



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

MODELOS ESTADÍSTICOS PARA EL PRONÓSTICO DE LOS PARÁMETROS FÍSICOQUÍMICOS DE HARINA DE ORIGEN ANIMAL DE UNA EMPRESA, USANDO EL SOFTWARE R

STATISTICAL MODELS FOR PHYSICOCHEMICAL PARAMETERS PREDICTION OF MEAL OF ANIMAL ORIGIN OF A COMPANY, USING THE R SOFTWARE.

Eduin Xavier Ruiz Monterroza.

Especialista en Estadística Aplicada Fundación Universitaria Los Libertadores.

Químico Puro.

Bogotá, Colombia.

Contacto exruizm@libertadores.edu.co

RESUMEN

Objetivo. Realizar pronósticos de los parámetros bromatológicos (porcentaje de proteína, ceniza, grasa y humedad de harina de origen animal). **Materiales y métodos.** Se utiliza como unidad de análisis una base de datos de harina de origen animal desde el año 2017 hasta el 2021, esta matriz cuenta con los resultados de los parámetros bromatológicos. Se considera que a través de la estadística de series de tiempo se puede ajustar un modelo a cada parámetro estudiado logrando pronosticarlo utilizando la metodología planteada por Box y Jenkins. **Resultados.** Se logra ajustar el modelo Arima (1,1,3) para el porcentaje de proteína, modelo Arima (1,1,2) para el porcentaje de ceniza, modelo Arima (0,1,2) para el porcentaje de grasa y modelo Arima (1,1,9) para el porcentaje de humedad. **Conclusión.** En los primeros cuatro (4) meses del 2022, se obtuvieron resultados para los modelos, Proteína= 0,69%; Ceniza= 1,37%; Grasa= 1,38% y Humedad= 7,39%.

Palabras clave: *Bromatológicos, Harina, Modelo Arima, Series de tiempo.*



ABSTRACT

Objective. To predict the bromatological parameters (protein, ash, fat and moisture percentage) from animal source meal. **Methodology.** Data base from 2017 to 2021 of animal source meal is used as analysis unit, which includes bromatological parameters results for this period. Through time series statistics a model can be adjusted to a specific parameter managing to predict it using the Box and Jenkins methodology. **Results.** An Arima Model was adjusted for each percentage as follows: protein (1,1,3), ash (1,1,2), fat (0,1,2), and moisture (1,1,9). **Conclusion.** The MAPE, is an indicator for an easy interpretation of the results, the lower the value, the better the accuracy of the prediction of the models. In the first four (4) months, results were obtained for the models, Protein= 0.69%; Ash= 1.37%; Fat= 1.38% and Moisture= 7.39%.

Keywords: *Bromatological, Meals, Arima Model, Time Series.*

INTRODUCCIÓN

La industria Rendering recicla subproductos de origen animal no aptos para el consumo humano, transformándolos en materia prima para la industria de alimentos balanceados para animales, cumpliendo dos funciones: Primero, Proteger el medio ambiente, evitando que estos subproductos sean dispuestos en rellenos o en los cuerpos de aguas. Segundo, suministrar energía y proteína a los alimentos balanceados para animales.

El proceso de recolección, pasan por las plantas de sacrificio, y en los puntos de distribución de carnes, los subproductos del sacrificio de animales, como son las plumas, las vísceras, la sangre y los huesos de especies avícolas, porcinas y bovinas. En las plantas se procesan estos materiales con tecnología de punta y de manera responsable con el medio ambiente. Los subproductos sufren un proceso térmico por encima de 130 grados centígrados, por espacio de dos horas o más y ninguna salmonella o entero bacterias sobreviven a esto (Blake J.P, 1999). Además, sufren cambios en el análisis proximal.

Para poder aplicar los métodos estadísticos de manera rápida y confiable se utilizan



programas estadísticos como Excel, SPSS, Minitab, Statgraphics, Statistica, entre otros, que permiten calcular la mayoría de metodologías estadísticas habituales, entre las que se cuentan: análisis exploratorio de datos, gráficos estadísticos, control de calidad, estadística no paramétrica, regresión y sus variantes, análisis multivariado de datos, etc. (Meier & Zund, 2000). Sin embargo, estos programas deben ser adquiridos a través de compra de licencias para poder acceder a todas sus aplicaciones. Por esta razón se busca optar por programas estadísticos gratuitos como es el caso del software estadístico R que puede accederse libremente, no tiene limitadas sus funciones, al contrario de lo que sucede con las herramientas estadísticas comerciales. R utiliza un entorno de programación formado por un conjunto de herramientas muy flexibles que pueden ampliarse fácilmente mediante paquetes, librerías o definiendo las propias funciones. Es una herramienta muy poderosa para todo tipo de procesamiento y manipulación de datos. Permite crear gráficos de alta calidad y resolver problemas estadísticos de gran complejidad que no puedes afrontar con otras herramientas estadísticas comerciales (Gardener, 2012).

La estadística ha desarrollado teorías y métodos que se han convertido en herramientas imprescindibles para la estimación, comparación y la predicción. Entre estos métodos estadísticos se encuentran la regresión, los métodos multivariados y los aplicados a las series temporales entre otros. Las series temporales se aplican, fundamentalmente, los métodos de pronósticos y, su mayor dificultad radica generalmente en la escasez de datos para analizar el problema y la falta de comunicación entre los estadísticos y los usuarios que emplean los resultados de estos métodos (Suárez, 2013).

Uno de los sectores que emplea los métodos estadísticos de las series temporales es la rama agropecuaria, en la cual se realizan pronósticos de condiciones agroclimáticas y de la situación comercial de las empresas, teniendo en cuenta a su vez el análisis de mercado. En el sector económico y financiero, los datos de las series temporales pueden ser muy variados y usualmente son usados para evaluar el comportamiento de las ventas de una empresa, o



para evaluar el comportamiento de los índices de precio de un país o de un tipo de producto, pero en general pueden utilizarse en cualquier área.

Se considera que a través del análisis de series de tiempo se logrará ajustar un modelo que pronostique el comportamiento de las variables bromatológicas: porcentaje de humedad, proteína, ceniza, grasa y así lograr tener un soporte en la toma de decisiones dentro de la planta.

MARCO TEÓRICO

Harinas de origen animal.

Esta empresa se dedica a la fabricación de harinas de origen animal y sebo industrial destinado a la producción de alimentos para animales. Los productos que actualmente ofrecen al mercado son: Harina de carne y hueso, Harina de víscera de pollo, Harina de pluma hidrolizada, Harina de sangre, Sebo industrial y Aceite de pollo.

Las harinas cuentan con registro ICA para la legalidad de la producción y comercialización. La empresa transforma subproductos de plantas de sacrificios de aves, bovinos y porcinos en harinas para alimentación animal las cuales son suministradas a las empresas productoras de concentrados para animales a nivel nacional. De esta forma se minimiza el impacto ambiental que estos subproductos generan si no se disponen adecuadamente.

Estas harinas a nivel nacional son una alternativa importante para aumentar el rendimiento nutricional en la alimentación animal, en especial por los altos contenidos proteicos presentes en algunas de ellas. En Colombia y otros países, la recuperación de estos desechos de matadero y su transformación industrial se han constituido en una fuente confiable de suministro de materias primas utilizadas en la elaboración de alimentos balanceados para animales; y a nivel municipal este tipo de industria genera grandes oportunidades laborales.

Por lo tanto, es de suma importancia el conocimiento de las características físico-químicas de los desechos de matadero es fundamental para su utilización en la alimentación animal.

Cuenta con la infraestructura para realizar los distintos análisis exigidos en las normas NTC 685 como son análisis bromatológicos y físico químicos, con respecto a las harinas de los diferentes productos se determina; humedad, proteína, grasa, acidez, ceniza, índice de peróxidos, digestibilidad y granulometría y con respecto al sebo se realiza humedad, acidez, índice de peróxidos, estos análisis se realizan a partir de una muestra representativa del lote y posteriormente se informa al área de producción si cada producto cumple con las fichas técnicas establecidas y es apto para despacho. Estos análisis son realizados bajo la norma NTC 685, y cada uno de estos ensayos se efectúa de acuerdo a las siguientes normas que indica la tabla 1.

Ensayos	Norma
Alimentos para animales. Determinación del contenido de humedad y materia volátil	NTC 4888
Alimentos para animales. Determinación de contenido de nitrógeno y cálculo del contenido de proteína cruda, método Kjeldahl	NTC 4657
Alimentos para animales. Determinación de ceniza cruda	NTC 4648
Alimentos para animales. Determinación del contenido de grasa	NTC 4969

Tabla 1: Normas NTC de análisis para de harinas.

Análisis proximal.

El análisis proximal es el método más usado para expresar la calidad nutritiva global de un alimento, mide la cantidad de nutrientes presentes, divididos en seis grupos: contenido de humedad, proteína bruta, fibra cruda, cenizas, extracto etéreo y los elementos libres de nitrógeno que constituyen una medida indirecta del contenido total de carbohidratos (Barrera, Tapia y Monteros, 2004).

Proteína.

Las proteínas son biomoléculas formadas básicamente por carbono, hidrógeno, oxígeno y nitrógeno, pueden además contener azufre y en algunos tipos de proteínas, fósforo, hierro, magnesio y cobre entre otros elementos (Bedoya, 2016).



Ceniza.

La materia orgánica en un alimento puede ser dividida en materia orgánica y inorgánica. Compuestos que contienen carbón (C), hidrógeno (H), oxígeno (O) y nitrógeno (N) son clasificados como orgánicos, los compuestos inorgánicos o minerales son los demás elementos químicos (calcio, fósforo etc.); cuando una muestra de alimento es colocada en un horno y mantenida a 550°C por 24 horas la materia orgánica esta quemada y la materia restante es la parte mineral, llamada ceniza (Wattiaux, Homan, Moreno y Rodríguez, 1998).

Grasa.

El contenido de grasa, también llamado extracto etéreo, puede estar formado por lípidos enlazado y lípidos libres; estos últimos básicamente consisten en grasas neutras (triglicéridos) y ácidos grasos libres los cuales pueden ser extraídos por disolventes como éter etílico o fracciones ligeras del petróleo, en cambio, los lípidos enlazados requieren de una hidrolisis con disolventes más polares para su extracción (Ortiz, 2006).

Humedad.

Los alimentos en su estado natural están compuestos por materia seca y agua en cantidades específicas; la humedad de un producto se expresa en forma porcentual en base húmeda o en base seca como la cantidad de agua contenida en el producto húmedo o seco (Puerta, 2006).

Series de Tiempo.

Los pronósticos son una herramienta que proporciona un estimado cuantitativo de la probabilidad de eventos futuros (Juárez, Zuñiga, Flores, y Partida, 2016), una de las técnicas para hacerlo son las series de tiempo, utilizadas en una amplia gama de disciplinas: finanzas, transporte, mercados, biología, medicina, ingeniería, etc., para analizar el comportamiento de una variable en el tiempo y facilitar la toma de decisiones (Moreno-García, Jimenez-Linares, y Rodriguez-Benitez, 2017). Estas toman como base datos históricos para obtener una caracterización estadística de los enlaces existentes en el pasado



y en el presente (Diebold, 2001). En las últimas dos décadas, el desafío de predecir el futuro mirando el pasado ha conducido a la creación de una variedad de algoritmos de predicción de series de tiempo (Egrioglu, E., Yolcu, U. y Aladag, C. 2015; Hyndman, R.J., Khandakar, Y 2008). Estos permiten considerar información temporal, estimar parámetros y proporcionar datos puntuales precisos sobre valores futuros (Lütkepohl, H. 2005; Pravilovic, Bilancia, Appice, y Malerba, 2017). Una serie de tiempo es una sucesión de variables aleatorias ordenadas de acuerdo a una unidad de tiempo, Y_1, \dots, Y_T . Estas son útiles para obtener pronósticos de valores futuros de la serie, con el fin de ayudar a tomar decisiones que tienen consecuencias importantes o para entender mejor el mecanismo de generación de los datos, que puede no ser claro inicialmente en una investigación (Giraldo, 2011), también son utilizadas para analizar el comportamiento dinámico de sistemas complejos (Zunino, Olivares, Bariviera, y Rosso, 2017).

Las series de tiempo se clasifican en estacionarias y no estacionarias. Las primeras se caracterizan por ser estables a lo largo del tiempo. En las series de tiempo no estacionarias la tendencia puede cambiar a lo largo del tiempo (Giraldo 2011). Los principales componentes de las series de tiempo: tendencia, estacionalidad y heterocedasticidad, indican el comportamiento periódico y a largo plazo de la serie. Dada una serie T_t , $t = 1, \dots, T$, el Modelo Aditivo Y_t consiste en asumir que se puede descomponer en tres componentes (Quintana y Jiménez, 2016). Ver Ecuación (1):

$$Y_t = T_t + S_t + \epsilon_t \quad (1)$$

Donde T_t es la tendencia, S_t es la componente estacional y ϵ_t es la componente de errores. Las componentes T_t y S_t son funciones de t determinísticas y su evolución es perfectamente predecible (Giraldo, 2011). Una serie de tiempo tiene componente estacional de período S , si presenta patrones que se repiten con una frecuencia constante (S), produciendo en ella un efecto periódico (Diebold, 2001).



Metodología Box-Jenkins

Desarrollar un modelo basado en series de tiempo no es algo para nada fácil, sin embargo, Box y Jenkins lograron crear una metodología bastante practica para estimar los modelos de series de tiempo, dicha metodología es la siguiente:

- 1.** Por medio de transformaciones y/o diferencias se estabiliza la varianza, y se eliminan la tendencia y la estacionalidad de la serie; obteniéndose así, una serie estacionaria.
- 2.** Para la serie estacionaria obtenida se identifica y se estima un modelo que explica la estructura de correlación de la serie con el tiempo.
- 3.** Al modelo hallado en 2.) Se aplican transformaciones inversas que permitan establecer la variabilidad, la tendencia y la estacionalidad de la serie original.
- 4.** El modelo estimado se valida a través de la correlación de sus residuales, si estos llegan a presentar algún tipo de correlación entonces se vuelve a estimar nuevos parámetros, es decir se regresa al punto 2.) Para realizar una nueva iteración. El modelo definitivo se usa para pronosticar (Box y Jenkins, 1976; Guerrero, 2003).



MATERIALES Y MÉTODOS.

Descripción de datos

Se utiliza como unidad de análisis una base de datos de harina de origen animal desde el año 2017 hasta el 2021, esta matriz cuenta con los resultados de los parámetros bromatológicos (proteína, ceniza, grasa y humedad).

Preparación de datos

Los datos fueron facilitados por la empresa por el sistema de información de análisis de laboratorio SIAL, descargamos y promediamos mensualmente por variables de estudio para su posterior estudio de serie de tiempo.

Exploración de datos

Se realiza análisis estadístico y desarrollo de gráficos con ayuda del Software RStudio, es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo.

Construcción de modelos

- La primera fase consiste en identificar el posible modelo ARIMA que sigue la serie, lo que requiere:
- Decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria.
- Determinar un modelo ARMA para la serie estacionaria, es decir, las órdenes p y q de su estructura autorregresiva y de media móvil.
- La segunda fase: estimación, En esta etapa se estiman los coeficientes de los términos autorregresivos y de medios móviles incluidos en el modelo, cuyo número de rezagos y ya han sido identificados en la etapa anterior.
- La tercera fase es el diagnóstico, donde se comprueba que los residuos no tienen



estructura de dependencia y siguen un proceso de ruido blanco.

- La cuarta fase es la predicción, una vez que se ha obtenido un modelo adecuado se realizan predicciones con el mismo.

RESULTADOS.

Proteína

Análisis Descriptivo

Los resultados de proteína obtenidos de los años estudiados se encuentran dentro de los rangos estipulados para este tipo de harinas, como se observa en la tabla 2.

Mínimo	46,79%
Primer cuartil	47,98%
Mediana	48,82%
Media	48,87%
Tercer cuartil	49,66%
Máximo	51,89%

Tabla 2: Resumen datos descriptivos Proteína.

En la figura 1 se observa que el histograma se comporta como una distribución normal y que los datos la mayoría se encuentran en la parte central entre 47% y 50%.

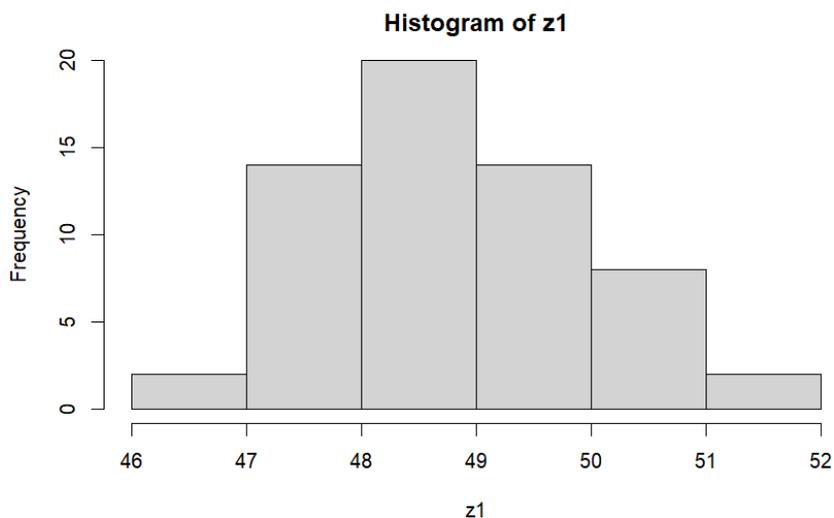


Figura 1: Histograma Proteína

Decomposition of additive time series

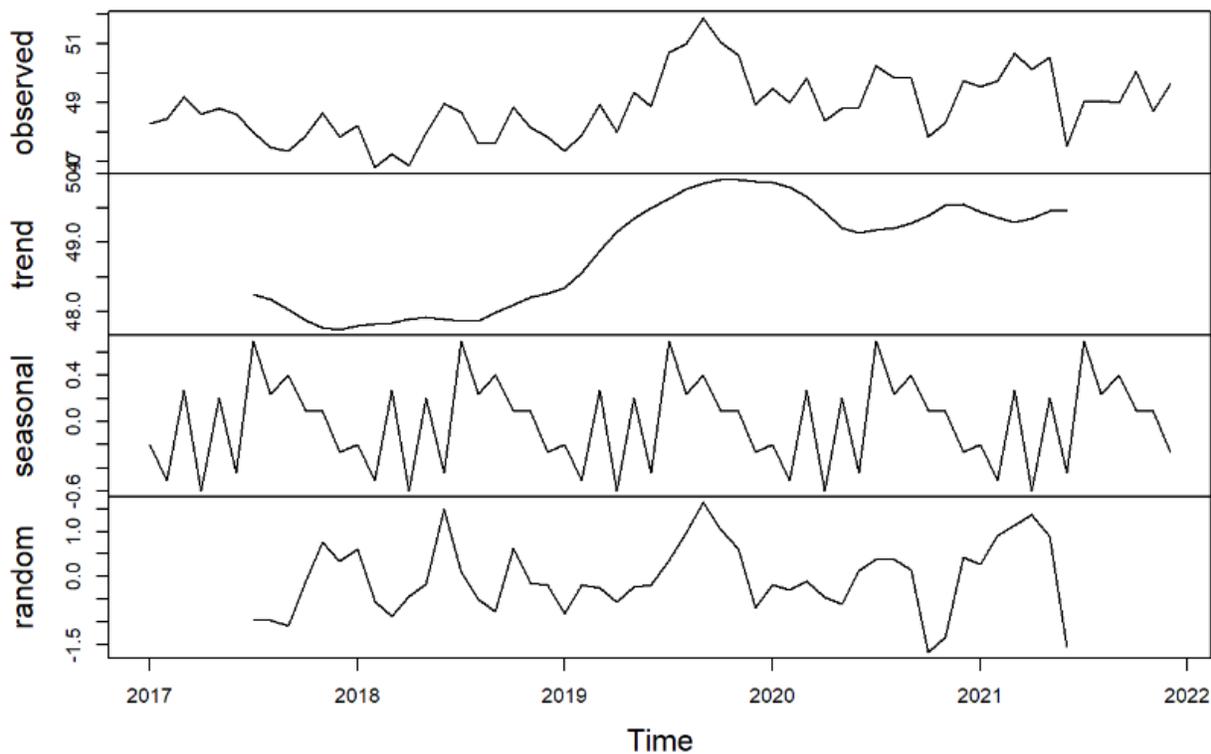


Figura 2: Componentes de la serie de tiempo para Proteína.

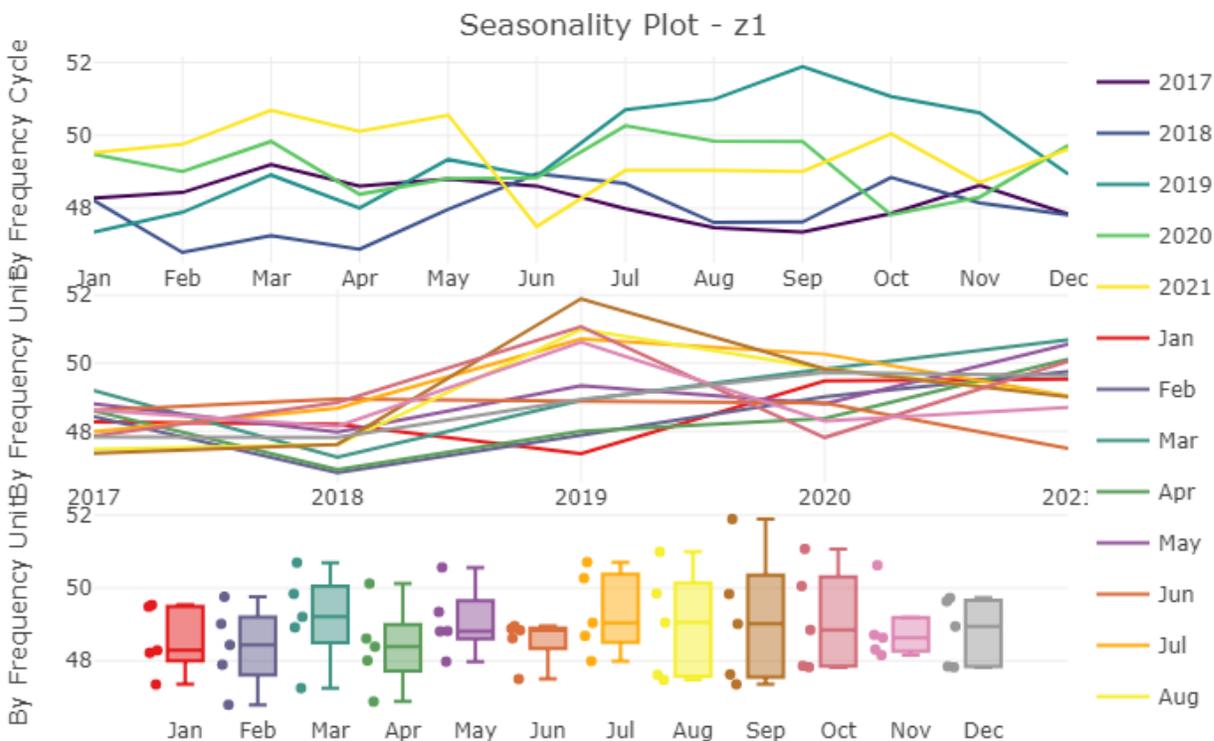


Figura 3: Comportamiento del periodo comprendido entre 2017 al 2021 de Proteína.

En las figuras 2 y 3 se observa que el comportamiento histórico de proteína, desde los años 2017 al 2019 se tuvo un promedio del 48%, en los años siguiente 2020 y 2021 el promedio fue de 49%, notándose su tendencia a incrementar. Los meses de mayor incremento de la variable es en marzo, julio, agosto y septiembre.

Construcción de modelos

Como se observó en los gráficos descriptivos la tendencia a incrementar para la variable. Se realiza la prueba de Dickey Fuller a la base principal que el P-valor >0.05 , indicando la No estacionariedad de la serie. Luego, se realiza la prueba de estacionariedad a la base con una diferencia y notamos que con la prueba de Dickey Fuller el P-valor $(0.01) <0.05$, indicando estacionariedad la serie.

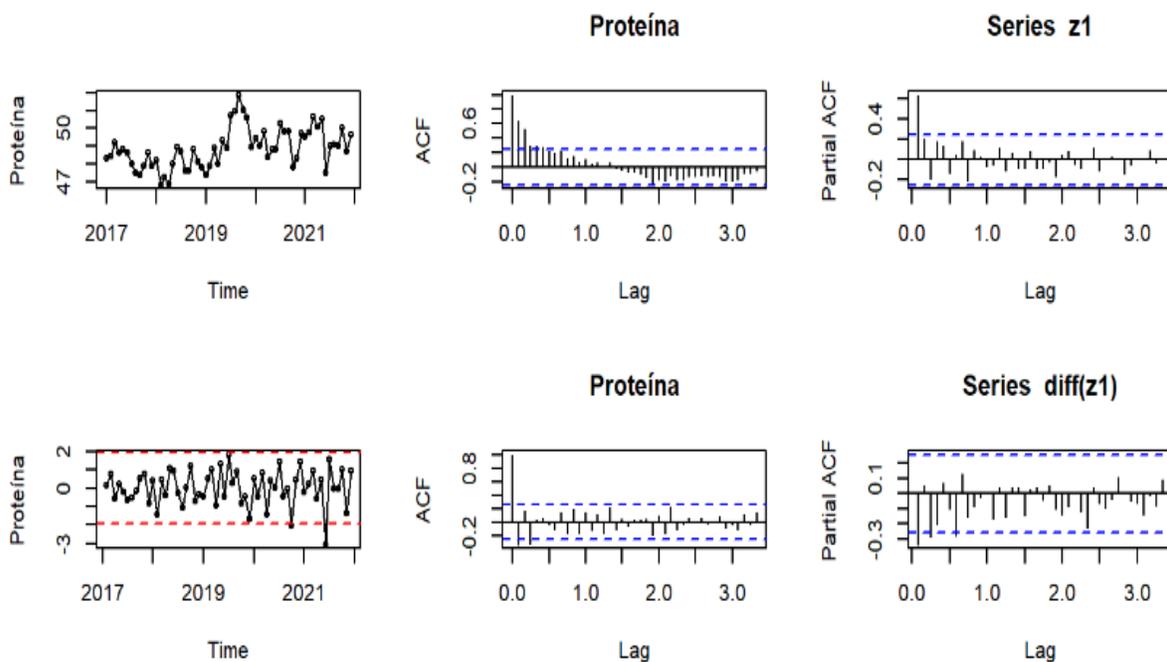


Figura 4: *Análisis de serie de tiempo diferenciada de Proteína.*

A partir de la figura 4, se plantean cinco (5) modelos arimas donde se escoge el de menor BIC, los resultados se muestran en la tabla 3.

Nro	Modelos	BIC	Ljung Box	Jarque Bera	Aleatoriedad	MAPE
1	1,1,1	162,6207	0,3215	0,1365	0,2579	0,63%
2	7,1,0	164,9859	0,9851	0,9808	0,7876	0,56%
3	0,1,3	153,2746	0,8897	0,9641	0,1204	0,55%
4	1,1,3	150,9416	0,9796	0,5704	0,1894	0,69%
5	3,1,1	158,7312	0,8496	0,4767	0,8875	0,64%

Tabla 3: Resumen Modelos Arima Proteína.

En base a los resultados, el modelo 4 Arima (1,1,3) es el que presenta mejores resultados, presenta menor BIC, además pasa las pruebas de Ljung Box, Jarque Bera, aleatoriedad y un MAPE (error porcentual absoluto medio) favorable. Las figuras 5 y 6 resultados obtenidas del software R studio del modelo seleccionado.

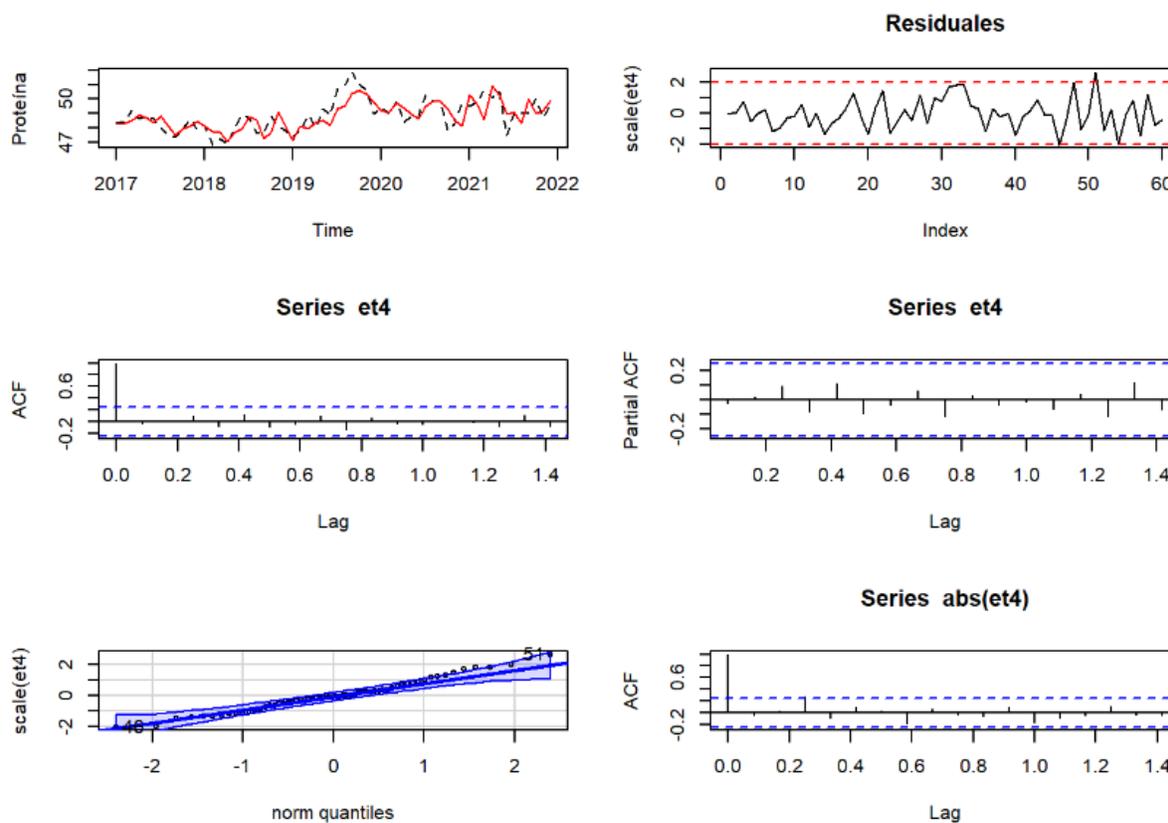


Figura 5: Gráficos residuales de Proteína.

En la figura 5, se observa los gráficos residuales, la línea punteada negra son los datos

reales y línea punteada roja son los ajustados. Tienen un comportamiento normal, no presentan tendencia, se asemeja a un ruido blanco y con varianza cero.

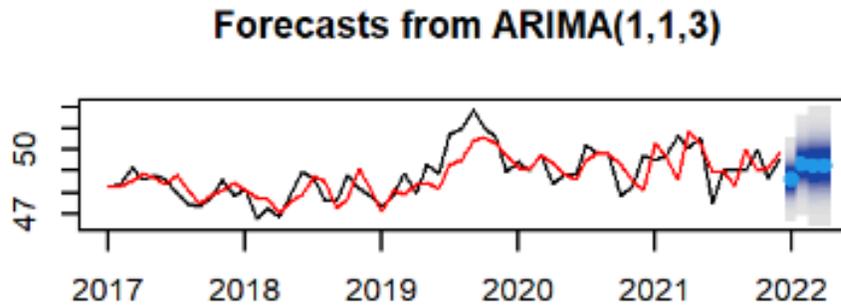


Figura 6: *Pronóstico de Proteína.*

En la figura 6, de acuerdo al modelo escogido, se pronostica para el 2022 en los primeros cuatro (4) meses estará con un MAPE (error porcentual absoluto medio) de 0,69%, se cuenta con la información real y se prueba el modelo. Por lo tanto, el diagnóstico es satisfactorio de cara a seleccionar el modelo y se plantea la siguiente ecuación (2).

Ecuación del modelo:

$$y_t = \phi_1 y_{t-1} + \rho_t + \theta_1 \rho_{t-1} + \theta_2 \rho_{t-2} + \theta_3 \rho_{t-3} \quad (2)$$

$$y_t = -0.3358 y_{t-1} + \rho_t - 0.6921 \rho_{t-3}$$

Ceniza

Análisis Descriptivo

Los resultados de ceniza obtenidos de los años estudiados se encuentran dentro de los rangos estipulados para este tipo de harinas, como se observa en la tabla 4.

Mínimo	30,70%
Primer cuartil	32,55%
Mediana	33,59%
Media	33,68%
Tercer cuartil	34,80%
Máximo	36,61%

Tabla 4: Resumen datos descriptivos Ceniza.

En la figura 7 se observa que el histograma se comporta como una distribución normal y que los datos la mayoría se encuentran en la parte central entre 31% y 36%.

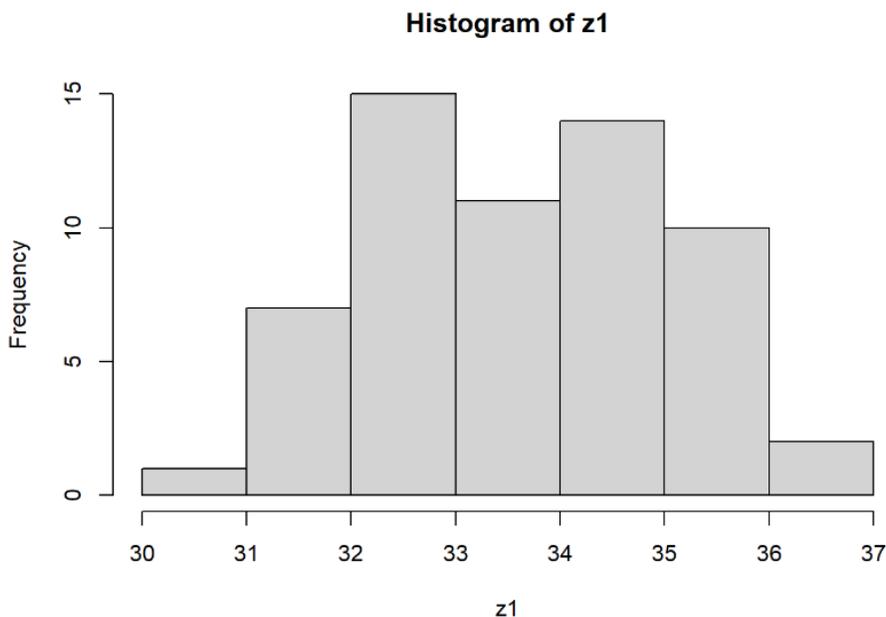


Figura 7: Histograma Ceniza.

Decomposition of additive time series

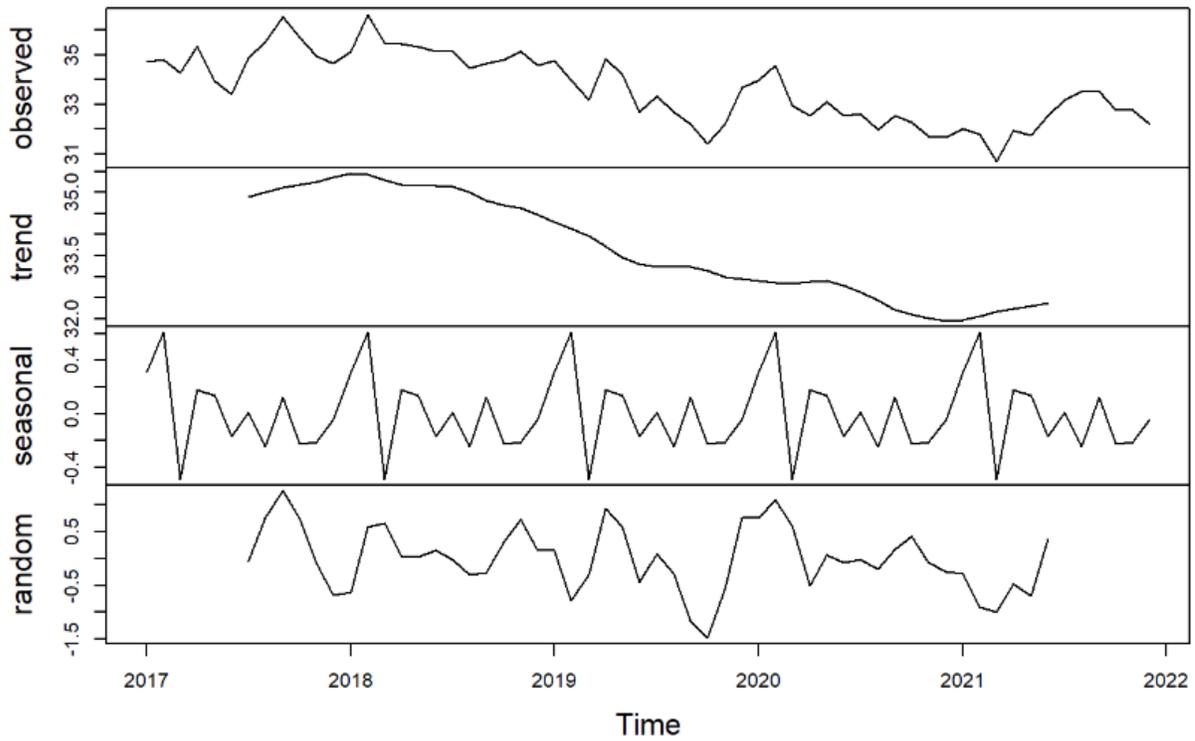


Figura 8: Componentes de la serie de tiempo para Ceniza.

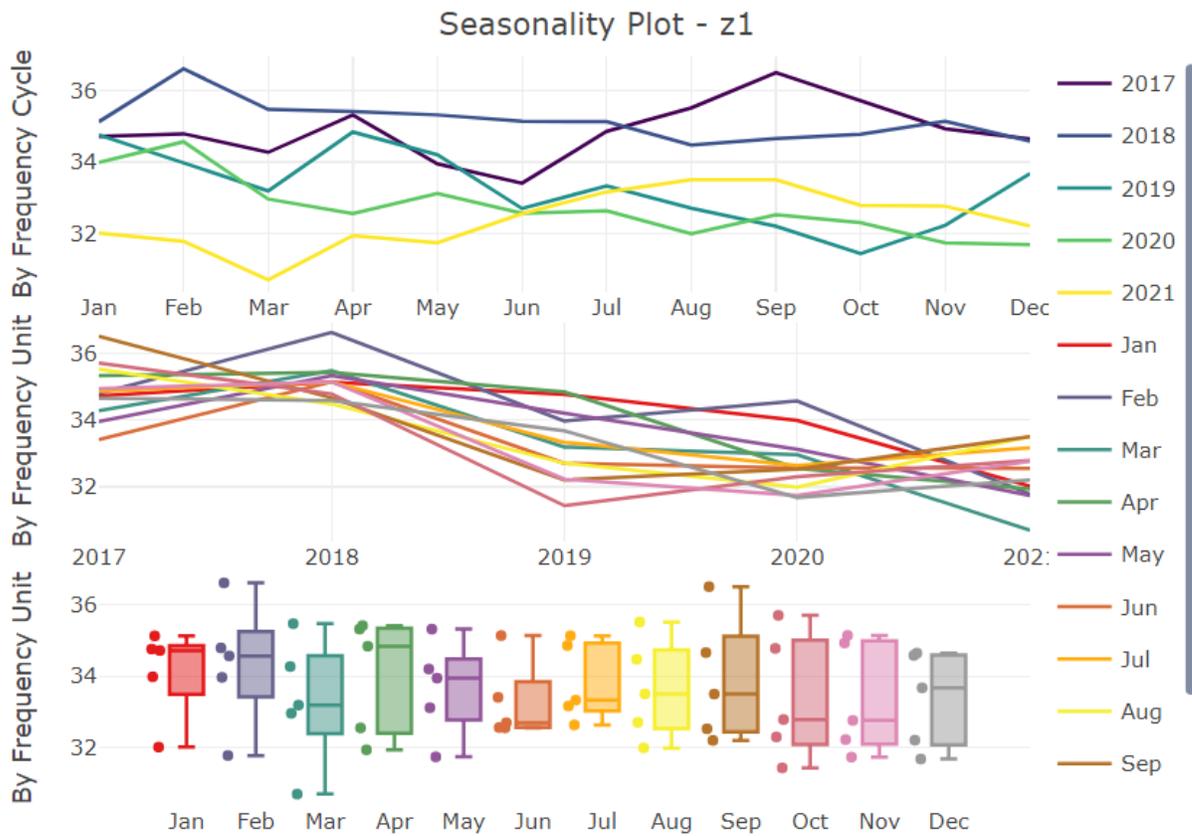


Figura 9: Comportamiento del periodo comprendido entre 2017 al 2021 de Ceniza.

En las figuras 8 y 9 se observa que el comportamiento histórico de ceniza, desde los años 2017 al 2019 se tuvo un promedio del 34%, en los años siguiente 2020 y 2021 el promedio fue decreciendo a 32%, notándose su tendencia a decrecer. Los meses de mayor incremento de la variable es en enero, febrero y abril.

Construcción de modelos

Como se observó en los gráficos descriptivos la tendencia a decrecer para la variable.

Se realiza la prueba de Dickey Fuller a la base principal que el P-valor >0.05 , indicando la No estacionariedad de la serie. Luego, se realiza la prueba de estacionariedad a la base con una diferencia y notamos que con la prueba de Dickey Fuller el P-valor $(0.01) <0.05$, indicando estacionariedad la serie.

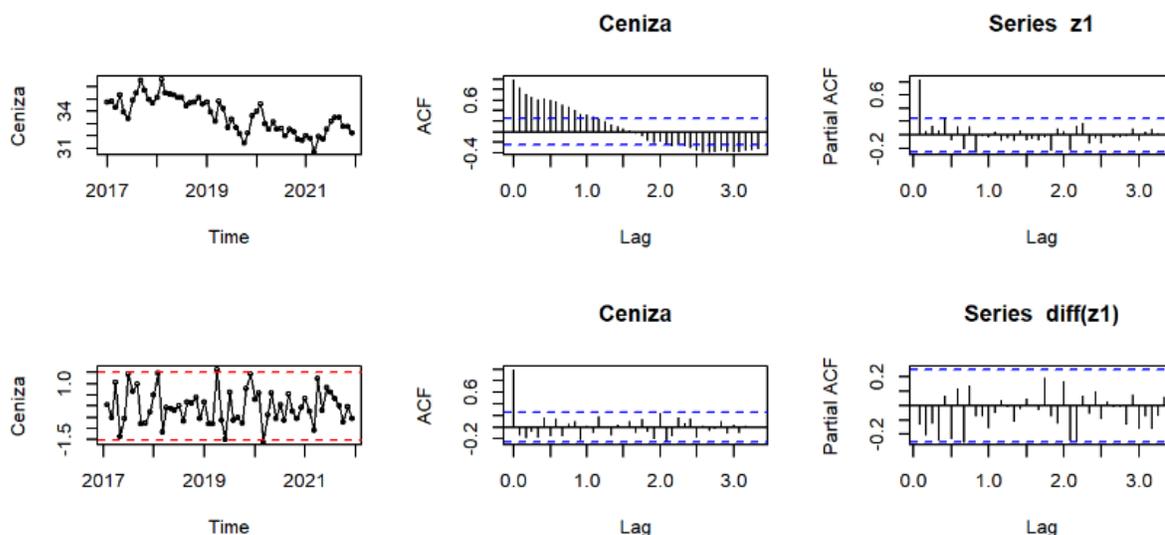


Figura 10: *Análisis de serie de tiempo diferenciada de Ceniza.*

A partir de la figura 10, se plantean cinco (5) modelos arimas donde se escoge el de menor BIC, los resultados se muestran en la tabla 5.

Nro	Modelos	BIC	Ljung Box	Jarque Bera	Aleatoriedad	MAPE
1	8,1,0	141,6449	0,9828	0,5115	0,3853	0,54%
2	0,1,2	140,8695	0,2901	0,3412	0,6843	1,20%
3	6,1,0	143,1739	0,3479	0,4908	0,3487	0,73%
4	1,1,2	138,7742	0,6470	0,2685	0,9644	1,37%
5	1,1,1	141,3511	0,6945	0,4973	0,9644	0,95%

Tabla 5: Resumen Modelos Arima Ceniza.

En base a los resultados, el modelo 4 Arima (1,1,2) es el que presenta mejores resultados, presenta menor BIC, además pasa las pruebas de Ljung Box, Jarque Bera, aleatoriedad y un MAPE (error porcentual absoluto medio) favorable. Las figuras 11 y 12 resultados obtenidas del software R studio del modelo seleccionado.

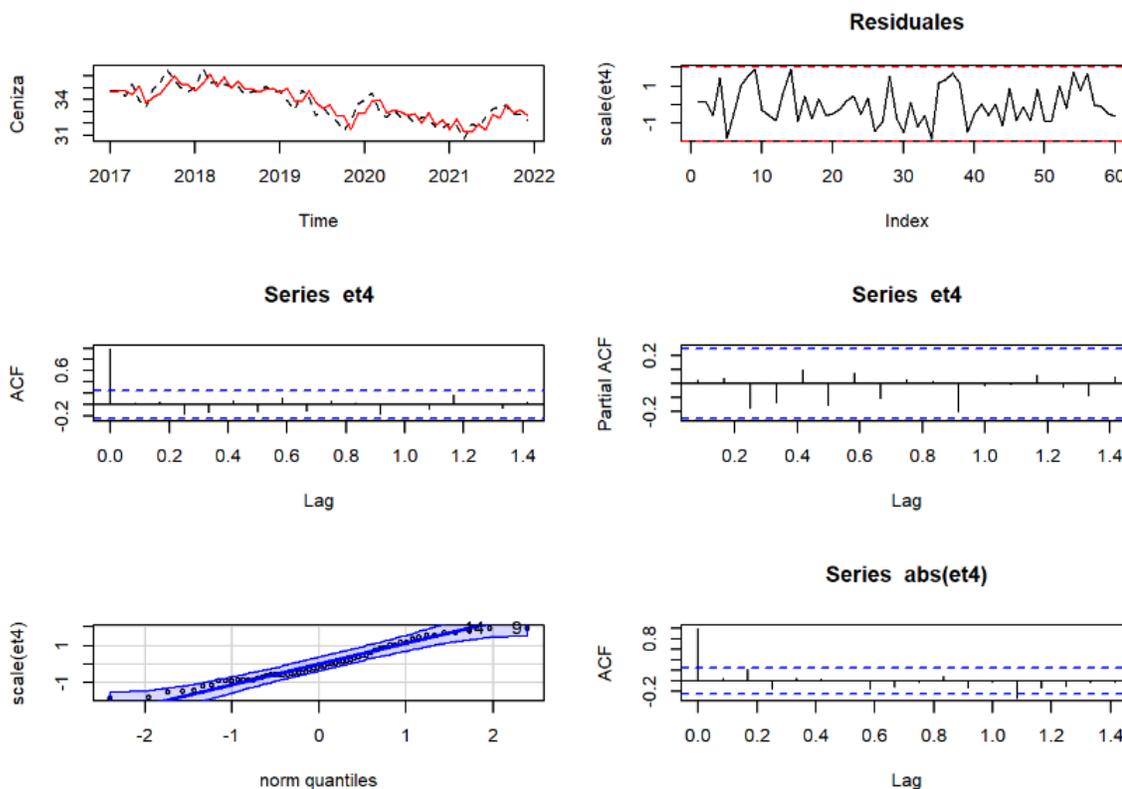


Figura 11: Gráficos residuales de Ceniza.

En la figura 11, se observa los gráficos residuales, la línea punteada negra son los datos reales y línea punteada roja son los ajustados. Tienen un comportamiento normal, no presentan tendencia, se asemeja a un ruido blanco y con varianza cero.

Forecasts from ARIMA(1,1,2)

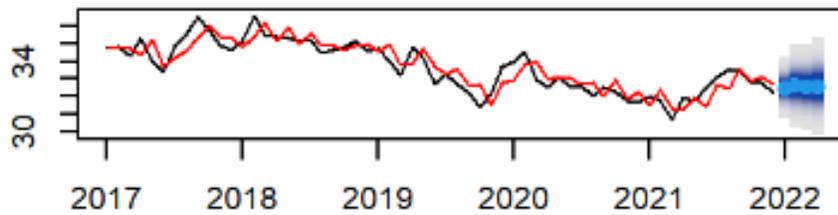


Figura 12: *Pronóstico de Ceniza.*

En la figura 12, de acuerdo al modelo escogido, se pronostica para el 2022 en los primeros cuatro (4) meses estará con un MAPE (error porcentual absoluto medio) de 1,37%, se cuenta con la información real y se prueba el modelo. Por lo tanto, el diagnóstico es satisfactorio de cara a seleccionar el modelo y se plantea la siguiente ecuación (3).

Ecuación del modelo:

$$y_t = \phi_1 y_{t-1} + \rho_t + \theta_1 \rho_{t-1} + \theta_2 \rho_{t-2} \quad (3)$$

$$y_t = -0.5179y_{t-1} + \rho_t + 0.4561\rho_{t-1} - 0.5439\rho_{t-2}$$

Grasa

Análisis Descriptivo

Los resultados de grasa obtenidos de los años estudiados se encuentran dentro de los rangos estipulados para este tipo de harinas, como se observa en la tabla 6.

Mínimo	9,52%
Primer cuartil	10,65%
Mediana	11,13%
Media	11,05%
Tercer cuartil	11,49%
Máximo	12,57%

Tabla 6: Resumen datos descriptivos Grasa.

En la figura 13 se observa que el histograma se comporta como una distribución normal y que los datos la mayoría se encuentran en la parte central entre 10,5% y 12%.

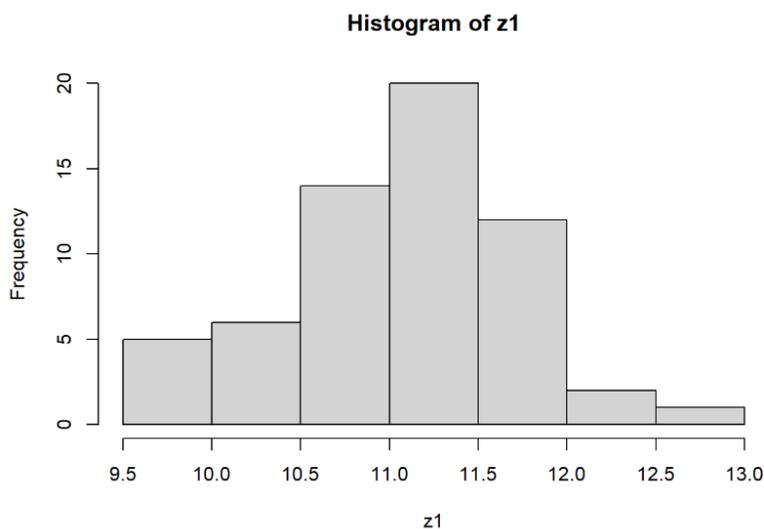


Figura 13: Histograma Grasa.

Decomposition of additive time series

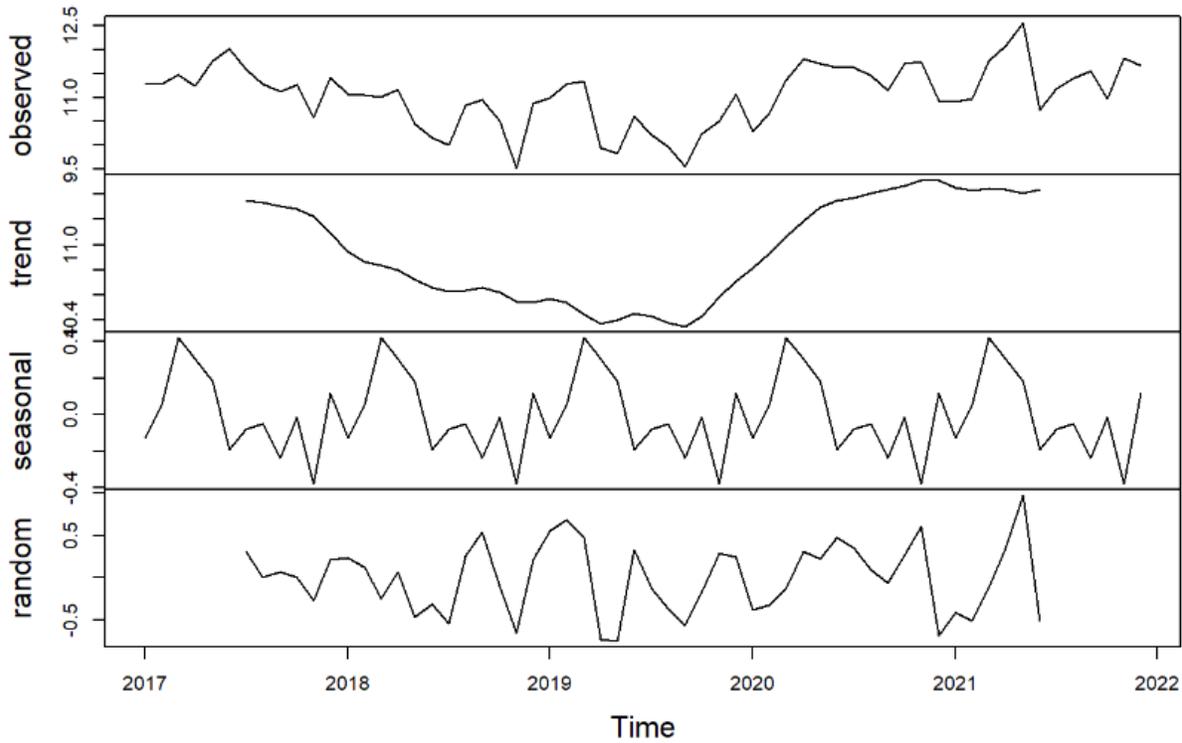


Figura 14: Componentes de la serie de tiempo para Grasa.

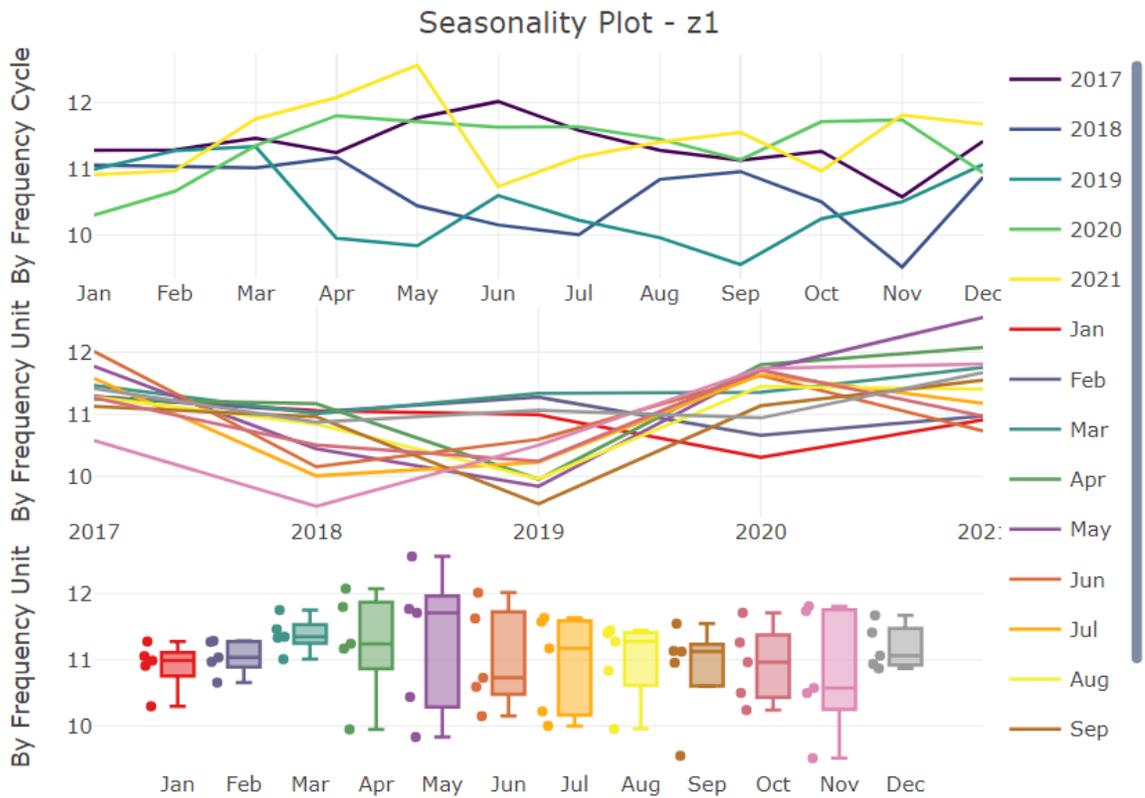


Figura 15: Comportamiento del periodo comprendido entre 2017 al 2021 de Grasa.

En las figuras 14 y 15 se observa que el comportamiento histórico de grasa, desde los años 2017 al 2020 se tuvo un decrecimiento llegando a 10,4%, en los años siguiente 2020 y 2021 el promedio fue incrementando 12%, notándose su tendencia a incrementar. Los meses de mayor incremento de la variable es en marzo, abril y mayo.

Construcción de modelos

Como se observó en los gráficos descriptivos la tendencia a decrecer y/o incrementar para la variable. Se realiza la prueba de Dickey Fuller a la base principal que el P-valor >0.05 , indicando la No estacionariedad de la serie. Luego, se realiza la prueba de estacionariedad a la base con una diferencia y notamos que con la prueba de Dickey Fuller el P-valor (0.01) <0.05 , indicando estacionariedad la serie.

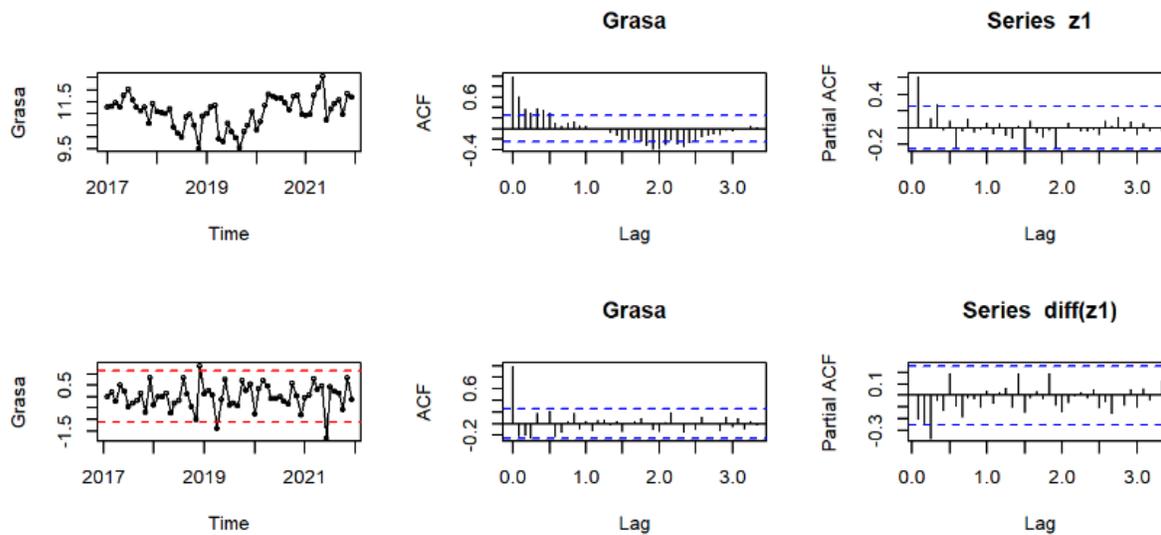


Figura 16: *Análisis de serie de tiempo diferenciada de Grasa.*

A partir de la figura 16, se plantean cinco (5) modelos arimas donde se escoge el de menor BIC, los resultados se muestran en la tabla 7.

Nro	Modelos	BIC	Ljung Box	Jarque Bera	Aleatoriedad	MAPE
1	3,1,0	100,7955	0,8492	0,5282	0,1249	1,75%
2	0,1,2	99,9662	0,5683	0,4083	0,3118	1,38%
3	1,1,1	100,2822	0,5034	0,2789	0,1471	1,21%
4	0,1,1	100,5653	0,1338	0,3644	0,1249	1,39%
5	1,1,0	104,7008	0,03766	0,05691	0,3118	1,77%

Tabla 7: Resumen Modelos Arima Grasa.

En base a los resultados, el modelo 2 Arima (0,1,2) es el que presenta mejores resultados, presenta menor BIC, además pasa las pruebas de Ljung Box, Jarque Bera, aleatoriedad y un MAPE (error porcentual absoluto medio) favorable. Las figuras 17 y 18 resultados obtenidas del software R studio del modelo seleccionado.

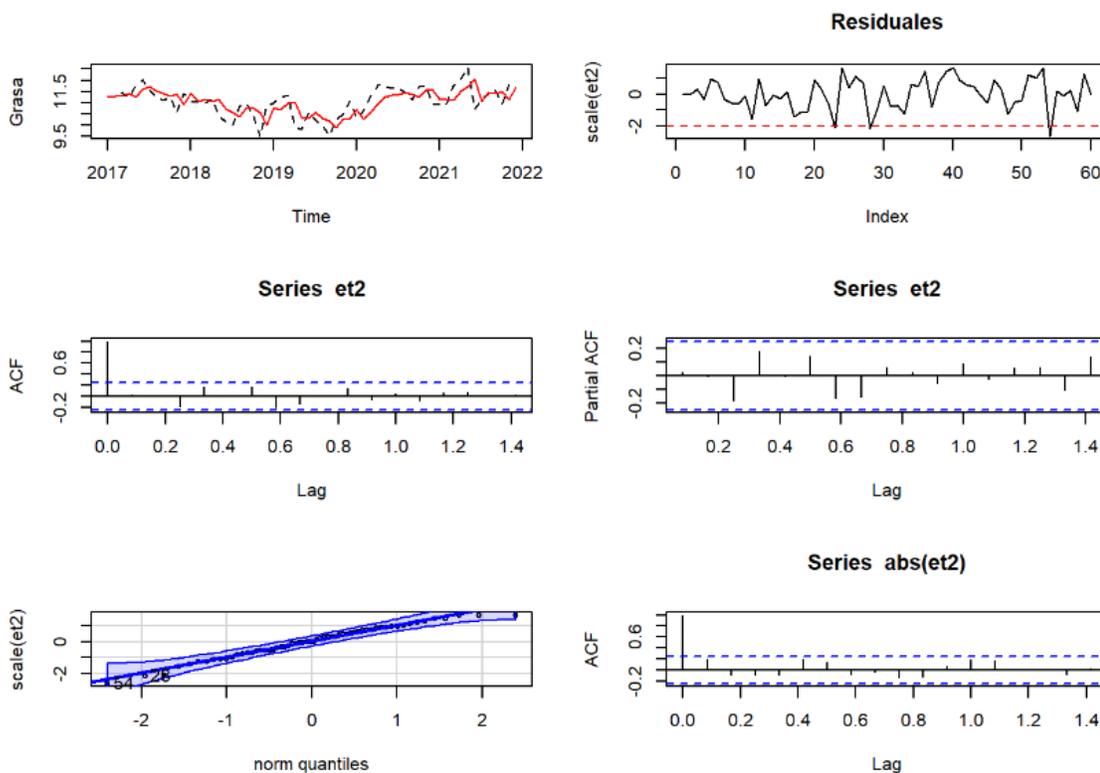


Figura 17: Gráficos residuales de Grasa.

En la figura 17, se observa los gráficos residuales, la línea punteada negra son los datos reales y línea punteada roja son los ajustados. Tienen un comportamiento normal, no presentan tendencia, se asemeja a un ruido blanco y con varianza cero.

Forecasts from ARIMA(0,1,2)

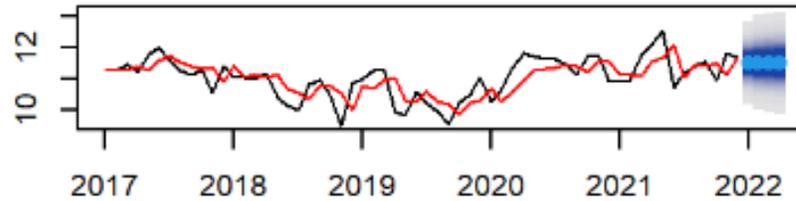


Figura 18: *Pronóstico de Grasa.*

En la figura 18, de acuerdo al modelo escogido, se pronostica para el 2022 en los primeros cuatro (4) meses estará con un MAPE (error porcentual absoluto medio) de 1,38%, se cuenta con la información real y se prueba el modelo. Por lo tanto, el diagnóstico es satisfactorio de cara a seleccionar el modelo y se plantea la siguiente ecuación (4).

Ecuación del modelo:

$$y_t = \rho_t + \theta_1\rho_{t-1} + \theta_2\rho_{t-2} \quad (4)$$

$$y_t = \rho_t - 0.4127\rho_{t-1} - 0.2450\rho_{t-2}$$

Humedad

Análisis Descriptivo

Los resultados de humedad obtenidos de los años estudiados se encuentran dentro de los rangos estipulados para este tipo de harinas, como se observa en la tabla 8.

Mínimo	1,49%
Primer cuartil	2,42%
Mediana	2,61%
Media	2,69%
Tercer cuartil	2,97%
Máximo	3,92%

Tabla 8: Resumen datos descriptivos Humedad.

En la figura 19 se observa que el histograma se comporta como una distribución normal y que los datos la mayoría se encuentran en la parte central entre 2% y 3,5%.

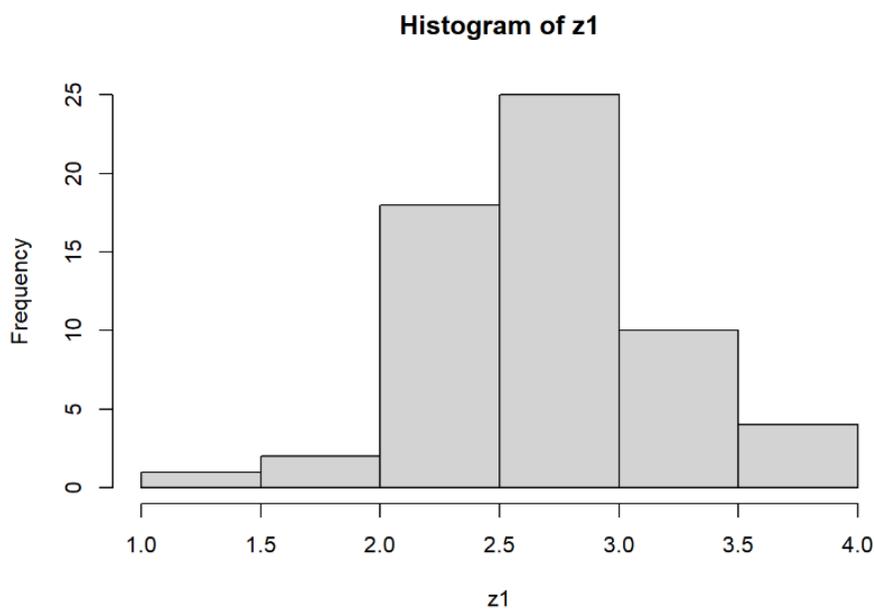


Figura 19: Histograma Humedad.

Decomposition of additive time series

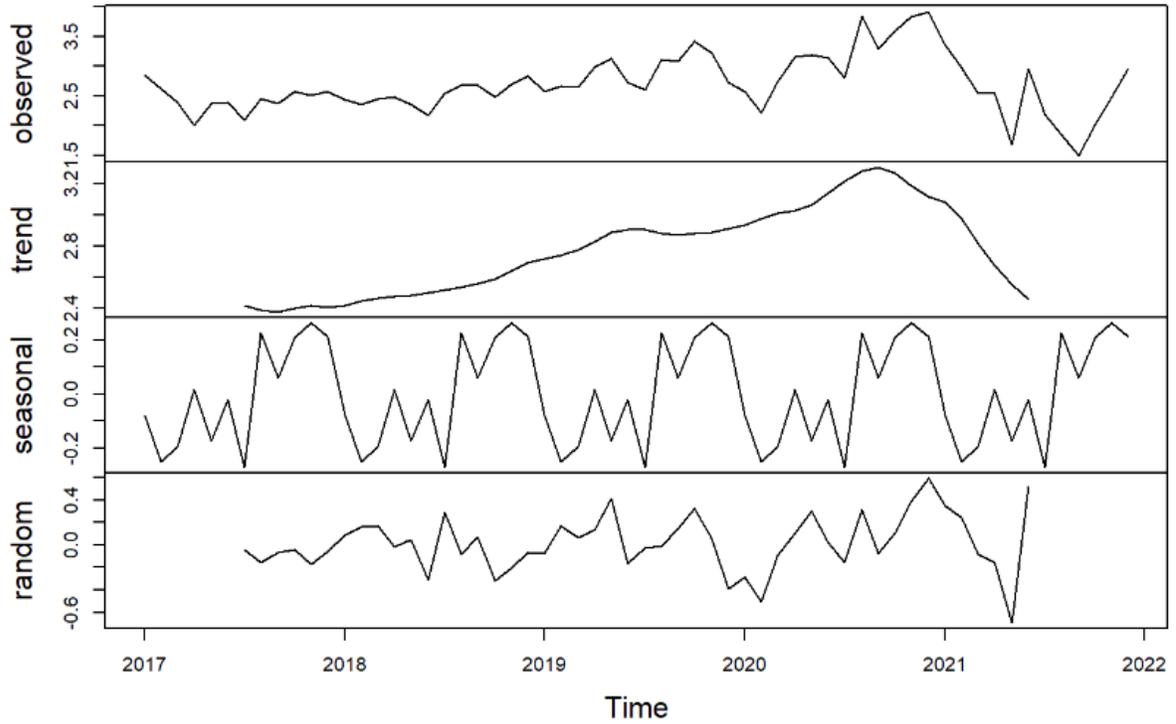


Figura 20: Componentes de la serie de tiempo para Humedad.

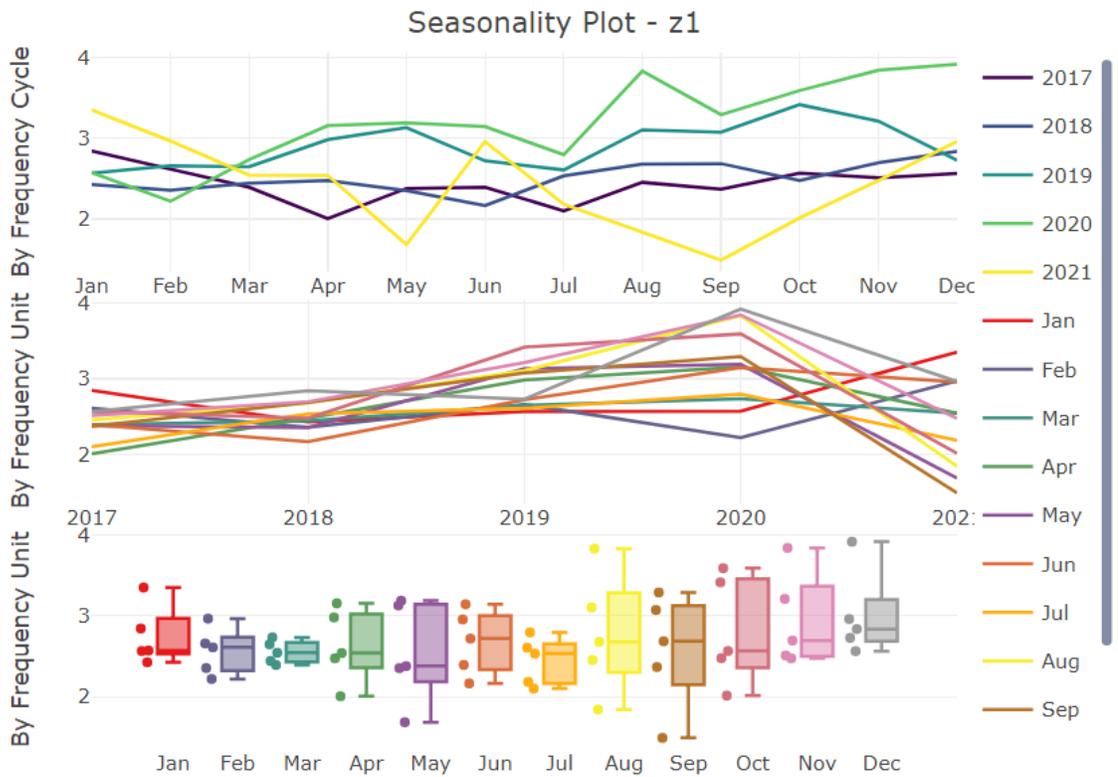


Figura 21: Comportamiento del periodo comprendido entre 2017 al 2021 de Humedad.

En las figuras 20 y 21 se observa que el comportamiento histórico de humedad, desde los años 2017 al 2020 se tuvo un incremento para un posterior decrecimiento en el 2021, notándose su tendencia. Los meses de mayor incremento de la variable es en julio, agosto y septiembre.

Construcción de modelos

Como se observó en los gráficos descriptivos la tendencia a incrementar y/o decrecer para la variable. Se realiza la prueba de Dickey Fuller a la base principal que el P-valor >0.05 , indicando la No estacionariedad de la serie. Luego, se realiza la prueba de estacionariedad a la base con una diferencia y notamos que con la prueba de Dickey Fuller el P-valor (0.01) <0.05 , indicando estacionariedad la serie.

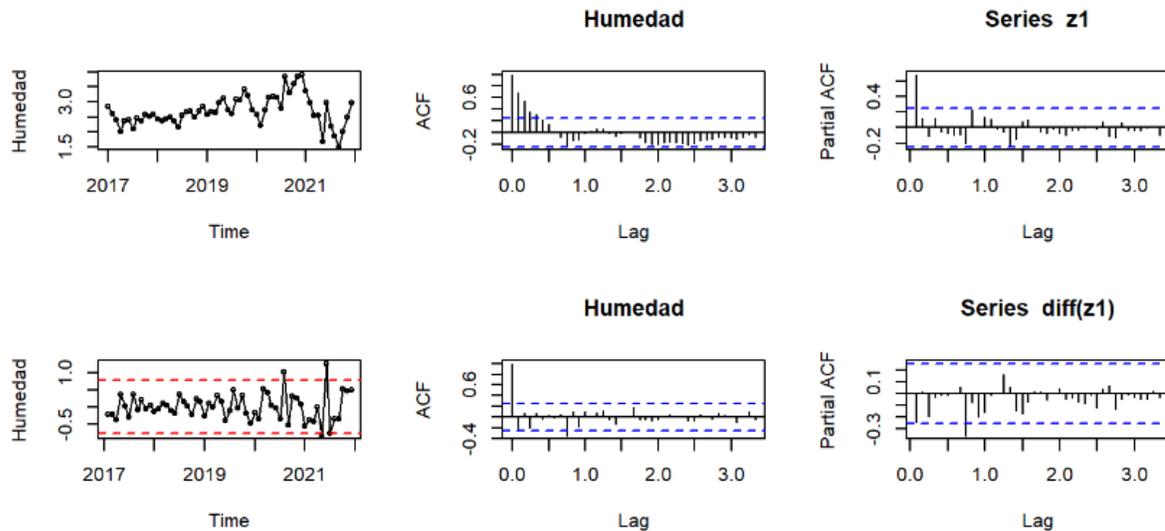


Figura 22: *Análisis de serie de tiempo diferenciada de Humedad.*

A partir de la figura 22, se plantean cinco (5) modelos arimas donde se escoge el de menor BIC, los resultados se muestran en la tabla 9.

Nro	Modelos	BIC	Ljung Box	Jarque Bera	Aleatoriedad	MAPE
1	3,1,0	65,1594	0,3899	0,6431	0,6253	17,51%
2	0,1,9	50,5734	0,7466	0,464	0,6081	7,76%
3	1,1,0	59,8371	0,3073	0,4398	0,3118	9,65%
4	0,1,1	59,6725	0,2515	0,6152	0,9931	12,19%
5	1,1,9	50,5638	0,8806	0,5116	0,9371	7,39%

Tabla 9: Resumen Modelos Arima Humedad.

En base a los resultados, el modelo 5 Arima (1,1,9) es el que presenta mejores resultados, presenta menor BIC, además pasa las pruebas de Ljung Box, Jarque Bera, aleatoriedad y un MAPE (error porcentual absoluto medio) el indicador más bajo. Las figuras 23 y 24 resultados obtenidas del software R studio del modelo seleccionado.

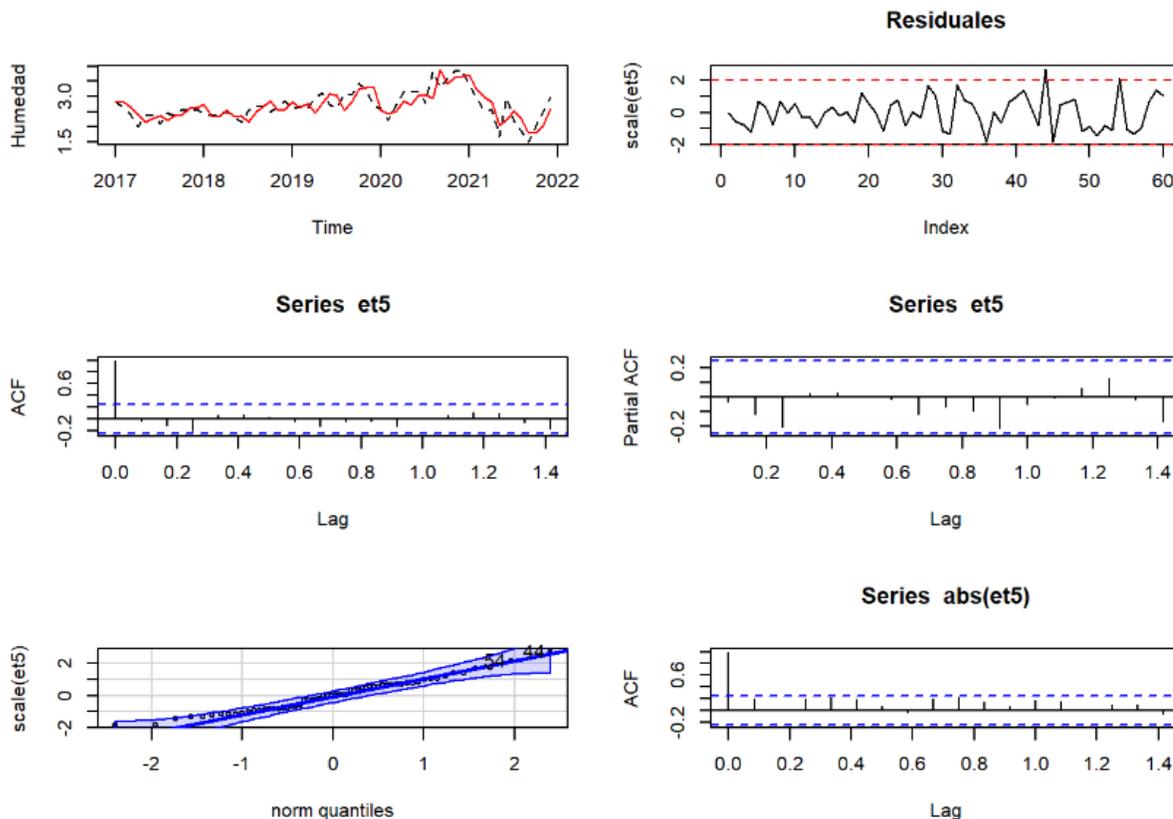


Figura 23: Gráficos residuales de Humedad.

En la figura 23, se observa los gráficos residuales, la línea punteada negra son los datos reales y línea punteada roja son los ajustados. Tienen un comportamiento normal, no presentan tendencia, se asemeja a un ruido blanco y con varianza cero.

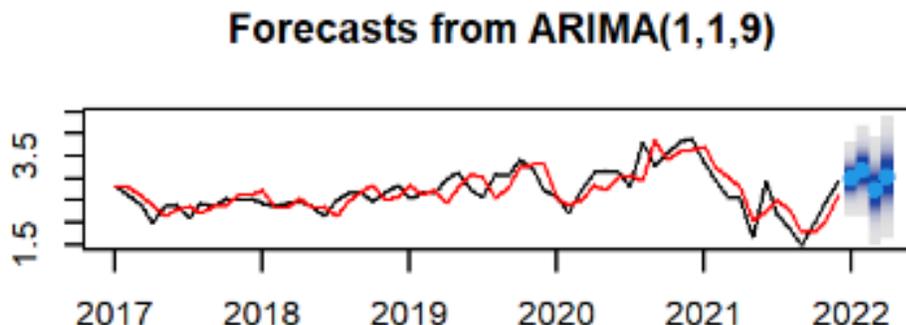


Figura 24: *Pronóstico de Humedad.*

En la figura 24, de acuerdo al modelo escogido, se pronostica para el 2022 en los primeros cuatro (4) meses estará con un MAPE (error porcentual absoluto medio) de 7,39%, se cuenta con la información real y se prueba el modelo. Por lo tanto, el diagnóstico es satisfactorio de cara a seleccionar el modelo y se plantea la siguiente ecuación (5).

Ecuación del modelo:

$$y_t = \phi_1 y_{t-1} + \rho_t + \theta_1 \rho_{t-1} + \theta_2 \rho_{t-2} + \theta_3 \rho_{t-3} + \theta_4 \rho_{t-4} + \theta_5 \rho_{t-5} + \theta_6 \rho_{t-6} + \theta_7 \rho_{t-7} + \theta_8 \rho_{t-8} + \theta_9 \rho_{t-9} \quad (5)$$

$$y_t = -0.2749y_{t-1} + \rho_t - 0.5675\rho_{t-9}$$



CONCLUSIONES

Al realizarse el análisis del comportamiento de las variables utilizadas en los periodos comprendidos entre 2017 y 2021, se aplicó una serie diferenciada y así obtener el mejor modelo para cada una de las variables: De la Proteína el modelo ARIMA (1,1,3), de la Ceniza el modelo ARIMA (1,1,2), de la Grasa ARIMA (0,1,2) y de la Humedad el modelo ARIMA(1,1,9).

Para los modelos establecidos de las variables Proteína, Ceniza, Grasa y Humedad, como resultados de Ljung Box se tomó en cuenta los modelos que presentaron una gran mayoría de residuales que superan las bandas de confianza. Para que el modelo fuese confiable al momento de su predicción.

En los primeros cuatro (4) meses del 2022, se obtuvieron resultados para los modelos, Proteína= 0,69%; Ceniza= 1,37%; Grasa= 1,38% y Humedad= 7,39%, siendo esta última, la variable que arrojó resultados superiores al 5%. Se cuenta con la información real y se prueban los modelos. Por lo tanto, los diagnósticos son satisfactorios de cara a seleccionar los modelos.

Los modelos ajustados resultan acertados para los pronósticos de los parámetros estudiados. Sin embargo, se abre la posibilidad de ajustar un solo modelo para todas las variables de la harina de origen animal, a través de las series de tiempo multivariada y se recomienda estudiar las variables que se dejaron por fuera de estudio, que inciden en las características de las harinas, como digestibilidad, acidez, índice de peróxidos.

REFERENCIAS BIBLIOGRÁFICAS.



- Barrera, V., Tapia, C., & Montero, A. (2004). Raíces y tubérculos andinos: alternativas para la conservación y uso sostenible en el Ecuador. Quito: INIAP.
- Bedoya, C. (2016). *Metodologías para el análisis bromatológico, físico y químico del cacao fermentado y seco, dentro del marco normativo internacional (Tesis de pregrado, Corporación Universitaria Lasallista, Caldas, Colombia)*. Obtenido de <http://repository.unilasallista.edu.co/dspace/handle/10567/2064>
- Belouafa, S., Habti, F., Benhar, S., Belafkih, B., Tayane, S., Hamdouch, S., Bennamara, A., and Abourriche, A. (2017). Statistical tools and approaches to validate analytical methods: methodology and practical examples. *Int. J. Metrol. Qual.*, Eng. 8, 9.
- Blake, J. P. (1999). Administración y procesamiento de los subproductos (desperdicios) de la industria avícola. Presentado en el XVI Congreso Latinoamericano de Avicultura, Septiembre. Lima.
- Box, G. E. P. y Jenkins, Gwilym M. (1976). Time series analysis: Forecasting and control. Revised Edition. Oakland, California: Editorial Holden-Day, 1976.
- Diebold, F. (2001). *Elements of forecasting 2nd ed.* South y W. C. Publishing., eds. Australia.
- Egrioglu, E., Yolcu, U. y Aladag, C. (2015). Red neuronal artificial multiplicativo modelo neuronal recurrente para el pronóstico de series de tiempo no lineales. *Proceso Neural. Lett.*, 41(2). 249-258.
- Galán Madruga, D., Fernández Patier, R. y Díaz Esteban A. (2012). Validación de métodos analíticos. Implementación al análisis rutinario. *Hig. Sanid. Ambient.* 12 (3). 882-888.
- Gardener, M. (2012). *Beginning R the statistical programming language*. Indianapolis, Indiana: John Wiley & Sons, Inc.
- Giraldo, N. (2011). *Modelado y Pronóstico de la Tendencia.*, pp.17–29. Obtenido de http://www.medellin.unal.edu.co/~ndgiraldo/Archivos%20Lectura/Archivos%20curso%20Series%20EIO/Trabajo_No1_EstadisticaIII_sem01_12.pdf [Accedido Marzo 04, 2017]
- Guerrero, V. (2003). *Análisis estadístico de series de tiempo económicas. Segunda edición*. México: Editorial Thomson.
- Hyndman, R.J., Khandakar, Y. (2008). Serie de Tiempo automático de predicción: el paquete pronóstico para R. *J. Stat. Software*, 27 (3). 1-22.
- Juárez, A. C., Zuñiga, C. A., Flores, J. L. M., y Partida, D. S. (2016). Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos. *Estudios Gerenciales*, 32(141), 387-396.
- Lütkepohl, H. (2005). *Nueva Introducción al Análisis Temporal de Series Múltiples*. Springer-Verlag Berlin Heidelberg.
- Magnusson, B. and Ornemark U. (2014). *Eurachem Guide: The fitness for purpose of analytical methods – A laboratory guide to method validation and related topics, 2nd ed.* . Obtenido de <https://www.eurachem.org/index.php/publications/guides/mv>
- Meier, P.C. and Zund, R.E. . (2000). *Statistical methods in analytical chemistry 2nd Edition. United States of America: John Wiley & Sons, Inc.*
- Moreno-García, A., Moreno-García, J., Jiménez-Linares, L., y Rodríguez-Benitez, L. (2017). Time series represented by means of fuzzy piecewise lineal segments. *Journal of Computational and Applied Mathematics*, 318, 156–167.
- Ortiz, S. (2006). *Determinación de la composición química proximal y fibra dietaria de 43 variedades criollas de maíz de 7 municipios del sureste del estado de Hidalgo (Tesis de pregrado, Universidad Autónoma del Estado de Hidalgo, Hidalgo, México)*. Obtenido de <http://dgsa.uaeh.edu.mx:8080/bibliotecadigital/handle/231104/630>
- Pravilovic, S., Bilancia, M., Appice, A., y Malerba, D. (2017). Using multiple time series analysis for geosensor data forecasting. *Information Sciences*, 380, 31–52.
- Puerta, G. (2006). La humedad controlada del grano preserva la calidad del café. CENICAFE. *Avances Técnicos No. 352*, 8.
- Quintana, M. J. G., y Jiménez, S. A. M. (2016). Modelos de series de tiempo aplicados a los expedientes de la Comisión de Derechos Humanos del Distrito Federal. *Economía Informa*, 398, 89-99.
- Suárez, L. S. (2013). Pronóstico de la producción de leche, mediante modelos ARIMA. Caso UBPC “Maniabo”. Universidad de La Habana.



- Wattiaux, M., Homan, J., Moreno, M. y Rodríguez, A. (1998). *Nutrición y alimentación*. Obtenido de http://soda.ustadistancia.edu.co/enlinea/2momento_alimentacionanimal_mariabelalcazar/Composicion_y_Analisis_de_los_Alimentos.pdf
- Zunino, L., Olivares, F., Bariviera, A. F., y Rosso, O. A. (2017). A simple and fast representation space for classifying complex time series. *Physics Letters A*, 381(11), 1021–1028.