

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS**

**DEPARTAMENTO DE POSTGRADOS**

**PROYECTO DE TITULACIÓN**

**PREVIO A LA OBTENCIÓN DEL TÍTULO DE:**

**“MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN CALIDAD Y PRODUCTIVIDAD”**

**TEMA:**

Construcción de un modelo estadístico para el pronóstico de la demanda de alimento balanceado para camarones en una empresa comercializadora de la ciudad de Guayaquil.

**AUTOR:**

**OSCAR WILLIAM SANCHEZ CORDOVA**

Guayaquil - Ecuador

2020

## **RESUMEN**

El sector camaronero actualmente es protagonista en la economía ecuatoriana, el 28.5% de las exportaciones no petroleras del año 2019 corresponden a este producto. El despunte de este sector beneficia a toda su cadena de suministro entre ellos a los proveedores de alimento para los crustáceos que es uno de los pilares fundamentales en el incremento de la producción. Para que el sector camaronero mantenga este rendimiento debe tener sincronizada toda su cadena de suministro, con ello se asegura la disponibilidad de recursos en cantidad, calidad y en el momento requerido. Para ello se debe contar con pronósticos los más exactos posibles para lograr cubrir o amortiguar cualquier cambio en la industria.

La serie de la demanda de alimento para camarones generalmente presenta tendencias y estacionalidades lo cual constituye un desafío para desarrollar pronósticos efectivos. En el presente trabajo se compara el rendimiento de pronóstico de los modelos ARIMA, de regresión estática y regresión dinámica. El rendimiento de las previsiones se aplica en uno de los principales productos que comercializa una conocida marca de alimento balanceado para camarones. Los modelos fueron evaluados a través del RMSE y MAE en los datos de entrenamiento y de prueba. Se observó que los modelos ARIMA y regresión dinámica son los que mejor ajustan los datos, pero en los datos de prueba no tuvieron los mejores resultados. Contrario a lo anterior se tiene a los modelos estáticos que arrojaron los ajustes más débiles, pero con pronósticos más acertados.

**Palabras clave: Arima, Demanda de alimento balanceado, Modelo predictivo, Regresión estática y dinámica**

## **ABSTRACT**

The shrimp sector is currently the protagonist in the Ecuadorian economy, 28.5% of non-oil exports in 2019 correspond to this product. The emergence of this sector benefits its entire supply chain, including crustacean food suppliers, which is one of the fundamental supports to have an increased production. For the shrimp sector to maintain this performance, its complete supply chain must be synchronized, thus ensuring the availability of quantity and quality resources at the time required. To do this, it must have the most accurate forecasts possible to cover or buffer any change in the industry.

The series of shrimp feed demand generally presents trends and seasonality, which constitutes a challenge to develop effective forecasts. This paper compares the forecast performance of the ARIMA, static regression and dynamic regression models. Forecast performance is applied to one of the leading products marketed by a well-known brand of shrimp feed. Models were evaluated through RMSE and MAE on training and test data. The ARIMA and dynamic regression models were found to be the best fit data, but they did not have the best results in the test data. In contrast with the above, static models that showed the weakest adjustments, having more accurate forecasts.

**Key words: Arima, Balanced feed demand, Predictive model, Static and dynamic regression**

## **DEDICATORIA**

Al único ser en el universo en el cual no existe incertidumbre: al Dios Vivo

## **AGRADECIMIENTO**

En estas líneas quiero manifestar mi gratitud a todos aquellos que directa e indirectamente influyeron en mi formación personal y académica.

A mis padres Gladys y Oscar por sus consejos que perdurarán por siempre. A mis hermanos gracias totales por su ayuda.

A mi compañera de vida ...Stephania

A mi tutora Nadia Cárdenas que con su consejería pude sacar adelante este trabajo.

A dos grandes amigas Kiana y Rosalva que fueron mis acompañantes por este sendero.

## DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

A handwritten signature in blue ink, appearing to read 'Oscar Sánchez', written over a horizontal line.

Oscar William Sánchez Córdova

## TRIBUNAL DE GRADUACIÓN



---

Holger Cevallos Valdiviezo, Ph.D.

PRESIDENTE



---

M.Sc. Nadia Cárdenas Escobar

DIRECTOR



---

M.Sc. Francisco Moreira Villegas

VOCAL 1



---

Dr. Francisco Vera Alcivar

VOCAL 2

## ABREVIATURAS O SIGLAS

ARIMA: Autoregressive integrated moving average (autorregresivo integrado media móvil)

AIC: Akaike's information criterion (Criterio de información Akaike)

AIC<sub>C</sub> : Corrected Akaike's information criterion (Criterio de información Akaike corregido)

ACF: Autocorrelation function (Función de autocorrelación)

BIC: Schwarz's Bayesian Information Criterion (Criterio de información Bayesiano de Schwarz)

MAE: Mean Absolute Error (Error absoluto medio)

MAPE: Mean Absolute Percentage Error (Error porcentual absoluto medio)

MSE: Mean squared Error (Error cuadrático medio)

PACF: Partial Autocorrelation Function (Función de autocorrelación Parcial)

PIB: Producto Interno Bruto

RMSE: Root Mean Squared Error (Raíz del error cuadrático medio)



<b>CAPÍTULO 1</b> .....	<b>1</b>
<b>INTRODUCCIÓN</b> .....	<b>1</b>
1.1 Antecedentes .....	1
1.2 Descripción del problema .....	5
1.3 Justificación .....	6
1.4 Objetivos .....	7
1.4.1. Objetivos Generales.....	7
1.4.2. Objetivos Específicos .....	7
1.5 Alcance .....	7
<b>CAPÍTULO 2</b> .....	<b>8</b>
<b>MARCO TEÓRICO</b> .....	<b>8</b>
2.1 Estado del Arte.....	8
2.1.1 Satisfacción del cliente en un mundo con incertidumbre .....	8
2.1.2 Insatisfacción por tiempo de entregas y flexibilidad en volumen. ....	9
2.1.3 Pronóstico de la demanda y su relación con la gestión de inventarios. ....	10
2.1.4 Aplicaciones de modelos de previsión de demanda. ....	12
2.2 Marco Teórico .....	16
2.2.1 Modelo .....	16
2.2.1.1 Modelo Estadístico .....	16

2.2.2 Métodos para el modelamiento predictivo .....	18
2.2.2.1 Modelos paramétricos .....	18
2.2.2.2. Modelos no paramétricos .....	19
2.2.3 Tipos de Pronósticos .....	19
2.2.4 Serie de tiempo .....	20
2.2.5 Proceso estocástico .....	21
2.2.5.1 Características de un proceso estocástico .....	22
2.2.5.2 Proceso Estocástico estacionario .....	23
2.2.5.3 Proceso Estocástico no estacionario .....	24
2.2.5.4 Raíz Unitaria .....	24
2.2.6 Modelos Lineales Estacionarios .....	26
2.2.6.1 Modelo de autorregresivo AR (p) .....	28
2.2.6.2 Proceso media móviles MA (q).....	29
2.2.6.3 Procesos Autorregresivos de Medias Móviles .....	30
2.2.6.4 Modelo Arima .....	32
2.2.7 Regresión lineal .....	34
2.2.7.1 Modelos de Regresión Dinámica .....	38
2.2.7.2 Criterios de selección de variables: .....	40
2.2.8 Errores de pronósticos .....	41
<b>CAPÍTULO 3 .....</b>	<b>43</b>
<b>METODOLOGÍA.....</b>	<b>43</b>

3.1 Análisis Descriptivo.....	43
3.1.1 Análisis descriptivo de variables macroeconómicas .....	47
3.2 Construcción y evaluación del modelo.....	54
3.2.1 Modelo ARIMA estacional.....	54
3.2.2 Modelo de Regresión Múltiple.....	64
3.2.3 Modelos de Regresión Dinámica.....	71
<b>CAPÍTULO 4 .....</b>	<b>74</b>
<b>VALIDACIÓN Y PRONÓSTICO DEL MODELO .....</b>	<b>74</b>
4.1 Validación de supuestos del modelo seleccionado.....	74
4.1.1 Errores con distribución normal.....	74
4.1.2 Errores no correlacionados entre si .....	75
4.1.3 Varianza constante de los errores .....	76
4.1.4 Multicolinealidad.....	78
4.2 Predicción a partir del modelo.....	79
<b>CONCLUSIONES Y RECOMENDACIONES.....</b>	<b>82</b>
5.1 Conclusiones.....	82
5.2 Recomendaciones.....	85
<b>BIBLIOGRAFÍA.....</b>	<b>86</b>
<b>ANEXOS .....</b>	<b>88</b>

## LISTADO DE FIGURAS

Gráfica 1. Ventas de Alimento Balanceado.....	44
Gráfico 2. Ventas estacionales de alimento balanceado.....	45
Gráfico 3. Ventas realizadas por referencia.....	46
Gráfico 4. PIB trimestral Miles de dólar (Constantes) 2007=100.....	48
Gráfico 5. Producto Interno Bruto-PIB 2007= 100, Tasas de variación trimestral.....	49
Gráfico 6. Nivel de Desempleo Nacional.....	50
Gráfico 7. Exportaciones de camarón en los últimos 5 años (Toneladas).....	51
Gráfico 8. Exportaciones de Camarón, Tasas de variación trimestral.....	52
Gráfico 9. Precio Promedio de la Toneladas de camarón exportado.....	53
Gráfico 10. Precio de Camarón, Tasas de variación trimestral.....	54
Gráfico 11. Demanda de la referencia 65.....	56
Gráfico 12. Demanda de la referencia 65.....	57
Gráfico 13. Demanda transformada de la referencia 65.....	58
Gráfico 14. Diagrama ACF y PACF de datos transformados.....	59
Gráfico 15. Residuos de ARIMA (1,1,1).....	61
Gráfico 16. Curva de ajuste de demanda para referencia 65,.....	62
Modelo ARIMA (1,1,1).....	62

Gráfico 17. Ajuste de curva de modelo Arima (0,1,1) .....	64
Gráfico 18. Ajuste de modelo de regresión múltiple (4 variables) .....	67
Gráfico 19. Ajuste de modelo de regresión múltiple (3 variables) .....	69
Gráfico 20. Ajuste de modelo de regresión múltiple (2 variables) .....	70
Gráfico 21. Ajuste de modelo de regresión (1 variable) .....	71
Gráfico 22. Regresión con errores ARIMA (0,1,1) con 4 variables regresoras.....	72
Gráfico 23. Regresión dinámica con errores ARIMA (0,1,1) dos variables predictoras....	74
Gráfico 24. Histograma de la distribución de los residuos .....	75
Gráfico 25. Función de autocorrelación de modelo de regresión estática .....	76
Gráfico 26. Diagrama de dispersión de residuos frente a cada variable predictora.....	77
Gráfico 27. Diagrama de dispersión de residuos frente a valores ajustados.....	78
Gráfico 28. Pronóstico de demanda del alimento referencia 65 .....	81

## LISTADO DE TABLAS

Tabla 1 Posibles modelos con su respectivo AIC .....	63
Tabla 2. Matriz de Correlación .....	65
Tabla 3. Modelos dinámicos con error ARIMA (0,1,1) .....	73
Tabla 4. Métricas de Precisión de ajuste y pronóstico de modelos propuestos .....	73
Tabla 5. Pronóstico de 5 periodos de prueba .....	79

# CAPÍTULO 1

## INTRODUCCIÓN

### 1.1 Antecedentes

Conocer lo que el futuro nos tiene preparado ha sido una de las inquietudes que ha pasado por la mente del ser humano por miles de años. Distintos métodos han sido utilizados para este fin, desde visiones, consultas a oráculos, observación de los astros, etc. Con el pasar de los siglos y el desarrollo de las ciencias abstractas se concibieron métodos analíticos para la toma de decisiones. Fue en la primera década del siglo XX en la que se publican los primeros trabajos empíricos relacionados a la teoría de la demanda en donde el modelo básico de regresión junto con el análisis de descomposición eran las técnicas utilizadas (Hernández Alonso & Zúñiga Rodríguez , 2013). En los años setenta la metodología ARIMA aparece como nuevo enfoque para modelar series de tiempo a través de modelos univariantes y posteriormente su generalización para la explicación de distintas variables de manera simultánea (modelos VARMA). Con el desarrollo en el campo de la computación se ha reforzado el tratamiento de los datos y la elaboración de modelos más complejos con aplicación en una serie de campos como en química, informática, física y en estadística. (Kuhn & Kjell, 2013) sostienen que, aunque en cada ciencia se le da un enfoque distinto el fin de un modelo de predicción es el mismo: predicciones precisas.

Una de las aplicaciones de modelamiento predictivo se ha realizado en la administración de demanda de productos / servicios en variadas industrias, tal como lo demuestra el trabajo realizado por Boada, A. & Vasconcelos, D. (2013) en la industria de

cosméticos, utilizaron un modelo de previsión de la demanda por cada SKU considerando factores mercadotécnicos como variables independientes. Empresas del sector público se han apoyado de modelos predictivos para determinar la demanda del servicio prestado, como ejemplo la propuesta de Kankal et al. (2010) modelaron el consumo de energía eléctrica en Turquía basado en variables socioeconómicas y demográficas utilizando el análisis de regresión y una red neuronal artificial. (Salazar Aguilar & Cabrera Ríos, 2007) exponen un modelo de redes neuronales para el pronóstico de demanda en la industria de telecomunicaciones, cuyos resultados comparados con los tradicionales métodos lineales fueron mejores (bajo el criterio de error cuadrado medio). Tal como las aplicaciones citadas, se pretende emular el empleo de los modelos predictivos en la industria acuícola que hoy por hoy es uno de los motores de la economía ecuatoriana. Teniendo en consideración el aforismo de que el futuro depende del pasado, es interesante conocer lo que ha sucedido en el mercado acuícola en estos últimos años para poder asimilar la problemática que se pretende investigar.

Corría el año 2000 y el virus mancha blanca ocasionaba la reducción de la producción de camarones en el país en un 30% y solamente 1200 fincas continuaron operativas (Bernabé Argandona, 2016). De acuerdo con el estudio realizado por la ESPAE Graduate School of Management (2015) el país se sobrepuso de este problema gracias a la investigación científica que contribuyó al desarrollo en lo concerniente al control de enfermedades, mejora de la calidad y mayor productividad. Otro punto importante que se destaca en dicho estudio es sobre el protagonismo que tuvo en la economía del país el sector camaronero en el año 2018 donde se constituyó como el segundo mayor generador de divisas de las exportaciones no petroleras al representar el 28% de dicho rubro. Ecuador está entre los tres



países latinoamericanos con mayor producción acuícola. Chile toma el liderazgo con la producción del salmón, mientras para Brasil su primordial actividad es la tilapia, en tanto que nuestro país lidera en la producción del camarón. En este informe también se destaca a los principales actores de la cadena de valor de la industria acuícola y pesquera, entre los cuales constan las empresas dedicadas a la fabricación de alimento balanceado, las mismas que han realizado inversiones para cubrir la demanda nacional con perspectivas de abastecer a mercados internacionales. El sector acuícola del Ecuador tiene al frente cinco desafíos que impactan en su evolución. El primero de ellos, es su marco legal e institucional, cuyos autores sugieren su actualización y reforma.

Como segundo punto consta la administración en el monitoreo, control, vigilancia e investigación. La transferencia de tecnología y el entrenamiento de productores de pequeña escala es el tercer punto por tener en consideración. El suministro de servicios públicos tales como energía, deben mejorar con miras al aumento de la competitividad. Consolidar mercados no tradicionales, ampliar mercados de gran valor y abrirse camino en nuevos mercados.

En el artículo de la revista Líderes (2014) destaca que el sector camaronero estimula a otras industrias tales como: laboratorios de larvas, fábricas de hielo, empacadoras y procesadoras de alimento, ésta última ha crecido proporcionalmente con las exportaciones de camarón. Cualquier amenaza a la industria del camarón afectará a toda su cadena de valor incluida a las procesadoras de alimento, por lo que los directivos de esta industria deben de contar con un monitoreo intenso de las variables que pueden causar estragos, tales como:

enfermedades, el precio del camarón que esta anidada con la oferta mundial de este crustáceo, políticas arancelarias, etc.

El reportaje realizado por Lizarzaburu (2018) destaca las proyecciones de la industria camaronera para el 2019 y 2020. En estos dos años se presentarían una baja de precio internacional del camarón, de acuerdo con el criterio de especialistas. La caída de los precios es debida al aumento de la producción mundial especialmente la de los países asiáticos. En el 2017 la oferta empezó a superar a la demanda, esto era señal de que los precios tiendan a bajar remarca Gorjan Nikolik director de Rabobank.

De acuerdo con datos de la CÁMARA DE ACUACULTURA en el 2015 en camaroneras ecuatorianas se obtenían 720.3 millones de libras y en el 2017 llegamos a 938,5 millones de libras. Para julio 2018 se llegó a 636,7 millones de libras y \$ 1867 millones (doble del presupuesto anual del Municipio de Guayaquil).

En este reportaje se señala que muchos camaroneros están cosechando a perdida o con márgenes de beneficio más bajo que la de los últimos 5 años debido a que elementos principales que integran la ecuación del costo (alimentación, trabajo y energía) no se han contraído.

## **1.2 Descripción del problema**

En el apartado precedente se subrayó la incertidumbre que advierten especialistas internacionales sobre los precios del camarón. Precios demasiados bajos podrían complicar la competitividad del camarón ecuatoriano en el mercado internacional lo cual generaría una desaceleración en la producción del crustáceo. Esta información no solo pone en jaque a los cultivadores de camarón sino también a las industrias dependientes de esta factoría como lo son las procesadoras de alimento balanceado puesto que sus directivos deberían reforzar su proceso de planificación para poder ajustar sus planes estratégicos a corto, mediano y largo plazo.

En la empresa en la cual se realiza el trabajo académico se han reportado retrasos en las entregas de pedidos realizados por clientes. Esto sin dudas ocasiona malestar en los consumidores del alimento ya que afecta su proceso de cultivo de larvas y se ven obligados a adquirir producto de otro ofertante. Estos sucesos ponen a tambalear a la alta directiva de la organización ya que atestiguan como la competencia gana mercado por el hecho de no contar con el producto a tiempo ni en la cantidad solicitada. Por los antecedentes se denota la deficiencia en la flexibilidad de la organización para cubrir la demanda de los clientes.

El desarrollar la capacidad de reaccionar a tiempo a las fluctuaciones de la demanda del mercado podría obtener dos beneficios para la organización: satisfacción de clientes y lograr una ventaja competitiva con relación a sus competidores. Entre los elementos indispensables para alcanzar flexibilidad en la oferta están las predicciones acerca de la

demanda y en el presente trabajo de investigación se propone un modelo estadístico para tratar la problemática.

### **1.3 Justificación**

El mercado del camarón y los negocios derivados de este sector, están marcados por el dinamismo económico, por tanto, las organizaciones deben poseer un alto nivel de flexibilidad para cubrir las necesidades en el momento oportuno y en el volumen requerido (Krajewki, Ritzman, & Malhotra, 2008). La problemática del desabastecimiento y poca flexibilidad motiva la revisión de los procesos estratégicos de la compañía y uno de los más destacables es la previsión de la demanda.

La previsión ha sido estudiada en varias industrias tradicionales, pero no abundan trabajos académicos relacionados al comercio de dietas balanceadas. El presente trabajo se justifica como un aporte a la aplicación teórica-práctica de modelos estadísticos en la industria acuícola pretendiendo así convertirse en una referencia bibliográfica para futuras investigaciones en este sector prominente de la economía ecuatoriana.

## **1.4 Objetivos**

### **1.4.1. Objetivos Generales**

Proponer un modelo estadístico para la previsión de demanda de alimento balanceado para camarones en una empresa comercializadora de la ciudad de Guayaquil.

### **1.4.2. Objetivos Específicos**

- Analizar los datos históricos del alimento balanceado con mayor demanda.
- Diseñar diferentes modelos estadísticos para el pronóstico.
- Evaluar los diferentes modelos y seleccionar aquel que presenta el mejor ajuste.
- Validar los pronósticos de la demanda con el modelo seleccionado.

## **1.5 Alcance**

El proyecto se realizará en una de las principales comercializadoras y productora de alimento para camarones en el Ecuador. El diseño del modelo estadístico predictivo se aplicará en el SKU con mayor demanda al mercado.

# CAPÍTULO 2

## MARCO TEÓRICO

### 2.1 Estado del Arte

#### 2.1.1 Satisfacción del cliente en un mundo con incertidumbre

El mundo de hoy se ha vuelto mucho más volátil, incierto, complejo y ambiguo en todo orden (político, económico, negocios, etc.), se utiliza el acrónimo VUCA para describir estas cuatro características del mundo actual (Pierre d' Huy & Jerome Lafon, 2018). En la primera década del siglo XXI hemos sido testigo de la rapidez con que la tecnología ha ido mejorando, lo que ha provocado modificaciones en el modelo de negocio de las organizaciones y en los hábitos de consumo, aunque esta última no siempre se deba a la aparición de tecnologías disruptivas. (Tundidor, y otros, 2019). Tres tendencias tecnológicas han beneficiado a los consumidores y permiten hacer posible lo imposible en todas las industrias: internet, aparatos móviles y la computación en la nube. (Eric Schmidt & Jonathan Rosenberg, 2016) . Gracias al internet tenemos apertura a información abundante, gratuita y perenne, los aparatos móviles han contribuido a la globalidad de la comunicación, la computación en la nube aumentó el poder de cómputo y almacenamiento infinito de información. La aparición de estas nuevas tecnologías también ha provocado un cambio en el comportamiento de los consumidores llegando a ser mucho más exigente.

Realizar previsiones ante escenarios con tales características es un reto que los líderes deben enfrentar, por lo que en la actualidad el horizonte temporal de las previsiones se considera más corto plazo (Moscoso & Lago, 2016). La determinación de la demanda (cantidad de producto que un mercado necesita para cubrir su necesidad) es una decisión que está sujeta al mundo VUCA (Pierre d' Huy & Jerome Lafon, 2018) y las organizaciones deben responder a dos cuestiones: la cantidad de producto que el cliente necesitará y en qué momento lo requerirá; teniendo en claro que ante un retraso en la entrega o la falta del producto implicaría una insatisfacción del cliente y hasta quizás pérdida del mismo (Krajewki, Ritzman, & Malhotra, 2008).

(Porter, 1991) plantea tres estrategias para competir: diferenciación, costo y enfoque. La diferenciación de un producto o servicio se puede lograr interviniendo en tres variables como lo son: la calidad, tiempo y flexibilidad (Krajewki, Ritzman, & Malhotra, 2008). Los clientes actuales son muchos más exigentes que en el antaño, ya no solo buscan calidad, sino que la flexibilidad (en volumen, variedad, personificación) y el tiempo (velocidad de entrega, desarrollo y entrega a tiempo) son uno de los factores que se ponderan. Para lograr el éxito comercial las empresas deben empeñarse en satisfacer todas las aristas de las necesidades de los clientes. (Eric Schmidt & Jonathan Rosenberg, 2016).

### **2.1.2 Insatisfacción por tiempo de entregas y flexibilidad en volumen.**

La pérdida de ventas es consecuencia de la insatisfacción del cliente por el producto o servicio. Entre los muchos motivos que a los clientes les induce a desertar en la adquisición

del producto/servicio que una organización oferta son las entregas atrasadas debido a falta de stock en el momento solicitado.

Los autores (Ballé, Jones, Chaize, & Fiume, 2018) en su libro Estrategia Lean, establecen que el desajuste entre la producción y las ventas impide cumplir los niveles de entrega puntual y productividad. Añaden también que una de las gigantes mundial como lo es Toyota identificó cuatro causas principales por las que se presentaban entregas deficiente y elevados inventarios: grandes lotes, flujos complicados, desajuste entre el ritmo de trabajo y de ventas, transporte y logística deficiente.

Para evitar entregas atrasadas y perdida de ventas por falta de inventario es necesario que se cuente con un pronóstico de la demanda de los clientes para así tener sincronización entre la cadena de abastecimiento y la producción.

### **2.1.3 Pronóstico de la demanda y su relación con la gestión de inventarios.**

De acuerdo con el libro Principios y Máximas de la cadena de abastecimiento (Mora, 2018) , los inventarios protegen a las compañías de las variaciones de la demanda y así pueden brindar un nivel de servicio adecuado con un costo equilibrado.



Se han realizado varios trabajos de investigación de gestión de existencias las mismas que dependen de un modelo preciso de pronóstico de la demanda para poder alcanzar el objetivo trazado de brindar un servicio adecuado a los clientes. Así tenemos el trabajo de (Castro Romero, 2015) donde recalca la necesidad de realizar un pronóstico de la demanda previo a determinar el tamaño óptimo de compra en una empresa importadora de juguetes en Perú.

Por otro lado, en el proyecto presentado por los autores (Aguirre Lasprilla, Ardila Rueda, Figueroa, & Romero Rodríguez, 2015) determinan una política de inventario en un Hospitales de Barranquilla- Colombia. Su trabajo se compone de cuatro fases: clasificación de ítems, entre las que se incluye el pronóstico de la demanda, parametrización de política de inventario y evaluación de las políticas. Una vez más podemos identificar de la relación entre ambos temas.

Otra investigación similar es la que desarrollaron (Betancurt Aristizabal, Fandiño Castillo, Gutierrez Ortegón, & Ospina Tenorio, 2018) en una distribuidora de instrumentación industrial Cali-Colombia. Aquí los autores definieron un sistema de control de inventario en la que también tuvieron que pronosticar la demanda para cada referencia seleccionada para poder determinar políticas de stocks.

(Cortes Gonzalez, 2019), en su trabajo de investigación comenta que la variación en el pronóstico de la demanda y los tiempos prolongados de reposición son dos puntos que inciden en el problema de la determinación del stock de seguridad. Aquí el autor también

realiza un análisis del comportamiento histórico de la demanda de cada familia de producto planteando los modelos de suavizado exponencial simple y el modelo Gompertz.

Como síntesis de las investigaciones presentadas podemos concluir que el punto de partida para una eficiente gestión de inventarios es el pronóstico muy preciso de la demanda. Con dicho dato es posible determinar políticas de inventarios, los niveles de stock de seguridad y punto de pedido de ítems.

#### **2.1.4 Aplicaciones de modelos de previsión de demanda.**

Existe una extensa investigación concerniente a la aplicación de modelos de pronóstico en distintas industrias. Son muchas las técnicas las que se han aplicado desde los clásicos modelos de regresión, series de tiempo y modelos contemporáneos como las redes neuronales y machine learning. En las siguientes líneas se hace referencia a investigaciones que han empleado estos métodos de previsión en distintos sectores comerciales e industriales.

(Yan, Chen Zheng , Zhang , & Xin, 2019) en su trabajo pronostican la demanda de bicicletas compartidas utilizando un modelo de redes neuronales del tipo memoria a corto y largo plazo (por sus siglas en inglés LSTM). En esta investigación se toman los datos de las estaciones de dos comunidades (New York y Jersey City). Como primer paso se realiza un procesamiento y análisis de los datos, en ella observan que la cantidad de bicicletas

alquiladas por hora varía de una ciudad a otra y que existen pocos viajes relacionados entre ambas metrópolis lo cual implica dos problemas: escasez de datos y secuencia de tiempos no significativa. En el modelo se utilizaron la hora de inicio y de finalización, la estación de inicio y de final de cada viaje, además se consideraron 3 indicadores meteorológicos como la temperatura, intensidad de precipitación y velocidad del viento. Se generan dos modelos uno LSTM y DNN (por sus siglas redes neuronales profundas). Los resultados del modelo LSTM fueron comparados con los datos reales y se obtuvo un RSME (raíz cuadrada del error cuadrático medio) 2,7069 para los datos de prueba y de los datos de entrenamiento igual a 3,6752. En la comparación de los datos de prueba del modelo LSTM es mejor que el DNN.

Kumar et al (2017) plantean dos modelos ARIMA univariante para el pronóstico de la producción de peces marino en Odisha-India. El primero incluye intervenciones (evento externo, mecanización de artesanía de pesca) y el segundo con la transformación de los datos. Kumar y su equipo contaban con una base de datos de 1985 a 2012, la misma que dividieron en dos, así que establecieron el periodo de pre-intervención (1985-2007) y periodo post-intervención (2008-2012). El criterio de información Akaike (AIC) y SBC (criterio de selección bayesiano) fueron utilizados para la elección del mejor modelo. Basados en dichos criterios el modelo con transformación fue escogido para realizar el pronóstico de los años 2013,2014 y 2015.

Por su parte Guerrero, M. & Medina, S. (2016) construye el pronóstico de los expedientes acumulados en la Comisión de Derechos Humanos del Distrito Federal de México (CDHDF). Ya que el comportamiento de las series de cada una de las visitadurías

es distinto procedieron con el modelamiento desagregado, es decir ajustaron modelos de series de tiempo para cada una de las visitadurías generales del CDHDF probando con distintas metodologías tales como ARIMA, RAW DRIFT y SARIMA.

Ramos et al (2015) realizaron un pronóstico de ventas de una tienda minorista de cinco categorías de calzado para mujeres. Esta investigación coteja el rendimiento de pronósticos de los modelos de espaciado de estado (ETS) y ARIMA (modelos promedio móvil autorregresivos integrados). En primera instancia los investigadores hallaron estacionalidad en la demanda de cada tipo de calzado, las botas y botines tiene su pico en el invierno, en tanto que las categorías pisos y sandalias son comercializadas en el verano, por su parte la categoría zapatos se vende todo el año. El equipo investigador diseña un modelo ETS y ARIMA para cada categoría de calzado los mismos que son evaluados con base a los resultados de las medidas de error. En este trabajo se concluyó que los ETS y ARIMA arrojan pronósticos muy similares.

Boada, A. & Vasconcelos, D. (2013) plantean un modelo de previsión de la demanda por cada SKU de productos de venta por catálogo, utilizando factores mercadotécnicos como variables independientes. La demanda la estiman con un modelo semilogarítmico aplicando regresión lineal múltiple en cada familia de productos. Para la inclusión de variables al modelo aplicaron el método de eliminación descendente.

Los investigadores (Hai, Hao, & Ping, 2011) aplicaron un modelo de control predictivo (MCP) en planificación de la cadena de suministro. Los investigadores para

determinar el nivel de inventario utilizaron una función la misma que dependía del límite de confianza superior de la CDS (demanda de cliente satisfactoria; la misma que fue calculada previamente) en un determinado momento y del pronóstico de la demanda. Para el pronóstico utilizaron regresión múltiple.

Kankal et al. (2010) modelaron el consumo de energía eléctrica en Turquía basado en variables socioeconómicas y demográficas utilizando el análisis de regresión y una red neuronal artificial. En el enfoque de regresión se generaron 4 modelos lineales múltiple y otros 4 modelos de potencia. Evaluados por los  $R^2$  y el máximo error absoluto se encontró que los modelos de potencia fueron mejores que los de regresión múltiple lineal. Se construyó también una red neuronal artificial la cual tuvo un error mucho mejor que los modelos de regresión.

Dado que el consumo de alimento balanceado está estrechamente relacionado con los niveles de exportación de camarones, es importante conocer los factores que impactan en tal sector de la economía ya que sus efectos decantan a la industria del balanceado. Molina S. (2009) modela la demanda del camarón tomando variables independientes al precio promedio importado por USA, ingreso promedio de los habitantes de Estados Unidos y el precio de la tilapia (producto sustituto).

## 2.2 Marco Teórico

### 2.2.1 Modelo

Un modelo es la representación de un sistema/situación de la vida real utilizado para proporcionar un resumen de baja dimensión de un conjunto de datos (Wickham & Golemund , 2017) .

#### 2.2.1.1 Modelo Estadístico

Un modelo estadístico es una extensión de un modelo matemático al cual se le añade una variable aleatoria (Téllez Piñerez & Morales Rivera , 2016). Un modelo está compuesto de dos partes: familia y ajuste. La familia expresa un patrón preciso pero genérico, mientras que el ajuste encuentra el modelo más cercano de una familia de modelos (Wickham & Golemund , 2017).

Un modelo estadístico tiene la forma siguiente:

$$Y = f(X) + \epsilon$$

Donde:

$f(X)$ : función que explica la relación entre la variable dependiente e independientes.

$\epsilon$  : termino de error o aleatorio que explica las fluctuaciones de la variable de respuesta que no puede ser explicadas por las variables de entrada.

En el análisis de datos se cuenta con conjunto de enfoques para estimar  $f$  . Los autores (James, Witten, Hastie, & Tibshirani, 2013) sostienen que las razones principales para estimar  $f$  es la predicción y la inferencia. La inferencia se centra en entender la relación entre las variables independientes y la dependiente. De acuerdo con (Kuhn & Kjell, 2013) el objetivo del modelado predictivo es desarrollar una herramienta o modelo matemático para generar una predicción precisa. En el modelado predictivo se busca estimar  $Y$  (variable de respuesta) utilizando:

$$\hat{Y} = \hat{f}(X)$$

Donde:

$\hat{Y}$ : predicción resultante para  $Y$

$\hat{f}$ : estimación de  $f$

$\hat{f}$  no es una estimación perfecta de  $f$  por tanto introduce un error. En un modelo predictivo se presentan dos errores: reducible e irreducible. El primer tipo de error podemos minimizarlo al mejorar la precisión de  $f$  por medio de alguna técnica de modelado. En el caso que se logre estimar perfectamente  $f$  , de manera que  $\hat{Y} = f(X)$  , aun así, el modelo tendría algún error en la predicción debido a que  $Y$  depende de  $\epsilon$  , generado por las variables que no fueron medidas para la predicción, este error es irreducible (James, Witten, Hastie, & Tibshirani, 2013). La diferencia al cuadrado entre el valor real y el pronosticado se denomina error cuadrático medio.

$$E(Y - \hat{Y})^2 = E[f(x) + \epsilon - \hat{f}(x)]^2 = [f(x) - \hat{f}(x)]^2 + \text{Var}(\epsilon)$$

## 2.2.2 Métodos para el modelamiento predictivo

Para poder realizar una predicción, inferencia o una combinación de ambas podemos optar por dos tipos de modelos: paramétricos y no paramétricos.

### 2.2.2.1 Modelos paramétricos

Es estos modelos se cuenta con un enfoque de dos fases:

Hacer una suposición sobre la forma funcional de  $f$ , ejemplo: función lineal

$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

Seguir un procedimiento que utilice datos de entrenamiento para ajustar o entrenar el modelo. En el caso del modelo lineal los parámetros  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  son estimados comúnmente con el enfoque de mínimos cuadrados. Asumir una forma determinada de  $f$  simplifica su estimación debido a que el problema se reduce a la determinación de un conjunto de parámetros. Pero si el modelo que se selecciona está muy lejos de la función verdadera la predicción será pobre, por lo que podemos optar por modelos más flexibles,



aunque estos requieren más parámetros lo que podría generar un sobreajuste en los datos (James, Witten, Hastie, & Tibshirani, 2013).

#### **2.2.2.2. Modelos no paramétricos**

En los modelos no paramétricos en contraste con los paramétricos, no se hacen suposiciones explícitas sobre la forma funcional de  $f$ . Se busca una estimación de  $f$  que se aproxime más a los puntos de los datos. Al evitar el supuesto de una forma funcional el modelo tiene potencial de ajustarse con precisión a un rango más amplio de formas posibles de  $f$ . Estos modelos requieren un mayor número de datos en comparación con los paramétricos.

#### **2.2.3 Tipos de Pronósticos**

Los pronósticos se pueden clasificar en cuatro tipos: cualitativo, análisis de serie de tiempo, relaciones causales y simulación (Chase, Jacobs, & Aquilano, 2009). Para (Hyndman & Athanasopoulos, 2018) los modelos de predicción se clasifican en cuatro: cualitativo, modelos explicativos, modelos de serie de tiempo y modelo Mixto. El enfoque cualitativo utiliza métodos estructurados para realizar pronósticos. Los modelos explicativos apuntan a explicar las causas de la variación de la variable dependiente debido a otras variables a las cuales se denominan variables predictoras. Los modelos de serie de tiempo predicen el futuro en valores pasados de una variable, pero no de variables externas. Los modelos mixtos adoptan características de los dos modelos explicados anteriormente.

Los autores (Hyndnan & Athanasopoulos, 2018) sostienen que entre las razones por las que se suele seleccionar un modelo de serie de tiempo en lugar de uno explicativo o mixto están:

- Dificultad de medir las relaciones que gobiernan el comportamiento del sistema a modelar.
- Necesidad de conocer los valores futuros de las variables predictoras o independientes, lo cual es sumamente difícil.
- La principal intención es pronosticar lo que sucederá, no se está interesado el por qué ocurre. En definitiva, el tipo de modelo que se utiliza está en función de los recursos y datos con los que se dispongan y en la forma que se utilizará el pronóstico.

#### **2.2.4 Serie de tiempo**

Se denomina serie de tiempo al conjunto de observaciones que ocurren secuencialmente en un tiempo específico (Brockwell & Davis, 2002). A diferencia de otros métodos estadísticos en el análisis de series de tiempo se tiene en cuenta el orden de los datos y la dependencia entre ellas. El análisis de las series puede ser univariante o multivariante, en el primer caso se analiza una serie temporal dependiendo en función de su pasado; mientras que para el caso multivariante se estudian a la vez varias series.

Los objetivos más frecuentes del análisis de las series temporales son la descripción de las características y la predicción de valores futuros. En un modelo univariante se formula

solamente con sus valores pasados y/o su posición con respecto al tiempo. Son modelos ingenuos debido a que las predicciones se obtienen de la extrapolación de los datos observados hasta el momento T. Al existir dependencia de los datos sucesivos los valores futuros pueden ser predichos a partir de las observaciones pasadas. Una serie es definida como determinista cuando la predicción es exacta, pero la realidad es que la mayoría de las series son estocásticas en la cual el futuro es definido parcialmente por sus valores pasados por lo cual las predicciones exactas son imposibles y deben ser sustituidas por el enfoque de que los valores futuros siguen una distribución de probabilidad condicionada a los valores del pasado. El análisis de series temporales utiliza la teoría de procesos estocásticos, por lo que en las siguientes líneas se expondrán todo lo relacionado a este tipo de procesos.

### **2.2.5 Proceso estocástico**

Es una familia de variables aleatorias que en general están relacionadas entre sí siguiendo una ley de distribución conjunta, se denota como:

$$\dots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}, \dots$$

Se pueden considerar secuencias de variables aleatorias ordenadas en el espacio, tiempo. Así en el contexto estadístico de los procesos estocástico una serie temporal  $Y_1, Y_2, \dots, Y_T$ , se puede considerar como una realización muestral de un proceso estocástico que se observa para un número finito de periodos  $t = 1, 2, \dots, T$ .

### 2.2.5.1 Características de un proceso estocástico

Un proceso estocástico queda caracterizado por su función de distribución o por sus momentos. La función de distribución de proceso estocástico incluye todas las funciones de distribución para cualquier subconjunto finito de variables aleatorias del proceso.

$$F(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}), \forall (t_1, t_2, \dots, t_n) \text{ con } n \text{ finito.}$$

Dada la complejidad de determinar la función de distribución del proceso estocástico se recurre a caracterizarlo a través de los dos primeros momentos. El primer momento viene dado por el conjunto de medias de todas las variables aleatorias del proceso.

$$E(Y_t) = \mu_t$$

El segundo momento queda determinado por el conjunto de varianzas de todas las variables aleatorias del proceso y por las covarianzas entre todo par de variables.

$$V(Y_t) = E[Y_t - \mu_t]^2 = \sigma_t^2 < \infty \quad t = 0, \pm 1, \pm 2, \dots,$$

$$\text{cov}(Y_t, Y_s) = E[Y_t - \mu_t][Y_s - \mu_s] = \gamma_{t,s} \quad \forall t, s (t \neq s)$$

Si el proceso tiene una distribución normal y se conoce los dos primeros momentos (medias, varianzas y covarianzas), el proceso queda caracterizado y definida su función de distribución. Ya hemos revisado el concepto de proceso de estocástico y las características que tiene, nuestro siguiente paso es exponer los tipos de procesos estocásticos. Esta revisión tiene como fin saber identificar el proceso estocástico que ha sido capaz de generar la serie bajo estudio, para caracterizar su comportamiento y predecir su futuro.

Los procesos estocásticos generadores de series de tiempo deben tener estabilidad para poder predecir, pero si por el contrario la serie presenta en cada momento comportamientos distintos e inestables no se podrá utilizar para pronosticar.

### 2.2.5.2 Proceso Estocástico estacionario

Las condiciones que deben tener los procesos estocásticos para que sean estables y así poder predecir se le llama estacionariedad. Un proceso estocástico es identificado con estacionariedad estricta si su función de distribución de cualquier conjunto finito de  $n$  variables aleatorias del proceso no se altera si se desplaza  $k$  periodos en el tiempo.

$$F[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}] = F[Y_{t_1+k}, Y_{t_2+k}, \dots, Y_{t_n+n}] \quad \forall (t_1, t_2, \dots, t_n) \text{ y } k$$

Un proceso estocástico es estacionario en covarianza si y solo si:

Las variables del proceso tienen misma media.

$$E(Y_t) = \mu_t < \infty, \forall t$$

Todas las variables del proceso tienen la misma varianza

$$V(Y_t) = E[Y_t - \mu]^2 = \sigma_Y^2 < \infty, \forall t$$

Las auto covarianzas sólo dependen del número de periodos de separación entre las variables y no del tiempo.

$$\text{cov}(Y_t, Y_s) = E[Y_t - \mu_t][Y_s - \mu_s] = \gamma_{|t-s|} = \gamma_k < \infty, \forall k$$

### **2.2.5.3 Proceso Estocástico no estacionario**

La mayoría de las series económicas no presentan comportamientos estacionarios por que suelen cambiar de nivel en el tiempo o porque la varianza no es constante, la solución es transformar la serie mediante algún método que estabilice la media y varianza. Las transformaciones como los algoritmos pueden ayudar a estabilizar la varianza de una serie temporal. Para estabilizar la media de la serie temporal se utiliza la diferenciación de observaciones consecutivas para eliminar los cambios en el nivel de una serie temporal y por lo tanto reduce la tendencia y la estacionalidad. En ciertos casos se debe realizar hasta una segunda diferenciación para lograr estabilidad en la media. Las diferenciaciones también se las realiza por estaciones, la cual es la diferencia entre una observación y la observación previa de la misma estación. Es necesario que las diferencias sean interpretables.

### **2.2.5.4 Raíz Unitaria**

El definir cuantas diferenciaciones requiere una serie para alcanzar la estacionariedad es un tema que cae en el plano de lo subjetivo, para evitarlo se realiza una prueba de raíz unitaria las cuales son hipótesis estadísticas de estacionariedad (Hyndnan & Athanasopoulos, 2018). En este trabajo se utiliza la prueba Kwiatkowski-Philips-Schmidt-Shin (KPSS). En esta prueba la hipótesis nula que se plantea es que los datos son estacionarios y la hipótesis alternativa es que los datos no son estacionarios.

La prueba funciona de la siguiente manera: se toman los residuos  $e_t$   $t = 1, 2, \dots, T$  al aplicar la regresión a la serie  $Y_t$  contra una constante y una tendencia lineal, es decir:

$$Y_t = \alpha + \beta t + \mu_t$$

Se estima la varianza de la regresión  $\widehat{\sigma^2}$  (igual a la suma residual entre el número de observaciones). Se define la suma parcia de los residuos como:

$$SP_t = \sum_{i=1}^T e_i \quad t=1,2, \dots, T$$

Quedando la prueba de autocorrelación LM expresado de la manera siguiente:

$$LM = \sum_{t=1}^T \frac{SP_t^2}{\sigma^2}$$

Pero esta expresión es validada para los casos en los que  $\mu_t$  se distribuyen de manera normal, media cero y cierta varianza; pero debido a que las series temporales suelen tener una correlación fuerte el supuesto no aplica. Por tal motivo se utiliza la estimación a largo plazo de varianza  $\sigma_{it}^2$ , quedando el estadístico de prueba KPSS en la siguiente expresión:

$$KPSS = T^{-2} \sum_{t=1}^T \frac{SP_t^2}{\sigma_{it}^2}$$

### 2.2.6 Modelos Lineales Estacionarios

El modelo de una serie temporal univariante se descompone en dos partes: sistemática y aleatoria, la primera de ella es predecible gracias a la información que se tiene. Al construir un modelo estadístico la parte más complicada es plantear la parte sistemática. Para una serie temporal de media cero, el valor de Y en el momento t depende de su pasado, un modelo teórico que lo describe puede ser expresado como:

$$Y_t = f(Y_{t-1} + Y_{t-2} + \dots + Y_{t-p}) + a_t \quad t=1,2,\dots$$

Para procesos estacionarios con distribución normal y media cero,  $Y_t$  se puede expresar como combinación lineal de los valores pasados infinitos de Y más un ruido blanco.

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \dots + a_t \quad \forall(t_1, t_2, \dots, t_n) \quad (1)$$

Este modelo se aplica si el proceso cumple con las siguientes condiciones:

- El proceso no es anticipante, esto quiere decir que el presente no dependa del futuro.
- El proceso debe ser invertible, es decir, que el presente dependa de forma convergente de su propio pasado, lo que significa que la influencia de  $Y_{t-k}$  en Y ha de ir disminuyendo conforme nos alejamos del presente. Esta condición se cumple si los parámetros del modelo general cumplen con la siguiente restricción:

$$\sum_{i=1}^{\infty} \pi_i^2 < \infty$$



Escribiendo el modelo 1 en términos del operador de retardo (L):

$$Y_t = (\pi_1 L + \pi_2 L^2 + \dots) Y_t + a_t$$

$$a_t = (1 - \pi_1 L - \pi_2 L^2 + \dots) Y_t$$

$$a_t = \prod_{\infty} (L) Y_t$$

Estos modelos son para series infinitas, pero en la práctica se trabaja con series finitas por lo que se deben simplificar estos modelos de manera que tengan un número finito de parámetros. De acuerdo con la teoría de polinomios se puede aproximar un polinomio de orden infinito por medio de un cociente de polinomios finitos:

$$\prod_{\infty} (L) \cong \frac{\varphi_p(L)}{\theta_q(L)}$$

Siendo  $\varphi_p(L)$  y  $\theta_q(L)$  operadores de retardo de orden p y q respectivamente:

$$\varphi_p(L) = (1 - \varphi_1 L - \varphi_2 L^2 + \dots + \varphi_p L^p)$$

$$\theta_q(L) = (1 - \theta_1 L - \theta_2 L^2 + \dots + \theta_q L^q)$$

Sustituyendo en la ecuación 1:

$$\prod_{\infty} (L) Y_t \cong \frac{\varphi_p(L)}{\theta_q(L)} Y_t = a_t \rightarrow \varphi_p(L) Y_t = \theta_q(L) a_t$$

Consecuentemente, el modelo general admite tres representaciones, todas igualmente provechosas bajo los supuestos remarcados:

Representación puramente autorregresiva AR ( $\infty$ ): el valor presente de la variable se representa en función de su propio pasado más un ruido blanco.

Puramente medias móviles MA ( $\infty$ ): el valor presente de la variable se representa en función del ruido blanco presente y pasado.

Representación finita: el valor de  $Y_t$  depende del pasado de  $Y$  hasta el momento  $t-p$  (parte autorregresiva), del ruido blanco y su pasado hasta el momento  $t-q$ .

Este modelo se denomina Autorregresivo de medias móviles de orden  $(p,q)$  y se denota como ARMA  $(p,q)$

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t$$

### 2.2.6.1 Modelo de autorregresivo AR (p)

El modelo ARMA  $(p, q)$  tiene dos casos particulares, el primero de ellos es cuando el orden del polinomio de medias móviles es de orden cero y queda representado por la parte autorregresiva AR(p)

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + a_t \quad a_t \sim \text{RB}(0, \sigma^2)$$

En términos de retardo:

$$a_t = (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) Y_t \quad \rightarrow \quad \varphi_p(L) Y_t = a_t$$

Siendo:

$\varphi_p(L)$  polinomio autorregresivo

$(\varphi_1, \varphi_2, \dots, \varphi_p)$  es el vector de parámetros autorregresivos

$a_t$ , es denominado ruido blanco, una variable aleatoria con media cero y varianza 1.

En primera instancia se debe comprobar que sí el proceso AR(p) cumple con la condición de estacionariedad para cualquier valor de los parámetros.

Se hace uso del teorema siguiente: Un proceso autorregresivo AR(p) es estacionario si y solo si el módulo de las raíces del polinomio autorregresivo  $\varphi_p(L)$  esta afuera de la unidad.

Las características del modelo estacionario AR(p) son:

- Media (para series con media distinta de cero)

$$E(Y_t) = \frac{\delta}{1 - \varphi_1 - \varphi_2 - \dots - \varphi_p}$$

- La función de autocorrelación  $\rho_k$ ,  $k = 1, 2$ , de un proceso AR(p) decrece exponencialmente hacia cero sin truncarse.

### 2.2.6.2 Proceso media móviles MA (q)

El segundo caso, el modelo se representa sólo con la parte medias móviles debido a que el orden del polinomio autorregresivo es cero:

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

En termino de retardo queda expresado como:

$$Y_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) a_t \rightarrow \theta_q(L) a_t = Y_t$$

Donde:

$\theta_q(L)$  : polinomio de medias móviles

$(\theta_1, \theta_2, \dots, \theta_q)$  : vector de parámetros de medias móviles

Debido a que el número de parámetros de un MA(q) es finito este modelo es siempre estacionario para cualquier valor de sus parámetros. Este modelo también es no anticipante ya que el presente esta desligado del futuro. En cuanto a la invertibilidad está se comprobará si y solo si el módulo de las raíces del polinomio de medias móviles  $\theta_q(L)$  está afuera del círculo de la unidad. Las características del modelo MA (q) son: media constante y cero, varianza constante y finita además su función de auto covarianza está truncada a partir del retardo q, es decir:

$$\rho_k = \begin{cases} \rho_k \neq 0 & k = 1, 2, \dots, q \\ \rho_k = 0 & k > q \end{cases}$$

Este modelo es una alternativa al modelo de autorregresión, explica la evolución de la variable temporal mediante una combinación adecuada de los valores del ruido blanco en instantes precedentes.

### **2.2.6.3 Procesos Autorregresivos de Medias Móviles**

Estos modelos determinan  $Y_t$  en función de su pasado hasta el retardo p, del ruido blanco y el pasado del ruido blanco hasta el retardo q:

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t \quad a_t \sim RB(0, \sigma^2)$$

Expresando el modelo en función de términos de retardo:

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) Y_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) a_t$$

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

Donde

$\varphi_p(L)$  es el polinomio de autorregresión

$\theta_q(L)$  es el polinomio de medias móviles

Las condiciones de estacionariedad del modelo ARMA (p, q) vienen definidas por la parte autorregresivas, dado que la parte de las medias móviles siempre es estacionaria. Por otro lado, las condiciones invertibilidad del modelo ARMA (p, q) vienen impuestas por la parte de las medias móviles debido a que la parte autorregresiva finita siempre es invertible.

El modelo ARMA (p, q) finito tienen media cero, varianza constante y finita además una función de autocovarianzas infinita. Su función de autocorrelación es infinita decreciendo rápidamente hacia cero sin truncarse.

#### 2.2.6.4 Modelo Arima

Los modelos descritos anteriormente deben ser aplicados a series temporales que sean estacionarias, lo que significa que tengan media y varianza constante. Los datos de las series que no cumplen con la hipótesis de estacionariedad pueden ser transformados adecuadamente para poder tratarlas como estacionarias (Hernández Alonso & Zúñiga Rodríguez , 2013)

Los modelos ARIMA (Promedio móvil integrado autorregresivo, por sus siglas en inglés) pretenden describir las autocorrelaciones en los datos de la serie. El modelo ARIMA no estacional se puede escribir:

$$Y'_t = \delta + \varphi_1 Y'_{t-1} + \dots + \varphi_p Y'_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t$$

$$Y'_t - \varphi_1 Y'_{t-1} - \dots - \varphi_p Y'_{t-p} = \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t$$

$$(1 - \varphi_1 L - \dots - \varphi_p L^p) (1-L)^d Y'_t = (1 + \theta_1 L + \dots + \theta_q L^q) a_t$$

Este modelo es conocido como ARIMA (p,d,q)

$Y'_t$ : es la serie diferenciada

p: orden autorregresivo del proceso regular

d: grado de primera diferenciación involucrada

q: orden de promedio móvil del proceso regular

L: Función de retardo

$a_t$ : ruido blanco

Las diferencias entre observaciones consecutivas se conocen como diferenciación. La estimación de los parámetros se la realiza utilizando el método de la máxima verosimilitud, que encuentra los valores de los parámetros que maximizan la probabilidad de obtener los datos que hemos observado. El Criterio de información de Akaike (AIC, por siglas en inglés) es útil para determinar el orden de un modelo ARIMA

Un modelo ARIMA también puede modelar datos estacionales, añadiéndole términos de esta naturaleza.

ARIMA (p, d, q) (P, D, Q)<sub>s</sub>

P: orden autorregresivo del proceso estacional

Q: orden promedio móvil del proceso estacional

S: amplitud anual de la serie

En los modelos de series temporales se utiliza estadísticos para identificar el modelo concreto. Uno de ellos es el coeficiente de correlación simple o parcial de orden k, que no es sino el coeficiente de correlación entre la variable  $Y_t$  y su retardo  $Y_{t-k}$ . Al conjunto de estos coeficientes, para distintos desfases temporales de la variable se le denomina función

de autocorrelación (ACF, por sus siglas en inglés) (Hernández Alonso & Zúñiga Rodríguez, 2013). La función de autocorrelación está definida como:

$$\rho = \frac{\gamma_X(k)}{\gamma_X(0)} = \text{Cor}(X_{t+k}, X_t)$$

Donde

$\gamma_X(k)$ : autocovarianza en el retraso  $k$ .

$\gamma_X(0)$ : autocovarianza sin retraso.

La función de autocorrelación parcial (PACF, por siglas en inglés) mide la correlación entre observaciones de una serie de tiempo que se encuentran distantes por  $k$  unidades de tiempo luego de ajustarse para la presencia de los demás términos de desfases más cortos. Los gráficos ACF y PACF de los datos nos ayuda identificar si pueden seguir un modelo ARIMA.

### 2.2.7 Regresión lineal

La regresión lineal es uno de los métodos más utilizados para realizar un análisis de serie de tiempo. El modelo lineal guarda la forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon_i$$

El método más común para la estimación de los parámetros es el de los mínimos cuadrados, aunque tenemos disponibles otras técnicas como la estimación por mínimos



cuadrados generalizados, máxima verosimilitud y mejor estimados lineal insesgado (Téllez Piñerez & Morales Rivera , 2016).

### Estimación de mínimos cuadrados

El principio de mínimos cuadrados proporciona un criterio para estimar los parámetros  $\beta_0, \beta_1, \dots, \beta_k$ . Dicho criterio consiste en minimizar la suma de los errores cuadrados

$$L = \sum_{t=1}^T \epsilon^2 = \sum_{t=1}^T (y_t - \beta_0 - \beta_1 X_{1,t} - \beta_2 X_{2,t} \dots - \beta_k X_{k,t})^2$$

Se minimiza L con respecto a  $\beta_0, \beta_1, \dots, \beta_k$ . Luego de derivar la expresión con relación a todos y cada uno de ellos se obtiene el sistema de ecuaciones normales de mínimos cuadrado:

$$\sum Y_t = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_{1t} + \hat{\beta}_2 \sum x_{2t} + \dots + \hat{\beta}_k \sum x_{kt}$$

$$\sum Y_t X_{1t} = \hat{\beta}_0 \sum X_{1t} + \hat{\beta}_1 \sum X_{1t}^2 + \hat{\beta}_2 \sum x_{1t} x_{2t} + \dots + \hat{\beta}_k \sum x_{kt} x_{1t}$$

.....  
 .....  
 .....

$$\sum Y_t X_{kt} = \hat{\beta}_0 \sum X_{kt} + \hat{\beta}_1 \sum X_{1t} X_{kt} + \hat{\beta}_2 \sum x_{2t} x_{kt} + \dots + \hat{\beta}_k \sum X_{kt}^2$$

Resolviendo este sistema de ecuaciones se determina los  $k$  estimadores de los parámetros de la ecuación y podemos expresar la ecuación estimada de los términos

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p + \epsilon_i$$

Los estimadores de MCO (mínimo cuadrado ordinarios) para series de tiempo son insesgados, de cumplir con todos los supuestos los estimadores MCO no podrán ser superados por otro estimador lineal insesgado. Los errores estándar de MCO, los estadísticos  $t$  y  $F$  pueden ser utilizados para la inferencia. Supuestos del modelo lineal para regresiones de series de tiempo

### Lineal en los parámetros

El proceso estocástico  $(X_{1t}, X_{2t}, \dots, X_{kt}, Y_t)$   $t: 1, 2, \dots, n$  sigue el modelo lineal

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{tk} + \epsilon_t$$

Donde

$\epsilon_t$ : secuencia de errores  $t = 1, 2, \dots, n$  es la secuencia de errores

$N$ : número de periodos

**No colinealidad:** en la muestra no hay variables independientes constantes ni hay combinación lineal perfecta de otras.

**Media condicional cero:** el valor esperado del error  $\epsilon_t$  dada las variables explicativas para todos los periodos, es cero.  $E(\mathbf{u}_t | \mathbf{X}) = 0$ ,  $t = 1, 2, \dots, n$

Este supuesto implica que el error en cada periodo t no está correlacionado con ninguna variable explicativa en ningún periodo.

**Homocedasticidad:** La varianza de  $\epsilon_t$  es la misma en todo t.

$$\text{Var}(\mathbf{u}_t|\mathbf{X}) = \text{Var}(\mathbf{u}_t) = \sigma^2, t=1,2,\dots,n$$

**No hay correlación serial:** Los errores de dos periodos consecutivos no están correlacionados.  $\text{Corr}(\mathbf{u}_t, \mathbf{u}_s|\mathbf{X}) = 0$  para todo  $t \neq s$

Para validar la no correlación se utilizan los gráficos ACF y la prueba Breusch – Godfrey.

Se trata de un contraste de dependencia lineal conjunta en la regresión, se calcula:

$$F = \frac{R^2(T - k - p)}{(1 - R^2)(p)}$$

Siendo

$R^2$  : coeficiente de determinación

P: orden autorregresivo

K: número de variables

T: número de periodos

**Normalidad:** Los errores son independientes X y son independientes e idénticamente distribuidos como normal  $(0, \sigma^2)$

### 2.2.7.1 Modelos de Regresión Dinámica

En los modelos ARIMA se permite la inclusión de información de observaciones pasadas de una serie, pero no admite la inclusión de otra información que sea relevante. Por el contrario, los modelos de regresión permiten la inclusión de información relevante de variables predictoras, pero no consiente la dinámica de series temporales sutiles que se puedan tratar con modelos ARIMA.

En los modelos estáticos se considera que los errores no se encuentran correlacionados. En los modelos dinámicos se permite que los errores de regresión tengan correlación. En este modelo se supone que los errores siguen un modelo ARIMA. Para un modelo con errores ARIMA (1,1,1)

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \epsilon_t$$

$$(1 - \varphi_p L) (1 - L) \epsilon_t = (1 + \theta_q L) a_t$$

Donde

$(1 - \varphi_p L)$  polinomio autorregresivo

$(1 - L)$  notación de retardo

$(1 + \theta_q L)$  polinomio medias móviles

$\epsilon_t$  error de regresión

$a_t$  error del modelo ARIMA

Cuando se estiman los parámetros de este modelo se debe minimizar los valores de  $a_t$  cuadrados. Si procedemos a minimizar los valores de  $\epsilon_t$ , tal como sucede cuando estimamos con regresión sin considerar las autocorrelaciones en los errores, se pueden presentar los siguientes problemas:

1. Los coeficientes estimados ya no son las mejores estimaciones.
2. Cualquier prueba estadística asociada al modelo será incorrecta
3. Los valores AIC de los modelos ajustados ya no serán buenas guías para seleccionar el mejor modelo para pronosticar
4. Los valores asociados con los coeficientes serán muy pequeños, por lo que se podrán considerar variables como significativas cuando no lo son, a esto se le denomina regresión espuria.

Para (Hyndnan & Athanasopoulos, 2018) estos problemas se pueden evitar minimizando los valores de  $a_t$  o utilizar la estimación de la máxima verosimilitud. Para aplicar este modelo se debe considerar que todas las variables sean estacionarias. De estimarse un modelo con una de las variables no estacionarias los coeficientes estimados no serán consistentes. De existir una combinación lineal de los no estacionarios y los predictores que son estacionarios, entonces los coeficientes estimados serán consistentes.

### 2.2.7.2 Criterios de selección de variables:

En seguida se describe tres criterios para selección de variables para un modelo determinado

#### **Criterio de información de Akaike (AIC):**

Es un estimador muestral de la esperanza de la log-verosimilitud

Está definido por:

$$AIC = T \log \left( \frac{SSE}{T} \right) + 2(k+2)$$

Siendo:

T: número de observaciones

SSE: suma de los cuadrados de los errores

K: número de variables

El AIC castiga el ajuste del modelo con el número de parámetros que deben estimarse.

El modelo con el AIC mínimo es a menudo el mejor para pronosticar. Para valores grandes de T minimizar el AIC es equivalente a minimizar el CV (coeficiente de variación).

#### **Criterio de información de Akaike corregido (AIC<sub>C</sub>)**

Para valores de T pequeños el AIC tiende a seleccionar muchas predictoras, por lo que se presenta un AIC corregido por el sesgo. El AIC<sub>C</sub> también se minimiza

$$AIC_c = AIC + \frac{2(k+2)(k+3)}{T-k-3}$$

### **Criterio de información Bayesiano de Schwarz (BIC)**

Este criterio penaliza más el número de parámetros que el AIC

Minimizar el BIC genera el mejor modelo, queda definido de la manera siguiente:

$$BIC = T \log \left( \frac{SSE}{T} \right) + (k+2) \log (T)$$

### **2.2.8 Errores de pronósticos**

El error de un pronóstico es la diferencia entre su valor real y el pronosticado, se puede expresar de la siguiente manera:

$$e_{T+h} = Y_{T+h} - \hat{Y}_{T+h|T}$$

El error de pronóstico es distinto al de los residuos. Este último se calcula con los datos de entrenamiento mientras que los primeros con los datos de prueba.

### **Errores dependientes de la escala**

Aquí los errores tienen la misma escala que los datos. Para las medidas que se basan solo de la escala no pueden compararse con otras series que involucran escalas diferentes.

Las dos medidas más conocidas son las que utilizan los errores absolutos o los errores al cuadrado. El Error absoluto medio (MAE, por sus siglas en inglés) , un método de pronóstico que minimice lo minimice conducirá pronósticos a la mediana , mientras que minimizar la raíz del error cuadrático medio (RMSE, por sus siglas en inglés) conducirá a pronóstico de la de la media.

$$\text{MAE} = \text{media} (|e_t|)$$

$$\text{RMSE} = \sqrt{\text{media} (e_t^2)}$$



## **CAPÍTULO 3**

### **METODOLOGÍA**

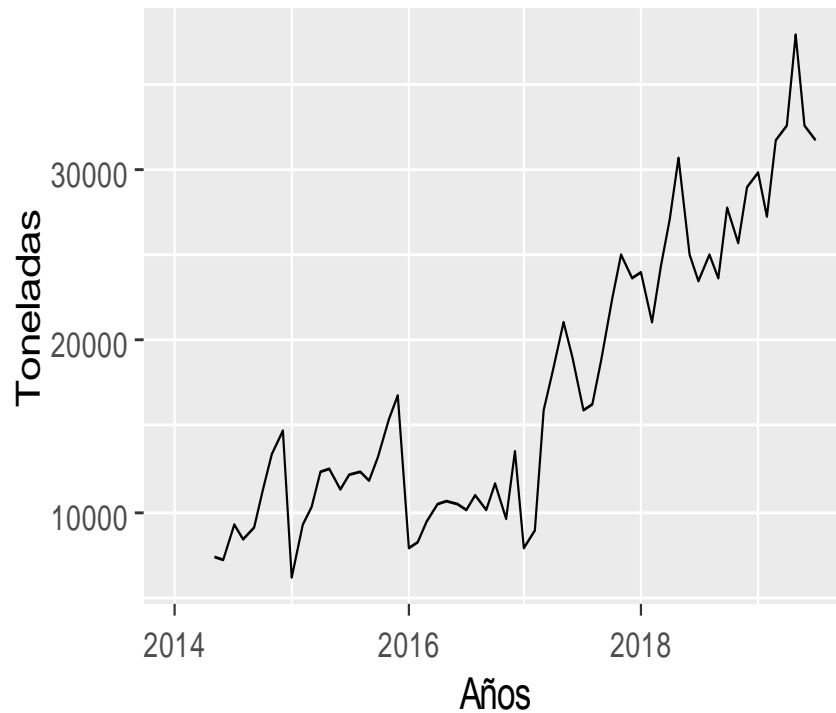
Las fases que comprenden la investigación son tres: comprensión de los datos del problema, construcción y evaluación del modelo y medición del rendimiento del modelo. En las siguientes líneas se describe lo realizado en cada una de las etapas. Entre las técnicas que se emplearán están las entrevistas a personal con experiencia en la comercialización del producto, con la finalidad de asimilar el comportamiento de la demanda de este producto.

#### **3.1 Análisis Descriptivo**

La comprensión de los datos implica entender el dinamismo de la demanda de este producto durante todo el tiempo en que la organización ha permanecido en el mercado. Para iniciar el análisis de la demanda se realizó un diagrama de tiempo (gráfica 1) en el cual se delinear los datos consecutivos contra el tiempo de observación. Las observaciones que se muestran corresponden a las ventas en toneladas de todo el portafolio de producto que la organización comercializa. Como se puede observar en la gráfica las ventas han ido incrementándose a medida que ha transcurrido el tiempo, excepto en el periodo 2016-2017 donde se aprecia un descenso en las ventas. El diagrama también nos revela que en ciertos

meses existen valles y picos en la demanda por lo que el siguiente punto a investigar es el comportamiento según estaciones del año.

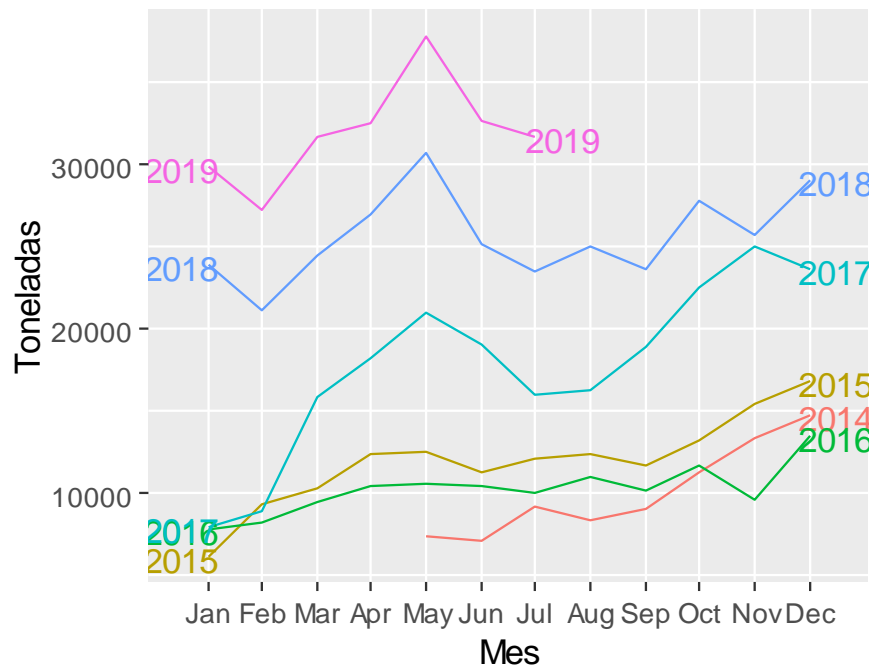
**Gráfica 1. Ventas de Alimento Balanceado**



Para estudiar el comportamiento de la serie de tiempo por estaciones del año recurrimos a entrevistas con profesionales con vasta experiencia en el sector acuícola y conocen el mercado del alimento balanceado para camarones. Nuestra primera entrevistada fue Miriam Mancero, que cuenta con experiencia en el área de nutrición animal en varias empresas reconocidas de esta industria, En la información levantada se nos comentó que la comercialización del alimento está determinada por la temporada climática.

En la temporada invernal donde hay un incremento de temperatura ambiental, se presenta los picos más altos de consumo de alimento debido al incremento en la actividad fisiológica y metabólica del crustáceo. Por lo contrario, en los meses de verano donde la temperatura desciende también lo hace el consumo de alimento debido a una disminución de la actividad fisiológica. Esta información revela un comportamiento estacional de la demanda de este producto. A continuación, el gráfico 2 se muestra las ventas por meses desde mayo 2014 a julio 2019. Analizando los datos de la serie de los 5 años podemos notar un patrón creciente de enero a mayo, luego se observar desde junio a septiembre un decrecimiento en la demanda del cliente. Desde octubre a diciembre la demanda vuelve a tener un repunte.

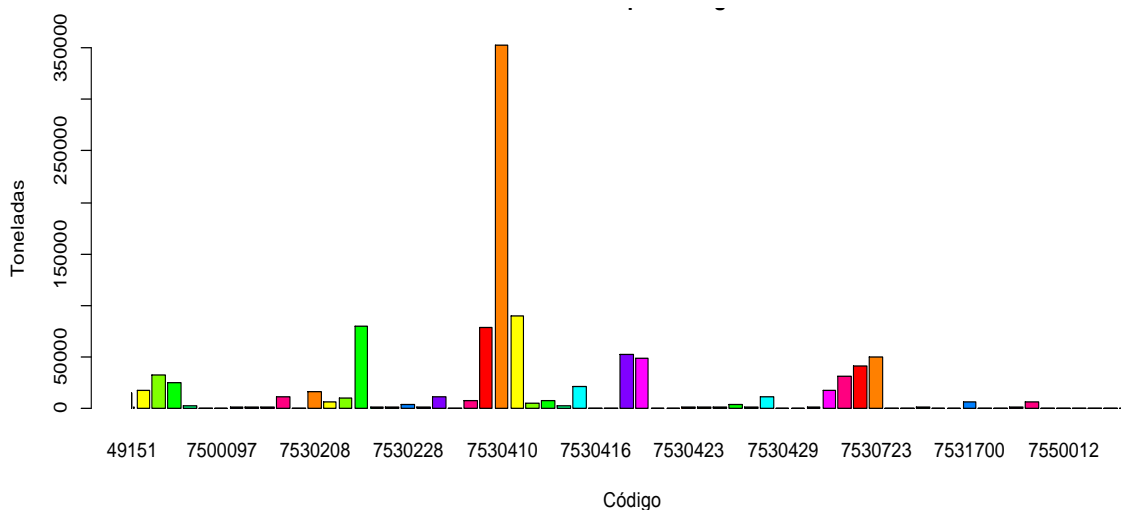
**Gráfico 2. Ventas estacionales de alimento balanceado**



La información que hasta el momento se ha detallado corresponde a ventas totales, es decir las toneladas demandas de todos los productos que la empresa comercializa. Enseguida se mostrará la estadística de las toneladas vendidas segregadas por cada ítem que se comercializa. En la gráfica 3 se puede observar que la referencia 7530410 (que de aquí en adelante del trabajo le mencionaremos como material 65) es la que ha sido mayormente demandada en todos estos años. Un total de 352.697 toneladas que corresponde el 32,96% de las ventas totales durante el periodo mayo 2014 a julio 2019.

Desde inicio de operaciones de la empresa (mayo 2014) el producto con código 7530411 acumula un total de 89.768 toneladas equivalente al 8.39%. de las ventas totales hasta julio 2019, mientras que el material 7530216 registra 79.157 toneladas vendidas que corresponde el 7.39%

**Gráfico 3. Ventas realizadas por referencia**



Como se estableció en la sección de objetivos, enfocaremos el análisis y modelado al producto con mayor demanda por los clientes, en este caso el codificado como 65.

### **3.1.1 Análisis descriptivo de variables macroeconómicas**

En la siguiente parte del trabajo se muestra el comportamiento de varias series económicas que utilizaremos para el análisis del pronóstico de nuestra variable de interés. Iniciaremos con uno de los indicadores más utilizados en los pronósticos, estamos hablando del PIB (Producto Interno Bruto).

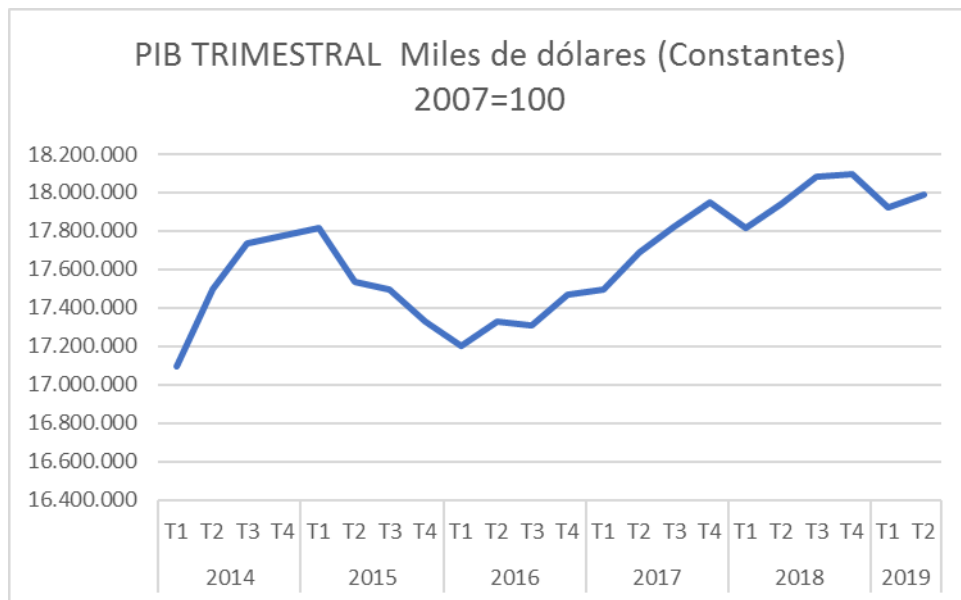
#### **Producto interno Bruto (PIB)**

El PIB, es uno de los indicadores básicos para medir la actividad económica de una nación. Se refiere al valor agregado bruto total de todos los productores residentes en una economía en un periodo determinado. La valoración del PIB puede darse de dos maneras: nominal y el real.

El primero de ellos toma el valor monetario de todos los bienes y servicios producidos en una economía determinada a precios corrientes en el año que son procesados. En cambio, el PIB real se calcula con precios constantes tomados de un año elegido como base. Se aplica al PIB nominal el deflactor del PIB, que es un índice que mide la variación de los precios entre el año base y el año en el que se calcula el PIB nominal.

La comparación del PIB nominal para distintos periodos no es efectiva debido a la diferencia de precios, lo cual es solucionado por el PIB real. Existen tres enfoques para el cálculo del PIB, a saber: por el gasto, por la producción y por el ingreso. Este último es el que utiliza el Banco Central del Ecuador para determinar su valor. Esta metodología consiste en la suma de remuneraciones de asalariados; el excedente bruto de explotación o remuneración del capital; el ingreso mixto y los impuestos netos sobre los productos. Enseguida se muestra la gráfica 4, donde se delinea la serie de los últimos 5 años de este indicador económico.

**Gráfico 4. PIB trimestral Miles de dólar (Constantes) 2007=100**

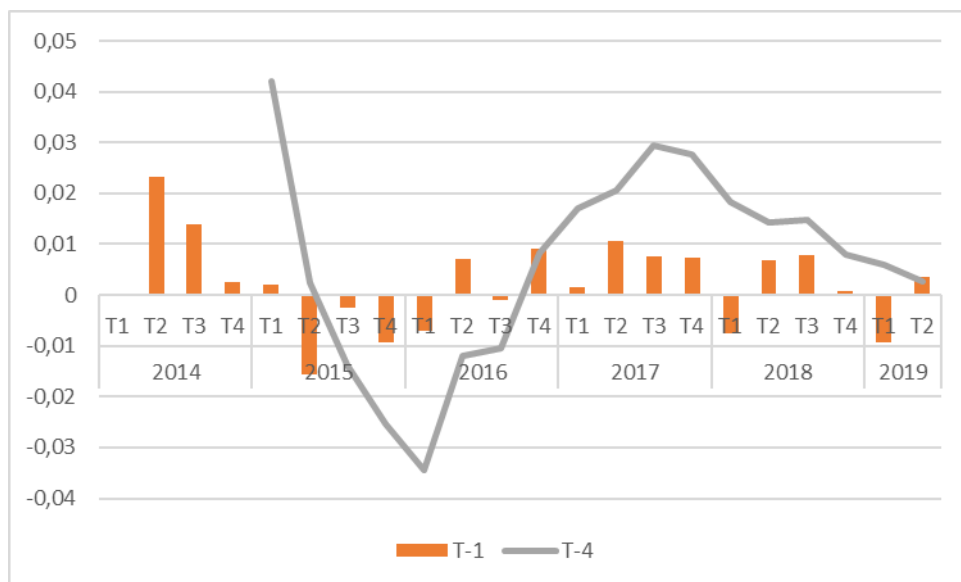


Como vemos el PIB real (en dólares constantes como base el 2007) ha tenido cambios en su nivel. Vemos por ejemplo que desde el 4to trimestre del 2014 tuvo una tendencia a la

baja. Este comportamiento cambia a partir del segundo trimestre del 2016 donde ha ido incrementándose inclusive en el segundo trimestre del 2019.

Si se analiza las tasas de variaciones entre trimestre de un mismo año y su valor correspondiente al periodo anterior se tiene otras conclusiones. Así tenemos que el PIB del segundo trimestre del 2019 creció en 0.3%, con respecto al segundo trimestre del 2018 (t/t-4) y reflejó una variación positiva de 0.4% respecto al trimestre anterior. Como se puede observar en la gráfica 5 las variaciones anuales de cada trimestre tienen una tendencia decreciente a partir del 4 trimestre del 2017.

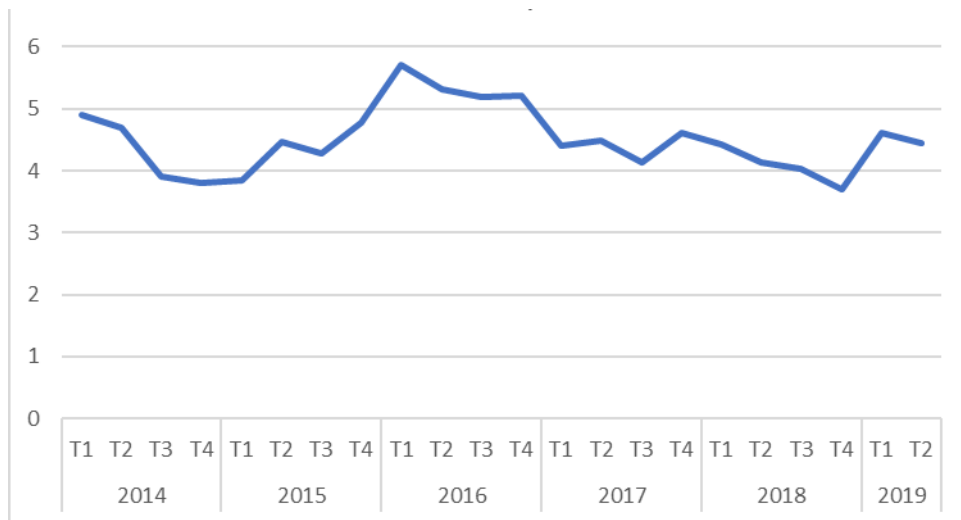
**Gráfico 5. Producto Interno Bruto-PIB 2007= 100, Tasas de variación trimestral**



## Desempleo

Otra de las variables consideradas para este trabajo es la tasa de desempleo en el país. Este indicador se calcula por la relación del número de personas desocupadas por Población Económicamente Activa (PEA). Enseguida se muestra la base histórica de los últimos 5 años de la tasa de desempleo nacional.

**Gráfico 6. Nivel de Desempleo Nacional**



Como ocurre con muchas otras variables macroeconómicas notamos que la tasa de desempleo varía de trimestre a trimestre y en cada año. La primera observación que se detecta en la serie que desde el segundo trimestre del 2015 hasta el primer trimestre del 2016 el desempleo nacional experimentó un crecimiento pasando de un 4.47 a 5.71% (el máximo de estos 5 años) es decir un incremento de 0.24 puntos porcentuales. Luego de ese periodo esta variable reporta valores inferiores trimestre a trimestre llegando a 3.69 en el IV trimestre del 2018 el cual es el más reducido en la serie de estos 5 años.



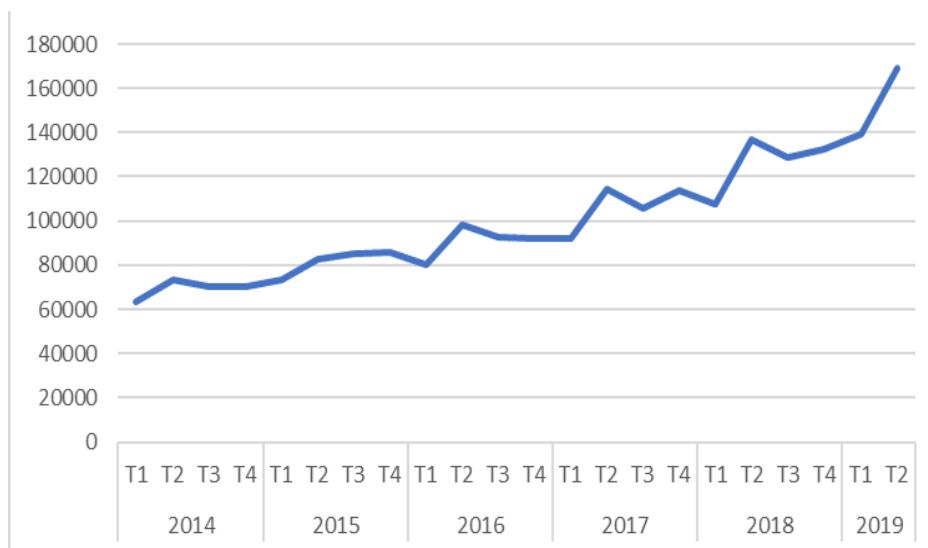
De acuerdo con los datos del BCE (Banco Central del Ecuador) en marzo del 2019 la tasa de desempleo nacional fue de 4.6%, a nivel urbano se ubicó en 5.8% y a nivel rural en 2.2%; además las variaciones tanto a nivel nacional, urbano y rural respecto a marzo del 2018 no fueron estadísticamente significativa.

### Exportaciones de camarón

El sector camaronero se ha convertido en la principal fuente de ingresos no petroleros para el país relegando a industrias que por tradición venían realizando aportes importantes en la economía ecuatoriana. Este sector ha venido creciendo en los últimos años, luego de haber sufrido la crisis de la mancha blanca al finalizar la década de los 90.

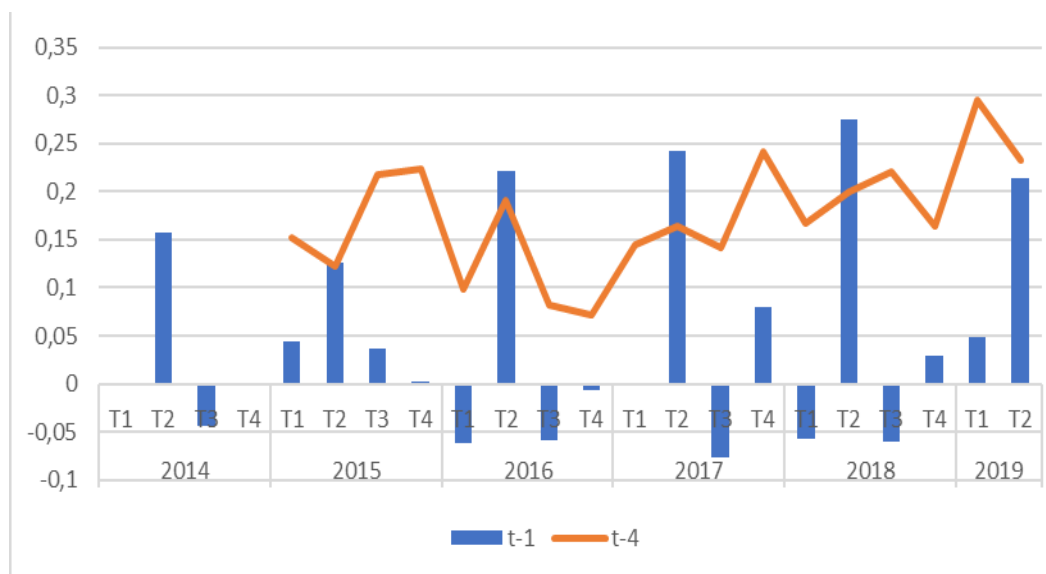
Hemos recopilado datos del comportamiento histórico de las exportaciones de camarón en los últimos 5 años (ver gráfico 7), nótese que la curva tiene una tendencia creciente. Se aprecia que en el segundo trimestre las exportaciones llegan al pico anual lo cual denota que esta variable puede tener efectos estacionales.

**Gráfico 7. Exportaciones de camarón en los últimos 5 años (Toneladas)**



El nivel de crecimiento de las exportaciones ha venido variando en estos años (gráfico 8); así podemos indicar que el año 2015 hubo un crecimiento del 17.88% en las toneladas exportadas; para el 2016 el crecimiento fue menor (alrededor del 11%) con respecto al año anterior. Para el 2017 el mercado logro un crecimiento 17.34% y en el año 2018 el crecimiento fue de 18.81%, dicho valor es mayor que su antecesor en 1.37 puntos porcentuales. Para lo que va del 2019, se observa crecimiento en el primer y segundo trimestre de 29.5 y 23.2% respetivamente (con respecto a los mismos periodos del 2018). Como se señaló en el marco teórico el crecimiento de este sector está asociado mucho con la desaceleración en la producción del crustáceo en Asia, que de acuerdo con las estadísticas de la Cámara Nacional de Acuicultura el 62% de las exportaciones del 2018 tuvieron como destino ese continente.

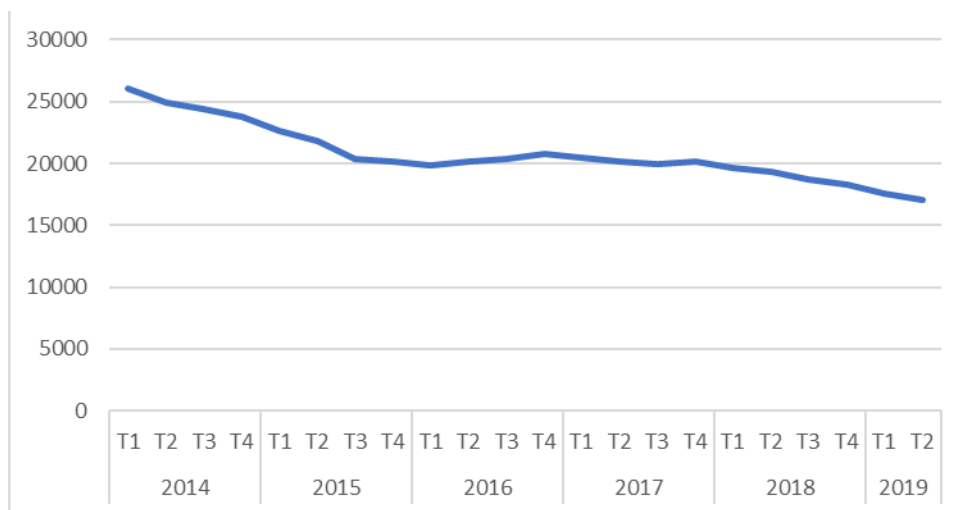
**Gráfico 8. Exportaciones de Camarón, Tasas de variación trimestral**



## Precio del camarón

Otra variable que tenemos en consideración para el análisis es el precio de la tonelada del camarón exportado. A diferencia del nivel de toneladas exportados el precio de este ha venido decreciendo tal como se observa en la gráfica 9, donde claramente se identifica una tendencia decreciente.

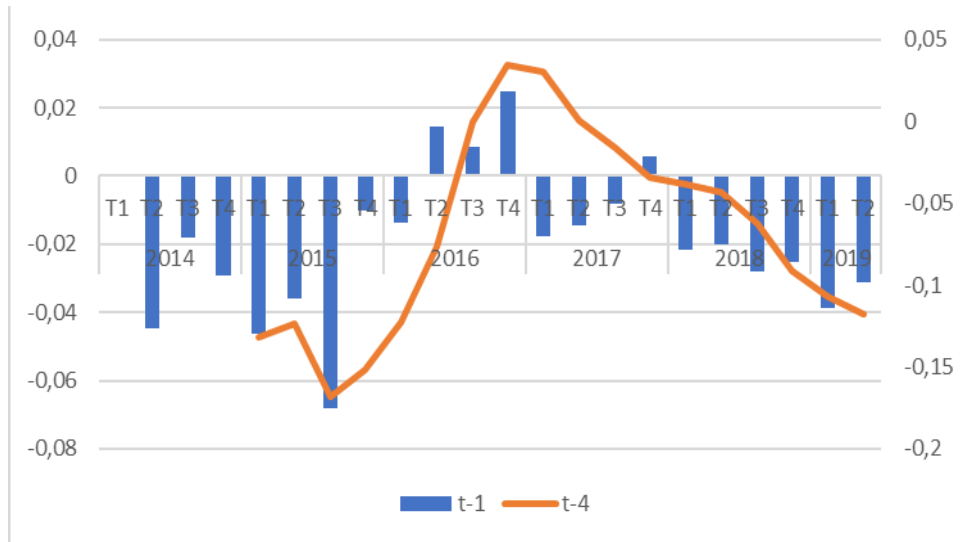
**Gráfico 9. Precio Promedio de la Toneladas de camarón exportado**



Este decrecimiento obedece al incremento de la demanda mundial del marisco por lo que es de esperarse una baja en sus precios (oferta de ley oferta-demanda). El cuarto trimestre del año 2016 reportó un crecimiento significativo en los precios al llegar a un alza del 3.44% (con respecto al mismo trimestre del año 2015). La variación del precio en un mismo periodo continuó la tendencia a decrecer y a partir del tercer trimestre del 2017 se reportó una caída del 1.63% con respecto al mismo periodo del 2016 y así se ha mantenido en estos años. En

lo que va del 2019 tenemos una caída del 10.72 y 11.75% en los dos primeros trimestres respectivamente.

**Gráfico 10. Precio de Camarón, Tasas de variación trimestral**



### 3.2 Construcción y evaluación del modelo.

#### 3.2.1 Modelo ARIMA estacional

El primer modelo que utilizaremos es el autorregresivo integrado medias móviles.

Para realizar este modelo seguiremos el siguiente procedimiento.

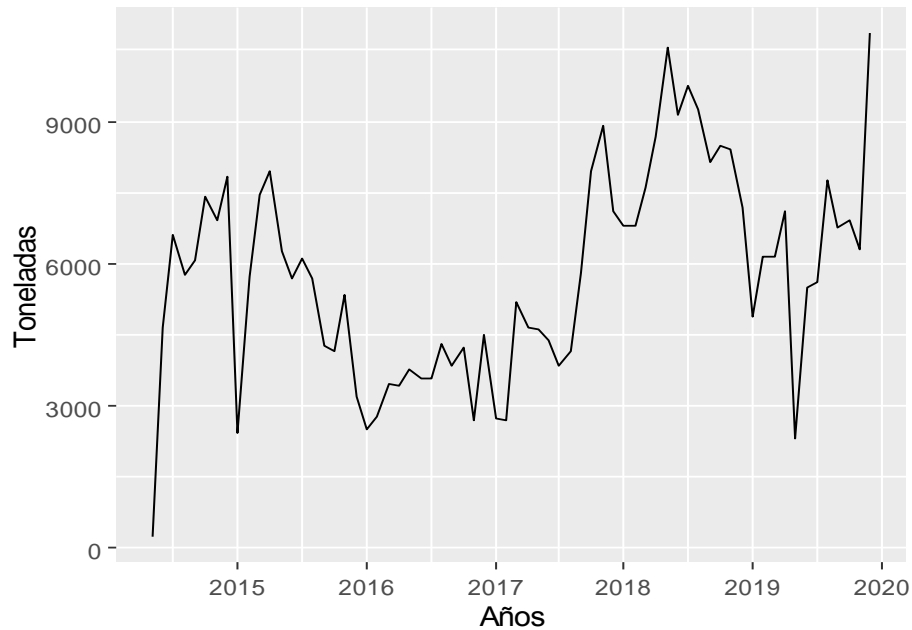
1. Trazar los datos y verificar si la serie es estacionaria.
2. De comprobarse que la serie no proviene de un proceso estocástico estacionario se debe transformar los datos según el tipo de no estacionariedad. Para estabilizar la varianza se utiliza la transformación Box Cox y para lograr estacionariedad en media y en caso de que la serie presente estacionalidad se calculan las primeras diferencias

estacionales de los datos de los datos transformados y en caso de que no logre obtener estacionariedad se procede a determinar la primera diferencia regular.

3. Determinar los órdenes de  $p$  y  $q$ , para ello comparamos las funciones de autocorrelación (FAC) muestrales con las teóricas de los modelos ARIMA.
4. Se estiman los parámetros desconocidos  $(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)$  con el método de máxima verosimilitud.
5. Probar los modelos seleccionados utilizando el AIC para buscar un modelo mejor.
6. Verificar los residuos del modelo elegido trazando el FAC de los residuos y generando una prueba de resumen de los residuos.
7. Luego que se valide que los residuos se comporten como ruido blanco, se calculan los pronósticos

Para iniciar el modelado trazamos la demanda histórica de la referencia 65. Al observar la serie (gráfico 11) se puede notar que la media varía al transcurrir el tiempo, adicionalmente hay que advertir la presencia de estacionalidad en los datos y de varianza no estable.

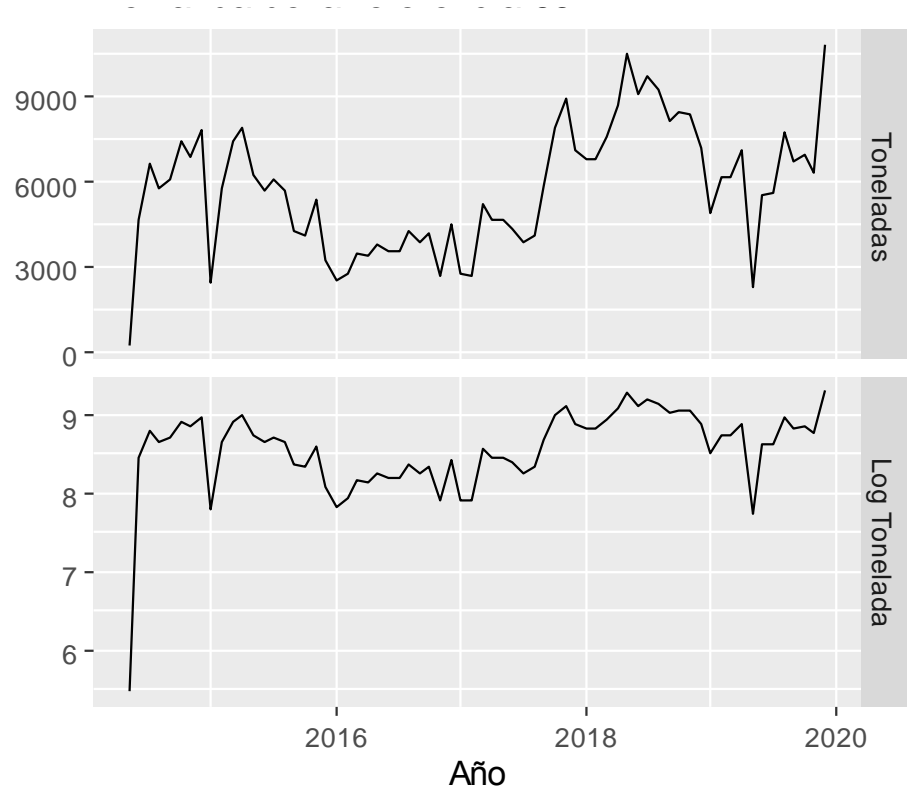
**Gráfico 11. Demanda de la referencia 65**



Dadas estas características en la serie podríamos determinar que los datos no provienen de un proceso estacionario. Para confirmar esta información se realiza un contraste de hipótesis utilizando la prueba *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) (Hyndnan & Athanasopoulos, 2018). Como hipótesis nula se plantea que la serie es estacionaria, mientras que la alternativa establece que el proceso no es estacionario, utilizando un nivel de significancia de 0,05.

El valor del estadístico de la prueba (0.5835) es mayor al valor crítico correspondiente al nivel de significancia planteado de 5% (0.463) por lo que no se acepta la hipótesis nula de que el proceso es estacionario. Dada la no estacionariedad de la serie es necesario aplicar transformaciones para lograr la estacionariedad en varianza y media. Para tal fin se procedió en primera instancia a realizar una transformación logarítmica de los datos. Los resultados de estas transformaciones se muestran en segundo panel la gráfica 12.

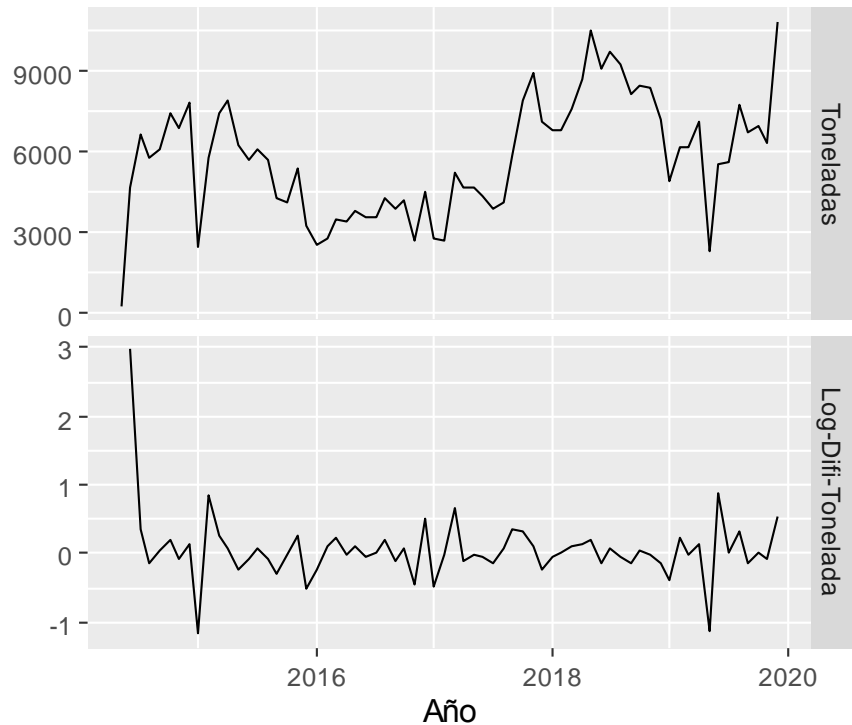
**Gráfico 12. Demanda de la referencia 65**



Los datos transformados presentan estabilidad en varianza, aunque la media aún presenta cierta inestabilidad en varios periodos. Para efectos de corroborar esta percepción volvemos a realizar una prueba de hipótesis KPSS.

El valor del estadístico de prueba (0.5939) es mayor a valor crítico al 5% de significancia (0.4630), por lo cual nos permite inferir que los datos transformados no provienen de un proceso estacionario. Una alternativa es adicionar una primera diferencia regular a los datos. Los datos con las diferencias realizada muestran en el panel segundo de la gráfica 13. Al comparar los datos originales con los de las transformaciones se puede notar mayor estabilidad en la media en el segundo conjunto de datos.

**Gráfico 13. Demanda transformada de la referencia 65**

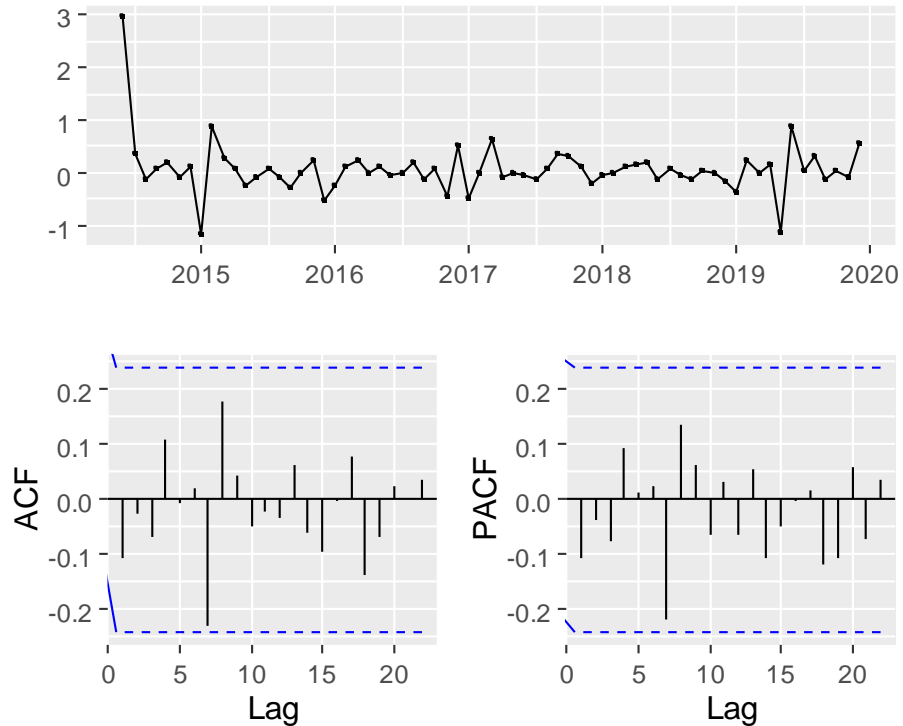


Aplicando una vez más la prueba KPSS obtenemos un valor de estadístico igual a 0.201 que es menor al valor crítico planteado igual a 0.463 con nivel de significancia del 5%. Con este resultado de la prueba podemos concluir que realizando una diferencia regular a los datos transformados logarítmicamente presentan estacionariedad.

Una vez que hemos realizado las transformaciones para obtener estacionariedad el siguiente objetivo es determinar un modelo ARIMA apropiado basados en los valores de  $p$  y  $q$ . Para ello nos valdremos de las gráficas ACF y PACF de los datos transformados, los cuales se muestran enseguida:



**Gráfico 14. Diagrama ACF y PACF de datos transformados**



Al observar la gráfica ACF (función de autocorrelación) de los datos transformados se observa que el primer término es distinto de cero. Esto sugiere un componente autorregresivo AR (1). En la PACF el primer término es distinto a cero y luego va reduciendo su valor lo cual refiere un término (MA) de 1. Por lo tanto, este análisis sugiere que un posible modelo para estos datos es un ARIMA (1,1,1) Con los códigos utilizados del software R-Studio, se estimaron los parámetros  $\delta$ ,  $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ , por medio el método de la máxima verosimilitud. Enseguida se muestran los resultados del ajuste del modelo:

Las medidas de ajuste que se utilizarán en el modelamiento son error medio (ME) raíz del error cuadrático medio (RMSE), error absoluto medio (MAE), error porcentual medio (MPE) y error porcentual absoluto medio (MAPE).

El coeficiente para la parte autoregresiva es de 0.3147 y para la media móvil  $-0.6598$ . Dado la estacionariedad de la serie la constante  $\delta$  es considerada como cero. El modelo cuenta con un AICc igual a 94.57 además el software nos devuelve varias medidas de error las cuales nos servirán para decidir en la selección del mejor modelo que ajuste los datos.

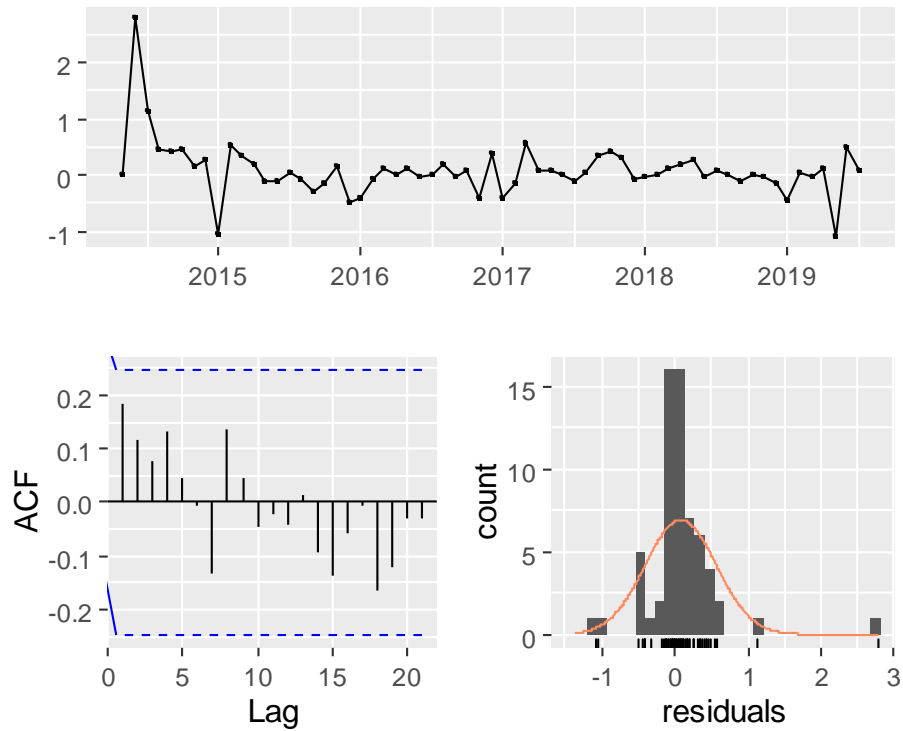
	ME	RMSE	MAE	MPE	MAPE
Test set	337.5377	1674.997	1188.154	-0.8360908	25.2781

Utilizando la notación ARIMA el modelo ajustado se puede expresar como:

$$Y'_t = 0.3147 Y_{t-1} - 0.6598 a_{t-1}$$

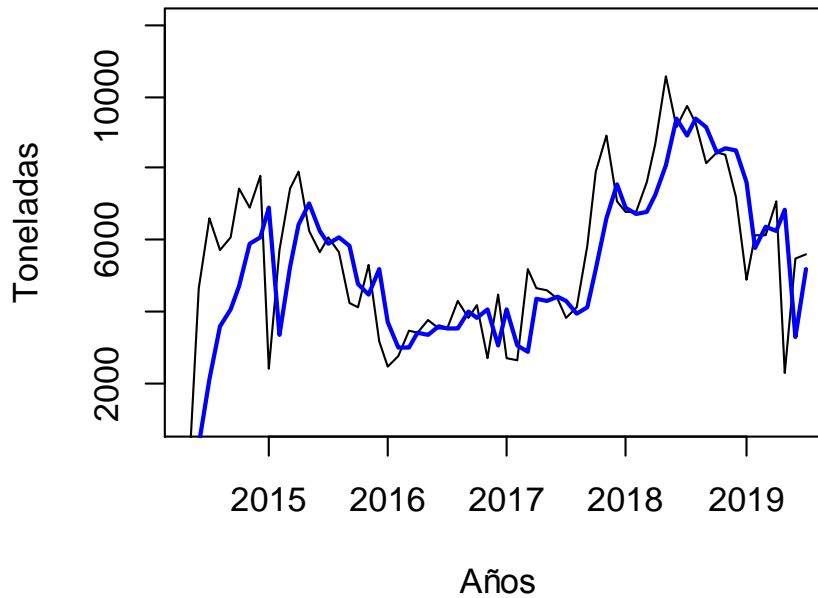
En el gráfico 15 se puede apreciar que en el ACF de los residuos del modelo ARIMA (1,1,1) todas las autocorrelaciones están dentro del umbral, lo cual significa que los residuos se comportan como proceso ruido blanco.

**Gráfico 15. Residuos de ARIMA (1,1,1)**



Este modelo está referido para los datos transformados logarítmicamente, por lo cual para realizar la estimación se debe realizar la conversión de los datos. En la gráfica 16 podemos observar la curva que ajusta los datos, cabe mencionar que el modelo captura el comportamiento de la serie original teniendo como medida de error medio absoluto (MAE) igual a 1188.154 y con un AIC (Criterio de información Akaike corregido) 94.15

**Gráfico 16. Curva de ajuste de demanda para referencia 65,  
Modelo ARIMA (1,1,1)**



Los valores  $p$  y  $q$  del modelo anterior fueron seleccionados interpretando las gráficas ACF y PACF, dado que este proceso es iterativo debemos probar varios modelos probando determinados ordenes de  $p$  y  $q$ . El software utilizado para este trabajo nos permite generar todos los modelos posibles y se informa el mejor modelo que ajuste los datos en base a su valor AIC. Enseguida se muestran una parte de los resultados AIC de los modelos posibles:

**Tabla 1 Posibles modelos con su respectivo AIC**

MODELO	AIC
ARIMA (0,1,0)	92,929
ARIMA (0,1,0) with drift	94,429
ARIMA (0,1,0) (0,0,1) [12]	94,969
ARIMA (0,1,0) (0,0,1) [12] with drift	96,510
ARIMA (0,1,0) (1,0,0) [12]	94,939
ARIMA (0,1,0) (1,0,0) [12] with drift	96,474
ARIMA (0,1,0) (1,0,1) [12]	97,010
ARIMA (0,1,0) (1,0,1) [12] with drift	98,623
ARIMA (0,1,1)	92,848
ARIMA (0,1,1) with drift	94,415
ARIMA (0,1,1) (0,0,1) [12]	95,054
ARIMA (0,1,1) (0,0,1) [12] with drift	96,688
ARIMA (0,1,1) (1,0,0) [12]	95,053
ARIMA (0,1,1) (1,0,0) [12] with drift	96,685
ARIMA (0,1,1) (1,0,1) [12]	97,304
ARIMA (0,1,1) (1,0,1) [12] with drift	99,008
ARIMA (0,1,2)	94,436
ARIMA (0,1,2) with drift	96,108
ARIMA (0,1,2) (0,0,1) [12]	96,532
ARIMA (0,1,2) (0,0,1) [12] with drift	98,208
ARIMA (0,1,2) (1,0,0) [12]	96,481

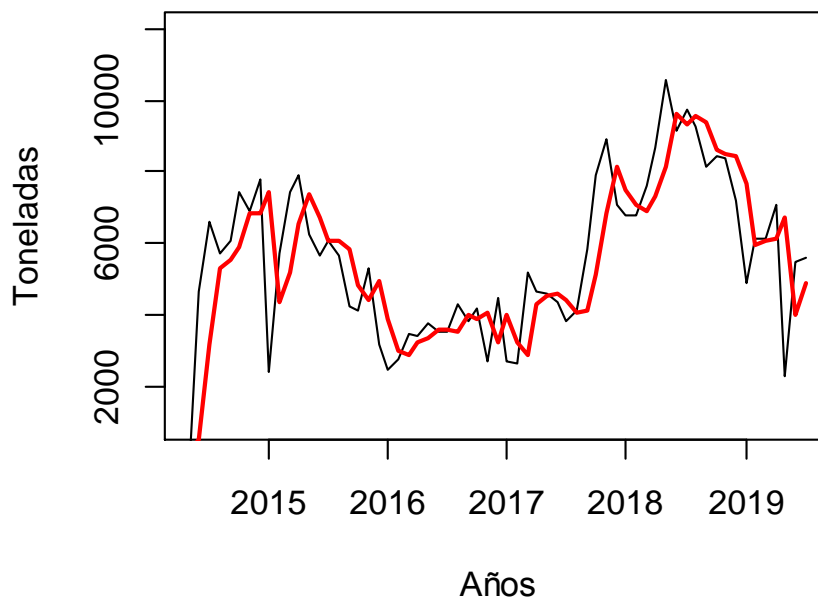
Best model: ARIMA(0,1,1)

Como se ilustra en la tabla el modelo con mejor AIC es el ARIMA (0,1,1) con valor de 92.92 el cual es ligeramente menor al primer modelo planteado. Ajustando los datos con este modelo tenemos los siguientes resultados:

	ME	RMSE	MAE	MPE	MAPE
Test set	120.4908	1540.694	1086.434	-4.410576	23.94868

El MAE y el RMSE de este modelo son mejores que los presentados por el primero. De igual manera el AIC que este modelo presenta es de 92.64 que comparado al modelo ARIMA (1,1,1) es ligeramente menor. En la gráfica 17 se observa la curva ajustada por el modelo la misma que se acerca a los datos reales.

**Gráfico 17. Ajuste de curva de modelo Arima (0,1,1)**



### 3.2.2 Modelo de Regresión Múltiple

En este apartado abordaremos nuestra variable de interés bajo el enfoque estructural o también conocido como causal en la cual el comportamiento de la variable dependiente se explica por la influencia de una o más variables independientes. Se tomarán datos generales de la economía ecuatoriana tales como el PIB (Producto Interno Bruto) y tasa desempleo, así como también de información relacionada a la industria acuícola como lo son las

toneladas exportadas del camarón y su precio promedio mensual. Para iniciar el análisis estructural se procedió con el cálculo de los coeficientes de correlación de las variables independientes o exógenas (PIB, tasa de desempleo, Exportaciones de camarón y precio promedio) con la demanda del alimento de código 65 (ver tabla 2).

**Tabla 2. Matriz de Correlación**

	<b>Demanda Alimento</b>	<b>PIB</b>	<b>Precio promedio (\$/Ton)</b>	<b>Exportaciones Camarón</b>	<b>Desempleo</b>
<b>Demanda Alimento</b>	1	0,74	-0,45	0,46	-0,56
<b>PIB</b>	0,74	1	0,54	0,71	-0,68
<b>Precio promedio (\$/Ton)</b>	-0,45	0,54	1	-0,83	0
<b>Exportaciones Camarón</b>	0,46	0,71	-0,83	1	-0,18
<b>Desempleo</b>	-0,56	-0,68	0	-0,18	1

Hay dos variables que guardan una relación positiva con la demanda del alimento balanceado estas son: PIB y las exportaciones de camarón. Podemos notar que la asociación entre la demanda y el PIB es la que posee un mayor coeficiente de correlación igual a 0.74 lo que implica una relación positiva fuerte.

Las exportaciones de camarón tienen un coeficiente de correlación de 0.46 que denota que esta variable también tiene una relación positiva débil con la demanda del producto. Este resultado era de esperarse debido a que estos productos son complementarios.

Por otro lado, el desempleo y el precio promedio de camarón tienen una moderada relación negativa con nuestra variable de estudio. Así tenemos que el desempleo tiene una correlación de -0.56, es decir que dado un incremento de la demanda del alimento se da apertura a nuevas plazas de empleo y con ello se reduce el indicador de personas desempleadas. La correlación del precio de la tonelada del camarón exportado y la demanda del alimento es igual a -0.45 dicho resultado nos indica que al existir un aumento en las exportaciones de camarón (por tanto, de la demanda de alimento) los precios se reducen. Los resultados obtenidos de este análisis de correlación concuerdan con los postulados de la teoría económica de oferta y demanda.

La matriz de correlación nos dio pautas acerca del tipo de relaciones que las variables económicas tienen con la demanda del alimento, en la siguiente fase del análisis de regresión se generarán modelos con todas aquellas variables potencialmente explicativas.

Al tener más de una variable debemos adoptar una estrategia para seleccionar aquellas que expliquen de la mejor manera la respuesta. En este trabajo hemos utilizado el método de eliminación hacia atrás, en el cual se modela en primera instancia con todos los predictores y se eliminan los de más alto valor  $-p$ . Luego el modelo se ajusta con las  $K-1$  variables de regresión restantes y se define la siguiente a eliminar. Los resultados del primer modelo, en el que se incluye todas las variables (PIB, desempleo, exportaciones de camarón y precio de la tonelada del crustáceo) se muestra enseguida:



$$\text{Demanda} = 0.00575\text{PIB} - 0.328\text{Precio} - 0.0489\text{Exportaciones} - 0.040\text{Desempleo} - 90350$$

Donde:

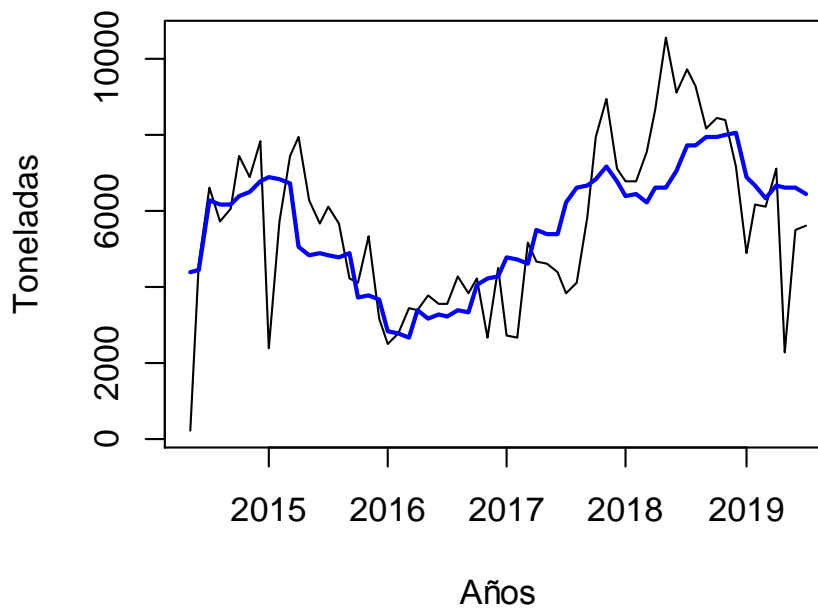
PIB22.ts : Variable PIB

PP.ts: Precio promedio de la tonelada del camarón

Ex.ts : Exportaciones de camarón

Ds.ts : Desempleo

**Gráfico 18. Ajuste de modelo de regresión múltiple (4 variables)**



Como podemos observar el coeficiente de la variable PIB es positiva lo cual quiere decir que ante un incremento en esta variable ocasionará un aumento en la demanda del alimento referencia 65 según el valor de su coeficiente, mientras que para las otras variables su coeficiente es negativo. Este modelo tiene un  $R^2$  ajustado igual a 0.4503 y la variable que debemos eliminar (debido a que tiene el mayor valor-p) es el desempleo. También se muestra las métricas de ajuste para los datos de entrenamiento y prueba.

	ME	RMSE	MAE	MPE	MAPE
Training set	-1.155e-13	1544.814	1131.468	-34.565	50.933
Test set	5.875e+02	1591.949	1104.852	4.519	12.332

Realizando la eliminación de la variable desempleo obtenemos los resultados que se muestran enseguida

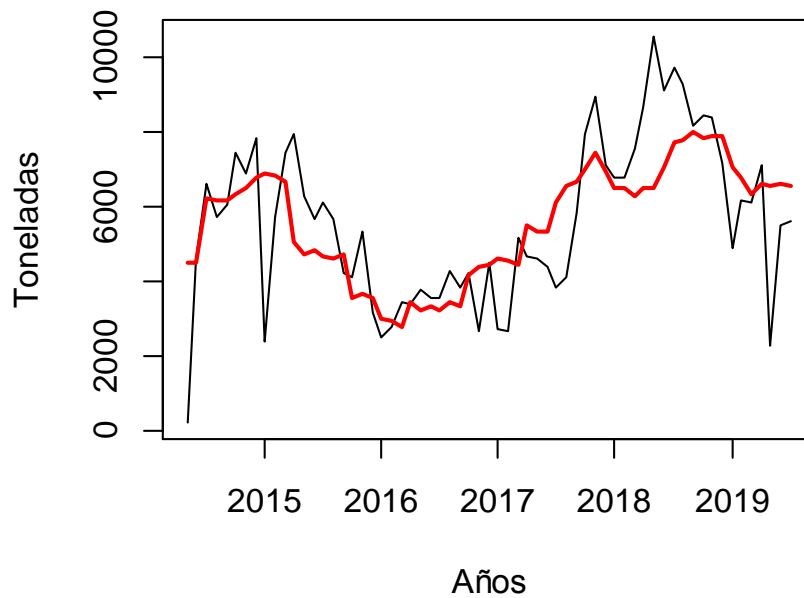
$$\text{Demanda} = 0.00656\text{PIB} - 0.0583\text{Exportaciones} - 0.271\text{Precio} - 106700$$

En este modelo se observa que el valor del  $R^2$  ajustado es igual al modelo inicial. Los signos de las variables PIB, precio y exportaciones son los mismo que en el modelo 1. Las mediciones de los errores de ajuste para los datos de entrenamiento se muestran en la siguiente tabla.

	ME	RMSE	MAE	MPE	MAPE
Training set	0.00	1548.49	1129.44	-35.25	51.51 0.
Test set	643.19	1651.59	1044.86	5.26	11.27 0.

Comparando estos valores con los del modelo 1, se observa que en casi todas las métricas el modelo de 3 variables tiene mejores resultados, solamente en el RMSE el modelo 2 es superado por el inicial. Esto nos da señales de un mejor ajuste de este modelo. La figura 19 muestra el ajuste (línea roja) de este modelo con tres variables.

**Gráfico 19. Ajuste de modelo de regresión múltiple (3 variables)**



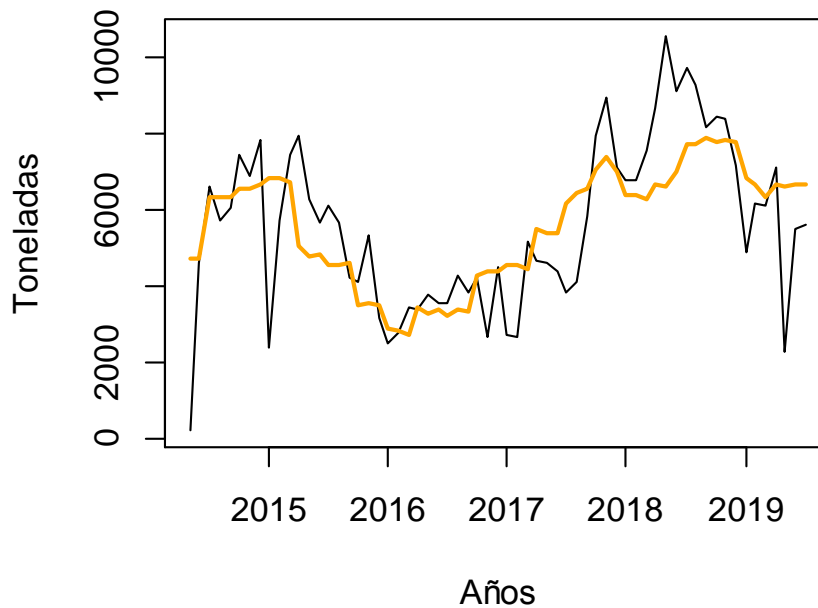
Dado el valor p de la variable precio del camarón será la siguiente en ser eliminada. Por lo que el tercer modelo planteado excluye a dicha variable. Se obtiene un  $R^2$  igual a 0.4643 el cual es ligeramente superior al de los otros dos modelos.

$$\text{Demanda} = 0.00636 \text{ PIB} - 0.0391 \text{ Exportaciones} - 105600$$

En este modelo de dos variables las métricas de ajuste (ME, RMSE, MAE, MPE Y MAPE) son ligeramente superadas por el modelo precedente. En la figura 14 se traza en línea anaranjada el ajuste de la curva real de demanda.

	ME	RMSE	MAE	MPE	MAPE
Training set	0.00	1551.12	1130.28	-36.75	53.01
Test set	620.28	1683.65	1074.25	4.80	11.68

**Gráfico 20. Ajuste de modelo de regresión múltiple (2 variables)**

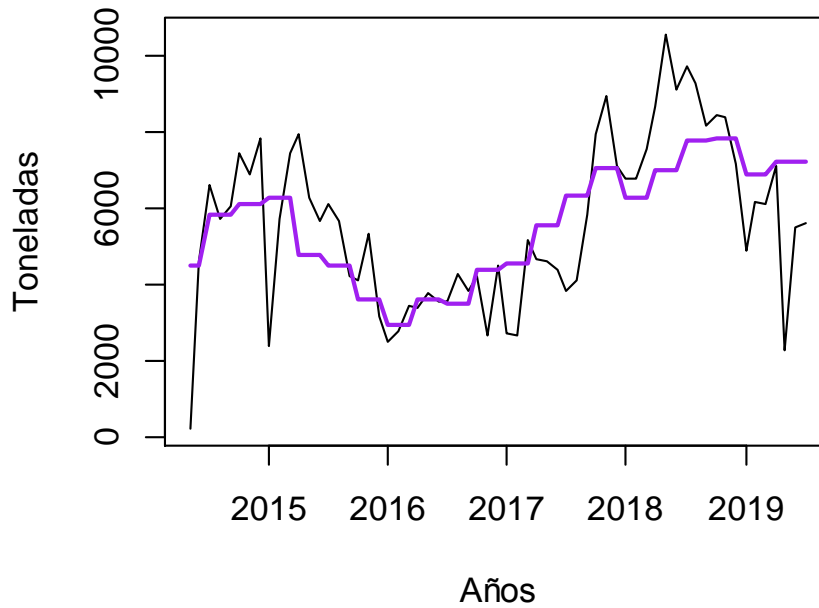


Si continuamos con la eliminación de variables por su valor p, la siguiente en ser retirada del modelo sería exportaciones. El modelo con una sola variable independiente queda definido con los siguientes resultados:

$$\text{Demanda} = 0.0055\text{PIB} - 91690$$

Como podemos observar este modelo junto con el primero es el que presenta el menor valor  $R^2$  de todos los ajustados bajo el enfoque de regresión múltiple. Como se observa en la gráfica 21 la curva del modelo no se ajusta en todos los puntos.

**Gráfico 21. Ajuste de modelo de regresión (1 variable)**



Tenemos enseguida las métricas de ajuste, los mismos que son los más altos que todos los modelos de regresión que hemos planteado.

	ME	RMSE	MAE	MPE	MAPE
Training set	0.00	1577.22	1167.76	-35.82	52.14
Test set	165.24	1678.07	1218.33	-1.59	14.41

### 3.2.3 Modelos de Regresión Dinámica

En este apartado se mostrarán varios modelos bajo el enfoque de regresión dinámica con errores Arima para ajustar los datos de nuestro caso de estudio. El primer modelo que

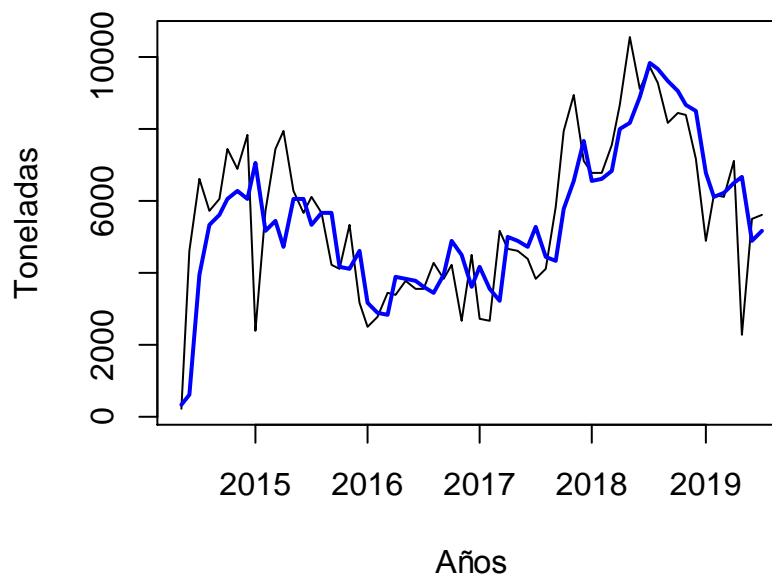
se propone el ajuste de regresión errores arima (0,1,1), tomamos estos parámetros basados a los resultados del modelo ARIMA (0,1,1) que se planteó en el apartado 3.2.1.

Enseguida se muestran los resultados de la regresión dinámica:

Regression with ARIMA(0,1,1) errors

$$\text{Demanda} = 0.0047\text{PIB} - 0.922 \text{ Precio} + 0.027 \text{ Exportaciones} - 126.784\text{Desempleo} - 0.5704 e_{t-1}$$

**Gráfico 22. Regresión con errores ARIMA (0,1,1) con 4 variables regresoras**



	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	115.14	1456.06	1009.36	-5.44	23.01	0.41
Test set	2083.50	2717.43	2083.50	24.00	24.00	0.85

Las métricas de ajuste del modelo dinámico con cuatro variables son mejores a los del modelo Arima (0,1,1) y de regresión múltiple. En cuanto a los errores de pronóstico el modelo de regresión dinámico es mejor al arima (0,1,1) pero no al de regresión múltiple Se probaron otros modelos con errores arima con variaciones en los parámetros p, d y q. Se

obtienen valores muy similares al considerado inicialmente, por ejemplo, en el modelo con errores Arima (1,1,1) se obtiene un valor de AIC igual 1093.93 y en el modelo Arima (1,1,0) tenemos 1092.48. También se ajustaron otros modelos con errores Arima (0,1,1) eliminando variables para observar sus valores AIC, en la siguiente tabla 3 se muestran los resultados con cada modelo.

**Tabla 3. Modelos dinámicos con error ARIMA (0,1,1)**

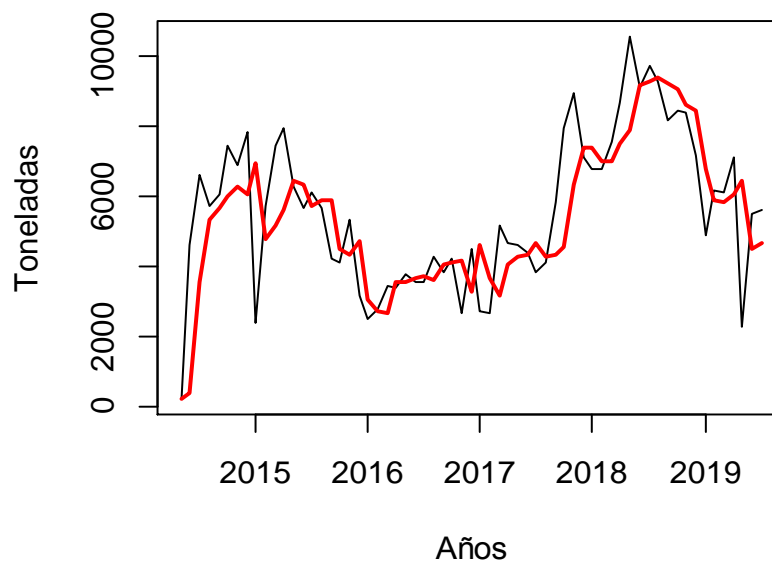
<b>VARIABLES PREDICTORAS</b>	<b>AIC</b>	<b>MAE (entrenamiento)</b>	<b>MAE (prueba)</b>
PIB, Precio, Exportaciones y desempleo	1093.93	1009.36	2083.50
PIB, Precio y Exportaciones	1090.51	1011.67	2117.70
PIB y Precio	1088.75	1013.62	2067.06
PIB y Exportaciones	1089.19	1010.44	2109.71
PIB y Desempleo	1089.27	1011.80	4273.12
Precio, Exportaciones y desempleo	1093.33	1037.73	2223.09
Precio y Desempleo	1092.42	1055.36	1532.08
Exportaciones y Desempleo	1092.65	1059.92	2111.41
PIB	1087.27	1013.07	2076.47
Exportaciones	1092.32	1064.48	2488.26
Desempleo	1091.40	1080.8	1994.50
Precio	1092.39	1069.45	2363.83

Como podemos observar el modelo dinámico que utiliza como única variable predictora el PIB tiene el AIC más bajo. También se ajustó un modelo con las dos variables (PIB y exportaciones) que tiene las correlaciones positivas mas fuertes con la demanda, el cual presenta un AIC igual a 1089,19. Basándonos en los valores del MAE, el mejor modelo dinámico que se ajusta a los datos de entrenamiento es aquel que utiliza las 4 variables, pero en los datos de prueba el modelo que incluye las variables precio y desempleo es el mejor

El ajuste de la curva obtenido en el modelo de regresión con errores Arima (0,1,1) con dos variables predictoras (PIB – Precio) es similar al modelo ARIMA (0,1,1) pero este último tiene un MAE igual a 1086.43 mientras que el modelo dinámico arroja un valor igual a 1013.62. En la gráfica 23 siguiente se muestra el ajuste del modelo dinámico con las variables PIB y precio como predictoras.

$$\text{Demanda} = 1.1622 \text{ Precio} - 953.028 \text{ Desempleo} - 0.5039 e_{t-1}$$

**Gráfico 23. Regresión dinámica con errores ARIMA (0,1,1) dos variables predictoras**





### 3.3 Selección del modelo

Luego de presentar los posibles modelos que ajusten la curva de demanda del alimento de referencia 65, proseguimos a seleccionar el mejor de ellos. En la tabla 2, mostraremos un resumen de las medidas de precisión de ajuste y predicción correspondiente a cada modelo ajustado. Las métricas con las que trabajaremos son la raíz del error cuadrático medio (RMSE) y el error absoluto medio (MAE). El valor “1” indica que el predictor se incluyó en el modelo, y “0” significa que el predictor se excluyó en el modelo.

En primer lugar, analizamos los resultados de los modelos ARIMA, con este enfoque se observa que el modelo Arima (0,1,1) es el que presenta los mejores ajustes con base en el MAE. Pero en los datos de prueba las métricas de precisión no son las mejores. En los modelos de regresión estática, en base al RMSE el mejor modelo que se ajusta es aquel que admite todas las variables ya que presenta el valor mínimo 1449.41, mientras que considerando el MAE el mejor ajuste se da con las variables que incluye PIB, precio y exportaciones. En la regresión dinámica, el modelo que mejor ajusta la curva con el MAE mínimo es el que utiliza todas las variables y considerando al RMSE también es este modelo el que presenta los mejores ajustes. Como podemos notar los modelos de regresión dinámica presentan los mejores ajustes, pero en los datos de prueba no son precisamente los que más se destacan. Contrario a ello tenemos los modelos de regresión estática, los cuales son los que menos ajustan los datos de entrenamiento pero que con los datos de prueba son los que presentan los MAE mínimos.

Para seleccionar el modelo definitivo tomaremos como base el valor mínimo del MAE de los datos de prueba, por lo que el modelo de regresión estática con tres variables Pib, Precio y exportaciones con un MAE igual a 1044.86 es el que seleccionaremos.

El modelo queda definido por siguiente expresión:

$$\text{Demanda} = 0.00656\text{PIB} - 0.0583\text{Exportaciones} - 0.271\text{Precio} - 10670$$

**Tabla 4. Métricas de Precisión de ajuste y pronóstico de modelos propuestos**

Método de Análisis	PIB	Precio	Exportaciones	Desempleo	ENTRENAMIENTO		PRUEBA	
					RMSE	MAE	RMSE	MAE
ARIMA (1,1,1)	0	0	0	0	1674,99	1188,15	2945,56	2443,47
ARIMA (0,1,1)	0	0	0	0	1540,69	1086,43	2891,28	2381,39
Regresión Estática	1	1	1	1	1544,81	1131,47	1591,95	1104,85
	1	1	1	0	1548,49	1129,44	1651,59	1044,86
	1	0	1	0	1551,12	1130,28	1683,65	1074,25
	1	0	0	0	1577,22	1167,76	1678,07	1218,33
Regresión Dinámica	1	1	1	1	1449,41	1007,91	2516,61	1871,64
	1	1	1	0	1456,42	1011,67	2750,91	2117,7
	1	0	0	1	1465,23	1011,8	4569,92	4273,12
	0	1	1	1	1491,03	1037,73	2775,07	2223,09
	0	1	0	1	1504,25	1055,36	1691,42	1532,08
	0	0	1	1	1507,6	1059,92	2729,56	2111,41
	1	1	0	0	1459,06	1013,62	2659,41	2067,06
	1	0	1	0	1464,48	1010,44	2718,3	2109,71
	1	0	0	0	1465,32	1013,07	2664,93	2076,47
	0	0	1	0	1529,07	1064,48	3075	2488,26
	0	1	0	0	1529,43	1069,45	2885,39	2363,83
0	0	0	1	1516,55	1080,8	2556,28	1994,5	

# CAPÍTULO 4

## VALIDACIÓN Y PRONÓSTICO DEL MODELO

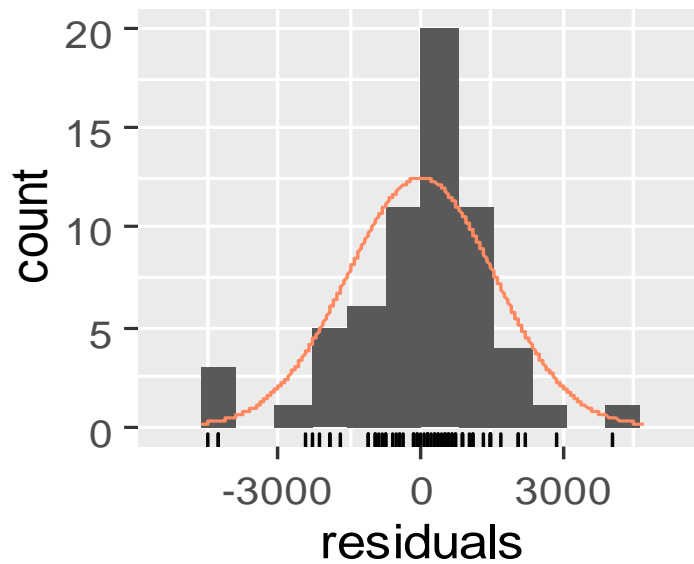
### 4.1 Validación de supuestos del modelo seleccionado

En el capítulo anterior se plantearon varios modelos con distintos enfoques para el ajuste de los datos de la demanda del alimento referencia 65 y luego con base al MAE, se determinó que el mejor modelo que predice es la regresión estática con tres variables. Para continuar con la metodología planteada, en este capítulo se revisará el cumplimiento de cada uno de los supuestos del modelo seleccionado (Hyndnan & Athanasopoulos, 2018)

#### 4.1.1 Errores con distribución normal

El primer supuesto que se procederá a contrastar es la distribución normal de los errores. Nos valdremos de una herramienta gráfica como lo es el histograma. El histograma de los residuos (gráfica 24), aunque aparentan un ligero sesgo hacia la derecha, la distribución de los residuos se aproxima a la normalidad.

**Gráfico 24. Histograma de la distribución de los residuos**



#### **4.1.2 Errores no correlacionados entre si**

Para comprobar la autocorrelación de los residuos se aplica la prueba Breusch - Godfrey en la cual la hipótesis nula que se plantea es que no existe autocorrelación en los residuos hasta determinado orden. Si el valor -p de la prueba es menor al nivel de significancia se rechaza la hipótesis nula, concluyendo así que los residuos están correlacionados. Los resultados de a prueba BG para el modelo que hemos seleccionado se muestran enseguida:

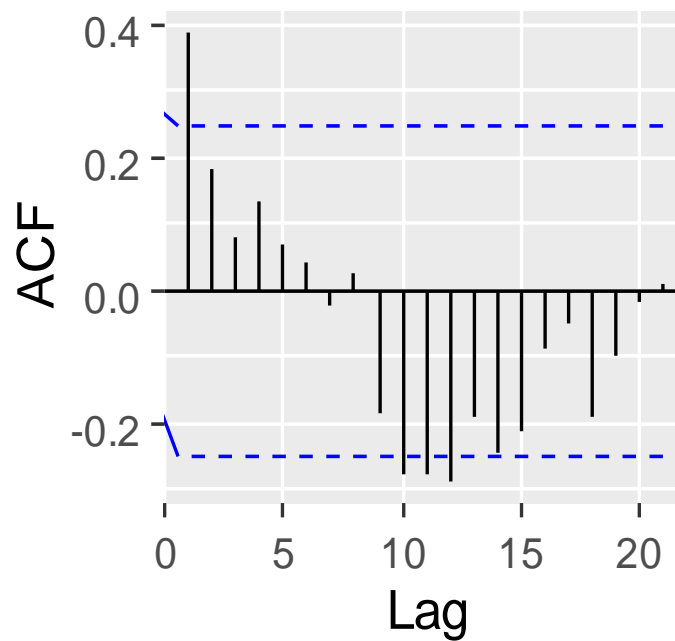
#### **Breusch-Godfrey test for serial correlation of order up to 13**

**data: Residuals from Linear regression model**

LM test = 22.775, df = 13, p-value = 0.04445

Se puede observar en el gráfico 25 de la función de autocorrelación que en el retraso 13 la autocorrelación es considera nula (por los límites de la función de autocorrelación)

**Gráfico 25. Función de autocorrelación de modelo de regresión estática**

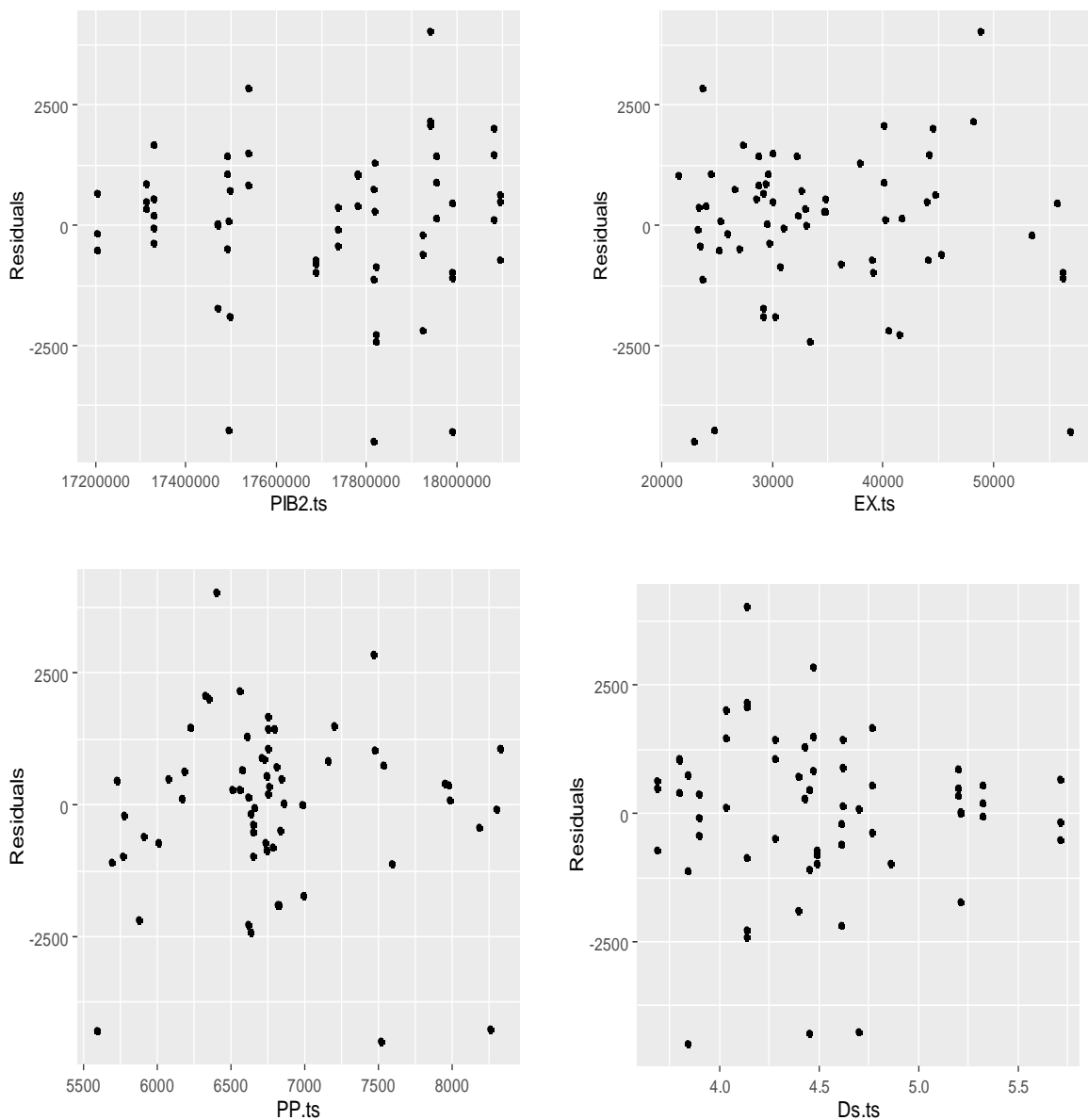


#### 4.1.3 Varianza constante de los errores

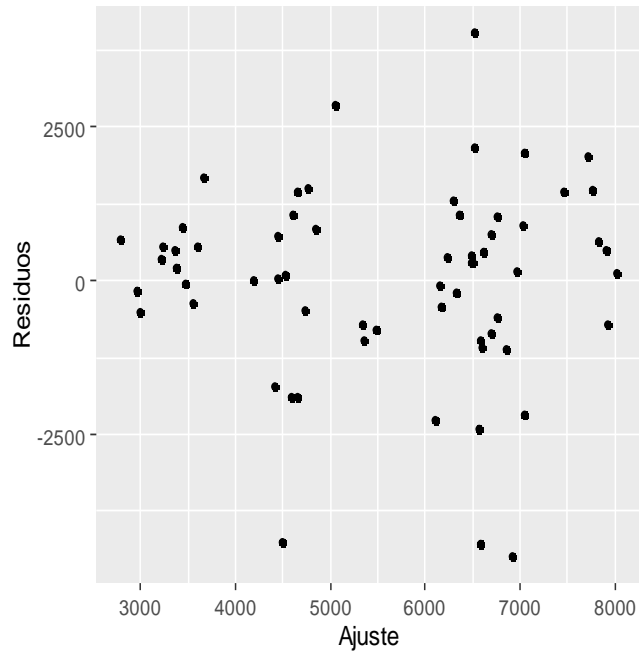
Otro supuesto que se debe revisar es la homocedasticidad de los residuos. Se utilizará los diagramas de dispersión de los residuos frente a cada variable predictora seleccionada y también los residuos frente a los valores ajustados.

En la gráfica 26 se muestran los tres diagramas de dispersión entre los residuos y cada una de las variables, no se observan patrones y presentan apariencia de dispersarse al azar. La distribución de los residuos frente a los valores ajustados también es importante para comprobar homocedasticidad (varianza constante), al igual que sucede con las variables predictoras, la dispersión de los errores frente a los valores ajustados no debe tener un patrón sino más bien debe estar sujeto al azar. Podemos observar en la gráfica 27 que los residuos no registran patrones definidos con los valores ajustados.

**Gráfico 26. Diagrama de dispersión de residuos frente a cada variable predictora**



**Gráfico 27. Diagrama de dispersión de residuos frente a valores ajustados**



#### **4.1.4 Multicolinealidad**

Comprobar que no existan dependencia lineal entre las variables predictoras del modelo es lo que se abordará en este punto de la investigación. Una de las formas para identificar existencia de la dependencia de las variables es por medio de los coeficientes de correlación simple. Como puede observarse en la tabla 1 para las variables consideradas en el modelo propuesto (PIB, exportaciones y precio promedio) no existen coeficientes de correlación significativos (mayores a 0,90 en valor absoluto) lo cual es indicativo de ausencia de multicolinealidad (Hernández Alonso & Zúñiga Rodríguez , 2013).



## 4.2 Predicción a partir del modelo

Verificados los supuestos del modelo, nos enfocaremos en la capacidad predictiva del mismo. Como ya se había mencionado en el capítulo tres, el modelo estático que utiliza tres variables presenta los mejores valores de MAE en comparación a los otros modelos planteados. A continuación, se muestran los pronósticos obtenidos con los datos de prueba.

**Tabla 5. Pronóstico de 5 periodos de prueba**

Mes	Pronóstico	Real	Límite Inferior	Límite Superior
Agosto 2019	7086	7752	3706	10465
Septiembre 2019	7475	6752	4144	10807
Octubre 2019	7025	6925	3702	10349
Noviembre 2019	6488	6308	2919	10058
Diciembre 2019	7309	10863	4002	10617

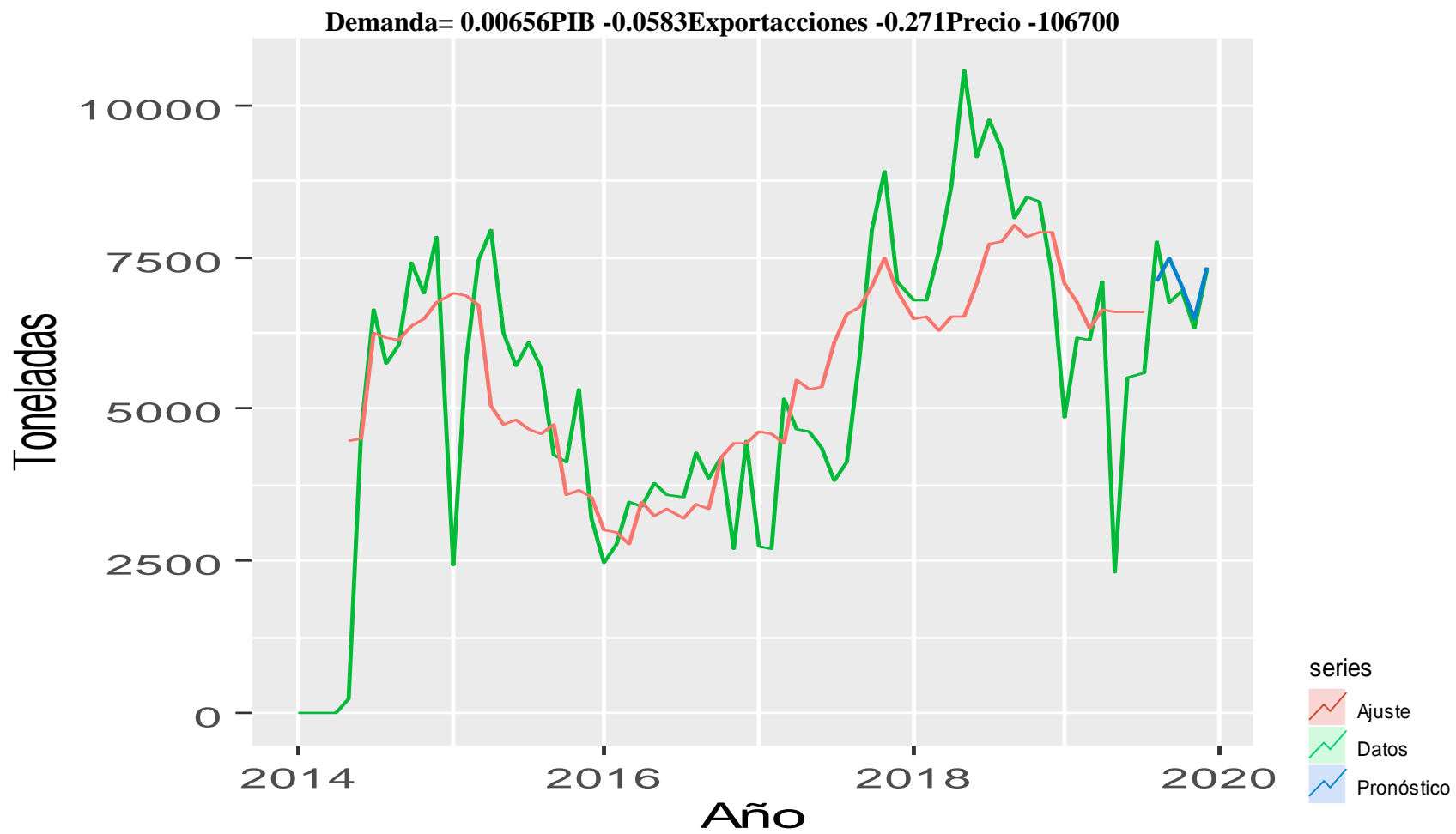
Como podemos notar en la tabla 5, los valores reales de la demanda de los meses de agosto a noviembre de la referencia 65, se ubican dentro del intervalo de predicción al 95% de confianza. La demanda real de diciembre es mayor al límite superior del intervalo de predicción e inclusive dicho valor (10863 toneladas) es mayor al promedio de los meses de diciembre de los dos últimos años. La explicación de esto se debe a que la organización

realizó una venta adelantada de 3583 toneladas que correspondía al mes de enero 2020, de ahí que en el mes de diciembre se haya inflado el valor de la demanda.

Al no considerar el valor de la compra adelantada, la demanda real (7280 toneladas) de dicho mes se ubica entre los límites de predicción del modelo. Así mismo podemos observar que el 25% de los periodos pronosticados fueron subestimados mientras que el 75% son sobreestimados.

Para los meses de noviembre y diciembre los datos reales presentan incremento con respecto al mes precedente, el modelo de pronóstico captura ese comportamiento en dichos meses. En la gráfica 28 se traza los datos reales, ajustados y pronosticados obtenido con el modelo estático.

Gráfico 28. Pronóstico de demanda del alimento referencia 65



## **CAPITULO 5**

### **CONCLUSIONES Y RECOMENDACIONES**

#### **5.1 Conclusiones**

La previsión de la demanda de alimento balanceado puede tener un gran impacto en la cadena de suministro del mercado acuícola del país. La serie de tiempo de la demanda de balanceado presenta tendencia fuerte y variaciones estacionales lo que los convierte en un desafío en el desarrollo de pronósticos efectivos.

El propósito de esta investigación es hallar un modelo que genere el pronóstico más ajustado a la realidad. Se han planteado varios enfoques como lo son el modelo autorregresivo integrado-media móvil ARIMA (por sus siglas en inglés), modelos de regresión estática y los que incluyen errores ARIMA. Los datos de entrenamiento fueron utilizados para ajustar un modelo de acuerdo con el enfoque. Así para identificar un modelo ARIMA apropiado, se realizaron las transformaciones necesarias para la estabilización de varianza y de la media, se aplicaron prueba de raíz unitaria con el fin de comprobar la estacionariedad del proceso. Definiendo los valores  $p$  y  $q$  de acuerdo con las gráficas PACF (función de autocorrelación parcial) y ACF (función de autocorrelación) y con la ayuda

del software se analizaron otros modelos con variantes en los valores p y q se pudo identificar al que mejor se ajusta a los datos.

En los modelos de regresión estática se utilizó la estrategia de eliminación hacia atrás. Para seleccionar las variables que se incluirían en el modelo, la selección se basó a su significancia a través del valor p. Los modelos dinámicos con errores ARIMA se planteó un modelo con los mismos valores p y q, que en la metodología ARIMA obtuvieron los mejores ajustes. Se probaron todas las posibles combinaciones de las variables para incluirlas en un modelo.

En cada modelo se determinó medidas de precisión de ajuste y de predicción, entre ellas se destaca el MAE (error absoluto medio) y RMSE (raíz de error cuadrático medio). Al comparar los mejores modelos de cada uno los métodos utilizados tenemos que el RMSE de los datos de entrenamiento de modelo dinámico con cuatro variables es 1449.41, dicho valor es menor al reportado por el ARIMA (0,1,1) y al modelo estático con cuatro variables.

En cuanto al MAE de los datos de entrenamiento, se observa nuevamente que el modelo dinámico con 4 variables es de 1007.91, en los modelos ARIMA el mejor tiene un MAE de 1086.43 y el modelo estático con tres variables es de 1129.44 además cabe resaltar que la superioridad de ajuste de los modelos dinámico está marcada sobre los otros métodos que aún el modelo con el peor ajuste los supera.

En lo que concierne a las medidas de pronóstico referidas a los datos de prueba, la balanza se inclina por los modelos de regresión estática. De los cinco modelos que tienen los RMSE más pequeños, cuatro pertenecen a la familia de modelos estáticos y uno al grupo de modelos dinámicos con 2 variables (precio y desempleo). Por su parte el MAE presenta el mismo comportamiento, los cuatro modelos que reportan los valores mínimos pertenecen al grupo de regresión estática.

Los resultados nuestra que existen modelos con grandes niveles de ajuste pero que los mismos no son necesariamente los mejores en pronosticar, por ejemplo, tenemos el caso del modelo dinámico con cuatro predictoras el ocupa el primer lugar en cuanto a valores en mínimos de MAE Y RMSE, pero en lo referente a pronóstico es el sexto mejor.

En contraste tenemos al modelo estático con tres variables predictoras (PIB, exportaciones y precio) que con base al MAE de datos de prueba es el que da mejores pronósticos y en cuanto al RMSE es el segundo mejor modelo, pero es todo lo contrario en los ajustes ya que en referencia a dicha métricas ocupa el lugar 14 y 15 respectivamente.

Dado que el presente trabajo de investigación está orientado al pronóstico nos inclinamos por el modelo estático con tres variables predictoras para la previsión de la demanda del alimento de referencia 65. Los residuos (datos de entrenamiento) de este modelo están distribuidos normalmente, no están correlacionados entre ellos y tiene una varianza constante tal como se demostró en apartado en capítulo 4. Los errores de pronóstico de los primeros dos meses son altos, pero a medida que se avanza en el tiempo

los mismos van reduciéndose. Este modelo tiene un porcentaje de error promedio del 4.8% el cual es aceptable de acuerdo con la literatura revisada.

## **5.2 Recomendaciones**

El modelado de pronóstico debe ser implementado en todos los productos comercializados por la empresa lo cual servirá como entrada a procesos relacionados directamente como la programación de la producción, definición de niveles de inventarios, etc.

El pronóstico de la demanda no debe ser exclusiva a una persona o departamento sino más bien debe ser un proceso multidisciplinar para poder recabar información de todas las aristas de las distintas funciones: técnicas, comerciales, fabriles y logísticas.

El disponer de abundante información relacionada a la industria del balanceado podrá contribuir a la elaboración de modelos de pronósticos más eficientes, por lo cual se debe robustecer su sistema de almacenamiento y procesamiento de información de variables.

Los modelos propuestos son de corte clásico, en futuros trabajos se podría aplicar métodos avanzados de inteligencia artificial para el pronóstico, por lo tanto, se debe contar con un programa de capacitación al personal de planificación en estas herramientas.

## BIBLIOGRAFÍA

- Aguirre Lasprilla, S., Ardila Rueda, W., Figueroa, L., & Romero Rodríguez, D. H. (2015). Parametrización y evaluación de Política de Inventario (s,Q) en Hospitales: Un caso de estudio en la ciudad de Barranquilla. *Prospectiva*, 99-105.
- Ballé, M., Jones, D., Chaize, J., & Fiume, O. (2018). *Estrategia Lean*. España: Profit.
- Bernabé Argandona, L. (2016). Sector Camaronero: Evolución y proyección a corto plazo. *FCSHOPINA*, 1-6.
- Betancurt Aristizabal, M. A., Fandiño Castillo, S., Gutierrez Ortegón, I. M., & Ospina Tenorio, C. (2018). *Diseño de un sistema de control de inventarios de productos para la empresa FERRONEUMATICA LTDA*. Cali.
- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time Series and Forecasting*. New York: Springer .
- Castro Romero, N. A. (2015). *Diagnóstico y propuesta de mejora en la gestión de inventarios y distribución de almacén en una importadora de juguetes aplicando el modelo Score y herramientas de pronóstico* . Lima.
- Chase , R. B., Jacobs, R. F., & Aquilano , N. J. (2009). *Administracion de Operaciones Produccion y Cadena de Suministro*. Mexico D.F: McGraw -Hill.
- Collen , C., & George, E. P. (2003). *Demand Management Best Practices: Process,Principles and Collaboration*. Boca Raton-Florida: J. Ross Publishing .
- Cortes Gonzalez, D. F. (2019). *Diseño y formulación de un sistema de gestión de inventarios para una compañía colombiana que comercializa productos tecnológicos con ciclos de vida corto*. CALI: PONTIFICIA UNIVERSIDAD JAVERIANA.
- Eric Schmidt , & Jonathan Rosenberg. (2016). *Cómo trabaja Google*. Bogotá D.C.: Penguin Random House Grupo Editorial.
- ESPAE Graduate School of Management. (2018). *Estudios Industriales: Orientación estratégica para la toma de decisiones – Industria de Acuicultura*. Guayaquil.
- Evans, J. R., & Lindsay, W. M. (2014). *Administracion y control de la calidad*. Mexico: Cengage Learning.
- Hai, D., Hao, Z., & Ping, L. (2011). Model Predictive Control for inventory Management in Supply Chain Planning. *ELSEVIER*, 1154-1159.
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Stanford,California: Springer.
- Hernández Alonso, J., & Zúñiga Rodríguez , J. (2013). *Modelos econométricos para el análisis económico*. Mdríd: ESIC.



- Hyndnan, R. J., & Athanasopoulos, G. (2018). *Forecasting Principles and Practice*. Melbourne, Australia: OTexts.
- ISO 9001. (2015). *Sistema de Gestión de Calidad: Requisitos*.
- ISO, 9. (2015). *Sistema de Gestión de la Calidad - Requisitos*.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning With Applications in R*. New York : Springer Science +Business Media .
- Krajewski, L. L., Ritzman, L. P., & Malhotra, M. K. (2008). *Administración de operaciones Procesos y cadena de valor*. Mexico : Pearson Prentice Hall.
- Kristian , L., & William , I. (2016). To predict and serve? *Significance*, 14-19.
- Kuhn, M., & Kjell, J. (2013). *Applied Predictive Modeling*. New York: Springer Science.
- Lizarzaburu, G. (23 de Agosto de 2018). Dos años duros para el camarón. *Diario Expreso*.
- Mora, L. A. (2018). *Principios y Máximas de la cadena de suministro*. Bogotá: Ecoe Ediciones.
- Moscoso, P., & Lago, A. (2016). *Gestión de operaciones para directivos*. Madrid: McGraw-Hill.
- Pierre d' Huy, & Jerome Lafon. (2018). *Innovación para dummies*. Bogota: Planeta Colombiana.
- Porter, M. (1991). *La ventaja competitiva de las naciones*. Buenos Aires.
- Revista Líderes. (2014). La industria nacional de camarón reflató con fuerza. *Líderes*.
- Salazar Aguilar, M. A., & Cabrera Ríos, M. (2007). Pronostico de demanda por medio de redes neuronales artificiales. *Ingenierías* , 6-12.
- Télez Piñerez, C., & Morales Rivera , M. A. (2016). *Modelos Estadísticos lineales con aplicaciones en R*. Bogota: Ediciones de la U.
- Tundidor, A., Hernández, E., Peña, C., Martínez , J., Campos, J., & Hernández, C. (2019). *Cadena de suministro 4.0 Beneficios y retos de las tecnologías disruptivas*. Bogota : Alfaomega Colombiana S.A.
- Wickham , H., & Golemund , G. (2017). *R for data Science Visualize, Model, Transform, Tidy and Import Data*. O'Reilly.
- Yan, P., Chen Zheng , R., Zhang , J., & Xin, Y. (2019). Predicción de la demanda de bicicletas compartidas utilizando redes neuronales recurrentes. *ELSEVIER*, 73-86.

## ANEXOS

### Códigos utilizados en software libre R-Studio

```
# se realiza la prueba KPSS  
dataar1 %>% ur.kpss() %>% summary
```

```
# Se procede con la transformación logaritmica .  
cbind("Toneladas" = dataar1,  
      "Log Tonelada"= log(dataar1)  
)%>%  
  autoplot(facets=TRUE) +  
  xlab("Año") + ylab("") +  
  ggtitle("Demanda de la referencia 65")
```

```
# Aplicando ARIMA  
dataar65<-window(balan65_ts,start = c(2014,5),end = c(2019, 7) )  
dataatra<- log(dataar65)  
salidajust<- Arima (dataatra,order = c(1,1,1))
```

```
ajuste<-fitted(salidajust)  
exp(ajuste)  
CV(salidajust)  
accuracy( exp(ajuste),exp(dataatra))
```

```

autoplot(DD.ts, series="Datos") +
  autolayer(exp(ajuste), series="Ajuste") +
  autolayer(exp(salidapredd$mean), series="Pronostico")
xlab("Año") + ylab("Toneladas") +
  ggtitle("Quarterly Beer Production")

```

```

# usando el algoritmo de R

automatico<- auto.arima(myts,trace = TRUE,seasonal = TRUE)

mod2<- Arima(myts, order=c(0,1,1),lambda = 0)
summary(mod2)
## AJUSTE MODELO 2
fcast2<-fitted(mod2)
plot(myts,xlab="Años",ylab="Toneladas",type="l",ylim=c(1000,12000))
lines(fcast2,lwd=1,col="blue")
# PRONOSTICO MODELO 2
mod2 %>% forecast(h=5,level=c(99.5)) %>% autoplot(ylim=c(0,25000))
myts%>% Arima(order=c(0,1,1),lambda = 0)%>% forecast(h=6)%>%
  autoplot(ylim=c(0,25000))+ylab("Toneladas")+xlab("Año")

prono2<-forecast(mod2,h=5,level=c(99.5))
summary(prono2)
plot(prono2,ylim=c(3000,25000))

# METRICAS DE EVALUACIÓN

accuracy( forecast(mod2,h=5),datatest)
accuracy( prono2,datatest)

round(accuracy( prono2,datatest),digits = 2)

```

```

# Modelo de Regresión
mrst1<-tslm(DD.ts~PIB2.ts+PP.ts+EX.ts+Ds.ts,data = modelmulent)
summary(mrst1)
library(modelr)
data.frame(
  R2 = rsquare(mrst1,data = modelmulent),
  RMSE = rmse(mrst1, data = modelmulent),
  MAE = mae(mrst1, data = modelmulent)
)

```

```

# ajuste
bb<-fitted(mrst1)
plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(bb,lwd=2,col="blue")
# Pronóstico
pronolin<-forecast(mrst1,data.frame(modelmultest[,2:5]),level=c(95))
autoplot(pronolin)

```

```

# tres gráficas
autoplot(DD.ts, series="Datos") +
  autolayer(bb, series="Ajuste") +
  autolayer(pronolin, series="Pronostico")
xlab("Año") + ylab("Toneladas") +
  ggtitle("Quarterly Beer Production")

```

```

# Modelo 2
mrst2<-tslm(DD.ts~PIB2.ts+EX.ts+PP.ts,data = modelmulent)
summary(mrst2)# Adjusted R-squared: 0.457
data.frame(
  R2 = rsquare(mrst2,data = modelmulent),
  RMSE = rmse(mrst2, data = modelmulent),
  MAE = mae(mrst2, data = modelmulent)
)

```

```

# residuos vs ajuste
cbind(Ajuste = fitted(mrst2),
      Residuos=residuals(mrst2)) %>%
  as.data.frame() %>%
  ggplot(aes(x=Ajuste, y=Residuos)) + geom_point()

```

```

autoplot(DD.ts, series="Datos") +
  autolayer(ba, series="Ajuste") +
  autolayer(pronolin2, series="Pronóstico")+
xlab("Año") + ylab("Toneladas")+
  ggtitle("Quarterly Beer Production")
round(accuracy( pronolin2,datatest),digits = 2)

```

```

mrst3a<-tslm(modelmulent[, "DD.ts"]~PIB2.ts+EX.ts,data = modelmulent)
summary(mrst3a)# Adjusted R-squared: 0.457
data.frame(
  R2 = rsquare(mrst2,data = modelmulent),
  RMSE = rmse(mrst2, data = modelmulent),
  MAE = mae(mrst2, data = modelmulent)
)

ba3<-fitted(mrst3a)
plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(ba3,lwd=2,col="red")
vif(mrst3a)
pronolin3a<-forecast(mrst3a,data.frame(modelmulent[,c("PIB2.ts", "EX.ts")],level=c(95)))

autoplot(pronolin2a)
accuracy( pronolin3a,datatest)

```

*# modelo pib + exportaciones*

```

mrst3<-tslm(DD.ts~PIB2.ts+EX.ts,data = modelmulent)
summary(mrst3) #Adjusted R-squared: 0.4643
data.frame(
  R2 = rsquare(mrst3,data = modelmulent),
  RMSE = rmse(mrst3, data = modelmulent),
  MAE = mae(mrst3, data = modelmulent)
)

# R2      RMSE      MAE
# 0.4815606 1551.12 1130.275
CV(mrst3)
b3<- fitted(mrst3)

plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(b3,lwd=2,col="orange")
ac<- modelmulent[,2:4]
pronolin3<-forecast(mrst3,data.frame(ac[,-2]),level = c(95))
autoplot(pronolin3)+autolayer(b3)

accuracy( pronolin3,datatest)

round(accuracy( pronolin3,datatest),digits = 2)

```

```

# modelo pib

mrst4<-tslm(DD.ts~PIB2.ts,data = modelmulent)
summary(mrst4) # Adjusted R-squared: 0.08967
data.frame(
  R2 = rsquare(mrst4,data = modelmulent),
  RMSE = rmse(mrst4, data = modelmulent),
  MAE = mae(mrst4, data = modelmulent)
)

# R2      RMSE      MAE
# 0.104355 2038.752 1755.012

CV(mrst4)

b4<- fitted(mrst4)
plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(b4,lwd=2,col="purple")

ac<- modelmultest[,2]
pronolin4<-forecast(mrst4,data.frame(ac),level=c(95),h=5)
autoplot(pronolin4)
accuracy( pronolin4,datatest)

round(accuracy( pronolin4,datatest),digits = 2)

```

```

# Regresión con errores arima (0,1,1)
fit<-Arima(modelmulent[,"DD.ts"],
           xreg= modelmulent[,2:5],order = c(0,1,1))

summary(fit)

# todas las variables
m6<- fitted(fit)
plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(m6,lwd=2,col="blue")

pronolin6<- forecast(fit,xreg= modelmulent[,2:5],h=5,level=c(95.0))
accuracy( pronolin6,datatest)

round(accuracy( pronolin6,datatest),digits = 2)

# Regresión con errores arima (0,1,1)
fit<-Arima(modelmulent[,"DD.ts"],
           xreg= modelmulent[,2:5],order = c(1,1,1))

summary(fit)

```

```

#### MODELO PIB,PRECIO Y EXPORTACIONES

fit2<-Arima(modelmulent[,"DD.ts"],
            xreg= modelmulent[,2:4],order = c(0,1,1))

summary(fit2)
pronolin7<- forecast(fit2,xreg= modelmulent[,2:4],h=5,level=c(95.0))
round(accuracy( pronolin7,datatest),digits = 2)

m7<- fitted(fit2)
plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(m7,lwd=2,col="green")

```

```

# MODELO PIB Y DESEMPLEO
X<-modelmulent[,-3]
head(X)

X.1<- X[,-3]
fitX.1<-Arima(X.1[, "DD.ts"],|
              xreg= X.1[,2:3],order = c(0,1,1))

summary(X.1)
pronolinX<- forecast(fitX.1,xreg= X.1[,2:3],h=5,level=c(95.0))
round(accuracy( pronolinX,datatest),digits = 2)

```

```

# MODELO DESEMPLEO PRECIO Y EXPORTACIONES
fDPE<-Arima(modelmulent[, "DD.ts"],
            xreg= modelmulent[,3:5],order = c(0,1,1))

summary(fDPE)
pronolinfDPE<- forecast(fDPE,xreg= modelmulent[,3:5],h=5,level=c(95.0))
round(accuracy( pronolinfDPE,datatest),digits = 2)

```

```

# MODELO PRECIO Y DESEMPLEO
PD<-modelmulent[,3:5]
PD.<- PD[,-2]
fPD<-Arima(modelmulent[, "DD.ts"],
            xreg= PD.,order = c(0,1,1))

summary(fPD)
pronolinfPD<- forecast(fPD,xreg= PD.,h=1,level=c(95.0))
round(accuracy( pronolinfPD,datatest),digits = 2)

plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(fitted(fPD),lwd=2,col="Red")

autoplot(pronolinfPD)

```

```

# EXPORTACIONES Y DESEMPLEO
fDE<-Arima(modelmulent[, "DD.ts"],
            xreg= modelmulent[,4:5],order = c(0,1,1))

summary(fDE)
pronolinfDE<- forecast(fDE,xreg= modelmulent[,4:5],h=5,level=c(95.0))
round(accuracy( pronolinfDE,datatest),digits = 2)

autoplot(pronolinfDE)

```

```

# MODELO DESEMPLEO
fD<-Arima(modelmulent[, "DD.ts"],
            xreg= modelmulent[,5],order = c(0,1,1))

summary(fD)
pronolinfD<- forecast(fD,xreg= modelmulent[,5],h=5,level=c(95.0))
round(accuracy( pronolinfD,datatest),digits = 2)

```



```

# MODELO PRECIO
fP<-Arima(modelmulent[, "DD.ts"],
          xreg= modelmulent[,3],order = c(0,1,1))
summary(fP)
pronolinFP<- forecast(fP,xreg= modelmulent[,3],h=5,level=c(95.0))
round(accuracy( pronolinFP,datatest),digits = 2)

```

```

# MODELO PIB + PRECIO |
fit3<-Arima(modelmulent[, 'DD.ts'],
            xreg= modelmulent[,2:3],order = c(0,1,1))
head(modelmulent[,2:3])
summary(fit3)
pronolin8<- forecast(fit3,xreg= modelmulent[,2:3],h=5,level=c(95.0))
round(accuracy( pronolin8,datatest),digits = 2)

m8<- fitted(fit3)
plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(m8,lwd=2,col="green")

```

```

# MODELO CON PIB
fit4<-Arima(modelmulent[, "DD.ts"],
            xreg= modelmulent[,2],order = c(0,1,1))
head(modelmulent[,2])
summary(fit4)
m9<- fitted(fit4)
plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(m9,lwd=2,col="green")
pronolin9<- forecast(fit4,xreg= modelmulent[,2],h=5,level=c(95.0))
round(accuracy( pronolin9,datatest),digits = 2)

```

```

# MODELO CON EXPORTACIONES
fit5<-Arima(modelmulent[, "DD.ts"],
            xreg= modelmulent[,4],order = c(0,1,1))
summary(fit5)

m10<- fitted(fit5)
plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(m10,lwd=2,col="orange")
pronolin10<- forecast(fit5,xreg= modelmulent[,4],h=5,level=c(95.0))
round(accuracy( pronolin10,datatest),digits = 2)

```

```

# MODELO PIB + EXPORTACIONES |
fit3.1<-Arima(modelmulent[, "DD.ts"],
             xreg= (modelmulent[,c("PIB2.ts", "EX.ts")]),order = c(0,1,1))
summary(fit3.1)
m3.1<-fitted(fit3.1)

plot(modelmulent[, 'DD.ts'],xlab="Años",ylab="Toneladas",type="l")
lines(m3.1,lwd=2,col="red")

fcastt <- forecast(fit3.1,xreg= (modelmulent[,c("PIB2.ts", "EX.ts")]),5,level=c(95.0))
round(accuracy( fcastt,datatest),digits = 2)

```