

UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN MATEMÁTICA APLICADA

Tema: “MODELO MATEMÁTICO DE LA PRODUCCIÓN DE LA
EMPRESA SALINERITO EN LA PROVINCIA BOLÍVAR”

Trabajo de titulación previo a la obtención del grado académico de Magíster en
Matemática Aplicada

Modalidad de titulación proyecto de desarrollo

Autor: Lic. Carlos Alfredo Sánchez Verdezoto

Director: Dr. Freddy Geovanny Benalcázar Palacios, Mg.

Ambato – Ecuador

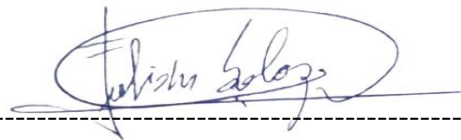
2021

APROBACIÓN DEL TRABAJO DE TITULACIÓN


A la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial. El Tribunal receptor de la Defensa del Trabajo de Titulación presidido por la Ing. Elsa Pilar Urrutia Urrutia Mg., e integrado por los señores: Ing. Fabián Rodrigo Salazar Escobar, Dr. y la Ing. Clara Augusta Sánchez Benítez Mg, designados por la Unidad Académica de Titulación de Posgrados de la Facultad de Ingeniería en Sistemas Electrónica e Industrial de la Universidad Técnica de Ambato, para receptor el Trabajo de Titulación con el tema: “Modelo matemático de la producción de la empresa Salinerito en la provincia Bolívar”, elaborado y presentado por el Lic. Carlos Alfredo Sánchez Verdezoto, para optar por el Grado Académico de Magister en Matemática Aplicada; una vez escuchada la defensa oral del Trabajo de Titulación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la Universidad Técnica de Ambato.



Ing. Elsa Pilar Urrutia Urrutia, Mg.
Presidente y Miembro del Tribunal de Defensa



Ing. Fabián Rodrigo Salazar Escobar, Dr.
Miembro del Tribunal de Defensa



Ing. Clara Augusta Sánchez Benítez, Mg.
Miembro del Tribunal de Defensa

AUTORÍA DEL TRABAJO DE TITULACIÓN

La responsabilidad de las opiniones, comentarios y críticas emitidas en el trabajo de titulación presentado con el tema: “Modelo matemático de la producción de la empresa Salinerito en la provincia Bolívar”, le corresponde exclusivamente a: Lic. Carlos Alfredo Sánchez Verdezoto, autor bajo la dirección de: Dr. Freddy Geovanny Benalcázar Palacios, Mg, director del trabajo de investigación; y el patrimonio intelectual pertenece a la Universidad Técnica de Ambato



Lic. Carlos Alfredo Sánchez Verdezoto

AUTOR



Dr. Freddy Geovanny Benalcázar Palacios, Mg

DIRECTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el trabajo de titulación, sirva como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos de mi trabajo de titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad Técnica de Ambato.

A handwritten signature in blue ink, appearing to read 'CARLOS SANCHEZ VERDEZOTO', is written over a horizontal dashed line.

Lic. Carlos Alfredo Sánchez Verdezoto

0202011854

ÍNDICE GENERAL

	Pág.
CONTENIDO	
PORTADA.....	i
APROBACIÓN DEL TRABAJO DE TITULACIÓN	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN	iii
DERECHOS DE AUTOR.....	iv
ÍNDICE DE TABLAS	vii
ÍNDICE DE FIGURAS.....	viii
AGRADECIMIENTO.....	ix
DEDICATORIA	x
RESUMEN EJECUTIVO	xi
EXECUTIVE SUMMARY.....	xii
CAPÍTULO I.....	1
EL PROBLEMA DE INVESTIGACIÓN.....	1
1.1. Introducción.....	1
1.2. Justificación.....	2
1.3. Objetivos.....	2
1.3.1. Objetivo General	2
1.3.2. Objetivos Específicos.....	2
CAPÍTULO II	3
ANTECEDENTES INVESTIGATIVOS.....	3
2.1. Estado del Arte	3
2.2. Fundamentos teóricos.....	8
2.2.1. Conceptos básicos	8
2.2.2. Metodología de Box-Jenkins.....	9
2.2.3. Procedimiento de la metodología Box-Jenkins.....	10
CAPÍTULO III.....	16

MARCO METODOLÓGICO	16
3.1. Ubicación.....	16
3.2. Equipos y materiales.....	18
3.3. Tipo de investigación	19
3.4. Problema e hipótesis.....	20
3.5. Población o muestra	20
3.6. Recolección de información	20
3.7. Procesamiento y análisis estadístico.....	20
3.8. Variables respuestas o resultados alcanzados.....	22
CAPÍTULO IV	23
RESULTADOS Y DISCUSIONES	23
4.1. Estructura de la base de datos.....	23
4.2. Aplicación de la metodología Box-Jenkins.....	23
4.2.1. Identificación del modelo.....	23
4.2.2. Estimación del modelo.....	29
4.2.3. Validación del modelo	30
4.2.4. Pronóstico.....	34
4.2.5. Comparación entre la producción y las ventas.....	36
4.3. Comparación entre la metodología Demand Driven Material Requirements Planning (DDMRP) VS SARIMA (1, 1, 1) (1,1,1) ¹²	38
CAPÍTULO V	39
5.1. CONCLUSIONES.....	39
5.2. RECOMENDACIONES	40
5.3. BIBLIOGRAFÍA.....	41
5.4. ANEXOS	46

ÍNDICE DE TABLAS

	Pág.
Tabla 1. Estructura de la base de datos	23
Tabla 2. Test de Dickey–Fuller Aumentada.	25
Tabla 3. Test de Dickey –Fuller Aumentada con una primera diferencia	26
Tabla 4. Modelos tentativos	28
Tabla 5. Resumen del modelo SARIMA (1, 1, 1) (1, 1, 1) 12.....	28
Tabla 6. Estimación del modelo ARIMA (1, 1, 1) (1, 1, 1)12	29
Tabla 7. Test de normalidad Jarque Bera.....	32
Tabla 8. Test de ruido blanco Ljung-Box	32
Tabla 9. Valores de entrenamiento y testeo	33
Tabla 10. Pronóstico de la cantidad de kilogramos producidos.....	35
Tabla 11. Pronóstico de kilogramos producidos y vendidos.....	37
Tabla 12. Metodología DDMRP VS SARIMA (1, 1, 1) (1, 1, 1) 12.....	38

ÍNDICE DE FIGURAS

	Pág.
Figura 1. Ubicación de la parroquia Salinas	16
Figura 2. Ubicación de la Empresa Confites El Salinerito.....	18
Figura 3. Metodología Box-Jenkins	18
Figura 4. Componentes de los valores históricos.....	24
Figura 5. Comportamiento de los valores con la primera diferencia	25
Figura 6. Función de autocorrelacion simple y parcial de la serie diferenciada	27
Figura 7. Diagnóstico del modelo SARIMA (1, 1, 1) (1, 1, 1) ¹²	30
Figura 8. Gráfico de los residuos del modelo	31
Figura 9. Valores de entrenamiento y testeo	34
Figura 10. Comportamiento de los pronósticos del modelo SARIMA.....	36
Figura 11. Mapa para llegar a Salinas	46
Figura 12. Parroquia Salinas	46
Figura 13. Fabrica Confites El Salinerito.....	47
Figura 14. Confites El Salinerito.....	47
Figura 15. Productos Terminados	48
Figura 16. Productos Terminados	48
Figura 17. Base de datos digitalizada de la empresa.....	49
Figura 18. Script del Modelo SARIMA (1, 1, 1) (1, 1, 1) ¹²	51

AGRADECIMIENTO

Agradezco en primer lugar a Dios que me ha dado la fuerza necesaria para no rendirme en el proceso educativo.

A mi madre que es el motivo de seguir luchando cada día para llegar hacer una persona de bien.

A mi esposa que me ha brindado confianza y soporte en los momentos más difíciles.

A mi familia que indistintamente me apoyan moralmente a lo largo de la vida.

A mis docentes y amigos quienes con sabiduría me han logrado formar como profesional y sobre todo con valores éticos y morales.

A mi director del trabajo de titulación quien con esfuerzo y dedicación me apoyó en el proceso de construcción del proyecto.

Carlos Alfredo Sánchez Verdezoto

DEDICATORIA

La presente investigación dedicó a tres mujeres maravillosas en primer lugar a mi madre quien me dio la vida y me forjó en los primeros años de vida, en segundo lugar, a mi esposa quien me ha acompañado en los momentos difíciles del proceso académico y por último a mi hermana que me apoya como si fuera mi segunda madre.

Carlos Alfredo Sánchez Verdezoto

UNIVERSIDAD TÉCNICA DE AMBATO
FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E
INDUSTRIAL
MAESTRÍA EN MATEMÁTICA APLICADA

TEMA:

“MODELO MATEMÁTICO DE LA PRODUCCIÓN DE LA EMPRESA SALINERITO EN LA PROVINCIA BOLÍVAR”

AUTOR: Lic. Carlos Alfredo Sánchez Verdezoto

DIRECTOR: Dr. Freddy Geovanny Benalcázar Palacios, Mg.

LÍNEA DE INVESTIGACIÓN: Modelamiento matemático

FECHA: 05 de enero 2021

RESUMEN EJECUTIVO

El presente trabajo investigativo es de tipo descriptivo, predictivo y longitudinal con tendencia. El objetivo fue desarrollar un modelo matemático para la producción de la empresa Confites El Salinerito en la provincia Bolívar, a partir de los valores de producción en el periodo de enero 2017 a julio 2020, los datos fueron obtenidos directamente de la empresa. Para alcanzar los objetivos de la investigación se aplicó la metodología propuesta por Box-Jenkins, la cual describe las características de las series de tiempo en términos de tendencia, estacionalidad y estacionariedad; para la estimación de los parámetros, el procesamiento y el análisis de los datos se aplicó el software libre RStudio versión 4.0.1. Se concluyó que el modelo matemático que se ajusta con mayor precisión a los valores de producción de la empresa fue el SARIMA (1, 1, 1) (1, 1, 1) 12, el mismo que permitió realizar los pronósticos de la producción mensual para el periodo de agosto 2020 a enero del 2021. También se realizó la comparación de los resultados que arrojan el modelo desarrollado y la metodología que utiliza la empresa para establecer su producción mensual.

Descriptor: Estacionalidad, estacionariedad, cantidad de kilogramos, Confites El Salinerito, Box-Jenkins, modelo matemático, producción mensual, pronósticos, series de tiempo, SARIMA

TECHNICAL UNIVERSITY OF AMBATO
FACULTY OF SYSTEMS, ELECTRONIC AND INDUSTRIAL
ENGINEERING
MASTER'S OF APPLIED MATHEMATICS

THEME:

"MATHEMATICAL MODEL OF THE PRODUCTION OF THE COMPANY
SALINERITO IN THE PROVINCE BOLÍVAR"

AUTHOR: Lic. Carlos Alfredo Sánchez Verdezoto

DIRECTED BY: Dr. Freddy Geovanny Benalcázar Palacios, Mg.

RESEARCH LINE: Mathematical modeling

DATE: January 05, 2021

EXECUTIVE SUMMARY

The present research work is descriptive, predictive and longitudinal with a trend. The objective was to develop a mathematical model for the production of the company Confites El Salinerito in the Bolívar province, from the production values in the period from January 2017 to July 2020, the data were obtained directly from the company. To achieve the research objectives, the methodology proposed by Box-Jenkins was applied, which describes the characteristics of the time series in terms of trend, seasonality and stationarity; The free software RStudio version 4.0.1 was applied for the estimation of the parameters, the processing and the analysis of the data. It was concluded that the mathematical model that most accurately adjusts to the production values of the company was SARIMA (1, 1, 1) (1, 1, 1) 12, the same one that allowed making the monthly production forecasts for the period from August 2020 to January 2021. The results obtained by the developed model and the methodology used by the company to establish its monthly production were also compared.

Keywords: Seasonality, stationarity, number of kilograms, Confites El Salinerito, Box-Jenkins, mathematical model, monthly production, forecasts, time series, SARIMA.

CAPÍTULO I

EL PROBLEMA DE INVESTIGACIÓN

1.1. Introducción

La investigación está orientada a la construcción de un modelo matemático que permita predecir la producción mensual de la empresa Confites El Salinerito. La metodología estadística seguida para la elaboración del modelo matemático fue la propuesta por Box-Jenkins que comprende de: identificación, estimación, validación y pronóstico.

En la etapa de identificación se verificó la estacionariedad, estacionalidad y la tendencia de la serie de tiempo; también se analizó la prueba de raíz unitaria y los correlogramas para seleccionar los valores de p , q , P , Q que forman a los modelos tentativos.

En la siguiente etapa se estimó los parámetros del modelo mediante el software libre RStudio versión 4.0.1, que utiliza el método de máxima verosimilitud. Mediante la estimación se comparó y analizó el error cuadrático medio (ECM), el criterio de información AKAIKE (AIC) y el criterio de información Bayesiano (BIC) de cada modelo tentativo, a partir de esto se eligió el modelo matemático SARIMA, formado por una parte estacionaria (1, 1, 1) y una parte estacional (1, 1, 1) con un periodo de estacionalidad 12. Una vez seleccionado el modelo y estimado los parámetros se construyó la ecuación que modela la producción mensual de la empresa.

En la etapa de validación se aplicó el estadístico Ljung-Box para comprobar que los residuos se comportan como ruido blanco. Por último, en la etapa de pronóstico se realizó la predicción de la producción mensual de la empresa para el periodo agosto de 2020 a enero de 2021.

Este documento se compone de cinco capítulos, dentro del primer capítulo se da a conocer la importancia y los objetivos de la investigación; en el segundo capítulo se describe el estado del arte y la fundamentación teórica; el tercer capítulo hace referencia al marco metodológico; en el cuarto capítulo se presentan los resultados y las discusiones; en el quinto y último capítulo se exponen las conclusiones y recomendaciones a las que se ha llegado mediante el desarrollo del presente trabajo investigativo.

1.2. Justificación

Actualmente la mayoría de las empresas analizan cada una de sus variables de respuesta o indicadores de desempeño de forma individual e independiente. Por esta razón, surge la necesidad de contar con recursos científicos que se traduzcan en modelos útiles que permitan obtener medidas de desempeño capaces de reflejar la situación real de las organizaciones [1].

Las empresas han desarrollado esfuerzos importantes en la construcción de modelos de planeación y programación de la producción, que se han constituido en una base fundamental para mejorar la competitividad de las empresas en términos de calidad, costo, flexibilidad, confiabilidad y desarrollo sostenible. Por esta razón, buscan como estrategia la construcción de modelos matemáticos y de programación [2].

A partir de lo expuesto anteriormente es necesario construir un modelo matemático que permita a la empresa Confites El Salinerito predecir la cantidad adecuada de kilogramos que debe producir mensualmente. Esta predicción se realiza con el fin de que la empresa pueda planificar las actividades productivas, aportándole mejor capacidad de respuesta y generando ahorro e incrementando las ganancias.

1.3. Objetivos

1.3.1. Objetivo General

Desarrollar un modelo matemático para la producción de la empresa Salinerito de la provincia Bolívar.

1.3.2. Objetivos Específicos

- Realizar un estudio del arte
- Construir un modelo matemático para la producción de la empresa El Salinerito de la parroquia Salinas de la provincia Bolívar.
- Evaluar el desempeño del modelo de producción.
- Comparar el modelo matemático construido frente al método actual de producción de la empresa El Salinerito.

CAPÍTULO II

ANTECEDENTES INVESTIGATIVOS

2.1. Estado del Arte

En la actualidad existen varios trabajos que documentan los modelos de predicción y los efectos que tienen sobre las variables, por ejemplo Sánchez [3] en su obra titulada “Modelo ARIMA para el pronóstico de la producción de cacao en el Perú” aplica la metodología propuesta por Box-Jenkins y llega a determinar un modelo de pronósticos que explica el comportamiento de la producción mensual a través de los datos históricos. El investigador establece que el modelo que predice y describe adecuadamente la producción mensual del cacao es el SARIMA compuesto por una parte estacionaria (7, 1, 7) y una parte estacional (1, 1, 0), con un periodo estacional 12, cuya ecuación de pronósticos estimada es: $\Delta Y_t = -0,87Y_{t-7} + 0,57\varepsilon_{t-7} - 0,43\varepsilon_{t-8} - 0,89Y_{t-12} + \varepsilon_t$.

Laurente & Quiñonez [4] en el trabajo titulado “Aplicación del modelo ARIMA para la producción de la papa en la región de Puno-Perú” emplean la metodología de Box-Jenkins y alcanzan a identificar un modelo que explica, describe y proyecta la producción de la papa. Los investigadores concluyen que el modelo más apropiado para explicar el comportamiento y la proyección de la producción de papa es el ARIMA compuesto por 3 regresores, 0 diferencias y 4 medias móviles.

Guerra & Medina [5] en el documento titulado “Modelos de series de tiempo aplicados a los expedientes de la Comisión de Derechos Humanos del Distrito Federal” utilizan los modelos de series de tiempo aditivas, en especial los modelos ARIMA y logran determinar un modelo para el pronóstico de los expedientes históricos acumulados. Los investigadores concluyen que al utilizar un modelo ARIMA se identifica cotas de comportamiento que sirven de referencia para la detección de comportamientos anormales en el número de expedientes acumulados.

Sánchez, Cabanas, Abad & Torres [6], en el trabajo titulado “Utilización de modelos ARIMA para la predicción de la producción de leche” aplican la metodología de Box-Jenkins que fue desarrollada por George Box y Gwilym Jenkins en 1970 y consiguen realizar un pronóstico a corto plazo de la producción de leche mediante modelos ARIMA. Los investigadores concluyen que el modelo que predice y explica la

producción de la leche es el SARIMA compuesto por una parte estacionaria (1, 0, 3) y una parte estacional (0, 1, 0) con un periodo estacional 12.

Para Botero & Cano [7] en la obra titulada “Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia” utilizan los modelos de series de tiempo y llegan a presentar una metodología para la implementación de modelos de regresión sobre una serie histórica de los precios de energía. Los investigadores concluyen que los modelos que siguen la metodología de Box-Jenkins son los adecuados para el modelamiento y la proyección de los precios de la energía.

Por otro lado, López, García & Herrera [8] en el documento titulado “Desarrollo de un modelo predictivo para la estimación del comportamiento de las variables en una infraestructura de red” emplean la metodología de Box-Jenkins y llegan a determinar un modelo que describe, explica y predice el comportamiento de las redes de telecomunicaciones. Los investigadores concluyen que el modelo ARIMA compuesto por 1 regresor, 1 diferencia y 3 medias móviles es el adecuado para estimar, describir y predecir el comportamiento de las variables en la infraestructura de una red.

Pavlyshenko [9] en su trabajo titulado “Modelo de aprendizaje automático para la previsión de series temporales de ventas” utiliza la metodología del efecto de la generalización del aprendizaje automático con un enfoque de apilamiento para construir modelos de regresión y consigue identificar los enfoques principales para el uso del aprendizaje automático en la previsión de ventas. El investigador concluye que al utilizar técnicas de apilamiento se mejora el rendimiento de los modelos predictivos.

Para Sánchez, Barreras, Pérez, Figueroa & Olivas [10], en la obra titulada “Aplicación de la metodología de Box-Jenkins para pronósticar la producción de leche de bovino en California México” aplican la metodología propuesta por Box-Jenkins y llegan a determinar un modelo univariante para el pronóstico de la producción lechera. Los investigadores establecen que el modelo que describe y predice adecuadamente el comportamiento de la producción lechera es el ARMA compuesto por 1 autorregresivo y 1 media móvil.

En la publicación de Fajardo, Romero & Ramírez [11] titulada “Un análisis de series de tiempo para la producción de carbón en Colombia” emplean la metodología de Box-Jenkins y llegan a determinar un modelo de pronósticos que se ajusta adecuadamente

a la producción de toneladas de carbón en el periodo comprendido de 1995 a 2015. Los investigadores concluyen que el modelo que explica y proyecta el comportamiento de la producción de carbón por toneladas es el SARIMA compuesto por una parte estacionaria (1, 1, 1) y una parte estacional (0, 1, 1) con un periodo estacional de 12.

Alarcón & Chuñir [12] en el trabajo titulado “Evaluación de modelos econométricos alternativos de series de tiempo para el pronóstico de la inflación en el Ecuador” aplican métodos invariados y multivariados de análisis como el modelo estacional SARIMA y el modelo de corrección de errores con variables exógenas VECX y llegan a determinar una serie de modelos con periodicidad mensual que permiten realizar proyecciones a corto plazo. Los investigadores concluyen que los modelos VECX compuesto por 2 regresores, 3 diferencias y 3 medias móviles, el SARIMA compuesto por una parte estacionaria (1, 0, 2) y un parte estacional (1, 0, 0) con un periodo de estacionalidad 12 y el SARIMA formado por una parte estacionaria (1, 0, 2) y una parte estacional (2, 0, 0) con un periodo estacional de 12, entregan proyecciones más estables que el resto de modelos y al comparar el valor promedio del error cuadrático medio del modelo VECX con 2 regresores, 3 diferencias y 3 medias móviles, con el SARIMA con una parte estacionaria (1, 0, 2) y una parte estacional (2, 0, 0) con un periodo estacional de 12, la diferencia es mínima (9,092 y 9,719), por lo que los dos modelos son útiles para realizar pronósticos.

Por su parte Heras [13] en su documento titulado “Factores determinantes y proyección de exportación de pitahaya en el Ecuador” aplica un modelo de mínimos cuadrados ordinarios (MCO) y la metodología de Box-Jenkins con lo cual logra analizar la incidencia de los factores que determinan la exportación de la pitahaya y su proyección en el mercado internacional. El investigador establece que el modelo que presenta mayor ajuste en la proyección de la exportación de pitahaya es el ARIMA compuesto por 2 regresores, 3 medias móviles y 1 diferencia.

Altamirano [14] en su obra titulada “Modelo econométrico del consumo privado para el Ecuador en el período 1990 - 2015” emplea la metodología propuesta por Box-Jenkins y consigue determinar un modelo econométrico que explica el comportamiento del consumo privado a través de la información estadística histórica. El investigador concluye que el modelo que predice y describe de forma adecuada el

consumo privado de las variables exógenas PIB e inflación es el modelo ARMA compuesto por 1 autorregresivo y 1 media móvil.

En la publicación de Damián & Sandoval [15] titulada “Modelo óptimo de pronóstico del índice mensual de producción de electricidad en Perú en el periodo 2006-2015” aplican la metodología propuesta por Box-Jenkins y llegan a determinar un modelo óptimo que predice el índice mensual de producción de electricidad. Los investigadores concluyen que el modelo SARIMA compuesto por una parte estacional (1, 1, 0) y una parte estacional (0, 1, 1) con un periodo estacional de 12, es el mejor modelo que explica y proyecta el comportamiento del índice mensual de producción eléctrica.

Pérez [16] en su trabajo titulado “Desarrollo de un modelo de estimación de depósitos monetarios para un banco privado” emplea los modelos de series de tiempo multivariados según las etapas de la metodología de Box-Jenkins y logra determinar una estimación de los depósitos monetarios mediante la comparación y validación de los modelos de series de tiempo. El investigador concluye que los modelos SARIMA son los adecuados para estimar los depósitos monetarios del banco y tienen mayor capacidad de predicción que los modelos ARIMA.

Arteaga [17] en su documento titulado “Construcción de un modelo econométrico para estimar las ventas mensuales de las bebidas gaseosas principales de la empresa Ecuador Bottling Company Corp” emplea la metodología univariada, multivariada y los vectores autorregresivos (VAR) logrando identificar un modelo econométrico que estima los pronósticos de las ventas mensuales. El investigador establece que los modelos implementados según la metodología de los vectores autorregresivos presentan menos errores de predicción y describen de forma adecuada las ventas mensuales de la empresa.

Coba [18] en su trabajo titulado “Pronóstico de ventas de la farmacéutica Sanofi usando series temporales” aplica la metodología de Box-Jenkins y llega a determinar un modelo econométrico que predice las ventas trimestrales de la farmacéutica. El investigador concluye que el modelo ARIMA formado de 6 regresores, 1 diferencia y 0 medias móviles, es el mejor modelo que describe y predice las ventas trimestrales de la farmacéutica.

En la publicación de Mora & Sempértégui [19] titulada “Proyección a corto plazo de la demanda eléctrica de los alimentadores de distribución de la empresa Eléctrica Regional Centro Sur” aplican la metodología propuesta por Box-Jenkins dentro de los procesos estocásticos autorregresivos integrados de medias móviles ARIMA y llegan a identificar un modelo matemático que predice a corto plazo la demanda de energía eléctrica de los alimentadores de distribución de la empresa. Los investigadores concluyen que el modelo ARIMA univariado es el apropiado para realizar los pronósticos de la demanda de energía eléctrica de cada alimentador de la empresa.

Angulo [20] en su documento titulado “Modelo ARIMA para el pronóstico de la liquidez monetaria mensual en el sistema financiero peruano” emplea la metodología de Box-Jenkins y llega a determinar un modelo de pronósticos para la liquidez monetaria en el periodo de enero 2003 - julio 2014. El investigador establece un modelo que describe y predice adecuadamente el comportamiento mensual de la serie de liquidez monetaria el cual es un ARIMA compuesto de 3 regresores, 1 diferencia y 12 medias móviles.

Por otro lado, Burgaentzle [21] en su obra titulada “Pronósticos y modelo de inventarios en una empresa láctea ecuatoriana” aplica la metodología de Box-Jenkins y el método de suavizamiento exponencial con lo cual consigue identificar varios modelos que proyectan la producción mensual de la empresa láctea ecuatoriana. El investigador concluye que la familia de modelos que están implementados bajo la metodología de Box-Jenkins, tienen mayor capacidad de predicción que los modelos que utilizan el suavizamiento exponencial.

Cuyo [22] en su trabajo titulado “Propuesta metodológica para el pronóstico de la demanda de agua potable en el norte de la ciudad de Quito” emplea los vectores autorregresivos y la metodología de Box-Jenkins, logra determinar una metodología eficaz para el pronóstico a corto plazo de la demanda de agua potable. El investigador concluye que la metodología adecuada para realizar los pronósticos de la demanda del agua es el método de vectores autorregresivos (VAR).

Según Aldaz & García [23] en el documento titulado “Utilidad de los modelos Box-Jenkins para el análisis y predicción de la parvovirus canina en la provincia Bolívar” aplican la metodología de Box-Jenkins y logran determinar un modelo SARIMA que evalúa la prevalencia y mortalidad de la parvovirus canina. Los investigadores

concluyen que la prevalencia y la mortalidad tienen un patrón estacional; por tal razón se pueden modelar y pronosticar con elevada precisión mediante la metodología de Box-Jenkins. El modelo ajustado para la prevalencia fue el SARIMA compuesto por una parte estacionaria (2, 0, 2) y una parte estacional (1, 0, 1), con un periodo de estacionalidad 6, mientras que para la mortalidad fue el SARIMA compuesto por una parte estacionaria (1, 0, 0) y una parte estacional (2, 0, 2) con un periodo de estacionalidad 6.

2.2. Fundamentos teóricos

2.2.1. Conceptos básicos

Modelos matemáticos

Un modelo matemático es una representación sintetizada de ecuaciones o funciones matemáticas que permiten relacionar dos o más variables de estudio. Mediante un modelo matemático se logra analizar, entender, predecir un fenómeno o las variables que se relacionaron dependiendo de los objetivos planteados por el investigador, en un modelo matemático no deben faltar las variables, los parámetros, las restricciones, la relación entre variables, las representaciones gráficas, [24].

Series de tiempo

Una serie de tiempo es una secuencia de observaciones medida en determinados momentos del tiempo, ordenados cronológicamente y espaciados entre sí de manera uniforme. Los modelos de series de tiempo tienen un enfoque netamente predictivo y en ellos los pronósticos se elaboran sólo con base al comportamiento pasado de la variable de interés [25]. Se distingue dos tipos de modelos de series de tiempo.

Modelos deterministas: Son modelos donde se conoce con certeza los datos y se tiene disponible toda la información necesaria para tomar decisiones.

Modelos estocásticos: También conocido como modelo probabilístico, es un modelo donde el tiempo (t) puede tomar cualquier valor entre 0 y ∞ . La variable t puede tomar su propia función de distribución convirtiéndose en un modelo aleatorio.

Componentes de una serie de tiempo.

En la obra de Ríos [25] se presentan los componentes de las series de tiempo:

Tendencia (T): Representa el comportamiento predominante de la serie.

Estacionalidad (E): Es un movimiento periódico que se produce dentro de un periodo corto y conocido.

Cíclica (C): Son las oscilaciones que se producen con un periodo superior al año.

Residual o Irregular (R): Son los movimientos originados por fenómenos imprevisibles como huelgas, catástrofes, que afectan a la variable de estudio.

2.2.2. Metodología de Box-Jenkins

En la publicación de George Box y Gwilym Jenkins titulada “Times series analysis forecasting and control” se da a conocer un conjunto de herramientas de predicción cuyo procedimiento se llamó metodología de Box-Jenkins, conocido también como modelo autorregresivo integrado de media móvil (ARIMA) [18].

Este método de predicción consiste en encontrar un modelo matemático que represente el comportamiento de una serie temporal de datos y realice pronósticos introduciendo únicamente el periodo de tiempo correspondiente [26].

Para Jiménez, Gázquez & Sánchez [26], la notación de los modelos ARIMA es la siguiente:

$$\text{ARIMA (p, d, q)} \quad (1)$$

donde:

p es el número de autorregresivos (AR),

d es el número de diferencias para que la serie sea estacionaria y

q es el número de medias móviles (MA).

La ecuación de los modelos ARIMA con dependencia regular es la siguiente:

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

donde:

Y_t es la variable dependiente,

ε_t es un proceso de ruido blanco,

$\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ son los parámetros del modelo [27].

Los modelos ARIMA también sirven para modelar datos estacionales. Un modelo ARIMA estacional o SARIMA se forma al incluir términos estacionales en el modelo [28].

En el documento de Villavicencio [28] la estructura del modelo SARIMA es:

$$\text{SARIMA } (p, d, q)(P, D, Q)m \quad (3)$$

donde:

(p, d, q) modela la dependencia regular de la serie,

(P, D, Q) modela la dependencia estacional de la serie,

m es el periodo estacional de la serie.

También menciona que la ecuación de un modelo SARIMA es:

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \Phi_1 Y_{t-s} + \dots + \Phi_p Y_{t-ps} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} - \dots - \Theta_1 \varepsilon_{t-s} - \dots - \Theta_q \varepsilon_{t-qs} + \varepsilon_t \quad (4)$$

donde:

Y_t es la variable dependiente,

$\varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p}$ representa la parte estacionaria AR (p),

$\Phi_1 Y_{t-s} + \dots + \Phi_p Y_{t-ps}$ representa la parte estacional del modelo SAR (P),

$\theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ representa la parte estacionaria en medias móviles MA (q),

$\Theta_1 \varepsilon_{t-s} - \dots - \Theta_q \varepsilon_{t-qs}$ representa la parte estacional en medias móviles SMA (Q) y

$\varphi_1, \dots, \varphi_p, \Phi_1, \dots, \Phi_p, \theta_1, \dots, \theta_q, \Theta_1, \dots, \Theta_q$ son los parámetros del modelo.

2.2.3. Procedimiento de la metodología Box-Jenkins

Las etapas que se debe seguir para la elaboración de un modelo ARIMA, con el fin de realizar pronósticos son las siguientes.

ETAPA 1. Identificación.

En la identificación del modelo se determina si la serie es estacionaria, para ello se utiliza el gráfico de la serie, los correlogramas y la prueba de raíz unitaria.

Gráfico de la serie. El análisis visual de la trayectoria de la serie a lo largo del tiempo da una idea de si es o no estacionaria. Si existe un valor en torno al cual va oscilando la serie y no se aleja de forma permanente, entonces se puede considerar que la serie es estacionaria. [3]

Análisis de los correlogramas. Son las gráficas de las funciones de autocorrelación simple y parcial que ayudan a determinar si una serie es estacionaria. Mediante estos diagramas se puede elegir los parámetros que forman un modelo ARIMA [18].

Función de autocorrelación (acf). Esta gráfica identifica la correlación entre una serie y sus retardos, mediante esta función se determina el orden de las medias móviles del modelo y se representa por $Ma(q)$ [29].

Función de autocorrelación parcial (pacf). Con esta gráfica se identifica la correlación entre una variable y sus retrasos que no se explica por rezagos anteriores. Mediante pacf se encuentra el orden de los autorregresivos del modelo y se representa por $AR(p)$ [29]

Prueba de raíz unitaria de Dickey Fuller (ADF). Es un procedimiento que ayuda a comprobar la existencia de una raíz unitaria en las series de tiempo. La hipótesis a probar es $H_0: \rho = 1$; para ello se utiliza el siguiente estadístico.

$$t = \frac{\rho - 1}{S / \sqrt{\sum_{t=1}^T Y_{t^2-1}}} \quad (5)$$

donde:

ρ es el coeficiente obtenido de la regresión,

S es el error estándar de la regresión y

Y_{t^2-1} es la variable desfasada [3].

Si el valor calculado del estadístico t es menor al valor crítico de Dickey Fuller, se rechaza la hipótesis nula y la serie de tiempo no tiene raíz unitaria y es estacionaria,

pero si es mayor al valor crítico no se rechaza la hipótesis, la serie no es estacionaria y será necesario diferenciarla.

El contraste de las hipótesis es:

La hipótesis nula H_0 : hay raíz unitaria (La serie no es estacionaria).

La hipótesis alternativa H_a : no hay raíz unitaria (La serie es estacionaria).

Se rechaza H_0 si $ADF < \text{valor crítico}$

Proceso de diferenciación

La mayoría de las series temporales no son estacionarias, es decir la media y la varianza no son constantes a lo largo del tiempo y es necesario transformarlas o diferenciarlas para que se conviertan en estacionarias. Para eliminar la tendencia de una serie aleatoria se utiliza la diferenciación, que consiste en restarle a cada dato el anterior, si la tendencia no se ha eliminado por completo al realizar la primera diferenciación, se realiza una segunda diferencia.

$$\Delta Y_t = Y_t - Y_{t-1} \quad (6)$$

Etapa 2. Estimación.

Una vez identificado el modelo y los valores apropiados de p , d y q , la siguiente fase es la estimación de los parámetros autorregresivos y de promedio de medias móviles que están dentro del modelo.

La estimación de modelos ARIMA es compleja y utiliza algoritmos de optimización no lineales para calcular los valores de los parámetros; para esto se utiliza el método de máxima verosimilitud. Este método maximiza la probabilidad de los parámetros de las funciones de densidad que dependen de la distribución de probabilidad y las observaciones de la muestra [30].

Una vez realizada la estimación se elegirá el modelo que tenga el error cuadrático medio (ECM), el criterio de información AKAIKE (AIC) y el criterio de información Bayesiano (BIC) más pequeño [3].

Criterio de información AKAIKE (AIC)

$$AIC = \log \sigma^2 + n + 2k/n \quad (7)$$

donde:

σ^2 es la suma cuadrática de los residuos sobre n,

k es el número de parámetros en el modelo y

n es el número de observaciones.

Criterio de información AKAIKE corregido (AICc)

$$AICc = \log \sigma^2 + n + 2k * n - k - 2 \quad (8)$$

donde:

σ^2 es la suma cuadrática de los residuos sobre n,

k es el número de parámetros en el modelo y

n es el número de observaciones.

Criterio de información Bayesiano (BIC).

$$BIC = \log \sigma^{k2} + (k * \log n)/n \quad (9)$$

donde:

σ^{k2} es la suma cuadrática de los residuos sobre n,

k es el número de parámetros en el modelo y

n es el número de observaciones

Etapa 3. Validación.

En la validación se analiza y se decide si el modelo es estadísticamente adecuado. Para determinar la significancia del modelo se debe comprobar que los residuos se comporten como un ruido blanco y que sigan una distribución normal [31].

Ruido blanco. Es una serie de datos que se caracteriza por tener distribución normal alrededor de una media de valor cero y una varianza constante [32].

Las series de tiempo que débilmente están autocorrelacionadas se denominan ruido blanco, para demostrar que los residuos de una serie se comportan como un ruido blanco se aplica el test de Ljung Box.

Test Ljung Box (LB). Se usa comúnmente para probar la calidad del ajuste de un modelo de series de tiempo. Si no se encuentra una autocorrelación significativa en los residuos del modelo, entonces se declara que los residuos se comportan como un ruido blanco.

Quesada [30] define al estadístico Ljung-Box de la siguiente manera:

$$Q = n(n + 2) \sum_{h=1}^m \frac{\rho_h^2}{n - r} \quad (10)$$

donde:

Q es el estadístico que se distribuye como una ji-cuadrada.

ρ_h es el coeficiente de autocorrelación de los residuos estimados,

n es el número de valores de la serie y

r es el número de parámetros estimados.

El contraste de las hipótesis es:

Ho: Los datos se distribuyen de forma independiente (existe ruido blanco).

Ha: Los datos no se distribuyen de forma independiente (no existe ruido blanco).

Si el p-value calculado del estadístico es mayor que el nivel de significancia prefijado (0,05) de Ljung Box, no se rechaza la hipótesis nula y existe ruido blanco.

Test de Jarque Bera (JB) para la normalidad. Es una prueba de bondad de ajuste que comprueba si una muestra de datos tiene la asimetría y la curtosis de una distribución normal. Esta prueba estadística se define como:

$$JB = n \left[\frac{A^2}{6} + \frac{(K - 3)^2}{24} \right] \quad (11)$$

donde:

n es el tamaño de la muestra,

A es la asimetría,

K es la curtosis.

El contraste de las hipótesis es:

Hipótesis nula H_0 : se distribuyen de forma normal.

Hipótesis alternativa H_a : no se distribuye de forma normal

La hipótesis nula de normalidad se rechaza cuando $JB > JB_{crítica}$. [33]

Si se determina que el modelo elegido es el adecuado se puede calcular las predicciones. Caso contrario se estudia el comportamiento de los residuos estimados que ayudan a identificar un nuevo modelo y se vuelve a la etapa 1 repitiendo todo el proceso [30].

Etapa 4. Pronóstico

Una vez determinado que la serie de tiempo es estacionaria, cumple los requisitos de varianza, autocorrelación y normalidad, se tendrá un modelo ARIMA adecuado. Con el modelo estimado y comprobada la bondad de ajuste, se llevan a cabo los pronósticos para uno o varios periodos.

CAPÍTULO III

MARCO METODOLÓGICO

3.1. Ubicación

Salinas se encuentra ubicada en la provincia de Bolívar a 35 minutos de la ciudad de Guaranda. Es una de las parroquias más prósperas del cantón y se ha convertido en un ejemplo de iniciativas comunitarias. Salinas está rodeada de prados y montañas del hermoso valle andino a una altura de 3550 metros [34]

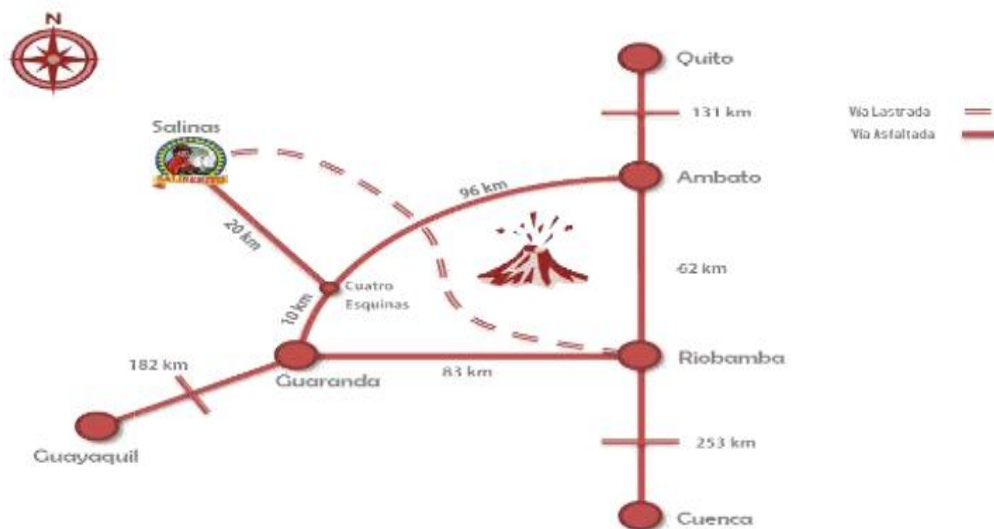


Figura 1. Ubicación de la parroquia Salinas
Fuente: Fundación Familia Salesiana Salinas

Para Tour Salinerito [34], la parroquia Salinas es un conjunto de microempresas. Es muy usual encontrar en las viviendas de la parroquia tiendas y almacenes que expenden los productos que generan sus pobladores como: Quesos, mantequilla, yogurt, turrone, mermeladas, fideos, galletas, trufas, hongos secos, pan, caramelos de miel de abeja, chocolates, chocolatines, lana, tejidos, balones, artesanías, etc.

Reseña histórica de Confites El Salinerito

La fábrica de Confites El Salinerito empezó en un taller agroindustrial en 1989 con la elaboración de turrone. Esta actividad se origina a raíz de un proyecto exitoso de apicultura a nivel provincial. Según Luis González uno de los fundadores de la fábrica manifiesta que las religiosas españolas que llegaron a Salinas fueron quienes instruyeron en la preparación de turrone. Las primeras pruebas eran realizadas por las

mujeres salineras en pailas de cobre y carbón, trabajo que era bastante pesado debido a la fuerza requerida para efectuar la mezcla [35].

Posteriormente por los años 90 se aplicaron las tinas de doble fondo de acero; el batido se realizaba manualmente con paletas de madera en un lapso de tres horas llegando a fabricar ochenta turronec diarios y para el envasado las fundas se confeccionaban manualmente. Actualmente, la fábrica posee una turronec mecánica con mayor capacidad de producción, logrando disminuir el tiempo de elaboración y requiriendo menos esfuerzo para los empleados que la operan. Este tipo de maquinaria llegó de Italia y actualmente se está reproduciendo en la mecánica de Funorsal.

Cuando el proyecto crece fructíferamente los directivos de ese entonces efectuaron un crédito de 4 millones de sucres para la creación de la primera planta de turronec, establecimiento donde se inició la producción de chocolate y que actualmente sigue funcionando. La intención de producir chocolate a 3500 metros de altura salía del contexto conceptual de Salinas. Sin embargo, con la llegada de un técnico italiano quien pertenecía a la red de Ferrero Rocher se logró concretar el proyecto chocolatero; las trufas de chocolate rellenas o “besos salineros” fue el primer producto en salir a la venta después de los turronec. Luego gracias al asesoramiento de los esposos Russhiem (técnicos suizos) se elaboraron nuevos productos en base al chocolate [35].

La elaboración de chocolates empezó como un complemento para los turronec y termino convirtiéndose en la actividad principal de la empresa Confites El Salinerito.

Producción

Todos los productos de la empresa Confites El Salinerito han sido desarrollados como proyectos pilotos y de experimentación. Actualmente, son cuatro los productos que se ofrece al mercado: chocolates, turronec, pasta de cacao y mermeladas. Los aportes para la creación de nuevos productos son de personas voluntarias extranjeras o turistas que visitan la fábrica. El producto perlas pasas es una contribución de un turista chileno. Los chocolates y turronec se clasifican en grupos según su elaboración e incluyen varias presentaciones [35].

Localización de la Empresa Confites el Salinerito

La Empresa Confites El Salinerito se encuentra ubicada en la Calle el Calvario y Samilagua s/n dentro de la parroquia Salinas de Guaranda.



Figura 2. Ubicación de la Empresa Confites El Salinerito
Fuente: Fundación Familia Salesiana Salinas

3.2. Equipos y materiales

La metodología empleada para la construcción del modelo matemático de producción de la empresa Confites El Salinerito es la propuesta por Box-Jenkins en 1970 y comprende de:

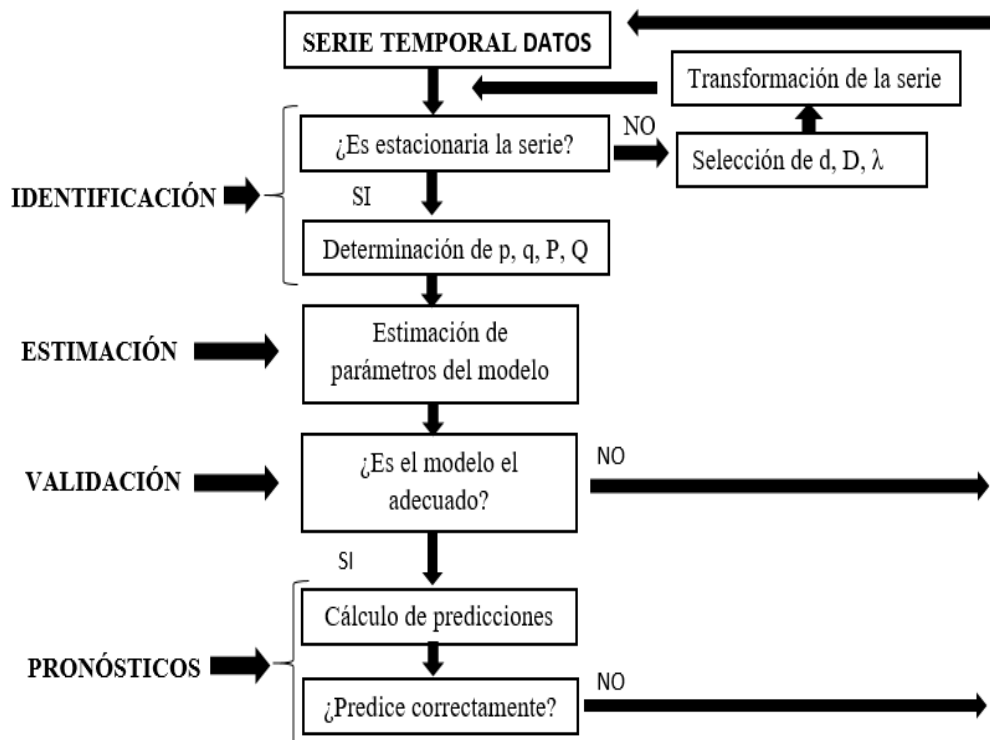


Figura 3. Metodología Box-Jenkins
Fuente: Gujarati & Porter. 2009. Pag. 778

Los equipos y materiales utilizados en la presente investigación fueron:

Datos de la producción mensual de la empresa Confites El Salinerito.

Una portátil Toshiba con un procesador Intel (R) Core (TM) i5 de tercera generación, con una RAM de 6GB.

Un celular Xiaomi Redmi Note 8 con un procesador Snapdragon 665 y una memoria RAM de 4GB.

Herramientas informáticas como: el software libre RStudio versión 4.0.1 donde se realizó el código informático que modela la producción de la empresa. También se utilizó Microsoft Word 2013 para la escritura del proyecto de investigación y Microsoft Excel 2013 se utilizó para la lectura de la base de datos, además esta herramienta sirvió para analizar y comparar los resultados que arroja el modelo construido y la metodología que utiliza la empresa para establecer su producción mensual.

3.3. Tipo de investigación

La investigación realizada es de tipo descriptiva, predictiva y longitudinal con tendencia.

Investigación descriptiva

La investigación descriptiva se enfoca en la caracterización de un hecho, fenómeno, individuo o grupo, con el objetivo de analizar su comportamiento. [36]

La investigación descriptiva ayudó al análisis detallado del comportamiento histórico de los valores de producción mensual de la empresa Confites El Salinerito.

Investigación predictiva.

La investigación predictiva tiene como propósito prever o anticipar situaciones futuras, requiere de la exploración, la descripción, la comparación, el análisis y la explicación [36]. El presente estudio se apoyó en este tipo de investigación ya que se quiere identificar, analizar y pronosticar la cantidad de kilogramos que debe producir mensualmente la empresa.

Investigación longitudinal con tendencia

En la investigación longitudinal con tendencia la unidad de análisis es observada en varios intervalos de tiempo. Los datos son recolectados a través del tiempo en periodos específicos para hacer inferencias acerca de la evolución, sus causas y sus efectos [37].

Se aplicó la investigación longitudinal con tendencia para estudiar, observar y analizar los procesos de cambio de la variable dependiente (cantidad de kilogramos producidos) de forma directa en el tiempo.

3.4. Problema e hipótesis

Problema

¿Qué modelo describe adecuadamente el comportamiento de la producción mensual de la empresa Confites El Salinerito?

Hipótesis

El modelo SARIMA (1, 1, 1) (1, 1, 1)₁₂ desarrollado según la metodología de Box-Jenkins permite realizar un pronóstico correcto de la producción mensual de la empresa.

3.5. Población o muestra

La población de estudio está constituida por la base de datos digitalizada que facilitó la empresa Confites El Salinerito.

La muestra utilizada está conformada por los valores de producción mensual de la empresa Confites El Salinerito, comprendida en el periodo de enero 2017 a julio 2020.

3.6. Recolección de información

En la presente investigación la información es obtenida directamente de la base de datos digitalizada de la empresa Confites El Salinerito. Dentro de la base de datos se encuentran variables como: el tiempo, cantidad de kilogramos producidos y cantidad de kilogramos vendidos mensualmente en el periodo enero 2017 - julio 2020.

3.7. Procesamiento y análisis estadístico

Para el procesamiento de los datos se utilizó los siguientes softwares estadísticos:

Microsoft Excel 2013: Mediante esta herramienta informática se analizó y se seleccionó los datos para la construcción del modelo matemático. También sirvió para registrar los valores del modelo SARIMA y el modelo actual de la empresa, logrando así realizar comparaciones entre los modelos.

Software RStudio versión 4.0.1: Con esta herramienta estadística se realizó el modelamiento de la serie de producción basándose en la metodología de Box-Jenkins. Mediante este software se desarrolló el código informático y se hizo pronósticos a corto plazo.

Para el análisis estadístico se empleó la metodología de Box-Jenkins, que comprende de las siguientes etapas.

Etapas 1. Identificación del modelo

En esta etapa se determinó si la serie es estacionaria; para ello se utilizó el análisis gráfico y el test de Dickey-Fuller. Si la serie no es estacionaria se aplica como máximo 2 diferencias. Mediante los correlogramas acf y pacf se determinó los valores de (p, d, q) , (P, D, Q) y m apropiados para formular los modelos SARIMA tentativos.

Etapas 2. Estimación de los parámetros del modelo

Para estimar los parámetros del modelo propuesto se empleó la herramienta estadística RStudio que utiliza la estimación de máxima verosimilitud. Se eligió este programa por la capacidad que tiene para el análisis estadístico y econométrico, tanto univariante como multivariante, permitiendo hacer pronósticos y los resultados se muestran como datos.

Para seleccionar el modelo SARIMA apropiado se observó el modelo que se ajuste con mayor exactitud a los valores de producción y en especial se centró en el modelo que tiene el menor valor en el criterio de información AKAIKE, criterio de información bayesiano (BIC) y el error cuadrático medio (ECM).

Etapas 3. Validación del modelo

Para validar el modelo seleccionado se aplicó el test de Jarque Bera del supuesto de normalidad, los correlogramas acf y pacf de los residuos y el test de Ljung-Box para comprobar que los residuales se comportan como un ruido blanco.

Etapa 4. Pronósticos con el modelo

Ya seleccionado el modelo SARIMA (p, d, q) (P, D, Q) m, e identificado la bondad y significancia se procedió a pronosticar para el periodo agosto 2020 a enero 2021.

3.8. Variables respuestas o resultados alcanzados

Dentro de la base de datos digitalizada se encontró las variables como: el tiempo, tipo de unidad, cantidad de kilogramos producidos mensualmente, cantidad de kilogramos vendidos mensualmente y las ventas. En este estudio la variable dependiente es la cantidad de kilogramos producidos mensualmente y la variable independiente es el tiempo que está medido en meses dentro del periodo enero del 2017 a julio 2020.

Para realizar la predicción de la cantidad de kilogramos a producirse mensualmente se diseñó un modelo matemático SARIMA, basado en la metodología de Box-Jenkins.

Una vez seleccionado el modelo matemático que se ajusta con mayor exactitud a los valores de producción, se realizó la validación de los pronósticos. Para ello se dividió a la muestra conformada de 43 datos en dos partes: el entrenamiento y testeo. En el entrenamiento existen 36 datos comprendidos desde enero 2017 a diciembre del 2019, que se emplearon para el proceso de estimación y pronóstico. El testeo comprendió el periodo enero a julio 2020 y consta de 7 datos, sirvió para comparar y validar el modelo. Luego se comparó los pronósticos generados en el entrenamiento con los valores del testeo (valores reales de la serie), al no existir grandes diferencias en sus resultados se verificó la eficiencia del modelo.

Ya confirmado el correcto funcionamiento del modelo se realizó las predicciones para seis meses. La producción en kilogramos para los meses de pronóstico es la siguiente: agosto 8223 kg, septiembre 5691 kg, octubre 5458 kg, noviembre 6939 kg, diciembre 8668 kg y en enero 6314 kg.

CAPÍTULO IV

RESULTADOS Y DISCUSIONES

4.1. Estructura de la base de datos

Para la construcción del modelo matemático se utilizó los valores históricos de producción de la empresa Confites El Salinerito en el periodo de enero 2017 a julio 2020. La estructura de los valores históricos se presenta en la siguiente tabla:

Tabla 1. Estructura de la base de datos

Variables		Características			
Tiempo (meses)	chr [1:43]	Enero	Febrero	Marzo	Abril...
Cantidad de kilogramos producidos	num [1:43]	5956	8543	7645	7893...
Cantidad de kilogramos vendidos	num [1:43]	5432	6341	10945	9578...

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 1 se presentan las variables y sus características. Cada variable consta de 43 datos, el tiempo es una variable de tipo categórica medida en meses y las variables cantidad de kilogramos producidos y cantidad de kilogramos vendidos son de tipo numérico; para el diseño del modelo matemático se emplea la metodología de Box-Jenkins, en donde la variable independiente es el tiempo y la variable dependiente es la cantidad de kilogramos producidos.

4.2. Aplicación de la metodología Box-Jenkins

4.2.1. Identificación del modelo

Componentes de los valores de producción históricos.

Una forma de identificar la tendencia, estacionalidad y estacionariedad de una serie de tiempo es a través de sus componentes. Por tal razón, se determina y se presenta los componentes de los valores históricos de la producción mensual de la empresa en el periodo enero 2017 a julio 2020.

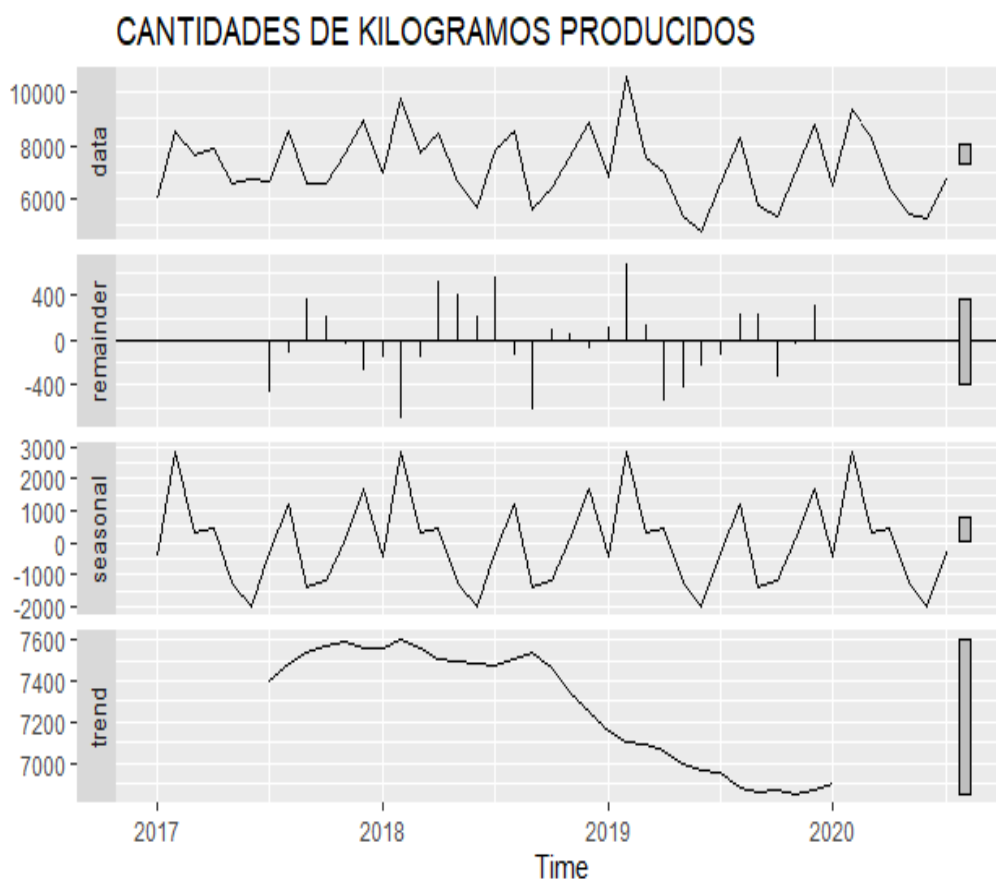


Figura 4. Componentes de los valores históricos

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

La figura 4 está compuesta por cuatro gráficos: el comportamiento de los valores de producción, los residuos, la estacionalidad y la tendencia. En el primer gráfico se observa el comportamiento y la información sobre los ciclos de la serie de los valores de producción. Dentro del gráfico de los residuos se identifica que la varianza no es constante en el tiempo, por tal motivo los valores históricos no son estacionarios; en el gráfico de la estacionalidad se aprecia que existe un patrón estacional que se repite cada 12 meses. Por último, en el cuarto gráfico se observa la tendencia de los valores de producción y tienen un comportamiento decreciente. Esta tendencia decreciente indica que los valores históricos no son estacionarios.

Prueba de raíz unitaria en los valores históricos

Para realizar la prueba de raíz unitaria y verificar si una serie de tiempo es estacionaria se aplica el test de Dickey-Fuller aumentada. Esta prueba estadística utiliza un nivel de significancia del 5% con un contraste de hipótesis:

$H_0: p\text{-value} > 0,05$ no existe estacionariedad.

Ha: $p\text{-value} < 0,05$ existe estacionariedad.

Tabla 2. Test de Dickey–Fuller Aumentada.

Test de Dickey-Fuller Aumentada			
data: produce			
Dickey-Fuller	$t = -3,1634$	Lag order = 3	p-value = 0,1141
alternative hypothesis: stationary			

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 2 se observa que el p-value del estadístico de Dickey–Fuller aumentada es de 0,1141 y es superior al nivel de significancia prefijado de 0,05 por lo que se concluye que los valores históricos de producción no son estacionarios y presentan raíz unitaria. Por lo que se hace necesario transformarlos u diferenciarlos para volverlos estacionarios y así aplicar la metodología de Box-Jenkins. Con el comando `ndiffs` se identifica el número de diferencias que se debe aplicar para obtener una serie estacionaria y con el comando `nsdiffs` se identifica el número de diferencias que debe llevar en la parte estacional.

En la figura 4 y la tabla 2 se determina que los valores históricos de producción no son estacionarios, por lo tanto, es necesario aplicar una primera diferencia para volverlos estacionarios. La serie diferenciada se presenta en la siguiente figura:

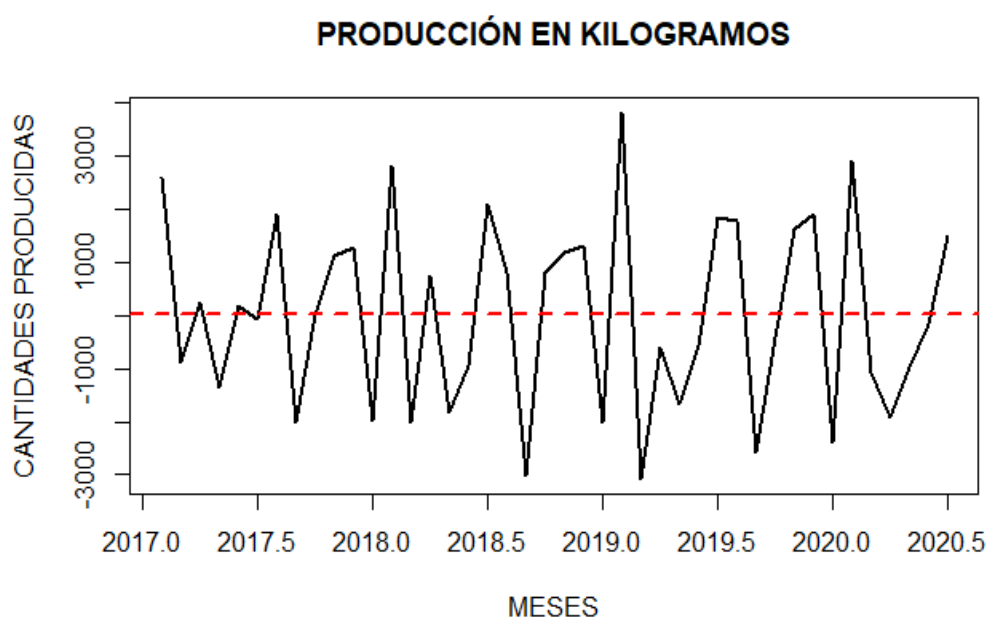


Figura 5. Comportamiento de los valores con la primera diferencia
Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito.

En la figura 5 se observa el comportamiento de los valores históricos de producción de la empresa con una primera diferencia, donde el valor de la media es 0 y la varianza está comprendida entre -3000 y 3000, con lo cual se verifica que la media y la varianza permanecen constantes a lo largo del tiempo. También podemos decir que los valores diferenciados no presentan signos de tendencia y por ende son estacionarios.

Para verificar si los valores históricos de producción con una primera diferencia son estacionarios, es necesario aplicar nuevamente el test de Dickey-Fuller Aumentada.

Tabla 3. Test de Dickey –Fuller Aumentada con una primera diferencia

Test de Dickey-Fuller Aumentada		
data: produce		
Dickey-Fuller = - 4,0687	Lag order = 3	p-value = 0,0165
alternative hypothesis: stationary		

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 3 se observa que el p-value del estadístico de Dickey-Fuller aumentada es de 0,0165 y es menor al nivel de significancia del 0,05, por tal razón los valores diferenciados son estacionarios y la serie de valores de producción no presenta raíz unitaria. Una vez conseguido que los valores históricos de producción cumplan con la estacionariedad, el siguiente paso es determinar los valores adecuados (p, d, q) (P, D, Q) m que forman el modelo.

Correlogramas de la serie histórica diferenciada

Para determinar los valores adecuados de (p, d, q) (P, D, Q) m, se utiliza la función de autocorrelación simple (acf) y la función de autocorrelación parcial (pacf). En la siguiente figura se muestra el comportamiento de las funciones de autocorrelación simple y parcial con una primera diferencia.

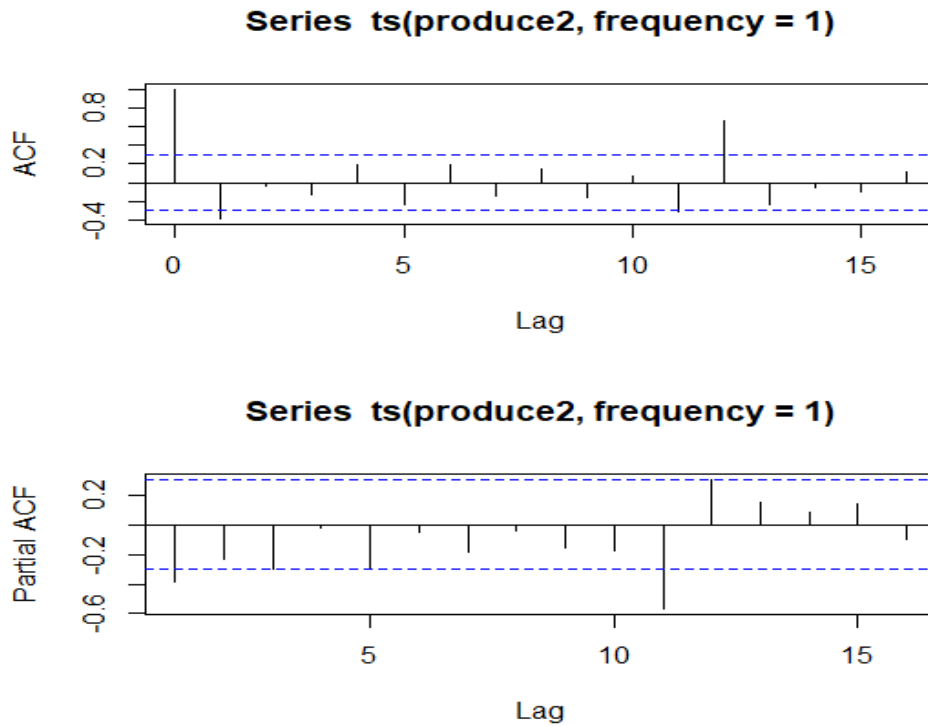


Figura 6. Función de autocorrelacion simple y parcial de la serie diferenciada
 Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la figura 6 se presentan los correlogramas de las funciones de autocorrelación simple y parcial. En la función de autocorrelación simple (acf) se observa que existen dos rezagos significativos, por lo que se sugiere implementar un modelo estacional que tenga una media móvil de orden $Q = 1$ o $Q = 2$. Mientras en la función de autocorrelación parcial (pacf) se visualiza que existe un rezago significativo y se propone un modelo estacional que tenga un autorregresivo de orden $P = 1$. Según las funciones acf y pacf debemos ajustar los modelos con un orden $p > 0$ y $q > 0$ para el comportamiento estacionario de la serie, a partir de esto los valores sugeridos para p y q que debe contener el modelo SARIMA es una media móvil de orden $p = 1$ y un autorregresivo de orden $q = 1$.

Modelos tentativos identificados

A través de las funciones acf y pacf se determina seis modelos tentativos. Para seleccionar el modelo adecuado se utiliza el valor del error cuadrático medio (ECM) y también los valores de los criterios de información AKAIKE (AIC) y Bayesiano (BIG). A continuación, se presenta los seis modelos seleccionados:

Tabla 4. Modelos tentativos

N°	Modelos	ECM	AIC	BIG
Mod1	AUTOARIMA (0,0,0) (0,1,0)12	582249	501,49	502,92
Mod2	SARIMA (2,1,2) (1,1,1)12	525817	493,69	503,50
Mod3	SARIMA (1,1,2) (1,1,1)12	495420	491,98	500,38
Mod4	SARIMA (2,1,1) (1,1,1)12	543789	492,27	500,67
Mod5	SARIMA (0,1,0) (1,1,1)12	592758	497,21	501,41
Mod6	SARIMA (1,1,1) (1,1,1)12	488744	491,04	498,05

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 4 se encuentra el resumen de los seis modelos seleccionados. Para determinar el modelo que se ajusta con mayor exactitud a los valores de producción, se elige el modelo que tiene el error cuadrático medio (ECM), el criterio de información AKAIE (AIC) y el criterio de información Bayesiano (BIC) más pequeño. Al realizar la comparación entre los modelos tentativos se concluye que el mejor modelo según ECM, AIC y BIC es el Mod6 SARIMA (1, 1, 1) (1, 1, 1)12, formado por una media móvil tanto en la parte estacionaria como en la parte estacional, por un regresor en la parte estacionaria y en la parte estacional, contiene una diferenciación y un periodo de estacionalidad de 12.

La siguiente tabla presenta el resumen del modelo SARIMA elegido.

Tabla 5. Resumen del modelo SARIMA (1, 1, 1) (1, 1, 1) 12

SARIMA (1,1,1) (1,1,1)12						
	ar1	ma1	sar1	sma1		
Coefficientes	0,1132	- 0,7768	0,5549	- 0,9999		
Error estándar Coef.	0,2595	0,1881	0,3311	0,6251		
ECM = 488744:	Probabilidad de registro = - 240,52					
AIC = 491,04	AICc = 493,54			BIC = 498,05		
Medidas de error en el conjunto de entrenamiento:						
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-97,13	543,61	359,45	-1,95	5,34	0,5975	-0,030

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 5 se observa la información del modelo seleccionado, donde se identifica los coeficientes del modelo, el error estándar de cada coeficiente, el error cuadrático medio, los criterios de información AKAIKE, Bayesiano y los errores de entrenamiento del modelo. El modelo SARIMA (1, 1, 1) (1, 1, 1) 12 será sometido a pruebas de validación.

4.2.2. Estimación del modelo

Para la estimación de los parámetros del modelo seleccionado se emplea el software libre RStudio que utiliza la estimación de máxima verosimilitud. Aplicada esta herramienta informática se identifica los coeficientes estimados, el error estándar de cada coeficiente, el estadístico z y la probabilidad del estadístico z, como se muestra en la siguiente tabla.

Tabla 6. Estimación del modelo ARIMA (1, 1, 1) (1, 1, 1)12

Parámetro	Coefficientes	Error estándar	Estadístico z	Probabilidad
ar1	0,1132	0,2595	2,4364	0,0066253
ma1	- 0,7768	0,1881	-4,1298	0,0000363
sar1	0,5549	0,3311	3,6756	0,0009381
sma1	- 0,9999	0,6251	-2,5956	0,0011050

Fuente: Elaborado por el autor en base a los datos de Confitos El Salinerito

En la tabla 6 se presenta la estimación de los parámetros del modelo. Se observa que la probabilidad de cada parámetro es inferior a 0,05, por tal razón se concluye que los parámetros son significativos y sirven para formar la ecuación del modelo.

Según la ecuación 4, el modelo SARIMA (1, 1, 1) (1, 1, 1) 12 se define de la siguiente forma:

$$Y_t = \varphi_1 Y_{t-1} + \Phi_1 Y_{t-12} - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

donde:

Y_t es la cantidad de kilogramos producidos mensualmente,

$\varphi_1, \Phi_1, \theta_1, \Theta_1$ son los parámetros estimados del modelo y

ε_t es el proceso de ruido blanco.

Reemplazando los coeficientes estimados del modelo según la tabla 6 la ecuación de pronósticos estimada se expresa de la siguiente forma:

$$\Delta Y_t = 0,1132Y_{t-1} + 0,5549Y_{t-12} + 0,7768\varepsilon_{t-1} + 0,9999\varepsilon_{t-12} + \varepsilon_t$$

4.2.3. Validación del modelo

Realizada la estimación la siguiente etapa es validar el modelo. Dentro de esta etapa se utiliza el diagnóstico del modelo, gráfico de los residuos, prueba de Jarque Bera y el test de Ljung-Box. A continuación, se presenta la figura del diagnóstico del modelo.

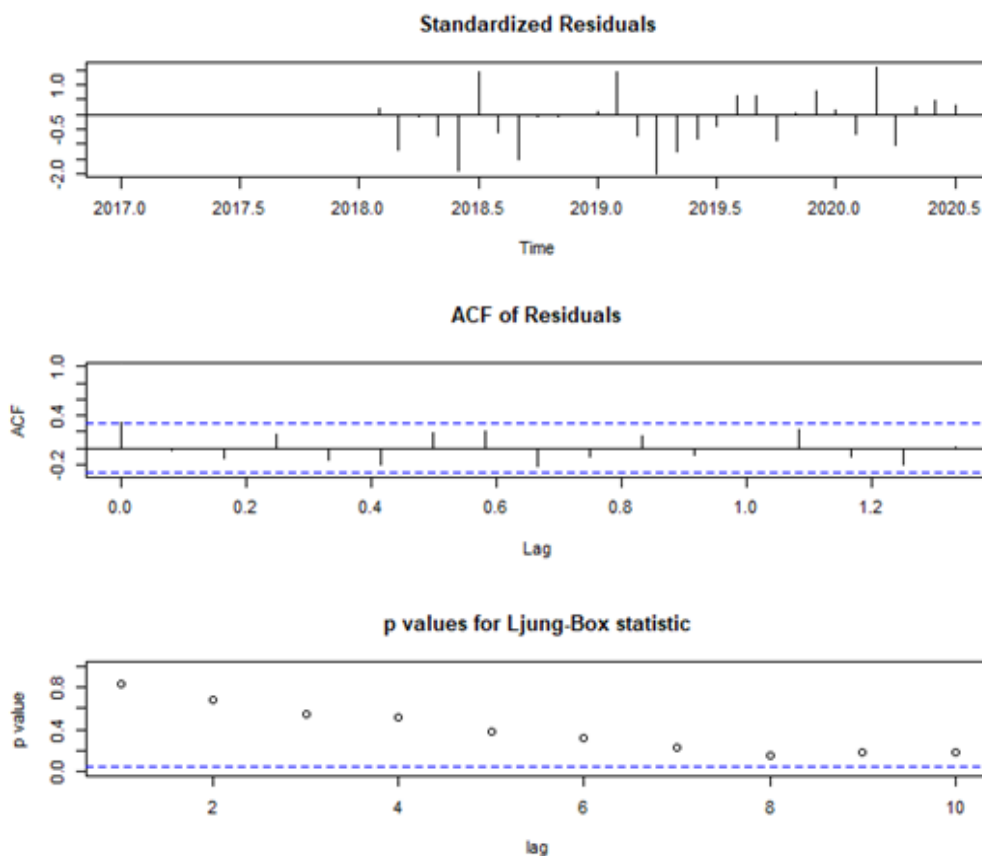


Figura 7. Diagnóstico del modelo SARIMA (1, 1, 1) (1, 1, 1)12

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la figura 7 se observa los errores estandarizados, la función de autocorrelación acf de los residuos y p-value del test de Ljung-Box. En el gráfico de los errores estandarizados se observa el comportamiento de los residuos del modelo. En la gráfica de la función acf se visualiza que no existe una correlación significativa en los residuos, su valor es pequeño y los retardos están dentro de la banda de confianza. Por último, en el gráfico del p-value de Ljung-Box se nota que los errores del modelo están

por arriba del nivel de significancia del 0,05 y se afirma que los residuos se comportan como ruido blanco.

De igual manera se presenta el gráfico de los residuos del modelo para determinar si los datos de la serie son aleatorios y los residuos generan un ruido blanco.

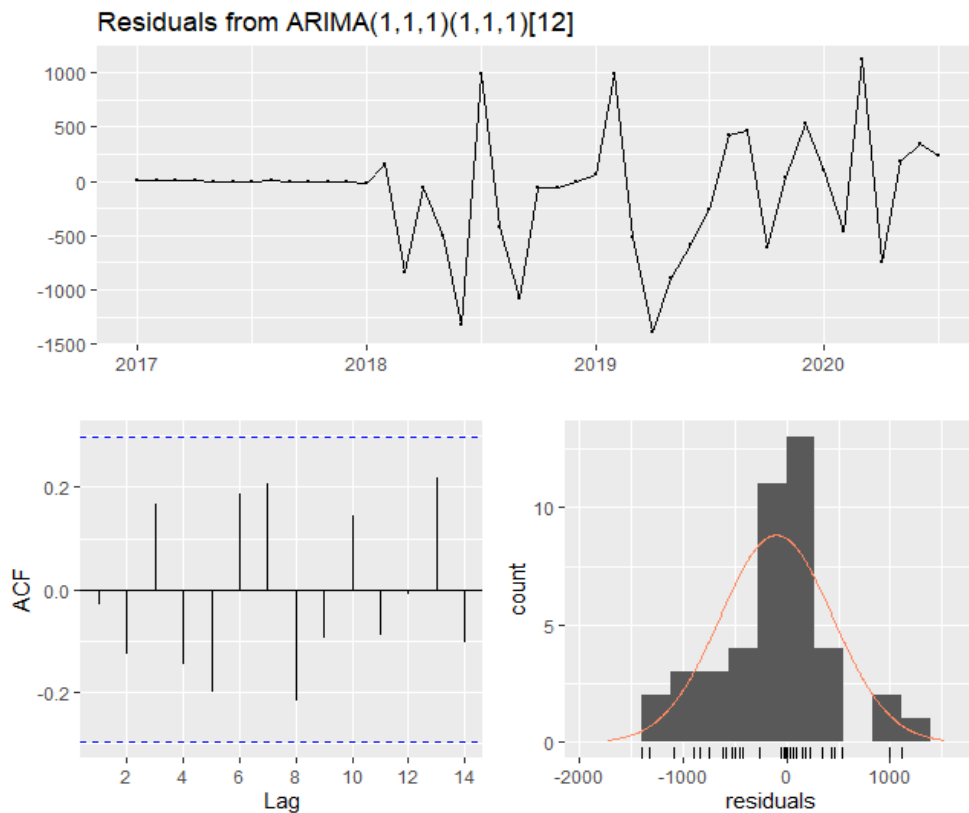


Figura 8. Gráfico de los residuos del modelo

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la figura 8 se visualiza tres gráficos. El primer gráfico presenta el comportamiento de los residuos que generan un ruido blanco. En el segundo gráfico se observa la función de acf de los residuos, evidenciando que ningún retardo es significativo y están dentro de los intervalos de confianza, por tal razón se concluye que los datos de la serie son aleatorios. Por último, en la tercera gráfica se observa el histograma y la curva de la distribución normal, donde se evidencia que los residuos se distribuyen de manera normal.

También se aplica las pruebas estadísticas como: el Test de normalidad Jarque Bera y el Test de ruido blanco Ljung- Box. Para comprobar que los residuos siguen una distribución normal se aplica el test de normalidad Jarque Bera que utiliza un nivel de significancia prefijado del 5% con un contraste de hipótesis:

Ho: $p\text{-value} > 0,05$ Existe normalidad.

Ha: $p\text{-value} < 0,05$ No existe normalidad.

Tabla 7. Test de normalidad Jarque Bera

Test Jarque Bera		
data: residuos (mod2)		
X-squared = 0,85146,	df = 2,	p-value = 0,6533

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 7 se presenta el test de normalidad Jarque Bera, donde se observa que el p-value es igual a 0,6533 y es superior al nivel de significancia del 0,05. Por lo tanto, se afirma que los residuos del modelo siguen una distribución normal.

Para verificar si los residuos del modelo se comportan como un ruido blanco se aplica el Test Ljung-Box con un nivel de significancia de 5% con un contraste de hipótesis:

Ho: $p\text{-value} > 0,05$ Existe ruido blanco.

Ha: $p\text{-value} < 0,05$ No existe ruido blanco.

Tabla 8. Test de ruido blanco Ljung-Box

Test Ljung- Box		
data: residuals(mod2)		
X-squared = 0,041806,	df = 1,	p-value = 0,838

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 8 se observa que el p-value del estadístico Ljung-Box tiene un valor de 0,838 y es superior al nivel de significancia prefijado del 0,05; por tal razón, se afirma que existe ruido blanco en los residuos.

Con los resultados de las pruebas estadísticas aplicadas al modelo elegido se concluye que está bien ajustado y es el adecuado para realizar la predicción de la cantidad de kilogramos a producirse mensualmente.

Validación de los pronósticos

Con el modelo validado es necesario medir la calidad de predicción realizada, para esto se divide a la muestra en dos partes el entrenamiento y testeo. Con el entrenamiento se realiza el pronóstico y el testeo sirve para realizar las comparaciones. En la siguiente tabla se presenta la validación de los pronósticos.

Tabla 9. Valores de entrenamiento y testeo

Meses	Valor real (kg)	Valor pronosticado (kg)
Enero 2020	6451	6814
febrero 2020	9343	10480
Marzo 2020	8281	7396
Abril 2020	6362	7166
Mayo 2020	5382	5363
Junio 2020	5205	4501
Julio 2020	6702	6777
Promedio	6818	6928

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 9 se presenta los valores reales y los valores pronosticados. Para la validación de los pronósticos se realiza una comparación entre los valores de testeo (valores reales) con los de entrenamiento (valores pronosticados) evidenciándose que no existe una marcada diferencia entre ellos. Según el promedio del entrenamiento y testeo, los valores pronosticados exceden en un 1,61% a los valores reales, este porcentaje nos indica que los pronósticos exceden en 110 kg a los reales, entonces se dice que el modelo planteado está bien ajustado y es el adecuado para realizar pronósticos para un periodo de corto plazo.

Para entender el comportamiento de los pronósticos se presenta el gráfico de los valores reales y los valores pronosticados.

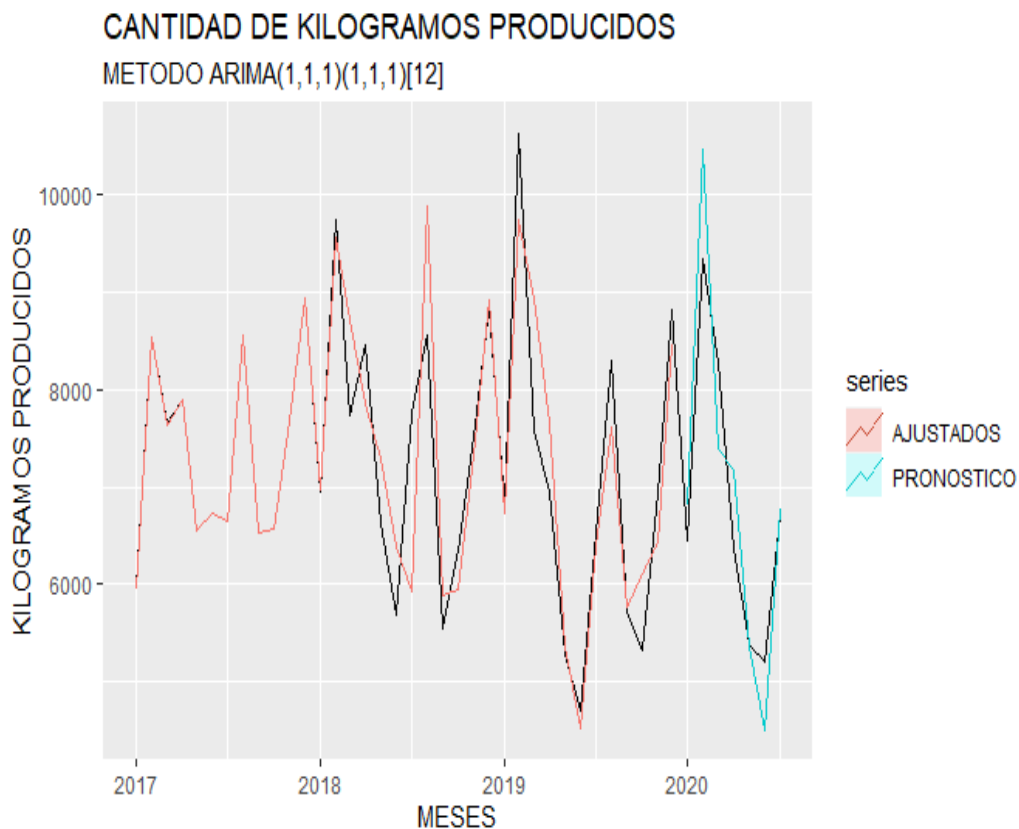


Figura 9. Valores de entrenamiento y testeo

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la figura 9 se observa el comportamiento de los valores reales y los valores pronosticados, donde la línea de color rojo representa el entrenamiento de la serie, la línea de color negro simboliza los valores reales de la serie y la línea de color turquesa representa el pronóstico realizado por el modelo. Se evidencia que la línea de los valores pronosticados es semejante a la línea de los valores reales, por tal razón se afirma que el modelo SARIMA (1, 1, 1) (1, 1, 1) 12 se ajusta con mayor exactitud a los valores históricos de producción y es el adecuado para realizar pronósticos.

4.2.4. Pronóstico

Una vez validado el modelo la última etapa es realizar el pronóstico de la cantidad de kilogramos producidos mensualmente por la empresa en un periodo de agosto 2020 a enero 2021. En la siguiente tabla se presenta el pronóstico de la producción de la empresa para un semestre.

Tabla 10. Pronóstico de la cantidad de kilogramos producidos

Meses	Media de Pronóstico (kg)	Límite inferior con el 95%	Límite superior con el 95%
Aug 2020	8223	6733,864	9713,71
Sep 2020	5691	4119,06	7263,05
Oct 2020	5458	3838,42	7077,33
Nov 2020	6939	5275,89	8601,25
Dic 2020	8668	6963,21	10372,54
Ene 2021	6314	4575,60	8053,96

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 10 se observa la media de los pronósticos y los intervalos de confianza de la parte superior e inferior con un nivel de significancia del 95%. La producción que la empresa debe realizar para el periodo agosto 2020 a enero 2021 es: en agosto 8223 kg con un límite inferior de 6733,864 kg y superior de 9713,71 kg, para septiembre 5691 kg con un intervalo de 4119,06 kg a 7263,05 kg, en octubre 5458 kg con un límite inferior de 3838,42 kg y superior de 7077,33 kg, para noviembre 6939 kg con un intervalo de 5275,89 kg a 8601,25 kg, en diciembre 8668 kg con un límite inferior de 6963,21 kg y superior de 10372,54 kg, por último en enero 6314 kg con un intervalo de 4575,60 kg a 8053,96 kg.

También se observa que los meses más productivos de la empresa son agosto y diciembre. Al tener una relación directa con las ventas se puede argumentar que la empresa aumenta la producción debido a que en esos meses existe mayor demanda, mientras el mes de octubre es uno de los meses que hay un decrecimiento en su producción. En base a esta información la empresa puede tomar una decisión para mejorar su productividad y de esa manera mantener una producción regular en todos los meses del año.

Para observar el comportamiento de los pronósticos realizados se presenta la siguiente figura:

**PRONÓSTICO DE LA CANTIDAD DE KILOGRAMOS PRODUCIDOS
MÉTODO SARIMA(1,1,1)(1,1,1)12**

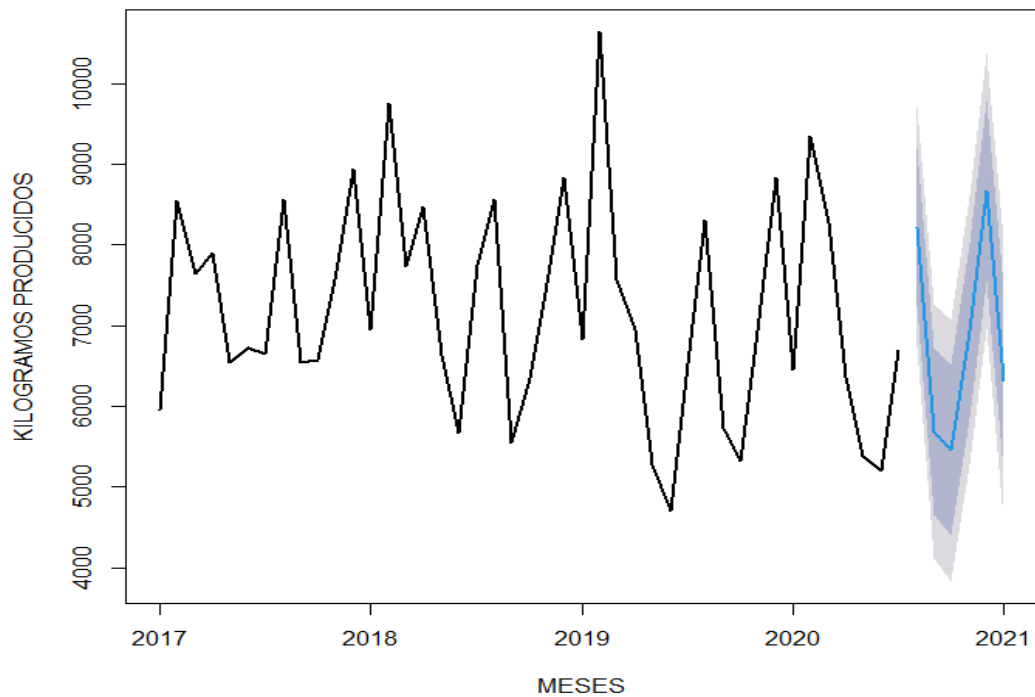


Figura 10. Comportamiento de los pronósticos del modelo SARIMA.

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la figura 10 se observa el comportamiento de los pronósticos. La línea azul simboliza la media de producción, el color gris oscuro representa la cantidad de producción con un nivel de significancia del 85% y el color gris claro representa la cantidad de producción con un nivel de significancia del 95%. En la tabla 10 y figura 10 se visualiza que el mes de octubre tiene una caída en la producción, pero agosto y diciembre son los meses donde la empresa tiene una producción alta. También se observa que las predicciones se asemejan a la serie de tiempo real histórica según los meses pronosticados. Entonces se concluye que la empresa produce más en el mes de diciembre, agosto y noviembre, mientras en los meses de septiembre y enero tiene una producción media y por último en octubre es el mes donde la empresa produce en menor cantidad.

4.2.5. Comparación entre la producción y las ventas

Al conocer que la empresa realiza los pronósticos de su producción en base a las ventas realizadas se vuelve necesario predecir la cantidad de kilogramos vendidos, para ello se utiliza el modelo SARIMA (1, 1, 1) (1, 1, 1) 12 con el objetivo de comparar y

observar el comportamiento de las predicciones realizadas. En la siguiente tabla se presenta el pronóstico de la cantidad de kilogramos vendidos.

Tabla 11. Pronóstico de kilogramos producidos y vendidos

Meses	PROD.	LO 95	HI 95	VEND	LO 95	HI 95
AUG 2020	8223	6733,864	9713,71	8066	5916,1	10217,4
SEP 2020	5691	4119,06	7263,05	6658	4469,8	8847,09
OCT 2020	5458	3838,42	7077,33	5125	2934	7316,2
NOV 2020	6939	5275,89	8601,25	8857	6665	11049,2
DEC 2020	8668	6963,21	10372,54	8871	6678,8	11063,5
JAN 2021	6314	4575,60	8053,96	7090	4970,8	9211,09

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

Una vez realizado el pronóstico de la cantidad de kilogramos a producirse mensualmente, se aplica el modelo SARIMA (1, 1, 1) (1, 1, 1) 12 a la cantidad de kilogramos vendidos con el objetivo de comparar y tener una idea general del proceso de producción. En la tabla 11 se presenta la cantidad de kilogramos producidos y vendidos en el periodo de agosto 2020 a enero 2021. Para el mes de agosto la producción esperada es de 8223 kg mientras el pronóstico de ventas es de 8066 kg, en septiembre se espera producir 5691 kg y el pronóstico de las ventas es de 6658 kg, para octubre uno de los meses con menor producción debido a que las ventas también son bajas, la producción esperada es de 5458 kg mientras que el pronóstico de las ventas es de 5125 kg, en el mes de noviembre la cantidad estimada a ser producida es de 6939 kg y la cantidad de ventas pronosticadas es de 8857 kg, en diciembre uno de los meses más comerciales en todo el mundo la producción esperada es de 8668 kg y las ventas pronosticadas es de 8871 kg, ya empezando un nuevo año en enero se presagia una producción de 6314 kg con un pronóstico de ventas de 7090 kg. Se concluye que las ventas pronosticadas y la producción esperada se encuentran muy relacionadas entre sí, por lo tanto, la producción pronosticada tiene un buen comportamiento. Cabe recalcar que este pronóstico puede verse afectado por factores que no pueden ser controlados como: catástrofes, huelgas, paros, etc.

4.3. Comparación entre la metodología Demand Driven Material Requirements Planning (DDMRP) VS SARIMA (1, 1, 1) (1,1,1)12

En la siguiente tabla se presentan los valores pronosticados del modelo desarrollado y los valores de la metodología que utiliza la empresa:

Tabla 12. Metodología DDMRP VS SARIMA (1, 1, 1) (1, 1, 1) 12

Meses	SARIMA (1,1,1) (1,1,1) 12 (kg)	Metodología DDMRP (kg)	Error
Aug 2020	8223	7972	3%
Sep 2020	5691	5567	2%
Oct 2020	5458	5683	-4%
Nov 2020	6939	6877	1%
Dic 2020	8668	8246	5%
Ene 2021	6314	6070	3%
Promedio	6882	6736	2%

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito

En la tabla 12 se observa los pronósticos del modelo SARIMA (1, 1, 1) (1, 1, 1) 12 y los valores de la metodología que utiliza la empresa para establecer su producción mensual. Al realizar la comparación entre el modelo SARIMA y la metodología DDMRP se demuestra que existe una semejanza en sus valores. Para el mes de agosto el modelo construido excede en un 3% a la metodología DDMRP, en septiembre el modelo supera un 2% a la metodología, para el mes de octubre la metodología excede en un 4% al modelo construido, en noviembre el modelo sobrepasa el 1% a la metodología de la empresa, para el mes de diciembre el modelo excede en un 5% a la metodología y para el mes de enero el modelo desarrollado supera en un 3% a la metodología que utiliza la empresa para el establecimiento mensual de su producción. En el promedio general de los modelos el SARIMA excede en un 2% a la metodología DDMRP. Por lo tanto, se concluye que el modelo SARIMA (1, 1, 1) (1, 1, 1) 12 sería de gran utilidad para tomar decisiones acertadas en los procesos de producción de la empresa.

CAPÍTULO V

5.1. CONCLUSIONES

Con el trabajo desarrollado se llegó a las siguientes conclusiones:

- ✓ En este trabajo se realizó un estudio del arte sobre la aplicación de la metodología de Box-Jenkins, dando sustento para la construcción del modelo ARIMA estacional. Esta metodología fue aplicada en la obra de Sánchez [3] donde logró identificar que los modelos con tendencia estacional o SARIMA pueden describir y predecir adecuadamente el comportamiento de una serie de tiempo.
- ✓ Se evaluó el desempeño del modelo matemático construido mediante el test de normalidad Jarque Bera con el cual se determinó que los residuos del modelo se distribuyen de manera normal, y con el test de ruido blanco Ljung-Box se afirmó que los residuos del modelo son aleatorios y se comportan como un ruido blanco. También se realizó la validación de las predicciones mediante la comparación del pronóstico del entrenamiento con los valores reales de la serie, observándose que los resultados del pronóstico del entrenamiento exceden en un 1,61% a los valores reales, demostrando que el modelo SARIMA (1, 1, 1) (1, 1, 1) 12 se ajusta adecuadamente a la tendencia de los valores de producción de la empresa y cumple con las condiciones para realizar predicciones.
- ✓ Se realizó la comparación entre los resultados de predicción a través del modelo SARIMA (1, 1, 1) (1, 1, 1) 12 construido y el criterio establecido mediante la metodología DDMPR que viene aplicando la empresa Confites El Salinerito en el establecimiento de su producción. A partir de dicha comparación se identificó que los resultados del modelo construido exceden en el pronóstico en un 2% a la metodología DDMRP, pudiendo ser este modelo de gran ayuda para la toma de decisiones en la empresa.
- ✓ Mediante la metodología de Box-Jenkins se desarrolló un modelo matemático que explica, describe y predice el comportamiento de la producción mensual de la empresa Confites El Salinerito de la provincia Bolívar. El modelo que cumplió con las pruebas de adecuación y de aleatoriedad en los residuos fue el SARIMA (1, 1, 1) (1, 1, 1) 12, cuya ecuación de pronósticos estimada es: $\Delta \hat{Y}_t = 0,1132Y_{t-1} + 0,5549Y_{t-12} + 0,7768 \epsilon_{t-1} + 0,9999\epsilon_{t-12} + \epsilon_t$.

5.2. RECOMENDACIONES

- ✓ Se recomienda utilizar este tipo de investigación cada cierto periodo de tiempo con el fin de ir ajustando el modelo y obtener resultados actualizados y así identificar el comportamiento que tiene la producción de la empresa.
- ✓ Se sugiere seguir utilizando las pruebas estadísticas como: el test de Dickey-Fuller que determina la estacionariedad de una serie de tiempo. El test de Jarque Bera que prueba la normalidad de los residuos. Y el test de Ljung-Box que determina si los residuos son aleatorios y si se comportan como un ruido blanco. Obteniendo así mejores resultados en los pronósticos de las variables de estudio.
- ✓ Se recomienda evitar el uso de varias diferencias y parametrizaciones en las series de tiempo, porque con ello se llega a la creación de modelos poco significativos.
- ✓ Según las características de la serie estudiada se recomienda utilizar los modelos SARIMA, porque son modelos que usan la parte estacional de una serie de tiempo y tienen mayor capacidad predictiva que los ARIMA con dependencia regular.

5.3. BIBLIOGRAFÍA

- [1] E. Díaz y C. Díaz, «Desarrollo de un modelo matemático para procesos multivariados mediante Balanced Six Sigma,» *Sciencedirect*, p. 12, 2015.
- [2] M. Valencia, «Innovación en procesos,» *Gestión Innovación y Competitividad*, p. 16, 2017.
- [3] D. Sánchez, Artist, *Modelo ARIMA para el pronóstico de la producción de cacao en el Perú 2012-2018*. [Art]. Universidad Nacional de Trujillo, 2018.
- [4] L. Laurente y F. Quiñonez, «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, La Paz*, vol. 6, nº 1, p. 11, 2019.
- [5] M. Guerrero y S. Medina, «Modelos de series de tiempo aplicados a los expedientes de la Comisión de Derechos Humanos del Distrito Federal,» *Economía Informa*, p. 11, 07 Mayo 2016.
- [6] L. Sánchez, G. Cabanas, Y. Abad y V. Torres, «Utilización de modelos ARIMA para la predicción de la producción de leche. Estudio de caso en la UBPC “Maniabo”, Las Tunas,» *Revista Cubana de Ciencia Agrícola*, vol. 4, nº 3, p. 7, 2014.
- [7] S. Botero y J. Cano, «Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia,» *Scielo*, vol. 27, nº 48, p. 14, 2008.
- [8] D. López, N. García y J. Herrera, «Desarrollo de un modelo predictivo para la estimación del comportamiento de las variables en una infraestructura de red,» *Scielo*, vol. 26, nº 5, p. 12, 24 Junio 2015.
- [9] B. Pavlyshenko, «Machine-Learning Models for Sales Time Series Forecasting,» *Instituto de Tecnología de la Sociedad*, vol. 4, p. 15, 01 2019.
- [10] E. Sánchez, A. Barreras, C. Pérez, F. Figueroa y J. Olivas, «Aplicación de la metodología de Box-Jenkins para pronosticar la producción de leche de bovino

en California México,» *Tropical and Subtropical Agroecosystems*, vol. 16, n° 3, p. 11, 2013.

- [11] J. Fajardo, H. Romero y M. Ramírez, «Un análisis de series de tiempo de la producción de carbón en Colombia para el periodo 1995-2015,» *Espacios*, vol. 39, n° 27, p. 5, 2018.
- [12] A. Alarcón y L. Chuñir, Artists, *Evaluación de modelos econométricos alternativos de series de tiempo para el pronóstico de la inflación en el Ecuador en el corto plazo*. [Art]. Universidad de Cuenca, 2012.
- [13] L. Heras, Artist, *Factores determinantes y proyección de exportación de pitahaya en el Ecuador, una estimación Arima y de Mínimos Cuadrados Ordinarios*. [Art]. Universidad Técnica de Ambato, 2018.
- [14] A. Altamirano, Artist, *Modelo econométrico de consumo privado para el Ecuador en el periodo 1990-2015*. [Art]. Universidad Técnica de Ambato, 2017.
- [15] M. Damián y N. Sandoval, Artists, *Modelo óptimo de pronóstico del índice mensual de producción de electricidad en Perú en el periodo 2006-2015*. [Art]. Universidad Nacional Pedro Ruiz Gallo, 2018.
- [16] P. F. Pérez Cargua, Artist, *Desarrollo de un modelo de estimación de depósitos monetarios para un banco privado*. [Art]. Universidad Andina Simón Bolívar sede Ecuador, 2017.
- [17] E. Arteaga, Artist, *Construcción de un modelo econométrico para estimar las ventas mensuales de las cuatro marcas principales de bebidas gaseosas de la empresa Ecuador Bottling Company Corp*. [Art]. Escuela Politécnica Nacional, 2010.
- [18] L. Coba, Artist, *Pronóstico de ventas de la farmaceutica Sanofi usando series temporales*. [Art]. Universidad Central del Ecuador, 2017.

- [19] S. Mora y E. Sempértégui, Artists, *Proyección a corto plazo de la demanda eléctrica de los alimentadores de distribución de la Empresa Eléctrica Regional Centro Sur C.A.*. [Art]. Universidad del Azuay, 2017.
- [20] J. Angulo, Artist, *Modelo ARIMA para el pronóstico de la liquidez monetaria en el sistema financiero peruano*. [Art]. Universidad Nacional de Trujillo, 2016.
- [21] F. Burgaentzle, Artist, *Pronósticos y modelo de inventarios en una empresa láctea ecuatoriana*. [Art]. Universidad San Francisco de Quito, 2016.
- [22] J. Cuyo, Artist, *Propuesta metodológica para el pronóstico de la demanda de agua potable en el norte de la ciudad de Quito*. [Art]. Universidad Central del Ecuador, 2016.
- [23] J. Aldaz, J. García, Quiñones y R, «Parvovirus canina en la provincia Bolívar, Ecuador. Utilidad de los modelos Box-Jenkins para su análisis y predicción,» *Scielo*, vol. 34, n° 3, p. 10, 2012.
- [24] K. Ogata, Ingeniería de control moderna, Cuarta ed., Madrid: Pearson Educación, 2003, p. 965.
- [25] G. Ríos, «Series de Tiempo,» de *modelos se series de tiempo*, Chile, Universidad de Chile, 2008, p. 52.
- [26] J. Jiménez, J. Gázquez y R. Sánchez, «La capacidad en los metodos de Box-Jenkisy Holt-Winters.,» *Revista Europea de dirección y economia* , vol. 15, n° 3, p. 13, 2006.
- [27] H. Rob y G. Athanasopoulos, *Forescasting: Principales and Practice*, Australia: Texts, 2018.
- [28] J. Villavicencio, *Introducción a las series de tiempo*, Universidad Nacional Pedro Ruiz Gallo, 2018.
- [29] M. P. Gonzales, *Análisis de Series Temporales Modelo ARIMA*, Vasco : Eman T Zabel Zazu, 2009.

- [30] M. Quesada, Artist, *Análisis de Series, modelos heterocedásticos*. [Art]. Universitas Granatensis, 2013.
- [31] G. Martínez y L. Chalita, «Aplicación de la metodología de Box-Jenkins para pronóstico de precios en jitomate,» *Scielo*, vol. 2, n° 4, p. 11, 2011.
- [32] A. Aguirre, Introducción al tratamiento de series temporales: aplicación a las ciencias de la salud, Madrid: Ediciones Díaz de Santos, 1994.
- [33] P. Muñoz, L. Muñoz y T. Sánchez, «Estudio de potencia de pruebas de normalidad usando distribuciones desconocidas con distintos niveles de no normalidad.,» *Perfiles*, vol. 21, n° 1, p. 8, 2019.
- [34] Tour Salinerito, «Turismo de Salinas Guaranda,» 24 07 2013. [En línea]. Available: <https://sites.google.com/site/turismosalinasguarandambpatino/>.
- [35] N. Almendariz y C. Sara, Artists, *Análisis de las Herramientas de Gestión que utilizan las unidades productivas comunitarias en la parroquia Salinas de la provincia Bolívar*. [Art]. Escuela Politécnica Nacional, 2012.
- [36] F. Arias, El proyecto de investigación. Introducción a la metodología científica, Venezuela: Editorial Episteme C.A., 2012.
- [37] H. Avila, Introducción a la metodología de la investigación, México : Eumed.net, 1999.
- [38] A. Rencher y B. Schaalje, *Linear models in statistics*, Canadá: Wiley - intersciencie, 2008.
- [39] T. C. A. Bernal, *Metodología de la investigación: para administración, economía, humanidades y ciencias sociales*, Segunda Edición ed., México: Pearson Educación, 2006, p. 304.
- [40] C. Olivera, *Introducción a la Educación Comparada*, Costa Rica: EUNED, 2008.
- [41] C. R. Álvarez, *Estadística aplicada a las ciencias de la salud*, España: Ediciones Díaz de Santos, 2007.

- [42] J. Sarabia y M. Pascual, Curso básico de estadística para economía y administración de empresas, Cantabria: Ed. Universidad de Cantabria, 2005, p. 378.
- [43] M. Gómez, Introducción a la metodología de la investigación científica, Córdoba: Editorial Brujas, 2006.
- [44] J. García, C. Ramos y G. Ruiz, Estadística empresarial, Cádiz: Servicio Publicaciones UCA, 2009.
- [45] F. Pérez, Introducción a las series de tiempo. Métodos paramétricos, Colombia: Universidad De Medellín, 2007.
- [46] J. Racine, Reproducible Econometric Using R, Estados Unidos: Oxford University Press , 2019.
- [47] F. Burgaentzle, Artist, *Pronósticos y modelo de inventarios en una empresa láctea ecuatoriana*. [Art]. Universidad San Francisco de Quito, 2016.

5.4. ANEXOS

Ubicación de la parroquia Salinas.



Figura 11. Mapa para llegar a Salinas

Fuente: Fundación Familia Salesiana Salinas



Figura 12. Parroquia Salinas

Fuente: Fundación Familia Salesiana Salinas

Empresa Confites el Salinerito



Figura 13. Fabbrica Confites El Salinerito
Fuente: Fundación Familia Salesiana Salinas



Figura 14. Confites El Salinerito
Fuente: Fundación Familia Salesiana Salinas

Productos Terminados de la empresa Confites El Salinerito



Figura 15. Productos Terminados
Fuente: Empresa Confites El Salinerito



Figura 16. Productos Terminados
Fuente: Empresa Confites El Salinerito

Base de datos digitalizada.


 CONTROL DE TRAZABILIDAD DE PRODUCTO TERMINADO				
Meses	Unidad	Unid. Producidas	unid. Vendidas	Ventas
Enero	kg	5956	5432	19867,85
Febrero	kg	8543	6341	16766,54
Marzo	kg	7645	10945	33607,70
Abril	kg	7893	9578	20073,21
Mayo	kg	6539	7534	20088,27
Junio	kg	6723	6567	15510,87
Julio	kg	6655	7230	19958,82
Agosto	kg	8567	6235	20798,82
Septiembre	kg	6537	8543	21060,08
Octubre	kg	6567	7345	17030,31
Noviembre	kg	7674	7523	24177,77
Diciembre	kg	8944	7201	22103,87
Enero	kg	6957	5432	6134,41
Febrero	kg	9756	8673	5142,40
Marzo	kg	7733	9673	15139,82
Abril	kg	8468	11312	9299,71
Mayo	kg	6632	8563	14699,54
Junio	kg	5673	6543	4923,79
Julio	kg	7753	7534	10515,88
Agosto	kg	8556	8260	14342,97
Septiembre	kg	5542	7423	6020,41
Octubre	kg	6345	6125	13978,16
Noviembre	kg	7523	9423	21023,13
Diciembre	kg	8834	9872	12447,85
Enero	kg	6831	6831	20788,33
Febrero	kg	10639	6162	23592,14
Marzo	kg	7565	11099	8671,50
Abril	kg	6948	7301	2868,34
Mayo	kg	5282	6393	17883,54
Junio	kg	4705	5670	7447,53
Julio	kg	6535	6092	16335,16
Agosto	kg	8303	6784	10255,93
Septiembre	kg	5729	7416	8238,31
Octubre	kg	5321	5799	8871,44
Noviembre	kg	6929	7525	21029,55
Diciembre	kg	8829	6686	18627,15
Enero	kg	6451	7147	6522,23
Febrero	kg	9343	8671	8857,42
Marzo	kg	8281	9864	13183,90
Abril	kg	6362	7043	12338,28

Figura 17. Base de datos digitalizada de la empresa
Fuente: Empresa Confites El Salinerito

Script del Modelo Matemático

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ - [Icons] Go to file/function [Grid] Addins
serie de tiempo.R x salinas x
← → [Icons] Source on Save [Search] [Run] [Refresh]
1
2 #LIBRERIAS A UTILIZAR
3
4 library(ggplot2)
5 library(tseries)
6 library(forecast)
7 library(TSstudio)
8 library(GGally)
9 library("tseries")
10 attach(salinas)
11 names(salinas)
12 str(salinas)
13
14 - #####PRONÓSTICO DE PRODUCCION #####
15 #TRANSFORMAR LA PRODUCCIÓN A SERIE DE TIEMPO
16
17 produce <- ts (salinas$`PRODUCIDO KG`, start =c(2017,1), frequency = 12 )
18 produce
19
20 #GRAFICO DE LA SERIE DE TIEMPO DE PRODUCCIÓN
21
22 plot(produce, ylab= "CANTIDAD DE UNIDADES", main = "SERIE DE PRODUCCIÓN")
23 abline(h = mean(produce), col = "red", lty = "dashed" , lwd = "2")
24
25 # OBSERVAMOS LA TENDENCIA Y ESTACIONALIDAD
26
27 dec <- decompose(produce)
28 autoplot(dec)+ ggtitle("CANTIDADES DE KILOGRAMOS PRODUCIDOS")
29
30 # COMPORTAMIENTO DE LA SERIE POR MESES Y AÑOS
31
32 ts_seasonal(produce,type = "all")
33
34 #PRUEBA DE RAIZ UNITARIA
35 #Ho > 0.05 SERIE NO ESTACIONARIA
36 #Ha <0.05 SERIE ESTACIONARIA
37
38 adf.test(produce,alternative = "stationary")
39
40 #NUMERO DE DIFERENCIACIONES A REALIZAR
41
42 ndiffs(produce) # 1
43 nsdiffs(produce) # 1
44
45 #DIFERENCIACION PARA ESTACIONARIEDAD
46
47 produce2 <- diff(produce)
48
49 # GRAFICO DE LA SERIE DE TIEMPO ESTACIONARIA
50
51 plot(produce2, xlab = "MESES",ylab= "CANTIDADES PRODUCIDAS",
52      main = "PRODUCCIÓN EN KILOGRAMOS", lwd = "2")
53 abline(h = mean(produce2), col = "red", lty = "dashed", lwd = "2" )
54
55 # PRUEBA DE RAIZ UNITARIA A SERIE ESTACIONARIA
56
57 adf.test(produce2,alternative = "stationary")
58
59 # LA SERIE ES ESTACIONARIA p-value < 0.05
60
61 #FUNCIÓN DE AUTOCORRELACIÓN Y AUTOCORRELACIÓN PARCIAL
62 #DETERMINAMOS LAS MEDIAS MOVILES Y REGRESORES
63
64 par(mfrow = c(2,1))
65 acf(ts(produce, frequency = 1))
66 pacf(ts(produce, frequency = 1))
67 acf(ts(produce2, frequency = 1)) # q =1 0 2
68 pacf(ts(produce2, frequency = 1))# p = 1
69
```

```

70 #MODELO AUTOARIMA
71
72 auto.arima(produce)
73
74 # 1. IDENTIFICACIÓN DEL MODELO
75
76 ##autoarima(0,0,0)(0,1,0) AIC 501,49; ECM=582249, BIC=502.92
77 ##Mod2(2,1,2)(1,1,1) AIC 493.69; ECM = 525817 ; BIC=503.5
78 ##Mod3(1,1,2)(1,1,1) AIC 491.98; ECM = 495420 ; BIC=500.38
79 ##Mod4 (2,1,1)(1,1,1) AIC 492.27; ECM = 543789 ; BIC=500.67
80 ##Mod5 (0,1,0)(1,1,1) AIC 497.21; ECM = 592758 ; BIC=501.41
81 #Mod6 (1,1,1)(1,1,1) AIC 491,04; ECM = 488744; BIC=498.05
82
83
84 #MODELO ARIMA
85
86 mod2 <- Arima(produce, order = c(1, 1, 1), seasonal= c (1, 1, 1), method="ML")
87 mod2
88
89 # 2. ESTIMACION DEL MODELO
90 #RESUMEN DEL MODELO
91
92 summary(mod2)
93 coefficients(mod2)
94 confint(mod2)
95 library(lmtest)
96 coeftest(mod2)
97 #3. VALIDACION DEL MODELO
98 #DIAGNOSTICO DEL MODELO
99
100 tsdiag(mod2)
101
102 #PRUEBA DE VALIDEZ DE RUIDO BLANCO
103
104 checkresiduals(mod2)
105
106 #PRUEBA DE NORMALIDAD JARQUE.BERA
107 #HO EXISTE NORMALIDAD SI p-value > 0.05
108 #Ha NO EXISTE NORMALIDAD SI p-value < 0.05
109
110 jarque.bera.test(residuals(mod2))
111
112 # PRUEBA DE LJUNG- BOX
113 #HO EXISTE RUIDO BLANCO SI p-value > 0.05
114 #Ha NO EXISTE RUIDO BALNCO S p-value < 0.05
115
116 Box.test(residuals(mod2), type= "Ljung-Box" )
117
118 # ENTRENAMIENTO
119 #SEPARAMOS UNA SERIE TRAIN Y SERIE TEST
120
121 serie_train <- window(produce, end = c(2019, 12))
122 serie_test <- window(produce, start = c(2020, 1))
123
124 # VALIDACION DEL MODELO CON PREDICCIONES ANTERIORES
125
126 mod <-Arima(serie_train, order = c(1,1,1), seasonal = c(1,1,1))
127 mod
128
129 pred <- forecast(mod, h = 7)
130 pred
131 round(accuracy(pred, serie_test),2)
132
133 autoplot(produce) +
134   autolayer(fitted(mod), series="AJUSTADOS") +
135   autolayer(pred, series="PRONOSTICO", PI = F) +
136   xlab("MESES") + ylab("KILOGRAMOS PRODUCIDOS") +
137   ggtitle("CANTIDAD DE KILOGRAMOS PRODUCIDOS",
138     subtitle = "METODO ARIMA(1,1,1)(1,1,1)[12]")
139
140 # PRONOSTICOS
141
142 #PREDECIR LAS CANTIDADES DE KILOGRAMOS A PRODUCIR
143
144 predecir <- forecast(mod2, h = 6)
145 predecir
146
147 # GRAFICO DE LAS PREDICCIONES
148
149 plot(predecir, xlab = "MESES", ylab = "KILOGRAMOS PRODUCIDOS",
150   main = "PRONÓSTICO DE LA CANTIDAD DE KILOGRAMOS PRODUCOS
151   MÉTODO ARIMA(1,1,1)(1,1,1)12", lwd = "2")

```

Figura 18. Script del Modelo SARIMA (1, 1, 1) (1, 1, 1)12

Fuente: Elaborado por el autor en base a los datos de Confites El Salinerito