, las cuales se conforma por los antecedentes internacionales y nacionales, así como las bases teóricas de la variable analizada. Ahora bien, el capítulo II, trata de la metodología y las herramientas utilizados en la tesis. Luego el capítulo III, muestra los resultados y la discusión del estudio, mientras que el capítulo IV, se encuentra las conclusiones y el capítulo V, trata de las sugerencias realizadas por las autoras.

# CAPÍTULO I. DISEÑO TEÓRICO

## 1.1. Antecedentes:

Se realizó una revisión de tesis y revistas a nivel internacional, las cuales se encontró el estudio de Fuentes (2017) que tuvo como propósito la evaluación y validación de un modelo de estimación de la producción por medio de imágenes de satélite Landsat 8 en la variedad CP73-1547 de caña de azúcar en los ingenios La Unión y Santa Ana. Concluye, la existencia de una regresión estadística muy baja, siendo el  $R^2 = 0.1269$  para el ingenio La Unión y el  $R^2 = 0.0349$  para el ingenio Santa Ana.

Además, se tiene el estudio de Pérez (2017), cuya finalidad principal fue evaluar el modelo de la producción de caña de azúcar por medio de imágenes del satélite Landsat 8 en Ingenio Madre Tierra. Se tiene como conclusión que, el modelo lineal de la zafra 2014-2015:  $y = 52.002x + 81.674$ , obtuvo un mayor  $R^2 = 0.0504$ ; los mismo para modelo por zafra y del primer tercio:  $y = 61.355x + 75.721$  tuvo un  $R^2 = 0.0493$ . El modelo lineal de la relación Índice de vegetación de diferencia normalizada (NDVI) y rendimiento de la gramínea de las zafras 2013-2015, fue:  $y = 28.366x + 93.244$  con  $R^2 = 0.0253$ . En este sentido los modelos no resultaron tener asociación, el NDVI y la producción de la caña.

Luego se encontraron estudios a nivel nacional, entre ellos se encuentra el artículo de Laurente y Laurente (2019), con el objetivo de aplicar un modelo ARIMA para el modelamiento y proyección de la producción de papa en Puno haciendo uso del método Box-Jenkins. La conclusión a la que se llegó es la obtención de un óptimo modelo el ARIMA (3, 0, 4) usando el AIC y SC.

También se tiene el estudio de Arévalo y Santisteban (2020), cuya finalidad principal fue determinar el Modelo óptimo de Pronóstico de la Cantidad de Producción de Electricidad con la Metodología de Box Jenkins, Región Lambayeque del periodo 2005– 2017. Los autores concluyen, la obtención de un modelo parsimonioso y con menor BIC (758.49), siendo el SARIMA (2,1,2)(1,1,0) teniendo como ecuación general: “ $(1 - \alpha_1 L - \alpha_2 L^2)(1 - \beta_1 L^{12})(Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) = (1 - \phi_1 L - \phi_2 L^2)e_t$ ”.

Por último, se tiene el estudio de Valverde (2021) que tuvo como propósito, encontrar el modelo óptimo de pronóstico de la producción de la caña de azúcar en la entidad Agroindustrial Laredo S.A.A, de enero 2012 - noviembre 2019. En el estudio se concluye, la obtención de un modelo de pronóstico, el ARIMA (1,1,12), siendo los coeficientes AR (-

0.229527) y MA (0.726968) que fueron muy significativos, con valores de p de 0.001 y 0.000 respectivamente.

## **1.2. Bases teóricas**

### **1.2.1. Serie de Tiempo**

Mauricio (2007) hace referencia que una serie tiempo se basa en la consecución de observaciones que se encuentran en un orden y están cronológicamente sobre un elemento o diversos elementos que puede ser observada en distintos momentos en el tiempo. La formulación matemática comúnmente utilizadas para describir series temporales de una sola variable.

$$Y_1, Y_2, \dots, Y_N; (Y_t)_{t=1}^N (Y_t: t = 1, \dots, N)$$

$Y_t$  es la observación  $t$  ( $1 \leq t \leq N$ ) y  $N$  representa la cantidad de observaciones presentes en la serie, lo cual se refiere al tamaño de la serie. Las  $N$  cantidades  $y_1, y_2, \dots, y_N$  se agrupan por un vector  $y = [y_1, y_2, \dots, y_N]'$ , de orden  $N \times 1$

#### **1.2.1.1. Proceso Estocástico**

Es una sucesión de elementos aleatorios  $\{y_t\}, t = -\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty$  (Novales, 1993).

#### **1.2.1.2. Ruido Blanco**

Son características aleatorias que tienen media cero, varianza uniforme y son mutuamente independientes a través en el tiempo. Se representa el ruido blanco como  $\{\varepsilon_t\}$ . (Novales, 1993)

#### **1.2.1.3. Estacionariedad**

Se considera estrictamente estacionario cuando el  $m$ -upla  $(t_1, t_2, \dots, t_m)$  y el  $k$  vector de variables  $(y_{t_1}, y_{t_2}, \dots, y_{t_m})$  son de igual probabilidad conjunta que el vector  $(y_{t_1 + k}, y_{t_2 + k}, \dots, y_{t_m + k})$ . (Novales, 1993)

#### **1.2.1.4. Función de autocovarianzas**

Es denotada por FAC, para el instante  $t$  y número entero  $k$ , representado por  $\gamma_k(t)$ , igual a la covarianza con  $y_t$  e  $y_{t-k}$ , es decir:  $\gamma_k(t) = \text{Cov}(y_t, y_{t-k})$  (Novales, 1993)

#### **1.2.1.5. Función de Autocorrelación Simple (FAS)**

Función para el instante  $t$  y  $k$  tiene un valor  $\rho_k(t)$  similar a la correlación de  $y_t$  e

$y_{t-k}$  (Novales, 1993).

$$\rho_k(t) = \frac{\text{Cov}(Y_t, Y_{t-k})}{\sqrt{\text{Var}(Y_t)} \sqrt{\text{Var}(Y_{t-k})}} = \frac{\gamma_k(t)}{\sqrt{\text{Var}(Y_t)} \sqrt{\text{Var}(Y_{t-k})}}$$

### 1.2.1.6. Función de autocorrelación parcial (FAP)

Función para el instante  $t$  y cada entero  $k$  donde tiene un valor similar a la correlación entre  $y_t$  e  $y_{t-k}$  precisa por el efecto de los retardos intermedios  $y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$  (Novales, 1993).

### 1.2.1.7. Modelos Autorregresivos

#### a. Autorregresivos: AR (p)

González (2009) explica que es próximo al modelo lineal general, donde se da un modelo finito limitando el modelo general:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t \quad t = 1, 2, \dots$$

Se representa por:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t$$

$$a_t \sim RB(0, \sigma^2) \quad t = 1, 2, \dots$$

#### b. Medias Móviles: MA(q)

González (2009) dice que se basa en las medias móviles de orden infinito y se representa como:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad a_t \sim RB(0, \sigma^2)$$

#### c. Autorregresivos de Medias Móviles: ARMA (p, q)

González (2009) menciona que se basa en la determinación de  $Y_t$  en función de su pasado hasta el retardo  $p$  y  $q$ , se representa por:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad a_t \sim RB(0, \sigma^2)$$

Se escribe de la siguiente manera:

$$(1 - \phi_1 L - \dots - \phi_p L^p) Y_t = (1 - \theta_1 L - \dots - \theta_q L^q) a_t$$

$$\phi_p(L) Y_t = \theta_q(L) a_t$$

Donde:

$\phi_p(L)$ : polinomio autorregresivo

$\theta_q(L)$ : medias móviles

### 1.2.2. Modelos ARIMA (p, d, q)

González (2009) menciona que tiene la denominación de autorregresivo de medias móviles p, d, q y se representa por:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d X_t = (1 - v_1 B - v_2 B^2 - \dots - v_q B^q) a_t$$

Permite la descripción de una serie de observaciones tras haber sido diferenciadas d veces, con el propósito de identificar y eliminar probables causas de no estacionariedad. Los modelos cíclicos se distinguen por presentar oscilaciones periódicas. Estas variaciones pueden superponerse en ocasiones a una tendencia secular. Los modelos ARIMA (p, d, q) (P, D, Q), considerando que los parámetros tienen una parte regular y no estacional.

### 1.2.3. Modelo SARIMA (p, d, q) (P, D, Q)

Arce y Mahia (2001) menciona que el modelo ARIMA (p, d, q) podría escribirse como: (p, d, q) (P, D, Q).

$$Y_T = \phi_1 Y_{T-1} + \phi_2 Y_{T-2} + \dots + \phi_{p_s+p+D_s+d} Y_{T-p_s-p-sD-d} + \delta + U_T + \theta_1 U_{T-1} + \dots + \theta_{q_s+q} U_{T-sQ-q}$$

Considerando que la serie obtiene más de un proceso, tanto en su componente estacional y no estacional y escribiendo los modelos MA(q) y AR(p) es necesario aplicar "d" diferencias en la parte no estacional y estacional para que la serie sea estacionario.

### 1.2.4. Metodología de Box Jenkins

En 1970, Box y Jenkins introdujeron una metodología rigurosa para la detección, estimación y evaluación de modelos dinámicos aplicados a datos de series temporales. Esta metodología ha ganado justificadamente prominencia como una herramienta esencial en el análisis de datos económicos. (Novales, 1993)

Puede definirse como un proceso cíclico iterativo que tiene como objetivo descubrir el modelo de ajuste óptimo para llevar a cabo pronósticos y estimaciones. (Box y Jenkins, 1970, como se citó en Osorio y Ángel, 2019)

De la Fuente (2013) menciona las siguientes fases:

- a. **La primera fase:** Se identifica el posible modelo ARIMA en la serie, y se transforma para que sea estacionaria encontrando una parte autorregresiva (p)

y media móvil (q).

✓ **Prueba de Dickey-Fuller (DF)**

Anchuelo (1993) menciona que DF realizaron un proceso para corroborar la existencia de raíz unitaria y por las cuales plantea las siguientes hipótesis:

- ✓ Hipótesis nula: Existe raíz unitaria (no es estacionario).
- ✓ Hipótesis alternativa: No existe raíz unitaria (estacionaridad).

El contraste se lleva a cabo por la estimación por mínimos cuadrados ordinarios de la regresión auxiliar:

$$\Delta x_t = \mu + \gamma x_{t-1} + e_t$$

$$\text{donde } \Delta = 1 - L$$

Además, Anchuelo (1993), menciona que la corrección de DF se basa en ampliar la regresión auxiliar mediante el uso de retardos de  $\Delta x$ :

$$\Delta x_t = \mu + \gamma x_{t-1} + \sum_{i=1}^n b_i \Delta x_{t-i} + e_t$$

Se elige “n” para desaparecer la autocorrelación en los residuos y se convierta en ruido blanco. Es crucial realizar una elección precisa: si “n” es muy pequeño, el problema no se resuelve, mientras que si el “n” es excesivamente grande entonces ocasiona que el contraste pierda poder.

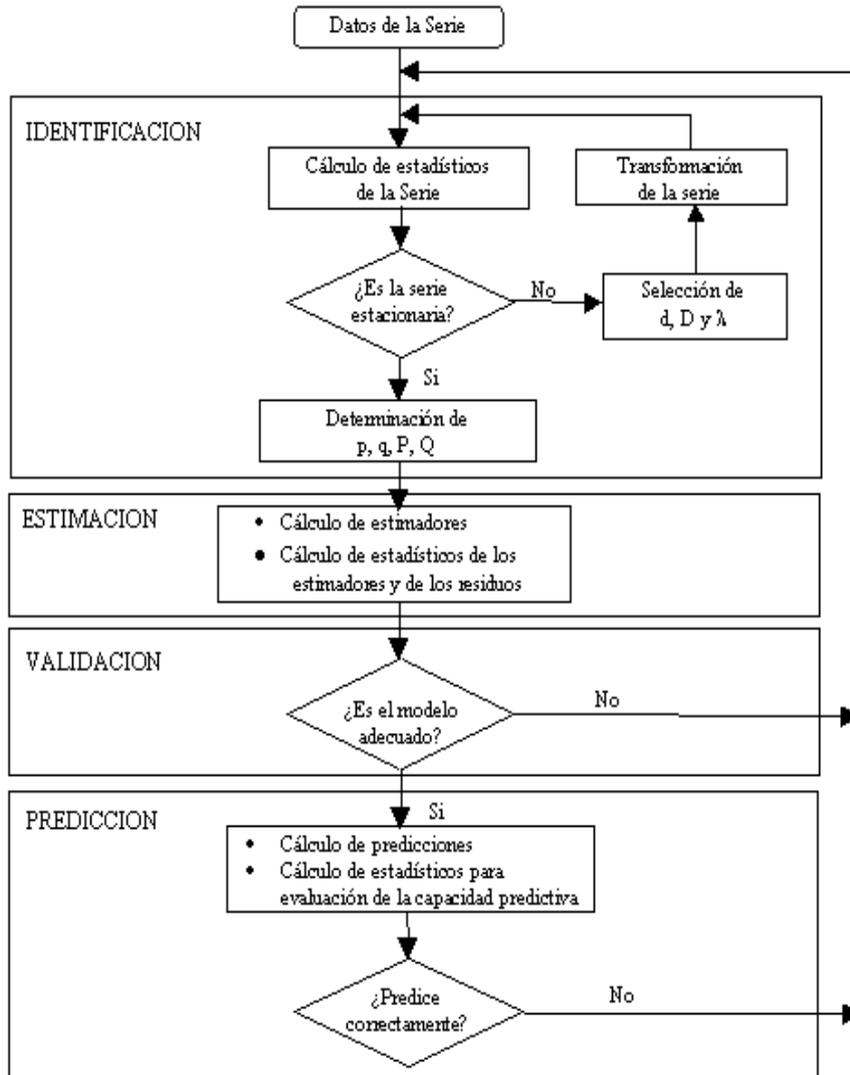
El contraste previamente explicado es conocido como DF aumentado, donde sus valores críticos para la estadística  $t(\gamma)$  son idénticos a la prueba DF, que permite utilizar las mismas tablas.

- b. Segunda fase:** Al presenciar estacionariedad la serie, los parámetros AR y MA del modelo se evalúan por máxima verosimilitud, se tiene los errores estándar y así como los residuales.
- c. La tercera fase:** Los residuos no tienen dependencia y son ruido blanco, caso contrario se modifica la estructura del modelo repitiendo los pasos anteriores hasta conseguir el correcto modelo.
- d. La cuarta fase:** Una vez obteniendo el modelo correcto se desarrollan las

predicciones pertinentes.

**Figura 1**

*Representación gráfica del proceso de realización de un modelo ARIMA*



Nota: Tomado de Series temporales. Modelos ARIMA por De la Fuente, 2013.

### 1.2.5. Caña de azúcar

Es una planta eficiente bajo una temperatura óptima del aire que oscila entre los 25 y 34 ° C. Conforme al informe del IV Censo Agropecuario 2012, efectuado por el Instituto Nacional de Estadística e Informática [INEI] en el 2013, menciona que en el Perú se cultiva aproximadamente 140,000 hectáreas de gramínea, generando casi 10 millones de toneladas de caña y un millón de toneladas de azúcar, en las regiones La Libertad y Lambayeque. A pesar de tener rendimientos elevados (10-12 toneladas por hectárea al mes), la producción de

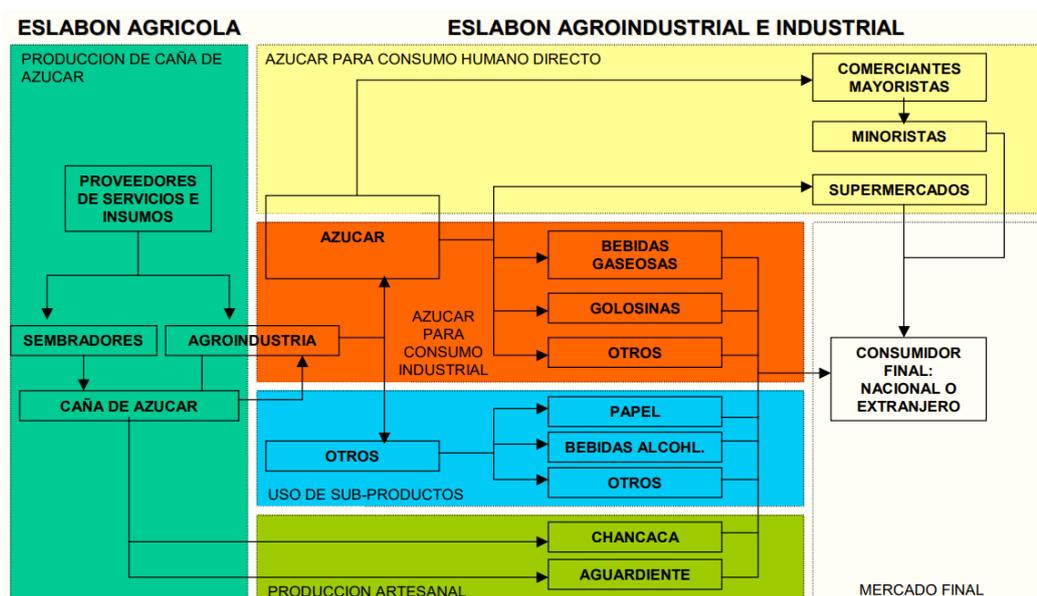
caña y su calidad se ven gravemente afectadas durante eventos de El Niño. (Pollack et al., 2018)

### 1.2.5.1. Proceso de producción

Se refiere al conjunto de actores involucrados en la producción, procesamiento y distribución de diversos productos de la gramínea. Estas actividades generan subcadenas, siendo la más destacada y relacionada con la producción de azúcar. A continuación, se presenta una estructura fundamental y se detallan los actores principales involucrados en esta cadena. (MIDAGRI, s.f.)

**Figura 2**

*Cadena productiva de la caña de azúcar en Perú*



Nota: Elaborado por DGPA-MINAG

### 1.2.5.2. Regionales de producción en el Perú

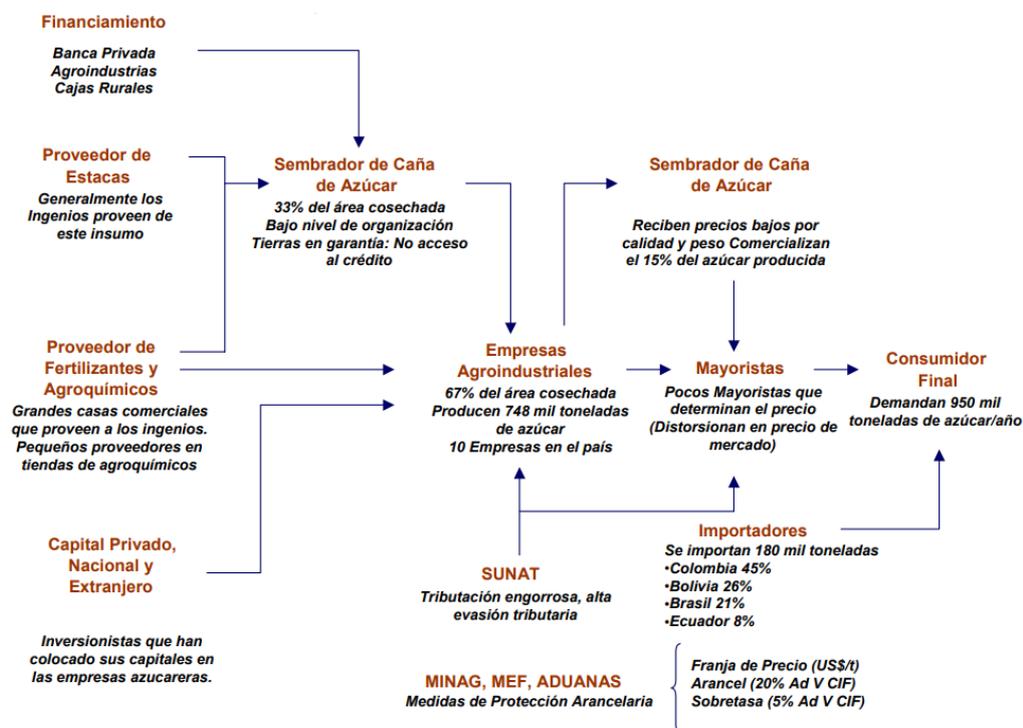
Está compuesta por las entidades agroindustriales que han surgido en el país debido a la transformación de las cooperativas como sociedades anónimas. Actualmente, solo funcionan 10 entidades en diferentes regiones de Perú: Lambayeque (Tumán, Pucalá y Pomalca); La Libertad (Casa Grande, Laredo y Cartavio); Ancash (San Jacinto); Lima (Paramonga y Andahuasi); Arequipa (Chucarapi). (MIDAGRI, s.f.)

Hoy en día, el estado de las empresas se caracteriza por una clara división, donde cuentan con una gestión técnica y administrativa efectiva, en su mayoría, han establecido asociaciones estratégicas. Además, se enfrentan a dificultades con

la gestión y experimentan problemas financieros, y se encuentran protegidas por el régimen patrimonial, impidiendo ser objeto de embargos. (MIDAGRI, s.f.)

**Figura 3**

*Proceso de Comercialización del Azúcar en el Perú*



Nota: Elaborado por MIDAGRI

Las compañías agroindustriales están afiliadas a la Asociación Peruana de Productores de Azúcar y Biocombustibles, que tiene la responsabilidad de elaborar las sugerencias al grupo agroindustrial para el estado y desempeña un papel activo en las conversiones comerciales internacionales en que el país tiene participación por asuntos relacionados con el azúcar. (MIDAGRI, s.f.)

### 1.2.5.3. Importación y exportación

El MIDAGRI (s.f.) menciona que en el 2021, Perú importó azúcar en bruto por un valor de \$78,6 millones, ubicándose en el puesto 74 como uno de los principales importadores internacionales. Durante ese año, el azúcar en bruto ocupó el puesto 133 en la lista de productos más importados en Perú. Las principales fuentes de importación de azúcar en bruto para Perú fueron Colombia (\$42,5 millones), Bolivia (\$12 millones), Guatemala (\$10,1 millones), Brasil (\$10,1 millones) y Nicaragua (\$1,88 millones). Entre 2020 y 2021, los mercados de importación de azúcar en bruto que registraron un mayor crecimiento en Perú

fueron Bolivia (\$4,38 millones), Paraguay (\$153,000) y Chile (\$96,200).

Además, en el 2021, Perú exportó azúcar en bruto por un valor de \$58,4 millones y se colocó en la posición 50 como exportador en el mundo. Asimismo, durante ese mismo año, el azúcar en bruto ocupó la posición número 75 de la lista de productos más exportados desde Perú. Los lugares más destacados a los que Perú exportó azúcar en bruto fueron Estados Unidos (\$35,2 millones), Colombia (\$16,2 millones), España (\$1,37 millones), Ecuador (\$1,36 millones) e Italia (\$1,1 millones). Entre 2020 y 2021, los mercados de exportación de azúcar en bruto que experimentaron un crecimiento más rápido para Perú fueron Ecuador (\$1,08 millones), Países Bajos (\$821,000) y Rumania (\$106,000).

#### **1.2.5.4. Características de la caña de azúcar en el Perú**

##### **a. Características biológicas**

Duarte y Gonzalez (2019) dice que es un cultivo semi-perenne que sigue produciendo durante muchos años, lo que implica la importancia de mantener la fertilidad del suelo a lo largo de su ciclo de vida. Además, se hace hincapié en la necesidad de incorporar materia orgánica y fertilizantes químicos para satisfacer los requerimientos nutricionales de la planta y garantizar un crecimiento saludable.

La fertilización se divide en dos etapas:

- ✓ Básica, se aplica en el fondo del surco y se mezcla con el suelo antes de plantar la caña.
- ✓ Cobertura, se aplica a una distancia de 5 a 10 cm de las hileras y luego se cubre con el suelo. Además, debe llevarse a cabo entre el cuarto y quinto mes posterior de la siembra, haciendo uso de la cultivadora en las hileras cada vez que el suelo tiene adecuada humedad. Es importante evitar retrasos en la fertilización de cobertura, dado que esto puede influir de manera desfavorable en el proceso de maduración.

##### **b. Características ambientales**

El cultivo la gramínea puede verse afectada por marcadas diferencias ambientales, lo que puede ocasionar daños a las plantas y, como consecuencia, pérdidas significativas en la cantidad producida. Algunos de los riesgos naturales

derivados de las condiciones meteorológicas incluyen eventos como la helada y la sequía. (Duarte y Gonzalez, 2019)

✓ Helada

Causa ruptura de las células debido a la congelación de su contenido líquido, detiene el proceso de maduración y gradualmente disminuye la cantidad de azúcar presente. La gravedad de los efectos está condicionada por la variedad de caña de azúcar, edad de la planta, rasgos del suelo y, en particular, la intensidad y duración de la helada. (Duarte y Gonzalez, 2019)

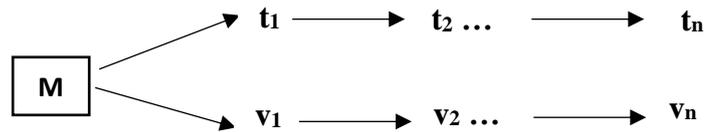
✓ Sequía

El crecimiento de la caña de azúcar puede ser afectado por tres tipos de sequías: la sequía fisiológica, causada por bajas temperaturas; la sequía atmosférica, resultado de temperaturas muy altas en climas tropicales secos; y la más perjudicial, la deficiencia de agua en el suelo, que se debe a la escasez de precipitación y tiene un efecto considerable en el desarrollo y productividad de la planta. (Duarte y Gonzalez, 2019)

## CAPITULO II. DISEÑO METODOLÓGICO

### 2.1. Diseño de contrastación de hipótesis/Procedimiento a seguir en la investigación.

Fue de diseño descriptivo longitudinal, dado que recoge información de la cantidad producida de caña de azúcar en el Perú, enero 1983 - abril 2021. El diseño siguiente fue adaptado del libro de Hernández et al. (2014):



Donde:

M: Muestra

$t_1, t_2, \dots, t_n$ : Tiempo

$v_1, v_2, \dots, v_n$ : Variable: Cantidad mensual producida de caña de azúcar en el Perú.

### 2.2. Población y muestra.

Se conformó 460 observaciones de la cantidad mensual producida de caña de azúcar en el Perú, de enero 1983 - abril 2021.

Se usó el muestreo no probabilístico por conveniencia ya que se tiene más facilidad de acceso a la muestra, por lo que posibilita a las autoras a realizar el estudio.

### 2.3. Técnicas e instrumentos de recolección de datos.

La técnica fue la observación documental indirecta estructurada de la variable.

El instrumento fue la Ficha electrónica en Excel “Registro de la cantidad mensual producida de caña de azúcar, enero 1983 – abril 2021” siendo conseguida de la página del INEI (ver anexo 1).

### 2.4. Técnicas estadísticas para el procesamiento de los datos.

Para el procesamiento de los datos se usó principalmente el programa estadístico SPSS. v.26 y con el apoyo del software libre RStudio, las cuales permitió determinar los estadísticos descriptivos como la media, desviación típica, coeficiente de variación, asimetría y curtosis; así mismo, se usó gráficas de línea y correlogramas de autocorrelación total y parcial, donde permitieron determinar diversos modelos basados en primera diferencia (ver anexo 2). Luego, se citaron aquellos modelos con coeficientes de estimación determinantes y, además se hizo uso de las pruebas de Q de Ljung-Box y kolmogorov

Smirnov para determinar cuál (es) es el modelo que cumplieron con los demás requisitos, es decir, residuales con distribución independiente y normal.

Se usó el Coeficiente de Determinación ( $R^2$ ), Error de raíz cuadrada media (RMSE), Error absoluto promedio (MAE), Error absoluto porcentual promedio (MAPE), Criterio de información de Akaike (AIC), Criterio de información de Akaike corregido (AICC), Criterio de Información Bayesiano (BIC) y el criterio de parsimonia, siendo estos dos últimos los más relevante, de acuerdo con la investigación realizada por Arévalo y Santisteban (2020), además, de Laurente y Laurente (2019) que tomó los criterios AIC y BIC.

## CAPITULO III. RESULTADOS

### 3.1. Principales características de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 abril 2021.

**Tabla 1**

*Estadísticos descriptivos de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021.*

Estadísticos	Valor
Media	650,8 MTM
Desviación típica	$\pm 198,5$ MTM
Coefficiente de variación	30.5%
Asimetría	0,272 MTM
Curtosis	-0,560 MTM
Mínimo	194 MTM
Máximo	1101,88 MTM
Observaciones	460

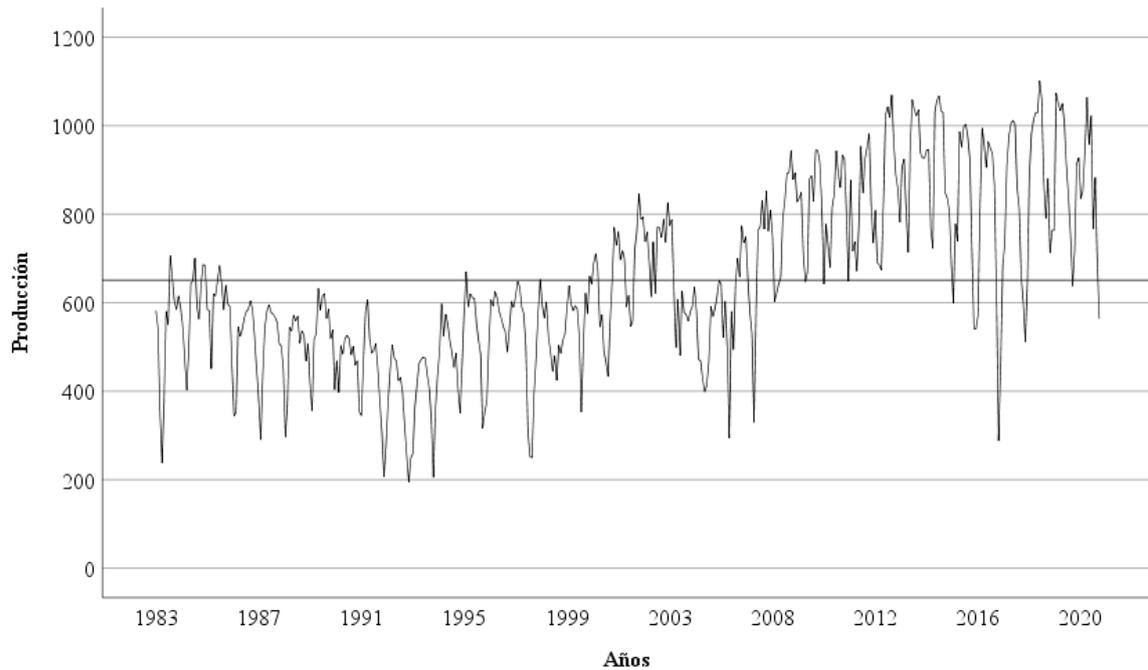
Nota: INEI, elaboración propia.

La tabla indica que la cantidad producida de caña de azúcar en el Perú contiene 460 observaciones correspondientes al periodo enero 1983 - abril 2021, las cuales la producción promedio fue de 650,8 MTM y tiene una alta variabilidad con  $\pm 198,5$  MTM, observando se también que los datos son heterogéneos con 30.5%. Por otro lado, la mayor producción de caña de azúcar fue 1101,88 MTM y la menor producción fue 194 MTM. Por último, se tiene una asimetría positiva de 0,272 MTM y una curtosis negativa de -0,560 MTM donde la distribución viene ser platicúrtica.

### 3.2. Estacionariedad y estacionalidad de cantidad mensual producida de caña de azúcar en el Perú, enero 1983 abril 2021.

**Figura 4**

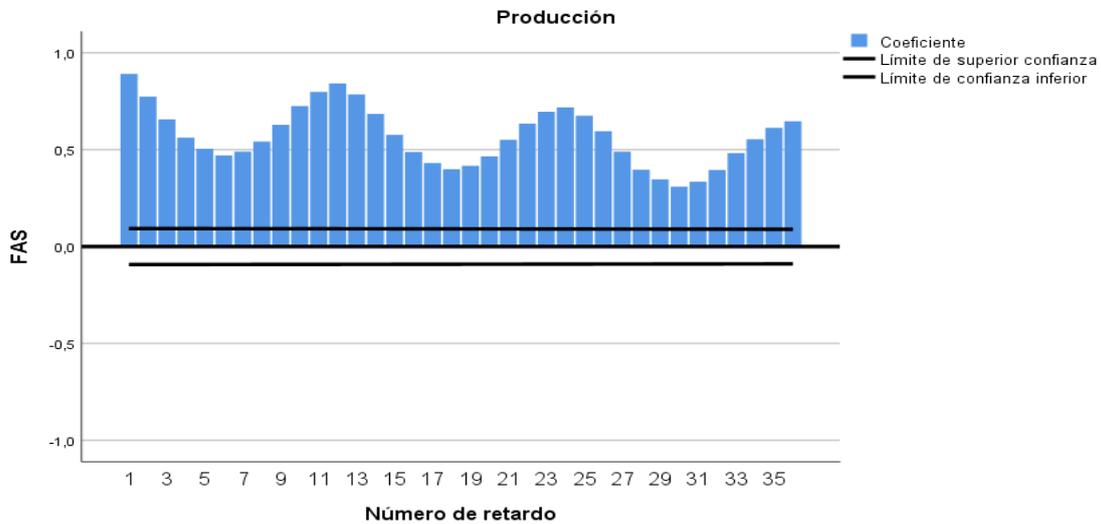
*Cantidad mensual producida de caña de azúcar en el Perú, periodo enero 1983 - abril 2021.*



La figura muestra que desde 1983 hasta el 2021 la cantidad mensual producida de caña de azúcar en los últimos años presenta una tendencia en crecimiento, pero sería irregular, dado que en cada año presenta picos altos y bajos.

**Figura 5**

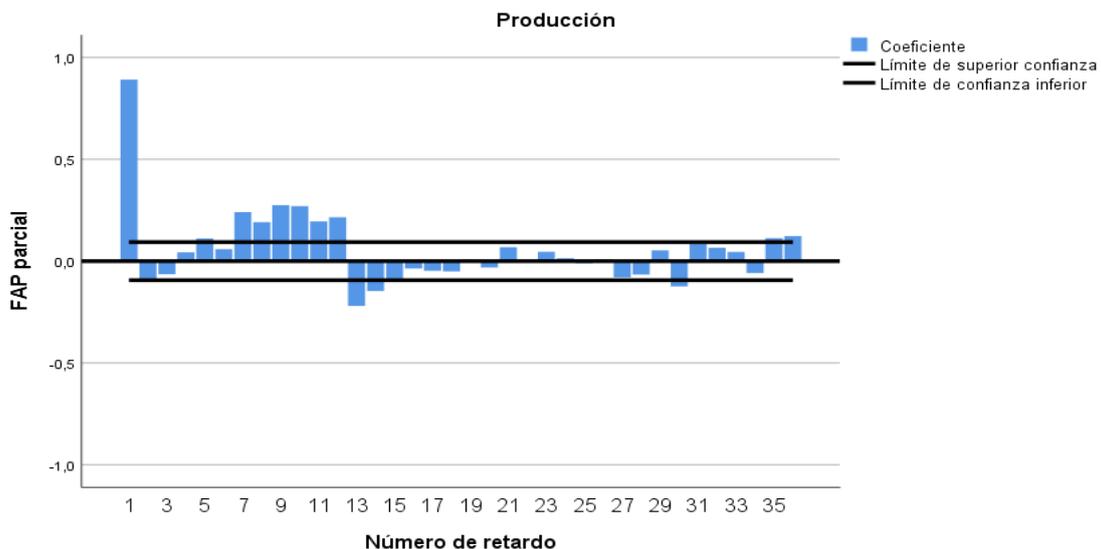
*Coefficientes de autocorrelación simple de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021.*



De la figura se puede observar un decaimiento rápido en los primeros coeficientes de autocorrelación simple de la serie, dando evidencia la ausencia de estacionariedad. Así mismo, se observa que la serie tiene un comportamiento estacional cada 12 periodos.

**Figura 6**

*Coefficientes de autocorrelación parcial de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021.*



La figura indica que los coeficientes tienen un decaimiento brusco en los primeros coeficientes de autocorrelación parcial y a su vez se observa la presencia de coeficientes significativos que se hallan fuera de los límites de confianza. Estos resultados indican que el modelo presenta componentes autorregresivos en la parte regular y estacional.

### 3.2.1. Análisis de la estacionariedad mediante Dickey – Fuller

**Tabla 2**

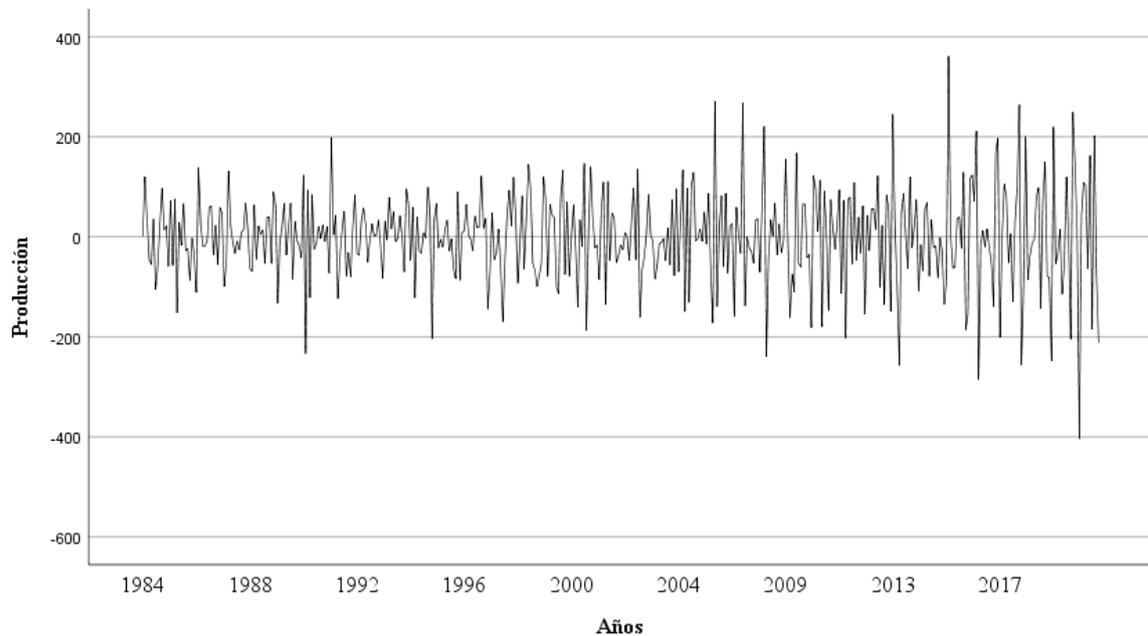
*Dickey Fuller Aumentado aplicada de la serie original.*

Variable	Estadísticos ADF	p-valor
Cantidad mensual producida de caña de azúcar	-1.5148	0,1335

En la tabla, el valor estadístico encontrado de la prueba ADF es -1.5148 y el p-valor= 0.1335 > 0.05, entonces se llega a la conclusión la no estacionariedad de la serie. En este sentido se usó las primeras diferencias de la serie.

### Figura 7

Cantidad mensual producida de caña de azúcar en el Perú, enero 1983 – abril 2021 en primeras diferencias.



Transformaciones: diferencia(1), diferencia estacional(1, periodo 12)

La figura indica que la cantidad mensual producida de caña de azúcar diferenciada en un periodo presenta un comportamiento cuyo promedio y varianza son constantes. Estos indicios sugieren que la serie podría ser estacionaria.

#### 3.2.2. Análisis de la estacionariedad mediante Dickey – Fuller en primeras diferencias.

Tabla 3

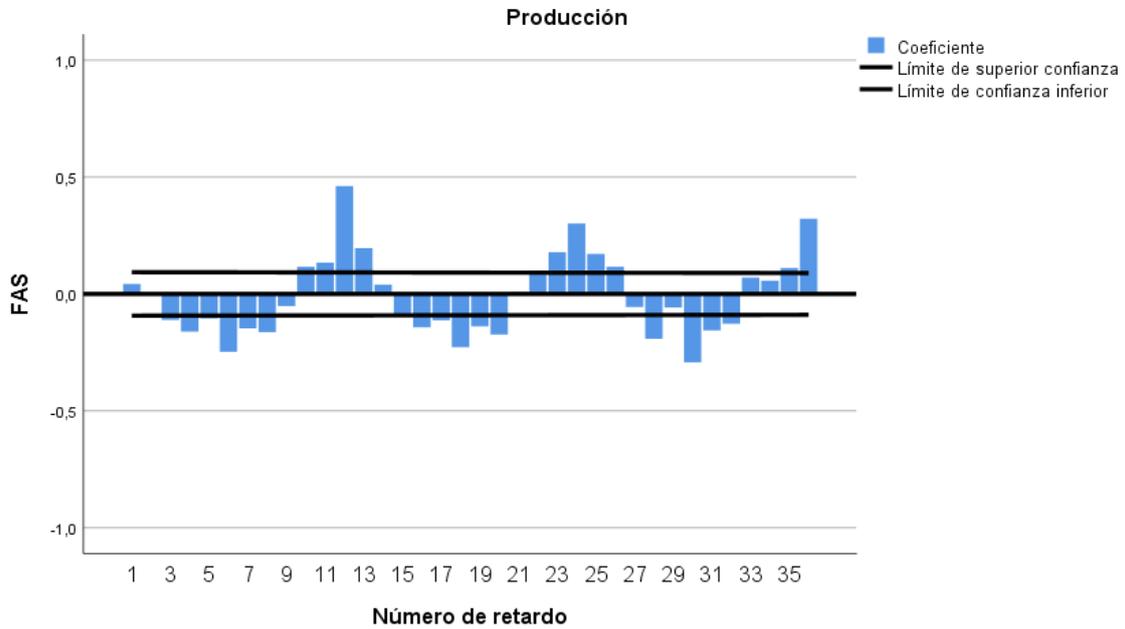
Dickey Fuller Aumentado aplicada de la serie original en primeras diferencias.

Variable	Estadísticos ADF	p-valor
Cantidad mensual producida de caña de azúcar	-14.7669	0,01

En la tabla, el valor estadístico encontrado de la prueba ADF es -14.7669 y el p-valor= 0.01 < 0.05, entonces se concluye que la serie en primeras diferencias es estacionaria.

### Figura 8

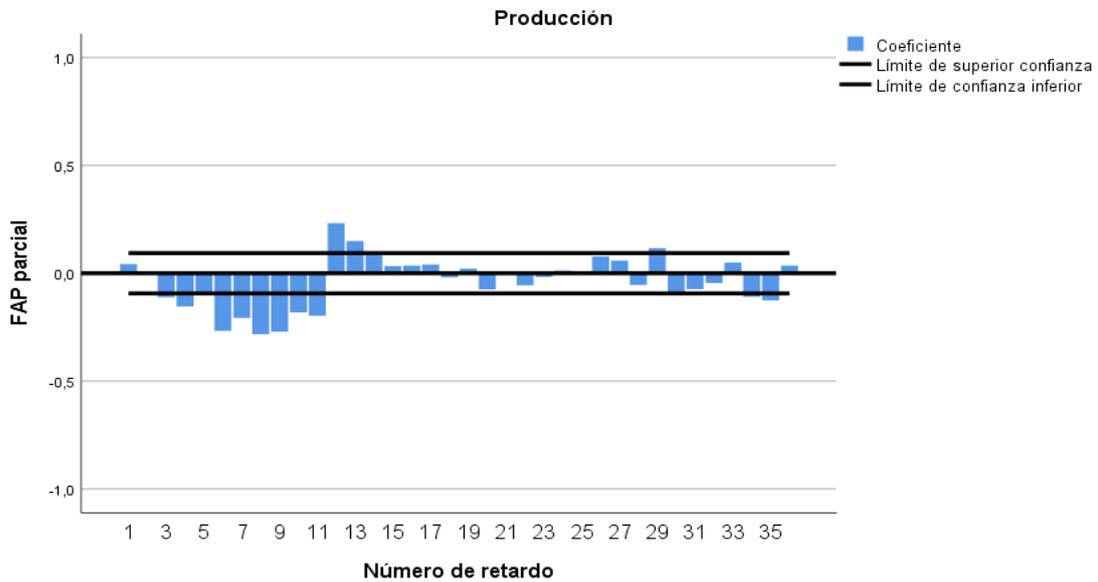
*Función de autocorrelación simple de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021 en primeras diferencias.*



La figura se visualiza los componentes de media móvil, las cuales se tienen los componentes 4, 6 y 8 para la parte regular y los componentes 12, 24 y 36 para la parte estacional; siendo los valores más significativos que se hallan fuera de los límites de confianza. En este sentido, el modelo presenta uno a tres componentes de media móviles (MA) para la parte regular y un componente para la parte estacional; permitiendo buscar los mejores modelos de pronóstico.

## Figura 9

*Función de autocorrelación parcial de la serie original diferenciada en un periodo tanto en la parte regular y estacional*



En la figura se visualiza los componentes autorregresivos, las cuales se tienen los componentes 6, 8 y 9 para la parte regular y solo el componente 12 para la parte estacional; siendo los valores más significativos que se hallan fuera de los límites de confianza. En este sentido, el modelo presenta uno a tres componentes autorregresivos (AR) regular y un componente estacional; permitiendo buscar los modelos de pronóstico más adecuados.

### 3.3. Modelo de serie de tiempo de la cantidad mensual producida de caña de azúcar en el Perú, enero 1983 abril 2021.

A continuación, se mencionan los modelos que satisficieron la significancia de los coeficientes de estimación, junto con sus respectivos parámetros. Además, se verificaron aquellos modelos cuyos residuales son independientes y se distribuyen normalmente.

#### SARIMA 4 (0, 1, 2) (1, 1, 1)

**Tabla 4**

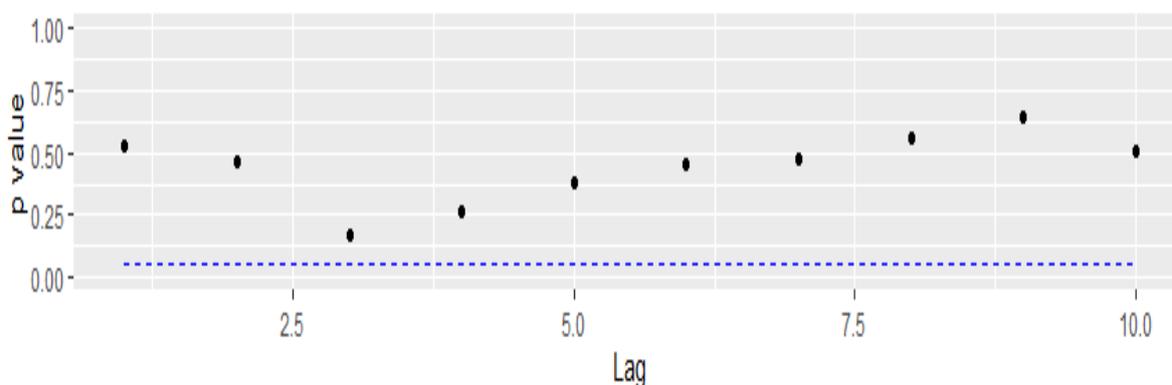
*Pruebas estadísticas aplicadas a los residuales con relación si tienden a una distribución independiente y normal del modelo SARIMA 4 (0, 1, 2) (1, 1, 1)*

Pruebas	Estadísticos	p-valor
Q de Ljung-Box	20,455	0,116
kolmogorov Smirnov	0.54783	0.9067

En la tabla se aprecia las pruebas estadísticas aplicados a los residuales con relación si tienden a una distribución independiente y normal. Para la prueba Q de Ljung-Box se encontró que el p-valor= 0.116 > 0.05, entonces se evidencia que los residuales se distribuyen independientemente. Para la prueba kolmogorov Smirnov se halló que el p-valor 0.9067 > 0.05, entonces se evidencia que los residuales se distribuyen normalmente.

**Figura 10**

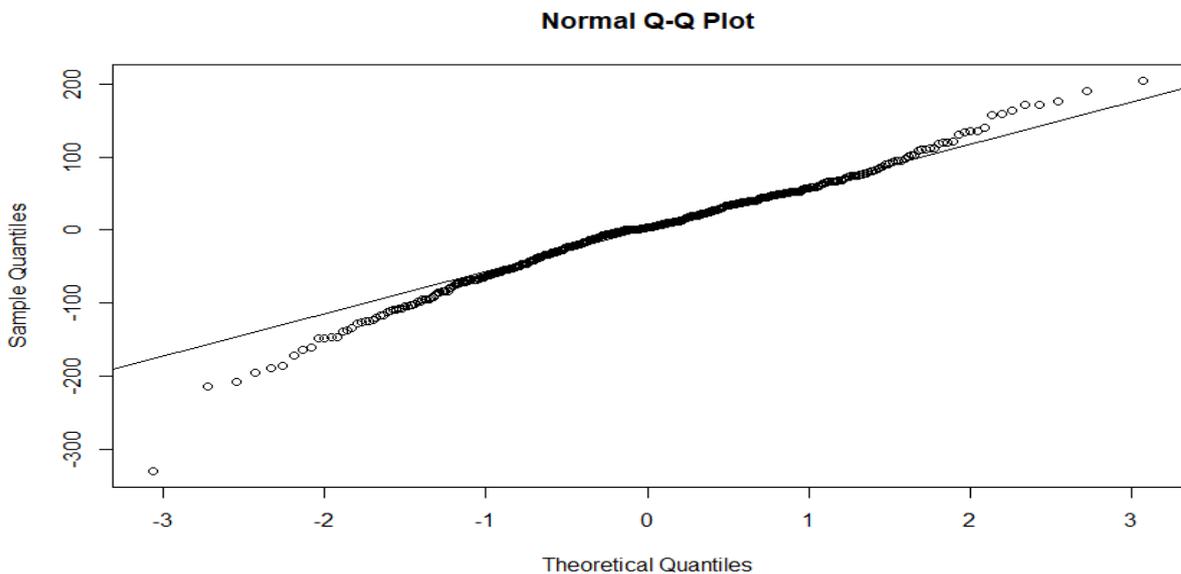
*Prueba de Ljung-Box de los residuales del modelo SARIMA 4 (0, 1, 2) (1, 1, 1)*



En la figura, se muestra que todos los coeficientes de autocorrelación conciernen a probabilidades mayores a 0.05, concluyendo que el modelo si es adecuado.

## Figura 11

Normalidad de los datos del modelo SARIMA 4 (0, 1, 2) (1, 1, 1) por la prueba de kolmogorov Smirnov



En la figura se aprecia la presencia de puntos que no se alinean con la recta, lo cual hace suponer que el modelo no se distribuye normalmente.

### Planteamiento de hipótesis del modelo SARIMA 4 (0, 1, 2) (1, 1, 1)

✓ Test de Ljung-Box corregido por Box - Pierce

$H_0$ : Los residuales del modelo SARIMA 4 (0, 1, 2) (1, 1, 1) se distribuyen de forma independiente

$H_a$ : Los residuales del modelo SARIMA 4 (0, 1, 2) (1, 1, 1) no se distribuyen independientemente

$\alpha = 0.05$

Decisión: Si  $p\text{-valor} > 0.05$ , entonces no se rechaza la  $H_0$ . Caso contrario se rechaza.

✓ Test de kolmogorov Smirnov.

$H_0$ : Los residuales del modelo SARIMA 4 (0, 1, 2) (1, 1, 1) proceden de una distribución normal.

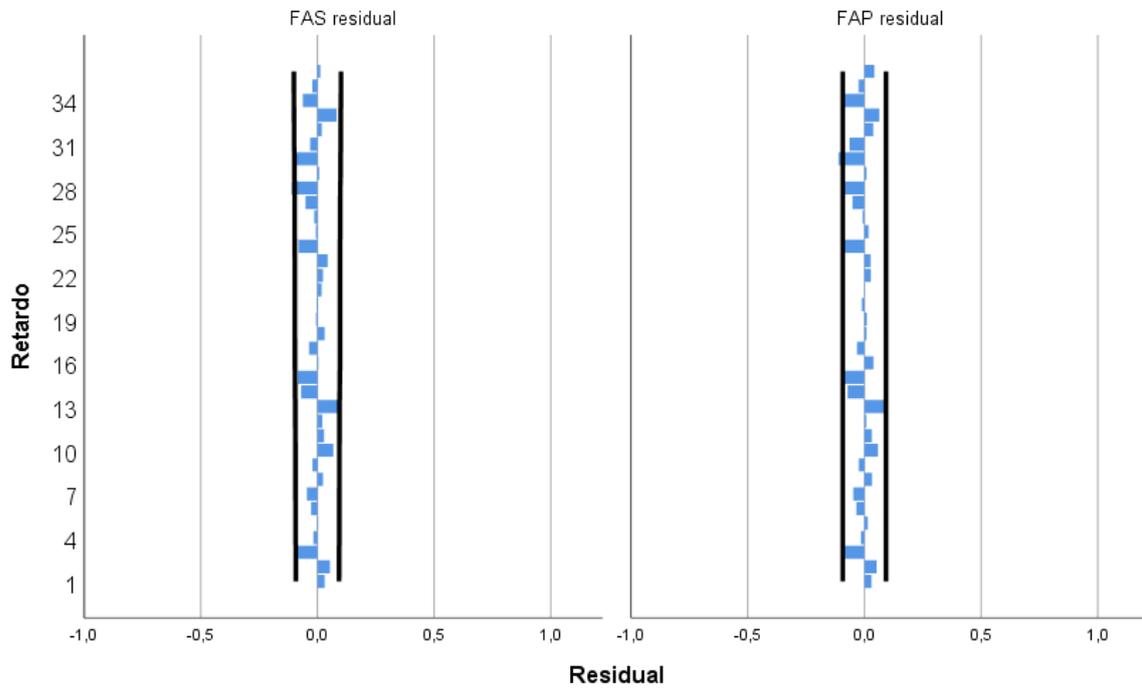
$H_a$ : Los residuales del modelo SARIMA 4 (0, 1, 2) (1, 1, 1) no se distribuyen normalmente.

$\alpha = 0.05$

Decisión: Si  $p\text{-valor} > 0.05$ , entonces no se rechaza la  $H_0$ . Caso contrario se rechaza.

**Figura 12**

*Función de autocorrelación simple y parcial de los residuales del modelo SARIMA 4 (0, 1, 2) (1, 1, 1)*



En la figura se observa que todos los coeficientes se hallan en los límites de confianza al 95%, esto demuestra que existen motivos para considerar que el modelo es adecuado.

En este sentido, de acuerdo con las pruebas aplicadas y los correlogramas de los residuales, se muestra que el modelo es adecuado.

### SARIMA 6 (0, 1, 3) (1, 1, 1)

**Tabla 5**

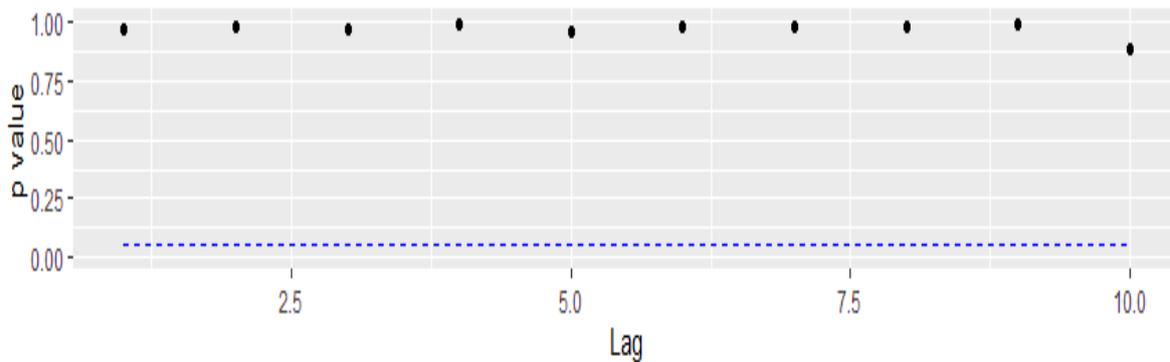
*Pruebas estadísticas aplicadas a los residuales con relación si tienden a una distribución independiente y normal del modelo SARIMA 6 (0, 1, 3) (1, 1, 1)*

Pruebas	Estadísticos	p-valor
Q de Ljung-Box	16,899	0,204
kolmogorov Smirnov	0.52826	0.9458

En la tabla se muestra las pruebas estadísticas aplicados a los residuales con relación si tienden a una distribución independiente y normal. Para la prueba Q de Ljung-Box se encontró que el p-valor= 0,204 > 0,05, entonces se evidencia que los residuales se distribuyen independientemente. Para la prueba kolmogorov Smirnov se halló que el p-valor 0.9458 > 0.05, entonces se evidencia que los residuales se distribuyen normalmente.

**Figura 13**

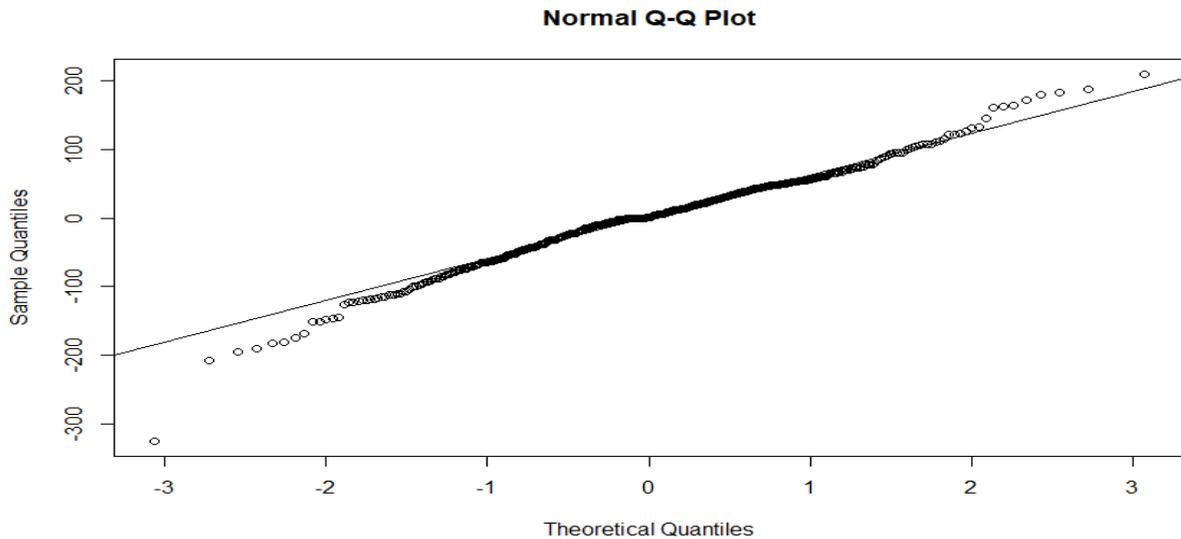
*Prueba de Ljung-Box de los residuales del modelo SARIMA 6 (0, 1, 3) (1, 1, 1)*



En la figura, que todos los coeficientes de autocorrelación conciernen a probabilidades mayores a 0.05, en este sentido se considera que el modelo si es adecuado.

### Figura 14

Normalidad de los datos del modelo SARIMA 6 (0, 1, 3) (1, 1, 1) por la prueba de kolmogorov Smirnov.



En la figura se aprecia la presencia de puntos que no se alinean con la recta, lo cual hace suponer que el modelo no se distribuye normalmente.

### Planteamiento de hipótesis del modelo SARIMA 6 (0, 1, 3) (1, 1, 1)

✓ Test de Ljung-Box corregido por Box - Pierce

$H_0$ : Los residuales del modelo SARIMA 6 (0, 1, 3) (1, 1, 1) se distribuyen de forma independiente

$H_a$ : Los residuales del modelo SARIMA 6 (0, 1, 3) (1, 1, 1) no se distribuyen independientemente

$\alpha = 0.05$

Decisión: Si  $p\text{-valor} > 0.05$ , entonces no se rechaza la  $H_0$ . Caso contrario se rechaza.

✓ Test de kolmogorov Smirnov.

$H_0$ : Los residuales del modelo SARIMA 6 (0, 1, 3) (1, 1, 1) proceden de una distribución normal.

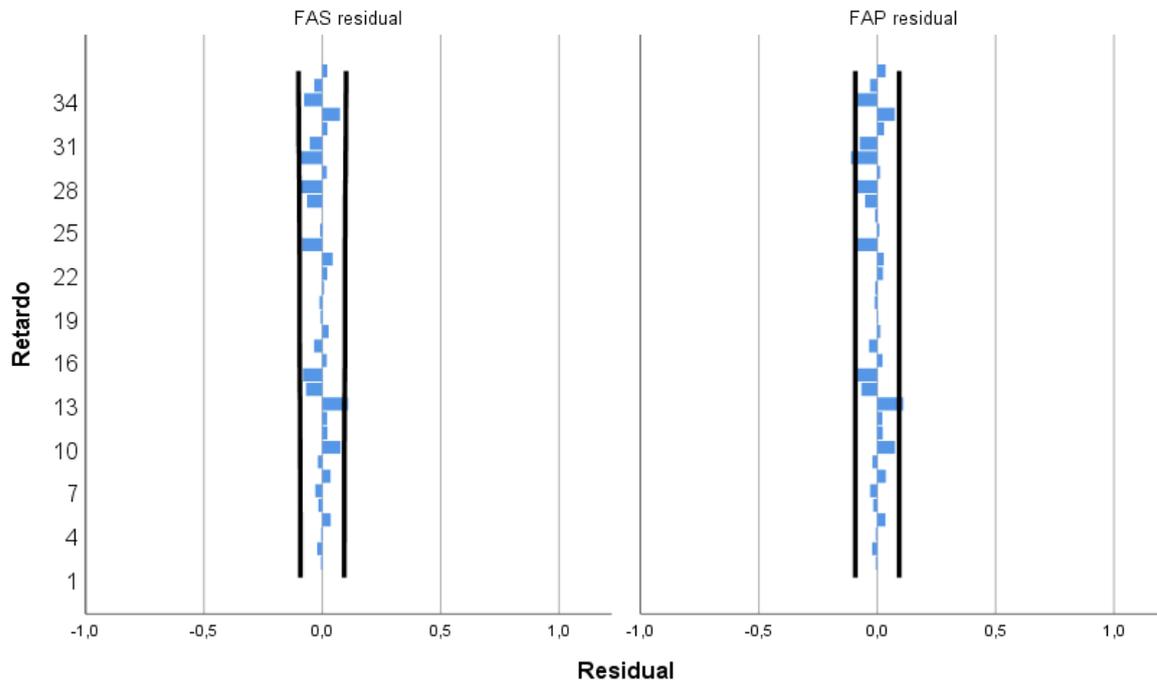
$H_a$ : Los residuales del modelo SARIMA 6 (0, 1, 3) (1, 1, 1) no se distribuyen normalmente.

$\alpha = 0.05$

Decisión: Si  $p\text{-valor} > 0.05$ , entonces no se rechaza la  $H_0$ . Caso contrario se rechaza.

**Figura 15**

*Función de autocorrelación simple y parcial de los residuales del modelo SARIMA 6 (0, 1, 3) (1, 1, 1)*



En la figura indica que los coeficientes se encuentran en los límites de confianza al 95%, esto demuestra que el modelo es adecuado.

En este sentido, de acuerdo con las pruebas aplicadas y los correlogramas de los residuales, se muestra claramente que el modelo es adecuado.

### SARIMA 8 (1, 1, 1) (1, 1, 1)

**Tabla 6**

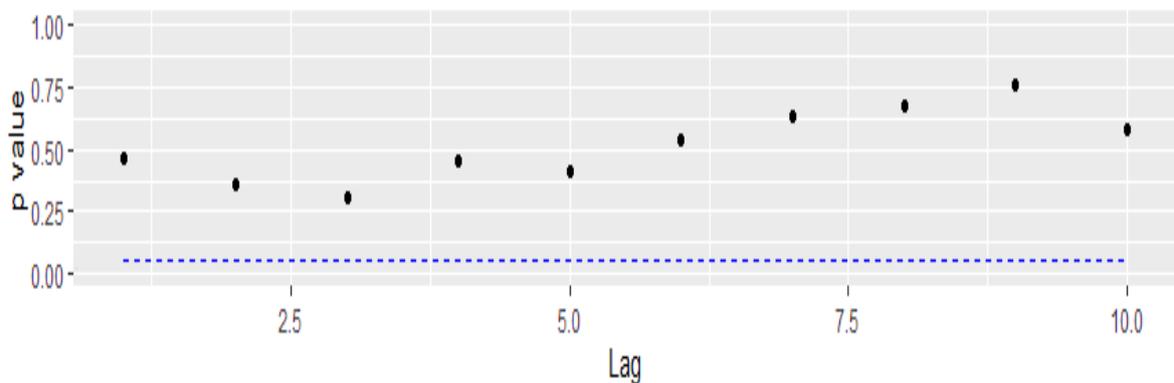
*Pruebas estadísticas aplicadas a los residuales con relación si tienden a una distribución independiente y normal del modelo SARIMA 8 (1, 1, 1) (1, 1, 1)*

Pruebas	Estadísticos	p-valor
Q de Ljung-Box	19,966	0,131
kolmogorov Smirnov	0.52391	0.9544

En la tabla se aprecia las pruebas estadísticas aplicados a los residuales con relación si tienden a una distribución independiente y normal. Para la prueba Q de Ljung-Box se encontró que el p-valor= 0,131 > 0,05, entonces se evidencia que los residuales se distribuyen independientemente. Para la prueba kolmogorov Smirnov se halló que el p-valor 0.9544 > 0.05, entonces se evidencia que los residuales se distribuyen normalmente.

**Figura 16**

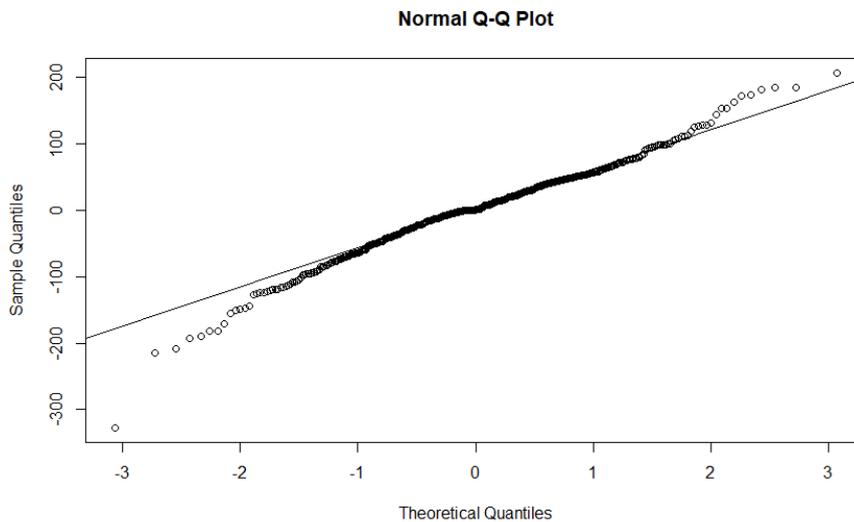
*Prueba de Ljung-Box de los residuales del modelo SARIMA 8 (1, 1, 1) (1, 1, 1)*



En la figura, se muestra que todos los coeficientes de autocorrelación conciernen a probabilidades mayores a 0.05, donde se concluye que el modelo si es adecuado.

## Figura 17

Normalidad de los datos del modelo SARIMA 8 (1, 1, 1) (1, 1, 1) por la prueba de kolmogorov Smirnov



En la figura se aprecia la presencia de puntos que no se alinean con la recta, lo cual sugiere la posibilidad que el modelo no se distribuye normalmente.

### Planteamiento de hipótesis del modelo SARIMA 8 (1, 1, 1) (1, 1, 1)

✓ Test de Ljung-Box corregido por Box - Pierce

H<sub>0</sub>: Los residuales del modelo SARIMA 8 (1, 1, 1) (1, 1, 1) se distribuyen de forma independiente

H<sub>a</sub>: Los residuales del modelo SARIMA 8 (1, 1, 1) (1, 1, 1) no se distribuyen independientemente.

$\alpha = 0.05$

Decisión: Si p-valor > 0.05, entonces no se rechaza la H<sub>0</sub>. Caso contrario se rechaza.

✓ Test de kolmogorov Smirnov.

H<sub>0</sub>: Los residuales del modelo SARIMA 8 (1, 1, 1) (1, 1, 1) proceden de una distribución normal.

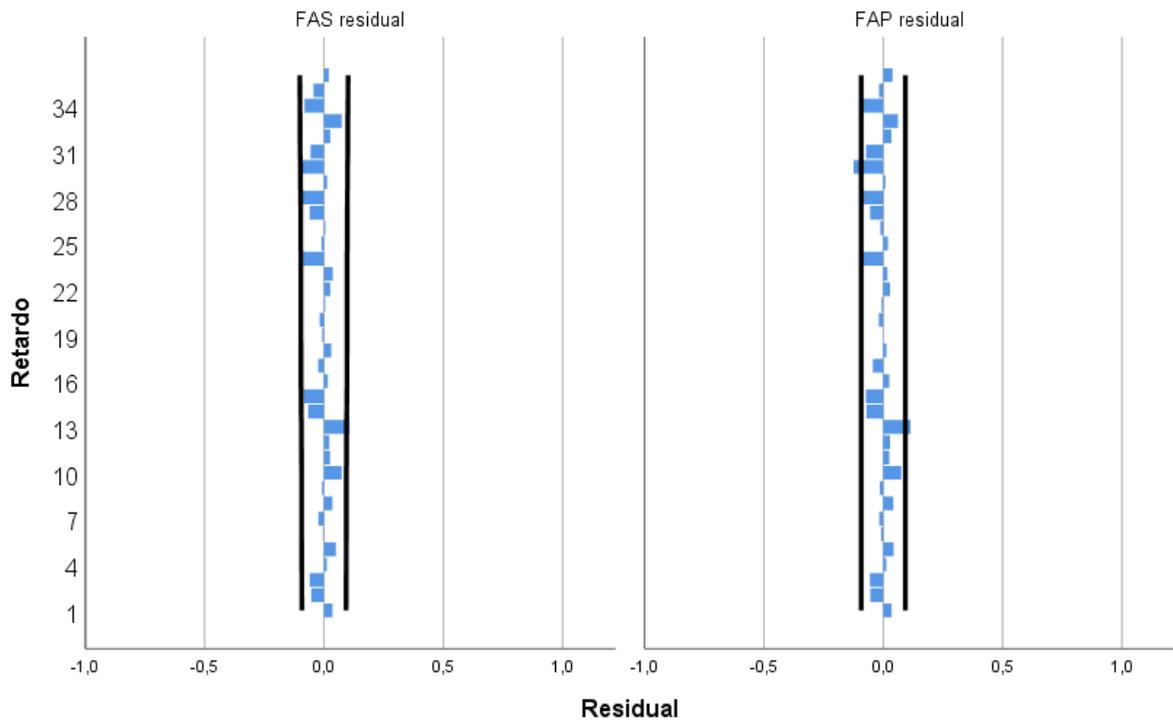
H<sub>a</sub>: Los residuales del modelo SARIMA 8 (1, 1, 1) (1, 1, 1) no se distribuyen normalmente.

$\alpha = 0.05$

Decisión: Si p-valor > 0.05, entonces no se rechaza la H<sub>0</sub>. Caso contrario se rechaza.

### Figura 18

Función de autocorrelación simple y parcial de los residuales del modelo SARIMA 8 (1, 1, 1) (1, 1, 1)



En la figura indica que todos los coeficientes se hallan en los límites de confianza al 95%, excepto dos que no son tan significativos, lo que demuestra que el modelo es el adecuado.

En este sentido, de acuerdo con las pruebas aplicadas y los correlogramas de los residuales, se muestra claramente que el modelo es adecuado.

### SARIMA 16 (2, 1, 2) (1, 1, 1)

**Tabla 7**

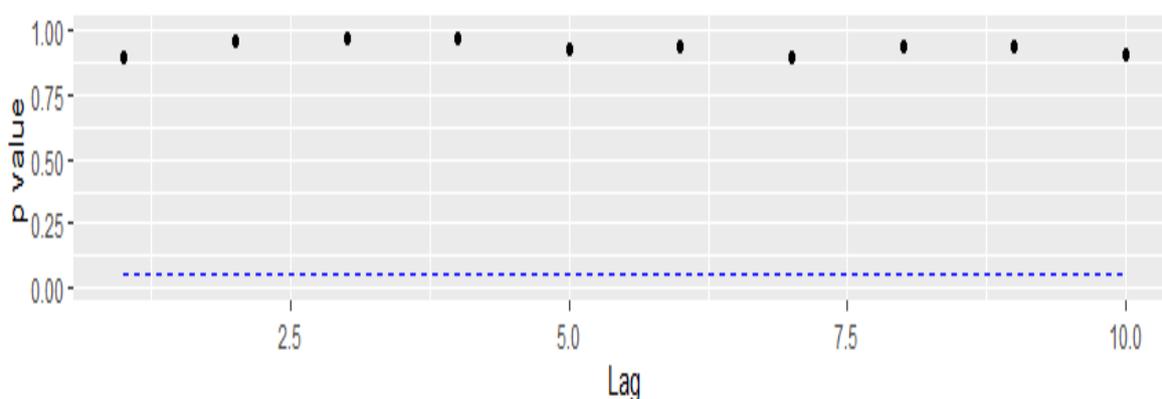
*Pruebas estadísticas aplicadas a los residuales con relación si tienden a una distribución independiente y normal del modelo SARIMA 16 (2, 1, 2) (1, 1, 1)*

Pruebas	Estadísticos	p-valor
Q de Ljung-Box	20,455	0,116
kolmogorov Smirnov	0.54348	0.9154

En la tabla se aprecia las pruebas estadísticas aplicados a los residuales con relación si tienden a una distribución independiente y normal. Para la prueba Q de Ljung-Box se encontró que el p-valor= 0,209 > 0,05, entonces se evidencia que los residuales se distribuyen independientemente. Para la prueba kolmogorov Smirnov se halló que el p-valor 0.9154 > 0.05, entonces se evidencia que los residuales se distribuyen normalmente.

**Figura 19**

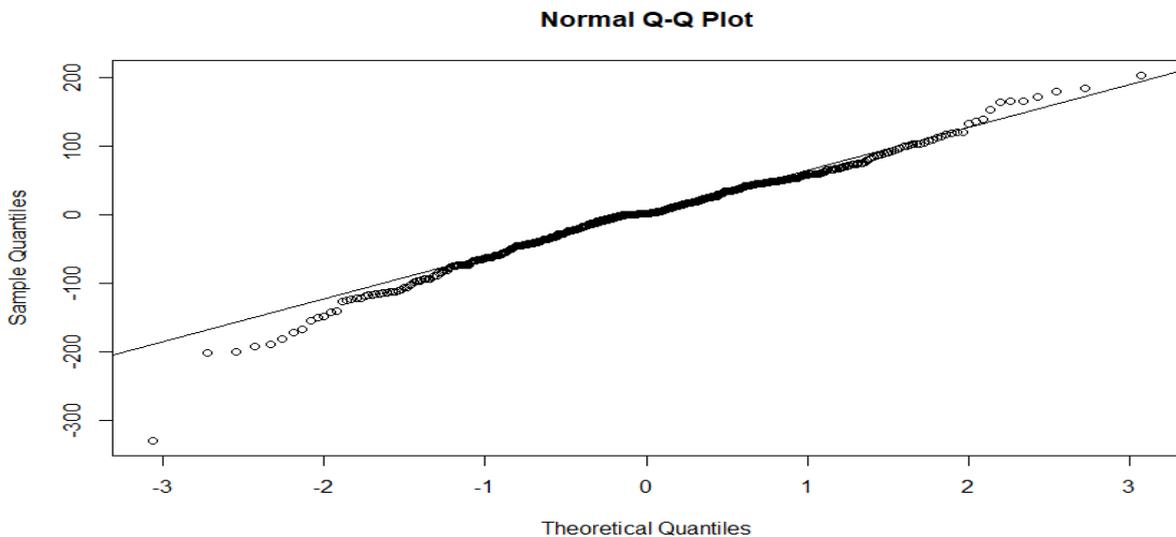
*Prueba de Ljung-Box de los residuales del modelo SARIMA 16 (2, 1, 2) (1, 1, 1)*



En la figura, se muestra que todos los coeficientes de autocorrelación conciernen a probabilidades mayores a 0.05, lo cual se considera que el modelo si es adecuado.

## Figura 20

Normalidad de los datos del modelo SARIMA 16 (2, 1, 2) (1, 1, 1) por la prueba de kolmogorov Smirnov



En la figura se aprecia la presencia de puntos que no se alinean con la recta, lo cual da entender que el modelo no se distribuye normalmente.

### Planteamiento de hipótesis del modelo SARIMA 16 (2, 1, 2) (1, 1, 1)

✓ Test de Ljung-Box corregido por Box - Pierce

$H_0$ : Los residuales del modelo SARIMA 16 (2, 1, 2) (1, 1, 1) se distribuyen de forma independiente.

$H_a$ : Los residuales del modelo SARIMA 16 (2, 1, 2) (1, 1, 1) no se distribuyen independientemente.

$\alpha = 0.05$

Decisión: Si p-valor  $> 0.05$ , entonces no se rechaza la  $H_0$ .

Si p-valor  $< 0.05$ , entonces se rechaza la  $H_0$ .

✓ Test de kolmogorov Smirnov.

$H_0$ : Los residuales del modelo SARIMA 16 (2, 1, 2) (1, 1, 1) proceden de una distribución normal.

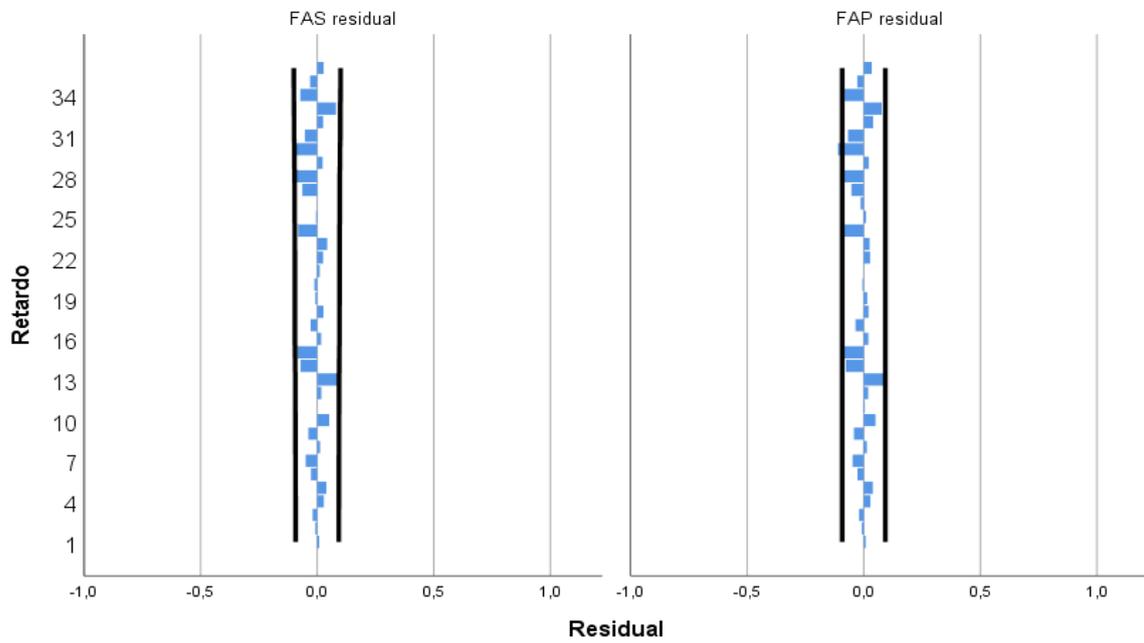
$H_a$ : Los residuales del modelo SARIMA 16 (2, 1, 2) (1, 1, 1) no se distribuyen normalmente.

$\alpha = 0.05$

Decisión: Si p-valor  $> 0.05$ , entonces no se rechaza la  $H_0$ . Caso contrario se rechaza.

### Figura 21

*Función de autocorrelación simple y parcial de los residuales del modelo SARIMA 16 (2, 1, 2) (1, 1, 1)*



En la figura indica que existen coeficientes en los límites de confianza al 95%, demostrando que el modelo es el adecuado.

En este sentido, de acuerdo con las pruebas aplicadas y los correlogramas de los residuales, se muestra claramente que el modelo es adecuado.

**Tabla 8***Evaluación de los modelos propuestos por medio de los errores de pronósticos.*

<b>Modelos</b>	<b>R<sup>2</sup></b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>	<b>AIC</b>	<b>AIC<sub>c</sub></b>	<b>BCI</b>
SARIMA 4 (0, 1, 2) (1, 1, 1)	0.875	70.501	53.718	9.145	5076.04	5076.17	5096.55
SARIMA 6 (0, 1, 3) (1, 1, 1)	0.886	67.756	52.037	8.732	5073.05	5073.24	5097.67
SARIMA 8 (1, 1, 1) (1, 1, 1)	0.885	67.988	51.973	8.721	5075.01	5075.14	5095.52
SARIMA 16 (2, 1, 2) (1, 1, 1)	0.886	67.844	52.165	8.765	5074.52	5074.77	5103.24

En la tabla se muestra los modelos que resultaron ser satisfactorios y se encontró que el SARIMA 8 (1, 1, 1) (1, 1, 1) es el mejor modelo que pronostica la variable, debido a que contiene los mejores errores de pronóstico, es parsimonioso, tiene cuatro parámetros por estimar, tiene a su favor el BCI y con un AIC muy cercano al modelo 6. Si bien es cierto los demás modelos contiene similares errores de pronósticos que el modelo seleccionado, estos tienen más de cuatro parámetros por estimar y tienen un mayor BIC.

✓ Coeficiente de determinación  $R^2$ :

- Gonzalez (2018) destaca que el  $R^2$  es una medida de la calidad y aptitud de un modelo, siendo comúnmente utilizado con propósitos descriptivos. Indica que las variables independientes elegidas son capaces de exponer la variabilidad en las variables dependientes.
- A su vez, este coeficiente tiene la propiedad intuitiva de una escala de 0 a 1, donde 0 indica que la predicción del modelo no experimenta mejoras sobre el modelo medio, mientras que 1 muestra una predicción óptima.

✓ Error cuadrático medio (RMSE):

- Mora (2023) destaca que el RMSE emerge como una métrica de evaluación ampliamente empleada en el análisis de series de tiempo y en el ámbito estadístico. Su función principal consiste en medir la precisión de las predicciones al compararlas con los valores reales.
- A su vez, habla que el RMSE calcula la raíz cuadrada de la media de los errores al cuadrado entre las predicciones del modelo y los valores observados de la serie de tiempo. En otras palabras, mide cuán dispersos están los errores.
- Mientras que Gonzalez (2018) considera que el RMSE proporciona una evaluación efectiva de la precisión con la cual el modelo realiza predicciones, siendo el criterio principal a considerar si el objetivo principal del modelo es la predicción. Valores más bajos de RMSE señalan un ajuste más óptimo.
- Por otra parte, Gonzalez (2018), considera que el RMSE se percibe como una métrica eficaz para evaluar la precisión con la que el modelo realiza predicciones, siendo considerado el criterio más crucial cuando el objetivo principal del modelo es la predicción. Siendo que un RMSE muy bajo sugiere ser el más óptimo.

✓ Error absoluto medio (MAE)

- Mora (2023) destaca que el MAE emerge como una métrica de evaluación ampliamente empleada tanto en el análisis de series de tiempo como en el ámbito estadístico. Su función principal radica en medir la precisión de las predicciones al compararlas con los valores reales. A diferencia del RMSE, el MAE se caracteriza por no considerar el cuadrado de las diferencias entre las predicciones y los valores reales. Esta particularidad

confiere al MAE una menor sensibilidad a valores atípicos y errores de magnitudes extremadamente grandes.

- Del mismo modo, se sostiene que el MAE se configura como una métrica de error que calcula el promedio de las diferencias absolutas entre las estimaciones del modelo y los valores observados. Un MAE más bajo, en este contexto, señala una mayor precisión en las predicciones del modelo.
  - Adicionalmente, el autor argumenta que el MAE muestra una menor sensibilidad a valores atípicos en comparación con el RMSE. Si hay valores extremos en los datos, el MAE se destaca al proporcionar una medida más precisa del modelo.
- ✓ El Porcentaje de Error Absoluto Medio (MAPE):
- Mora (2023) menciona que el MAPE se erige como una métrica ampliamente empleada en el análisis de series de tiempo y en el ámbito estadístico para la evaluación de la veracidad de las estimaciones en contraste con los valores reales. En contraposición al MAE y al RMSE, el MAPE se presenta como un porcentaje, simplificando así la interpretación de la precisión de las predicciones en términos relativos.
  - Además, el MAPE mide el error promedio como un porcentaje de los valores reales. En este contexto, un valor más bajo de MAPE refleja una mayor precisión en las predicciones del modelo.
  - También el autor destaca que, al presentarse como un porcentaje, el MAPE posibilita una interpretación más intuitiva de la precisión de las predicciones en relación con el tamaño de los valores reales.
- ✓ Índice de Información de Akaike (AIC)
- Mora (2023) señala que el AIC desempeña un papel crucial como criterio de selección de modelos en estadísticas y análisis de series de tiempo. Su función primordial consiste en comparar diversos modelos estadísticos y discernir cuál de ellos se adapta de manera óptima a los datos observados.
  - Así mismo, el AIC persigue hallar un equilibrio entre la capacidad del modelo para ajustarse adecuadamente a los datos y su complejidad intrínseca. El propósito es

seleccionar un modelo que exhiba un buen ajuste a los datos sin incurrir en una complejidad excesiva, evitando así el sobreajuste.

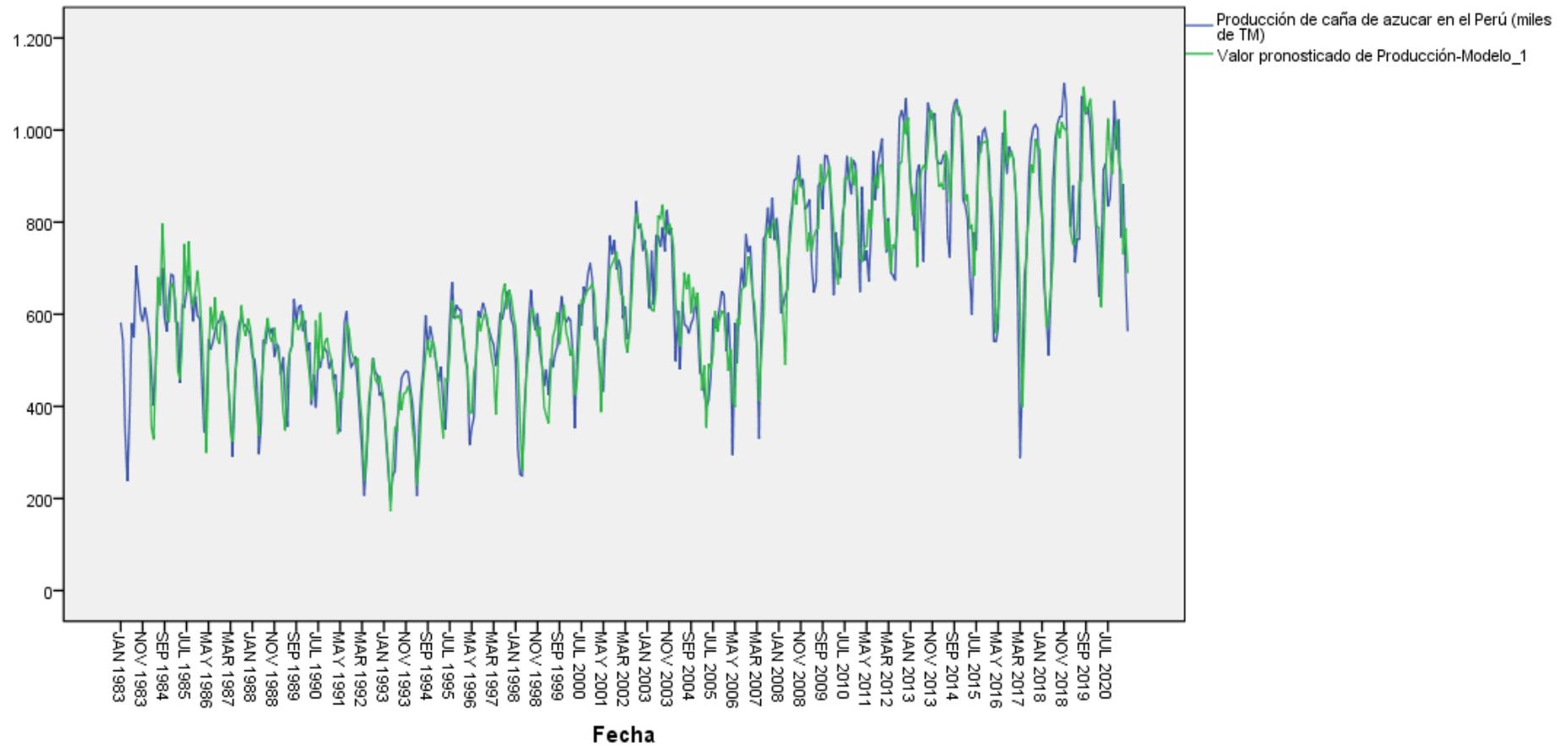
- Además, se prefiere un valor más bajo del AIC. De entre diversos modelos candidatos, aquel que exhiba el AIC más reducido se considera superior en términos de lograr un equilibrio adecuado entre el ajuste a los datos y la complejidad del modelo.

✓ Criterio de Información Bayesiano (BCI)

- Mora (2023) habla que el BCI, constituye una métrica empleada en estadísticas y análisis de series de tiempo, compartiendo similitudes con el AIC. Su función principal es facilitar la elección entre modelos estadísticos alternativos. El BIC se hace uso para analizar la calidad de ajuste de un modelo a los datos observados y, de manera análoga al AIC, impone una penalización a la complejidad del modelo.
- Así mismo el BIC comparte similitudes con el AIC al penalizar modelos de mayor complejidad mediante la inclusión de un término que depende del número de parámetros (segundo término en la fórmula). No obstante, el BIC impone una penalización más intensa a la complejidad del modelo en comparación con el AIC. En consecuencia, el BIC tiende a favorecer de manera más pronunciada modelos más simples que su contraparte, el AIC.
- Por último, se prefiere un valor más bajo de BIC. De entre diversos modelos candidatos, aquel que exhiba el BIC más reducido se considera el mejor en términos de lograr un ajuste óptimo a los datos y mantener la simplicidad del modelo.

**Figura 22**

Valores de la serie original y del pronóstico del modelo SARIMA 8 (1, 1, 1) (1, 1, 1)



En la figura se aprecia que los valores del pronóstico del modelo seleccionado se ajustan a los valores de la serie original en estudio con relación a los otros modelos anteriormente mencionados.

**3.4. Pronosticar la cantidad mensual producida de caña de azúcar en el Perú, para los años mayo 2021, diciembre 2023.**

**Tabla 9**

*Pronóstico de cantidad mensual producida de caña de azúcar en el Perú, para los años mayo 2021, diciembre 2023*

<b>MESES</b>	<b>SARIMA 8 (1, 1, 1) (1, 1, 1)</b>
Mayo 2021	732.40
Junio 2021	807.70
Julio 2021	842.35
Agosto 2021	909.96
Setiembre 2021	929.44
Octubre 2021	983.62
Noviembre 2021	947.52
Diciembre 2021	952.81
Enero 2022	819.53
Febrero 2022	800.25
Marzo 2022	721.20
Abril 2022	636.74
Mayo 2022	740.18
Junio 2022	805.58
Julio 2022	861.67
Agosto 2022	936.69
Setiembre 2022	934.84
Octubre 2022	971.23

---

Noviembre 2022	952.98
Diciembre 2022	942.49
Enero 2023	840.90
Febrero 2023	786.50
Marzo 2023	736.19
Abril 2023	663.59
Mayo 2023	749.95
Junio 2023	812.78
Julio 2023	874.47
Agosto 2023	951.42
Setiembre 2023	944.03
Octubre 2023	975.81
Noviembre 2023	962.21
Diciembre 2023	947.63

---

La tabla indica los pronósticos del modelo que mejor pronostica la variable, para los años mayo 2021 hasta diciembre 2023.

## CAPITULO IV. DISCUSIÓN DE RESULTADOS

La finalidad del estudio fue encontrar el mejor modelo de pronóstico de la cantidad mensual de la caña de azúcar en el Perú aplicando la metodología de Box Jenkins, y se logró identificar al modelo 8: SARIMA (1, 1, 1) (1, 1, 1) debido a que tuvo los menores errores de pronósticos y por ser parsimonioso.

A continuación, se presenta la ecuación del modelo:

$$(1 - \alpha_1 L)(1 - B_1 L^{12})((Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})) = (1 - \phi_1 L)(1 - \varphi_1 L^{12})e_t$$

Tipo	Coeficiente	p-valor
AR 1	0,439	0,000
MA 1	0,863	0,000
SAR 1	0,239	0,000
SMA 1	0,945	0,000

Los resultados conseguidos contrastan con los siguientes estudios:

En el estudio de Laurente y Laurente (2019) no coincide con el modelo hallado en el presente estudio, dado que aplicaron el modelo ARIMA para la producción de la papa en Puno, teniendo como el mejor modelo de pronóstico, el ARIMA (3, 0, 4) que se fundamentó con la evaluación de los criterios de AIC y BIC.

La investigación citada ayudó a las autoras en la selección del modelo de pronóstico más apropiado mediante los criterios de AIC y BIC, debido a que los errores de pronósticos como el RMSE, MAE y MAPE, además, del  $R^2$  son muy similares entre sí.

Mientras que si coincide con el estudio de Arévalo y Santisteban (2020), quienes aplicaron el método Box Jenkins y encontraron como modelo óptimo de pronóstico el SARIMA (2,1,2) (1,1,0)<sub>12</sub> teniendo como ecuación general:  $(1 - \alpha_1 L - \alpha_1 L^2)(1 - \beta_1 L^{12})((Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})) = (1 - \phi_1 L - \phi_2 L^2)e_t$ . Para la elección del mejor modelo se basaron por medio del principio de parsimonia y de los errores de pronósticos, sobre todo del criterio de BIC.

La investigación citada ayudó a las autoras a la elección del mejor modelo de pronóstico, tomando en consideración al criterio de parsimonia y del menor BIC, dado que se encontró valores similares de los errores de pronósticos.

Además, se tiene el estudio de Valverde (2021) que no coincide con el modelo de pronóstico encontrado en la presente investigación, esto debido a que halló un modelo de pronóstico ARIMA (1,1,12).

A pesar de que esta investigación citada no considera el criterio de parsimonia ni el BIC, a diferencia de las autoras que, si determinaron el mejor modelo de pronóstico bajo esos criterios, por el motivo que se halló valores de los errores de pronósticos similares. Valverde considera importante los coeficientes AR y MA su significancia ( $p$ -valor  $< 0.05$ )

A su vez, se tiene el estudio de Fuentes (2017) que no coincide con el modelo obtenido en la investigación, debido a que obtuvo un modelo para el ingenio la Unión:  $y = 82.624x + 65.889$ , con un  $R^2 = 0.1269$  muy bajo y un modelo para el ingenio Santa Ana:  $y = 24.74x + 99.074$  el  $R^2 = 0.0349$  muy bajo.

Por último, se tiene el estudio de Pérez (2017), que tampoco no coincide con la investigación, dado que se halló un modelo lineal de la Zafra 2014-2015:  $y = 52.002x + 81.674$ , con un  $R^2 = 0.0504$ ; lo mismo para modelo por Zafra y del primer tercio:  $y = 61.355x + 75.721$  que tuvo un  $R^2 = 0.0493$ . El modelo lineal de la relación NDVI y rendimiento de la gramínea de las zafra 2013-2015, fue:  $y = 28.366x + 93.244$  con  $R^2 = 0.0253$ . En este sentido los modelos no resultaron tener asociación, el NDVI y la producción de la caña.

Las investigaciones citadas proporcionan una base sólida para la evaluación y la identificación de un  $R^2$  apropiado, permitiendo a las autoras la determinación del mejor modelo de pronóstico.

## CAPITULO V. CONCLUSIONES

1. El modelo estadístico para el pronóstico de cantidad mensual producida de caña de azúcar en el Perú fue el SARIMA (1, 1, 1) (1, 1, 1) y su modelo respectivo es:  
$$(1 - \alpha_1 L)(1 - B_1 L^{12})((Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})).$$
2. La cantidad mensual producida de caña de azúcar en el Perú, enero 1983 – abril 2021, muestra 460 observaciones las cuales la cantidad mínima es 194 MTM y la cantidad máxima es 1101,88 MTM. La media hallada fue 650,8 MTM, con una desviación típica de  $\pm 198,5$  MTM, un coeficiente de variación de 30.5% donde alude a datos heterogéneos, una asimetría positiva de 0,272 MTM y una curtosis negativa de -0,560 MTM siendo la distribución platicúrtica.
3. La cantidad mensual producida de caña de azúcar en el Perú, enero 1983 - abril 2021 mostró una ligera tendencia decreciente en los primeros años de estudio y una tendencia creciente en los últimos años de estudio, además no fue estacionaria y presentó estacionalidad cada 12 meses. Sin embargo, al realizar la primera diferencia la serie fue estacionaria aplicando de la prueba aumentada de DF.
4. Los modelos que resultaron adecuados fueron: SARIMA (0, 1, 2) (1, 1, 1), SARIMA (0, 1, 3) (1, 1, 1), SARIMA (1, 1, 1) (1, 1, 1) y el SARIMA (2, 1, 2) (1, 1, 1).
5. Los valores del pronóstico del modelo SARIMA (1, 1, 1) (1, 1, 1) son cercanos a los valores reales de la serie, dado que el modelo es parsimonioso, tiene cuatro parámetros por estimar y tiene un menor el BCI.

## **Recomendaciones**

Se recomienda al jefe del INEI, hacer uso del modelo de pronóstico de la cantidad mensual producida de caña de azúcar en el Perú que se logró encontrar en la investigación, siendo el SARIMA (1, 1, 1) (1, 1, 1).

Actualizar y evaluar la base de datos para identificar los posibles cambios del modelo de pronóstico encontrado.

Esta investigación encontró resultados óptimos de pronósticos y destaca la importancia de llevar a cabo futuros estudios de modelos predictivos con la finalidad de observar y comprender a largo plazo la tendencia de las variables en la economía peruana, incluyendo aspectos como los costos, inversiones, exportaciones, entre otros.

## Bibliografía

- Anchuelo. (1993). Series integradas y cointegradas: Una introducción. *Revista de Economía Aplicada*, *I*(1), 151-164.  
<https://www.revecap.alde.es/revista/numeros/01/pdf/anchuelo.pdf>
- Arce, R. y Mahia, R. (2001). *Modelos ARIMA*. Universidad de Investigación y Desarrollo.  
<https://docplayer.es/15112140-Modelos-arima-prof-rafael-de-arce-prof-ramon-mahia-dpto-economia-aplicada-u-d-i-econometria-e-informatica.html>.
- Arévalo S. y Santisteban C. (2020). *Modelo de pronóstico de la cantidad de Cantidad de Producción de Electricidad con la Metodología de Box Jenkins de la Región Lambayeque en el periodo 2005–2017*. (Tesis de pregrado). Universidad Nacional Pedro Ruiz Gallo, Lambayeque, Perú.  
<https://repositorio.unprg.edu.pe/handle/20.500.12893/9022>
- De la Fuente (2013). *Series temporales. Modelos ARIMA*. Universidad Autónoma de Madrid, España. <https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>
- Duarte, O., & Gonzalez, J. (2019). *Guía técnica cultivo de caña de azúcar*.  
[https://www.jica.go.jp/Resource/paraguay/espanol/office/others/c8h0vm0000ad5gk-e-att/gt\\_01.pdf](https://www.jica.go.jp/Resource/paraguay/espanol/office/others/c8h0vm0000ad5gk-e-att/gt_01.pdf)
- Fuentes, R. (2017). *Modelo de estimación de la producción de caña de azúcar a través de imágenes de satélite en los ingenios La Unión y Santa Ana* (Tesis de pregrado). Universidad Rafael Landívar, Escuintla, Guatemala.  
<http://biblio3.url.edu.gt/publijrcifuentes/TESIS/2018/06/17/Fuentes-Rodolfo.pdf>
- González, M. (2009). *Análisis de Series Temporales: Modelos ARIMA*. Vasco: Sarriko ON.  
<https://addi.ehu.es/handle/10810/12492>
- Gonzalez, L. (2018). *Evaluando el error en los modelos de regresión*.  
<https://aprendeia.com/evaluando-el-error-en-los-modelos-de-regresion/>
- Hernández, R. (2014). *Metodología de la investigación*. McGRAW-HILL/Interamericana.  
[https://periodicooficial.jalisco.gob.mx/sites/periodicooficial.jalisco.gob.mx/files/metodologia\\_de\\_la\\_investigacion\\_-\\_roberto\\_hernandez\\_sampieri.pdf](https://periodicooficial.jalisco.gob.mx/sites/periodicooficial.jalisco.gob.mx/files/metodologia_de_la_investigacion_-_roberto_hernandez_sampieri.pdf)

- Instituto Nacional de Estadística e Informática (2021, 24 de agosto). Producción de caña de azúcar se redujo 20,1%. <https://www.inei.gob.pe/prensa/noticias/produccion-de-quinua-crecio-752-durante-junio-de-2021-13075/#:~:text=Producci%C3%B3n%20de%20ca%C3%BAa%20de%20az%C3%BAcar,por%20las%20menores%20cosechas%20realizadas>.
- Laurente L. y Laurente F. (2019). Aplicación del modelo ARIMA para la producción de la papa en la región de Puno-Perú. *Revista de Investigación e Innovación Agropecuaria y de Recursos Naturales*. (6). [http://www.scielo.org.bo/scielo.php?pid=S2409-16182019000100006&script=sci\\_abstract](http://www.scielo.org.bo/scielo.php?pid=S2409-16182019000100006&script=sci_abstract)
- Mauricio, JA. (2007). *Introducción al Análisis de Series de Tiempo*. Universidad Complutense de Madrid, España. <https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf>
- Ministerio de Desarrollo Agrario y Riego. (2020). *Observatorio de Commodities: Azúcar octubre- diciembre 2020*. [Boletín, Ministerio de Desarrollo Agrario y Riego]. <https://repositorio.midagri.gob.pe/bitstream/20.500.13036/839/1/Commodities%20Az%C3%BAcar%20oct-dic%202020.pdf>
- Ministerio de Desarrollo Agrario y Riego. (s.f.). *Caracterización de la Cadena*. <https://www.midagri.gob.pe/portal/download/pdf/especiales/bioenergia/otros/caracterizacion-delacadena-productiva-d-canad-azucarenelperu.pdf>
- Mora, M. (20 de Noviembre de 2023). Tiempedia: Evaluación y Métricas de Rendimiento en Pronósticos Temporales. <https://www.linkedin.com/pulse/tiempedia-evaluaci%C3%B3n-y-m%C3%A9tricas-de-rendimiento-en-mora-caballero-u5wpe/?originalSubdomain=es>
- Novales, A. (1993) *Econometría* (2da ed.). McGraw-Hill. <https://econometriai.files.wordpress.com/2011/04/econometrc2a1a-2c2a6-ed-2000-alfonso-novales-mcgraw-hill.pdf>
- Osorio, J., & Ángel, V. (2019). Aplicación del modelo Box-Jenkins en la estimación del flujo de llamadas en servicios de atención al cliente en el sector contact center. [Tesis de posgrados, Fundación Universitaria Los Libertadores]. [https://repository.libertadores.edu.co/bitstream/handle/11371/2798/Ochoa\\_%20V%C3%ADctor\\_Hern%C3%A1ndez\\_%20Jony\\_2019.pdf?sequence=1&isAllowed=y](https://repository.libertadores.edu.co/bitstream/handle/11371/2798/Ochoa_%20V%C3%ADctor_Hern%C3%A1ndez_%20Jony_2019.pdf?sequence=1&isAllowed=y)

- Pérez, M. (2017). *Modelo de estimación de la producción de caña de azúcar a través de imágenes de satélite en Ingenio Madre Tierra*. (Tesis de pregrado). Universidad Rafael Landívar, Escuintla, Guatemala.  
<http://biblio3.url.edu.gt/publijrcifuentes/TESIS/2018/06/17/Perez-Misael.pdf>
- Valverde L. (2021). *Modelo pronóstico para la producción de caña de azúcar en la Empresa Agroindustrial Laredo S.A.A periodo enero 2012 – noviembre 2019*. (Tesis de pregrado). Universidad Nacional de Trujillo, Trujillo, Perú.  
<https://dspace.unitru.edu.pe/handle/UNITRU/16484>
- Pollack, M., Helfgott, S., & Tejada, J. (2018). El cultivo de caña de azúcar en la Costa del Perú durante los eventos de El Niño 1982-83 y 1997-98. *Ecología Aplicada*, 17(1), 77-84. <http://dx.doi.org/10.21704/rea.v17i1.1176>

## Anexos

Anexo 1: Registro mensual de la cantidad mensual producida de caña de azúcar, enero 1983 – abril 2021. Perú.

	<b>ENE</b>	<b>FEB</b>	<b>MAR</b>	<b>ABR</b>	<b>MAY</b>	<b>JUN</b>	<b>JUL</b>	<b>AGO</b>	<b>SET</b>	<b>OCT</b>	<b>NOV</b>	<b>DIC</b>
<b>1983</b>	582	542	346	238	377	581	550	706	656	602	585	615
<b>1984</b>	590	551	475	402	495	644	649	700	591	563	642	686
<b>1985</b>	683	585	581	451	619	616	650	684	641	585	640	597
<b>1986</b>	591	459	344	352	546	524	540	565	580	586	605	585
<b>1987</b>	523	450	382	290	434	544	581	595	578	575	567	556
<b>1988</b>	508	502	450	296	371	544	536	572	559	569	508	536
<b>1989</b>	528	468	507	414	355	518	530	633	583	615	621	564
<b>1990</b>	586	519	539	404	469	397	503	484	519	526	518	482
<b>1991</b>	501	458	469	354	345	473	582	606	518	486	493	508
<b>1992</b>	448	374	304	207	283	376	449	505	474	468	424	431
<b>1993</b>	397	324	258	194	249	260	364	413	462	471	477	475
<b>1994</b>	438	407	340	205	356	433	489	598	524	574	551	516
<b>1995</b>	487	454	486	403	350	461	584	670	591	620	611	609
<b>1996</b>	552	515	483	316	353	377	508	607	593	625	611	581
<b>1997</b>	565	546	534	489	542	603	589	615	649	635	589	574
<b>1998</b>	499	310	252	250	397	478	584	652	593	565	602	522
<b>1999</b>	489	445	480	424	504	485	515	529	590	639	596	582
<b>2000</b>	593	587	523	353	507	622	575	660	641	689	711	674
<b>2001</b>	545	573	489	466	433	556	650	770	730	761	697	717
<b>2002</b>	698	590	616	546	560	721	763	846	788	794	738	760
<b>2003</b>	693	613	737	620	771	770	747	789	736	826	774	788
<b>2004</b>	636	498	608	480	627	578	573	558	580	592	636	580
<b>2005</b>	472	468	428	398	414	467	592	569	587	615	650	643
<b>2006</b>	521	603	496	294	580	494	641	700	658	774	735	749
<b>2007</b>	654	577	530	330	583	764	772	831	766	853	761	809
<b>2008</b>	749	602	622	643	656	790	831	892	894	944	878	894
<b>2009</b>	828	836	849	708	647	669	879	887	829	945	944	917
<b>2010</b>	815	642	777	730	679	815	844	944	892	860	934	925
<b>2011</b>	798	648	877	716	738	671	775	954	848	925	952	982
<b>2012</b>	823	735	809	691	685	673	834	1026	1043	1019	1069	963
<b>2013</b>	888	856	782	909	925	811	714	956	1059	1035	1022	1036
<b>2014</b>	938	927	927	945	945	763	722	1033	1058	1067	1034	1029
<b>2015</b>	848	836	807	690	599	778	739	987	951	997	1003	975
<b>2016</b>	924	726	542	541	573	822	994	957	905	964	949	937
<b>2017</b>	863	612	288	454	684	732	909	977	1005	1012	1003	861
<b>2018</b>	815	660	600	511	644	892	983	1013	1029	1030	1102	1058
<b>2019</b>	868	790	880	712	763	764	1074	1050	1034	1050	1007	910
<b>2020</b>	839	753	637	719	915	927	834	853	946	1064	958	1023
<b>2021</b>	772	887	693	564								

Anexo 2: Modelos considerados en el análisis que cumplieron y no cumplieron con los requisitos.

<b>Modelo</b>	<b>Parámetros</b>	<b>Normalidad</b>	<b>Ruido blanco</b>
SARIMA 1 (0, 1, 1) (1, 0, 1)	Significativo	Si cumple	No cumple
SARIMA 2 (0, 1, 1) (1, 1, 1)	Significativo	Si cumple	No cumple
SARIMA 3 (0, 1, 2) (1, 0, 1)	Significativo	Si cumple	No cumple
SARIMA 4 (0, 1, 2) (1, 1, 1)	Significativo	Si cumple	Si cumple
SARIMA 5 (0, 1, 3) (1, 0, 1)	Significativo	No cumple	No cumple
SARIMA 6 (0, 1, 3) (1, 1, 1)	Significativo	Si cumple	Si cumple
SARIMA 7 (1, 1, 1) (1, 0, 1)	Significativo	Si cumple	No cumple
SARIMA 8 (1, 1, 1) (1, 1, 1)	Significativo	Si cumple	Si cumple
SARIMA 9 (1, 1, 2) (1, 0, 1)	No todos los coeficientes son significativos	Si cumple	No cumple
SARIMA 10 (1, 1, 2) (1, 1, 1)	No todos los coeficientes son significativos	Si cumple	Si cumple
SARIMA 11 (1, 1, 3) (1, 0, 1)	No todos los coeficientes son significativos	Si cumple	No cumple
SARIMA 12 (1, 1, 3) (1, 1, 1)	No todos los coeficientes son significativos	Si cumple	Si cumple
SARIMA 13 (2, 1, 1) (1, 0, 1)	No todos los coeficientes son significativos	Si cumple	No cumple
SARIMA 14 (2, 1, 1) (1, 1, 1)	No todos los coeficientes son significativos	Si cumple	Si cumple
SARIMA 15 (2, 1, 2) (1, 0, 1)	Significativo	Si cumple	No cumple
SARIMA 16 (2, 1, 2) (1, 1, 1)	Significativo	Si cumple	Si cumple
SARIMA 17 (2, 1, 3) (1, 0, 1)	No todos los coeficientes son significativos	Si cumple	No cumple
SARIMA 18 (2, 1, 3) (1, 1, 1)	No todos los coeficientes son significativos	Si cumple	Si cumple
SARIMA 19 (3, 1, 1) (1, 0, 1)	No todos los coeficientes son significativos	Si cumple	No cumple
SARIMA 20 (3, 1, 1) (1, 1, 1)	No todos los coeficientes son significativos	Si cumple	Si cumple

SARIMA 21 (3, 1, 2) (1, 0, 1)	No todos los coeficientes son significativos	Si cumple	No cumple
SARIMA 22 (3, 1, 2) (1, 1, 1)	No todos los coeficientes son significativos	Si cumple	Si cumple
SARIMA 23 (3, 1, 3) (1, 0, 1)	No todos los coeficientes son significativos	Si cumple	No cumple
SARIMA 24 (3, 1, 3) (1, 1, 1)	No todos los coeficientes son significativos	Si cumple	Si cumple

Anexo 3: Cuadro de operacionalización de variables.

<b>Variab</b> les	<b>Indicador</b>	<b>Escala de medida</b>	<b>Instrumento de recolección de datos</b>
<b>Variable Dependiente:</b> Cantidad mensual producida de caña de azúcar en el Perú	Miles de toneladas métricas. (MTM)	De razón	
			Ficha electrónica
<b>Variable Independiente:</b> Tiempo Enero 1983-Abril 2021	Mes-Año	Ordinal	