



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría en Artes en Estadística Aplicada

**DISEÑO DE UN MODELO ESTADÍSTICO PARA PRONOSTICAR LA PRODUCCIÓN DE
PALMA DE ACEITE MEDIANTE ANÁLISIS CORRELACIONALES Y SERIES DE TIEMPO DE
LA PRODUCCIÓN DURANTE EL PERIODO 2010 AL 2020**

Ing. Ronald Ricardo Ordoñez Marroquín

Asesorado por Mtro. Luis Carlos Leonardo Bolaños Méndez

Guatemala, abril de 2024

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE UN MODELO ESTADÍSTICO PARA PRONOSTICAR LA PRODUCCIÓN DE
PALMA DE ACEITE MEDIANTE ANÁLISIS CORRELACIONALES Y SERIES DE TIEMPO DE
LA PRODUCCIÓN DURANTE EL PERIODO 2010 AL 2020**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ING. RONALD RICARDO ORDOÑEZ MARROQUÍN
ASESORADO POR EL MTRO. LUIS CARLOS LEONARDO BOLAÑOS MÉNDEZ

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, ABRIL DE 2024

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. José Francisco Gómez Rivera (a. i.)
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton De León Bran
VOCAL IV	Ing. Kevin Vladimir Armando Cruz Lorente
VOCAL V	Ing. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

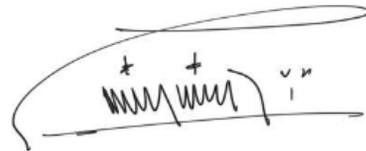
DECANO	Ing. José Francisco Gómez Rivera (a. i.)
EXAMINADORA	Mtra. Inga. Aurelia Anabella Cordova Estrada
EXAMINADOR	Mtro. Ing. Edwin Adalberto Bracamonte Orozco
EXAMINADORA	Dra. Aura Marina Rodriguez Pérez
SECRETARIO	Mtro. Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

DISEÑO DE UN MODELO ESTADÍSTICO PARA PRONOSTICAR LA PRODUCCIÓN DE PALMA DE ACEITE MEDIANTE ANÁLISIS CORRELACIONALES Y SERIES DE TIEMPO DE LA PRODUCCIÓN DURANTE EL PERIODO 2010 AL 2020

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 10 de noviembre de 2022.

A handwritten signature in black ink, consisting of a series of loops and flourishes, enclosed within a large, sweeping, hand-drawn oval shape.

Ing. Ronald Ricardo Ordoñez Marroquín

LNG.DECANATO.OI.154.2024

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Posgrado, al Trabajo de Graduación titulado: **DISEÑO DE UN MODELO ESTADÍSTICO PARA PRONOSTICAR LA PRODUCCIÓN DE PALMA DE ACEITE MEDIANTE ANÁLISIS CORRELACIONALES Y SERIES DE TIEMPO DE LA PRODUCCIÓN DURANTE EL PERIODO 2010 AL 2020**, presentado por: **Ing. Ronald Ricardo Ordoñez Marroquín**, que pertenece al programa de Maestría en artes en Estadística aplicada después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Ing. José Francisco Gómez Rivera

Decano a.i.

Guatemala, abril de 2024

JFGR/gaoc



Guatemala, abril de 2024

LNG.EEP.OI.154.2024

En mi calidad de Directora de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

“DISEÑO DE UN MODELO ESTADÍSTICO PARA PRONOSTICAR LA PRODUCCIÓN DE PALMA DE ACEITE MEDIANTE ANÁLISIS CORRELACIONALES Y SERIES DE TIEMPO DE LA PRODUCCIÓN DURANTE EL PERIODO 2010 AL 2020”

presentado por **Ing. Ronald Ricardo Ordoñez Marroquín** correspondiente al programa de **Maestría en artes en Estadística aplicada** ; apruebo y autorizo el mismo.

Atentamente,

“Id y Enseñad a Todos”


Mtra. Inga. Aurelia Anabela Cordova Estrada
Directora
Escuela de Estudios de Postgrado
Facultad de Ingeniería





Guatemala, 26 de octubre de 2023

M.A. Inga. Aurelia Anabela Cordova Estrada
Directora
Escuela de Estudios de Postgrado
Presente

Estimada M.A. Inga. Cordova Estrada

Por este medio informo a usted, que he revisado y aprobado el **INFORME FINAL y ARTÍCULO CIENTÍFICO** titulado: **DISEÑO DE UN MODELO ESTADISTICO PARA PRONOSTICAR LA PRODUCCIÓN DE PALMA DE ACEITE MEDIANTE ANALISIS CORRELACIONALES Y SERIES DE TIEMPO DE LA PRODUCCIÓN DURANTE EL PERIODO 2010 AL 2020.** del estudiante **Ronald Ricardo Ordoñez Marroquín** quien se identifica con número de carné **202290265** del programa de Maestria En Estadística Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el **Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014.** Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.



Msc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador
Maestria En Estadística Aplicada
Escuela de Estudios de Postgrado

Oficina Virtual



Guatemala, 26 de octubre de 2023

M.A. Inga. Aurelia Anabela Cordova Estrada
Directora
Escuela de Estudios de Postgrados
Presente

Estimada M.A. Inga. Cordova Estrada

Por este medio informo a usted, que he revisado y aprobado el Trabajo de Graduación y el Artículo Científico: **"DISEÑO DE UN MODELO ESTADISTICO PARA PRONOSTICAR LA PRODUCCIÓN DE PALMA DE ACEITE MEDIANTE ANALISIS CORRELACIONALES Y SERIES DE TIEMPO DE LA PRODUCCIÓN DURANTE EL PERIODO 2010 AL 2020."** de el/la estudiante **Ronald Ricardo Ordoñez Marroquín** del programa de **Maestria En Estadística Aplicada** identificado(a) con número de carné 202290265.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.



Mtro. Ing. Luis Carlos Leonardo Bolaños Méndez

Colegiado No. 7653

Asesor de Tesis

Luis Carlos Leonardo Bolaños Méndez
Ingeniero Electrónico
colegiado 7653

ACTO QUE DEDICO A:

- Dios** Por ser nuestro creador y ser supremo.
- Mis padres** Mayra Marroquín y Ricardo Ordoñez, quienes por su amor y sacrificio infinito, fueron el inicio y la base para llegar al lugar donde estoy, quienes son mi mejor ejemplo para seguir.
- Mi esposa** Eunice Cerón, quien se convirtió en pieza fundamental en este proyecto, quien me ha tenido tanta paciencia y me ha brindado todo el apoyo y su amor incondicional, quien ha sido mi soporte en los momentos más difíciles.
- Mis hijos** Ricardo y Mayeli Ordoñez, quienes se han vuelto mi todo, lo más lindo que tengo en mi vida, mi inspiración y mi motivación.
- Mis hermanos** Alejandro Ordoñez, Hugo Morales, Herminio y Daniel Marroquín y Andrea Castañeda, por su apoyo incondicional en todo momento.
- Mis tíos** Blanca, Hugo, Zoila y en especial a Zuly Marroquín, quien ha sido mi segunda madre.
- A mis abuelos** En especial a Mirtala Natareno (q. e. p. d.).

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ayudar a mi crecimiento profesional y permitir ser parte de su selecto grupo de profesionales.
Escuela de Estudios de Postgrados de la Facultad de Ingeniería	Por aceptarme y brindarme todos los conocimientos adquiridos durante esta etapa de preparación en mi vida.
Asesor	Mtro. Luis Carlos Leonardo Bolaños Méndez, quien con sus conocimientos y experiencia aportaron al logro de este proyecto.
Maestros	Por la enseñanza en cada uno los cursos de la carrera, en especial a la Doctora Aura Marina Rodríguez Pérez, quien con sus conocimientos y tutorías fue muy importante en la elaboración de este trabajo de tesis.
Mis amigos	Quienes siempre me brindaron su apoyo incondicional.

	2.1.3.1.2.	Métodos de proyección histórica.....	12
	2.1.3.1.3.	Métodos causales	13
2.1.3.2.		Análisis correlacionales	13
	2.1.3.2.1.	Coeficiente de correlación.....	13
	2.1.3.2.2.	Regresión lineal simple .	14
	2.1.3.2.3.	Regresión lineal múltiple.....	15
	2.1.3.2.4.	Coeficiente de determinación R cuadrado	15
	2.1.3.2.5.	Coeficiente de determinación ajustado R^2_a	16
	2.1.3.2.6.	Identificación de datos atípicos.....	17
2.1.4.		Análisis de serie de tiempo.....	18
	2.1.4.1.	Metodología Box-Jenkins (ARIMA)	21
	2.1.4.1.1.	Modelos autorregresivos	23
	2.1.4.1.2.	Modelos de promedio móvil.....	23
	2.1.4.1.3.	Modelos ARIMA no estacionales	24
	2.1.4.1.4.	Modelos ARIMA estacionales	25
2.1.5.		Prueba de normalidad	26
	2.1.5.1.	Shapiro-Wilk	26

2.1.6.	Pruebas de diagnóstico de los modelos	27
2.1.6.1.	Criterio de información de Akaike (AIC)	27
2.1.6.2.	Criterio de información Bayesiano (BIC)	28
2.1.6.3.	Factor de inflación de la varianza (VIF)	28
2.1.6.4.	Cp de Mallows	29
2.1.6.5.	Prueba de Breusch Pagan.....	30
2.1.6.6.	Prueba de Durbin-Watson	30
2.1.6.7.	Prueba de Ljung-Box.....	30
2.1.6.8.	Prueba de Dickey Fuller aumentado....	31
2.1.6.9.	Función de autocorrelación (ACF)	31
2.1.6.10.	Función de autocorrelación parcial (PACF).....	31
2.2.	La palma de aceite	32
2.2.1.	La importancia del cultivo	32
2.2.1.1.	A nivel mundial	33
2.2.1.2.	En Guatemala.....	35
2.2.1.3.	En el departamento de Escuintla	35
2.2.2.	Aspectos botánicos.....	35
2.2.2.1.	El sistema radicular	36
2.2.2.2.	El tallo.....	37
2.2.2.3.	Hojas	38
2.2.2.4.	Inflorescencias.....	39
	2.2.2.4.1. Flor femenina.....	39
	2.2.2.4.2. Flor masculina	40
2.2.2.5.	Frutos	40
2.2.2.6.	Semilla.....	40

2.2.3.	Factores climáticos que influyen en la formación de frutos	40
3.	PRESENTACIÓN DE RESULTADOS.....	43
3.1.	Objetivo 1. Estimar el grado de asociación y dependencia estadística entre la variable producción con respecto a las variables climáticas entre los años 2010 y 2020 a través del coeficiente de correlación	43
3.1.1.	Análisis exploratorio	43
3.1.1.1.	Análisis descriptivo	43
3.1.1.2.	Pruebas de normalidad	45
3.1.1.3.	Análisis correlacional.....	46
3.2.	Objetivo 2. Construir un modelo estadístico de predicción de producción utilizando las variables ambientales mediante regresión lineal múltiple	48
3.2.1.	Modelo 1.....	48
3.2.2.	Modelo 2.....	50
3.2.3.	Modelo 3.....	53
3.2.4.	Comparación de modelos de regresión múltiple	55
3.3.	Objetivo 3. Construir un modelo estadístico que se ajuste aceptablemente al comportamiento de la producción de estudio mediante un análisis de series temporales para obtener predicciones aceptables	56
3.3.1.	Análisis exploratorio de datos.....	57
3.3.2.	Selección del modelo ARIMA	59
3.3.2.1.	Diagnóstico del modelo 1, ARIMA(2,0,0)(0,1,1)[12]	62
3.3.2.1.1.	Análisis de residuos	62

3.3.2.2.	Diagnóstico del modelo 2, ARIMA(3,0,0)(0,1,1)[12].....	65
3.3.2.2.1.	Análisis de residuos.....	66
3.3.2.3.	Diagnóstico del modelo 3, ARIMA(4,0,0)(0,1,1)[12].....	68
3.3.2.3.1.	Análisis de residuos.....	69
3.3.2.4.	Diagnóstico del modelo 4, ARIMA(4,0,0)(1,1,1)[12].....	71
3.3.2.4.1.	Análisis de residuos.....	72
3.3.2.5.	Selección del mejor modelo.....	74
3.3.3.	Pronóstico del modelo	75
3.4.	Objetivo general. Diseñar un modelo estadístico para pronosticar la producción de palma de aceite mediante análisis correlacionales y series de tiempo de la producción durante el periodo 2010 al 2020.....	76
4.	DISCUSIÓN DE LOS RESULTADOS	79
4.1.	Análisis interno del estudio	79
4.2.	Análisis externo del estudio	82
	CONCLUSIONES	87
	RECOMENDACIONES.....	89
	REFERENCIAS	91
	APÉNDICE.....	97

ÍNDICE DE ILUSTRACIONES

FIGURAS

Figura 1.	Flujograma del esquema de solución	XXXI
Figura 2.	Serie temporal de la producción mensual kg/ha, periodo 2010-2020.....	19
Figura 3.	Serie temporal de la producción mensual kg/ha, periodo 2010-2020.....	20
Figura 4.	Serie temporal de la producción mensual kg/ha, periodo 2010-2020.....	21
Figura 5.	Diagrama de flujo del método de Box-Jenkins.....	22
Figura 6.	Principales países productores de palma de aceite, año 2021	34
Figura 7.	Sistema radicular de palma de aceite	37
Figura 8.	Matriz de correlación de las variables en estudio	46
Figura 9.	Serie temporal de la producción mensual kg/ha, periodo 2010-2020.....	57
Figura 10.	Descomposición de la serie temporal de producción, periodo del 2010 al 2020	58
Figura 11.	Análisis de residuos de la serie de tiempo original	59
Figura 12.	Gráfico de autocorrelación y autocorrelación parcial de la parte estacional.....	60
Figura 13.	Gráfico de autocorrelación y autocorrelación parcial de la parte no estacional.....	61
Figura 14.	Análisis de residuos del modelo ARIMA(2,0,0)(0,1,1)[12].....	63
Figura 15.	Gráfico de Q-Q de los residuos del modelo ARIMA(2,0,0)(0,1,1)[12].....	64

Figura 16.	Gráfico de valores-p para la prueba Ljung-Box	65
Figura 17.	Análisis de residuos del modelo ARIMA(3,0,0)(0,1,1)[12]	66
Figura 18.	Gráfico de Q-Q de los residuos del modelo ARIMA(3,0,0)(0,1,1)[12]	67
Figura 19.	Gráfico de valores-p para la prueba Ljung-Box	68
Figura 20.	Análisis de residuos del modelo ARIMA(4,0,0)(0,1,1)[12]	69
Figura 21.	Gráfico de Q-Q de los residuos del modelo ARIMA(4,0,0)(0,1,1)[12]	70
Figura 22.	Gráfico de valores-p para la prueba Ljung-Box	71
Figura 23.	Análisis de residuos del modelo ARIMA(4,0,0)(1,1,1)[12]	72
Figura 24.	Gráfico de Q-Q de los residuos del modelo ARIMA(4,0,0)(1,1,1)[12]	73
Figura 25.	Gráfico de valores-p para la prueba Ljung-Box	74
Figura 26.	Pronóstico de 24 meses con el modelo ARIMA(4,0,0)(0,1,1)[12]	76

TABLAS

Tabla 1.	Operativización de las variables	XXX
Tabla 2.	Análisis descriptivo de las variables en estudio	44
Tabla 3.	Resultados de pruebas de normalidad de las variables en estudio	45
Tabla 4.	Análisis de regresión múltiple del modelo 1	48
Tabla 5.	Coeficientes de regresión del modelo 1	49
Tabla 6.	Análisis de regresión múltiple del modelo 2	51
Tabla 7.	Coeficientes de regresión del modelo 2	51
Tabla 8.	Análisis de regresión múltiple del modelo 3	53
Tabla 9.	Coeficientes de regresión del modelo 3	54
Tabla 10.	Comparación de modelos de regresión lineal múltiple	56

Tabla 11. Tabla resumen de criterios y características de los 4 modelos
evaluados 75

LISTA DE SÍMBOLOS

Símbolo	Significado
n-1	Grados de libertad
ha	Hectárea
h	Hora
kg	Kilogramo
n	Muestra
N	Población
%	Porcentaje
Pmes	Precipitación mensual
RFFm	Racimos de fruta fresca mensual
Rmed	Radiación mensual
Tmax	Temperatura máxima
Tmed	Temperatura media
Tmin	Temperatura mínima

GLOSARIO

Abiótico	Son los elementos del ecosistema que no tienen vida, pero se necesitan para vivir, como lo es el agua, el oxígeno, temperatura, y la luz solar.
Agroindustria	Explotación agraria organizada como una industria.
Agronomía	Es el conjunto de conocimientos de diversas ciencias aplicadas que rigen la práctica de la agricultura.
Antesis	Es el tiempo de expansión de una flor hasta que está completamente desarrollada y en estado funcional, durante el cual ocurre el proceso de polinización.
Ápice	Es el extremo superior o punta de la hoja, del fruto, entre otros.
Base de datos	Una base de datos es una recopilación organizada de información o datos estructurados, que normalmente se almacena de forma electrónica en un sistema informático.
Biomasa	Cantidad de productos obtenidos por fotosíntesis, susceptibles de ser transformados en combustible útil para el hombre y expresada en unidades de superficie y de volumen.

Biótico	De los organismos vivos o relacionado con ellos.
Coefficiente de Pearson	Es una prueba que mide la relación estadística entre dos variables continuas.
Diagrama de flujo	Es un gráfico que representa la esquematización gráfica de un algoritmo, el cual muestra gráficamente los pasos o procesos a seguir para alcanzar la solución de un problema.
Empírico	Que está basado en la experiencia y en la observación de los hechos.
Endocarpio	Es la capa interior del pericarpio; es decir, la parte del fruto que rodea las semillas, conocido también como carozo o hueso.
Espiguilla	Es la inflorescencia elemental de las gramíneas o poáceas. Presenta en la base dos glumas o brácteas, luego siguen los antecios dispuestos dísticamente.
Estipe	El tronco de un árbol sin ramificación lateral, como en las arecáceas.
Exocarpio	Es la capa más externa, de función protectora, a modo de epidermis, pero que puede estar muy

transformada y adaptada para la dehiscencia o la dispersión del fruto.

Foliolo

Hoja o parte de una hoja compuesta.

Fotosíntesis

Proceso químico que tiene lugar en las plantas con clorofila y que permite, gracias a la energía de la luz, transformar un sustrato inorgánico en materia orgánica rica en energía.

Inferir

Extraer un juicio o conclusión a partir de hechos, proposiciones o principios, sean generales o particulares.

Mesocarpio

En la mayoría de las frutas, es una capa gruesa y succulenta; corresponde a la “pulpa” en el aguacate, y mango; en frutas como la naranja, es lo “blanco” que rodea los gajos.

Modelo estadístico

El Modelo o Modelado Estadístico es una herramienta basada en las matemáticas y el análisis de datos que busca representar a la realidad mediante métodos estadísticos sobre los datos disponibles.

Monoica

En botánica, se denomina monoicas a las especies en las cuales ambos sexos se presentan en una misma planta.

Pecíolo	Estructura que une a la lámina de la hoja con la rama o tallo.
Presupuesto	Conjunto de los gastos e ingresos previstos para un determinado período de tiempo.
Pronóstico	Predicción de la evolución de un proceso o de un hecho futuro a partir de criterios lógicos o científicos.
Red neuronal (informática)	Una red neuronal es un método de la inteligencia artificial que enseña a las computadoras a procesar datos de una manera que está inspirada en la forma en que lo hace el cerebro humano.
Sistematizar	Organizar un conjunto de elementos de manera que formen un sistema.
Técnica <i>Mann-Kendall</i>	Es una prueba no paramétrica sugerido para evaluar la tendencia en series de datos ambientales.

RESUMEN

El propósito de esta investigación fue diseñar un modelo estadístico a través del análisis de regresión lineal múltiple y análisis de series de tiempo que permita pronosticar la producción de fruta de palma de aceite de acuerdo con el comportamiento a través del tiempo de la producción y variables climáticas durante el periodo del año 2010 al 2020.

El estudio propuesto pretende explicar la problemática desde un enfoque cuantitativo.

El diseño a utilizado es no experimental con un alcance descriptivo y correlacional, que permitió conocer el nivel de influencia entre las variables:

- Producción
- Temperatura máxima mensual
- Temperatura media mensual
- Temperatura mínima mensual
- Precipitación mensual
- Radiación solar
- Humedad relativa promedio mensual

La recolección de información se hizo a través de base de datos históricas de la finca, las mismas fueron analizadas por medio de las herramientas Infostat, Rstudio y Microsoft Excel.

Para cumplir con los objetivos propuestos en la investigación, se utilizó:

- Análisis de correlación de Pearson
- Medidas de tendencia central
- Medidas de dispersión
- Análisis de regresión múltiple
- Análisis de series de tiempo

Comprobando cada uno de los objetivos, realizando pruebas estadísticas y validación de modelos.

Como resultados, se identificó:

- Qué el grado de correlación de la producción con las otras variables es bajo.
- Que existe una relación lineal significativa entre la producción y la temperatura mínima mensual.
- Precipitación mensual y humedad relativa promedio mensual.
- El modelo solo explica un 20 % de la variación.

De acuerdo con el análisis de series temporales, la producción no tiene tendencia, pero si tiene estacionalidad de 12 meses, y el mejor modelo de pronóstico fue un $ARIMA(4,0,0)(0,1,1)[12]$.

Se hace necesario continuar con estudios utilizando otro tipo de métodos estadísticos, agregando otras variables que no se consideraron para esta

investigación y que permitan explicar el comportamiento de la producción para poder construir un modelo más ajustado.

PLANTEAMIENTO DEL PROBLEMA

La agroindustria de palma de aceite en Guatemala está establecida en tres zonas del país. Este cultivo es perenne y su producción es permanente, puede iniciar entre los 2 y 4 años, y se puede extender hasta los 35 años del cultivo.

- Contexto general

Dentro de las actividades que conlleva el cultivo de palma de aceite se encuentra la elaboración del presupuesto anual de ingresos y gastos para el departamento de Producción Agrícola e Industrial, para lo cual la base es la producción de fruta de palma. Del pronóstico de esta depende el ejercicio financiero para el periodo anual próximo, por lo que es sumamente importante tener un ajuste aceptable.

Con el uso de métodos empíricos puede existir una variación del pronóstico de producción con respecto a la producción real fuera de los parámetros establecidos, esto debido a las variaciones causadas por la falta de conocimiento de la interacción entre esta y las variables ambientales: humedad relativa, precipitación, temperatura y radiación solar. Lo cual genera varias deficiencias operativas y financieras para las empresas.

- Descripción del problema

Derivado de la problemática que se genera por los pronósticos inadecuados, se hace importante evaluar el comportamiento de la producción de palma de aceite en función del tiempo y de factores ambientales que influyen

directamente en esta. Uno de los principales problemas es la falta de modelos estadísticos aplicados a la estimación de la producción de palma, esto implica la utilización del conocimiento empírico para poder realizarlas. Esta falta de conocimiento estadístico no permite conocer los efectos de la interacción que existe entre la producción y las variables ambientales: precipitación, radiación solar, temperatura y humedad relativa.

- Formulación del problema

El problema se plantea mediante la formulación de preguntas, las cuales se dividen en pregunta central: la cual tiene como función contextualizar todo el problema para buscar un objetivo general de solución. Y las preguntas auxiliares: las cuales son parte de la pregunta central y tienen como función el apoyo a la resolución de ésta, cumpliendo objetivos específicos.

- Pregunta central

¿Cuáles son los factores ambientales que se relacionan con la producción de fruta de palma y qué efecto directo generan en el comportamiento de la producción en el tiempo?

- Preguntas auxiliares

- ¿Cuál es el efecto individual que generan directamente los factores ambientales sobre el comportamiento en el tiempo de la producción de fruta de palma?
- ¿Qué efecto genera la interacción de los factores ambientales sobre el comportamiento en el tiempo de la producción de fruta de palma?

- ¿Qué modelo de serie de tiempo se ajusta de manera confiable para pronosticar la producción de palma de aceite?
- Delimitación del problema

El objeto de estudio es el proceso de producción de la fruta de palma de aceite monitoreando el periodo comprendido entre el año 2010 al 2020 en una finca de la costa sur de Guatemala, ubicada en el municipio de Sipacate, Escuintla, con coordenadas Latitud 13°57'21.87"N y Longitud 91°12'20.82"O. La población de la finca es palma de aceite (*Elaeis guineensis* Jacq) de la variedad Deli x Ekona sembrada en el año 1990.

OBJETIVOS

General

Diseñar un modelo estadístico para pronosticar la producción de palma de aceite mediante análisis correlacionales y series de tiempo de la producción durante el periodo 2010 al 2020.

Específicos

- Estimar el grado de asociación y dependencia estadística entre la variable producción con respecto a las variables climáticas entre los años 2010 y 2020 a través del coeficiente de correlación.
- Construir un modelo estadístico de predicción de producción utilizando las variables ambientales mediante regresión lineal múltiple.
- Elaborar un modelo estadístico que se ajuste aceptablemente al comportamiento de la producción de estudio mediante un análisis de series temporales para obtener predicciones aceptables.

RESUMEN DEL MARCO METODOLÓGICO

En esta parte del informe, se expone la metodología aplicada en este estudio donde se detalla el enfoque, diseño, tipo de estudio, alcance, variables y la fase del estudio.

La fuente principal de información para desarrollar este trabajo fueron los registros de producción de la finca y los datos de la estación meteorológica representativa para el área de la finca.

- Enfoque

El enfoque de la investigación es cuantitativo, ya que se trabajó con variables numéricas que consistieron en los registros de producción, registros de temperatura mínima mensual, registros de temperatura máxima mensual, registros de temperatura promedio mensual, registros de radiación promedio mensual, registros humedad relativa promedio mensual, registros precipitación mensual y dentro de la variable tiempo que fueron periodos mensuales del año 2010 al año 2020.

- Diseño

El diseño es no experimental, porque no se realizaron ensayos en laboratorio para obtener los registros de las variables, si no que utilizo la información de registros históricos mensuales dentro del periodo 2010 al 2020, los cuales se observaron, describieron y analizaron en búsqueda de una relación estadística de las variables en estudio.

- Tipo de estudio

El estudio fue de tipo descriptivo, ya que toda la información de las variables en estudio fue tomada de registros históricos y el investigador no tuvo interferencia de ningún tipo para que estas variaran experimentalmente.

- Alcance

El alcance del estudio es descriptivo longitudinal, ya que se realizó la recolección de datos para poder describir las características de la variable dependiente (producción), y es correlacional porque se analizó la relación que tiene la variable dependiente (producción) con las independientes (tiempo, precipitación, temperatura promedio, máxima y mínima, radiación solar y humedad relativa) con la finalidad de elaborar un modelo de pronóstico de producción.

- Unidades de análisis

La población de estudio fueron los 132 registros mensuales de la producción de fruta de palma de aceite y de los registros de variables climáticas registradas por la estación meteorológica (temperatura media, máxima y mínima, radiación promedio mensual, humedad relativa promedio y precipitación mensual) para los años 2010 al 2020.

- Operativización de variables

Las variables utilizadas para lograr los objetivos fueron de tipo cuantitativa continua, con escala de razón.

Las variables clasificadas como variables dependientes son:

- Producción de racimos de fruta fresca mensual

Las variables clasificadas como variables independientes son:

- Temperatura mínima mensual
- Temperatura máxima mensual
- Temperatura promedio mensual
- Radiación promedio mensual
- Humedad relativa promedio mensual
- Precipitación mensual
- Tiempo

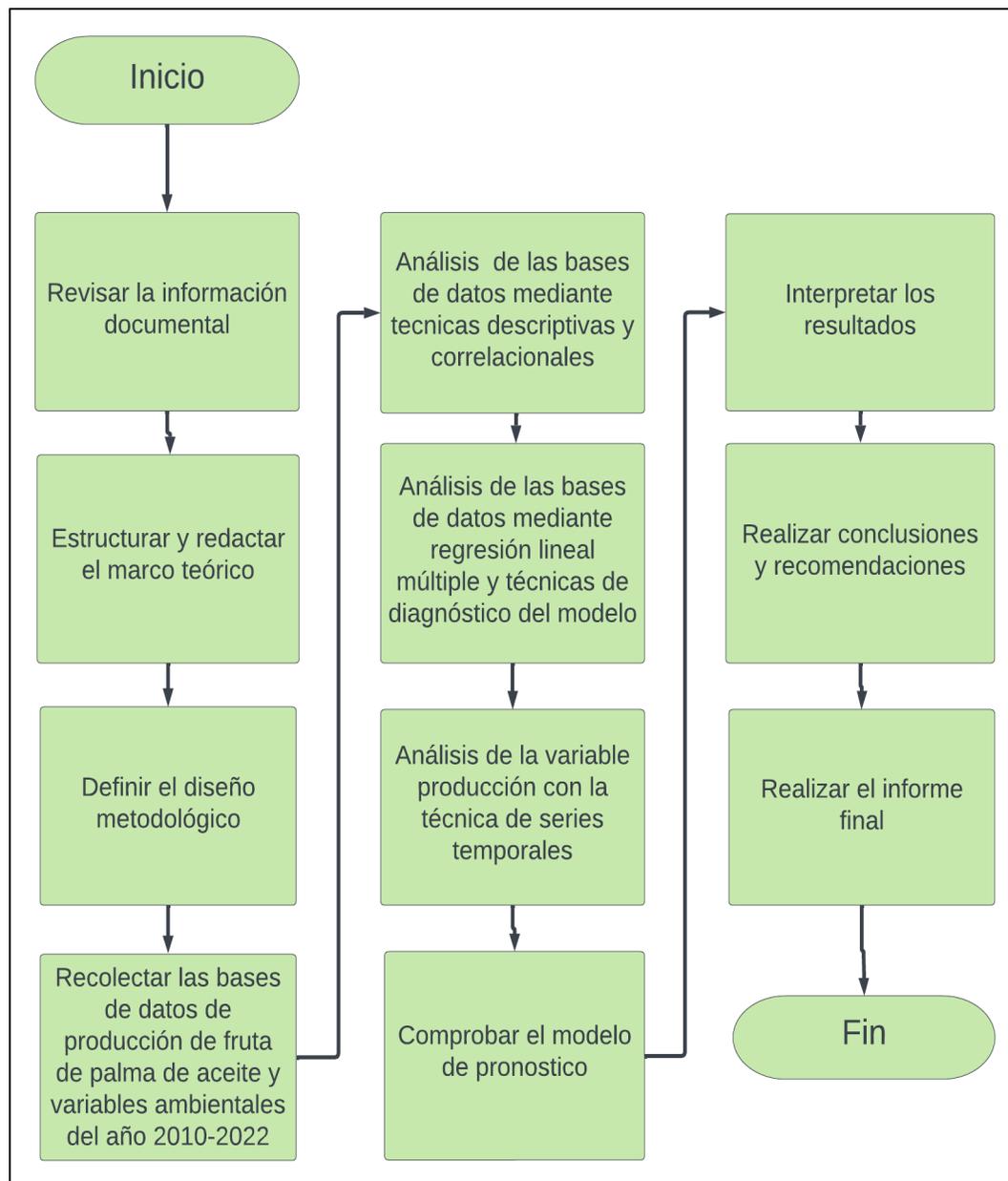
Tabla 1.*Operativización de las variables*

Variable	Definición Teórica	Definición Operativa
Producción de racimos de fruta fresca mensual (RFFm)	Es la cantidad de la masa de los racimos de fruta fresca producidos en una hectárea en un periodo mensual. Variable cuantitativa continua	Número real en kg ha ⁻¹ por mes, representa la producción de una masa por un área en un determinado periodo de tiempo. Escala de razón
Temperatura mínima mensual (Tmin)	Es la menor temperatura alcanzada en un lugar en un mes. Variable cuantitativa continua.	Medición directa: escala de intervalos °C: grados Celsius. Escala de razón.
Temperatura Máxima mensual (Tmax)	Es la mayor temperatura alcanzada en un lugar en un mes. Variable cuantitativa continua.	Medición directa: escala de intervalos °C: grados Celsius. Escala de razón.
Temperatura promedio mensual (Tmed)	Es el promedio de la temperatura registrada en un lugar en un mes. Variable cuantitativa continua.	Medición directa: escala de intervalos °C: grados Celsius. Escala de razón.
Radiación promedio mensual (Rmed)	Es el conjunto de radiaciones electromagnéticas emitidas por el sol. Variable cuantitativa continua.	Número real en W/m ² . Escala de razón
Humedad relativa promedio mensual (Hr)	Es la relación entre la presión parcial del vapor de agua y la presión de vapor de equilibrio del agua a una temperatura dada. Variable cuantitativa continua.	Expresada en porcentaje (%). Escala de razón
Precipitación mensual (Pmes)	Cantidad de agua que se precipita de la atmosfera a la superficie terrestre. Variable Cuantitativa continua.	Es un número real expresado en milímetros (mm) acumulados por mes. Escala de razón
Tiempo (t)	Es el periodo en el que se analizara la información de producción con intervalos mensuales. Variable cuantitativa continua.	El estudio tomara una referencia de doce años (2010 al 2020) y se tendrán intervalos mensuales. Escala de razón.

Nota. Descripción técnica y operativa de las variables restantes. Elaboración propia, realizado con Microsoft Word.

Figura 1.

Flujograma del esquema de solución



Nota. Se explica la ruta seguida para la realización de este trabajo de investigación. Elaboración propia, realizado con Lucid chart.

Para el presente estudio se han definido las siguientes fases de estudio:

- Fase 1. Revisión de información documental

Esta fue la fase inicial de la investigación, aquí se realizó la revisión bibliográfica, se seleccionó la que fue útil para fundamentar los temas que se analizaron en esta investigación.

- Fase 2. Recopilación de la información

En esta fase se realizó la búsqueda de las bases de datos históricas del periodo de enero 2010 a diciembre 2020, las cuales se analizaron posteriormente.

- Fase 3. Análisis de la información

En esta fase del estudio se realizaron los análisis de regresión lineal múltiple, y análisis de series temporales de los datos para poder llegar al modelo estadístico del pronóstico de la producción de fruta fresca de palma de aceite.

- Fase 4. Interpretación de resultados

En esta fase se procedió al análisis de los resultados obtenidos en la fase 3, llegando a una conclusión y a la construcción del modelo estadístico de pronóstico.

- Fase 5. Redacción del informe final

En esta fase se procedió a redactar el informe final con todos los elementos obtenidos en las fases anteriores, usando como soporte teórico y realizando los análisis y discusiones de los resultados respectivamente.

INTRODUCCIÓN

La presente investigación es una sistematización del proceso de pronóstico de producción de fruta de palma de aceite (*Elaeis guineensis* L.), en una finca de la costa sur, en el municipio de Sipacate, Escuintla, Guatemala.

La finca realiza los pronósticos de producción de una forma empírica, por lo que se plantea realizar un análisis de las variables (climáticas), que se consideran más importantes en la productividad de las plantas para poder explicar y generar un modelo que ayude a genera pronósticos con base en la estadística.

La importancia de los pronósticos de producción de fruta de palma de aceite radica en ser una herramienta que ayudará al o a los productores a estimar los ingresos y los recursos necesarios para el proceso productivo (presupuestos).

La metodología que se utilizó en este estudio fue de enfoque cuantitativo, diseño no experimental (observacional), longitudinal y un alcance descriptivo correlacional.

Las variables que se utilizaron fueron registros históricos de la producción de fruta de palma de aceite y variables climáticas correspondientes al periodo del año 2010 al 2020.

Los resultados obtenidos en este estudio contribuirán con la agroindustria de palma de aceite, ya que se obtuvo un modelo estadístico confiable que aportó

a la disminución del margen de error de los pronósticos de producción, y a que los presupuestos sean más ajustados a la realidad.

Generará más oportunidades de crecimiento empresarial al eficientizar el uso de los recursos a través de la planificación lo que constituye su relevancia social.

El trabajo de investigación fue factible ya que se contó con todos los recursos necesarios para desarrollar todas las fases del estudio.

El informe final se estructuró de la siguiente forma:

Capítulo 1: marco referencial, este contiene las investigaciones previas que tienen relación con este estudio y que sirvieron como guía o base metodológica y técnica para la elaboración de esta investigación.

Capítulo 2: marco teórico, el cual se divide en dos secciones principales, la primera, la parte estadística donde se detallan los fundamentos y teorías de las técnicas que se aplicaron. La segunda, donde se describen los conceptos agronómicos donde se detalla el funcionamiento y la importancia del cultivo de palma de aceite.

Capítulo 3: presentación de resultados, en este apartado se evidencia la baja relación estadística existente entre la variable producción y las variables climáticas, también se agrega la evaluación de modelos estadísticos basados en regresión lineal múltiple y serie de tiempo seleccionando la mejor opción para el pronóstico.

Capítulo 4: discusión de resultados, apartado en el cual se hace un análisis interno donde se argumenta la validación de la investigación y análisis externo de los resultados, donde se contrastan los resultados del estudio con resultados de otros estudios.

Y por último están las secciones de conclusiones y recomendaciones.

1. MARCO REFERENCIAL

El estudio sobre la predicción de fenómenos futuros siempre ha sido un tema bastante interesante, a lo largo de la historia han existido bastantes estudios sobre cómo reducir el margen del error de estos.

A continuación, se analizan algunos realizados con el fin de entender dentro del marco estadístico como en el marco de la naturaleza propia de la producción de fruta de palma y su posible relación de variables climáticas que inciden en ella.

Cordero et al. (1994), indican que existe poca información para la zona tropical de América, y que algunas investigaciones han permitido establecer que las fluctuaciones del rendimiento de la palma de aceite (*Elaeis guineensis* L.), estas obedecen a la compleja interacción de factores ambientales y fisiológicos de la planta. Por lo que desarrollaron un modelo estadístico para una plantación de palma de aceite del sur de Costa Rica con información del periodo entre 1981 y 1992, para predecir rendimientos de producción.

El modelo predice la producción mensual de 1 hasta 50 meses antes de la cosecha, y su diseño se llevó mediante regresión lineal múltiple escalonada considerando solo humedad relativa, producción y brillo solar de 13 variables ambientales utilizadas.

Esta investigación fue una guía técnica para la identificación de las variables a utilizar para la realización de la presente investigación. Además, sirvió como guía metodológica al manejo de la información y la aplicación de la

regresión lineal múltiple como parte de las técnicas estadísticas que se aplican en este estudio.

Erales (2017), describe que las estimaciones de producción de fruta de palma de aceite se ven influenciadas por factores conocidos como la edad, variedad, condiciones de suelo y manejo de la nutrición. Además, menciona que hay factores que no se conocen como la precipitación, temperatura, radiación solar, entre otros.

Este método de estimación consiste en un muestreo del 5 al 10 % de la población en el cual se realiza conteo de inflorescencias femeninas y racimos que estarán disponible en 6 meses según la fisiología de la planta.

La investigación fue utilizada para la identificación de las variables (temperatura media, máxima y mínima, radiación promedio mensual, humedad relativa promedio y precipitación mensual) que se utilizaron para la elaboración de esta investigación.

Nda, et al. (2018) realizaron un estudio en el cual fueron evaluadas las variables precipitación, temperatura y producción de racimos de fruta de palma de aceite, en el cual utilizaron la técnica Mann-Kendall, coeficiente de correlación de Pearson y la regresión lineal. El estudio mostró que las fluctuaciones de lluvia y temperatura son escalonadas y que pueden ayudar a identificar la dirección en la cual se están dando estas. Además, los resultados de correlación y de regresión lineal muestran que no existe un efecto significativo sobre la producción, y que existen otros factores que pueden determinar esta.

El estudio ayudó a la investigación como guía de aplicación de la técnica de la regresión lineal a la producción de racimos de fruta de palma y factores ambientales.

Ramos-Escalante et al. (2018) realizaron un estudio con el cual se definió un modelo para estimar la cantidad de Carbono (variable dependiente), almacenado en la biomasa según el diámetro a la altura del pecho (variable independiente). La técnica para estimar el Carbono fue de digestión húmeda de Walkley y Black, se utilizó un muestreo destructivo aleatorizado, utilizando 12 palmas de muestras.

Utilizaron análisis de regresión lineal simple para el desarrollo de dos ecuaciones para la estimación de la cantidad de Carbono una en función de la variable altura y otra en función de la variable diámetro de palma, obteniendo coeficientes de ajuste de 0.5195 y 0.7621 respectivamente.

Esta investigación aportó al estudio una guía de aplicación de la técnica estadística de la regresión lineal simple, ayudando a interpretar el ajuste de esta y la creación de modelos estadísticos.

Woittiez et al. (2018) describieron la visión general sobre los factores que determinan, reducen y limitan la productividad del cultivo de palma de aceite. Afirman que el rendimiento potencial se encuentra determinado por la radiación fotosintéticamente activa, temperatura, concentración de CO₂, el material de siembra, déficit hídrico y la densidad de siembra.

La investigación aportó al estudio una guía para identificar los factores o variables que limitan la productividad del cultivo, ayudando a elegir las variables a estudiar.

L. Laurente y F. Laurente (2019) utilizaron la metodología ARIMA de Box y Jenkins para poder pronosticar la producción de papa en la región de Puno, Perú. Para la cual se utilizaron datos anuales de producción comprendidos entre el periodo del año 1950 a 2017 obtenidos del Ministerio de Agricultura y del Banco Central del Perú.

Con los datos se utilizaron test de estacionariedad ADF de Dickey y PP de Philips y Perron con un nivel de significancia del 1 %, el cual mostró que la producción de papa no es estacionaria.

Se evaluaron varios métodos para la estimación como modelos autorregresivos (AR), medias móviles (MA) y modelos autorregresivos integrados y de medias móviles (ARIMA) utilizando la metodología de mínimos cuadrados.

Se eligió el modelo ARIMA (3, 0, 4) mediante el Criterio de Información de Akaike (AIC) y Criterio de Información de Schwarz (SC).

La investigación sirvió como una guía para la investigación de cómo aplicar las series temporales para poder realizar pronósticos de la producción a través de datos históricos.

Ramirez y Potes (2019) realizaron un estudio para construir un modelo que se adaptaba a la estimación del rendimiento del cultivo de maracuyá a partir de datos históricos del periodo comprendido entre los años 2007 y 2014 de plantaciones ubicadas en el municipio de la Plata en el departamento del Huila, en Colombia. Evaluaron el análisis de regresión múltiple y el modelo ARIMA, este último se ajustó más a las series temporales analizadas.

El estudio aportó la experiencia de la utilización de un modelo de series temporales, por lo que se utilizó como guía para la aplicación de estos modelos de serie temporales.

García-Arteaga et al. (2020), realizaron experimentos computacionales para poder obtener predicciones de los cultivos mediante el aprendizaje automático. Las bases de datos utilizadas comprenden entre el 1970 al 2020 con los productos sembrados y cosechados en Argentina. Utilizaron 13 variables, de las cuales 2 se trataron como dependientes (superficie cosechada y producción), y el resto como independientes (velocidad del viento, precipitación, temperatura promedio, máxima y mínima; latitud, longitud, y superficie por encima del nivel del mar; hectáreas sembradas y rendimiento).

La presente investigación aportó la experiencia ganada en el manejo de información en bases de datos para generar pronósticos, esto se utilizó como guía en el manejo de las bases de datos de la información utilizada.

Jaramillo (2020) indica que la palma africana (*Elaeis guinensis* L.) es la planta de las oleaginosas que tiene el mayor potencial de producción de aceite por unidad de área. Este cultivo responde fácilmente a cualquier tipo de estrés biótico o abiótico durante 40 meses, periodo de tiempo que se toma la planta desde que se inicia el botón floral hasta que se cosecha el racimo maduro, reduciendo su producción.

La investigación aportó al estudio la descripción detallada de la formación de los racimos, esta información se utilizó para comprender el proceso de formación y poder estructurar la investigación desde el punto de vista fisiológico de la planta.

Luis-Rojas et al. (2020) desarrollaron un modelo para la predicción de la producción de la vainilla verde en México durante el periodo comprendido entre los años 1961 al 2016. Evaluaron series temporales ARIMA con la metodología Box y Jenkins, se eligió el mejor modelo mediante el Criterio de Información de Akaike (AIC) y el criterio bayesiano de Shwartz (SBC) los cuales comparan la bondad de ajuste de los diferentes modelos. El modelo que más se ajustó fue el ARIMA (1,1,1) mostrando una variación a corto plazo de $\pm 8.32\%$.

La investigación aportó experiencia para seleccionar modelos de series temporales a la presente investigación, en la cual se comparan varios de estos.

2. MARCO TEÓRICO

2.1. Fundamentos estadísticos

A continuación, se presentan conceptos estadísticos que sustentan teóricamente este estudio.

2.1.1. La estadística como ciencia

La estadística es la ciencia de los datos, la que se encarga de recogerlos, organizarlos e interpretarlos. Es fundamental para otras ramas de la ciencia, pero lo más importante es el aporte que tiene en la interpretación de los datos de la investigación científica (Gorgas et al., 2011).

2.1.2. Tipos de estadística

Por lo general la estadística se divide en dos categorías: descriptiva e inferencial (Lind et al., 2012).

2.1.2.1. Estadística descriptiva

Según Lind et al. (2012), es la parte de la estadística que tiene métodos para organizar, resumir y presentar datos solo para informar las características de estos.

2.1.2.1.1. Medidas de tendencia central

Las medidas de tendencia central indican la ubicación del conjunto de datos, resumiendo el total de estos. Las más comunes son la media, la mediana y la moda (Galindo-Domínguez, 2020).

La media: es el proceso de promediar los datos, se realiza mediante la sumatoria de todos los valores y se divide dentro del número de valores. En estadística se pueden clasificar en dos tipos de media, la media de la muestra o media muestral y la media de la población o media poblacional (Hanke y Reitsch, 1996).

Según los autores mencionados anteriormente, la media poblacional se representa por la letra griega μ (mu) y su ecuación es:

$$\mu = \frac{\sum X}{N} \quad (1)$$

Donde:

μ : media poblacional

$\sum X$: la suma de todos los valores de la población

N: el tamaño de la población

Según Hanke y Reitsch (1996) la media muestral se representa por el símbolo \bar{X} y la ecuación es:

$$\bar{X} = \frac{\sum X}{n} \quad (2)$$

Donde:

\bar{X} : media muestral

$\sum X$: la suma de todos los valores de la población

n: el tamaño de la muestra

La mediana: esta medida indica el valor del punto medio de los datos, tanto por encima como por debajo el 50 % de los casos. Cuando el número de casos es par, la mediana es el promedio de los dos casos centrales (Galindo-Domínguez, 2020).

Moda: es el valor de una variable dentro de una muestra que tiene una frecuencia máxima, o dicho de otra forma es la que más veces se repite (Gorgas et al., 2011).

2.1.2.1.2. Medidas de dispersión

Según Galindo-Domínguez (2020), las medidas de dispersión son las que ayudan a cuantificar cuanto se dispersan los datos. Los más utilizados son la varianza, la desviación estándar o desviación típica y el rango.

La varianza: es una medida de dispersión que se basa en la diferencia de los datos con la media para cada uno de ellos. Existen dos tipos o formas de ellas, cuando se tiene la totalidad de los datos de la población se le llama varianza poblacional, la cual se representa con la letra griega sigma al cuadrado (σ^2) (Posada, 2016), su ecuación es la siguiente:

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} \quad (3)$$

En donde:

- σ^2 : varianza poblacional
- x_i : cada dato u observación de la variable X
- μ : media poblacional
- N: número total de datos de la población

Además, también se puede utilizar la varianza de la muestra, la cual se representa con S^2 , y esta tiene como objetivo convertirse en un estimador de la variación para la población (Posada, 2016), su ecuación es la siguiente:

$$S^2 = \frac{\sum(x_i - \bar{X})^2}{n-1} \quad (4)$$

En donde:

- S^2 : varianza muestral
- x_i : cada dato u observación de la variable X
- \bar{X} : media muestral
- n: número de datos de la muestra

Desviación estándar: esta medida de dispersión es una de las más utilizadas en la estadística descriptiva, ya que su función es indicar cuantos puntos hacia arriba o hacia abajo se alejan los datos de la media. Cuando el resultado de la desviación estándar es bajo indica que la variabilidad de los datos es baja y se encuentra más cercanos a la media, caso contrario si el resultado de la desviación estándar es alto el conjunto de datos se encuentra agrupado más lejano a la media (Galindo-Domínguez, 2020).

Según Posada (2006), si el objetivo es calcular la desviación estándar para la población se representa con la letra griega σ , y su ecuación es la siguiente:

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \quad (5)$$

En donde:

- σ : desviación estándar poblacional
- x_i : cada dato u observación de la variable X
- μ : media poblacional
- N: número total de datos de la población

Y si el objetivo es la desviación estándar para la muestra se representa con un “S” (Posada, 2016), y su ecuación es la siguiente:

$$S = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} \quad (6)$$

En donde:

- S: desviación estándar muestral
- x_i : cada dato u observación de la variable X
- \bar{X} : media muestral
- n: número de datos de la muestra

Rango: este es la diferencia entre el valor máximo y el valor mínimo del conjunto de datos, por tal razón también es conocido como recorrido o amplitud. Su uso mayormente es en variables cuantitativas (Galindo-Domínguez, 2020).

2.1.2.2. Estadística inferencial

Según Lind et al. (2012), indican que es la parte de la estadística que utiliza métodos para determinar una propiedad de una población basada en la información proporcionada por una parte de ella.

2.1.3. Pronósticos

Se define el pronóstico como una estimación de un dato a futuro, el pronóstico puede ser construido con métodos cuantitativos, cualitativos o una combinación de ambos métodos, y puede ser basado con factores externos e internos (Frausto, 2009).

Anderson et al. (2008,) afirman que “Un pronóstico es simplemente un pronóstico de lo que ocurrirá en el futuro. Los administradores deben aceptar que, sea cual sea la técnica que usen, no podrán obtener pronósticos perfectos” (p. 766).

2.1.3.1. Clasificación de los métodos de pronóstico

Existen varios métodos los cuales se pueden clasificar en tres grupos: cualitativos, de proyección histórica y casuales. La precisión relativa de cada grupo difiere sobre el pronóstico entre el largo y corto plazo (Frausto, 2009).

2.1.3.1.1. Métodos cuantitativos

Frausto (2009), indica que son métodos utilizados para pronósticos a mediano o largo plazo, y para este método se utiliza el juicio, encuestas o técnicas comparativas para hacer estimados cuantitativos acerca del futuro.

2.1.3.1.2. Métodos de proyección histórica

Estos métodos requieren de un volumen alto de datos históricos que muestren estabilidad en su estacionalidad y de tendencia sean estables y bien

definidas en la serie de tiempo. Estos modelos son bien efectivos a corto plazo y se basan en la premisa que el patrón del futuro replicará al del pasado. Estos modelos logran rastrear los cambios a medida que se va actualizando la información, por lo que son capaces de adaptarse a los cambios de patrones de tendencia y estacionalidad, aunque si el cambio es repentino los modelos no emiten una señal de cambio (Frausto, 2009).

2.1.3.1.3. Métodos causales

Estos métodos son desarrollados por un modelo de causa y efecto entre dos o más de dos variables. Para desarrollar estos métodos se debe recopilar la información previa y realizar análisis para constar que sean válidos. El modelo causal más conocido y usado es el de la regresión lineal (Schroeder et al., 2011).

Frausto (2009), indica que los métodos causales con frecuencia no encuentran verdaderas variables causales y su asociación con la variable que pronostica es baja.

2.1.3.2. Análisis correlacionales

Estos hacen referencia al grado de asociación o correlación que existen entre dos o más variables (Galindo-Domínguez, 2020).

2.1.3.2.1. Coeficiente de correlación

Este coeficiente describe la fuerza de asociación que tiene dos conjuntos de variables, su símbolo es r y comúnmente llamado r de Pearson, debido a su creador Karl Pearson en los años 1900 aproximadamente. Su valor va de -1 a 1, los cuales indicarían una correlación perfecta, siendo para 1 la relación en un

sentido lineal positivo y para -1 la relación en un sentido lineal inverso (Lind et al. 2012).

El coeficiente de correlación de Pearson está dado por:

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{(n-1)S_x S_y} \quad (7)$$

Donde:

- r : coeficiente de correlación de Pearson
- X : son los valores de la variable en comparación
- Y : son los valores de la variable en comparación
- $(n-1)$: son los grados de libertad
- \bar{X} : es la media de la variable X
- \bar{Y} : es la media de la variable Y

2.1.3.2.2. Regresión lineal simple

Este consiste en un método estadístico en el cual se encuentra la relación entre dos variables, a las cuales se les llamará variable dependiente o variable Y , y variable independiente o variable X . Este método ayuda a la predicción de la variable Y , mediante la estimación de la ecuación de regresión basado en el método de mínimos cuadrados (Lind et al. 2012).

$$E(y) = \beta_0 + \beta_1 x \quad (8)$$

Donde:

- $E(y)$: es la variable dependiente que se quiere predecir
- x : es la variable independiente o predictora
- β_1 : es la pendiente de la línea de regresión
- β_0 : es el intercepto

2.1.3.2.3. Regresión lineal múltiple

Esta técnica puede verse como una herramienta descriptiva o inferencial. Desde la óptica descriptiva se puede: encontrar la mejor ecuación de predicción lineal, controlar algunos para poder evaluar la contribución a un modelo lineal de unas variables específicas y encontrar algunas relaciones estructurales. Desde la óptica inferencial se puede decir: que no hay relación entre una variable dependiente y un conjunto de variables independientes, que una variable independiente no tiene efecto lineal sobre la variable dependiente y la relación entre la variable dependiente y una particular variable independiente no es lineal (Silva et al., 2017).

$$E(y) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p \quad (9)$$

Donde:

$E(y)$: es la variable dependiente que se quiere predecir

x_1, x_2, x_p : son las variables independientes o predictoras

$\beta_1, \beta_2, \beta_p$: son los coeficientes de la regresión

β_0 : es el intercepto

2.1.3.2.4. Coeficiente de determinación R cuadrado

Este es un coeficiente de análisis de la varianza del modelo, comúnmente llamado R cuadrado. Este coeficiente representa la proporción de la variación de la variable respuesta que explica el modelo. Este coeficiente es positivo y se encuentra entre 0-1, cuanto más cercano este a un valor de 1 mejor será el desempeño del modelo.

$$R^2 = \frac{SSRes}{SSTo} = 1 - \frac{SSRes}{SSTo} \quad (10)$$

Donde:

SSRes: suma de cuadrados de los residuos

SSTo: suma de cuadrados total

Cuando se agregan variables explicativas (X) al modelo de regresión el R cuadrado solo aumenta, y nunca tiende a bajar. Esto debido a la suma de cuadrados de residuos (SSRes), no puede aumentar con más covariables X y la suma de cuadrados total (SSTo), siempre vale lo mismo para un conjunto fijo de respuestas Y_i . Por tal razón se recomienda cuando se incluyen más covariables X hacer uso también del coeficiente R cuadrado ajustado (Szretter, 2017).

2.1.3.2.5. Coeficiente de determinación ajustado R_a^2

El coeficiente de determinación ajustado o R cuadrado ajustado a diferencia del coeficiente de determinación R cuadrado, este si puede disminuir o aumentar cuando se agregan covariables, ya que la SSRes puede ser más compensada por la pérdida de un grado de libertad en el denominador n-p:

$$R_a^2 = 1 - \frac{\frac{SSRes}{\frac{n-p}{n-1}}}{\frac{SSTo}{n-1}} = 1 - \left(\frac{n-1}{n-p} \right) \frac{\frac{SSRes}{n-p}}{\frac{SSTo}{n-1}} \quad (11)$$

Donde:

SSRes: suma de cuadrados de los residuos

SSTo: suma de cuadrados total

n - 1: grados de Libertad

n - p: grados de Libertad del denominador

Cuando se explica Y en un modelo y se compara con otro modelo que tiene las mismas covariables más otra covariable y se ve el aumento de R^2_a , es indicativo que la covariable adicional es importante para predecir a Y. Sin embargo, cuando R^2_a no aumenta o incluso disminuye al incorporar al modelo la covariable adicional, es indicativo que la covariable adicional no aporta valor a la predicción del modelo y puede predecirse solo con la covariables iniciales (Szretter, 2017).

2.1.3.2.6. Identificación de datos atípicos

Es importante la identificación de datos *outliers* o atípicos, estos datos no siguen los patrones de las otras observaciones. Dentro de las metodologías para identificar estas, se tiene el estadístico de Leverage, y los residuos estudentizados y la distancia de Cook.

Leverage es una medida que indica la posición o distancia entre X_i y su media muestral de las X, por lo que, indica cual es el aporte de la observación i -ésima a la varianza muestral de las X. Este no depende del valor Y_1 observado, y su traducción al español significa palanca o influencia (Szretter, 2017).

- Distancia de Cook

Este estadístico permite identificar los datos más influyentes en la estimación del modelo de regresión lineal múltiple el cual fue introducido en el 1977 por Ralph Dennis Cook basada en la función de influencia (Tan et al., 1999).

$$D_i = \frac{(y_i - \hat{y}_i)^2}{(p-1)s^2} \left[\frac{h_i}{(1-h_i)^2} \right] \quad (12)$$

Donde:

D_i : medida de la distancia de Cook para la observación i

$y_i - \hat{y}_i$: residual de la observación i

h_i : influencia de la observación i

p : número de variables independientes

s : error estándar

2.1.4. Análisis de serie de tiempo

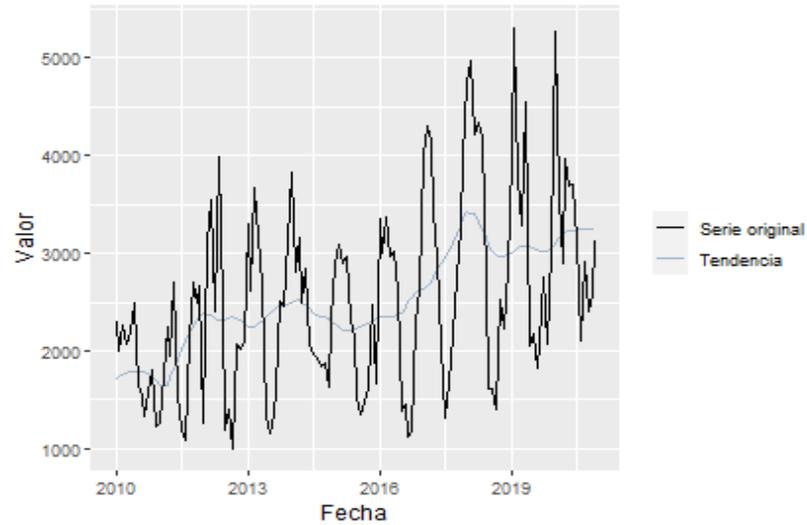
Una serie de tiempo es un conjunto observado de datos en periodo de tiempo o periodo sucesivo de tiempo, y su objetivo es la predicción de valores futuros de la misma variable, identificando patrones y después extrapolarlos para conseguirlos.

En las series temporales los datos siguen un comportamiento que están dados por cuatro componentes que se conjugan para dar los valores a estas, los cuales se detallan a continuación:

- Componente de tendencia: este es el desplazamiento gradual a valores relativamente altos o bajos, y este dado por factores de largo plazo.
- Componente cíclico: es toda sucesión de puntos recurrente que se encuentran por abajo y por arriba de la tendencia, que sea mayor a un periodo de un año, (ver figura 2).

Figura 2.

Serie temporal de la producción mensual kg/ha, periodo 2010-2020

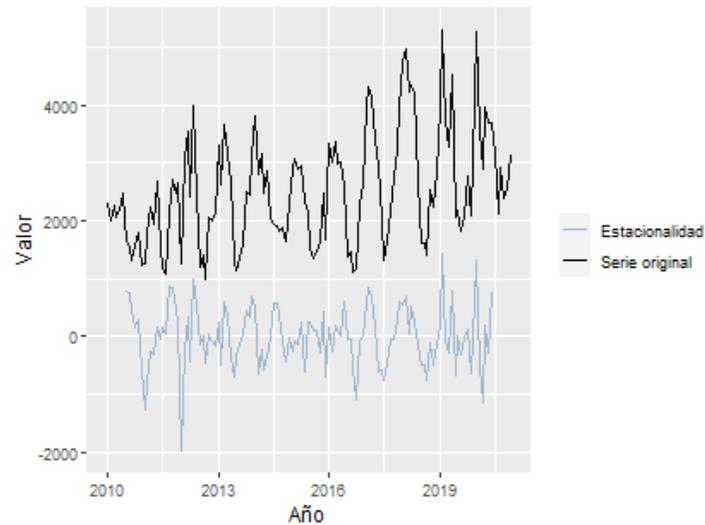


Nota. En la gráfica se observa el componente de tendencia de la serie en color negro, el cual se muestra al alza. Y el componente cíclico con periodos de 12 meses. Elaboración propia, realizado con Rstudio

- Componente estacional: es la variabilidad de la serie de tiempo menor o igual a un año, y muestra variaciones regulares por temporadas marcadas, (ver figura 3).

Figura 3.

Serie temporal de la producción mensual kg/ha, periodo 2010-2020

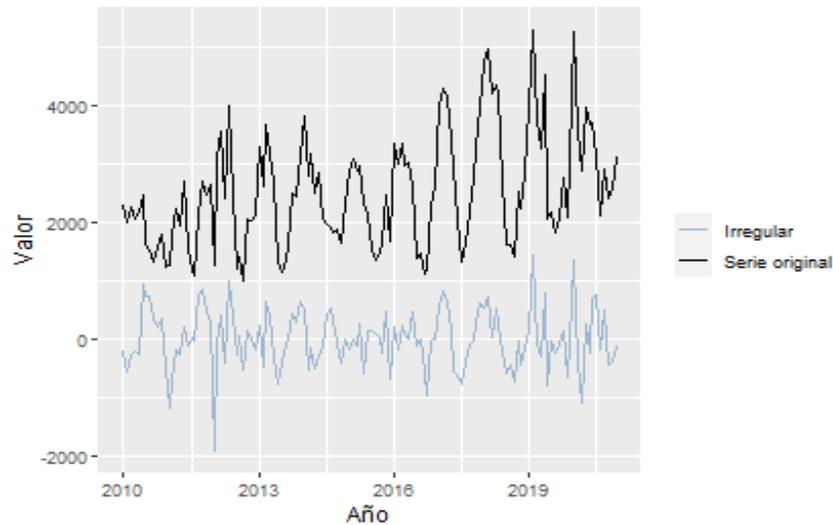


Nota. En la gráfica se observa el componente estacional de la serie en color gris azulado y la serie original en color negro. Elaboración propia, realizado con Rstudio

- **Componente irregular:** este es un componente aleatorio, impredecible ya que es ocasionado a corto plazo por algunos factores, por algunos imprevistos y que no son recurrentes. Se considera como un factor residual de los tres anteriormente descritos y muestra las desviaciones de los valores reales (Anderson et al., 2008), (ver figura 4).

Figura 4.

Serie temporal de la producción mensual kg/ha, periodo 2010-2020



Nota. En la gráfica se observa el componente irregular de la serie en color gris azulado y la serie original en color negro. Elaboración propia, realizado con Rstudio

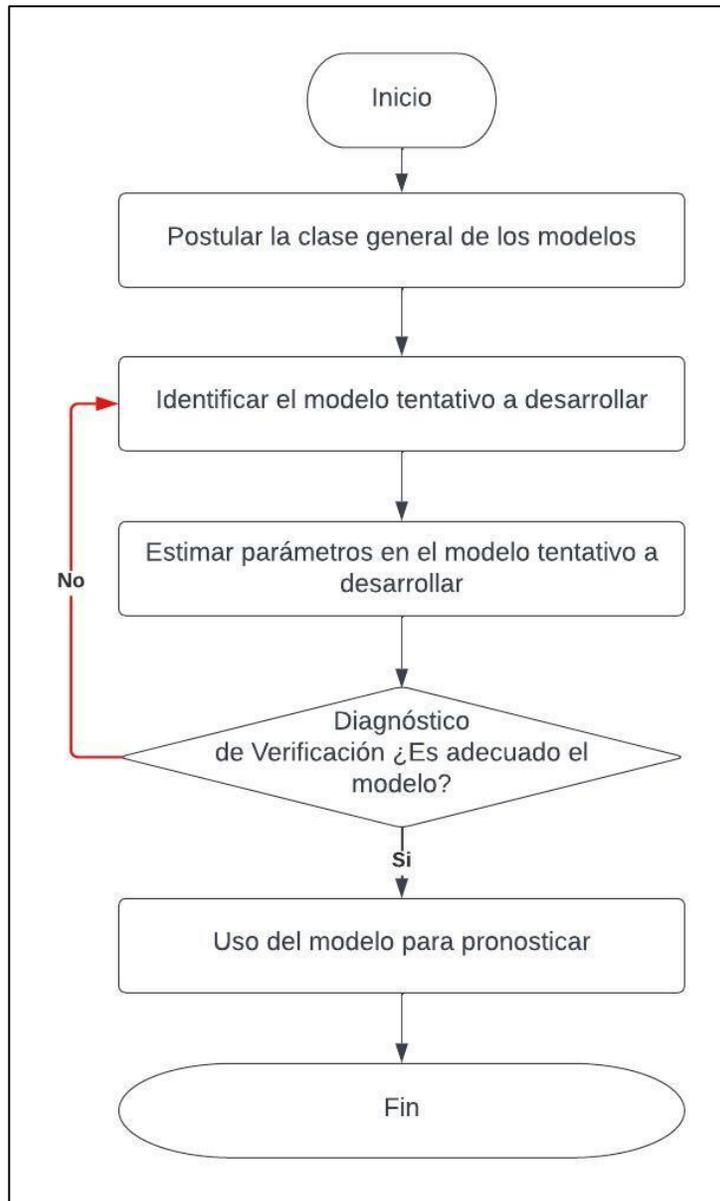
2.1.4.1. Metodología Box-Jenkins (ARIMA)

Este método no asume ningún patrón en particular de los datos, utiliza el enfoque de identificación de un modelo a partir de algunos de tipo general, el cual verifica contra los históricos para verificar la precisión de la descripción. Las series temporales ARIMA, son de tipo Box-Jenkins.

Hanke y Reitsch (1996) afirman “Las técnicas de Box-Jenkins aplican métodos autorregresivos y de promedio móvil a los problemas de pronóstico de serie de tiempo” (p. 432).

Figura 5.

Diagrama de flujo del método de Box-Jenkins



Nota. Se explica la ruta seguida para la realización de un modelo de serie de tiempo con el método Box-Jenkins. Obtenido de J. Hanke y A. Reitsch (1996). *Pronóstico en los negocios*. (p. 432). Prentice Hall.

2.1.4.1.1. Modelos autorregresivos

Estos modelos pronostican la variable al utilizar una combinación lineal de predictores. El termino autorregresivo dice que se trata de una regresión de la variable contra ella misma (Hyndman y Athanasopoulos, 2018).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (13)$$

Donde:

y_t : es el valor de la serie temporal en el momento t

c : es una constante

$y_{t-1}, y_{t-2}, \dots, y_{t-p}$: son los valores pasados de la serie temporal en los momentos $t-1, t-2, \dots, t-p$

$\phi_1, \phi_2, \dots, \phi_p$: son los coeficientes autorregresivos que representan la relación de autocorrelación entre la serie en diferentes retrasos.

ε_t : ruido blanco

2.1.4.1.2. Modelos de promedio móvil

Este modelo utiliza errores de pronóstico pasados en un modelo similar a una regresión, este se puede considerar como un promedio móvil ponderado. Este modelo no debe confundirse con un modelo de suavizado de promedio móvil, ya que el modelo de promedio móvil se utiliza para pronosticar valores a futuro, en cambio el modelo suavizado de promedio móvil se utiliza para estimar los ciclos de tendencia de los valores pasados (Hyndman y Athanasopoulos, 2018):

$$y_t = C + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_p \varepsilon_{t-p} \quad (14)$$

Donde:

y_t : es el valor de la serie temporal en el momento t

c : es la media de la serie temporal

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$: son los valores pasados de los términos de error en los momentos $t-1, t-2, \dots, t-q$

$\theta_1, \theta_2, \dots, \theta_p$: son los coeficientes de la media móvil que representan la relación de autocorrelación entre los términos de error en diferentes retrasos.

ε_t : ruido blanco

2.1.4.1.3. Modelos ARIMA no estacionales

Este modelo se obtiene de la combinación de diferenciación con autorregresivo y un modelo de media móvil (Hyndman y Athanasopoulos, 2018).

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_p \varepsilon_{t-q} + \varepsilon_t \quad (15)$$

Donde:

y'_t : es el valor de la serie temporal en el momento t

c : es una constante

ϕ_1, \dots, ϕ_p : son los coeficientes autorregresivos que representan la relación de autocorrelación entre la serie en diferentes retrasos.

y'_{t-1}, y'_{t-p} : son los valores pasados de la serie temporal en los momentos $t-1, \dots, t-p$

- $\theta_1, \dots, \theta_p$: son los coeficientes de la media móvil que representan la relación de autocorrelación entre los términos de error en diferentes retrasos.
- $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$: son los valores pasados de los términos de error en los momentos $t-1, t-2, \dots, t-q$
- ε_t : ruido blanco

2.1.4.1.4. Modelos ARIMA estacionales

Este modelo se obtiene de agregar componentes estacionales adicionales a un modelo ARIMA. Su representación son ARIMA (p,d,q) (P,D,Q)m, el primer término es la parte no estacional del modelo, el segundo término es la parte estacional del modelo (Hyndman y Athanasopoulos, 2018). Un ejemplo para un modelo ARIMA (1,1,1)(1,1,1)[4] es:

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t - \Phi_1 Y_{t-s} - \dots - \Phi_P Y_{t-Ps} + \theta_1 \varepsilon_{t-s} + \dots + \theta_Q \varepsilon_{t-Qs} \quad (16)$$

Donde:

- Y_t : es el valor de la serie temporal en el momento t
- c : es una constante
- ϕ_1, \dots, ϕ_p : son los coeficientes autorregresivos no estacionales.
- Y_{t-1}, Y_{t-p} : son los valores pasados de la serie temporal en los momentos $t-1, \dots, t-p$ no estacionales
- $\theta_1, \dots, \theta_p$: son los coeficientes de la media móvil no estacional.
- $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$: son los valores pasados de los términos de error no estacionales en los momentos $t-1, t-2, \dots, t-q$
- ε_t : ruido blanco

$\Phi_1, \dots, \Phi_p:$	son los coeficientes autorregresivos estacionales
$Y_{t-s}, Y_{t-ps}:$	son los valores pasados de la serie temporal en los momentos $t-1, \dots, t-ps$ estacionales, donde s es el periodo de la estacionalidad
$\theta_1, \dots, \theta_Q:$	son los coeficientes autorregresivos estacionales
$\varepsilon_{t-s}, \dots, \varepsilon_{t-Qs}:$	son los valores pasados de los términos de error estacionales en los momentos $t-1, t-2, \dots, t-Qs$

2.1.5. Prueba de normalidad

La normalidad de los datos es esencial para la validez de las pruebas de hipótesis paramétricos y modelos de regresión. La falta de normalidad puede llevar a la ineficiencia del estimador mínimo-cuadrático y a la aproximación en intervalos de confianza y controles de significancia. Aunque la prueba de normalidad es un supuesto crítico, si se tiene certeza de la normalidad en las poblaciones de origen, los resultados paramétricos pueden ser válidos. En resumen, la verificación de la normalidad es crucial antes de aplicar estadísticos (C. Flores y K. Flores, 2021).

2.1.5.1. Shapiro-Wilk

Según Carmona y Carrión (2015) (como se cita en C. Flores y K. Flores, 2021), esta es una prueba estadística que se utiliza para evaluar normalidad de una muestra, la hipótesis nula (H_0) y la hipótesis alternativa (H_1) para la prueba de Shapiro-Wilk son las siguientes:

- $H_0:$ La muestra sigue una distribución normal.
- $H_1:$ La muestra no sigue una distribución normal.

La prueba de Shapiro-Wilk se representa por una letra W y su ecuación es la siguiente:

$$W = \frac{\sum_{i=1}^n \alpha_i Y_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (17)$$

Donde:

- W: prueba Shapiro-Wilk
- n: tamaño de la muestra
- Y_i : valores ordenados de menor a mayor
- \bar{Y} : media de la muestra
- α_i : coeficientes de ponderación

2.1.6. Pruebas de diagnóstico de los modelos

Las pruebas de diagnóstico de modelos buscan evaluar la capacidad del modelo analítico para expresar eficientemente la información contenida en los datos observados. Aunque el objetivo ideal es lograr una traslación perfecta, uno a uno, para evitar pérdida de información, esta tarea es imposible debido a la finitud de los datos. En consecuencia, el objetivo práctico es identificar el modelo que mejor se ajuste a los datos, minimizando la pérdida de información (Martínez, 2009).

2.1.6.1. Criterio de información de Akaike (AIC)

Es un criterio de selección de modelos utilizado en estadística. Esta es una medida que tiene en cuenta tanto la bondad de ajuste del modelo como su complejidad, por lo que se busca el equilibrio entre el ajuste y la simplicidad del modelo (Martínez, 2009). Esta prueba se representa por las letras "AIC" y su ecuación es la siguiente:

$$AIC = -2 \log(L) + 2k \quad (18)$$

Donde:

AIC: criterio de información de Akaike

L: función de verosimilitud

k: número de parámetros en el modelo

2.1.6.2. Criterio de información Bayesiano (BIC)

Este método fue propuesto por Schwarz en 1978, y ha sido uno de los más utilizados en la selección de modelos, y se basa en la evaluación de los modelos en términos de sus probabilidades posteriores (Montesinos, 2011). Esta prueba se representa por las letras "BIC" y su ecuación es la siguiente:

$$BIC = 2 \log(L) + k(\log(n)) \quad (19)$$

Donde:

AIC: criterio de información Bayesiano

L: función de verosimilitud

k: número de parámetros en el modelo

n: tamaño de la muestra

2.1.6.3. Factor de inflación de la varianza (VIF)

El Factor de Inflación de Varianza (VIF), es una métrica utilizada en estadísticas para evaluar la colinealidad entre variables predictoras en un modelo de regresión. Se calcula examinando el cuadrado del coeficiente de correlación múltiple (R^2) de una variable predictora específica (X_i) en comparación con todas las demás variables predictoras. Si X_i tiene una fuerte relación lineal con otras

variables, resultando en un R^2 cercano a 1, el VIF correspondiente será grande. Valores de VIF superiores a 10 indican problemas de colinealidad. En ausencia de relaciones lineales entre las variables predictoras, el VIF es 1, señalando ortogonalidad. El VIF también cuantifica cómo la varianza del coeficiente de regresión de X_i aumenta debido a su asociación lineal con otras variables, comparado con la varianza si no hubiera relación lineal (Chatterjee & Hadi, 2012).

$$VIF = \frac{1}{1-R_i^2} \quad (20)$$

Donde:

VIF: factor de inflación de varianza

R_i^2 : coeficiente de determinación resultante de la regresión de X_i , en función de las demás variables predictoras

2.1.6.4. Cp de Mallows

Cp (Estadístico de Proceso de Predicción Correcta), es una métrica utilizada para evaluar el desempeño de un modelo de regresión en términos del error cuadrático medio total estandarizado de predicción para los datos observados. Un valor Cp cercano a p indica un buen rendimiento del modelo, y la selección de subconjuntos de variables se realiza gráficamente, trazando Cp frente al número de términos (p) en el modelo. La línea Cp = p en el gráfico ayuda a identificar subconjuntos deseables de variables para la construcción del modelo (Chatterjee & Hadi, 2012). Se calcula considerando la varianza y el sesgo de la predicción, y se expresa mediante la siguiente ecuación:

$$Cp = \frac{SSE_p}{MSE_{full}} - (n - 2p) \quad (21)$$

Donde:

SSE_p : suma de los errores cuadráticos para el modelo con p predictores

MSE_{full} : error cuadrático medio para el modelo completo

n : tamaño de la muestra

p : número de predictores en el modelo

2.1.6.5. Prueba de Breusch Pagan

La prueba de Breusch-Pagan es una herramienta crucial en econometría para evaluar la heterocedasticidad en modelos de regresión. Se basa en la relación entre los residuos cuadráticos y las variables explicativas, ofreciendo una valiosa indicación sobre la presencia de variabilidad no constante de errores (Breusch & Pagan, 1979).

2.1.6.6. Prueba de Durbin-Watson

La prueba de Durbin-Watson es una estadística utilizada para evaluar la presencia de autocorrelación en los residuos de un modelo de regresión. Se basa en la comparación de las autocorrelaciones de primer orden de los residuos con el objetivo de detectar patrones sistemáticos de dependencia temporal. El estadístico de Durbin-Watson (DW) toma valores entre 0 y 4, donde valores cercanos a 2 indican ausencia de autocorrelación, mientras que desviaciones significativas de 2 sugieren la presencia de autocorrelación positiva o negativa (Wooldridge, 2010).

2.1.6.7. Prueba de Ljung-Box

La prueba de Ljung-Box, desarrollada por Greta M. Ljung y George E. P. Box, es una prueba estadística utilizada para evaluar si hay autocorrelación

significativa en una serie temporal. Su enfoque principal es examinar si las autocorrelaciones de una serie temporal hasta cierto rezago son estadísticamente diferentes de cero. El estadístico de prueba se compara con los valores críticos de la distribución chi-cuadrado para determinar si se puede rechazar la hipótesis nula de ausencia de autocorrelación (Tunnickliffe, 2016).

2.1.6.8. Prueba de Dickey Fuller aumentado

La prueba de Dickey-Fuller Aumentado (ADF), es una herramienta estadística utilizada para evaluar la presencia de raíces unitarias en una serie temporal, lo que implica determinar si la serie es estacionaria o no. Desarrollada por David Dickey y Wayne Fuller, esta prueba extiende la prueba Dickey-Fuller básica al incluir términos de rezagos y tendencias en el modelo, proporcionando una evaluación más robusta de la estacionariedad (Enders, 2015).

2.1.6.9. Función de autocorrelación (ACF)

La función de autocorrelación (ACF, por sus siglas en inglés), es una medida estadística que evalúa la correlación lineal entre los valores de una serie temporal y sus rezagos. Esta herramienta ayuda a identificar patrones repetitivos o ciclos en la serie, proporcionando información sobre la dependencia temporal de los datos (Montgomery, Peck & Vining, 2012).

2.1.6.10. Función de autocorrelación parcial (PACF)

La función de autocorrelación parcial (PACF, por sus siglas en inglés), es una medida estadística que evalúa la correlación directa entre dos valores en una serie temporal, excluyendo la influencia de los rezagos intermedios. Esta herramienta es esencial para identificar la dependencia temporal específica entre

puntos de datos, ayudando en el análisis detallado de patrones en la serie (Chatfield, 2016).

2.2. La palma de aceite

La palma de aceite (*Elaeis guineensis* Jacq), es un cultivo oleaginoso de alto valor productivo y económico, su origen se atribuye al Golfo de Guinea y su producción comercial actualmente se encuentra en todas las regiones tropicales del mundo.

Sus productos tienen mucha versatilidad y tienen un alto valor en la industria de alimentos, química y energética, ya que sus tres productos principales los cuales son aceite rojo que se extrae del mesocarpio, aceite de la nuez o de palmiste que se extrae del endospermo o nuez y la torta o harina desgrasada de palmiste que es la parte sólida del endospermo o nuez que queda de residuo de la extracción del aceite de palmiste.

En comparación con otros cultivos oleaginosos (soya, canola, girasol, algodón, entre otros), la palma de aceite puede llegar a producir hasta 10 veces más aceite por unidad de área.

2.2.1. La importancia del cultivo

La importancia del cultivo de la palma de aceite radica en su característica como cultivo oleaginoso, el cual es el más productivo por unidad de área (Gremial de Palmicultores de Guatemala [GREPALMA], 2021), a continuación, se describen aspectos importantes sobre este cultivo.

2.2.1.1. A nivel mundial

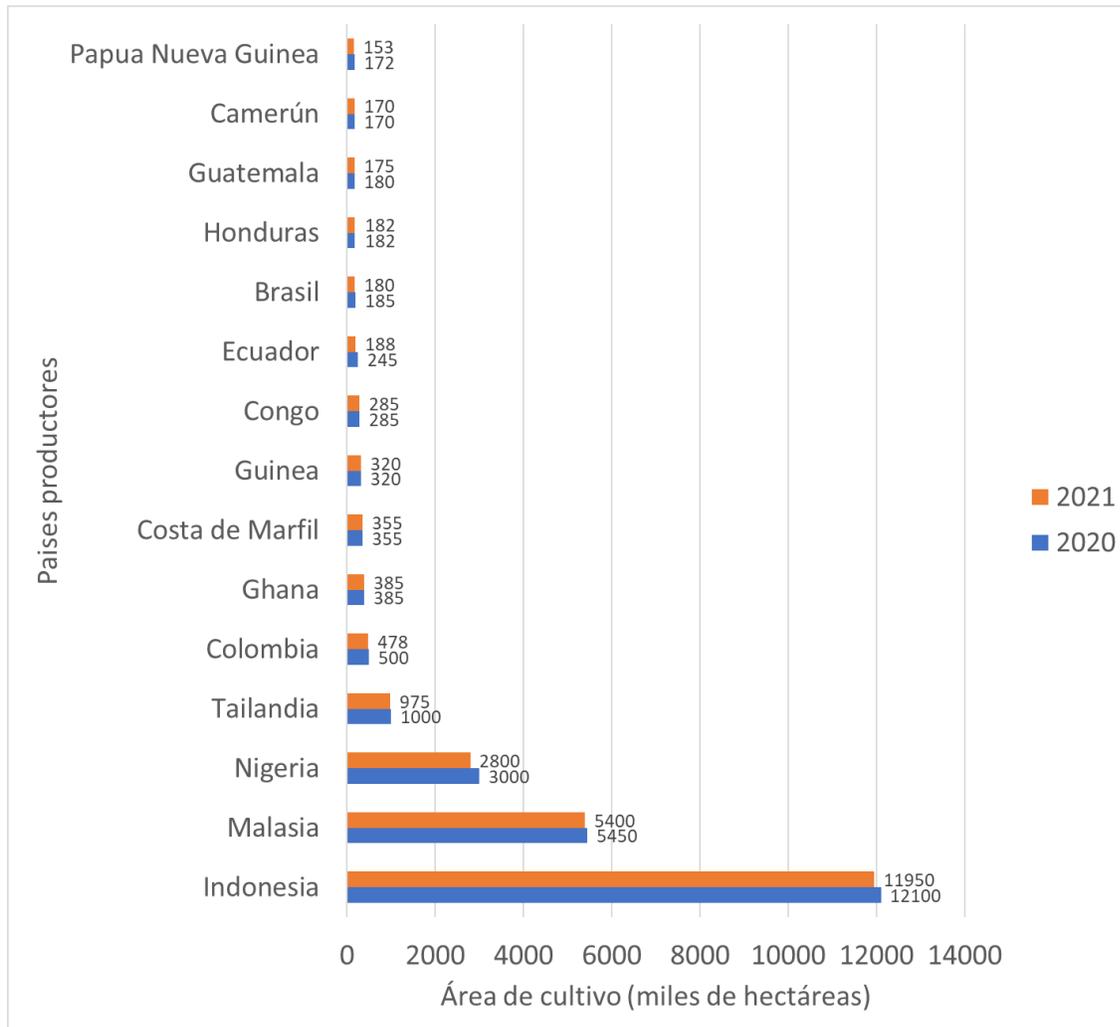
La Organización Mundial de Alimentos (2021) (como se cita en GREPALMA, 2021) indica que al año 2021 se contaban con 25,058,000 de hectáreas de cultivo, mostrando un 2.1 % de crecimiento con respecto al año 2020. El promedio mundial de productividad se estima en 18 t de RFF ha⁻¹.

Para el año 2021 la mayor proporción del cultivo a nivel mundial se encuentra en el continente asiático con el 74 % del total mundial del área. Los tres países con mayor área cultivada son: Indonesia con 12,100,000 hectáreas, Malasia con 5,450,000 hectáreas y Tailandia con 1,000,000 de hectáreas (GREPALMA, 2021).

En el continente americano los 5 países con mayor área sembrada son: “Colombia con 500,000 hectáreas (2 % en relación al área mundial cultivada); le sigue Ecuador con 245,000 hectáreas (0.98 %), Brasil con 185,000 hectáreas (0.74 %), Honduras con 182,000 (0.73 %) y Guatemala con 180,614 hectáreas” (GREPALMA, 2021, p. 18).

Figura 6.

Principales países productores de palma de aceite, año 2021



Nota. Descripción de los países con más área sembrada con cultivo de palma de aceite en el mundo al año 2021. Obtenido de Gremial de Palmicultores de Guatemala (2021). *Estadísticas socioeconómicas al año 2021.* (p. 18). GREPALMA.

2.2.1.2. En Guatemala

Según GREPALMA (2021), para el año 2021 Guatemala cuenta con 180,164 hectáreas de cultivo de palma de aceite, representando el 1.66 % del territorio nacional. Este cultivo se encuentra establecido en tres regiones: región Norte la cual representa el 57.78 % del total del área con este cultivo, región Sur la cual representa el 22.19 % del total del área con este cultivo y región Nororiente con el 20.03 %. Del total de área cultivada, el 96.45 % corresponde a palma en etapa productiva (> 5 años) con un promedio nacional de producción de 25.61 t de RFF ha⁻¹.

La producción de aceite rojo crudo de palma total en el país asciende a 941 mil toneladas, la cual aporta el 1.25 % del total de la producción mundial (GREPALMA, 2021).

2.2.1.3. En el departamento de Escuintla

En el departamento de Escuintla la palma de aceite ocupa el 2.61 % del área cultivable del departamento. Son 11,753.92 hectáreas distribuidas en 7 municipios del departamento (Ministerio de Agricultura y Ganadería, 2021).

Para el año 2021 la región sur produjo el 22.19 % del total de fruta fresca de palma de aceite en Guatemala (GREPALMA, 2021).

2.2.2. Aspectos botánicos

La palma de aceite (*Elaeis guineensis*), presenta un crecimiento a partir de un único punto de crecimiento o meristemo apical, dando origen a una sucesión continua de yemas foliares.

El desarrollo de las hojas es inicialmente lento, tomando de dos a tres años para abrirse en el centro de la corona.

- Caracterizada por un tallo único de tipo pleonántico, las inflorescencias aparecen en las axilas de las hojas a medida que la planta continúa su crecimiento vegetativo.
- El estípite, o tallo principal, es erecto y muestra cicatrices amplias dejadas por las bases peciolares de las hojas caídas.
- La especie produce inflorescencias femeninas y masculinas en la misma planta, con ciclos alternados de duración variable.
- La fructificación se inicia tras 4,5 a 6 meses desde la antesis de la inflorescencia femenina, resultando en racimos maduros que pueden alcanzar más de 50 cm de largo y 35 cm de ancho.
- El fruto, una drupa sésil, presenta variaciones en la apariencia externa durante su desarrollo, siendo clasificado según el grosor del cuesco en Dura (grosso), Tenera (delgado) y Piscifera (sin cuesco) (Forero et al., 2012).

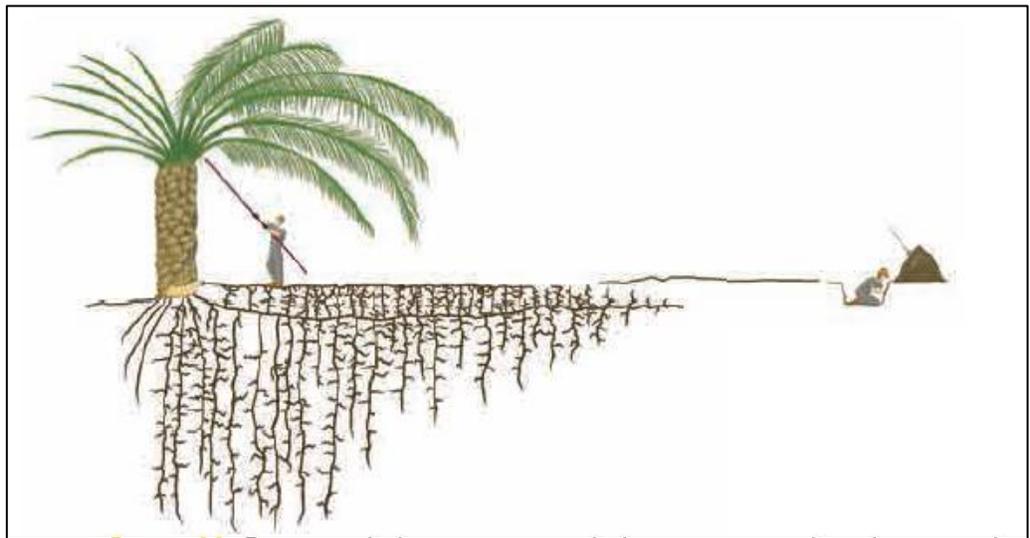
2.2.2.1. El sistema radicular

Las raíces forman un sistema fibroso y adventicio, muy extenso, que se origina desde la base del estipe y se extiende entre 0.4 y 0.5 m en el suelo (ver figura 2). La biomasa radicular se estima entre 30 a 40 t ha⁻¹ de materia seca, la cual se encuentra en su mayoría dentro de 1 m a partir de la superficie del suelo.

Se estima que los nutrientes y el agua se absorben en los ápices no lignificados del sistema radicular (Siew et al., 2012).

Figura 7.

Sistema radicular de palma de aceite



Nota. Descripción grafica del sistema radicular de una planta de palma de aceite. Obtenido de Forero et al. (2012). *Generalidades sobre la morfología y fenología de la palma de aceite.* (p. 32). Cenipalma.

2.2.2.2. El tallo

El tronco o estipe es erecto, en el cual se producen las hojas según el crecimiento de este, y en su parte inferior tiene forma es dilatada y es el punto de origen del sistema radicular (Forero et al., 2012).

Su crecimiento puede alcanzar los 30 metros, consta de tres capas: la corteza o capa externa, el periciclo el cual está dentro de la corteza y el cilindro

central o núcleo, que consiste en el sistema vascular de la planta (Siew et al., 2012).

Entre sus funciones se pueden enlistar:

- Sirve de sostén a las hojas
- Es un órgano transportador de agua y nutrientes a la planta a través de su sistema vascular.
- Sirve como almacén de nutrientes, mayormente Potasio (Forero et al., 2012).

2.2.2.3. Hojas

La palma de aceite tiene una tasa constante de producción de hojas de 20 a 25 por año, aunque puede variar según la edad. En condiciones normales una palma puede tener entre 30 a 49 hojas funcionales de entre 5 a 7 m de longitud, con un peso aproximado de 5 a 8 kg.

Las partes principales de una hoja de palma de aceite son:

- Pecíolo
- Raquis
- Foliolos

Su disposición alrededor del eje (estipe), da lugar a la formación de espirales.

Su función más importante es realizar la fotosíntesis, en el cual realiza funciones de intercambio de gases entre la planta y el ambiente. Tiene órganos que realizan dicho intercambio, a los cuales se les denomina estomas, estos están ubicados mayormente en la superficie inferior de la hoja (Forero et al., 2012).

2.2.2.4. Inflorescencias

Es una planta monoica, por lo que produce inflorescencias de ambos sexos en la misma planta.

El primordio o yema de una inflorescencia se produce en la axila de cada hoja, de la cual puede desarrollar una inflorescencia femenina o masculina, por esto la producción está directamente relacionada con la tasa de producción de hojas.

Se estima que la formación de las inflorescencias se inicia a partir de la cuarta hoja emitida por la planta y su periodo de madurez oscila entre los 30- 35 meses (Henry, 1955 como se cita en Forero et al., 2012).

2.2.2.4.1. Flor femenina

Esta llega a medir 30 centímetros previo a su apertura, sus flores están alrededor de las espiguillas colocadas en forma de espiral, cuentan con una bráctea superficial. El número varía de una a otra, encontrando menor cantidad en las jóvenes que en las adultas (10 a más de 15 respectivamente) (Corley & Tinker, 2009 como se cita en Corzo et al., 2021).

2.2.2.4.2. Flor masculina

El pedúnculo es más largo que la femenina, las espiguillas son cilíndricas, digitiformes, sin espinas, estas miden entre 10-20 centímetros. Durante la anthesis se esparce su polen, con una duración de 2-5 días (Corley & Tinker, 2009 como se cita en Corzo et al., 2021).

2.2.2.5. Frutos

Son tipo drupa sésil de forma esférica con tendencia a ovoide, sus frutos oscilan entre 2-7 centímetros de largo y consisten en un exocarpio delgado, mesocarpio, endocarpio duro (cuesco) y un endospermo (almendra). El aceite rojo está contenido en su mesocarpio (25 % en racimo) (Siew et al., 2012).

2.2.2.6. Semilla

Está formada por un cuesco duro y de una almendra, aunque en algunos casos puede tener dos o hasta tres almendras. Cada nuez cuenta con tres poros germinales que corresponden a tres partes del ovario tricarpelar, la almendra está dentro de la nuez y consiste en capas de endospermo aceitoso, de color blanco grisáceo rodeado por un tegumento pardo oscuro cubierto con una red de fibras (Siew et al., 2012).

2.2.3. Factores climáticos que influyen en la formación de frutos

Según Corzo et al. (2020), la producción de racimos de fruta fresca se puede ver afectada por algunas condiciones climáticas, entre las cuales se pueden mencionar:

- Sequías
- Inundaciones
- Temperaturas altas
- Temperaturas bajas

Estas condiciones afectaran el número de inflorescencias femeninas siendo la causa principal de la baja de producción.

Dentro de las principales variables climática que se consideran afectan al rendimiento son:

- la precipitación
- la humedad relativa
- la radiación solar
- los vientos (Turne & Gillbanks, 2003, como se cita en Corzo et al., 2020).

3. PRESENTACIÓN DE RESULTADOS

De acuerdo con los objetivos establecidos se presentan los resultados obtenidos en esta investigación. Este apartado está formado por tablas y las figuras realizadas con los datos de interés y siguiendo la metodología y los objetivos planteados.

3.1. Objetivo 1. Estimar el grado de asociación y dependencia estadística entre la variable producción con respecto a las variables climáticas entre los años 2010 y 2020 a través del coeficiente de correlación

Para el desarrollo de este objetivo, se realizó un análisis de correlación de Pearson para la variable de producción y cada una de las variables climáticas en estudio. Como primer paso se realizó un análisis exploratorio de los datos.

3.1.1. Análisis exploratorio

A continuación, se presentan resultados descriptivos que ayudan a la comprensión del comportamiento de las variables en estudio.

3.1.1.1. Análisis descriptivo

En este apartado se presentan los resultados para la media, mediana, desviación estándar y Rango.

Tabla 2.*Análisis descriptivo de las variables en estudio*

Variable	n	Media	Mediana	Desviación estándar	Rango
Producción de racimos de fruta fresca mensual	132	2537.07	2456.00	967.03	4318.00
Temperatura mínima mensual	132	21.99	22.36	1.31	6.25
Temperatura máxima mensual	132	33.84	33.84	0.89	5.00
Temperatura promedio mensual	132	27.07	27.17	0.82	4.09
Radiación promedio mensual	132	651484.57	648484.35	55167.12	279680.30
Humedad relativa promedio mensual	132	87.07	87.56	3.49	16.76
Precipitación mensual	132	118	73.30	148.20	775.00

Nota. En la tabla se muestran las medidas estadísticas descriptivas de las variables en estudio. Elaboración propia, realizado con Infostat.

En la tabla 2 se observa el resumen descriptivo de 7 variables con 132 observaciones para cada una de ellas, las cuales son los registros mensuales desde enero 2010 hasta diciembre 2020.

3.1.1.2. Pruebas de normalidad

A continuación, se presentan los resultados de la prueba de normalidad Shapiro-Wilk para las variables en estudio.

Tabla 3.

Resultados de pruebas de normalidad de las variables en estudio

Variable	Shapiro-Wilk	
	Estadístico	Sig.
1 Producción de racimos de fruta fresca mensual	0.95	<0.0001
2 Temperatura mínima mensual	0.98	0.1827
3 Temperatura máxima mensual	0.93	<0.0001
4 Temperatura promedio mensual	0.97	0.0491
5 Radiación promedio mensual	0.77	<0.0001
6 Humedad relativa promedio mensual	0.98	0.4401
7 Precipitación mensual	0.95	0.0010

Nota. Resultados de la prueba estadística de normalidad Shapiro-Wilk. Elaboración propia, realizado con Infostat.

En la tabla 3 se observa que las variables producción de racimos de fruta fresca mensual, temperatura máxima mensual, radiación promedio mensual, y precipitación mensual no siguen una distribución normal ya que se encuentran por debajo 5 % del nivel de significancia según la prueba de Shapiro-Wilk realizada en el programa Infostat.

3.1.1.3. Análisis correlacional

En este apartado se presenta el grado de correlación de las variables en estudio mediante el método de Pearson.

Figura 8.

Matriz de correlación de las variables en estudio

Matriz de correlación, coeficiente de Pearson

PROD	***	-0.12	-0.24*	0.08	-0.33**	-0.03	0.01
T_MED	-0.12	***	0.84***	0.69***	0.15	0.31**	0.31**
T_MIN	-0.24**	0.84***	***	0.28*	0.51***	-0.03	0.65***
T_MAX	0.08	0.69***	0.28**	***	-0.37***	0.48***	-0.03
PREC	-0.33***	0.15	0.51***	-0.37***	***	-0.40***	0.54***
RADI	-0.03	0.31***	-0.03	0.48***	-0.40***	***	-0.42***
H_REL	0.01	0.31***	0.65***	-0.03	0.54***	-0.42***	***
	PROD	T_MED	T_MIN	T_MAX	PREC	RADI	H_REL

Nota. La matriz muestra el grado de correlación entre todas las variables en estudio. Elaboración propia, realizado con Rstudio.

Como se puede observar en la figura 8, el grado de correlación de la variable producción de racimos de fruta fresca mensual es bastante bajo, teniendo el nivel más alto de correlación negativa con la precipitación mensual (-0.33) y temperatura mínima promedio mensual (-0.24).

- Para el caso de la temperatura media mensual tiene fuerte correlación con la temperatura máxima promedio mensual (0.69), temperatura mínima promedio mensual (0.84) y una correlación media con la radiación promedio mensual (0.31) y la humedad relativa promedio mensual (0.31).
- Para el caso de la temperatura mínima promedio mensual se puede observar una correlación fuerte con la humedad relativa promedio mensual (0.65) y con la precipitación mensual (0.51).
- Para el caso de la temperatura máxima promedio mensual se puede observar una correlación fuerte con temperatura media mensual (0.69), una correlación media con radiación promedio mensual (0.48), temperatura mínima promedio mensual (0.28) y precipitación mensual (-0.37).
- Para el caso de la precipitación promedio mensual se puede observar una correlación media con humedad relativa promedio mensual (0.54) y radiación promedio mensual (-0.40).
- Y para el caso de la radiación promedio mensual tiene una correlación media con la humedad relativa promedio mensual (-0.42).

3.2. Objetivo 2. Construir un modelo estadístico de predicción de producción utilizando las variables ambientales mediante regresión lineal múltiple

En este apartado se construyó un modelo de predicción de la producción mediante regresión lineal múltiple evaluando los supuestos y los criterios de ajuste para elegir el mejor modelo y así poder cumplir con el objetivo número 2.

3.2.1. Modelo 1

A continuación, se presenta el primer modelo de regresión lineal múltiple, este modelo se creó utilizando la variable producción de racimos de fruta fresca mensual como variable dependiente y las seis variables climáticas como variables independientes.

Tabla 4.

Análisis de regresión múltiple del modelo 1

Variable	N	R ²	R ² Aj	ECMP	AIC	BIC
Producción	132	0.320	0.290	735118.830	2153.570	2176.630

Nota. Se presentan los resultados de la regresión lineal múltiple del modelo 1, el cual contiene la variable producción de racimos de fruta fresca utilizando como variables regresores seis variables climáticas. Elaboración propia, realizado con Infostat.

En la tabla 4 se puede observar los resultados del análisis de la regresión lineal múltiple, el modelo tiene un R² de 0.32 y un R² ajustado de 0.29. Esto indica que el modelo explica un 32 % de la variabilidad del modelo, el criterio de

información de Akaike de 2153.70 y el criterio de información Bayesiano de 2176.63.

Tabla 5.

Coefficientes de regresión del modelo 1

Coef	Est.	E.E.	LI (95 por ciento)	LS (95 por ciento)	T	p-valor	CpMallows	VIF
(Intercepto)	-	4505.90	-	-	-	0.0010		
Temperatura media mensual	15178.70	467.30	24096.50	6260.90	3.40	<0.0001	23.41	28.99
Temperatura mínima mensual	2004.91	268.06	1080.06	2929.75	4.29	<0.0001	29.54	24.25
Temperatura máxima mensual	-1327.98	216.77	-1858.51	-797.44	-	<0.0001	16.43	7.34
Precipitación mensual	-732.79	0.73	-4.01	-1.11	-	0.0006	17.29	2.30
Radiación promedio mensual	-2.56	0.00	-0.01	0.00	-	0.2419	6.38	1.79
Humedad relativa promedio mensual	0.00	39.70	140.46	297.59	1.18	<0.0001	35.45	3.77

Nota. Se muestra los coeficientes de regresión del modelo 1, en los cuales se identifican los significativos. Elaboración propia, realizado con Infostat.

En la tabla 5 se observan los coeficientes de la regresión de las variables incluidas en el modelo. Según el valor-p de la prueba indica que no existe significancia para la variable radiación promedio mensual (0.24), por lo que se indica que no existe relación entre esta variable y la producción. También se observa que el criterio Cp de Mallows el valor más bajo coincide en la variable radiación promedio mensual (6.38). Observando la variable temperatura media mensual tiene el mayor (28.99) factor de inflación de la varianza (VIF), esto indica que esta variable tiene multicolinealidad con una o más variables en el modelo. Este modelo cuenta con un Error porcentual absoluto medio (MAPE por sus siglas en inglés) = 29.61 %.

Se realizó la prueba de Breusch-Pagan para probar homocedasticidad en los residuos del modelo, el valor del estadístico es de 4.433 y el valor-p de la prueba es 0.6183 el cual no indica significancia, por lo que existe suficiente evidencia estadística para aceptar la H_0 = la varianza de los residuos es constante, por lo que se indica que hay homocedasticidad.

También se realizó la prueba de Shapiro-Wilk para probar normalidad en los residuos, el valor del estadístico es de $W=0.978$ con un valor-p = 0.0294, el cual indica significancia, por lo que existe suficiente evidencia estadística para rechazar H_0 y aceptar H_a = la muestra no sigue una distribución normal.

Se realizó la prueba de Durbin-Watson para probar independencia en los residuos, el valor del estadístico es de $DW= 1.1531$ con un valor-p = <0.00001 el cual indica significación, por lo que existe suficiente evidencia estadística para rechazar la H_0 y aceptar H_a = existe autocorrelación de primer orden en los residuos.

3.2.2. Modelo 2

A continuación, se presenta el modelo 2, creado con regresión lineal múltiple, este modelo contiene la variable producción de racimos de fruta mensual como variable dependiente y cuatro variables climáticas como variables independientes. En este modelo no se incluyeron las variables temperatura media mensual y radiación promedio mensual, ya que se tomó en consideración la significancia y los criterios C_p de Mallows y VIF.

Tabla 6.*Análisis de regresión múltiple del modelo 2*

Variable	N	R ²	R ² Aj	ECMP	AIC	BIC
Producción	132	0.220	0.190	814080.200	2167.860	2185.160

Nota. Se presentan los resultados de la regresión lineal múltiple del modelo 2, el cual contiene la variable producción de racimos de fruta fresca utilizando como variables regresores cuatro variables climáticas. Elaboración propia, realizado con Infostat.

En la tabla 6 se puede observar los resultados del análisis de la regresión lineal múltiple, el modelo tiene un R² de 0.22 y un R² ajustado de 0.19, estos coeficientes de ajuste aumentaron con respecto al modelo 1, debido a la reducción de 2 variables. También se puede observar que el AIC y BIC aumentaron lo cual es un criterio para indicar que el modelo bajo de ajuste.

Tabla 7.*Coefficientes de regresión del modelo 2*

Coef	Est.	E.E.	LI (95 por ciento)	LS (95 por ciento)	T	P- valor	CpMallows	VIF
(Intercepto)	-	3756.04	-	3640.49	-	0.31		
Temperatura mínima mensual	3792.02	94.82	11224.54	-67.13	1.01	0.01	10.22	2.68
Temperatura máxima mensual	-254.77	113.16	-442.40	270.70	2.69	0.68	3.17	1.77
Precipitación mensual	46.78	0.77	-177.15	-0.95	0.41	0.00	13.34	2.27
Humedad relativa promedio mensual	-2.48	30.33	62.20	182.23	3.22	0.00	19.24	1.95

Nota. Se muestra los coeficientes de regresión del modelo 2, en los cuales se identifican los significativos. Elaboración propia, realizado con Infostat.

En la tabla 7 se observan los coeficientes de la regresión de las variables incluidas en el modelo. Según el valor-p de las pruebas indica que la variable temperatura máxima mensual no tiene significancia (0.68), por lo que no existe suficiente evidencia estadística para aceptar H_0 y se puede indicar que dicha variable no tiene relación con la variable dependiente (producción de racimos de fruta fresca mensual).

Se puede observar que los coeficientes VIF se encuentran por debajo de 5, lo cual indica que no existe colinealidad entre las variables independientes en el modelo, y el coeficiente C_p Mallows para la variable temperatura máxima mensual se encuentra por debajo de los demás (3.17). Este modelo cuenta con un Error porcentual absoluto medio (MAPE por sus siglas en inglés) = 32.48 %.

Se realizó la prueba de Breusch-Pagan para probar homocedasticidad en los residuos del modelo, el valor del estadístico es de 2.362 y el valor-p de la prueba es 0.6695 el cual no indica significancia, por lo que existe suficiente evidencia estadística para aceptar la H_0 = la varianza de los residuos es constante, por lo que se indica que hay homocedasticidad.

También se realizó la prueba de Shapiro-Wilk para probar normalidad en los residuos, el valor del estadístico es de $W=0.9766$ con un valor-p = 0.0222, el cual indica significancia, por lo que existe suficiente evidencia estadística para rechazar H_0 y aceptar H_a = la muestra no sigue una distribución normal.

Se realizó la prueba de Durbin-Watson para probar independencia en los residuos, el valor del estadístico es de $DW= 0.9412$ con un valor-p = <0.00001 el cual indica significación, por lo que existe suficiente evidencia estadística para rechazar la H_0 y aceptar H_a = existe autocorrelación de primer orden en los residuos.

3.2.3. Modelo 3

A continuación, se presenta el modelo 3, creado con regresión lineal múltiple, este modelo contiene la variable producción de racimos de fruta mensual como variable dependiente y tres variables climáticas como variables independientes. En este modelo no se incluyó la variable temperatura máxima mensual (como en el modelo 2), ya que se tomó en consideración la significancia y el criterio Cp de Mallows para descartarla del modelo.

Tabla 8.

Análisis de regresión múltiple del modelo 3

Variable	N	R ²	R ² Aj	ECMP	AIC	BIC
Producción	132	0.220	0.200	810016.340	2166.040	2180.460

Nota. Se presentan los resultados de la regresión lineal múltiple del modelo 3, el cual contiene la variable producción de racimos de fruta fresca utilizando como variables regresores tres variables climáticas. Elaboración propia, realizado con Infostat.

En la tabla 8 se puede observar los resultados del análisis de la regresión lineal múltiple, el modelo tiene un R² de 0.22 y un R² ajustado de 0.20, se presenta una leve mejora con respecto al modelo 2, pero queda por debajo de la cantidad de variación que explica el modelo 1. Se puede observar que el AIC disminuyó con respecto del modelo 2, pero sigue por arriba del modelo 1, al igual que el BIC que disminuyó con respecto al modelo 2, pero estando por arriba del modelo 1 (2180.46).

Tabla 9.*Coefficientes de regresión del modelo 3*

Coef	Est.	E.E.	LI (95 por ciento)	LS (95 por ciento)	T	p-valor	CpMallows	VIF
(Intercepto)	-2546.55	2235.51	-6969.89	1876.79	-1.14	0.26		
Temperatura mínima mensual	-232.83	78.33	-387.83	-77.84	-2.97	0.00	10.83	1.84
Precipitación mensual	-2.67	0.63	-3.91	-1.43	-4.25	<0.0001	20.07	1.51
Humedad relativa promedio mensual	120.80	30.04	61.37	180.24	4.02	0.00	18.17	1.92

Nota. Se muestra los coeficientes de regresión del modelo 3, en los cuales se identifican los significativos. Elaboración propia, realizado con Infostat.

En la tabla 9 se observan los coeficientes de la regresión de las variables incluidas en el modelo 3. Para este modelo el valor-p de las pruebas indica que las tres variables incluidas como independientes tienen significancia por encontrarse por debajo de 0.05, por lo que existe suficiente evidencia estadística para aceptar H_0 y se puede indicar que todas las variables tienen relación con la variable dependiente (producción de racimos de fruta fresca mensual).

Se puede observar que los coeficientes VIF se encuentran por debajo de 5, lo cual indica que no existe colinealidad entre las variables independientes en el modelo, y el coeficiente Cp Mallows para todas las variables se encuentra por arriba de 10. Este modelo cuenta con un Error porcentual absoluto medio (MAPE por sus siglas en inglés) = 32.59 %.

Se realizó la prueba de Breusch-Pagan para probar homocedasticidad en los residuos del modelo, el valor del estadístico es de 2.3391 y el valor-p de la prueba es 0.5051 el cual no indica significancia, por lo que existe suficiente

evidencia estadística para aceptar la H_0 = la varianza de los residuos es constante, por lo que se indica que hay homocedasticidad.

También se realizó la prueba de Shapiro-Wilk para probar normalidad en los residuos, el valor del estadístico es de $W=0.9764$ con un valor- $p = 0.02147$ el cual indica significancia, por lo que existe suficiente evidencia estadística para rechazar H_0 y aceptar $H_a =$ la muestra no sigue una distribución normal.

Se realizó la prueba de Durbin-Watson para probar independencia en los residuos, el valor del estadístico es de $DW= 0.93542$ con un valor- $p = <0.00001$ el cual indica significación, por lo que existe suficiente evidencia estadística para rechazar la H_0 y aceptar $H_a =$ existe autocorrelación de primer orden en los residuos.

3.2.4. Comparación de modelos de regresión múltiple

A continuación, se presenta un resumen de los indicadores más relevantes de los resultados de los tres modelos de regresión múltiple desarrollados para pronosticar la producción de racimos de fruta fresca mensual.

Tabla 10.*Comparación de modelos de regresión lineal múltiple*

Criterio	Modelo 1	Modelo 2	Modelo 3	Mejor modelo
R ²	0.32	0.22	0.22	1
R ² ajustado	0.29	0.19	0.20	1
AIC	2153.57	2167.86	2166.04	1
BIC	2176.63	2185.16	2180.46	1
Rango de valores Cp Mallows	(6.40 – 35.40)	(3.17 – 19.24)	(10.83 – 20.07)	3
Rango de valores VIF	(1.80 – 29.00)	(1.77 – 2.68)	(1.51 – 1.92)	3
Número de variables independientes	6	4	3	3

Nota. En esta tabla se muestra el resumen de los coeficientes más importantes de los tres modelos de regresión. Elaboración propia, realizado con Infostat.

En la tabla 10 se puede observar que los mejores modelos son el 1 y el 3, sin embargo, por la simplicidad del modelo (tres variables independientes) se puede trabajar con el modelo 3.

3.3. Objetivo 3. Construir un modelo estadístico que se ajuste aceptablemente al comportamiento de la producción de estudio mediante un análisis de series temporales para obtener predicciones aceptables

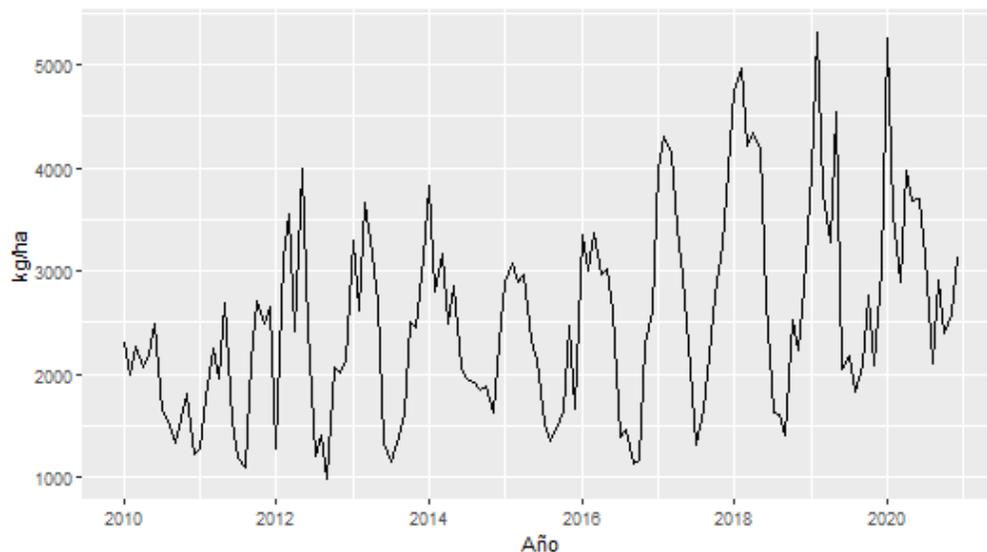
A continuación, se realizó la exploración, caracterización, construcción y diagnóstico de un modelo de pronóstico basado en series de tiempo para el cumplimiento del tercer objetivo planteado en este trabajo de investigación.

3.3.1. Análisis exploratorio de datos

Como se observa en la figura 9, durante el periodo del 2010 al 2020, la serie de tiempo de la producción mensual kg/ha muestra un comportamiento cíclico y una tendencia a muy leve al alza.

Figura 9.

Serie temporal de la producción mensual kg/ha, periodo 2010-2020

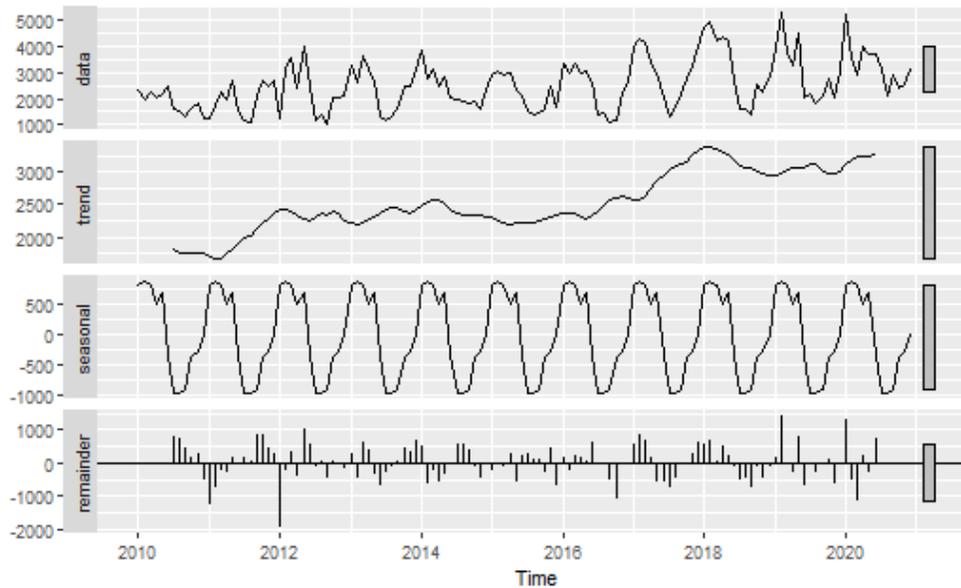


Nota. En la gráfica se observa el comportamiento de la producción mensual a través del periodo de 2010 al 2020. Elaboración propia, realizado con Rstudio.

Se realizó una descomposición de la serie temporal para la visualización de los componentes de tendencia y hacer un diagnóstico gráfico de la misma.

Figura 10.

Descomposición de la serie temporal de producción, periodo del 2010 al 2020

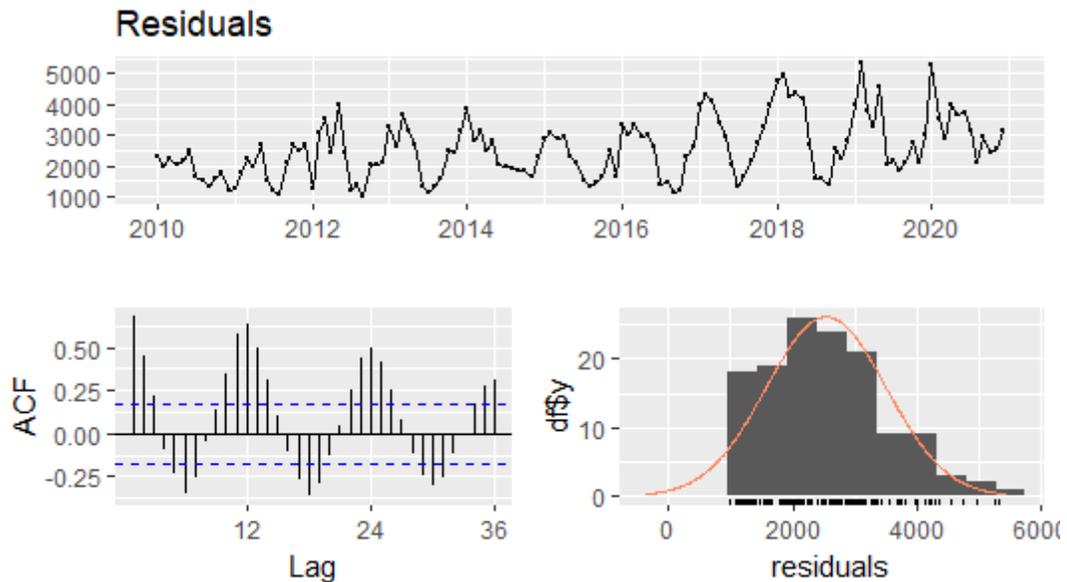


Nota. Se la gráfica se puede observar tendencia y estacionalidad. Elaboración propia, realizado con Rstudio.

En la figura 10 se puede observar que la serie temporal tiene tendencia al alza, la serie es no estacionaria. La prueba de Dickey-Fuller Aumentada (ADF, por sus siglas en inglés) da como resultado para el estadístico -8.7662 , con un valor-p de 0.01, esto indica significancia y se tiene suficiente evidencia estadística para poder rechazar la H_0 y aceptar la H_1 = que la serie de tiempo es estacionaria y que no tiene una raíz unitaria.

Figura 11.

Análisis de residuos de la serie de tiempo original



Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 11 se puede observar que los residuos siguen la distribución normal, y la función de autocorrelación muestra un desvanecimiento lento lo que puede dar indicios de tendencia estocástica, además muestra significancia en muchos valores, pero siendo el mayor el 12, lo que podría ser una estacionalidad anual, esto confirma el resultado de la prueba de Dickey-Fuller la cual indicó que la serie de tiempo es estacionaria y que no tiene una raíz unitaria.

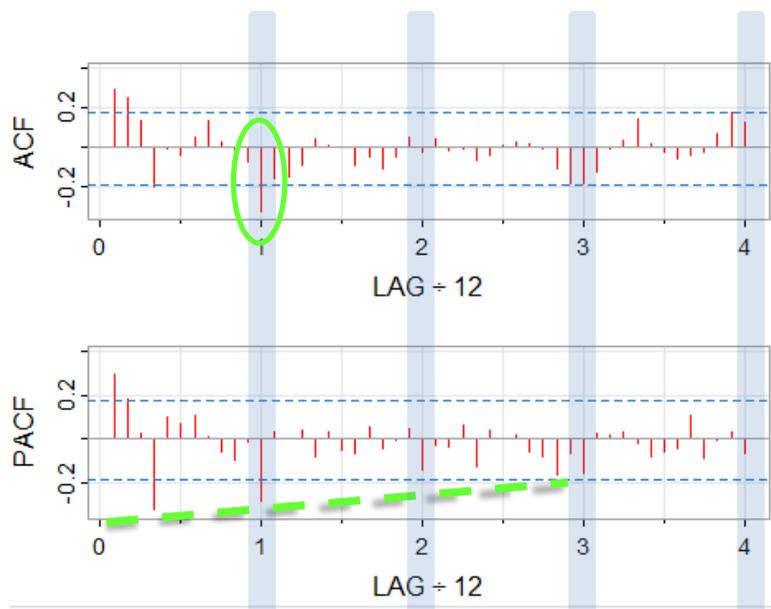
3.3.2. Selección del modelo ARIMA

A pesar de que la prueba de Dickey Fuller mostró significancia y se concluyó que la serie es estacionaria, se pudo observar en la figura 9 la serie

temporal de la producción mensual cuenta con tendencia o es no estacionaria, por lo que se utilizará una diferencia estacional para iniciar con el diagnóstico y selección del mejor modelo para el pronóstico buscado en el objetivo 3.

Figura 12.

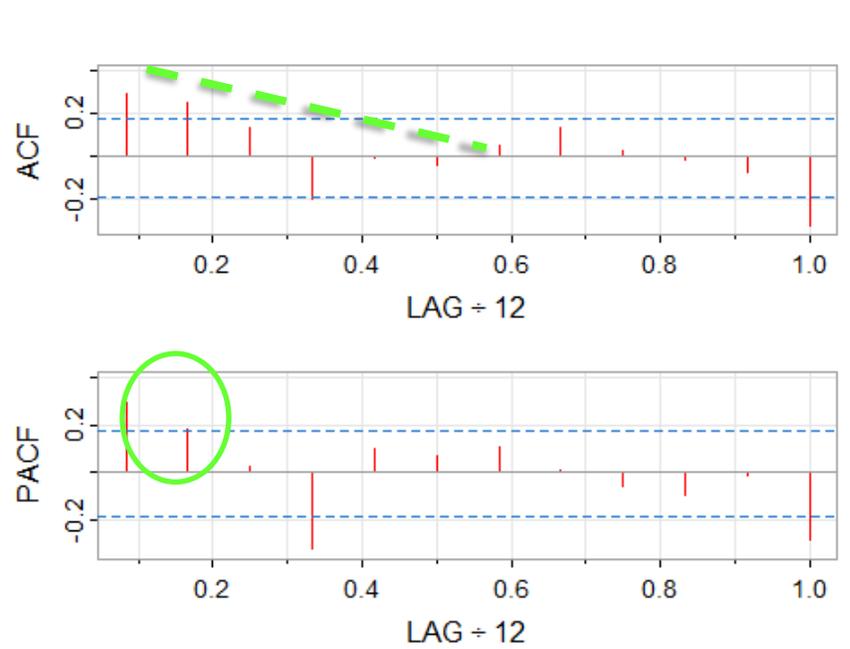
Gráfico de autocorrelación y autocorrelación parcial de la parte estacional



Nota. En las gráficas de autocorrelación y autocorrelación parcial se presenta el análisis de la parte estacional del modelo. Elaboración propia, realizado con Rstudio.

Figura 13.

Gráfico de autocorrelación y autocorrelación parcial de la parte no estacional



Nota. En las gráficas de autocorrelación y autocorrelación parcial se presenta el análisis de la parte no estacional del modelo. Elaboración propia, realizado con Rstudio.

Según lo observado en las figuras 12 y 13, el modelo que mejor se puede ajustar a la serie temporal es el $ARIMA(2,0,0)(0,1,1)[12]$. Por lo que se realizó diagnóstico a este modelo y a otros 3 más. En el segundo modelo se le agregó un coeficiente autorregresivo(p), siendo el modelo $ARIMA(3,0,0)(0,1,1)[12]$. En el tercer modelo se le agregó otro coeficiente autorregresivo (p), siendo el modelo $ARIMA(4,0,0)(0,1,1)[12]$. Y al cuarto modelo evaluado se le agregó un coeficiente autorregresivo estacional (P), siendo el modelo $ARIMA(4,0,0)(1,1,1)[12]$.

3.3.2.1. Diagnóstico del modelo 1, ARIMA(2,0,0)(0,1,1)[12]

A continuación, se realizaron pruebas de diagnóstico al modelo para poder comprobar que se ajuste correctamente a los datos.

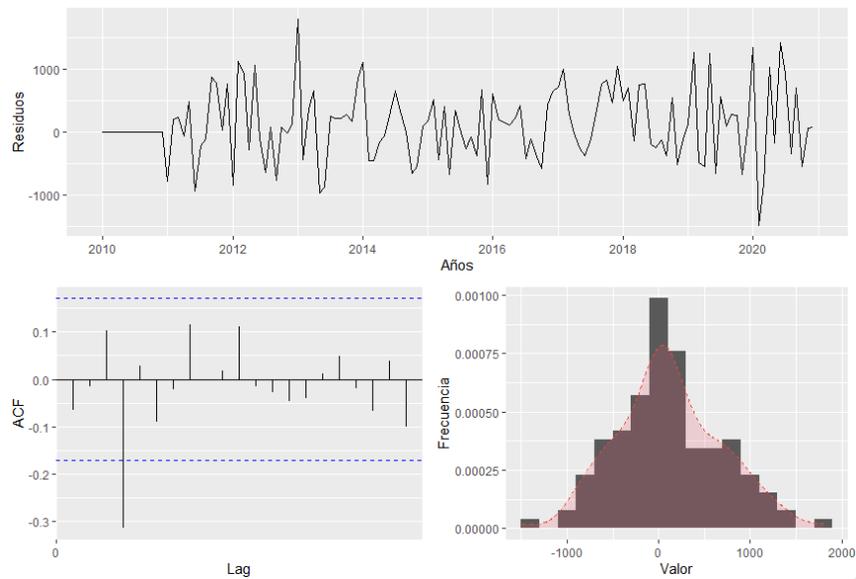
El coeficiente de ajuste AIC del modelo es = 1894.18. Las medidas de error del modelo son las siguientes: Error medio (ME por sus siglas en inglés) = 113.19, Error cuadrático medio (RMSE por sus siglas en inglés) = 584.13, Error absoluto medio (MAE por sus siglas en inglés) = 446.00, Error porcentual medio (MPE por sus siglas en inglés) = 0.06, Error absoluto porcentual medio (MAPE por sus siglas en inglés) = 18.25 y Error medio absoluto escalado medio (MASE por sus siglas en inglés) = 0.74.

3.3.2.1.1. Análisis de residuos

Se realizaron pruebas a los residuos para comprobar que estos tengan cumplan con los supuestos del modelo.

Figura 14.

Análisis de residuos del modelo ARIMA(2,0,0)(0,1,1)[12]

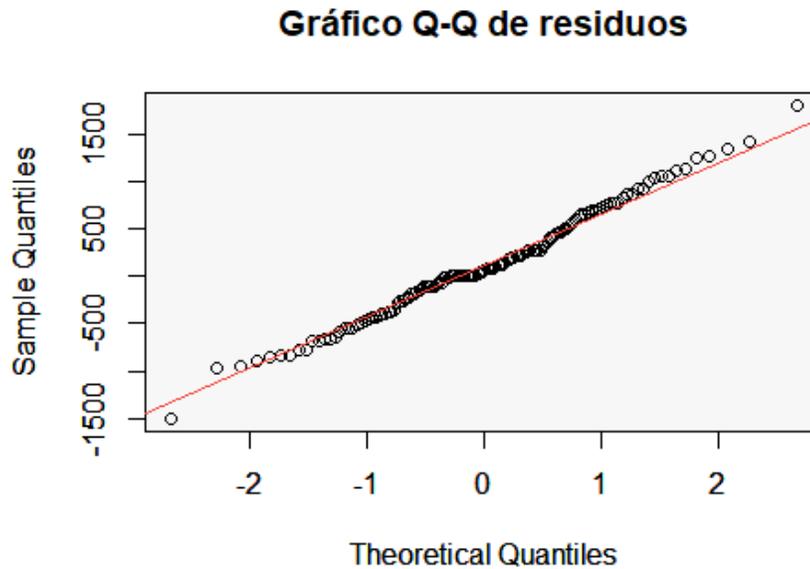


Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 14 se puede observar que los residuos del modelo ARIMA(1,0,0)(0,1,1)[12] tienen media cero, no existe autocorrelación y siguen la distribución normal.

Figura 15.

Gráfico de Q-Q de los residuos del modelo ARIMA(2,0,0)(0,1,1)[12]

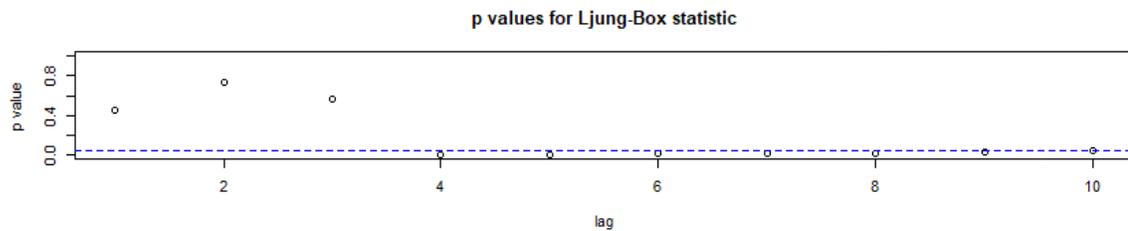


Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 15 se puede observar que los residuos se apegan a la línea diagonal por lo que se confirma que siguen una distribución normal, se confirma la normalidad de los residuos.

Figura 16.

Gráfico de valores-p para la prueba Ljung-Box



Nota. En la figura se puede evidenciar que no todos los valores se encuentran fuera de la zona de rechazo. Elaboración propia, realizado con Rstudio.

En la figura 16 se puede observar que no todos los valores-p se encuentran fuera de la zona de rechazo, por lo que se rechaza la H_0 = los residuos son independientes, se acepta la H_a = los residuos no son independientes.

3.3.2.2. Diagnóstico del modelo 2, ARIMA(3,0,0)(0,1,1)[12]

A continuación, se realizaron pruebas de diagnóstico al modelo para poder comprobar que se ajuste correctamente a los datos.

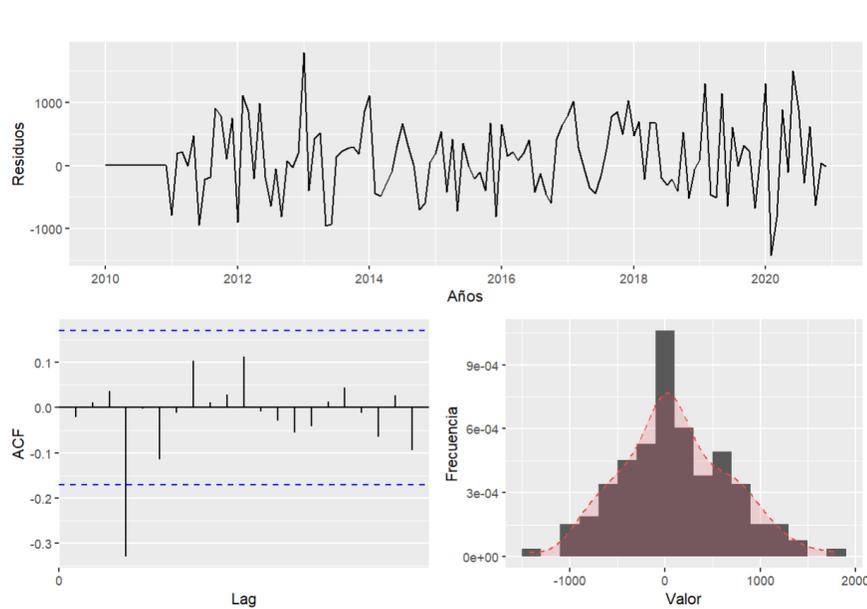
El coeficiente de ajuste AIC del modelo es = 1895.39. Las medidas de error del modelo son las siguientes: Error medio (ME por sus siglas en inglés) = 105.24, Error cuadrático medio (RMSE por sus siglas en inglés) = 581.73, Error absoluto medio (MAE por sus siglas en inglés) = 446.19, Error porcentual medio (MPE por sus siglas en inglés) = -0.39, Error absoluto porcentual medio (MAPE por sus siglas en inglés) = 18.42 y Error medio absoluto escalado medio (MASE por sus siglas en inglés) = 0.74.

3.3.2.2.1. Análisis de residuos

Se realizaron pruebas a los residuos para comprobar que estos tengan cumplan con los supuestos del modelo.

Figura 17.

Análisis de residuos del modelo ARIMA(3,0,0)(0,1,1)[12]

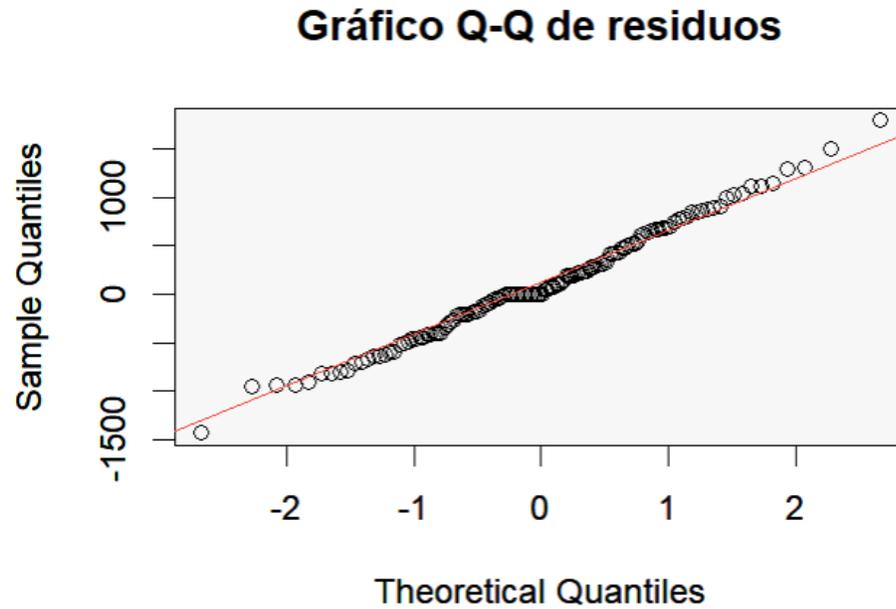


Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 17 se puede observar que los residuos del modelo ARIMA(3,0,0)(0,1,1)[12] tienen media cero, no existe autocorrelación y siguen la distribución normal.

Figura 18.

Gráfico de Q-Q de los residuos del modelo ARIMA(3,0,0)(0,1,1)[12]

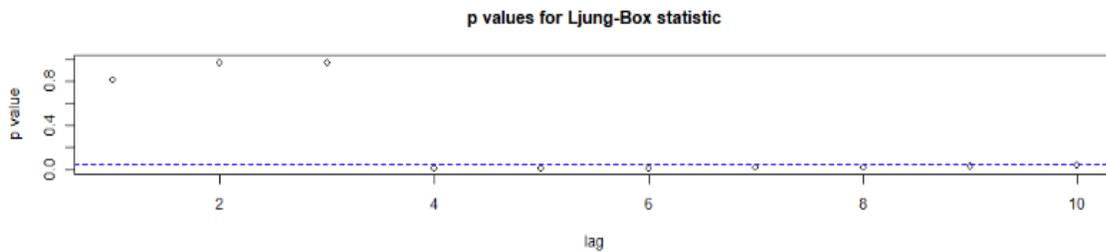


Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 18 se puede observar que los residuos se apegan a la línea diagonal por lo que se confirma que siguen una distribución normal.

Figura 19.

Gráfico de valores-p para la prueba Ljung-Box



Nota. En la figura se puede evidenciar que no todos los valores se encuentran fuera de la zona de rechazo. Elaboración propia, realizado con Rstudio.

En la figura 19 se puede observar que no todos los valores-p se encuentran fuera de la zona de rechazo, por lo que se rechaza la H_0 = los residuos son independientes, se acepta la H_a = los residuos no son independientes.

3.3.2.3. Diagnóstico del modelo 3, ARIMA(4,0,0)(0,1,1)[12]

A continuación, se realizaron pruebas de diagnóstico al modelo para poder comprobar que se ajuste correctamente a los datos.

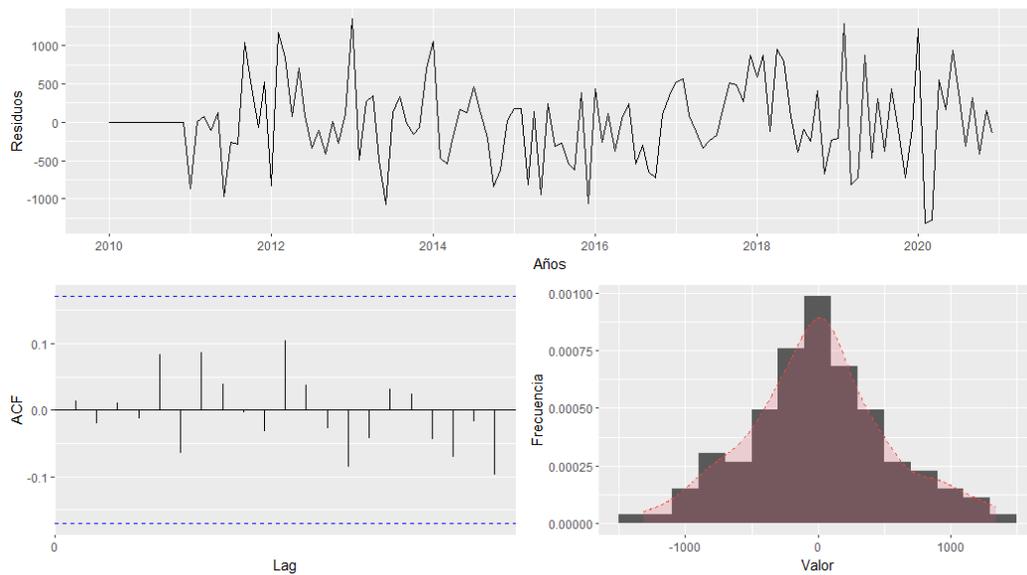
El coeficiente de ajuste AIC del modelo es = 1888.93. Las medidas de error del modelo son las siguientes: Error medio (ME por sus siglas en inglés) = 124.83, Error cuadrático medio (RMSE por sus siglas en inglés) = 563.67, Error absoluto medio (MAE por sus siglas en inglés) = 429.58, Error porcentual medio (MPE por sus siglas en inglés) = 1.25, Error absoluto porcentual medio (MAPE por sus siglas en inglés) = 17.59 y Error medio absoluto escalado medio (MASE por sus siglas en inglés) = 0.71.

3.3.2.3.1. Análisis de residuos

Se realizaron pruebas a los residuos para comprobar que estos cumplan con los supuestos del modelo.

Figura 20.

Análisis de residuos del modelo ARIMA(4,0,0)(0,1,1)[12]

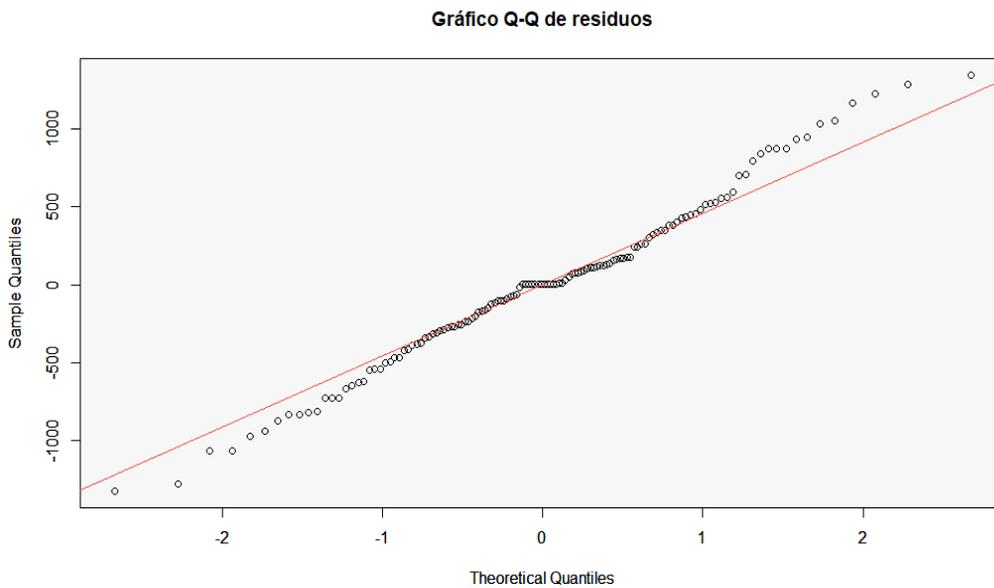


Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 20 se puede observar que los residuos del modelo ARIMA(4,0,0)(0,1,1)[12] tienen media cero, no existe autocorrelación y siguen la distribución normal.

Figura 21.

Gráfico de Q-Q de los residuos del modelo ARIMA(4,0,0)(0,1,1)[12]

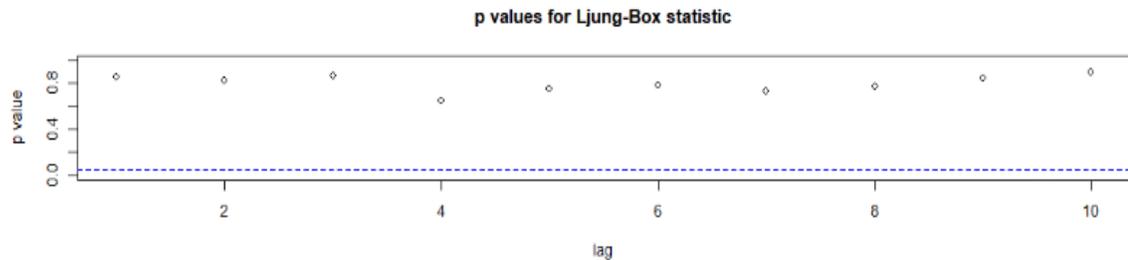


Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 21 se puede observar que los residuos se apegan a la línea diagonal por lo que se confirma que siguen una distribución normal.

Figura 22.

Gráfico de valores-p para la prueba Ljung-Box



Nota. En la figura se puede evidenciar que todos los valores se encuentran fuera de la zona de rechazo. Elaboración propia, realizado con Rstudio.

En la figura 22 se puede observar que todos los valores-p se encuentran fuera de la zona de rechazo, por lo que se acepta la H_0 = los residuos son independientes.

3.3.2.4. Diagnóstico del modelo 4, ARIMA(4,0,0)(1,1,1)[12]

A continuación, se realizaron pruebas de diagnóstico al modelo para poder comprobar que se ajuste correctamente a los datos.

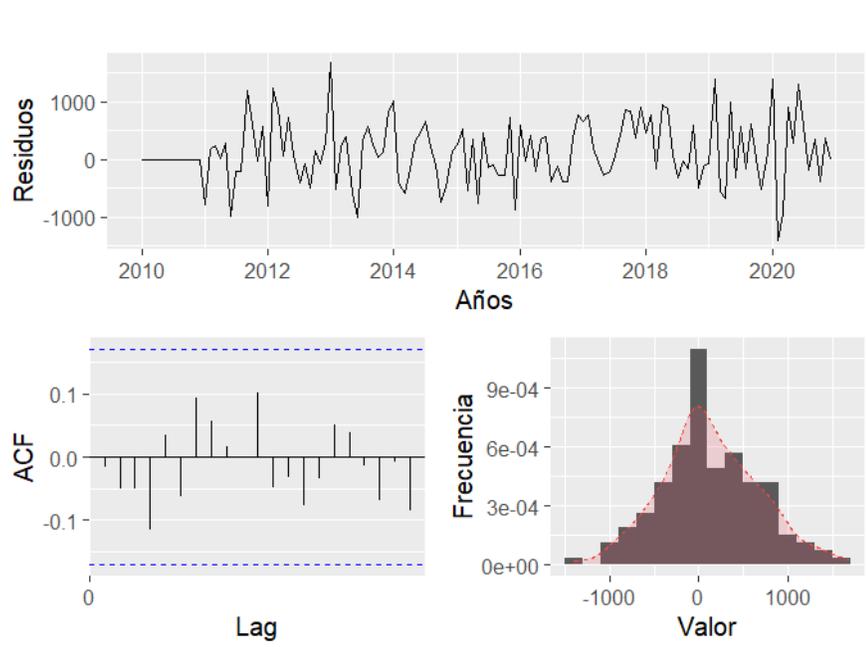
El coeficiente de ajuste AIC del modelo es = 1890.84. Las medidas de error del modelo son las siguientes: Error medio (ME por sus siglas en inglés) = 125.50, Error cuadrático medio (RMSE por sus siglas en inglés) = 563.38, Error absoluto medio (MAE por sus siglas en inglés) = 429.58, Error porcentual medio (MPE por sus siglas en inglés) = 1.27, Error absoluto porcentual medio (MAPE por sus siglas en inglés) = 17.60 y Error medio absoluto escalado medio (MASE por sus siglas en inglés) = 0.71.

3.3.2.4.1. Análisis de residuos

Se realizaron pruebas a los residuos para comprobar que estos tengan cumplan en los supuestos del modelo.

Figura 23.

Análisis de residuos del modelo ARIMA(4,0,0)(1,1,1)[12]

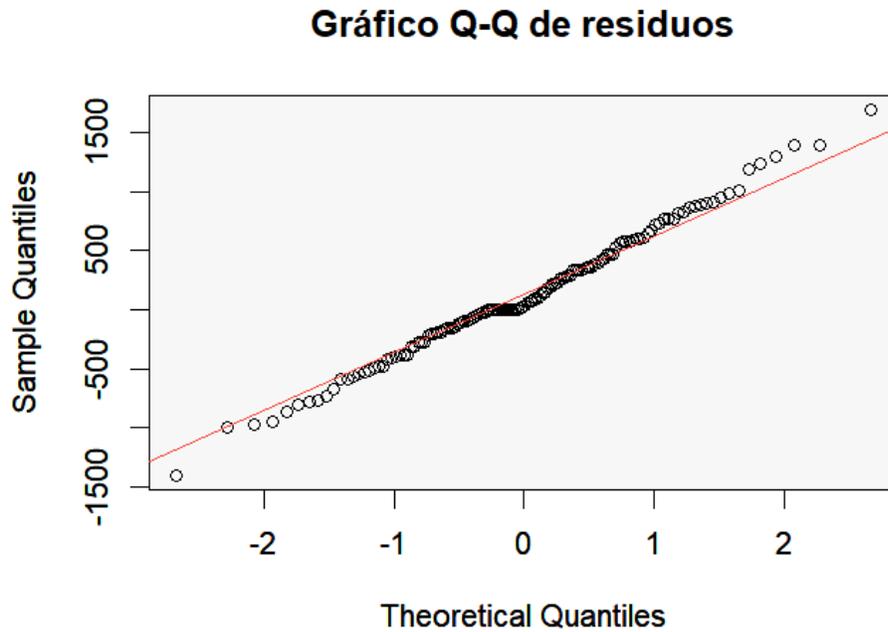


Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 23 se puede observar que los residuos del modelo ARIMA(4,0,0)(1,1,1)[12] tienen media cero, no existe autocorrelación y siguen la distribución normal.

Figura 24.

Gráfico de Q-Q de los residuos del modelo ARIMA(4,0,0)(1,1,1)[12]

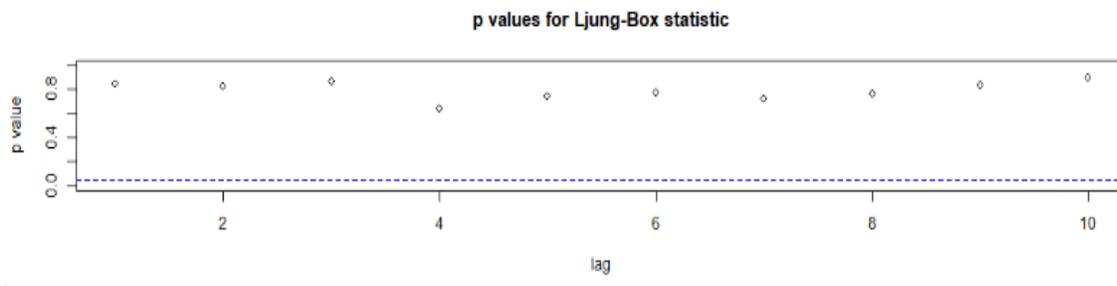


Nota. En la figura se puede observar que los residuos presentan una distribución normal. Elaboración propia, realizado con Rstudio.

En la figura 24 se puede observar que los residuos se apegan a la línea diagonal por lo que se confirma que siguen una distribución normal.

Figura 25.

Gráfico de valores-p para la prueba Ljung-Box



Nota. En la figura se puede evidenciar que todos los valores se encuentran fuera de la zona de rechazo. Elaboración propia, realizado con en Rstudio.

En la figura 25 se puede observar que todos los valores-p se encuentran fuera de la zona de rechazo, por lo que se acepta la H_0 = los residuos son independientes.

3.3.2.5. Selección del mejor modelo

Para poder elegir el mejor modelo se sometieron a evaluación algunos criterios de información de los cuatro modelos, estos se presentan en un cuadro comparativo (ver tabla 11).

Tabla 11.*Tabla resumen de criterios y características de los 4 modelos evaluados*

Coeficiente	Modelo 1 ARIMA (2,0,0)(0,1,1)[12]	Modelo 2 ARIMA (3,0,0)(0,1,1)[12]	Modelo 3 ARIMA (4,0,0)(0,1,1)[12]	Modelo 4 ARIMA (4,0,0)(1,1,1)[12]
valor-p ar1	0.002 **	0.0025 **	0.0011 **	0.0012 **
valor-p ar2	0.168 ns	0.2072 ns	0.0782 ns	0.0877 ns
valor-p ar3		0.7994 ns	0.2042 ns	0.2098 ns
valor-p ar4			0.0004 **	0.0004 **
valor-p sar1				0.4107 ns
valor-p sma1	<0.0001 **	<0.0001 **	<0.0001 **	<0.0001 **
AIC	1894.18	1895.39	1888.93 **	1890.84
MAPE	18.26	18.42	17.59 **	17.60

Nota. En la tabla se muestran el valor-p para los coeficientes de cada modelo. Además, se muestra el AIC y MAPE señalando con doble asterisco el mejor. Elaboración propia, realizado con Microsoft Word.

En la tabla 11 se puede observar un resumen de los coeficientes aplicados a cada modelo y su valor-p mostrando su significancia. Considerando que el modelo 3 cuenta con el AIC y el MAPE más bajo se considera el mejor modelo el número 3, el cual es un ARIMA(4,0,0)(0,1,1)[12], el cual consta de un componente autorregresivo de orden 4 y un componente de media móvil de estacionalidad de orden 1. Además, se aplicó una diferencia estacional de primer orden a los datos.

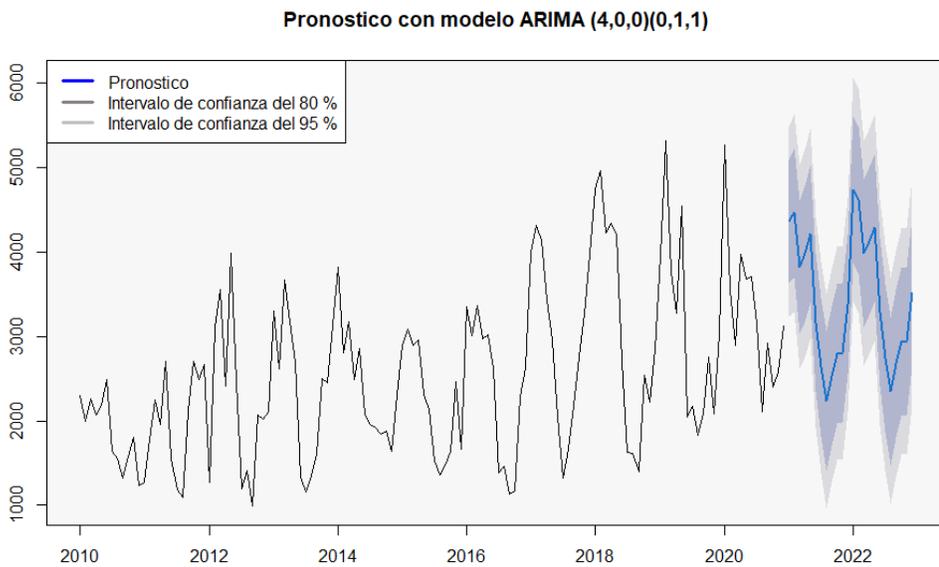
3.3.3. Pronóstico del modelo

Continuando con el análisis del comportamiento de la serie de tiempo de la producción mensual de racimos de fruta fresca de palma aplicando el modelo

ARIMA(4,0,0)(0,1,1)[12] se procedió a pronosticar 24 meses siguientes (periodo enero 2021 a diciembre 2022).

Figura 26.

Pronóstico de 24 meses con el modelo ARIMA(4,0,0)(0,1,1)[12]



Nota. Gráfica con la serie temporal con pronóstico de 24 meses utilizando el modelo ARIMA(4,0,0)(0,1,1)[12]. Elaboración propia, realizado con Rstudio.

3.4. Objetivo general. Diseñar un modelo estadístico para pronosticar la producción de palma de aceite mediante análisis correlacionales y series de tiempo de la producción durante el periodo 2010 al 2020

De acuerdo con el objetivo general, se pudo constar que el grado de asociación de la variable producción de racimos de fruta fresca mensual tiene un bajo grado de correlación de Pearson con las variables climáticas: temperatura

media mensual (-0.12), temperatura máxima promedio mensual (0.08), radiación solar promedio mensual (-0.03) y humedad relativa promedio mensual (0.01). Y una asociación media-baja con las variables temperatura mínima promedio mensual (-0.24) y precipitación mensual (-0.33).

Se construyeron tres modelos de regresión múltiple, los cuales fueron evaluados y se tomó como el mejor modelo el número 3, el cual incluye tres variables independientes (temperatura media mensual, precipitación mensual y humedad relativa promedio), predominando por tener los mejores criterios de información VIF (1.51-1.92) y Cp de Mallows (10.83-20.07). Adicionalmente se estimó un MAPE de 32.59 %.

Se propone utilizar para la creación de pronósticos de producción el modelo ARIMA(4,0,0)(0,1,1)[12], el cual presentó un MAPE de 17.18 %. Este modelo se denota de la siguiente forma:

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4) \cdot (1 - \Phi_1 L^{12}) \cdot (1 - L^{12}) Y_t = (1 + \theta_1 L^{12}) \cdot a_t \quad (22)$$

Donde:

L es el operador de rezago

Y_t : es el valor de la serie de tiempo t

$\phi_1, \phi_2, \phi_3, \phi_4$: son los coeficientes de autorregresión

Φ_1 : es el coeficiente de autorregresión estacional

a_t : es el término de ruido blanco en el tiempo t

θ_1 es el coeficiente de media móvil estacional

4. DISCUSIÓN DE LOS RESULTADOS

El objetivo de la investigación fue el de construir o ajustar un modelo estadístico para poder estimar la producción de racimos de fruta fresca de palma de aceite.

Para poder llegar a esto se realizaron pruebas comparativas de modelos de regresión lineal múltiple y modelos de series temporales. Así mismo se realizaron análisis correlacionales para poder identificar la relación que se tenía entre la variable de producción con variables climáticas y poder incluirla en los modelos de regresión lineal múltiple.

El conjunto de datos analizados comprende registros mensuales de 7 variables históricas comprendidos en el periodo de los años 2010 al 2020, obtenidos de los registros históricos de la producción y registros climáticos de una estación meteorológica.

4.1. Análisis interno del estudio

Para la realización del análisis de la información se debió recolectar toda la información de las variables en estudio, la variable de producción se debió estandarizar o normalizar a kilogramos por hectárea, esto para tener una estructura métrica comparable en cualquier lugar y en cualquier época si se requiere realizar una comparación con esta investigación.

Con la recolección de las variables ambientales, se debió tomar los registros electrónicos de una estación meteorológica ubicada en el área de

influencia de la finca, la cual está ubicada 3600 metros de distancia de la finca en estudio.

Para esta investigación las variables se estudiaron desde la perspectiva estadística, no se consideraron los aspectos biológicos de las plantas, esto por ser un estudio longitudinal y retrospectivo no se cuenta con toda la información necesaria para poder realizar un análisis considerando otras variables a parte de las ambientales.

Los métodos que se utilizaron para realizar el pronóstico fueron dos:

- Método de regresión lineal múltiple
- Método de series temporales

Del primero se puede indicar que su ajuste no fue el mejor ya que la variable en estudio depende de muchos factores externos como los climáticos, pero también depende de otros que son internos ya que la palma de aceite un organismo vivo reacciona a diversos estímulos, esto hace que la producción no tenga un comportamiento lineal y por ende que no se tenga un buen ajuste del modelo.

Por otra parte, la metodología de series temporales es un poco más versátil y de enfoque diferente que el de regresión lineal múltiple, ya que este analiza la variable respuesta con la variable tiempo.

Aunque el modelo de serie de tiempo muestra un mejor ajuste, se considera que no es el mejor método para realizar pronósticos debido a que no considera otros factores que inciden directamente sobre el individuo en estudio.

Sin embargo, la investigación se puede considerar válida ya que el objetivo era construir un modelo de predicción de la producción, y no entender las fluctuaciones de producción por factores externos.

Los resultados de ambos modelos fueron revisados y evaluados con los datos de producción reales, mostrando variaciones interanuales más ajustadas que el ajuste, analizando el total de los resultados como lo indica la metodología de estadística utilizada.

Esto puede indicar que existe alguna variación que no se esté tomando en consideración que hace variar fuertemente los resultados globales.

La elección del mejor modelo se considera válida, ya que se realizaron tres rondas de comparaciones.

- La primera fue comparar tres modelos de regresión múltiple con diferentes variables independientes, las cuales se fueron depurando según los criterios de información (R^2 , R^2 ajustado, AIC, VIF, C_p de Mallows), significancia mostrada en el modelo por cada una de ellas (valor-p) y cantidad de variables independientes, determinando que el mejor modelo fue el que tuvo los mejores de estos.
- Para la segunda ronda se compararon 4 modelos de series de tiempo, construyendo el primero según el análisis exploratorio de la serie, utilizando técnicas estadísticas como la prueba de Dickey Fuller Aumentada, diferenciando estacionalmente la serie, analizando gráficas de autocorrelación, autocorrelación parcial, Q-Q, descomposición de la serie, y luego agregando coeficientes autorregresivos y autorregresivos

estacionales para poder comparar los modelos con la significancia, criterio de información BIC y el error MAPE.

- La tercera ronda fue comparar el mejor modelo de regresión lineal múltiple con el mejor modelo de serie de tiempo para poder determinar cuál se ajustaba mejor según la tabla promedio de indicadores.

4.2. Análisis externo del estudio

Para lograr cumplir con los objetivos del estudio se utilizaron softwares estadísticos como Rstudio e Infostat, además de softwares de apoyo para la organización de información y redacción del informe final como Microsoft Excel y Microsoft Word.

Además de estos se utilizaron fuentes de información para lograr sustentar de forma teórica y práctica con literatura y ensayos anteriormente realizados relacionados con el tema de investigación.

Cordero et al. (1994), realizaron una investigación en la cual se presenta como resultado dos modelos de regresión lineal múltiple. El autor indica que no pueden ser tomados literalmente, ya que para realizarlos se llevó a cabo un promedio móvil de 3 periodos a la variable producción, lo cual sesga el resultado debido a que no se tomó la producción real mensualmente, perdiendo información en la suavización de la curva al utilizar este método.

Para este estudio se aplicó la producción real mensual por hectárea, considerando poder realizar su aplicación para el pronóstico de esta. Se coincide con el autor, donde se indica que esta metodología no puede sustituir a un

experto y es más una herramienta para poder tener un punto comparativo con el criterio de este.

Además, Erales (2017), propone una metodología en la cual se realiza un muestreo, pero esta tiene un menor alcance, ya que solo se puede proyectar la producción por cuatro meses.

La metodología estadística es el muestreo y los cálculos son basados en promedios históricos, no tiene estimado el error del pronóstico lo que da una incertidumbre muy amplia. La técnica estadística de serie de tiempo puede aumentar el alcance del pronóstico a un mayor periodo, de hasta 24 meses con un MAPE de 17.59 %.

La presente investigación tiene elementos similares con los del autor, para poder lograr un pronóstico de la producción ajustado.

El uso de la regresión lineal simple y coeficiente de Pearson se ha dado en otros estudios. Nda et al. (2018), realizó estudios de pronósticos, en este estudio se utilizaron las variables temperatura, precipitación para la comprensión de la variable producción. Los resultados muestran un bajo nivel de correlación en la relación las variables temperatura y precipitación con la producción.

En este estudio, el autor se limita bastante con la evaluación de dos variables, considerando que las plantas son organismos vivos, y la producción se ve influenciada por diferentes factores como lo son climáticos, agronómicos, genéticos, edáficos, entre otros.

Sin embargo, se considera que el modelo de regresión lineal múltiple presenta mayor grado de confiabilidad en la predicción para estas variables

climáticas considerando que su efecto no es aislado, y considera el efecto de la interacción de estas.

Ramirez y Potes (2019) realizaron un estudio para construir un modelo de pronóstico del rendimiento del cultivo de maracuyá a partir de datos históricos, en el cual utilizaron la técnica de regresión lineal múltiple y la de series de tiempo.

Los autores determinaron que por el nivel bajo de correlación entre la variable dependiente y las independientes el método de regresión múltiple no es el adecuado para realizar el modelo de pronóstico, por lo que utiliza la técnica de series de tiempo.

Para esta técnica se realizaron pruebas de estacionariedad, estacionalidad y descomposición de la serie, dando como el modelo más ajustado el ARIMA(0,1,0). Los autores indican que las dos metodologías no son comparables debido a su ajuste por diferentes metodologías.

Para construir el modelo de predicción de la producción palma de aceite se aplicó un proceso similar a que utilizaron Ramirez y Potes (2019). Se concuerda con los criterios de comparación de modelos, a pesar de ser diferentes cultivos en los estudios, las características son similares por lo que se pudo aplicar los métodos utilizados.

Sin embargo, a pesar de que indican que las metodologías de regresión lineal múltiple y de series de tiempo no son comparables, se realizó una comparación en términos del error porcentual absoluto medio para poder diagnosticar el modelo con mejor ajuste.

Para la construcción del modelo de pronóstico fue sumamente importante el conocimiento construido en estudios anteriores como los publicados por Luis-rojas et al. (2020), quienes utilizaron la técnica de serie de tiempo seleccionando un modelo ARIMA(1,1,1), con el cual se pronosticó la producción en el cultivo de vainilla verde en México, con información histórica del periodo de 1961 a 2014.

Para la realización del pronóstico de producción de palma se utilizó información de 132 meses, utilizando la técnica de series de tiempo se desarrolló un modelo ARIMA con estacionalidad.

Se coincide con los autores Luis-rojas et al. (2020), que es necesario ampliar el número de variables y realizar análisis multivariado para poder tener una predicción más confiable debido a que la técnica de serie temporales solo considera las fluctuaciones de la variable a través del tiempo.

CONCLUSIONES

1. Se estimó que el grado de asociación o correlación de los factores ambientales con la producción es estadísticamente bajo o nulo, siendo la precipitación (-0.33) y la temperatura mínima promedio mensual (-0.24) los que mayor coeficiente de Pearson presentan.
2. Se construyó un modelo estadístico con regresión lineal múltiple, en el cual puede observar que existe relación lineal baja debido al coeficiente de determinación relativamente bajo ($R^2 = 0.22$), y la relación significativa se presenta con las variables “temperatura mínima mensual”, “precipitación mensual” y “humedad relativa promedio mensual”, descartando el resto de las variables del modelo de pronóstico con este método estadístico.
3. Se construyó un método de predicción ajustando un modelo de series temporales. En el cual se identificó que la variable producción no tiene tendencia y si tiene estacionalidad, por esta razón se utilizó un modelo ARIMA estacional. Donde se tomaron en cuenta los criterios de información y análisis de autocorrelación y autocorrelación parcial se eligió el modelo ARIMA(4,0,0)(0,1,1)[12].
4. Se diseñó un modelo estadístico por el método de serie de tiempo, el cual es el modelo ARIMA(4,0,0)(0,1,1)[12], este tuvo un error porcentual absoluto medio (MAPE) = 17.59 %. Por lo que estadísticamente se puede indicar que para el pronóstico de producción es más ajustado un modelo correlacional con la variable tiempo (series de tiempo) y no con variables climáticas.

RECOMENDACIONES

1. Continuar con estudios que incluyan la estimación de la relación estadística con la aplicación de otras técnicas, para poder encontrar la relación entre las variables ambientales y la variable producción, para poder explicar la asociación y el efecto que estos tienen sobre esta y ampliar con nuevas variables independientes que afecten directa e indirectamente el proceso de formación de frutos de palma de aceite.
2. Construir modelos de pronóstico de producción de palma de aceite con la técnica de regresión lineal múltiple, agregando otras variables climáticas y de manejo agronómico (ej. la nutrición, riego, entre otros), y así para poder mejorar el nivel de ajuste del modelo.
3. Elaborar un modelo de pronóstico de serie temporal ARIMA estacional implica tomar en cuenta el ajuste y la cantidad de componentes incluidos para contar con un modelo no complejo y preciso en su aplicación. Además, hay que considerar que se amplía el período a pronosticar se pierde la precisión.
4. Diseñar un modelo de pronóstico utilizando las técnicas de análisis correlacionales y análisis de serie de tiempo sugiere la consideración de un periodo amplio de información histórica. En el caso específico de la regresión lineal múltiple, se debe contar con variables independientes que incidan directa o indirectamente en la fisiología de la planta dado que se trata de individuos vivos y susceptibles a cualquier estímulo interno o externo. Además, es necesario tener en cuenta que el modelo debe

actualizarse constantemente debido a la fluctuación o cambios en los patrones de las variables, los cuales conllevan a los estímulos mencionados anteriormente.

REFERENCIAS

- Anderson, D.R., Sweeney, D.J., & Williams, T.A. (2008). *Estadística para la administración y economía*. Manson, OH: Thomson South-Western.
- Breusch, T., & Pagan, A. (1979). A simple test for heteroscedasticity and random coefficient variation [Prueba de heterocedasticidad y variación de coeficientes aleatorios]. *Econometrika*, 47(5), 1287-1294. <https://doi.org/10.2307/1911963>
- Chatfield, C. (2016). *The analysis of time series: an introduction* [El análisis de series temporales: una introducción]. Chapman and Hall/CRC.
- Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example* [Análisis de regresión mediante ejemplos]. John Wiley & Sons.
- Cordero, L., Barboza, R., y Sterling, F. (1994). Un modelo estadístico para la predicción del rendimiento en una plantación de palma aceitera en Costa Rica. *Revista Agronomía Costarricense*, 18(2), 189-196. http://www.mag.go.cr/rev_agr/v18n02_indice.html
- Corzo, J. M., Juracán, R., Ical, B., & López, L. M. (2020). Factores importantes en la formación y conformación del racimo en el cultivo de palma de aceite. *Boletín La Palma* (18), 1-11
- Enders, W. (2015). *Applied econometric time series* [Series temporales econométricas aplicadas]. Wiley.

- Erales, R. (2017). Estimado de producción. *Boletín Técnico La Palma*, (9), 1-12. <https://www.grepalma.org/wp-content/uploads/2018/09/boletin-no9-febrero2017-vf.pdf>
- Forero H., D. C., Hormaza M., P. A., Moreno C., L. P., & Ruiz R., R. (2012). *Generalidades sobre la morfología y fenología de la palma de aceite*. Colombia. Centro de Investigación en Palma de Aceite.
- Flores T, C. E., y Flores C, K. L. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov. *Societas. Revista de Ciencias Sociales y Humanísticas*, 23(2), 83-97.
- Frausto E, J. H. (2009). *Pronósticos de ventas para la administración estratégica de los recursos en una empresa del sector automotriz*. [Tesis de maestría. Instituto Tecnológico y de Estudios Superiores de Monterrey]. Archivo digital. https://repositorio.tec.mx/bitstream/handle/11285/569469/DocsTec_10231.pdf?seq
- Galindo-Domínguez, H. (2020). *Estadística para no estadísticos. Una guía básica sobre la metodología cuantitativa de trabajos académicos*. Editorial área de Innovación y Desarrollo, S.L.
- García-Arteaga, J. J., Zambrano-Zambrano, J. J., Alcivar-Cevallos, R., & Zambrano-Romero, W. D. (2020). Predicción del rendimiento de cultivos agrícolas usando aprendizaje automático. *Revista Arbitrada Interdisciplinaria KOINONIA*, 5(2), 145-160. 144. <https://doi.org/10.35381/r.k.v5i2.1013>

- Gorgas G. J., Cardiel, L., N., & Zamorano, C, J. (2011). *Estadística básica para estudiantes de ciencias*. https://webs.ucm.es/info/Astrof/users/jaz/ESTADISTICA/libro_GCZ2009.pdf
- Gremial de Palmicultores de Guatemala. (2021). *Estadísticas socioeconómicas al año 2021*. <https://www.grepalma.org/wp-content/uploads/2022/05/Anuario-Estadistico-2020-2021-22-5-2022.pdf>
- Hanke J. E., & Reitsh, A. G. (1996). *Pronóstico en los negocios*. Prentice Hall.
- Hyndman R. J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*. [Pronóstico: principios y práctica]. https://play.google.com/books/reader?id=_bBhDwAAQBAJ&pg=GBS.PA1
- Jaramillo V., R. (2020). Cambio climático y manejo de la nutrición en palma de aceite. *Boletín Técnico La Palma*, (17), 1-15. https://www.grepalma.org/wp-content/uploads/2020/12/BOLETIN-17_ANO-5_-DIC-2020-VF.pdf
- Laurente Blanco, L. F., y Laurente Quiñonez, 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(1), 30-40. http://www.scielo.org.bo/pdf/riiarn/v6n1/v6n1_a06.pdf
- Lind, D. A., Marchal, W. G., & Wathen, S. A. (2012). *Estadística aplicada a los negocios y la economía*. México: McGrawHill/Interamericana editores, S.A. de C.V.

- Luis-Rojas, S., Ramírez-Valverde, B., Díaz-Bautista, M., Pizano-Calderón, J. & Rodríguez-López, C. (2020). La producción de vainilla (*Vanilla planifolia*) en México: análisis y pronóstico. *Revista Mexicana de Ciencias Agrícolas*, 11(1), 175-187. <https://www.scielo.org.mx/pdf/remexca/v11n1/2007-0934-remexca-11-01-175.pdf>
- Martínez, D. R., Albín, J. L., Cabaleiro, J. C., Pena, T. F., Rivera, F. F., & Blanco, V. (2009) El criterio de información de Akaike en la obtención de modelos estadísticos de rendimiento. *Actas de las XX jornadas de paralelismo*. Universidad da Coruña. https://www.researchgate.net/publication/236279245_El_criterio_de_informacion_de_Akaike_en_la_obtencion_de_modelos_estadisticos_de_Rendimiento/citation/download
- Ministerio de Agricultura Ganadería y Alimentación. (2021). *Determinación de la cobertura vegetal y uso de la tierra a escala 1:50,000 de la República de Guatemala, año 2,020*. MAGA.
- Montesinos, L., A. (2011). *Estudio del AIC y BIC en la selección de modelos de vida con datos censurados* [Tesis de Maestría, Centro de Investigación en Matemáticas, A.C.]. <http://cimat.repositorioinstitucional.mx/jspui/handle/1008/293>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* [Introducción al análisis de regresión lineal]. Wiley.
- Nda, M., Adnan, M. S., Suhadak, M. A., Zakaria, M. S., & Lopa, R.T. (2018). Effects of hydrological parameteres on palm oil fresh fruit bunch yield [Efectos de los parámetros hidrológicos en el rendimiento del racimo]

fresco de frutas de palma aceitera]. *IOP Conference Series: Earth and Environmental Science* [Serie de Conferencias del OIP: Ciencia de la Tierra y el Medio Ambiente], 140(1), 12-51. <https://iopscience.iop.org/article/10.1088/1755-1315/140/1/012051/pdf>

Posada H, G. J. (2016). *Elementos básicos de estadística descriptiva para el análisis de datos*. Fondo Editorial Luis Amigó.

Ramos-Escalante, G., Ley de-Coss, A., Arce-Espino, C., Escobar-España, J. C., Raj-Aryal, D., Pinto-Ruiz, R., & Guerra-Medina, C. E. (2018). Ecuaciones alométricas para estimar biomasa y carbono en palma de aceite (*Elaeis guineensis* Jacq.) en el trópico húmedo de Chiapas, México. *Agrociencia*, 52(5), 671-683.

Ramirez C, L. N., y Potes, S. (2019). Estimación del rendimiento del cultivo de *Passiflora Edulis* (Maracuyá) a partir de modelos estadísticos. *Inventum*, 14(26), 33-42. <https://revistas.uniminuto.edu/index.php/Inventum/article/view/1962/1807>

Schroeder, R. G., Meyer Goldstein, S., & Rungtusanatham, M. J. (2011). Editor Mares Chacón, Jesus. *Administración de operaciones: conceptos y casos contemporáneos*. México: McGrawHill/Interamericana editores, S.A. de C.V.

Siew, K, N., von Uexküll, H., & Härdter, R. (2012). Aspectos botánicos de la palma de aceite pertinentes al manejo del cultivo. En Fairhurst, T. y

Hårdter, R. (Eds.), *Palma de aceite, manejo para rendimientos altos y sostenibles* (pp 33-46).

Silva, C., Jadue M., J., Crivelli M., A. & Buccioni P., E. (2017). Regresión Lineal. Universidad de Santiago de Chile. *Revista Contribuciones Científicas y Tecnológicas*, 38-48.
<https://www.revistas.usach.cl/ojs/index.php/contribuciones/article/view/3026>

Szretter N, M. E. (2017). *Apunte de regresión lineal*. Argentina: Universidad de Buenos Aires.

Tunncliffe, W., G. (2016). *Time series analysis: forecasting and control* [Análisis de series temporales: pronóstico y control]. *Journal of Time Series Analysis* [Revista de Análisis de Series Temporales], 37(5), 709-711.
<https://doi.org/10.1111/jtsa.12194>

Woittiez, L. S., Van Wijk, M. T., Slingerland, m. Van Noordwijk, M., & Giller, K. E. (2018). Brechas de rendimiento en el cultivo de palma de aceite: una revisión cuantitativa de factores determinante. *Palmas*, 39(1), 16-68.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* [Análisis econométrico de datos de sección cruzada y panel]. MIT Press.

APÉNDICE

Apéndice 1. *Matriz de coherencia*

DISEÑO DE UN MODELO ESTADÍSTICO PARA PRONOSTICAR LA PRODUCCIÓN DE PALMA DE ACEITE MEDIANTE ANÁLISIS CORRELACIONALES Y SERIES DE TIEMPO DE LA PRODUCCIÓN DURANTE EL PERIODO 2010 AL 2020							
ELEMENTOS	PROBLEMA DE INVESTIGACIÓN	PREGUNTAS DE INVESTIGACIÓN	OBJETIVOS	METODOLOGIA	RESULTADOS	CONCLUSIONES	RECOMENDACIONES
GENERAL O CENTRAL	Se desconoce el efecto individual de la interacción de los factores ambientales que inciden sobre el comportamiento de la producción de fruta de palma en el tiempo.	¿Cuáles son los factores ambientales que se relacionan con la producción de fruta de palma y qué efecto directo generan en el comportamiento de la producción durante el período 2010 al 2020?	Diseñar un modelo estadístico para pronosticar la producción de palma de aceite mediante análisis correlacionales y series de tiempo de la producción durante el período 2010 al 2020.	Las mismas que en los específicos	En el análisis comparativo de los modelos por regresión lineal y por series de tiempo el mejor modelo fue el de series de tiempo ARIMA(4,0,0)(0,1,1)[12] con un MAPE de 17.59 %	Se estimó que el grado de asociación o correlación de los factores ambientales con la producción estadísticamente es bajo o nula, siendo la precipitación (-0.33) y la temperatura mínima promedio mensual (-0.24) los que mayor coeficiente de Pearson presentan.	1. Es importante continuar la estimación de la relación estadística con la aplicación de otras técnicas, para poder encontrar la relación entre las variables ambientales y la variable producción, para poder explicar la asociación y el efecto que estos tienen sobre esta y ampliar con nuevas variables independientes que afecten directa e indirectamente el proceso de formación de frutos de palma de aceite.
ESPECÍFICOS O AUXILIARES	1. Se desconoce el efecto individual que generan los factores ambientales que generan un efecto directo sobre el comportamiento en el tiempo de la producción de fruta de palma.	1. ¿Cuál es el efecto individual que generan directamente los factores ambientales sobre el comportamiento en el tiempo de la producción de fruta de palma?	1. Estimar el grado de asociación y dependencia estadística entre la variable producción con respecto a las variables climáticas entre los años 2010 y 2020 a través del coeficiente de correlación	Análisis de correlación de Pearson y medidas de tendencia central y de dispersión	Se puede observar el grado de correlación bajo para la variable producción con las variables precipitación y temperatura mínima mensual. Para el resto de variables no existe correlación.	Se construyó un modelo estadístico con regresión lineal múltiple, en el cual puede observarse que existe relación lineal baja debido al coeficiente de determinación relativamente bajo ($R^2 = 0.22$), y la relación significativa se presenta con las variables "temperatura mínima mensual", "precipitación mensual" y "humedad relativa promedio mensual", descartando el resto de las variables del modelo de pronóstico con este método estadístico.	2. Es necesario al construir modelos de pronóstico de producción de palma de aceite con la técnica de regresión lineal múltiple, agregar o incluir otras variables climáticas y de manejo agronómico (ej. la nutrición, riego, etc.), y así para poder mejorar el nivel de ajuste del modelo.
ESPECÍFICOS O AUXILIARES	2. Se desconoce el efecto de la interacción de los factores ambientales sobre el comportamiento en el tiempo de la producción de fruta de palma.	2. ¿Qué efecto genera la interacción de los factores ambientales sobre el comportamiento en el tiempo de la producción de fruta de palma?	2. Construir un modelo estadístico de predicción de producción utilizando las variables ambientales mediante regresión lineal múltiple.	Análisis de regresión lineal múltiple, coeficientes de información VIF, Cp Mallows, MAPE, AIC y BIC	El mejor modelo de regresión lineal múltiple es el que incluye como variables independientes a la precipitación, temperatura mínima mensual y humedad relativa promedio mensual con un MAPE de 32.59 %	Se construyó un modelo con la técnica de series temporales, en el cual se identificó que no tiene estacionalidad o tendencia y si tiene estacionalidad, por lo que el mejor modelo que se ajusta para el pronóstico de la variable fue el ARIMA(4,0,0)(0,1,1)[12].	3. Al construir un modelo de pronóstico de serie temporal ARIMA estacional, se debe tomar en cuenta el ajuste y la cantidad de componentes incluidos para contar con un modelo no complejo y preciso en su aplicación. Adema hay que considerar que el periodo a pronosticar cuando crece se pierde la precisión.
ESPECÍFICOS O AUXILIARES	3. No se cuenta con un modelo estadístico confiable para pronosticar la producción de palma de aceite.	3. ¿Qué modelo de serie de tiempo se ajusta de manera confiable para pronosticar la producción de palma de aceite?	3. Construir un modelo estadístico que se ajuste aceptablemente al comportamiento de la producción de estudio mediante un análisis de series temporales para obtener predicciones aceptables.	Análisis de series de tiempo, técnica de autocorrelación y autocorrelación parcial, MAPE, BIC.	En el análisis de la serie de datos de producción se pudo observar que no tenía tendencia, pero si tenía estacionalidad por lo que el modelo que mas se ajusto fue el ARIMA(4,0,0)(0,1,1)[12] con un MAPE de 17.59 %	Se diseño un modelo estadístico por el método de serie de tiempo, el cual es el modelo ARIMA(4,0,0)(0,1,1)[12]. este tuvo un error porcentual absoluto medio (MAPE = 17.59 %). Por lo que estadísticamente se puede indicar que para el pronóstico de producción es más ajustado un modelo correlacional con el tiempo (series de tiempo) y no con factores climáticos.	4. Se sugiere que al diseñar un modelo de pronóstico utilizando las técnicas de análisis correlacionales y análisis de serie de tiempo se debe utilizar un periodo amplio de información histórica, y en el caso específico para regresión lineal múltiple se debe contar con variables independientes que incidan directa o indirectamente fisiológicamente (ya que para este caso se trata de individuos vivos y obedecen a cualquier estímulo interno o externo. Y se debe tomar en cuenta que el modelo debe estar actualizando constantemente debido a la fluctuación o cambios en los patrones de las variables que conllevan a los estímulos antes mencionados.

Nota. En el apéndice 1 se puede observar resumido el trabajo. Elaboración propia, realizado con Microsoft Excel.

