

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

ESCUELA PROFESIONAL DE ESTADÍSTICA



TESIS

“MODELO DE SERIES DE TIEMPO PARA DESCRIBIR LA PRODUCCIÓN

MENSUAL DE ESPÁRRAGO EN EL PERÚ, PERIÓDO

ENERO 2007 – DICIEMBRE 2018”.

AUTORAS:

Bach. Acosta Bravo Roxana Estela

Bach. Chapoñan Bances Deyci Yris

ASESOR:

Dr. Rodríguez López Wilver Omero

FECHA DE SUSTENTACION: 20 de abril del 2022

PARA OPTAR POR EL TITULO PROFESIONAL DE: Licenciada en Estadística

Perú – Chiclayo, 2022



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



**“MODELO DE SERIES DE TIEMPO PARA DESCRIBIR LA PRODUCCIÓN
MENSUAL DE ESPÁRRAGO EN EL PERÚ, PERÍODO
ENERO 2007 – DICIEMBRE 2018”.**

TESIS

**PARA OPTAR POR EL TÍTULO PROFESIONAL DE LICENCIADA EN
ESTADÍSTICA**

PRESENTADO POR:

Bach. Roxana Estela Acosta Bravo

Autora

Bach. Chapoñan Bances Deyci Yris

Autora

Dr. Rodríguez López Wilver Omero

Asesor

Perú – Chiclayo, 2022



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



**“MODELO DE SERIES DE TIEMPO PARA DESCRIBIR LA PRODUCCIÓN
MENSUAL DE ESPÁRRAGO EN EL PERÚ, PERÍODO
ENERO 2007 – DICIEMBRE 2018”.**

TESIS

**PARA OPTAR POR EL TÍTULO PROFESIONAL DE LICENCIADA EN
ESTADÍSTICA**

APROBADA POR:

Lic. José Lorenzo Núñez Rodríguez
PRESIDENTE

Dr. José Luis Rodas Cabanillas
SECRETARIO

Lic. Luis Enrique Tuñoque Gutiérrez
VOCAL

Perú – Chiclayo, 2022

Declaración jurada de Originalidad

Nosotras, Acosta Bravo Roxana Estela, Chapoñan Bances Deyci Yris, investigadoras principales, y Dr. Rodríguez López Wilver Omero, asesor del trabajo de investigación “Modelo de serie de tiempo para describir la producción mensual de espárrago en el Perú, período enero 2007 – diciembre 2018”, declaramos bajo juramento que este trabajo no ha sido plagiado, ni contiene datos falsos. En caso se demostrará lo contrario, asumo responsablemente la anulación de este informe y por ende el proceso administrativo a que hubiera lugar. Que puede conducir a la anulación del título o grado emitido como consecuencia de este informe.

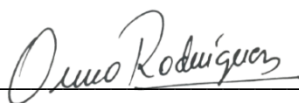
Lambayeque, __ de enero del 2021



Bach. Acosta Bravo Roxana Estela
Autora



Bach. Chapoñan Bances Deyci Yris
Autora



Dr. Rodríguez López Wilver Omero
Asesor

DEDICATORIA

A Dios, Por darme vida, salud para lograr mis objetivos.

A mis padres y hermanas, *ARMANDO Y MIRTA* por ser el pilar fundamental en mi educación y en todo lo que soy.

A mi esposo Neptali, por su apoyo incondicional para la consecución de mis logros profesionales.

A mi hijo Estefano Fernando, por ser mi fortaleza y fuente de inspiración para seguir adelante.

¡Gracias!

Deyci Yris Chapañan Bances

A Dios, ya que gracias a él he podido concluir mi carrera profesional.

A mis padres Máximo y Luisa, por su apoyo incondicional para salir adelante.

A mis hermanos Antonio y Karen, por ser mi mayor motivación para no rendirme.

¡Gracias!

Roxana Estela Acosta Bravo

AGRADECIMIENTO

Agradecemos a Dios por bendecirnos y darnos fuerzas para culminar esta etapa académica e impulsarnos a seguir adelante y seguir creciendo como profesionales.

A nuestro asesor Dr. Wilver Omero Rodríguez López por su comprensión, guía y constante apoyo en nuestra formación, por sus consejos que han sido el punto clave para lograr el éxito en el desarrollo de nuestro proyecto de investigación.

De manera especial agradecer a nuestros docentes que con sus valores y enseñanzas nos han formado para ser profesionales exitosos.

A la Universidad Nacional “Pedro Ruiz Gallo” por habernos cedido un lugar en sus aulas para estudiar y permitirnos tener una excelente formación profesional.

¡Muchas Gracias!

RESUMEN

La presente investigación que tuvo por finalidad determinar el modelo de series de tiempo para describir la producción mensual de espárrago en el Perú, periodo enero 2007 – diciembre 2018, empleo una metodología observacional retrospectivo y prospectivo, con un tipo de investigación teórica, empírica y longitudinal. La muestra estuvo constituida por la producción mensual en miles de toneladas de espárrago en el Perú, en el periodo de enero a diciembre durante los años 2007-2018; siendo la fuente secundaria de donde se ha obtenido la producción mensual de espárrago en el Perú. Los resultados evidenciaron que el modelo que pronostica de forma correcta la producción de espárragos fue un SARIMA (1,1,1) (2,1,0), asimismo, sus parámetros fueron significativos, mientras que sus errores fueron: RMSE igual a 2.2035, el MAE fue 1.5899; el MPE igual a -1.7234, y el MAPE fue 5.4501. Este modelo fue elegido gracias a tener menos errores. Se concluye que la metodología de Box Jenkins mediante el modelo SARIMA (1,1,1) (2,1,0), pronostica correctamente la producción de espárragos, ello debido a que sus residuales cumplieron independencia y normalidad ($p < 0.05$).

Palabras clave: metodología de box Jenkins, espárragos, producción, SARIMA

ABSTRACT

The purpose of this research was to determine the time series model that describes the monthly production of asparagus in Peru, period January 2007 - December 2018, used a retrospective and prospective observational methodology, with a type of theoretical, empirical research and longitudinal. The sample consisted of the monthly production in thousands of tons of asparagus in Peru, in the period from January to December during the years 2007-2018; being the secondary from which the monthly asparagus production in Peru has been obtained. The results showed that the model that correctly predicts the production of asparagus was a SARIMA (1,1,1) (2,1,0), likewise, its parameters were significant, while its errors were: RMSE equal to 2,2035, the MAE was 1.5899; the MPE equal to -1.7234, and the MAPE was 5.4501. This model was chosen thanks to having fewer errors. It is concluded that the Box Jenkins methodology through the SARIMA (1,1,1) (2,1,0) model, correctly predicts the production of asparagus, due to the fact that its residuals fulfilled independence and normality ($p < 0.05$).

Key words: box Jenkins methodology, asparagus, production, SARIMA.

INDICE GENERAL

DEDICATORIA.....	v
AGRADECIMIENTO.....	vi
RESUMEN.....	vii
ABSTRACT.....	viii
INDICE GENERAL.....	ix
ÍNDICE DE TABLAS.....	xi
ÍNDICE DE FIGURAS.....	xii
INTRODUCCIÓN.....	1
CAPITULO I. DISEÑO TEÓRICO.....	4
1.1. Antecedentes.....	4
1.1.1. Internacionales.....	4
1.1.2. Nacional.....	5
1.1.3. Locales.....	6
1.2. Base Teórica.....	8
1.2.1. Pronóstico.....	8
1.2.2. Serie de Tiempo.....	8
1.2.3. Modelos de Predicción.....	12
1.2.4. Modelos ARIMA, Metodología Box-Jenkins.....	15
1.2.5. Etapa de estimación.....	16
1.2.6. Identificación del Modelo.....	23
1.2.7. Estadístico de Prueba.....	25
1.2.8. Producción de espárragos.....	30
CAPITULO II. MÉTODOS Y MATERIALES.....	31
2.1. Diseño de contrastación de Hipótesis.....	31
2.2. Variables.....	31
2.3. Población y muestra.....	31
2.4. Técnicas e instrumentos de recolección de datos.....	32
2.5. Análisis y Procesamiento Estadístico de los datos.....	32
2.5.1. Identificación del Modelo.....	33
2.5.2. Estimación.....	33
2.5.3. Adecuación del Modelo.....	34
2.5.4. Pronóstico con el Modelo.....	34
2.5.5. Validez del Pronóstico.....	34
CAPITULO III. RESULTADOS Y DISCUSIONES.....	36

3.1. Resultados	36
3.2. Discusión de los resultados	48
CAPÍTULO IV. CONCLUSIONES.....	50
CAPÍTULO V. RECOMENDACIONES.....	52
REFERENCIAS BIBLIOGRAFICAS	53
ANEXOS	56
ANEXO 2. CÓDIGOS EN RSTUDIO.....	58

ÍNDICE DE TABLAS

Tabla 1 Estimación de coeficientes de los parámetros tentativos de los posibles modelos SARIMA.....	44
Tabla 2 Criterios AIC y BIC de los posibles modelos SARIMA.....	46
Tabla 3 Comparación de los errores de pronósticos de los modelos SARIMA propuestos.	47
Tabla 4 Resultados de ruido blanco, independencia y normalidad de los residuos para el modelo SARIMA (1,1,1) (2,1,0) [12].....	45
Tabla 5 Valores pronosticados para los siguientes meses.	46
Tabla 6. Producción agropecuaria (miles de toneladas) - Agrícola – Espárrago	56

ÍNDICE DE FIGURAS

Figura 1. Producción de espárragos a nivel nacional en el periodo 2007 – 2018.	36
Figura 2. Función de autocorrelación simple y parcial de la producción de espárragos 2007-2018.	37
Figura 3. Descomposición de los datos de producción de espárragos 2007-2018.	39
Figura 4. Serie transformada en primeras diferencias tanto en la parte regular y estacional de la producción de espárragos 2007-2018.	40
Figura 5. Función de autocorrelación simple y parcial de la serie producción de espárragos en primeras diferencias tanto en la parte regular y estacional.	42
Figura 6. Función de autocorrelación simple y parcial de los residuales del modelo SARIMA (1,1,1) (2,1,0) [12]	47
Figura 7. Residuales del modelo SARIMA (1,1,1) (2,1,0) [12]	44
Figura 8. Pronóstico de la producción de espárrago para los años 2019 - 2021.	47

INTRODUCCIÓN

El espárrago es un cultivo perenne que se presta bien a las operaciones agrícolas a pequeña escala y a tiempo parcial con muchas oportunidades en el mercado. (Penn State, 2017)

A nivel mundial el país que tiene la mayor producción de espárragos es China con 57 000 hectáreas. En Norteamérica hay 44.600 hectáreas. Los Estados Unidos y México son los principales productores (23.000 y 20.000 hectáreas, respectivamente), seguidos de lejos por Canadá (1.600 ha). En América Latina, hay 31.300 hectáreas en total y Perú es el mayor productor (20.000 ha). (Fresh Plaza, 2018)

A nivel nacional en el año 2017, el Perú produjo aproximadamente 390 mil toneladas de espárragos, lo que representó un ligero incremento de 3.3% respecto a las 377.500 toneladas producidas el año pasado, según señala un informe del Departamento de Agricultura de Estados Unidos. La superficie cultivada en Perú es de 28.000 hectáreas. La región de Ica representa el 45% de la producción, La Libertad concentra el 43 %, mientras que Lima y regiones del sur representan la producción restante. (Agraria, 2017)

El rendimiento promedio del espárrago asciende a 14 toneladas por hectárea, aunque se obtienen rendimientos de hasta 18 toneladas por hectárea. En la última década el rendimiento promedio se incrementó gracias a la utilización de mejor tecnología, sistemas más eficientes y un mejor conocimiento de los productores. (Agraria, 2017)

Ante ello, surge la problemática para la producción de espárragos en el Perú, pues está se hace cada vez más importante, esto sucede por ser uno de los cultivos que más se exporta a nivel nacional. Se considera de mucha importancia predecir el comportamiento futuro de la producción de espárragos en el Perú, y se hace de necesidad identificar, el pronóstico correcto que tendrá los próximos años, ello permitirá generar mejores estrategias para la producción y exportación, generando mayor rentabilidad para productores y comerciantes. Para el modelo de pronósticos cuantitativos existen diferentes técnicas y metodologías; una de las más conocidas es la metodología de box Jenkins, la cual requiere el cumplimiento de ciertos supuestos, y hace que se limite en algunos aspectos con respecto a las variables.

Ante ello surgió el problema general: ¿Cuál es el modelo de serie de tiempo para pronosticar la producción mensual de espárrago en el Perú, periodo enero 2007 – diciembre 2018?

Para solución al problema se estableció como objetivo general: Determinar un modelo de series de tiempo para describir la producción mensual de espárrago en el Perú, periodo enero 2007 – diciembre 2018; mientras los objetivos específicos: 1. Describir el comportamiento de la producción mensual de espárragos en el Perú, período enero 2007 – diciembre 2018; analizar la estacionariedad y estacionalidad de la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018; 2. identificar el modelo de serie de tiempo para describir la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018; estimar los parámetros de series de tiempo de la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018; 3. validar el modelo de series de tiempo para describir la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018; y, 4. pronosticar la producción mensual de espárrago en el Perú mediante

el modelo obtenido para los años 2019-2020.

Siendo la hipótesis: El modelo de series de tiempo para describir la producción mensual de espárrago en el Perú período enero 2007 – diciembre 2018 es un modelo ARIMA.

CAPITULO I. DISEÑO TEÓRICO

1.1. Antecedentes

1.1.1. Internacionales

Mota (2016) en su trabajo realizado en Puebla, denominado “*Pronostico del precio del crudo de exportación mexicano con la Metodología de Box-Jenkins para series de tiempo*”, buscó establecer un modelo estadístico con fines de pronóstico para el precio del crudo de exportación. La metodología utilizada fue predictiva de enfoque cuantitativo, tomando por modelo al Box Jenkins. Los resultados evidenciaron que el mejor modelo de pronóstico fue el ARIMA (1,1,1). El modelo al inicio presentó un comportamiento creciente para las 6 semanas próximas; se esperaban precios constantes, sin embargo, hay factores que afectaron el cierre del 2015 y apertura de 2016.

Sanchez y Varillas (2014), en su investigación “*Modelos de serie de tiempo mediante la Metodología de Box y Jenkins para pronosticar la llegada de turistas nacionales y extranjeros a la ciudad sagrada de Caral. Periodo Enero a diciembre, 2014*”, su objetivo principal fue realizar un modelo para pronosticar la llegada de turistas del ámbito nacional como extranjeros, empleando el método de Box y Jenkins. La investigación fue de tipo teórica empírica con un diseño no experimental; como resultados se obtuvo el modelo SARIMA (3, 1,0) (1, 1,0), el cual llegó a predecir para el año 2014, 55378 visitantes nacionales y extranjeros que llegarían a la ciudad sagrada de Caral.

1.1.2. Nacional

Angulo (2013), en su investigación “*Predicción del PBI Real Nacional Trimestral: Redes Neuronales Autorregresivas vs Metodología Box Jenkins*”, estudio de grado para recibir el título de Ingeniero Estadístico, Universidad Nacional de Trujillo, que tuvo como objetivo determinar que metodología facilita un modelo más ideal de serie de tiempo para pronosticar trimestralmente el “PBI” de todo el país, la investigación tuvo un diseño no experimental, longitudinal con tendencia. Para lograr con el propósito de la investigación, se trabajó con 74 datos de PBI registrados en el INEI desde año 1994. Como resultados se encontró que el AR - NN 5, es el óptimo modelo que resultó de las “redes neuronales”, cuya estructura estuvo dada por el cuarto y segundo rezago como variable de entrada y dos nodos en la capa oculta; con respecto al modelo más ideal empleando el método de Box - Jenkins se encontró al modelo “SARIMA (2,1,2) (2,1,2) s” con constante. En conclusión, se obtuvo que el mejor modelo se encontró empleando la metodología Box - Jenkins, el cual tiene mayor ajuste en el periodo de evaluación y de calibración.

Cruz (2014), la presente investigación “*Pronóstico para la exportación de espárrago de la Región La Libertad enero 2010 – agosto 2014*”. tuvo como objetivo, determinar el modelo que de una mejor explicación del comportamiento sobre la exportación de espárrago mensual y estimar pronósticos, la investigación aplicada fue longitudinal, los datos se obtuvieron de la empresa de “inteligencia comercial de comercio exterior VERITRADE S.A.C.” desde enero del 2010 hasta agosto del 2014. Los modelos resultantes empleando la metodología de Box - Jenkins son: para las exportaciones total de espárrago, se halló un ARIMA (1, 1, 9), para las exportaciones de espárragos frescos o refrigerados, un ARIMA (1, 1, 3), para las exportaciones de espárrago conservado, no congelado, se obtuvo un ARIMA

(0, 1, 1) y para los espárragos congelados, cocidos en agua o vapor, se encontró un ARIMA (2, 1, 12).

Delacruz y Aguirre (2015), en su investigación “*Análisis comparativo de redes Neuronales Artificiales y la metodología de box Jenkins en el pronóstico del índice de precios al consumidor*”. Huaraz - Perú. que tuvo como objetivo comparar la metodología de Box - Jenkins con las redes neuronales artificiales para pronosticar el índice de precios del consumidor, Huaraz, para el periodo 2000 hasta 2014. La metodología de esta investigación fue de tipo retrospectiva predictiva con un diseño no experimental. Los resultados constataron que el método de “Box y Jenkins” generó un menor valor del ECM, mientras que el método de “redes neuronales” se obtuvo un mayor valor del ECM, siendo este último el más adecuado para pronosticar, pero se encontró que, empleando Box Jenkins, se obtuvo una mayor precisión en la predicción del IPC en donde para los meses agosto y setiembre fue de 115.96 y 116.02.

1.1.3. Locales

Gil y Guevara (2016), En la presente investigación se busca encontrar un “*Modelo óptimo de serie de tiempo para el pronóstico del crédito de consumo total mensual otorgado por la Banca Múltiple del Perú.2001-2014, mediante el método de “Box y Jenkins”*”. Del análisis de los errores de pronóstico se eligieron como óptimos los modelos ARIMA (2, 1,3) para la serie nominal y ARIMA (3, 1,3) para la serie deflactada. De la elección del modelo, podemos decir que rechazamos la hipótesis, por lo que concluimos que el modelo óptimo de pronóstico no es de tipo ARIMA (0,2,2) suavizada por el método exponencial doble de Holt.

Vásquez (2017), en su estudio *“Metodología de Box Jenkins vs Redes Neuronales Artificiales para pronosticar el precio del dólar del Sistema Bancario en moneda nacional año 2015”*, que tuvo como objetivo realizar la estimación de un modelo que proporcione predecir el costo del dólar en moneda nacional del sistema bancario mediante la metodología de Box - Jenkins y redes neuronales artificiales. La investigación fue de tipo descriptiva, predictiva, longitudinal; la muestra bajo estudio estuvo conformada por la serie de datos respecto al valor del dólar del sistema bancario, enero del 2004 hasta diciembre del 2014. Entre los principales resultados encontrados fueron: para la metodología de Box y Jenkins, para la compra, se obtuvo un ARIMA (2,2,3) un RMSE = 0.983, el MAE = 0.008 y para la venta se halló un ARIMA (2,2,3) un RMSE = 0.982 y con MAE = 0.008 indicando un coeficiente bueno; respecto al modelo de redes neuronales, para la compra, un modelo de estructura (12:1:5:1), su RMSE = 0.042 y para la venta se halló un (12:1:6:1) con RMSE = 0.038. Se concluye que la Metodología de Box y Jenkins es más viable, dado que, para la venta se encontró el ARIMA (2,2,3) y la venta un ARIMA (3,1,2), el valor obtenido MAE fue menor.

Villalobos (2014), en su investigación, *“Pronostico de Ingresos Tributarios recaudados de Impuestos generales a las ventas a nivel nacional, por la SUNAT ,mediante la Metodologia de Box y Jenkins y Redes Neuronales Artificiales para el año 2014.* Tuvo como finalidad obtener el modelo adecuado para predecir el ingreso de tributos por “IGV” para el 2015, el cual ayudará a la planificación presupuestal del fisco 2015. La serie obtenida de los ingresos mensuales fue de la página web de la SUNAT. El modelo utilizando el método “Box y Jenkins” es un SARIMA (0,1,1) x (1,1,0)₁₂ con un RMSE = 66 921.9 y un MAPE = 3.56 %; mientras que, con las redes neuronales artificiales el modelo encontrado está estructurado

como 26:1:4:1 con un RMSE = 17 511.1 y un MAPE = 0.85%. Se concluyó que, con las redes neuronales artificiales, el modelo es más exacto para pronosticar el IGV percibido por la Sunat a nivel nacional.

1.2. Base Teórica

1.2.1. Pronóstico

Según Villareal (2016), nos dice “Es una estimación cuantitativa o cualitativa de uno o más variables que componen un evento futuro, en base a información actual o pasada”.

La finalidad de los pronósticos consiste en disminuir el nivel de incertidumbre de donde se toma decisiones que afectan la continuidad de la empresa, así como de todos los involucrados. Sin embargo, el pronóstico no reemplaza la razón administrativa en la toma de decisiones, únicamente viene a ser un apoyo en ese proceso. (Polis, 2002)

1.2.2. Serie de Tiempo

Diversos autores lo definen como:

Es el análisis de un estudio estadístico con valores cogidos a lo largo del tiempo, teniendo en cuenta periodos que pueden estar comprendidos por días, meses o años, en una escala ordenada (Ollé, 2017). Según Ruiz (2004), nos indica que “una serie temporal es un conjunto de datos, que corresponden a un conjunto de fenómenos que van ordenados de acuerdo con el tiempo en el que se los establece”.

Formalmente, es una secuencia ordenada cronológicamente de valores de medición sobre el estado de una variable de un fenómeno o proceso; a menudo, dependen unas de otras. Esta dependencia, jugará un papel importante en el análisis de los datos de la serie.

Así entonces, la temperatura máxima diaria, los procesos de un artículo, tasas de desempleo, velocidad del viento hasta los esfuerzos y temperaturas en diferentes puntos de una obra civil instrumentada se pueden representar en una serie. (Ollé, 2017)

Las observaciones de una serie de tiempo pueden ser denotadas como:

$$Y_1 + Y_2 + Y_3 + \dots + Y_{t-1} + Y_t$$

$$t = 1, 2, 3, \dots, \text{etc.}$$

Y_t: es el valor que toma el proceso en el instante *t*

Modelos de series de tiempo

Modelos deterministas: Vienen a ser métodos de extrapolación simples, en los que no se refieren a las fuentes ni a la naturaleza de la aleatoriedad subyacente en la serie. Su relativa sencillez suele ir acompañada de una menor precisión. Un ejemplo de un modelo determinista es un modelo de promedio móvil en el que la predicción de una variable se calcula a partir del promedio de los valores "n" inmediatamente anteriores. (Ollé,

2017)**Modelos estocásticos:** Es la descripción del proceso aleatorio subyacente en la serie de manera esquematizada. Es decir, se asume que la

serie observada Y_1, Y_2, \dots, Y_t se obtiene de un grupo de variables aleatorias con cierta distribución conjunta complicado de establecer, por lo cual, que se crean modelos aproximados que sean de utilidad para la formación de pronósticos. (Ríos, 2008)

Tipos de Series

Serie no estacionaria: es donde cuyas características de media, varianza y covarianza varía en el tiempo, lo que dificulta el modelado. No obstante, en muchas veces, si la serie es diferenciada una o más veces, esta resulta ser estacionaria.

Serie estacionaria: es donde la media y varianza no son constantes a través del tiempo, y su covarianza sólo está dado en función del rezago. Gracias a ello, se puede modelar la serie mediante una ecuación con coeficientes fijos estimados de los datos pasados. (Ríos, 2008)

Objetivos del análisis de una serie de tiempo.

Ríos (2008) fundamenta que un proceso para la realización de un modelo ARIMA orden (p, d, q) se realiza de forma iterativa, en donde se identifican 4 etapas:

Identificación: Usando valores ordenados en el tiempo, intentaremos recomendar el modelo ARIMA de orden (p, d, q) que sea digna de ser investigada. El objetivo radica en establecer (p, d, q) que sean adecuados para el desarrollo de la serie. Se identificarán varios modelos candidatos.

Estimación: para la inferencia de los parámetros, se debe realizar a partir del modelo apropiado de la serie de tiempo.

Validación: en esta etapa se realiza la contrastación del diagnóstico para validar si el modelo obtenido se ajusta a los datos, si resulta ser negativa, se escoge otro modelo y se repite el proceso

Predicción: se utiliza el modelo seleccionado para realizar los pronósticos.

Componentes de una serie de tiempo

Ruiz (2004), afirma que hay cuatro clases básicas de variación dentro de una serie, los cuales se superponen u operan simultáneamente, contribuyendo a las modificaciones que se observan durante un lapso de tiempo generando que la serie parezca errática. Asimismo, asumiremos que se relaciona multiplicativamente los 4 componentes; lo que significa, cualquier valor dado en la serie, es el resultado del producto de factores que tienen la facilidad de atribuir a cuatro componentes, estos son:

Tendencia: La tendencia de una serie se da por los factores a largo plazo. Representa los patrones graduales y consistentes de las alteraciones de la misma serie, resultado de fuerzas persistentes que causan ya sea el incremento o decrecimiento, estas pueden ser: modificaciones de la población, en sus cualidades demográficas, en la salud, modificaciones en sus ingresos, su nivel educativo y tecnológico.

Variaciones Cíclicas: las series de tiempo muestran con regularidad sucesiones de puntos alterno por encima o debajo de la línea de tendencia, estas pueden durar más de un año, esta alteración u variación continúa después de que se han quitado las variaciones estacionales e irregulares.

Variaciones Estacionales: El componente estacional representa la variabilidad que hay en los datos, esto se da debido a las influencias de las estaciones. Esta variación recurre año tras año en los mismos meses o en los mismos trimestres, se pueden presentar con más o menos intensidad.

Variaciones Irregulares: suceden por las causas a corto tiempo, repentinas y poco frecuentes que influyen a la serie. Dado que este elemento (componente) expresa la variación casual de la serie, es indispensable, no dejar esperar para predecir su impacto en la serie. Se divide en: a) Las variaciones que se identifican con facilidad, las cuales son causadas por acontecimientos especiales, como, por ejemplo, huelgas, terremotos, inundaciones. b) Variaciones que no se pueden identificar de manera precisa, pero que se equilibran con el tiempo, éstas se llaman variaciones por casualidad.

1.2.3. Modelos de Predicción

Según Ollé (2017), los modelos más sencillos, estos son los siguientes:

Modelos Intuitivos: Se usan muy a menudo, ya sea que todos los usemos conscientemente o no, tiene la gran desventaja, al basarte en la intuición (un

aspecto completamente individualizado), la designación de los valores a una variable puede variar mucho si más de una persona se involucra en el proceso. Por lo tanto, el método utilizado más simple es asumir que la variable tendrá el mismo comportamiento que la presentada en un período anterior, se expresa como:

$$Y_{i+1} = Y_i$$

Donde:

Y_{i+1} es el valor de la variable que toma en el período siguiente.

Y_i es el valor de la variable que toma en el periodo anterior.

Estos modelos son muy importantes en circunstancias nuevas, en donde no hay registros de cómo se comporta la variable y así puedan complementarse de mejor manera a la medida que se pueda registrar datos sobre tendencia y estacionalidad.

Modelos de media móvil (MA): Utilizan el valor promedio de un conjunto de datos de un período anterior, así como del valor de la variable en el futuro. Se denominan móviles porque al disponer de un nuevo dato se elimina el más antiguo, permitiendo ir actualizando el valor pronosticado, el modelo puede expresarse matemáticamente de la manera siguiente:

$$Y_{i+1} = \frac{Y_i + Y_{i-1} + Y_{i-2} + Y_{i-3} + \cdots + Y_{i-m+1}}{m}$$

Donde:

Y_{i+1} : valor del pronóstico en siguiente período.

Y_i : valor en el periodo anterior.

m: número de datos.

La desventaja radica en que el valor de las variables que participan se considera iguales, repercutiendo en la pérdida de alguna característica singular de la variable en estudio.

Método de Suavizamiento Exponencial: Se llama así porque las ponderaciones tomadas a períodos anteriores en el pronóstico disminuyen exponencialmente, lo que significa que los valores de mayor antigüedad participarán en menos medida en la estimación, siendo cercanas a cero, pero jamás se deja de considerar. Matemáticamente se expresa de la siguiente manera:

$$P_{i+1} = \beta Y_i + (1 - \beta)P_i$$

Donde:

P_{i+1} : viene a ser valor de pronóstico en el nuevo periodo.

Y_i : valor real actual.

P_i : Valor del pronóstico para el período anterior.

β : es la constante, su valor se encuentra de 0 hasta 1.

β es la constante clave de suavizamiento, su elección determinara la respuesta del modelo a los cambios en el valor de la variable, el cual dependerá de la experiencia pasada, los valores que se acerquen a 1 favorecerán al más reciente una mayor participación en la predicción, en cambio, cerca de 0, lo evitan, obteniendo una previsión del mismo valor que

el valor del período anterior. En cualquier modelo sencillo elegido, es fundamental medir el error que se comete al hacer la comparación de su valor con el que en realidad ocurrió, su propósito se centra en hacer cada vez mejor estimaciones.

1.2.4. Modelos ARIMA, Metodología Box-Jenkins

La metodología ARIMA, comúnmente se le conoce como, “la metodología de Box-Jenkins”, este se fundamenta sobre el estudio de las cualidades estocásticas de las series de tiempo económicas.

Mota (2016), refirió que, Box y Jenkins realizaron un método para identificar, estimar y validar modelos dinámicos de series temporales, donde, el tiempo es la variable base. Una sección de esta metodología está realizada para evitar la tarea de especificar el modelo al azar y dejar que los datos históricos de la variable señalen las particularidades de la estructuración subyacente probabilística. Box y Jenkins plantearon una familia de modelos de series de tiempo que se ajustaran para expresar cómo evoluciona a través del tiempo una variable.

Asimismo, menciona que en 1976 Box y Jenkins formalizó la metodología de los modelos, también conocida como Box-Jenkins. Este punto de vista nace del hecho de que la serie temporal que quiere predecir se genera por un proceso estocástico cuya esencia puede ser caracterizada por un modelo. Para estimar el modelo ARIMA se necesita de una serie de tiempo, ya sea mensual o trimestral con una gran cantidad de observaciones.

Principalmente, consiste en hallar un modelo matemático que simbolice cómo se comporta una serie de tiempo, y permita realizar predicciones solo introduciendo el periodo de tiempo que corresponde. En los modelos ARIMA univariantes el comportamiento de una serie temporal se explica en base a sus propias observaciones pasadas y en base de los errores pasados de predicción. Su connotación es de la siguiente manera: *ARIMA* (p, d, q).

$$Y_t = \phi_0 + \phi_1 y_{t-1} + L + \phi_p y_{t-p} + a_t - \phi_1 a_{t-1} - L - \phi_q a_{t-q}$$

Donde:

p: n° de parámetros AR

d: n° de diferencias

q: n° de parámetros de MA

Su ecuación es de la siguiente manera:

1.2.5. Etapa de estimación

Las estimaciones de los parámetros se realizan de acuerdo al modelo identificado en la primera etapa; por lo tanto, aparece una familia de procesos estacionarios, como son AR (p), MA(q), ARMA (p, q), ARIMA (p, d, q) y SARIMA (p, d, q) (P, D, Q); los cuales se describen a continuación.

Proceso autorregresivo “AR (p)”

Para Ollé (2017), estos modelos son útiles para representar la dependencia de los valores de una serie temporal respecto a su pasado. Este proceso se obtiene como caso particular de la forma de la ecuación de diferencias en donde el polinomio de medias móviles es unitario $\theta(B) = 1$. Y desde mucho

tiempo este proceso goza de una popularidad, debido a su fácil manipulación, sobre todo en la etapa de estimación.

Definición: Una serie de tiempo (Z_t), es llamado un proceso autorregresivo de orden "p" o abreviadamente denotado por AR (p), si el valor de la variable en un instante cualquiera de tiempo puede ser explicada por los valores en los instantes anteriores más un término error, el caso más simple ARIMA (1, 0, 0) o AR (1):

Modelo autorregresivo de orden "p"

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t; \quad t = 1, 2, 3, \dots, t$$

Donde:

Φ : coeficiente del modelo.

Θ_0 : constante del modelo (está presente en el modelo siempre que sea significativo).

p: orden del proceso.

ε_t termino error ($\varepsilon_t \sim N(0, \sigma^2 \varepsilon)$)

Modelo de orden "p" en su forma compacta

A través de algunas transformaciones, se puede obtener el modelo AR (p) en una forma compacta;

$$\phi(B)Z_t = \theta_0 + \varepsilon_t; \quad t = 1, 2, \dots, T \quad \varepsilon_t \sim N(0, \sigma^2 \varepsilon)$$

Modelo autorregresivo estimado de orden “p”

$$\hat{Z}_t = \hat{\theta}_0 + \hat{\phi}_1 Z_{t-1} + \hat{\phi}_2 Z_{t-2} + \dots + \hat{\phi}_p Z_{t-p}$$

Dónde:

\hat{Z}_t : Valor de predicción para la variable Z en el periodo “t”

Propiedades del proceso autorregresivo de orden “p”

1. La media de la serie es calculada por;

$$E(Z_t) = \mu = \theta_0 / (1 - \phi_1 - \phi_2 - \dots - \phi_p)$$

2. La varianza de la serie es dada por;

$$VAR(Z_t) = VAR(\theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t)$$

$$Y_0 = \sigma^2 \varepsilon / (1 - \phi_1^2 - \phi_2^2 - \dots - \phi_p^2)$$

Proceso de media móvil “MA (q)”

Es la explicación del valor de una variable durante un t-período determinado, sobre la base de términos independientes y una secuencia de errores correspondientes a períodos precedentes, que se ponderaron oportunamente. Tiene “q” números de error, según Ollé (2017) se expresa de la forma siguiente:

Modelo media móvil de orden "q":

$$Z_t = \theta_0 + \varepsilon_{t-1} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}; \quad t = 1, 2, 3, \dots, t$$

Donde:

θ : Coeficientes del modelo.

Θ_0 : Constante del modelo (está presente en el modelo siempre que sea significativo).

q : Orden del proceso.

ε_t : Terminio error ($\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$)

Modelo de media móvil de orden “q” en su forma compacta

A través se algunas transformaciones, se puede obtener el modelo MA

(q) en una forma compacta;

$$Z_t = \theta_0 + \theta(B)\varepsilon_t; t = 1, 2, 3, \dots, t; \quad \varepsilon_t \sim N(0, \sigma^2_\varepsilon)$$

Modelo de MA estimado, Orden “q”

$$\hat{Z}_t = \hat{\theta}_0 - \hat{\theta}_1\varepsilon_{t-1} - \hat{\theta}_2\varepsilon_{t-2} - \dots - \hat{\theta}_q\varepsilon_{t-q}$$

\hat{Z}_t : Valor de predicción para la variable Z en el periodo “t”.

Propiedades del proceso de MA de orden “q”

1. La media de la serie es calculada por;

$$\mu = E(Z_t)$$

$$\mu = E(\theta_0 + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}) = \theta_0$$

2. La varianza de la serie es calculada por;

$$\gamma_0 = Var(Z_t) = (1 + \theta^2_1 + \theta^2_2 + \dots + \dots + \theta^2_q)\sigma^2_\varepsilon$$

Proceso mixto autorregresivo de media móvil, ARMA (p, q).

Una serie de tiempo Z_t , es llamado un proceso autorregresivo, con media móvil p y q o abreviadamente denotado por ARMA (p; q), si el valor de la variable en un instante cualquiera de tiempo puede ser explicada por los valores en los instantes anteriores y de errores secuenciales correspondientes a períodos anteriores más un término error. (Ruiz, 2004)

Modelo mixto ARMA (p, q)

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots - \phi_p Z_{t-p} = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Donde:

Θ y Φ : Coeficientes del modelo.

θ_0 : Constante del modelo (está presente en el modelo siempre que sea significativo).

p y q: Valores del proceso.

ε_t : Término error ($\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$).

Modelo mixto ARMA (p, q) en su forma compacta

A través de algunas transformaciones, se puede obtener el modelo

ARMA (p, q) en una forma compacta;

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

$$\Phi(B) Z_t = \theta_0 + \theta(B) \varepsilon_t$$

Donde los polinomios $\Phi(B)$ y $\theta(B)$, son definidos de la siguiente forma:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

Modelo mixto ARMA (p, q) estimado

$$Z_t = \hat{\theta}_0 + \hat{\phi}_1 Z_{t-1} + \hat{\phi}_2 Z_{t-2} + \dots + \hat{\phi}_p Z_{t-p} - \hat{\theta}_1 \varepsilon_{t-1} - \hat{\theta}_2 \varepsilon_{t-2} - \dots - \hat{\theta}_q \varepsilon_{t-q}$$

Z_t: Valor de predicción para la variable Z en el periodo “t”

Proceso autorregresivo integrado de medias móviles ARIMA (p, d, q)

Dado que la no estacionariedad puede darse tanto en el nivel medio de la serie, como también en la varianza, entonces tales efectos deben ser considerados en la estructura del modelo definido a continuación (Ríos, 2004).

Modelo ARIMA (p, d, q)

$$\begin{aligned} & (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Z_t \\ & = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \end{aligned}$$

Modelo ARIMA (p, d, q) en su forma compacta

A través de algunas transformaciones, se puede obtener el modelo ARIMA (p, d, q) en una forma compacta:

$$\phi(B) \Delta^d Z_t = \theta(B) \varepsilon_t, \quad t = 1, 2, 3, \dots, t$$

Donde; $\theta(B)$ y $\Phi(B)$, son los operadores autorregresivos y medias móviles respectivos,

Δ^d : es el operador de diferencias, y se aplica cuando existe no estacionariedad en media.

$$\Delta^d Z_t = (1 - B)^d$$

Z_t ; Por lo general $d=1$ o 2

Generalmente es aplicable a cualquier modelo. Si existe alguna de las componentes p, d, q igual a 0, se procede a eliminar el término de la fórmula general.

Proceso autorregresivo integrado de media móviles con parte estacional “SARIMA (p, d, q) (P, D, Q)”

Se caracterizan por fluctuaciones cíclicas, también conocidas como variaciones estacionales. Las series con tendencias seculares y variaciones cíclicas se pueden representar mediante los modelos SARIMA (p, d, q) (P, D, Q). las letras en minúsculas hacen referencia la parte regular y las letras en mayúsculas hacen referencia a la variación estacional de la serie, hay que tener en cuenta que la parte estacional puede cambiar, si la estacionalidad es trimestral será $S=4$ y si la estacionalidad es mensual será $S=12$. (Ollé, 2017)

Modelo SARIMA (p, d, q) (P, D, Q)₁₂

$$\phi_p(B)\phi_p(B^S)(1 - B^S)^D(1 - B)^d Z_t = \phi_0 + \theta_q(B)\theta_q(B^S)\varepsilon_t$$

Es decir, es un modelo autorregresivo con parte media móvil, de orden "p, d

y q", el cual presenta estacionalidad de mes a mes de orden "P" para el lado autorregresivo y "Q" en la parte de media móvil, después de haber sido diferenciada D veces. (Sanchez & Varillas, 2014)

1.2.6. Identificación del Modelo

En esta primera parte se calcula la función de autocorrelación total y parcial de la serie (tener la seguridad de que cumplen con las propiedades que describen a un proceso estacionario estocástico) y comparar sus correlogramas con los equivalentes a modelos teóricos MA (q), AR(p), y ARMA (p, q).

Una vez determinado el modelo, lo estimaremos, y los correlogramas de los residuos deben tener a un ruido blanco, caso contrario no suceda eso, se procederá a realizar nuevamente la estimación, en la que se debe incorporar la estructura más cercana al modelo teórico.

Para tener conocimiento cuándo estamos entre un ruido blanco, se puede realizar las comparaciones que se menciona a continuación (Sanchez & Varillas, 2014):

Media Nula: verificar en la gráfica de los residuales, si el valor del error se desplaza a cero, caso contrario, calcular el cociente de los residuos que surge entre la media y varianza muestral, si el valor de la ratio obtenido es menor a 2 y $e = 0.05$, se concluirá que la media no es estadísticamente significativa diferente a cero.

Varianza constante: se tiene que verificar el gráfico de residuales que la varianza sea constante. En caso de heterocedasticidad, se recomienda transformar logarítmicamente la serie inicial.

Incorrelación: el valor de los coeficientes de autocorrelación de la muestra de los residuales no debe sobrepasar el valor de las bandas de significancia al 5% ($\pm 1.96 (1/ T^{1/2})$). En donde:

$T^{1/2}$: valor aproximado de la varianza asintótica; valores grandes de “j”.

“Para los términos que están cerca a cero, se recomienda hacer uso de una distinta amplitud de bandas como por ejemplo $\pm (1/T^{1/2})$ ”. (Angulo, 2013)

Prueba de Dickey – Fuller (DF)

Se emplea para realizar el contraste de no estacionariedad, debido a que H_0 indica que existe una raíz unitaria en el proceso que genera valores en la serie estudiada. (Ríos, 2008)

Definición básica:

$$H_1: y_t = a_1 y_{t-1} + \varepsilon_t$$

$$H_0: y_t = y_{t-1} + \varepsilon_t$$

Para el parámetro estimado “ a_1 ” su distribución de probabilidad exhibe una discontinuidad cuando $a_1 = 1$, dada la no estacionariedad en varianza de y_t el modelo utilizado es:

$$Y_t - y_{t-1} = a_0 + a_1 y_{t-1} - y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = a_0 + (a_1 - 1)y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = a_0 + \gamma \cdot y_{t-1} + \varepsilon_t$$

$$H_0: \gamma = 0 \quad \text{respecto a} \quad H_1: \gamma < 0$$

Los datos base para contrastar el parámetro γ depende de la elección que se hizo del proceso generador de valores.

- Modelo 1 (simple) (τ):

$$\Delta Y_t = \gamma \cdot y_{t-1} + \varepsilon_t$$

- Modelo 2 (con constante) ($\tau\mu$):

$$\Delta Y_t = a_0 + \gamma \cdot y_{t-1} + \varepsilon_t$$

- Modelo 3 (con constante y tend. determinística) ($\tau\mu$):

$$\Delta Y_t = a_0 + \gamma \cdot y_{t-1} + a_2 t + \varepsilon_t$$

1.2.7. Estadístico de Prueba

Se realiza la estimación del modelo planteado y se procede a calcular el valor estimado de “t” del parámetro estudiado. Dicho valor resulta de la estimación de los 3 modelos empleando el MCO. Para ello, se divide el coeficiente estimado de y_{t-1} para todos los casos con su error estándar.

1) Criterio de decisión

Después del cálculo, se compara con el valor experimental base obtenido utilizando las “tablas de Dickey y Fuller o de MacKinnon”. Si la estimación de $\hat{\rho}$ es menor en valor absoluto de la tabla con un cierto nivel de confianza, se toma la decisión de aceptar la hipótesis nula, es decir, se acepta la presencia de raíz unitaria.

2) Contraste de la “Q” de Box – Pierce.

Se realiza el análisis de la hipótesis nula planteada:

$$H_0: \rho_1(a) = \rho_2(a) = \rho_3(a) = \dots \dots \rho_M(a) = 0$$

Se supone la expresión:

$$Q = T \sum_{j=1}^M r^2 j (\hat{a})_t$$

O la alternativa propuesta por Ljung-Box:

$$Q^* = T(T + 2) \sum_{j=1}^M (T - j)^{-1} r^2 j (\hat{a})_t$$

tiene una distribución $\sim X^2$, con $g = M - k$

3) Criterios de selección de modelos

La construcción de los principios de información se realizó como estimadores insesgados de la log - verosimilitud esperada $\eta(G; \hat{\theta}) \equiv E_{G(Z)}[\ln f(Z/\hat{\theta})]$ o si lo vemos desde el punto de vista predictivo, es equivalente a la discrepancia de la información de Kullback-Leibler entre la distribución verdadera de g estadístico $f(z/\hat{\theta})$.

Una particularidad importante de los criterios informativos es que se combinan con los métodos de máxima verosimilitud, en modelos causales se basa en la suposición que las variables que se analizan tienen distribución normal multivariante. Asimismo, para fines de comparación, R^2 ajustado resulta ser una mejor medida que R^2 , porque no penaliza cuando se agregan más variables regresoras.

4) R^2 ajustada

Es la capacidad explicativa que tiene la variable X sobre la variable Y. Cuando otra variable regresora se incluye al modelo, se obtiene un mayor nivel explicativo que con una sola, por lo que la variable X seguirá siendo explicativa”.

En la literatura econométrica podemos encontrar varias soluciones al problema del incremento artificial del valor de R^2 . Como una solución se propone un coeficiente de determinación corregido o ajustado, y se denota por $\overline{R^2}$, y definido como:

$$\overline{R^2} = 1 - \frac{\frac{s_e^2}{n-k}}{\frac{s_y^2}{n-1}}$$

En donde:

k: n° de parámetros que hay en el modelo, se cuenta el término independiente.

Realizar la comprobación que mantiene R^2 y $\overline{R^2}$ es fácil.

$$\overline{R^2} = 1 - (1 - R^2) \frac{n-1}{k-1}$$

Observemos que para $k > 1$, $\overline{R^2} < R^2$, lo cual implica que a medida que el número de variables exógenas aumenta, $\overline{R^2}$ ajustado aumenta menos que R^2 no ajustado. Observemos que, en este caso, ($k > 1$), si $\overline{R^2} = 0$ R^2 puede ser negativa, a pesar de que R^2 sea una magnitud no negativa. Si esto ocurre $\overline{R^2}$, se interpreta como si su valor fuese 0. (Ruiz, 2004)

Establecida esta relación entre los dos coeficientes, podemos afirmar que R^2 corregido tiene la característica de ser neutral cuando se agregan variables adicionales. En opinión de algunos autores, es mejor utilizar $\overline{R^2}$ en lugar de R^2 , porque R^2 alcanza a dar un aspecto muy optimista del ajuste de la regresión, particularmente surge cuando la cantidad de variables explicativas no es muy pequeño comparado con el número de observaciones, lo que podríamos considerar como un grado de libertad del modelo inadecuado. (Ruiz, 2004)

5) Criterio de información de Akaike (CIA)

Es una serie de principios que surgen de la información estadística. La principal idea es penalizar el exceso del número de parámetros ajustados, lo que no puede lograr hacer la prueba chi-cuadrado. El AIC es el estimador muestral de $E[\ln f(X|\theta)]$, según Ruiz (2004) su expresión general es:

$$CIA(k) = -2\ln\mathcal{L}[\hat{\theta}(k)] + 2k \quad \text{donde:}$$

$\mathcal{L}[\theta(k)]$ representa la función de verosimilitud de las observaciones.

$\hat{\theta}(k)$ valor máximo de estimación máximo verosímil del vector de parámetros θ (en donde, $\theta = (\mu, L, \Psi)$).

k : n° de parámetros independientes que se estiman dentro del modelo.

6) Criterio de información Schwartz (CIB)

En los procedimientos basados en la verosimilitud, el CIA podría no ser asintóticamente aceptable, por lo que se presenta un criterio de información alternativo en base al enfoque bayesiano, “Bayesian Information Criterion” (BIC), con esta regla se penaliza la cantidad de parámetros con \ln , en lugar de 2.

$$BIC(k) = -2\ln\mathcal{L}[\hat{\theta}(k)] + (\ln n)k$$

Donde $\mathcal{L}[\theta(k)]$ representa la función de verosimilitud de las observaciones;

$\hat{\theta}(k)$, representa el valor verosímil máximo estimado del vector de

parámetros $\theta = (\mu, L, \Psi)$ y k , representa el número de parámetros independientes que se estiman dentro del parámetro es el tamaño muestral.

1.2.8. Producción de espárragos

Es un cultivo permanente rentable para la producción agrícola a tiempo parcial y menor proporción con alta coyuntura en el mercado. Puede ser exportado a diferentes partes del mundo.

CAPITULO II. MÉTODOS Y MATERIALES

2.1. Diseño de contrastación de Hipótesis

El diseño de investigación fue observacional retrospectivo y prospectivo, de tipo teórica, empírica y longitudinal dado que se enfoca en encontrar justificación lógica del comportamiento de la producción agropecuaria mensual de espárrago, con el propósito de crear modelos estocásticos y matemáticos que permitan realizar pronósticos.

A su vez fue no experimental, debido a que no existe intención del investigador por manipular alguna de las variables en estudio, más si observa lo que ocurre dentro del contexto. (Sampieri, Batista, & Fernandez, 2014)

2.2. Variables

Variable dependiente: Producción mensual de espárragos periodo enero 2007 – diciembre 2018.

Variables independientes: Tiempo (año/meses)

2.3. Población y muestra

Población

La población estuvo constituida en miles de toneladas por toda la producción mensual de espárrago en el Perú. Esta población mensual está definida a nivel de un Proceso Estocástico.

Muestra

La muestra estuvo constituida por la producción mensual en miles de toneladas de espárrago en el Perú, en el periodo de enero a diciembre durante los años 2007-2018.

2.4. Técnicas e instrumentos de recolección de datos.

La técnica a emplear será el análisis documental y su instrumento será guía del análisis documental. Para recolectar los datos se empleó una fuente secundaria (página web del “BCR”), de donde se ha obtenido la producción mensual de espárrago en el Perú, periodo enero 2007 – diciembre 2018.

Metodología Box y Jenkins

Se empleó el método de Box – Jenkins para encontrar el modelo óptimo para pronosticar la producción mensual de espárragos en el Perú periodo enero 2007- diciembre 2018.

2.5. Análisis y Procesamiento Estadístico de los datos

Para el análisis descriptivo de las series, se calculará el promedio, varianza, desviación estándar y el coeficiente de variación, empleando el software RStudio versión 4.1.

A continuación, se detallan las etapas de la metodología ARIMA, en la cual se incluyen los modelos estadísticos descriptivos como inferenciales que se emplearán:

2.5.1. Identificación del Modelo

Se realiza el gráfico original de la serie con el objetivo de observar cómo se comporta la estacionariedad, tendencia y estacionalidad.

Se determinarán los componentes p , d y q del modelo ARIMA en donde:

p : n° de componentes no estacionales para la parte autorregresiva AR.

d : n° de diferenciaciones no estacionales.

q : n° de componentes no estacional para la parte media móvil (MA).

en esta primera etapa, también se analizará:

con el uso del programa estadístico RStudio 4.1, se verificará si la serie aplicando diferencias es estacionaria.

2.5.2. Estimación

Para estimar del modelo propuesto se empleó el Rstudio 4.1, se eligió el programa por ser versátil en la facilitación del método de prueba y error para identificar el modelo, el cual permite realizar las predicciones, grabar valores de los coeficientes estimados, los datos que se estiman, así como de los residuales, lo que posteriormente servirán para constatar la bondad de ajuste del modelo estimado.

El análisis de los gráficos de autocorrelación total y parcial, ayudarán a reconocer los componentes AR y MA.

2.5.3. Adecuación del Modelo

Para realizar este proceso, se desarrollarán los indicadores que se describen a continuación:

- Las pruebas de significancia de los coeficientes del modelo
 - Cuantificar el error de estimación del modelo
 - Analizar los residuos, estudiar qué es lo que comprende
- a) Para determinar si existe independencia de los errores en el tiempo, se empleará la prueba de Ljung - Box. El cual se desarrollará de manera progresiva para los rezagos 12,24 y 36.
 - b) Analizar las funciones de autocorrelación de los residuales, el total (FAC) y el parcial (FACP).
 - c) Verificar si los residuales tienen distribución normal mediante la técnica de Kolmogorov Smirnov. Asimismo, se obtendrán la hoja probabilística, el histograma y ploteo con los datos de la predicción y repartidos a través del tiempo.

2.5.4. Pronóstico con el Modelo

Con la estimación del modelo se pronosticará la producción mensual de espárrago en el Perú, para el año 2019-2020.

2.5.5. Validez del Pronóstico

Se corroboran los datos reales de la producción mensual de espárrago en el Perú con los pronósticos que se realicen con el modelo estimado para los meses de enero 2019 hasta diciembre del 2020. La evaluación de los errores del pronóstico, se realizará con los

siguientes métodos.

- Error absoluto promedio (MAE).
- Raíz del error cuadrático promedio (RMSE).
- Error absoluto porcentual promedio (MAPE).

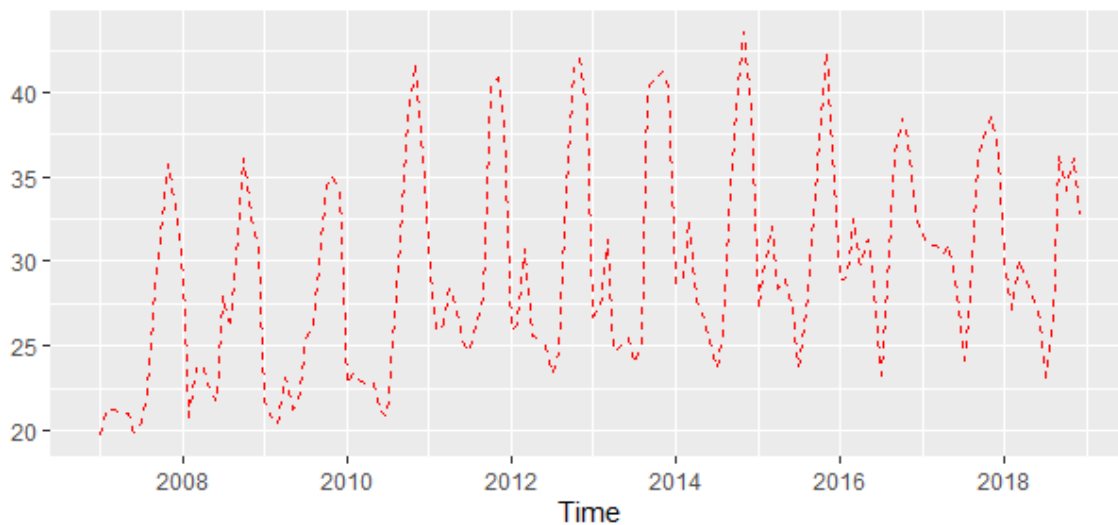
CAPITULO III. RESULTADOS Y DISCUSIONES

3.1. Resultados

En este apartado se muestran los datos de la serie de producción de espárragos en el periodo 2007 – 2018, donde se construyeron los mejores modelos de pronóstico ARIMA, buscando conocer cómo será la producción en los últimos 2 años. Antes de ello se conoce la secuencia de los datos originales.

Figura 1

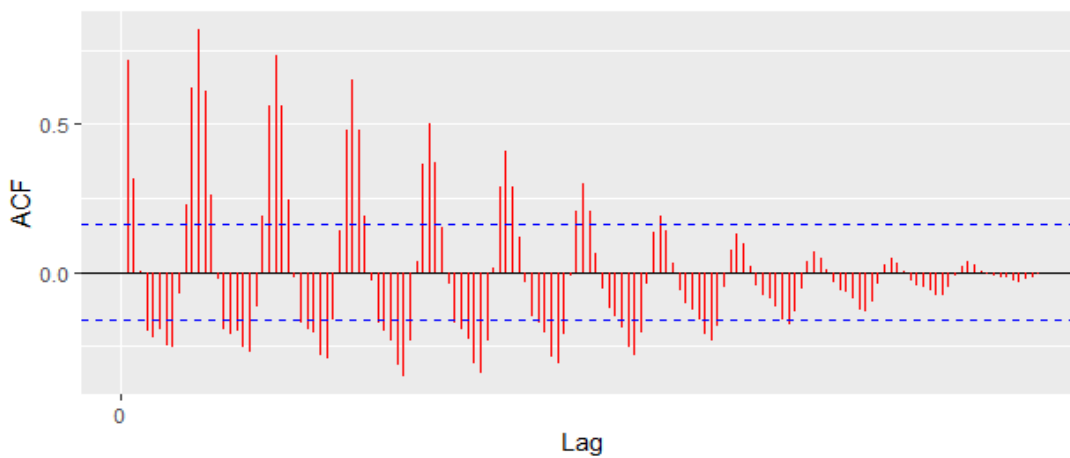
Comportamiento de la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018.



En la figura 1 se muestra la secuencia de la producción de espárragos en el periodo 2007 – 2018, donde claramente se visualiza un periodo de estacionalidad y un crecimiento lento, es decir, no es constante ni en media ni en varianza. La serie no es estacionaria por motivos que la producción durante los años 2007 – 2018, no se presentan constantes ni en media ni en varianza, ante ello, surgen esos picos en la serie, dando indicios que la serie puede ser transformada tanto en la parte estacional, como la parte regular.

Figura 2

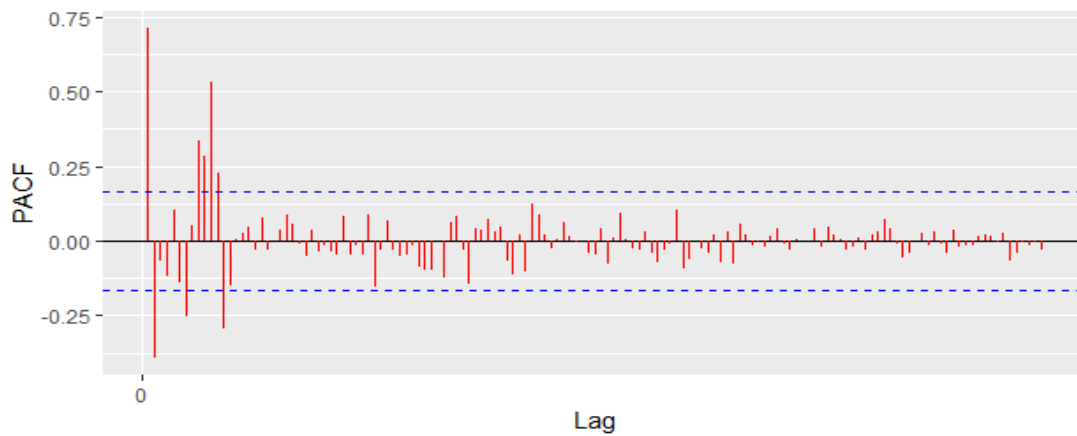
Función de autocorrelación simple de la producción de espárragos 2007-2018.



La figura 2 muestra cómo la serie, mediante la función de autocorrelación tiene un decaimiento lento, decrece lentamente, esto da más indicios que la serie no es estacionaria, además se nota claramente como las bandas de confianza son superadas por los retardos. Los datos de producción de espárragos en el periodo 2007 – 2018, debe ser transformada en primeras diferencias en la parte regular y estacional.

Figura 3

Función de autocorrelación parcial de la producción de espárragos 2007-2018.

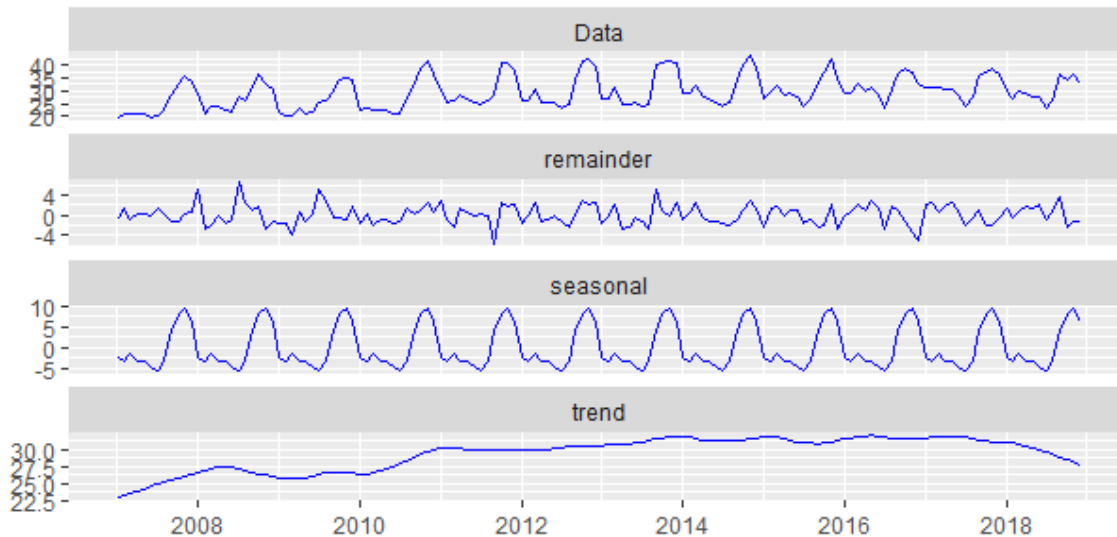


En cuanto a la figura 3, la función de autocorrelación parcial, evidencia componentes autorregresivos, es decir la estructura del modelo ARIMA, puede presentar combinación de componentes autorregresivos y media móvil.

OE 2. Analizar la estacionariedad y estacionalidad de la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018.

Figura 4

Descomposición de los datos de producción de espárragos, período 2007 – 2018.



En la figura 4 claramente se muestra cómo se descomponen los datos de producción de espárragos durante periodo 2007 - 2018, llamando la atención la parte estacional que se observa, pues se va contrastando visualmente que debemos quitar las estaciones de producción para poder realizar la metodología de Box Jenkins. En cuanto a la parte regular, no se evidencia mucha tendencia creciente, pero sería necesario conocer si se debe aplicar primeras diferencias. Ante ello, presenta estacionalidad, conllevando al estudio a realizar diferenciación en la parte estacional del modelo.

Número de diferencias en la parte regular y parte estacional

```
ndiffs = 1
nsdiffs = 1
```

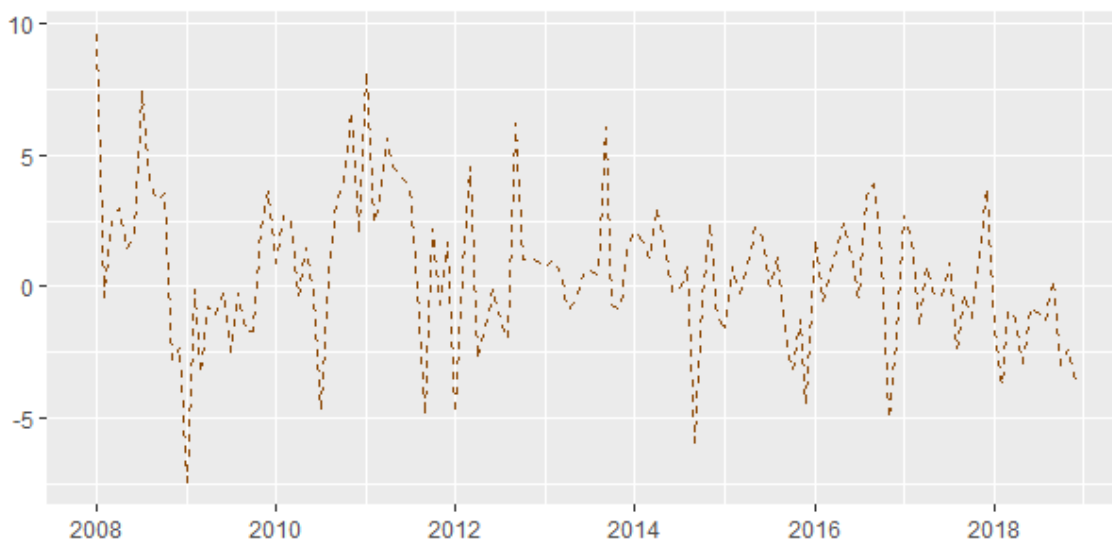
Los comandos de R Studio, ndiffs y nsdiffs, representan el número de diferencias

tanto en la parte regular, como en la parte estacional que debe contener la serie de producción de espárrago en el periodo 2007 – 2018, para que se convierta en estacionaria. Dicho esto, los comandos señalan que la serie debe ser transformada en primeras diferencias, tanto en la parte regular como en la parte estacional.

OE 3. Identificar el modelo de serie de tiempo para describir la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018.

Figura 5

Serie transformadora en primeras diferencias tanto en la parte regular y estacional de la producción de espárragos en el período 2007 - 2018.



En la figura 5 se muestra a la serie de datos de producción de espárragos transformada en primeras diferencias para la parte regular y estacional, evidenciando, que ya no tiene parte estacional, y quedando dudas para ver si es constante en media. Para ello, debe aplicarse la prueba de Dickey-Fuller, la cual contrasta si la serie tiene raíz unitaria o no.

Prueba de estacionariedad Dickey-Fuller

H_0 : Los datos de producción de espárragos en primeras diferencias no es estacionaria,

tiene raíz unitaria.

H_1 : Los datos de producción de espárragos en primeras diferencias es estacionaria, no tiene raíz unitaria.

$\alpha = 5\%$

Solución

Test Dickey Fuller

p-valor = 0.01237455

Prueba KPSS

p-valor = 0.04661168

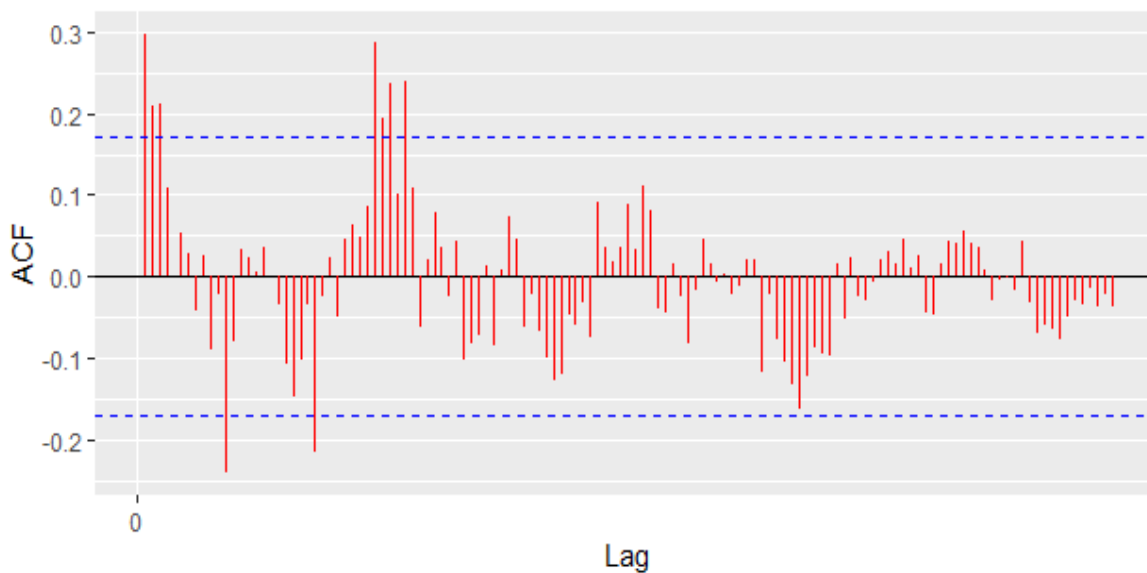
Test de Phillips-Perron

p-valor < 0.01

La prueba KPSS confirma que la serie es estacionaria, esto gracias a su p-valor el cual es igual a 0.0466, siendo un número menor al 0.05, asimismo, el p-valor del test de Dickey-Fuller y Phillips-Perron, también tiene un valor menor al 5%, por lo tanto, la serie de producción de espárragos a nivel nacional es estacionaria, luego de ser transformada en sus primeras diferencias tanto en la parte regular como la parte estacional.

Figura 6

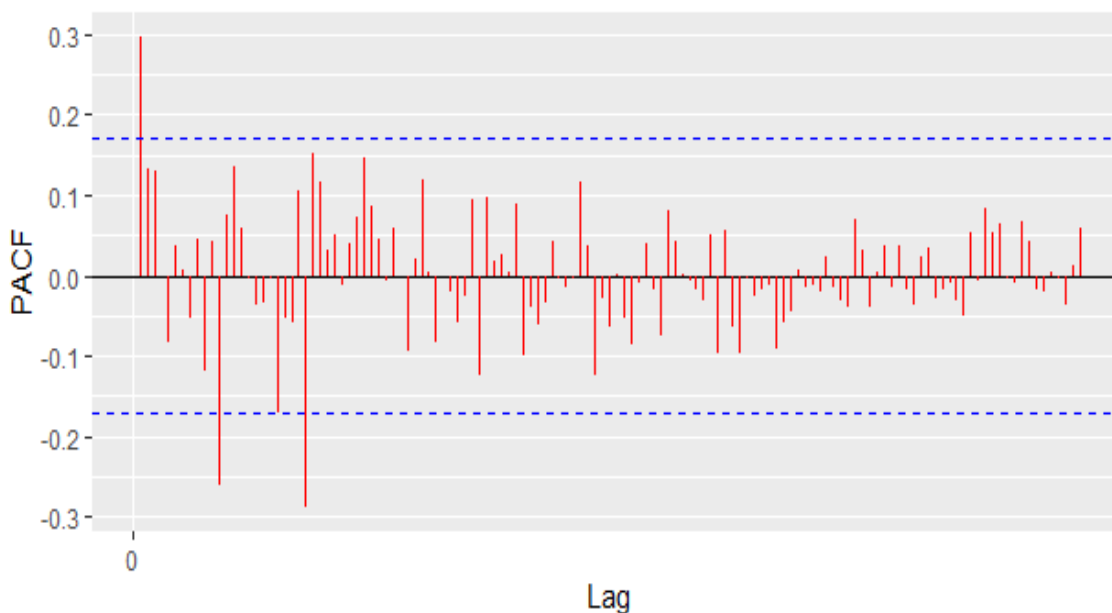
Función de autocorrelación simple de la serie producción de espárragos en primeras diferencias tanto en la parte regular y estacional.



Analizando la figura 6, acerca de la función de la autocorrelación simple, para identificar componentes autorregresivos AR, se visualiza los componentes 1, 11, y 21 fuera de las bandas de confianza, mientras que para media móvil (MA) se visualiza los componentes 1,2,3,4,13,25,33,34,35 y 37.

Figura 3

Función de autocorrelación parcial de la producción de espárragos en primeras diferencias de la parte regular y estacional.



Entonces mediante la figura 7, la función de autocorrelación parcial (AR), se identifica un componente (1) para la parte regular y ningún componente para la parte estacional (2), asimismo, en la función de autocorrelación simple se identifica un componente (2) en la parte regular y 0 componentes en la parte estacional. Sin embargo, en base a ello, se buscan los mejores modelos.

OE 4. Estimar los parámetros de series de tiempo de la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018.

Tabla 1

Estimación de coeficientes de los parámetros tentativos de los posibles modelos SARIMA

MODELO (1): SARIMA (1,1,1) (0,1,0) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	0.290334	0.095415	3.0428	0.002344 **
ma1	0.956445	0.037650	-25.4034	< 2.2e-16 ***
MODELO (2): SARIMA (1,1,1) (0,1,1) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	0.267784	0.099544	2.6901	0.007143 **
ma1	-0.921539	0.048893	-18.8480	< 2.2e-16 ***
sma1	-0.542485	0.091607	-5.9219	3.182e-09 ***
MODELO (3): SARIMA (1,1,1) (0,1,2) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	0.263040	0.117969	2.2297	0.025765 *
ma1	-0.888790	0.078002	-11.3944	< 2.2e-16 ***
sma1	-0.384895	0.117138	-3.2858	0.001017 **
sma2	-0.323451	0.143085	-2.2606	0.023787 *
MODELO (4): SARIMA (1,1,1) (1,1,0) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	0.273639	0.095576	2.8631	0.0041957 **
ma1	-0.948149	0.039506	-23.9998	< 2.2e-16 ***
sar1	-0.306514	0.088982	-3.4447	0.0005718 ***
MODELO (5): SARIMA (1,1,1) (2,1,0) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	0.258287	0.096616	2.6733	0.00751 **
ma1	-0.927777	0.041752	-22.2210	< 2.2e-16 ***
sar1	-0.390644	0.085411	-4.5737	4.792e-06 ***
sar2	-0.411818	0.084077	-4.8981	9.677e-07 ***
MODELO (6): SARIMA (1,1,1) (2,1,1) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	0.257656	0.093732	2.7489	0.00598 **
ma1	-0.938266	0.036047	-26.0288	< 2.2e-16 ***
sar1	-0.753884	0.192255	-3.9213	8.808e-05 ***
sar2	-0.506338	0.081891	-6.1831	6.287e-10 ***
sma1	0.443493	0.219877	2.0170	0.04370 *
MODELO (7): SARIMA (1,1,1) (2,1,0) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	0.73196	0.15970	4.5834	4.576e-06 ***
ma1	-1.44543	0.19403	-7.4495	9.372e-14 ***
ma2	0.45989	0.18503	2.4855	0.01294 *
MODELO (8): SARIMA (0,1,2) (0,1,1) [12]				
	Estimate	Std. Error	Z value	Pr (> z)

ma1	-0.665098	0.089574	-7.4252	1.126e-13 ***
ma2	-0.208333	0.094495	-2.2047	0.02748 *
sma1	-0.548968	0.090901	-6.0392	1.549e-09 ***
<hr/>				
MODELO (9): SARIMA (0,1,2) (2,1,0) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ma1	-0.685233	0.087103	-7.8669	3.635e-15 ***
ma2	-0.207055	0.088509	-2.3394	0.01932 *
sar1	-0.394442	0.085661	-4.6047	4.131e-06 ***
sar2	-0.416207	0.083919	-4.9596	7.063e-07 ***
<hr/>				
MODELO (10): SARIMA (1,1,0) (0,1,0) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	-0.427523	0.082279	-5.196	2.036e-07 ***
<hr/>				
MODELO (11): SARIMA (1,1,0) (0,1,1) [12]				
	Estimate	Std. Error	Z value	Pr (> z)
ar1	-0.403263	0.081753	-4.9327	8.109e-07 ***
sma1	-0.536239	0.098123	-5.4650	4.630e-08 ***

En la tabla 1 se evidencian los posibles modelos tentativos que permiten pronosticar la producción de espárragos. Aunque todos se presentan significativos, deben pasar por conocer sus estimaciones de errores a fin de saber cuál tiene el menor error para pronosticar la producción de espárragos.

Tabla 2*Crterios AIC y BIC de los posibles modelos SARIMA*

Modelo	AIC	BIC
MODELO (1): SARIMA (1,1,1) (0,1,0) [12]	635.7728	644.3984
MODELO (2): SARIMA (1,1,1) (0,1,1) [12]	614.5208	626.0216
MODELO (3): SARIMA (1,1,1) (0,1,2) [12]	611.4517	625.8277
MODELO (4): SARIMA (1,1,1) (1,1,0) [12]	626.7012	638.202
MODELO (5): SARIMA (1,1,1) (2,1,0) [12]	608.4675	622.8435
MODELO (6): SARIMA (1,1,1) (2,1,1) [12]	608.6593	625.9104
MODELO (7): SARIMA (1,1,2) (0,1,0) [12]	634.7211	646.2219
MODELO (8): SARIMA (0,1,2) (0,1,1) [12]	615.3451	626.8459
MODELO (9): SARIMA (0,1,2) (2,1,0) [12]	609.2871	623.6631
MODELO (10): SARIMA (1,1,0) (0,1,0) [12]	655.8679	661.6183
MODELO (11): SARIMA (1,1,0) (0,1,1) [12]	636.8255	645.451

La tabla 2 evidencia los criterios de AIC y BIC, los cuales deben ser los menores posibles para elegir el mejor modelo que realice el pronóstico de la producción de espárragos. En este caso, se toma como mejor modelo el SARIMA (1,1,1) (2,1,0) [12]; sin embargo, deben mostrarse también los errores, ya que ellos, también ayudan a identificar el mejor modelo que permita pronosticar nuestra serie.

OE 5. Validar el modelo de series de tiempo para describir la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018.

Tabla 3

Comparación de los errores de pronósticos de los modelos SARIMA propuestos

Modelo	ME	RMSE	MAE	MPE	MAPE	AIC	BIC
MODELO (1): SARIMA (1,1,1) (0,1,0) [12]	-0.423473	2.535057	1.776996	-1.855281	6.148713	635.7728	644.3984
MODELO (2): SARIMA (1,1,1) (0,1,1) [12]	-0.37766	2.283428	1.630327	-1.747074	5.631423	614.5208	626.0216
MODELO (3): SARIMA (1,1,1) (0,1,2) [12]	-0.397163	2.217492	1.591451	-1.816862	5.496284	611.4517	625.8277
MODELO (4): SARIMA (1,1,1) (1,1,0) [12]	-0.392524	2.417642	1.69434	-1.77543	5.856681	626.7012	638.202
MODELO (5): SARIMA (1,1,1) (2,1,0) [12]	-0.38775	2.203511	1.589956	-1.7234	5.450104	608.4675	622.8435
MODELO (6): SARIMA (1,1,1) (2,1,1) [12]	-0.379569	2.17604	1.587954	-1.67172	5.454996	608.6593	625.9104
MODELO (7): SARIMA (1,1,2) (0,1,0) [12]	-0.405823	2.506366	1.736309	-1.783324	5.95463	634.7211	646.2219
MODELO (8): SARIMA (0,1,2) (0,1,1) [12]	-0.364169	2.291476	1.634073	-1.708825	5.653073	615.3451	626.8459
MODELO (9): SARIMA (0,1,2) (2,1,0) [12]	-0.379923	2.210387	1.596228	-1.701294	5.480892	609.2871	623.6631
MODELO (10): SARIMA (1,1,0) (0,1,0) [12]	-0.121238	2.776052	1.983677	-0.738076	6.746148	655.8679	661.6183
MODELO (11): SARIMA (1,1,0) (0,1,1) [12]	-0.147393	2.522574	1.845098	-0.921553	6.282574	636.8255	645.451

Mediante la tabla 3, se contrasta que el mejor modelo para realizar el pronóstico de la producción de espárragos es el modelo SARIMA (1,1,1) (2,1,0) [12], ya que posee los menores errores. Ante ello, debe realizar diversas pruebas en los residuales del modelo, es decir, debe probarse normalidad, ruido blanco y la linealidad de los residuales.

Figura 4

Función de autocorrelación simple de los residuales del modelo SARIMA (1,1,1) (2,1,0)
[12].

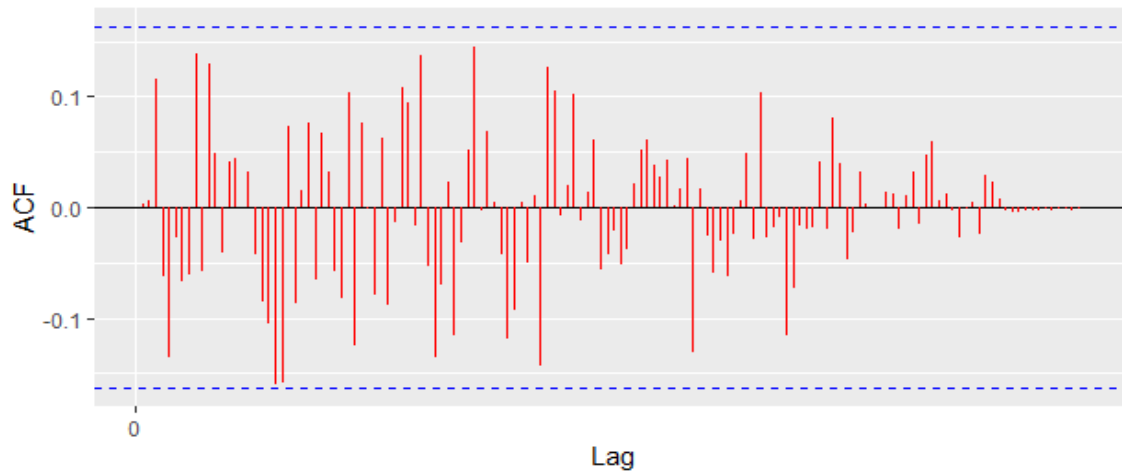
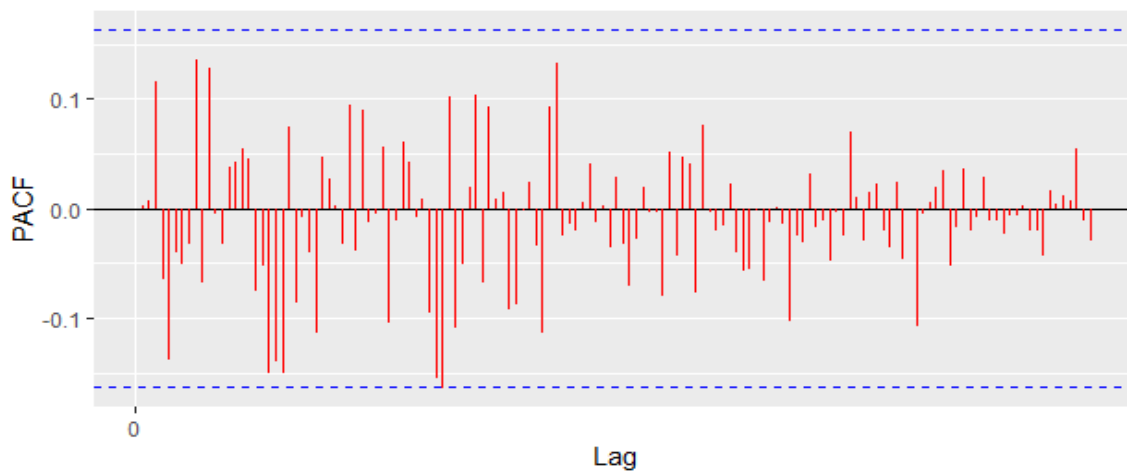


Figura 5

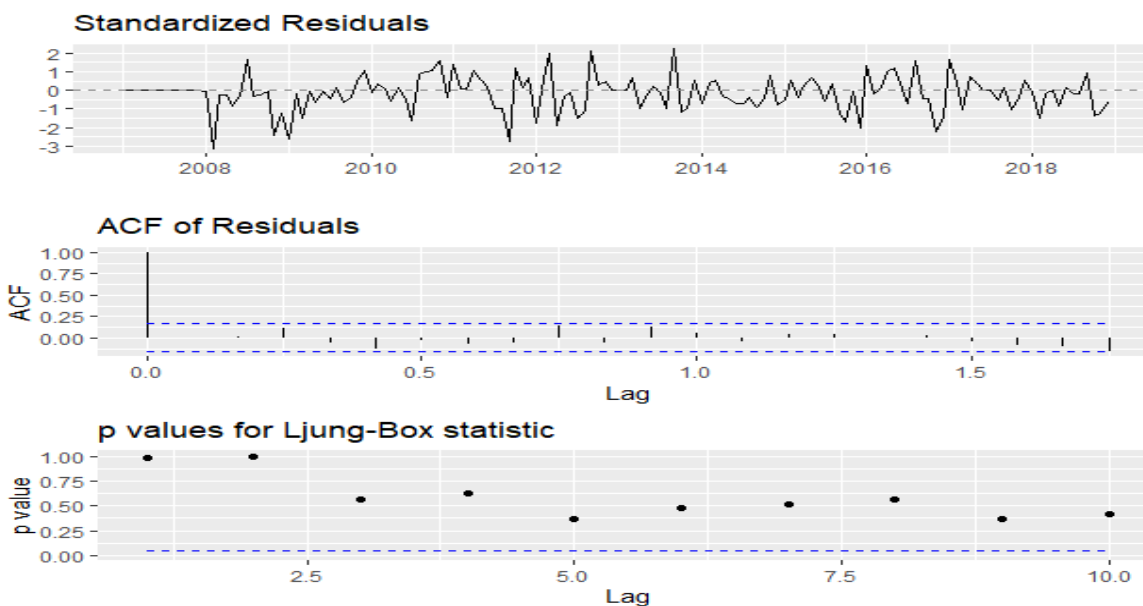
Función de autocorrelación parcial de los residuales del modelo SARIMA (1,1,1) (2,1,0)
[12]



En la figura 8 y 9, se puede observar que no existen rezagos que sobresalen de la banda de confianza, esto demuestra que el modelo es el adecuado para realizar pronósticos.

Figura 10

Residuales del modelo SARIMA (1,1,1) (2,1,0) [12]



La figura 10 evidencia los residuales del modelo SARIMA (1,1,1) (2,1,0) [12], donde se observa que los residuales estandarizados se distribuyen alrededor del promedio; asimismo, observamos que no existen rezagos fuera de la banda de confianza.

Prueba de hipótesis de acuerdo a los TEST

Test de Ljung – Box

Tabla 4

Resultados de ruido blanco, independencia y normalidad de los residuos para el modelo SARIMA (1,1,1) (2,1,0) [12].

SARIMA (1,1,1) (2,1,0)	P-valor	Decisión	Conclusión
Test de Box-Pierce	0.9762	No rechazamos H ₀	Los residuos se distribuyen de forma independiente
Kolmogorov-Smirnov	0.0592	No rechazamos H ₀	Los residuos proceden de una distribución normal

Fuente: Datos obtenidos del Software Rstudio y de la base de espárragos del BCR.

En la tabla 4 se muestran los valores de los test de independencia Box-Pierce y el de normalidad Kolmogorov-Smirnov, quien mediante una confianza del 95% argumentan un p-

valor mayor a 0.05, infiriendo que son independientes y proceden de una distribución normal, siendo aptos para pronosticar correctamente la producción de espárragos.

OE 6. Pronosticar la producción mensual de espárrago en el Perú mediante el modelo obtenido para los años 2019-2020.

Tabla 5

Valores pronosticados para los siguientes meses.

	Point.Forecast	Lo.95	Hi.95
Ene 2019	28.07096	23.54294	32.59897
Feb 2019	26.79656	22.02764	31.56548
Mar 2019	30.06312	25.24111	34.88513
Abr 2019	28.41857	23.56952	33.26761
May 2019	28.34072	23.46999	33.21144
Jun 2019	26.85026	21.95919	31.74133
Jul 2019	22.18601	17.275	27.09702
Ago 2019	27.10416	22.17337	32.03495
Sep 2019	35.39963	30.44915	40.3501
Oct 2019	34.97738	30.00731	39.94745
Nov 2019	35.66009	30.6705	40.64968
Dic 2019	31.77654	26.76751	36.78558
Ene 2020	28.57277	22.62879	34.51675
Feb 2020	27.57749	21.4815	33.67348
Mar 2020	29.50832	23.34975	35.66689
Abr 2020	28.28973	22.08536	34.4941
May 2020	28.48551	22.23942	34.7316
Jun 2020	26.44183	20.15524	32.72841
Jul 2020	22.02771	15.70112	28.3543
Ago 2020	26.49162	20.12535	32.8579
Sep 2020	34.72255	28.31684	41.12825
Oct 2020	35.00106	28.55618	41.44595
Nov 2020	35.9198	29.43597	42.40363
Dic 2020	32.76604	26.24349	39.28858
Ene 2021	28.27172	21.35719	35.18625
Feb 2021	26.49879	19.47509	33.5225
Mar 2021	28.78652	21.69759	35.87545
Abr 2021	27.72829	20.58389	34.8727
May 2021	27.35518	20.15796	34.5524
Jun 2021	25.86055	18.61146	33.10964

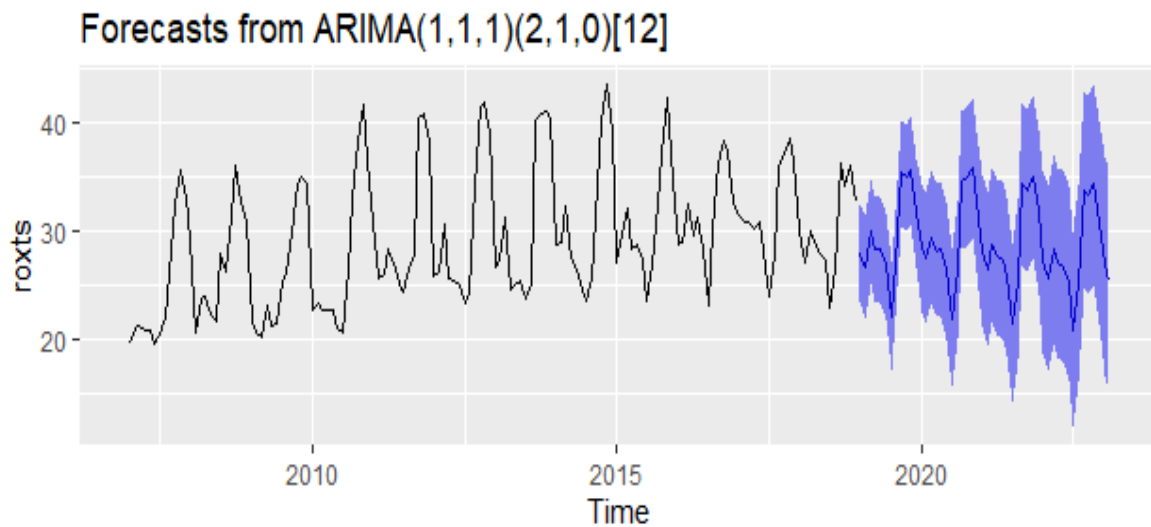
Jul 2021	21.5361	14.23566	28.83654
Ago 2021	25.56319	18.21179	32.91459
Sep 2021	34.42098	27.01899	41.82298
Oct 2021	33.77065	26.31841	41.22289
Nov 2021	35.09811	27.59595	42.60026
Dic 2021	31.86521	24.31347	39.41695

En la tabla 5 se muestran los valores pronosticados para los años 2019 a 2021 según el modelo SARIMA (1,1,1) (2,1,0). En ese aspecto se precisa que los datos pronostican bien a la producción de espárragos en los últimos años.

A continuación, se muestra la figura con los valores pronosticados.

Figura 11

Pronóstico de la producción de espárrago para los años 2019 - 2021.



Cómo se muestra en la figura 11, los datos fueron pronosticados luego de cumplir con los supuestos de independencia y normalidad.

3.2. Discusión de los resultados

Luego de haber obtenido los resultados según los objetivos e hipótesis previamente establecidos, pasan a ser discutidos con la teoría y los trabajos previos. Para ello, se empieza por cumplir con el objetivo general, el cual fue: determinar el modelo de series de tiempo para describir la producción mensual de espárragos en el Perú; donde podemos afirmar que el modelo queda definido por un SARIMA (1,1,1) (2,1,0). Este modelo precisa que la serie tuvo parte regular y estacional. Este modelo mediante la metodología box Jenkins es adecuado para pronosticar la producción de espárragos. Sánchez y Varillas (2014), indican conclusiones similares, quienes afirman que la metodología de Box Jenkins es una técnica estadística muy útil para realizar pronósticos. Angulo (2013), también afirma su preferencia por la utilización de Box Jenkins.

El primer objetivo específico fue describir el comportamiento de la producción mensual de espárragos en el Perú, período enero 2007 – diciembre 2018, el cual mediante la figura 1 la secuencia de la producción de espárragos 2007 – 2018, donde claramente se visualiza un periodo de estacionalidad y un crecimiento lento, es decir, no es constante ni en media ni en varianza. La serie no es estacionaria por motivos que la producción de espárragos tiene su temporada, ante ello, surgen esos picos en la serie, dando indicios que la serie puede ser transformada tanto en la parte estacional, como la parte regular.

Cabe mencionar que para que una serie sea trabajada mediante la metodología de Box Jenkins debe ser constante en media y varianza.

Como segundo objetivo se tuvo que analizar la estacionariedad y estacionalidad de la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018; donde

mediante los estadísticos $ndiff$ y $nsdiff$ se evidenció que la serie tenía parte estacional y no era estacionaria.

Luego de probar varios modelos, el más adecuado según los errores, el AIC y BIC, se llegó a la conclusión que usando un modelo SARIMA (1,1,1) (2,1,0). Cabe mencionar que el AIC obtuvo un valor igual a 608.4675, mientras que el BIC fue igual a 622.8435, valores menores a los demás modelos estudiados. El modelo arrojó un RMSE igual a 2.2035, el MAE fue igual a 1.5899; el MPE igual a -1.7234, y el MAPE fue 5.4501. Este modelo fue elegido gracias a tener menos errores.

El modelo fue validado y pronosticado, pues cumplió su test de independencia y normalidad. Asimismo, permitió pronosticar de forma correcta a la producción de espárragos.

CAPÍTULO IV. CONCLUSIONES

1. El modelo de series de tiempo que describe correctamente a la producción mensual de espárrago en el Perú, periodo enero 2007 – diciembre 2018, fue un SARIMA $(1,1,1) (2,1,0)_{12}$; pues permite pronosticar correctamente la producción de espárragos en el Perú.
2. Al describir el comportamiento inicial de la producción de espárragos en el Perú se observó un comportamiento creciente, con picos cada cierto tiempo, evidenciando estacionalidad cada cierto periodo, infiriendo que la serie no fue estacionaria, es decir, no era constante ni en media ni en varianza.
3. Analizando la estacionariedad y estacionalidad, la serie tuvo que ser transformada en primeras diferencias tanto en la parte regular como estacional para llegar a ser constante en media y en varianza y poder aplicar la metodología de Box Jenkins.
4. Mediante la contrastación de los test de ruido blanco, independencia y normalidad, el mejor modelo que permite describir la producción de espárragos, fue el SARIMA $(1,1,1) (2,1,0)_{12}$, pues fue el que menos errores presentó.
5. Se estimaron los parámetros de series de tiempo de la producción mensual de espárragos en el Perú, periodo enero 2007 – diciembre 2018, siendo también contrastados con el AIC y BIC quienes fueron 608.4675 y 622.8435 respectivamente, los parámetros del modelo SARIMA $(1,1,1) (2,1,0)_{12}$ fue significativo.
6. El modelo fue validado por medio de los test de ruido blanco, independencia y normalidad de los residuales, donde ambos tuvieron un p-valor mayor a 0.05.
7. Se pronosticó la producción de espárragos en el Perú para los años 2019 a 2021, donde se debe tener en cuenta al mes de julio del 2019, por ser la producción más

baja, seguir la proyección del mes de noviembre del 2021.

CAPÍTULO V. RECOMENDACIONES

1. Se recomienda a los futuros investigadores realizar una comparación de la metodología empleada con redes neuronales u otro estadístico que realice este tipo de pronósticos, ello a fin de obtener valores pronosticados más exactos.
2. A las entidades interesadas en la producción de espárragos, considerar lo que se avecina en los siguientes años de producción, a fin de tomar medidas y acciones para poder comprar y vender la cosecha de espárragos.
3. Realizar una investigación más amplia, pues se conoce que mientras mayor cantidad de datos, mayor exactitud de información para el pronóstico.

REFERENCIAS BIBLIOGRAFICAS

Agraria. (2017). Produccion de espárragos en el Perú crecería ligeramente en el presente año.

<http://agraria.pe/noticias/produccion-de-esparragos-en-peru-creceria-ligeramente-14588>

Angulo. (2013). Prediccion del PBI Real Nacional Trimestral : Redes Neuronales

Autorregresivas vs Metodologia Box Jenkins. Trujillo.

<http://dspace.unitru.edu.pe/bitstream/handle/UNITRU/8660/ANGULO%20ELORREAGA%2c%20Luis%20Alfredo.pdf?sequence=1&isAllowed=y>

Cruz, S. (2014). Pronostico para la exportación de espárrago de la Region La Libertad enero

2010 - agosto 2014. La Libertad.

<https://dspace.unitru.edu.pe/bitstream/handle/UNITRU/8338/CRUZ%20IRIGOIN%2c%20Sherlly%20Wuendy.pdf?sequence=1&isAllowed=y>

De la Cruz, y Aguirre. (2015). Analisis comparativo de redes NeuronalesArtificiales y la

metodologia de box Jenkins en el pronóstico del índice de precios al consumidor.

Huaraz - Perú.

<http://repositorio.unasam.edu.pe/bitstream/handle/UNASAM/1135/T05-EI%200052%202015.pdf?sequence=1&isAllowed=y>

Fresh Plaza. (2018). Produccion mundial de espárragos.

<https://www.freshplaza.es/article/3116004/produccion-mundial-de-esparragos/>

Gil, M., & Guevara, G. (2016). Modelo óptimo de serie de tiempo para pronostico del credito

de consumo total mensual otorgado por la banca multiple del Perú 2001-2014.

Lambayeque : Universidad Nacional Pedro Ruiz Gallo .

Mota, A. (2016). Pronostico del precio del crudo de exportación mexicano con la Metodología de Box-Jenkins para series de tiempo. Puebla.
<https://repositorioinstitucional.buap.mx/handle/20.500.12371/12812>

Ollé, J. (2017). Descubre qué es una serie temporal y cómo interpretarla. Conceptos Claros:
<https://conceptosclaros.com/serie-temporal/>

Penn State. (2017). Alternativas agriculas. producción de espárrago . Penn State extensión.
<https://extension.psu.edu/produccion-de-esparragos>

Polis, G. (18 de Diciembre de 2002). ¿Que es un pronóstico? Características y metodos. GestioPolis.com:
<https://www.gestiopolis.com/que-es-un-pronostico-caracteristicas-y-metodos/>

Ríos, G. (2008). Series de Tiempo. Santiago. https://www.ucursos.cl/ingenieria/2010/1/CC52A/1/material_docente/bajar?id_material=296003

Ruiz, D. (2004). Manual de estadísticas. España: eumed.net.
<https://www.eumed.net/cursecon/libreria/drm/24.pdf>

Sampieri, Batista, & Fernandez. (2014). Metodología de la Investigación. Mexico: Mc Graw Hill.
<http://observatorio.epacartagena.gov.co/wp-content/uploads/2017/08/metodologia-de-la-investigacion-sexta-edicion.compressed.pdf>

Sanchez, & Varillas, R. (2014). Modelos de serie de tiempo mediante la metodologia de Box

y Jenkins para el pronóstico de la llegada de visitantes nacionales y extranjeros a la ciudad sagrada de Caral para los meses de enero a diciembre del año 2014. Huacho - Peru. <http://repositorio.unjfsc.edu.pe/handle/UNJFSC/59>

Vásquez, F. (2017). Modelo de Box y Jenkins y Redes para pronosticar el precio del dólar del sistema bancario en moneda nacional año 2015 . Lambayeque: Universidad Nacional Pedro Ruiz Gallo .
<https://repositorio.unprg.edu.pe/bitstream/handle/20.500.12893/1860/BC-TES-TMP-695.pdf?sequence=1&isAllowed=y>

Villalobos. (2014). Pronóstico de Ingresos Tributarios recaudados de Impuestos generales a las ventas a nivel nacional, por la SUNAT ,mediante la Metodología de Box y Jenkins y Redes Neuronales Artificiales para el año 2014 . Lambayeque - Perú.
<https://repositorio.unprg.edu.pe/bitstream/handle/20.500.12893/459/BC-TES-4208.pdf?sequence=1&isAllowed=y>

Villarreal. (2016). Introducción a los modelos pronósticos. Argentina.
https://www.matematica.uns.edu.ar/uma2016/material/Introduccion_a_los_Modelos_de_Pronosticos.pdf

ANEXOS

Tabla 6

Producción agropecuaria (miles de toneladas) - Agrícola – Espárrago

MES/AÑO	2007	2008	2009	2010	2011	2012
ENERO	19.6236	29.2222	21.76946	22.6895	30.80188	25.95005
FEBRERO	21.2164	21.785	18.66679	23.36415	25.7557	26.26332
MARZO	21.24845	23.7307	20.3627	22.8577	31.2116	30.76402
ABRIL	20.9382	23.94173	23.1621	22.78428	35.41493	25.61188
MAYO	18.985	22.40165	21.2424	22.72924	30.20371	25.5231
JUNIO	14.7205	21.7113	21.4766	22.14822	20.26131	25.16585
JULIO	15.4742	27.9119	25.45096	22.7204	24.5503	22.34019
AGOSTO	22.1696	26.3187	26.0583	26.54515	26.52469	24.55824
SETIEMBRE	28.2739	31.5624	29.82792	32.98895	31.02236	34.19388
OCTUBRE	32.5516	36.07428	36.4068	38.2884	47.48895	47.4954
NOVIEMBRE	35.72357	32.86868	34.991	41.6345	47.93373	49.00122
DICIEMBRE	33.1783	30.8451	34.46467	36.45878	41.13699	39.11056

MES/AÑO	2013	2014	2015	2016	2017	2018
ENERO	26.68537	28.75743	27.16615	28.90982	31.6252	30.04991
FEBRERO	27.26218	28.95823	29.70839	29.06673	30.9451	27.15195
MARZO	31.32231	32.38205	32.09619	32.49926	30.94757	30.0181
ABRIL	24.7164	27.67454	28.37866	29.66912	30.37816	29.167
MAYO	24.991	26.61249	28.88046	31.26517	30.90708	27.9673
JUNIO	25.59317	25.3794	27.28442	28.50707	28.11482	27.2853
JULIO	20.91962	20.71009	23.65927	23.20086	24.08134	23.07605
AGOSTO	25.03442	25.76205	26.89299	30.26514	27.8198	26.5026
SETIEMBRE	40.22727	34.31749	32.50303	36.39455	36.03365	36.25905
OCTUBRE	47.8185	42.27255	37.09842	38.40936	37.23125	34.24605
NOVIEMBRE	50.18854	45.60925	42.36765	37.32491	38.54964	36.14513
DICIEMBRE	39.44495	39.2654	34.74742	32.7938	36.4639	32.7617

ANEXO 2

CÓDIGOS EN RSTUDIO

```
View(ROX_2020)
attach(ROX_2020)
```

```
ROX_2020
```

```
roxts<-ts(ROX_2020$P_Espárrago, start = c(2007,1), frequency = 12)
```

```
print(roxts)
```

```
library(forecast)
```

```
library(tseries)
```

```
autoplot(roxts, colour = "red", linetype = "dashed")
```

```
autoplot(acf(roxts, plot = FALSE, lag.max=144), colour = "red", )
```

```
autoplot(pacf(roxts, plot = FALSE, lag.max =144), colour="red", )
```

```
autoplot(stl(roxts, s.window = "periodic"), ts.colour = "blue")
```

```
ndiffs(roxts)
```

```
nsdiffs(roxts)
```

```
diff.roxts<-autoplot(diff(roxts), linetype = "dashed", colour = "darkmagenta")
```

```
diff.roxts
```

```
diff.roxts.12<-diff(roxts, lag = 12)
```

```
autoplot(diff.roxts.12, colour = "darkorange4", linetype = "dashed")
```

```
adf<-adf.test(diff.roxts.12)
```

```
adf$p.value
```

```
kpss<-kpss.test(diff.roxts.12)
```

```
kpss$p.value
```

```
autoplot(acf(diff.roxts.12, lag.max =144), plot = FALSE, colour="red", conf.int.colour = "blue")
```

```
autoplot(pacf(diff.roxts.12, lag.max =144), plot = FALSE, colour="red", conf.int.colour = "blue")
```

```
auto.arima(roxts, stepwise = FALSE, approximation = FALSE)
```

```
arima1<-arima(roxts, order = c(1,1,1), seasonal = list(order=c(2,1,0),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,1), seasonal = list(order=c(0,1,0),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,1), seasonal = list(order=c(0,1,1),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,1), seasonal = list(order=c(0,1,2),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,1), seasonal = list(order=c(1,1,0),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,1), seasonal = list(order=c(2,1,0),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,1), seasonal = list(order=c(2,1,1),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,2), seasonal = list(order=c(0,1,0),period=12))
```

```
arima1<-arima(roxts, order = c(0,1,2), seasonal = list(order=c(0,1,1),period=12))
```

```
arima1<-arima(roxts, order = c(0,1,2), seasonal = list(order=c(2,1,0),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,0), seasonal = list(order=c(0,1,0),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,0), seasonal = list(order=c(0,1,1),period=12))
```

```
arima1<-arima(roxts, order = c(1,1,1), seasonal = list(order=c(2,1,1),period=12))
```

```
AIC(arima1)
```

```
BIC(arima1)
```

```
coefest(arima1)
```

```

summary(arima1)
autoplot(acf(arima1$residuals, plot = FALSE, lag.max = 144), colour = "red", lag = 12)
autoplot(pacf(arima1$residuals, plot = FALSE, lag.max = 144), colour = "red", lag = 12)
ggtsdiag(arima1)
bp <- Box.test(arima1$residuals) # Test de Box-Pierce
bp$p.value

lb <- Box.test(arima1$residuals, type="Ljung-Box") # Test de Ljung-Box
lb$p.value

jb <- jarque.bera.test(arima1$residuals) # Test de Jarque-Bera
jb$p.value

sht<-shapiro.test(arima1$residuals) # Test de Shapiro-Wilk
sht$p.value

forecast1<-forecast(arima1, level = c(95), h = 50)
autoplot(forecast1)

hist(arima1$residuals, freq=FALSE, col="deepskyblue4", main = "Histograma de los
residuales
del SARIMA (1,1,1) (2,1,0)", ylim=c(0,3), xlim = c(-1,1), xlab="")

curve(dnorm(x, mean=mean(arima1$residuals), sd=sd(arima1$residuals)),
from=2, to=2, add=TRUE, col="red2", lwd=2)

qqnormPlot(arima1$residuals)

ggplot(arima1$residuals, aes(x=arima1$residuals))+geom_density(fill="blue")

```



ACTA DE SUSTENTACIÓN VIRTUAL N.º 014-2022-D/FACFyM

Siendo las 4.00 pm del miércoles 20 de abril del 2022, se reunieron vía plataforma virtual, meet.google.com/gvo-exoi-xmt los miembros del jurado evaluador de la Tesis titulada:

“Modelo de series de tiempo para describir la producción mensual de espárrago en el Perú, periodo enero 2007 – diciembre 2018”.

Designados por Resolución N° 357-2022-VIRTUAL-D/FACFyM de fecha 11 de abril de 2022

Con la finalidad de evaluar y calificar la sustentación de la tesis antes mencionada, conformada por los siguientes docentes:

Lic. José Lorenzo Núñez Rodríguez	presidente
Dr. José Luis Rodas Cabanillas	Secretario
M.Sc. Luis Enrique Tuñoque Gutiérrez	Vocal

La tesis fue asesorada por el Dr. Wilver Omero Rodríguez López, nombrado por Resolución N° 310-2019-D/FACFyM de fecha 8 de marzo de 2019.

El Acto de Sustentación fue autorizado por Resolución N° 014-2022-D/FACFyM de fecha 20 de abril del 2022. La Tesis fue presentada y sustentada por los Bachilleres: Acosta Bravo Roxana y Chapoñan Bancas Deyci Yris, y tuvo una duración de 30 minutos. Después de la sustentación, y absueltas las preguntas y observaciones de los miembros del jurado se procedió a la calificación respectiva, otorgándole el Calificativo de 17 (DIECISIETE) en la escala vigesimal, mención Bueno.

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

Siendo las 16:47 pm se dio por concluido el presente acto académico, dándose conformidad al presente acto con la firma de los miembros del jurado.

Lic. José Lorenzo Núñez Rodríguez
Presidente

Dr. José Luis Rodas Cabanillas
Secretario

M.Sc. Luis Enrique Tuñoque Gutiérrez
Vocal

Dr. Wilver Omero Rodríguez López
Asesor





CONSTANCIA DE APROBACIÓN DE ORIGINALIDAD DE TESIS

Yo, Wilver Omero Rodríguez López, Asesor de la Tesis, Revisor del trabajo de Investigación, de las estudiantes, Acosta Bravo Roxana Estela y Chapoñan Bances Deyci Yris

Tituladas:

“Modelo de Series de Tiempo para describir la producción mensual de esparrago en el Perú, periodo enero 2007 – diciembre 2018”.

Luego de la revisión exhaustiva del documento constato que la misma tiene un índice de similitud de 17 % verificable en el reporte de similitud del Programa Turnitin.

El suscrito analizó dicho reporte y concluyó que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad Nacional Pedro Ruiz Gallo.

Lambayeque, 29 de setiembre del 2021

Wilver Omero Rodríguez López
DNI: 17636159.
ASESOR

INFORME DE TESIS

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	repositorio.unprg.edu.pe Fuente de Internet	4%
2	rstudio-pubs-static.s3.amazonaws.com Fuente de Internet	1%
3	agraria.pe Fuente de Internet	1%
4	hdl.handle.net Fuente de Internet	1%
5	www.slideshare.net Fuente de Internet	1%
6	www.murata.eb.waseda.ac.jp Fuente de Internet	1%
7	repositorio.lamolina.edu.pe Fuente de Internet	1%
8	dspace.unitru.edu.pe Fuente de Internet	1%
9	repositorio.unh.edu.pe Fuente de Internet	1%

10	1library.co Fuente de Internet	1 %
11	repositorio.unp.edu.pe Fuente de Internet	1 %
12	repository.libertadores.edu.co Fuente de Internet	1 %
13	www.freshplaza.es Fuente de Internet	<1 %
14	red.uao.edu.co Fuente de Internet	<1 %
15	repositorio.ucv.edu.pe Fuente de Internet	<1 %
16	repositorio.unheval.edu.pe Fuente de Internet	<1 %
17	Diana Irene Chavasse, Rafael Santiago Seoane. "Asociación determinístico-estocástica para predicción de caudales", Ingeniería del agua, 1997 Publicación	<1 %
18	Repository.Libertadores.Edu.Co Fuente de Internet	<1 %
19	extension.psu.edu Fuente de Internet	<1 %
20	repositorio.unsm.edu.pe Fuente de Internet	<1 %

21	Ugochinyere Ihuoma Nwosu, Chukwudi Paul Obite. "Modeling Ivory Coast COVID-19 cases: Identification of a high-performance model for utilization", Results in Physics, 2021 Publicación	<1 %
22	www.scribd.com Fuente de Internet	<1 %
23	upcommons.upc.edu Fuente de Internet	<1 %
24	repositorio-aberto.up.pt Fuente de Internet	<1 %
25	repositorio.unsaac.edu.pe Fuente de Internet	<1 %
26	www.cesma.usb.ve Fuente de Internet	<1 %
27	ecampus.fca.unam.mx Fuente de Internet	<1 %
28	repositorio.ucss.edu.pe Fuente de Internet	<1 %
29	Rey Francisco Ayala Castrejon, Christian Bucio Pacheco. "Modelo ARIMA aplicado al tipo de cambio peso-dólar en el periodo 2016-2017 mediante ventanas temporales deslizantes", Revista Mexicana de Economía y Finanzas, 2020 Publicación	<1 %
<hr/>		
30	repositorio.uss.edu.pe Fuente de Internet	<1 %
31	html.rhhz.net Fuente de Internet	<1 %

Excluir citas

Activo

Excluir coincidencias < 15 words

Excluir bibliografía

Activo



Recibo digital

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

La primera página de tus entregas se muestra abajo.

Autor de la entrega:	Roxana Y Deyci Acosta Bravo_chapoñan Bances
Título del ejercicio:	INFORME DE TESIS ACOSTA-CHAPOÑAN
Título de la entrega:	INFORME DE TESIS
Nombre del archivo:	INFORME_DE_TESIS_-corr1_ACOSTA_Y_CHAPO_AN.pdf
Tamaño del archivo:	844.34K
Total páginas:	71
Total de palabras:	12,982
Total de caracteres:	68,790
Fecha de entrega:	24-feb.-2022 03:30a. m. (UTC-0500)
Identificador de la entre...	1769802610

