

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE MATEMÁTICAS**

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

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

TEMA:

**PRONÓSTICO DE LA DEMANDA EN UNA EMPRESA
COMERCIALIZADORA DE ADITIVOS PARA LUBRICANTES**

AUTOR:

MARÍA ROSA REYES ACOSTA

Guayaquil - Ecuador

2020

RESUMEN

El presente estudio se llevó a cabo dentro de una empresa importadora y comercializadora de aditivos ubicada en la ciudad de Guayaquil, Ecuador.

Teniendo en cuenta que las ventas de aditivos para la fabricación de aceites lubricantes de motores alimentados por diésel, gasolina u otros combustibles, representan cerca del 90% de la facturación de la empresa, la planificación de las importaciones se convierte en un factor crítico para lograr inventarios ajustados acordes a la capacidad física y logística de la empresa de acuerdo a los requerimientos mensuales de los clientes, permitiendo que el capital de trabajo pueda mantenerse en movimiento. Por lo anteriormente expuesto, se tiene como objetivo determinar modelos estadísticos que ajusten de forma adecuada a la demanda de aditivos de los diferentes segmentos, con el fin de realizar predicciones a corto plazo.

Los datos en toneladas fueron recopilados de los registros de ventas de la empresa comercializadora de aditivos a través del sistema XASS durante el periodo 2014-2019.

Los resultados obtenidos durante la presente investigación son modelos univariantes ARIMA utilizando la metodología Box-Jenkins y el Software R para pronosticar: ventas de totales de los aditivos, ventas del aditivo para la fabricación de lubricantes "Fuera de Borda", ventas de aditivos para la fabricación de lubricantes alimentados por "Gasolina" y ventas de aditivos para la fabricación de lubricantes alimentados por "Diésel".

PALABRAS CLAVE: Box-Jenkins, ARIMA, aditivos, predicción, lubricantes.

ABSTRACT

Present investigation was carried out in additive importing and trading company located in Guayaquil, Ecuador. Taking into account that sales of additives for the manufacture of lubricating oils for diesel, gasoline or other fuels engines represent close to 90% of the company's income, import planning becomes a critical factor in achieving inventories adjusted to the physical and logistical capacity of the company considering customers' monthly requirements, allowing working capital to be kept in motion. For this reason, the objective is to determine univariate models that adequately fit the time series that represent the demand for additives in the different segments, in order to make short-term predictions.

The data was collected from the sales records of the additive marketing company through the XASS system during the 2014-2019 periods.

The results obtained during current investigation are univariate ARIMA models using the Box-Jenkins methodology and Software R to forecast: total additives sales, sales of the additive for "Outboard" lubricants manufacturing, sales of additives for lubricants fueled by "Gasoline" manufacturing and sales of additives for lubricants fueled by "Diesel" manufacturing.

KEY WORDS: Box-Jenkins, ARIMA, additives, prediction, lubricants.

DEDICATORIA

Este estudio es dedicado a mi mamá, a la memoria de mi papá, a mis hijos y a mis hermanas, a toda mi familia; quienes no lo leerán, pero quienes han sido mi motivación y apoyo incondicional a lo largo de mi desarrollo como persona y como profesional.

A mis superiores: Ing. Andrés Ballerino, Ing. Ana Armas, Lcda. Jessel Burgos y a mi jefe; el Ing. Manuel Fernández Freire, quienes en el año 2016 decidieron confiar en una novata y sin saberlo cambiaron la vida de una madre y sus dos hijos.

A mis compañeros de trabajo y a mis amigos, quienes, con cada café, cada momento y cada consejo, permiten que toda dificultad que se presente pueda ser superada y todo momento feliz sea digno de ser recordado o repetido.

A Jesús, quien entregó su vida por nosotros y gracias a Él somos salvos; que para Él sea siempre toda honra y toda gloria.

AGRADECIMIENTO

A través del presente escrito quisiera agradecerle primero a Dios por hacer todo posible.

A mi mamá, por haberme dicho que estudie esta maestría mientras ella cuidaba a mis hijos y por impulsarme siempre hacia adelante a pesar de cualquier inconveniente.

Gracias a María De Los Ángeles, mi hermana menor, mi mejor amiga, mi socia, mi mano derecha y en momentos difíciles también mi mano izquierda, mi cabeza, mis pies, todo. Gracias por tus palabras de aliento, por tus exhortaciones y por no permitir que me rinda jamás.

Gracias a Elisa y Alexa, mis hermanas mayores, quienes siempre han sido ejemplo de generosidad y fortaleza. Gracias por siempre preocuparse por mí y por mis hijos.

Quisiera agradecerle también a cada integrante de la empresa Hidromecánica Andina C.Ltda., porque me enseñaron a que ser feliz en tu lugar de trabajo es posible y que tus compañeros se transforman con el tiempo en tu familia, velando siempre por el bienestar de todos.

Al Ing. Andrés Ballerino y a la Ing. Ana Armas, porque a pesar de que me conocían por apenas 9 meses me apoyaron desde el inicio de mi preparación tanto académica como profesional.

Gracias a mis profesores, especialmente a la M.Sc. Sandra González y a la memoria de su amada madre, por compartirme sus conocimientos y guía durante estos años de estudio y formación.

Y finalmente, pero no menos importante, a Gabriel y a Benjamín; mis hijos, quienes me recuerdan todos los días que tengo que levantarme a darles desayuno, a llevarlos a la escuela y a quienes no les importa si hay o no feriado. Gracias por mostrarme que lo hermoso de la vida está en los detalles y que los triunfos se celebran con pizza y en familia, porque no hay mejor premio que sus sonrisas.

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.



María Rosa Reyes Acosta

TRIBUNAL DE GRADUACIÓN



M.Sc. John Ramírez Figueroa
PRESIDENTE



M.Sc. Sandra González Camba
DIRECTOR



Mgtr.. Jacobo Mero Herrera
VOCAL 1



Mgtr. Sergio Bauz Olvera
VOCAL 2

ABREVIATURAS O SIGLAS

Abreviatura	Significado
ACF	Función de Autocorrelación
ACF1	Coefficiente de Autocorrelación de Primer Orden
AIC	Criterio de Información de Akaike
AR	Autorregresivo
ARIMA	Modelo Autorregresivo Integrado de Medias Móviles
ARMA	Modelo Autorregresivo de Media Móvil
BIC	Criterio de Información Bayesiano
ERP	Planificación de Recursos Empresariales
KG	Kilogramos
MA	Media Móvil
MAE	Error Medio Absoluto
MAPE	Error Medio Absoluto Porcentual
MASE	Error Medio Absoluto de Escala
MCO	Mínimos Cuadrados Ordinarios
ME	Error Medio
MPE	Error Porcentual Medio
MT	Toneladas Métricas
PACF	Función de Autocorrelación Parcial
RMSE	Error Cuadrático Medio
SE	Error Estándar
TBN	Número Básico Total
VM	Modificador de Viscosidad

TABLA DE CONTENIDO

CAPÍTULO 1	1
1. INTRODUCCIÓN.....	1
1.1. Antecedentes	2
1.2. Descripción del problema.....	3
1.3. Objetivos	4
1.3.1. Objetivo General	4
1.3.2. Objetivos Específicos	4
1.4. Hipótesis	4
1.5. Alcance.....	4
CAPÍTULO 2	5
2. MARCO TEÓRICO.....	5
2.1. Aceite Lubricante.....	5
2.2. Definición de aditivo	5
2.3. Series Temporales	5
2.3.1. Proceso Estocástico	5
2.3.2. Definición de Series Temporales:	6
2.3.3. Estacionariedad	7
2.3.4. Función de correlación y autocorrelación	9
2.3.4.1. Función de Autocorrelación (ACF)	10
2.3.4.2. Prueba de ACF individual	11
2.3.4.3. Ruido Blanco.....	11
2.3.4.4. Procesos Autorregresivos AR(p)	12
2.3.4.5. Procesos de Medias Móviles: MA(q)	12
2.3.4.6. Procesos Autorregresivos de Medias Móviles: ARMA(p,q)	13
2.3.4.7. Modelos ARIMA (p,d,q):.....	14
2.3.4.8. Modelización utilizando ARIMA:	15
2.3.4.9. Contraste de Dickey-Fuller:	16
2.3.4.10. Modelos ARIMA estacionales:	16
2.3.4.11. Modelos ARIMA multiplicativos, ARIMA (p,d,q)x(P,D,Q)s:.....	17
2.3.4.12. Prueba de Shapiro Wilks:.....	17
2.3.4.13. Prueba de Normalidad Jarque-Bera:	17

CAPÍTULO 3	18
3. METODOLOGÍA	18
CAPÍTULO 4	21
4.1.1. Ventas Totales	26
4.1.2. Segmento Diésel.....	26
4.1.3. Segmento Fuera de Borda.....	27
4.1.4. Segmento Gasolina	28
4.2. SEGMENTACIÓN DE PRODUCTOS.....	28
4.3. MODELO: VENTAS TOTALES.....	29
4.3.1. ANÁLISIS EXPLORATORIO DE DATOS.....	29
4.3.2. ANÁLISIS DE TENDENCIA	33
4.3.3. AJUSTE DEL MODELO	35
4.3.3.1. Modelos ARIMA Estacionales (p,d,q)x(P,D,Q)s.....	36
4.3.4. ELECCIÓN DEL MEJOR MODELO	37
4.3.5. PREDICCIÓN UTILIZANDO ARIMA	38
4.4. MODELO: FUERA DE BORDA.....	41
4.4.1. ANÁLISIS EXPLORATORIO DE DATOS.....	41
4.4.2. ANÁLISIS DE TENDENCIA	45
4.4.3. AJUSTE DEL MODELO	47
4.4.3.1. Modelos ARIMA Estacionales (p,d,q)x(P,D,Q)s.....	47
4.4.4. ELECCIÓN DEL MEJOR MODELO	49
4.4.5. PREDICCIÓN UTILIZANDO ARIMA	50
4.5. MODELO: GASOLINA.....	52
4.5.1. ANÁLISIS EXPLORATORIO DE DATOS.....	52
4.5.2. ANÁLISIS DE TENDENCIA	56
4.5.3. AJUSTE DEL MODELO	58
4.5.3.1. Modelos ARIMA Estacionales (p,d,q)x(P,D,Q)s.....	59
4.5.4. ELECCIÓN DEL MEJOR MODELO	60
4.5.5. PREDICCIÓN UTILIZANDO ARIMA	61
4.6. MODELO: DIESEL.....	63
4.6.1. ANÁLISIS EXPLORATORIO DE DATOS.....	63
4.6.2. ANÁLISIS DE TENDENCIA	68
4.6.3. AJUSTE DEL MODELO	70
4.6.3.1. Modelos ARIMA Estacionales (p,d,q)x(P,D,Q)s.....	71
4.6.4. ELECCIÓN DEL MEJOR MODELO	72
4.6.5. PREDICCIÓN UTILIZANDO ARIMA	73

CAPÍTULO 5	77
5. CONCLUSIONES Y RECOMENDACIONES.....	77
5.1. CONCLUSIONES.....	77
5.2. RECOMENDACIONES	78
6. Referencias	79
7. Apéndices y anexos	82
7.1. Base de datos utilizada	82
7.2. Visualizaciones en Power Bi.....	88
7.3. Modelos estudiados	93
7.3.1.1. Modelo AR(p).....	93
7.3.1.1.1. Modelo AR(1)	93
7.3.1.1.2. Modelo AR(2)	94
7.3.1.2. Modelo ARIMA (p,d,q).....	96
7.3.2. Serie Temporal: Fuera de Borda.....	97
7.3.2.1. Modelo AR(p).....	97
7.3.2.1.1. Modelo AR(1)	97
7.3.2.1.2. Modelo AR(3)	99
7.3.2.2. Modelo ARIMA (p,d,q).....	100
7.3.3. GASOLINA.....	102
7.3.3.1. Modelo AR(p).....	102
7.3.3.1.1. Modelo AR(1)	102
7.3.3.1.2. Modelo AR(2)	103
7.3.3.2. Modelo ARIMA (p,d,q).....	105
7.4. DIESEL.....	106
7.4.1. Modelos ARIMA (p,d,q).....	106

LISTADO DE FIGURAS

Figura No.4.1 Participación porcentual por adquisición de aditivos 2018 (%) ...	21
Figura No.4.2 Participación porcentual por adquisición de aditivos 2019(%)	22
Figura No.4.3 Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019	23
Figura No.4.4 Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019. Aditivos: Diésel, Fuera de Borda y Gasolina.	24
Figura No.4.5 Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019. Aditivos: Diésel.....	24
Figura No.4.6. Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019. Aditivo: Fuera de Borda.....	25
Figura No.4.7. Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019. Aditivos: Gasolina.....	25
Figura No.4.8. Ventas totales anuales y mensuales de aditivos. Período: 2014-2019.	26
Figura No.4.9. Ventas totales anuales y mensuales, segmento: diésel. Período: 2014-2019.	26
Figura No.4.10. Ventas totales anuales y mensuales, segmento: fuera de borda. Período: 2014-2019.....	27
Figura No.4.11. Ventas totales anuales y mensuales, segmento: Gasolina. Período: 2014-2019.....	28
Figura No.4.12. Histograma de Ventas Mensuales Totales de Aditivos, período 2014-2019.	31
Figura No.4.13. Serie temporal de las ventas totales mensuales de aditivos expresada en toneladas métricas. Período: 2014-2019.	31
Figura No.4.14. Descomposición de la Serie Temporal de las ventas totales mensuales de aditivos. Período: 2014-2019.....	32
Figura No.4.15. Diagrama de caja de la serie temporal de las ventas totales de aditivos segmentada por mes, en toneladas métricas. Período: 2014-2019.	32
Figura No.4.16. Serie temporal de las ventas totales de aditivos, en toneladas métricas. Período: 2014-2018.	33
Figura No.4.17. Logaritmo de la serie temporal de las ventas totales de aditivos. Período: 2014-2018.....	34
Figura No.4.18. Diferencia del logaritmo de la serie temporal de ventas totales de aditivos. Período: 2014-2018.....	35
Figura No.4.19. ACF y PACF de la serie temporal seriedif. Ventas Totales de Aditivos	35
Figura No.4.20. Gráfico de diagnóstico: ajuste 4. Serie temporal de ventas totales de aditivos.	37

Figura No.4.21. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de ventas totales de aditivos. Período: 2014-2018. ..	38
Figura No.4.22. Pronóstico de ventas totales a través de Arima $(2,0,0) \times (1,1,1)_{12}$, expresado en toneladas métricas. Período: 2014-2019.	39
Figura No.4.23. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de ventas totales de aditivos. Período: 2014-2019. ..	40
Figura No.4.24. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de ventas totales de aditivos. Período: 2019.....	41
Figura No.4.25. Histograma de Ventas Mensuales Totales del aditivo para la fabricación de lubricantes fuera de borda, período 2014-2019.....	43
Figura No.4.26. Serie temporal de las ventas totales mensuales del aditivo para la fabricación de lubricantes fuera de borda, expresada en toneladas métricas. Período: 2014-2019.....	43
Figura No.4.27. Descomposición de la Serie Temporal de las ventas totales mensuales del aditivo para la fabricación lubricantes fuera de borda. Período: 2014-2019. 44	44
Figura No.4.28. Diagrama de caja de la serie temporal de las ventas totales del aditivo fuera de borda segmentada por mes, en toneladas métricas. Período: 2014-2019. 44	44
Figura No.4.29. Serie temporal de las ventas totales del aditivo fuera de borda, en toneladas métricas. Período: 2014-2018.....	45
Figura No.4.30. Logaritmo de la serie temporal de las ventas totales mensuales del aditivo fuera de borda. Período: 2014-2018.....	46
Figura No.4.31. Diferencia del logaritmo de la serie temporal de ventas totales mensuales del aditivo fuera de borda. Período: 2014-2018.	46
Figura No.4.32. ACF y PACF de la serie temporal seriedif. Aditivo Fuera de Borda. 47	47
Figura No.4.33. Gráfico de diagnóstico: ajuste 4. Serie temporal aditivo para Fuera de Borda.....	49
Figura No.4.34. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de las ventas del aditivo fuera de borda. Período: 2014-2018. 50	50
Figura No.4.35. Pronóstico de ventas totales del aditivo fuera de borda a través de Arima $(0,0,0) \times (0,1,1)_{12}$. Período: 2014-2019.	50
Figura No.4.36. Comparación de la serie temporal pronosticada y la serie temporal de datos reales del aditivo fuera de borda. Período: 2014-2019.	51
Figura No.4.37. Comparación de la serie temporal pronosticada y la serie temporal de datos reales del aditivo fuera de borda. Período: 2019.	52
Figura No.4.38. Histograma de Ventas Mensuales Totales de los aditivos: Gasolina 1, Gasolina 2, Gasolina 3 y Gasolina, período 2014-2019.	54
Figura No.4.39. Series temporales de las ventas de los aditivos para la fabricación de lubricantes alimentados por gasolina, expresadas en toneladas métricas. Período: 2014-2019.	55

Figura No.4.40. Descomposición de la Serie Temporal Gasolina. Período: 2014-2019.	55
Figura No.4.41. Diagrama de caja de la serie temporal gasolina, segmentada por mes, en toneladas métricas. Período: 2014-2019.	56
Figura No.4.42. Serie temporal gasolina, en toneladas métricas. Período: 2014-2018.	57
Figura No.4.43. Logaritmo de la serie temporal gasolina. Período: 2014-2018.	57
Figura No.4.44. Diferencia del logaritmo de la serie temporal gasolina.....	58
Figura No.4.45. ACF y PACF de la serie temporal seriedif. Aditivos Gasolina. .	58
Figura No.4.46. Gráfico de diagnóstico: ajuste 4. Serie temporal Aditivo para Gasolina.	60
Figura No.4.47. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de las ventas del segmento gasolina. Período: 2014-2018.	61
Figura No.4.48. Pronóstico de ventas de los aditivos del segmento gasolina a través de Arima $(0,0,0) \times (0,1,1)_{12}$. Período: 2014-2019.	61
Figura No.4.49. Comparación de la serie temporal pronosticada y la serie temporal de datos reales del segmento gasolina. Período: 2014-2019.....	62
Figura No.4.50. Comparación de la serie temporal pronosticada y la serie temporal de datos reales del segmento gasolina. Período: 2019.....	63
Figura No.4.51. Histograma de Ventas Mensuales Totales de los aditivos: Diésel 1, Diésel 2, Diésel 3, Diésel 4 y Diésel 5, período 2014-2019.....	65
Figura No.4.52. Series temporales de las ventas de los aditivos para la fabricación de lubricantes alimentados por diésel, expresadas en toneladas métricas. Período: 2014-2019.	67
Figura No.4.53.Descomposición de la Serie Temporal Diesel. Período: 2014-2019.	67
Figura No.4.54. Diagrama de caja de la serie temporal diésel, segmentada por mes, en toneladas métricas. Período: 2014-2019.	68
Figura No.4.55. Serie temporal diesel, en toneladas métricas. Período: 2014-2018.	69
Figura No.4.56. Logaritmo de la serie temporal diesel. Período: 2014-2018.....	69
Figura No.4.57. Diferencia del logaritmo de la serie temporal diesel. Período: 2014-2018.	70
Figura No.4.58. ACF y PACF de la serie temporal seriedif. Aditivos Diésel.	70
Figura No.4.59. Gráfico de diagnóstico: ajuste 5. Serie Temporal Aditivos para Diesel	72
Figura No.4.60. Comparación de la serie temporal pronosticada arima $(3,0,0) \times (0,1,1)_{12}$ y la serie temporal de datos reales de las ventas del segmento diésel. Período: 2014-2018.....	73
Figura No.4.61. Pronóstico de ventas de los aditivos del segmento diésel a través de arima $(3,0,0) \times (0,1,1)_{12}$. Período: 2014-2019.....	74

Figura No.4.62. Comparación de la serie temporal pronosticada con arima(3,0,0)x(0,1,1) ¹² y la serie temporal de datos reales del segmento diésel. Período: 2014-2019.....	75
Figura No.4.63. Comparación de la serie temporal pronosticada con arima (3,0,0)x(0,1,1) ¹² y la serie temporal de datos reales del segmento diésel. Período: 2019.....	76

LISTADO DE TABLAS

Tabla No. 4.1. Participación porcentual por adquisición de aditivos	22
Tabla No. 4.2. Comparación de ventas de aditivos 2018-2019.....	23
Tabla No. 4.3. Aplicación de Pareto a las ventas totales de aditivos del año 2019, expresadas en kilogramos.	28
Tabla No. 4.4. Serie histórica mensual de las ventas totales de aditivos para la fabricación de lubricantes expresadas en kilogramos, periodo 2014-2019.....	29
Tabla No. 4.5. Serie histórica mensual de las ventas totales de aditivos para la fabricación de lubricantes expresadas en toneladas métricas, periodo 2014-2019.	30
Tabla No. 4.6. Resumen Estadístico - Orden $(2,0,0) \times (1,1,1)_{12}$,	36
Tabla No. 4.7. Estadístico Z - Orden $(2,0,0) \times (1,1,1)_{12}$,	37
Tabla No. 4.8. Prueba Box-Ljung- Orden $(2,0,0) \times (1,1,1)_{12}$,	37
Tabla No. 4.9. Comparación de modelos resultantes. Serie temporal ventas totales de aditivos.	38
Tabla No. 4.10. Pronósticos resultantes del ajuste 4 (Periodo 2019). Serie temporal ventas totales	39
Tabla No. 4.11. Comparación de ventas pronosticadas y reales durante el año 2019. Serie temporal ventas totales de aditivos.....	40
Tabla No. 4.12. Serie histórica mensual de las ventas del aditivo para la fabricación de lubricantes "Fuera de Borda" expresadas en kilogramos, periodo 2014-2019.....	41
Tabla No. 4.13. Serie histórica mensual de las ventas del aditivo para la fabricación de lubricantes "Fuera de Borda" expresadas en toneladas métricas, periodo 2014-2019.....	42
Tabla No. 4.14. Resumen Estadístico - Orden $(0,0,0) \times (0,1,1)_{12}$,	48
Tabla No. 4.15. Estadístico Z - Orden $(0,0,0) \times (0,1,1)_{12}$,	48
Tabla No. 4.16. Prueba Box-Ljung- Orden $(0,0,0) \times (0,1,1)_{12}$,.....	49
Tabla No. 4.17. Comparación de modelos resultantes. Serie temporal Aditivo Fuera de Borda.	49

Tabla No. 4.18. Pronósticos resultantes del ajuste 4 (Periodo 2019). Serie temporal fuera de borda.....	51
Tabla No. 4.19. Comparación de ventas pronosticadas y reales durante el año 2019. Serie temporal aditivo fuera de borda.	52
Tabla No. 4.20. Serie histórica mensual de las ventas de los aditivos para la fabricación de lubricantes para motores alimentados por "Gasolina", expresadas en kilogramos, periodo 2014-2019.....	53
Tabla No. 4.21. Serie histórica mensual de las ventas de los aditivos para la fabricación de lubricantes para motores alimentados por "Gasolina" expresadas en toneladas métricas, periodo 2014-2019.	53
Tabla No. 4.22. Resumen Estadístico - Orden $(0,0,0) \times (0,1,1)_{12}$,	59
Tabla No. 4.23. Estadístico Z - Orden $(0,0,0) \times (0,1,1)_{12}$,	59
Tabla No. 4.24. Prueba Box-Ljung- Orden $(0,0,0) \times (0,1,1)_{12}$,.....	60
Tabla No. 4.25. Comparación de modelos resultantes. Serie temporal gasolina. .	60
Tabla No. 4.26. Pronósticos resultantes del ajuste 4 (Periodo 2019). Serie temporal gasolina.....	62
Tabla No. 4.27. Comparación de ventas pronosticadas y reales durante el año 2019. Serie temporal gasolina.....	63
Tabla No. 4.28. Serie histórica mensual de las ventas de los aditivos para la fabricación de lubricantes para motores alimentados por "Diésel", expresadas en kilogramos, periodo 2014-2019.....	64
Tabla No. 4.29. Serie histórica mensual de las ventas de los aditivos para la fabricación de lubricantes para motores alimentados por "Diésel" expresadas en toneladas métricas, periodo 2014-2019.	64
Tabla No. 4.30. Resumen Estadístico - Orden $(3,0,0) \times (0,1,1)_{12}$,	71
Tabla No. 4.31. Estadístico Z - Orden $(3,0,0) \times (0,1,1)_{12}$,	72
Tabla No. 4.32. Prueba Box-Ljung- Orden $(3,0,0) \times (0,1,1)_{12}$,.....	72
Tabla No. 4.33. Comparación de modelos resultantes. Serie temporal diesel.....	73
Tabla No. 4.34. Matriz de pronósticos Ajuste 5. Serie temporal diesel.	74
Tabla No. 4.35. Comparación de ventas pronosticadas con arima $(3,0,0) \times (0,1,1)_{12}$ y ventas reales durante el año 2019. Serie temporal diesel.....	75
Tabla No. 4.25. Prueba Box-Ljung- Orden $(1,0,0)$, Serie Temporal Ventas Totales	94

CAPÍTULO 1

1. INTRODUCCIÓN

Los modelos de predicción para la planificación de importación de productos e inventarios, en el presente los costos de inventarios excesivos reducen el capital de trabajo que las empresas comercializadoras pueden utilizar en inversiones o nuevas oportunidades de negocios. A través de la utilización de modelos Box-Jenkins se consiguieron modelos de predicción que facilitan la toma de decisiones acertadas respecto a la importación de materias primas a través de la aplicación de esta rama de la ciencia, soportadas en técnicas de prueba de hipótesis a partir de series analizadas, que permiten predecir la forma en la que se comporta la variable de estudio.

En primera instancia, se expone acerca de la ejecución de la presente tesis, planteamiento del problema, los objetivos correspondientes y la hipótesis analizada.

Con la finalidad de permitir la adecuada comprensión del presente estudio se explican los fundamentos teóricos y nociones conceptuales. Dentro de estos conceptos se encuentran: series temporales, estacionariedad, funciones de correlación, autocorrelación y autocorrelación parcial, ruido blanco, procesos autorregresivos, procesos de medias móviles, procesos ARIMA y procesos ARIMA multiplicativos y estacionales.

Posteriormente, se describe la metodología utilizada a lo largo de la presente investigación, así como información referente a las bases de datos y herramientas que hicieron el estudio posible.

Así también encontraremos la aplicación de la metodología Box-Jenkins a las series temporales de los segmentos de aditivos: ventas totales, fuera de borda, gasolina y diésel; obteniendo modelos de pronósticos de demanda de estos productos a través de sus datos originales de ventas de los periodos 2014 al 2019.

Finalmente, se detallan las conclusiones respecto a las variables y procesos estudiados y las recomendaciones para mejorar la utilización de este tipo de herramientas dentro del campo estadístico puro como aplicativo en el sector industrial.

1.1. Antecedentes

De acuerdo a lo expresado por (León y Romo, 2018) en Ecuador las actividades relacionadas al sector automotriz se encuentran muy cercanamente concatenadas a diferentes áreas de la economía: agricultura, industria, comercio tanto nacional como internacional; cada una de ellas tiene la necesidad de utilizar medios de transporte y vehículos para la movilización respectiva de mercancías, a nivel de turismo los automóviles son fundamentales para el transporte de personas a las diversas regiones del país, de manera similar otras áreas productivas como son banca y salud. A nivel nacional, es viable identificar una cantidad superior a 29000 empresas cuyas actividades principales se encuentran intrínsecamente relacionadas con producción, servicios y comercio automotriz que en consecuencia aportan con el progreso económico ecuatoriano.

En Ecuador, el mercado de lubricantes asociado al segmento automotriz está caracterizado por abrir las puertas tanto para marcas de lubricantes nacionales como para aquellas marcas que son producidas fuera del país, acorde a las necesidades mercantiles. De acuerdo a lo indicado por (APEL, 2019) en Ecuador se comercializaron 27.4 millones de galones en el 2019, de los cuales 59% del producto comercializado era de origen nacional, estos son fabricados localmente con aditivos importados y aceites básicos de origen nacional o extranjero. Si estimamos un precio de venta al público cercano a \$21 conforme a lo indicado por (León y Romo, 2018) podríamos indicar que el sector de lubricantes en su totalidad significaría una contribución a la economía ecuatoriana de quinientos setenta y cinco mil millones de dólares, de los cuales trescientos cuarenta mil millones de dólares corresponderían al mercado de lubricantes de origen nacional.

Se han elaborado trabajos previos de planificación de la demanda aplicados en esta empresa. Zuñiga (2018) realizó un estudio titulado "Desarrollo de un modelo de planificación de demanda mediante el análisis de aditivos para la empresa Hidromecánica Andina CIA. LTDA."

En este proyecto se desarrolló un modelo de planificación de demanda tomando como información base el registro histórico de ventas generado por el ERP de la compañía, esta información sirvió para proyectar estadísticamente el comportamiento del mercado en sus distintos segmentos, para desarrollar una planificación en base a presupuesto. Se definió la importancia de realizar una planificación de demanda, la cual debe ser tratada y revisada de manera periódica, además se indicaba que la correcta planificación de demanda permitiría al departamento de operaciones estar preparado para los posibles escenarios de venta y evitar la pérdida de facturación por escasez de inventario. (Zuñiga, 2018)

Durante los últimos 6 años el mercado ecuatoriano de aceites lubricantes ha presenciado la importación de aceites lubricantes fabricados en el extranjero; debido a agresivas campañas de marketing, incremento en las tecnologías de

lubricantes, reducción de precios; entre otros, alcanzando una participación de ventas a nivel nacional durante el año 2019, del 41% (APEL, 2019). Este fenómeno, trajo consigo la reducción del volumen de los productores de lubricantes nacionales y en consecuencia reducción de demanda de aditivos. Debido a la fuerte competencia a la que se enfrentan los productores nacionales, existe una creciente necesidad de disponibilidad de productos con precios competitivos acorde a la demanda de los clientes, en términos de tiempos de entrega y volúmenes.

1.2. Descripción del problema

La empresa comercializadora de aditivos para lubricantes; es una institución proveedora de materias primas para grandes fábricas de lubricantes y otras industrias a nivel nacional, que en su mayoría requieren que sus proveedores tengan disponibilidad inmediata de la mayoría de sus ítems de venta (Hidromecánica, 2019).

Considerando que la bodega de almacenamiento de la empresa comercializadora de aditivos tiene un área de 1068 metros cuadrados (Guamán, 2016); no es viable proceder con la importación de grandes volúmenes de cada uno los 100 aditivos que conforman su portafolio. A su vez, es imperativo mantener el capital de trabajo en movimiento, lo cual puede sólo ser alcanzado si se importan los productos que demanden los clientes al tiempo que lo requieran "Just in time". Cabe mencionar, que en diferentes ocasiones se ha procedido con la importación excesiva de productos que posteriormente no han podido ser colocados en el mercado generando eventos de "SMI" (Slow Moving Inventory) en donde el aditivo debe ser revendido o reexportado, generando pérdidas por almacenamiento prolongado y reducción del capital del trabajo. Por el contrario, han existido casos en los que a pesar de que el producto mantiene una demanda constante, este no ha sido importado oportunamente generando inconvenientes severos entre el cliente y el proveedor. (Zúñiga, 2018)

Como un insumo para la planificación de las ventas de la compañía, se plantea realizar un modelo de series temporales que incluyan la estimación de los montos de ventas totales de la compañía, así como de sus principales aditivos, de tal forma que se planifique de forma oportuna su importación minimizando los inconvenientes que se suscitan entre el cliente y el proveedor.

El modelo de series de temporales será generado a partir de los valores de las ventas mensuales del producto o productos de mayor volumen de la empresa, basados en la metodología Box-Jenkins.

1.3. Objetivos

1.3.1. Objetivo General

Diseñar un modelo de series de tiempo de las ventas mensuales en una empresa comercializadora de aditivos para lubricantes.

1.3.2. Objetivos Específicos

- Fundamentar teóricamente la aplicación de los modelos de series de tiempo.
- Desarrollar diferentes escenarios de modelos de Series de Tiempo para estimar las ventas en una empresa comercializadora de aditivos para lubricantes.
- Evaluar y comparar el rendimiento de los modelos auto-regresivos propuestos.
- Analizar las ventas de aditivos con la aplicación de modelos de series de tiempo autorregresivos incluyendo: AR, MA, ARIMA y SARIMA.

1.4. Hipótesis

Los modelos univariantes de series de temporales proporcionan mejores ajustes para el pronóstico de demanda en una empresa comercializadora de aditivos para lubricantes, al evaluar las diferencias entre los conjuntos de entrenamiento y evaluación de los modelos planteados.

1.5. Alcance

El presente estudio se llevará a cabo en una empresa comercializadora de productos aditivos para lubricantes ubicada de Guayaquil-Ecuador, a partir de los datos de ventas de aditivos para lubricantes del periodo 2014-2019. Se analizarán cuatro series de tiempo, incluyendo: ventas totales de aditivos, ventas de aditivos para los principales lubricantes: fuera de borda, gasolina y diésel.

Para este estudio se utilizarán datos desde enero 2014 a diciembre 2019.

CAPÍTULO 2

2. MARCO TEÓRICO

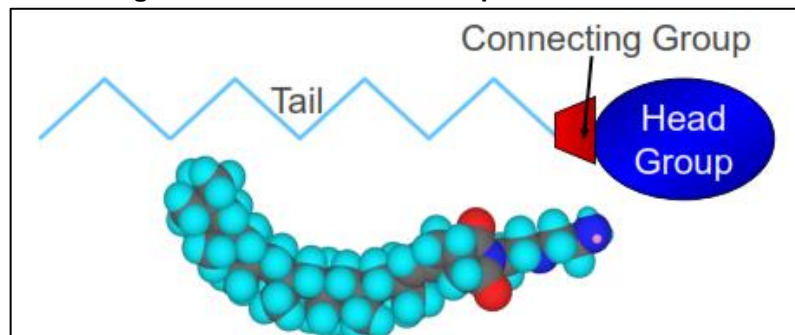
2.1. Aceite Lubricante

Un aceite lubricante, es un fluido cuya finalidad es evitar el contacto entre superficies en movimiento en búsqueda que no se produzca un roce directo entre ellas. De esta forma, se incrementa la vida útil del mecanismo, mejorando su rendimiento, evitando pérdidas por fricción, promoviendo el hermetismo adecuado, disipando el calor producido, facilitando la limpieza de las piezas y evacuando impurezas. Principalmente, los aceites lubricantes líquidos automotrices e industriales, se encuentran compuestos de: aceites básicos y aditivos. (Lubrizol, 2017).

2.2. Definición de aditivo

Por otra parte, los aditivos son sustancias químicamente activas o inertes que se añaden a los aceites para formar el lubricante final y así mejorar las propiedades existentes o añadirle nuevas, estos están formados generalmente por un grupo funcional de conexión, un grupo funcional de "cabeza" y una cola.

Figura No.2.1 Estructura de típica de un aditivo



Fuente: The Lubrizol Cooperation.

2.3. Series Temporales

A continuación, se detallan los principales conceptos asociados a las series temporales, los cuales dan sustento a la aplicación de este tipo de series para estudiar las relaciones entre diferentes variables y sus variaciones respecto al tiempo.

2.3.1. Proceso Estocástico

De acuerdo a lo indicado por (Quezada, Rodríguez, Pérez, Rodríguez-Picón, 2018) podemos definir como proceso estocástico a un modelo en donde una o más variables cambian de forma aleatoria a lo largo del tiempo. Es también una familia " $\{D_t\}$ ", siendo un conjunto no vacío en donde t adquiere valores (representados por

subíndices) de variables aleatorias, las cuales a su vez tienen distribución normal y varianzas cero. Indica también, que los elementos que distinguen principalmente a un proceso estocástico son el conjunto de subíndices, la forma en la que se relacionan las variables aleatorias y el espacio de los estados.

De forma similar (Tsay, 2010) indica que un proceso estocástico constituye también una familia de variables aleatorias que, de manera general, están relacionadas entre sí y a su vez siguen una ley de distribución conjunta. Se encuentra representada por: $\dots Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}, \dots$, ó sencillamente Y_t . Teniendo en cuenta que, la presente investigación se enfoca en series temporales, consideraremos únicamente secuencias de variables ordenadas aleatorias en el tiempo; no obstante, conforme a lo expuesto por los autores (Tsay, 2010) y (Quezada, Rodríguez, Pérez, Rodríguez-Picón, 2018), para el caso de un proceso estocástico, esta condición no es necesaria, podría utilizarse en lugar del tiempo un ordenamiento en función del espacio.

2.3.2. Definición de Series Temporales:

Los autores (Parra y Vicente, 2019) y (Parra, 2019) indican que una serie temporal se caracteriza por ser una sucesión respecto a una variable efectuada en intervalos regulares de tiempo. Así también, (Tsay, 2010) expresa que una serie de tiempo o serie temporal es básicamente una secuencia de observaciones, que han sido recopiladas, registradas u observadas. Estas observaciones $\{Z_1, \dots, Z_t, \dots, Z_n\}$ son recopiladas considerando aumentos sucesivos de tiempo, de características regulares; de donde Z_t es la observación de nuestra variable de interés en el instante t .

Este instante temporal t anteriormente mencionado, puede considerarse un año, trimestre, mes, semana u otros, lo cual determina la frecuencia de la observación o registro; siendo esta, anual, trimestral, etc. Es decir, todas las observaciones detalladas en la serie son obtenidas en instantes equidistantes de tiempo; así también, en las series temporales lineales deben considerarse los supuestos que se describen a continuación (Parra, 2019):

Existe un comportamiento estable respecto al fenómeno estudiado y para esto es requerido que los periodos estudiados sean homogéneos o que se encuentren muy próximos a la homogeneidad.

Del mismo modo, los datos deben ser homogéneos respecto al tiempo, para esto es necesario conservar la forma en la que se define y mide cada magnitud estudiada.

Aquellas que no cumplen los supuestos anteriormente expuestos, entran dentro de otro universo de series temporales que no serán incluidas en el presente estudio, series temporales no lineales. De acuerdo a lo expuesto por (Sánchez y Velásquez, 2005) existen también un sin número de series temporales cuyo comportamiento es no lineal; y que, en gran cantidad de casos, no es viable especificar un modelo estadístico que los represente a través de leyes físicas o económicas existentes.

2.3.3. Estacionariedad

La base del análisis de las series temporales es la estacionariedad (TSay, 2010). Se puede definir la estacionariedad de la siguiente forma:

Se dice que una serie temporal $\{r_t\}$ es *estrictamente estacionaria* si la distribución conjunta de $(r_{t_1}, r_{t_2} \dots r_{t_k})$ es idéntica a $(r_{t_1+t}, r_{t_2+t} \dots r_{t_k+t})$ para todo t , donde k es un entero positivo arbitrario y $(t_1, t_2 \dots t_k)$ es una colección de k enteros positivos. (Tsay, 2010).

Así también, expresa que la estacionariedad estricta o fuerte exige que la distribución $(r_{t_1}, \dots, r_{t_k})$ de una serie de tiempo $\{r_t\}$ sea invariable respecto al tiempo. Acorde a la literatura revisada, es bastante complejo comprobar este estado de forma empírica. Por lo que generalmente se intenta determinar o asumir una versión más débil de estacionariedad. El autor anteriormente mencionado, señala que una serie temporal $\{r_t\}$ es débilmente estacionaria si tanto la media de r_t como la covarianza entre r_t y $r_{t-\ell}$ no varían en relación al tiempo, donde ℓ es un número entero arbitrario. Dicho de otra manera, $\{r_t\}$ tiene estacionariedad débil si $E(r_t) = \mu$ y $Cov(r_t, r_{t-\ell}) = \gamma_\ell$, en donde este último enunciado depende estrictamente de ℓ . (Tsay, 2010)

De manera similar (Ramírez, 2019) expresa que la estacionariedad corresponde a que las características de una serie no cambian en relación al tiempo, es decir que su variación no presenta cambios en función del tiempo. Matemáticamente, tomaremos como referencia los siguientes conceptos:

Un proceso estocástico que cumple la propiedad descrita a continuación, es fuertemente estacionario:

$$F_X(x_{t_1+r}, \dots, x_{t_n+r}) = F_X(x_{t_1}, \dots, x_{t_n}) \quad (2.1)$$

En donde $X \in \mathbb{Z}$ y $T \in \mathbb{N}$.

En otras palabras, la probabilidad de que el proceso adquiriera determinado valor en cada uno de sus estados, se mantiene igual respecto al tiempo.

Por otra parte, la estacionariedad débil se cumple si:

1. $E[X_t] = \mu \quad \forall t \in T$, expresión que indica que la media es constante.
2. $E[X_t^2] < \infty \quad \forall t \in T$ siendo el segundo momento finito en cualquier momento, se asegura que también sea finita la varianza.
3. $Cov(X_{t_1}, X_{t_2}) = Cov(X_{t_1+h}, X_{t_2+h}) \quad \forall t \in \mathbb{N}, \forall h \in \mathbb{Z}$ De tal forma que, si la varianza es constante, podríamos decir que la varianza a lo largo de los estados del proceso tampoco varía:

$$Cov(X_{t_1}, X_{t_2}) = Cov(X_{t_1+h}, X_{t_2+h}) \Leftrightarrow Var[X_t] = Var[X_{t+h}] \quad \forall t \in \mathbb{N}, \forall h \in \mathbb{Z}$$

A fin de aplicar estos conceptos a ejercicios prácticos, podemos asumir que hemos observado T puntos de datos, $\{r_t | t = 1, \dots, T\}$. De forma gráfica o visual la estacionariedad débil significaría que nuestros datos, los T valores de la serie oscilan presentando cambios constantes respecto a un mismo nivel. (Tsay, 2010)

En las aplicaciones, la estacionariedad débil permite hacer inferencia convergiendo observaciones futuras (por ejemplo, predicción).

Por otra parte, en los estados en donde se presente estacionariedad débil, debemos suponer que los dos primeros momentos de r_t se encuentran definidos o tienen características finitas. Teniendo en consideración las definiciones revisadas, si r_t posee estacionariedad estricta y los dos primeros momentos que la conforman son de carácter finito, entonces podemos decir que r_t es también débilmente estacionaria. Ambos autores (Parra, 2019) y (Tsay, 2010) son congruentes al indicar que el estado contrario no es cierto generalmente; es decir, el hecho de una serie de tiempo no tenga estacionariedad fuerte no significa que esta tenga estacionariedad débil. No obstante, existen excepciones en donde si la serie de tiempo r_t presenta adicionalmente una distribución normal, entonces la estacionariedad débil es también equivalente a estacionariedad estricta. En el presente estudio encontraremos principalmente series débilmente estacionarias.

En relación al presente estudio, durante la etapa de identificación en la modelación con ARIMA(p,d,q) se ejecutarán los análisis respectivos de estacionariedad, por lo que a continuación se exponen los tipos de estacionariedad aplicados a estos modelos. De acuerdo a lo expuesto por González (2009), detallamos a continuación los tipos de estacionariedad que podríamos encontrar típicamente en el análisis del comportamiento en series temporales:

Estacionariedad en varianza: Generalmente podríamos decir que una serie es estacionaria respecto a su varianza siempre y cuando sea factible conservar el supuesto de que se mantiene una misma varianza a lo largo de la serie temporal, dicho de otra manera, cuando se cumple que la variabilidad de una serie en relación a su media es constante con el transcurso del tiempo. De no ser así, se recomienda aplicar transformaciones que contribuyan a la estabilización de la varianza, estas son también llamadas transformaciones de Box-Cox, González (2009). Considerando que las variables analizadas son positivas y no contienen valores de cero, la transformación utilizada será la logarítmica.

Estacionariedad en media: Decimos que una serie es estacionaria en media si fluctúa en relación a un nivel constante. Con el fin de ampliar este texto, expondremos las características de las series estacionarias y no estacionarias:

- **Serie estacionaria en media:** Se dice que una serie es estacionaria en media cuando es factible conservar el supuesto de que existe una misma media a lo largo de la serie; así también, su función de autocorrelación teórica decae de forma rápida y exponencial.

- **Serie no estacionaria en media:** González (2009) indica también que estamos antes la presencia de una serie no estacionaria en media cuando se evidencia en su comportamiento a lo largo del tiempo una tendencia o en su defecto tramos diferentes con valores de medias variados. Así también, una serie es no estacionaria en media cuando su función de autocorrelación de muestra tiene un decaimiento muy lento; no obstante, no es requerido que se encuentre cercano a la unidad.

En caso de que la serie sea no estacionaria en media, esta puede alcanzarse transformando la serie aplicando diferencias. De esta forma, si la serie es no estacionaria en términos de media, se emplearán d continuas diferencias de orden 1 a la serie estudiada con el objetivo de obtener finalmente una serie estacionaria. González (2009) resume una serie y las aplicaciones de las d diferencias en el siguiente enunciado:

$$Z_t = (1 - L)^d Y_t \quad (2.2)$$

2.3.4. Función de correlación y autocorrelación

(Tsay, 2010) expresa que en relación a las series temporales se define un coeficiente de correlación entre dos variables aleatorias X e Y a través de la siguiente ecuación:

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E(X - \mu_x)^2 E(Y - \mu_y)^2}} \quad (2.3)$$

Aquí observamos que μ_x y μ_y equivalen a las medias de X e Y , de forma correspondiente y se supone que se evidencia también una varianza. La intención de definir este factor es medir que tan linealmente dependientes son X e Y el uno del otro, (Tsay, 2010) también indica que se puede demostrar que $-1 \leq \rho_{x,y} \leq 1$ y $\rho_{x,y} = \rho_{y,x}$. En caso de que $\rho_{x,y} = 0$ decimos también que estas dos variables aleatorias no se encuentran correlacionadas entre sí. Finalmente, podemos asumir que $\rho_{x,y} = 0$, si tanto X como Y poseen una distribución normal y además no dependen linealmente entre sí.

Tomando como referencia el hecho de que sea posible utilizar una muestra $\{(x_t, y_t)\}_{t=1}^T$, la correlación puede ser estimada a través de la siguiente expresión:

$$\rho_{x,y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2 \sum_{t=1}^T (y_t - \bar{y})^2}} \quad (2.4)$$

(Tsay, 2010) indica que las expresiones $\bar{x} = \sum_{t=1}^T x_t / T$ y $\bar{y} = \sum_{t=1}^T y_t / T$ corresponden a las medias muestrales de cada una de las variables de X e Y .

2.3.4.1. Función de Autocorrelación (ACF)

Los autores (Carvajal, 2014) y (Tsay, 2010) son congruentes en que cuando es necesario estudiar si una serie de tiempo r_t posee o no dependencia lineal con sus valores pasados r_{t-i} es imperativo generalizar concepto de correlación hacia autocorrelación. Llamamos autocorrelación Lag- ℓ de r_t al factor ρ_ℓ que expresa la relación entre r_t y $r_{t-\ell}$, lo anteriormente expuesto es viable bajo el supuesto de estacionariedad débil, dependiendo solamente de ℓ , tal como se describe a continuación:

$$\rho_\ell = \frac{Cov(r_t, r_{t-\ell})}{\sqrt{Var(r_t)Var(r_{t-\ell})}} = \frac{Cov(r_t, r_{t-\ell})}{Var(r_t)} = \frac{\gamma_\ell}{\gamma_0} \quad (2.5)$$

Según (Gujarati y Porter, 2010) esta expresión no es otra cosa que el cociente entre la covarianza del rezago ℓ y la varianza. Así también nos indica que, al tener las mismas unidades, el resultado de este cociente ρ_ℓ es adimensional y por su naturaleza sus valores oscilarán entre -1 y +1, del mismo modo que ocurre generalmente en la estimación correcta de correlaciones.

En el enunciado anteriormente expuesto por (Tsay, 2010), usamos la propiedad Donde se usa la propiedad $Var(r_t) = Var(r_{t-\ell})$ aplicable a series temporales con estacionariedad débil. En relación a la definición $\rho_0 = 1$, $\rho_l = \rho_{-l}$ y $-1 \leq \rho_l \leq 1$, decimos que una serie r_t débilmente estacionaria no se encuentra correlacionada si y solo si $\rho_\ell = 0$ para todo $\ell > 0$. En adición, tomando como referencia una muestra de retornos $\{r_t\}_{t=1}^T$, siendo \bar{r} la media de nuestra muestra ($\bar{r} = \sum_{t=1}^T r_t / T$). Definimos nuestra autocorrelación de la muestra en función de su retardo lag-1 de r_t , conforme se expone a continuación:

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2} \quad (2.6)$$

Finalmente podríamos indicar que, considerando ciertas condiciones generales que $\hat{\rho}_1$ es una estimación estable respecto de ρ_1 . (Tsay, 2010) menciona que si tenemos en consideración una serie de tiempo r_t , la cual se caracterizaría por ser una secuencia o sucesión distribuida de forma independiente e idéntica (iid) y considerando también que $E(r_t^2) < \infty$, es factible también decir que $\hat{\rho}_1$ es normal de forma asintótica con una varianza $1/T$ y media equivalente a cero. Exponemos lo anteriormente mencionado con el fin de que pueda usarse para comprobar tanto la hipótesis nula $H_0: \rho_1 = 0$ en relación a la hipótesis alternativa $H_a: \rho_1 \neq 0$. Usaremos como estadístico de esta comprobación la relación *t ratio*, la cual corresponde a $\sqrt{T} \hat{\rho}_1$ y que a su vez se aproxima de forma asintótica a la distribución normal de características estándar. Para este caso, rechazamos la hipótesis nula H_0 si *t ratio* es significativo en magnitud o, de manera similar, si el valor p del *t ratio* es inferior a 0.05. Tanto (Tsay, 2010) como (Carvajal, 2014) definen la autocorrelación de muestral lag-l de r_t como a continuación se describe:

$$\widehat{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, 0 \leq \ell < T - 1. \quad (2.7)$$

Es importante también resaltar de qué se trata la Función Parcial de Autocorrelación (PACF). Los autores (Gujarati y Porter, 2010) así como (Tsay, 2010) indican que es una herramienta que facilita el análisis de modelos de tipo AR en donde por definición decimos que la aplicación de PACF corta el rezago p de una serie de tipo AR(p).

Para el caso de los modelos MA; no obstante, es recomendable aplicar ACF debido a que una vez más, por definición, esta función corta en cambio el rezago q en una serie de tipo MA(q).

(Gujarati y Porter, 2010) indican que al utilizar ACF, muestral o poblacional, medimos la forma en la que se modifica el valor medio de la variable regresada considerando un cambio unitario, conservando la influencia del resto de variables a las que se les ha aplicado retraso.

2.3.4.2. Prueba de ACF individual

Para concluir los conceptos básicos relacionados a las pruebas de autocorrelación, (Tsay, 2010) indica también que si consideramos un entero positivo dado ℓ , es posible utilizar el resultante de la ecuación anterior para verificar que $H_0: \rho_\ell = 0$ contra $H_a: \rho_\ell \neq 0$. Se define también que el estadístico de la prueba, como a continuación se expone:

$$t \text{ ratio} = \frac{\widehat{\rho}_\ell}{\sqrt{(1 + 2 \sum_{i=1}^{\ell-1} \widehat{\rho}_i^2)/T}} \quad (2.8)$$

Adicionalmente, el autor (Tsay, 2010) indica que si una serie temporal r_t es definida como Gaussiana, y que a su vez cumple con el enunciado $\rho_j = 0$ para $j > \ell$, $t \text{ ratio}$, tiene una distribución con forma asintótica al igual que una variable aleatoria de características normales y estándar. Debemos tener en consideración que la base de la toma de decisión de esta prueba es rechazar la hipótesis nula si el valor absoluto de $t \text{ ratio}$ es mayor que $Z_{\alpha/2}$, aquí encontramos que este $Z_{\alpha/2}$ se define también como el percentil número $100(1 - \alpha/2)$ de la distribución normal estándar. Con el fin de reducir este complejo procedimiento, la mayoría de paquetes de software actuales aplican simplemente $1/T$ en símil de la varianza asintótica $\widehat{\rho}_\ell$ en relación a todo ℓ diferente de cero.

2.3.4.3. Ruido Blanco

Según (González, 2009) definimos el proceso de Ruido Blanco, como el proceso estocástico que se expresa como una sucesión de variables aleatorias cuya media es igual a cero, manteniendo la misma varianza a lo largo del proceso y la nulidad

de sus covarianzas. Se encuentra usualmente denotada de la siguiente manera: $a_t, t = 0, \pm 1, \pm 2, \dots$:

$$E(a_t) = 0, \forall t \quad V(a_t) = \sigma^2, \forall t \quad Cov(a_t a_s) = 0, \forall t \neq s \quad (2.9)$$

De esta manera, encontramos que un proceso de ruido blanco, $a_t \sim RB(0, \sigma^2)$, se define como estacionario siempre y cuando su varianza σ^2 tiene valores finitos, en relación a la función de auto covarianzas:

$$\gamma_k = \sigma^2, k = 0 \quad y \quad \gamma_k = 0, k > 0 \quad (2.10)$$

Así como su función de autocorrelación:

$$\gamma_k = 1, k = 0 \quad y \quad \gamma_k = 0, k > 0 \quad (2.11)$$

2.3.4.4. Procesos Autorregresivos AR(p)

Siendo directamente aplicables al presente estudio, según (González, 2009) modelo de características AR(p), que se define como autorregresivo finito de orden p, es una aproximación natural al modelo lineal general:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \Phi_3 Y_{t-3} + \dots + a_t \quad \forall t, \quad t = 1, 2, \dots \quad (2.12)$$

Por otra parte, vemos también que de manera general los procesos autorregresivos cuyo orden es p, exponen a Y_t como una función desde su pasado, acercándose al retardo $t-p$, considerando también una innovación contemporánea a_t :

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + a_t \quad a_t \sim RB(0, \sigma^2) \quad t = 1, 2, \dots \quad (2.13)$$

Transformando la ecuación anterior aplicando operadores de retardos se tiene:

$$(1 - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p) Y_t = a_t \quad \Phi_p(L) Y_t = a_t \quad (2.14)$$

Donde $\Phi_p(L)$ toma la denominación de polinomio autorregresivo y $(\Phi_1, \Phi_2, \dots, \Phi_p)$ constituye ser el vector de parámetros autorregresivos. En primera instancia se requiere verificar que el proceso AR(p) alcanza las condiciones de estacionariedad para cada uno de sus parámetros. Por lo anteriormente expuesto, la literatura expresa que sólo de presentarse un polinomio autorregresivo $\Phi_p(L)$ cuyo módulo de raíces se localice en el exterior del círculo unitario podría decirse que un proceso AR(p) posee características estacionarias. González (2009)

2.3.4.5. Procesos de Medias Móviles: MA(q)

Otro de los procesos aplicados en la presente investigación son los procesos de medias móviles, en donde tanto (González, 2009) como (Gujarati y Porter, 2010)

en resumen nos indican que un modelo de medias móviles MA(q), de orden finito, puede ser expuesto bajo el siguiente enunciado:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad a_t \sim RB(0, \sigma^2) \quad (2.15)$$

Que expresado en términos del operador de retardos se tiene:

$$Y_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) a_t \rightarrow Y_t = \theta_q(L) a_t \quad (2.16)$$

En esta ecuación $\theta_q(L)$ toma el nombre de polinomio de medias móviles; mientras que $(\theta_1, \theta_2, \dots, \theta_q)$ se convierte en los parámetros de medias móviles expresados a través de un vector.

$$Y_t = a_t - \theta_1 a_{t-1} \quad a_t \sim RB(0, \sigma^2) \quad t = 1, 2, \dots \quad (2.17)$$

Esta última ecuación es un modelo típico MA(1), los modelos MA(q) son generalizaciones de los modelos MA(1), razón por la cual tienen características análogas; no obstante, conforme se agregan más retardos de la perturbación en los modelos, la memoria incrementa y la estructura dinámica que representa el modelo es más óptima. De acuerdo a lo indicado por (González, 2009), si una interferencia ingresa en un momento t , esta tendrá consecuencias desde Y_t hasta Y_{t+q} . De esta manera, la interferencia a_t , dentro de un MA(q) persiste q periodos en el procedimiento. Este fenómeno nos indica que un modelo MA(q) tiene una memoria más larga, lo cual veremos reflejado en las funciones de autocovarianzas y autocorrelación.

2.3.4.6. Procesos Autorregresivos de Medias Móviles:

ARMA(p,q)

Una vez más acorde a la literatura revisada de los autores (Gujarati y Porter, 2010) los procesos autorregresivos de medias móviles (ARMA) tiene características de AR y MA a la vez, por otra parte, acorde a lo indicado por (González, 2009) los modelos ARMA (p,q) definen Y_t como función de su pasado llegando al retardo p , así como de la innovación presentada en relación t de este mismo periodo y su pasado correspondiente llegando al retardo q :

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \\ a_t \sim RB(0, \sigma^2) \quad (2.18)$$

Que una vez más, expresado en función del operador de retardos se tiene:

$$(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) a_t = (1 - \phi_1 L - \dots - \phi_p L^p) Y_t \quad (2.19)$$

$$\theta_q(L) a_t = \phi_p(L) Y_t \quad (2.20)$$

Logrando así que $\Phi_p(L)$ sea el polinomio autorregresivo, mientras que $\theta_q(L)$ conformaría el polinomio de medias móviles.

Para el caso de los modelos ARMA(p,q) se tiene también el siguiente teorema respecto al análisis de estacionariedad para cualquiera valor de sus parámetros, en donde se expresa que podemos decir que estamos en presencia de un proceso ARMA(p,q), de características estacionarias cuando su polinomio $\Phi_p(L)$ tiene raíces cuyo módulo se encuentra en el exterior del círculo unitario. (González, 2009). Es importante mencionar que también que las condiciones de estacionariedad de los modelos ARMA(p,q) son dadas estrictamente por la parte AR, debido a que la parte de MA siempre es estacionaria.

Un modelo ARMA(p,q) tiene propiedades y características en común con los fundamentos establecidos para los modelos autorregresivos y de medias móviles, debido a que utiliza las dos estructuras al mismo tiempo. De manera general, se dice que un modelo ARMA(p,q) consta de una media igual a cero, conservando una varianza constante al igual que finita, mientras que su función de autocovarianzas se caracteriza por ser infinita. Así también, su ACF es infinita; no obstante, sin presentar cortes o truncamientos esta decrece de forma rápida hacia cero.

2.3.4.7. Modelos ARIMA (p,d,q):

O también llamados procesos autorregresivos integrados de promedios móviles, de acuerdo a lo indicado por (Gujarati y Porter, 2010) se fundamentan en el supuesto de que las series temporales analizadas son estacionarias de forma débil. (González, 2009) por otra parte, define un modelo Y_t de orden d es de carácter integrado, $Y_t \sim I(d)$, cuando no es estacionario Y_t ; no obstante, al aplicar diferencias de orden d , $\Delta^d Y_t$, este se transforma en un proceso de cuyas propiedades corresponde a un ARMA(p-d,q) que además es invertible y también estacionario. La cantidad de diferencias que hay que aplicar al proceso para alcanzar la estacionariedad en media corresponde al orden de integración del proceso es, o de forma equivalente, el número de raíces unitarias del proceso. Este tipo de series puede representarse de la siguiente manera:

$$\Phi_p(L)\Delta^d Y_t = \delta + \theta_q(L)a_t \quad (2.21)$$

En esta expresión, $\Phi_p(L)$ es el polinomio autorregresivo estacionario y $\theta_q(L)$ corresponde al polinomio invertible de medias móviles sin raíces comunes. (González, 2009) indica que el modelo anteriormente mencionado toma el nombre de modelo Autorregresivo Integrado de Medias Móviles de orden (p,d,q), también denominado ARIMA(p,d,q). Como fue explicado en los modelos AR y MA, p no es otra cosa que el orden del polinomio autorregresivo estacionario, mientras d corresponde al orden de integración de la serie o dicho de otra manera, es la cantidad de diferencias que se aplican a la serie hasta que se logre la estacionariedad en la misma, y finalmente q es el orden del polinomio MA invertible.

2.3.4.8. Modelización utilizando ARIMA:

Con el fin de proceder con la aplicación de la metodología Box-Jenkins es imperativo conocer los valores de la serie de tiempo o serie temporal Y_1, Y_2, \dots, Y_t con el objetivo de definir qué estructura $ARIMA(p,d,q)$, fue causal de origen (González, 2009).

La generación de modelos ARIMA se realiza de manera iterativa a través detallado a continuación:

- **Identificación:** (Gujarati y Porter, 2010) indican que es necesario hallar los valores adecuados de p , d y q a través de los análisis de correlogramas y correlogramas parciales. (González, 2009) indica que para este proceso de identificación es se utilizan tanto de datos como otra clase de información disponible que contenga detalles de cómo fue generada una serie, se trata de recomendar una subclase de modelos ARIMA (p,d,q) que tenga sustento suficiente para ser investigado. Lo anteriormente expuesto, tiene la finalidad de definir los órdenes p,d y q que sean apropiados para producir las características de la serie motivo de investigación y si tiene o no pendiente. Es en esta etapa pueden definirse varios modelos candidatos posiblemente capaces de generar la serie.
- **Estimación:** Cuando se utilizan de correcta los datos u observaciones es factible efectuar inferencias acerca de los parámetros con las condiciones necesarias para que el modelo motivo de estudio sea el adecuado. Determinando un proceso propuesto, se intenta cuantificar los parámetros del modelo: $\theta_1 \dots \theta_q, \phi_1, \dots \phi_p, \sigma^2$ y de presentarse δ .
- **Validación:** o también llamado diagnóstico (Gujarati y Porter, 2010), busca verificar si el modelo generado representa o no de forma adecuada a la serie temporal original. (González, 2009) menciona que es en esta etapa se llevan a cabo contrastes de diagnóstico con el fin de verificar si el modelo se ajusta a nuestras observaciones o datos, o de lo contrario, analizar las incongruencias del modelo en búsqueda de mejoras.
- **Predicción:** Finalmente, se realiza la obtención de pronósticos, expresados en términos de probabilidad, de los valores venideros o futuros de la variable de respuesta. Es de vital importancia durante la ejecución de la predicción, analizar la capacidad predictiva del modelo propuesto.

La metodología anteriormente expuesta, se basa en los siguientes fundamentos:

- 1) Elección del modelo de manera iterativa. A lo largo de los periodos de identificación, estimación, validación y predicción, se bosqueja la posibilidad de volver a ejecutar fases anteriores.
- 2) Aplicación del principio de parametrización (parsimonia). Este principio tiene la finalidad de exponer un modelo con la capacidad de representar el conjunto de datos o serie temporal con el menor número de parámetros posibles y solamente utilizar la amplificación del modelo si esto fuera rigurosamente requerido para lograr una mejor descripción de la serie (González, 2009).

2.3.4.9. Contraste de Dickey-Fuller:

En búsqueda de explicar de forma sencilla el principio sobre el que se fundamenta el contraste de Dickey-Fuller (Rodó, 2019) expresa que este contraste es una comprobación de raíz única que identifica de forma estadística que existe evidencia de un comportamiento tendencial estocástico aplicado a las series temporales de las variables a través de contraste de hipótesis. Es decir, este contraste nos permite comprender si existen tendencias representativas en nuestras series mediante el contraste de hipótesis antes mencionado.

Para el presente estudio, haremos referencia a lo indicado por (González, 2009), considerando que un proceso ARMA(p,q) parametrizado:

$$\Delta Y_t = \beta Y_{t-1} + \alpha_1 \Delta Y_{t-2} + \dots + \alpha_{p-1} \Delta Y_{t-p+1} + at \quad (2.21)$$

En donde:

$$\beta = \sum_{i=1}^p \phi_i - 1 \quad y \quad \alpha_i = \sum_{j=1}^i \phi_{p-i+j} \quad (2.22)$$

Teniendo en cuenta que un proceso AR(p) contiene una raíz unitaria al presentarse $\sum_{i=1}^p \phi_i = 1$, y que todo proceso ARMA (p,q) puede ser aproximado hasta el grado de bondad o similitud requerido de un AR(p), el realizar el contraste de hipótesis nula acerca de la existencia de una raíz unitaria es igual a contrastar $H_0: \beta = 0$ en la regresión del proceso parametrizado. El contraste de raíz unitaria que antecede toma el nombre de Dickey-Fuller Aumentado (ADF) y se fundamenta en el estimar los mínimos cuadrados ordinarios (MCO) del parámetro β en el modelo parametrizado y en su correspondiente estadístico t. Se conoce que este estadístico tiene una distribución igual para el caso de un AR(1), y por lo tanto se utilizan los valores críticos determinados por Dickey-Fuller (González, 2009).

2.3.4.10. Modelos ARIMA estacionales:

Debido a que las series de tipo económico se observan generalmente de forma mensual, trimestral, anual, etc; denotan evidencia de comportamientos estacionales; teniendo en consideración factores ajenos a la variable de respuesta como ambiente, turismo, costumbres entre otros. Para el análisis de series temporales ARIMA se recogen dos tipos de dependencia inter-temporal que representan: un comportamiento de tendencia o regular que es as su vez la relación lineal que existe en observaciones continuas o sucesivas y por otra parte el comportamiento estacional que es la relación lineal que se presenta entre observaciones de un mismo mes en años continuos. Para el caso de los modelos ARIMA estacionales se determinará el número de periodos estacionales a través de s, donde tendremos s=4 de obtener una serie trimestral o s=12 , para series de tipo mensual (González, 2009).

2.3.4.11. Modelos ARIMA multiplicativos, ARIMA

(p,d,q)x(P,D,Q)s:

Dentro de los modelos ARIMA(p,d,q)x(P,D,Q)s; P es el orden del polinomio AR estacional estacionario, Q por otra parte es el orden del polinomio MA estacional estacionario, mientras que D es la cantidad de diferencias de tipo estacional ($1-L^s$), requerido aplicar a la serie Y_t se estacionaria, dicho de otra manera, el orden de integración estacional de la serie. Este tipo de modelos son modelos flexibles debido a que denotan estacionalidades y tendencias estocásticas, recogiendo posibles interacciones entre estos componentes (González, 2009).

2.3.4.12. Prueba de Shapiro Wilks:

Durante el presente estudio nos vemos en la necesidad de analizar si nuestros datos globales, así como los datos de nuestras series temporales específicas tienen o no distribuciones normales. Para esto, aplicaremos el test de Shapiro-Wilks que se fundamenta en el planteamiento de la hipótesis nula de que una muestra se origina a partir de una distribución normal (Dietrichson, 2019).

Prueba de Hipotesis de Shapiro Wilks:

$H_0 =$ Los datos siguen una distribución Normal

Vs

$H_1 =$ Los errores no siguen una distribución Normal

Rechazar H_0 en favor de H_1 si el p-valor $< 0,05$

2.3.4.13. Prueba de Normalidad Jarque-Bera:

De acuerdo a lo expresado por (Flores, Muñoz y Sánchez, 2019) la prueba de normalidad de Jarque-Bera considera una muestra aleatoria X_1, X_2, \dots, X_n , en donde se define como el estadístico de la prueba:

$$JB = n \left(\frac{s^2}{6} + \frac{(k-3)^2}{24} \right) \quad (2.23)$$

En esta ecuación, s es la representación del coeficiente de asimetría, mientras que k nos indica el coeficiente de curtosis. Para este caso específico, la hipótesis nula de normalidad para llevar a cabo el contraste de hipótesis, es rechazada siempre y cuando se cumpla que $JB > JB_{c,\alpha}$ siendo $JB_{c,\alpha}$ el valor crítico dentro de una distribución chi cuadrado con un área α de 2 grados de libertad a la derecha.

CAPÍTULO 3

3. METODOLOGÍA

La presente investigación tiene un enfoque cuantitativo considerando que los modelos de Series de tiempo son una metodología estadística ampliamente utilizados para realizar proyecciones de variables económicas.

La investigación planteada, utilizando procesos deductivos, busca explicar de forma objetiva el comportamiento de las ventas de una empresa comercializadora de aditivos para lubricantes mediante el uso de metodologías de series temporales. Es necesario mencionar que la investigación es realizada sobre registros contables y administrativos que nos dan la facultad de que los datos se encuentren disponibles y aptos para ser comparados a través de la estadística analítica; específicamente mediante el desarrollo de modelos de series de tiempo.

Se lleva a cabo la utilización del Software RStudio para la elaboración de los modelos de series de tiempo; así también, se realizan análisis descriptivos de las variables involucradas en el desarrollo. Con el fin de darle validez estadística al estudio propuesto se verifica el orden de integración de las variables implicadas, realizando la comprobación de supuestos, estacionalidad y autocorrelación, para cada uno de los modelos o escenarios presentados a través de las pruebas estadísticas que mejor apliquen basados en los datos mensuales de al menos periodo de 6 años con el fin de evitar la identificación errónea de patrones y relaciones (González, 2009). En búsqueda cumplir el primero de los objetivos específicos, se procede con la argumentación teórica de los modelos de series de tiempo (autorregresivos) (Box, Jenkins 1970).

En relación al siguiente objetivo se desarrollan diferentes modelos y escenarios de series de tiempo: Modelo Autorregresivos (AR), Modelo Medias Móviles (MA), Modelos Autorregresivos y Promedios Móviles (ARIMA); y Modelos Autorregresivos y Promedios Móviles Integrados con componente cíclico (SARIMA) (Tsay, 2010). Posteriormente se lleva a cabo la verificación del modelo a través del análisis de los residuales, verificando que se distribuyan de forma normal, y su posterior comparación a través de los criterios AIC, BIC y MAPE determinando cuál de los modelos propuestos describe de forma más adecuada el comportamiento de la variable de respuesta.

Finalmente se verifica si a través de los modelos construidos es posible generar proyecciones de ventas más precisas y usarlas como referencia para la toma de decisiones de importaciones de aditivos oportunas.

La base de datos utilizada es extraída mensualmente a través del sistema XASS utilizado en la empresa importadora y comercializadora de aditivos en donde se registran todas las ventas. Estos datos son extraídos en un documento a través de Excel consolidado y archivado en formato CSV. La base de datos en mención incluirá los periodos mensuales desde el año 2014 hasta el año 2019, expresadas en kilogramos. Dentro del presente desarrollo se estudiarán específicamente 14 unidades de observación, cada una con 72 observaciones para los periodos señalados.

Inicialmente se realiza un análisis exploratorio de datos previo a la ejecución de la serie temporal de las ventas totales, esta serie temporal es construida a partir de la sumatoria de las ventas mensuales de todos los productos.

Se analizan también los principales productos vendidos por la empresa motivo de estudio en base a la aplicación de Pareto a los datos de ventas totales de aditivos del año 2019, en donde se determina que alrededor del 80% de las ventas se centran en 13 productos que al ser divididos en segmentos resultan en 5: Fuera de Borda, Diésel, Gasolina, VM (Modificador de Viscosidad) y BN Booster.

Posteriormente se lleva a cabo la formación de la serie temporal Fuera de borda a partir de los datos de ventas mensuales de este producto; mientras que, las series temporales de Gasolina y Diesel son constituidas a partir de la suma de las ventas mensuales de los productos de estos segmentos, agrupados como se describe a continuación:

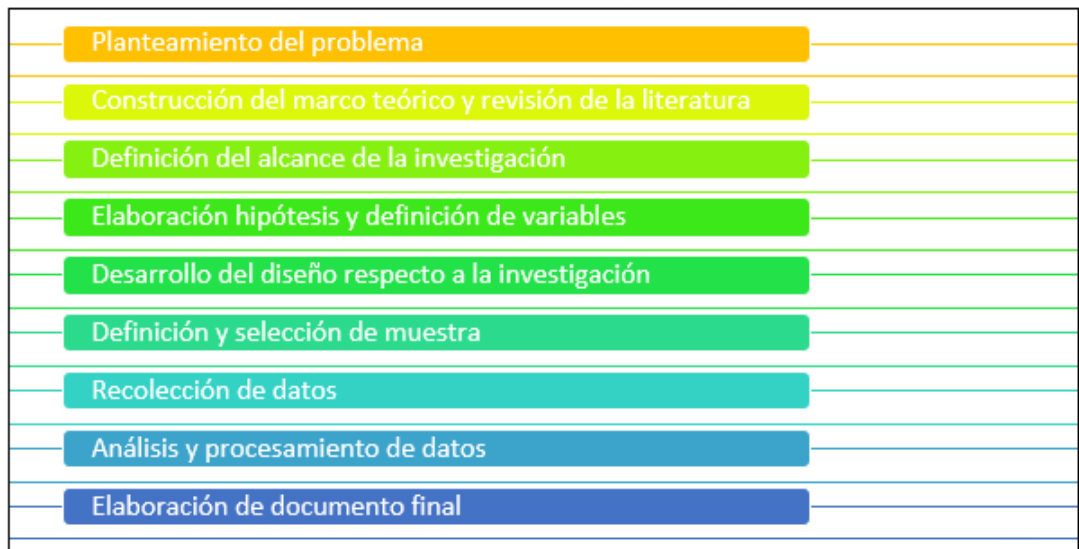
Gasolina: Gasolina 1, Gasolina 2 y Gasolina 3.

Diésel: Diésel 1, Diésel 2, Diésel 3, Diésel 4 y Diésel 5.

En base a las series temporales obtenidas y sus modelos generados correspondientes se llevan a cabo pronósticos anuales que inician en enero 2019 y finalizan en diciembre 2019, cada uno de estos pronósticos es comparado con los datos reales que se tienen de este mismo periodo con la finalidad de determinar si los modelos obtenidos representan el comportamiento de nuestra variable de respuesta asociada a la serie temporal. PI

A continuación, se detallan las fases de investigación aplicadas al presente estudio:

Figura No.3.1 Fases de investigación aplicada



Fuente: (Huilca, 2020)

Figura No.3.2 Arquitectura de Ciencia de Datos



Fuente: Autor

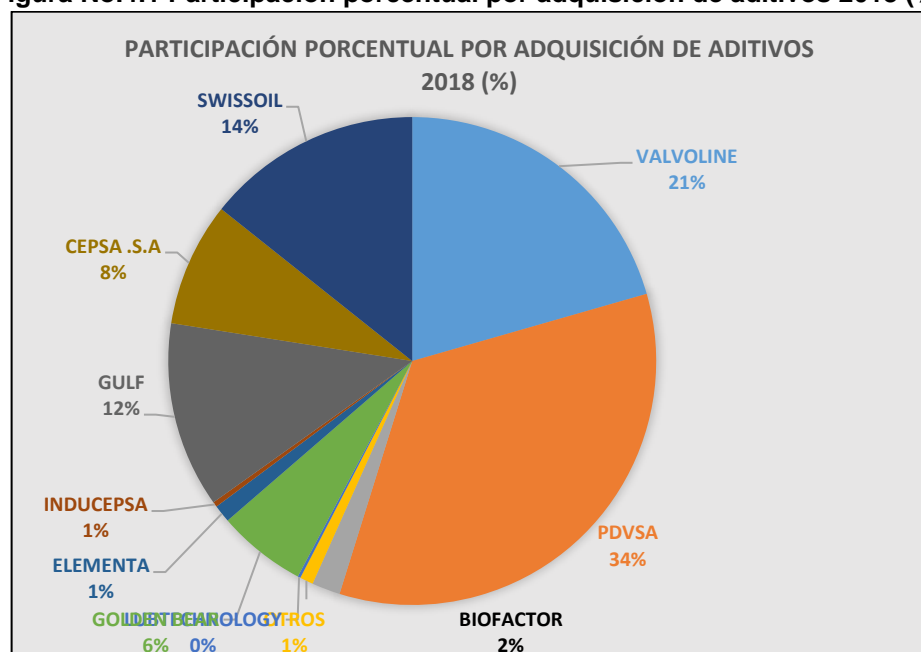
CAPÍTULO 4

4. RESULTADOS: APLICACIÓN DE LA METODOLOGÍA BOX-JENKINS

4.1. ANÁLISIS ESTADÍSTICO DESCRIPTIVO DE DATOS

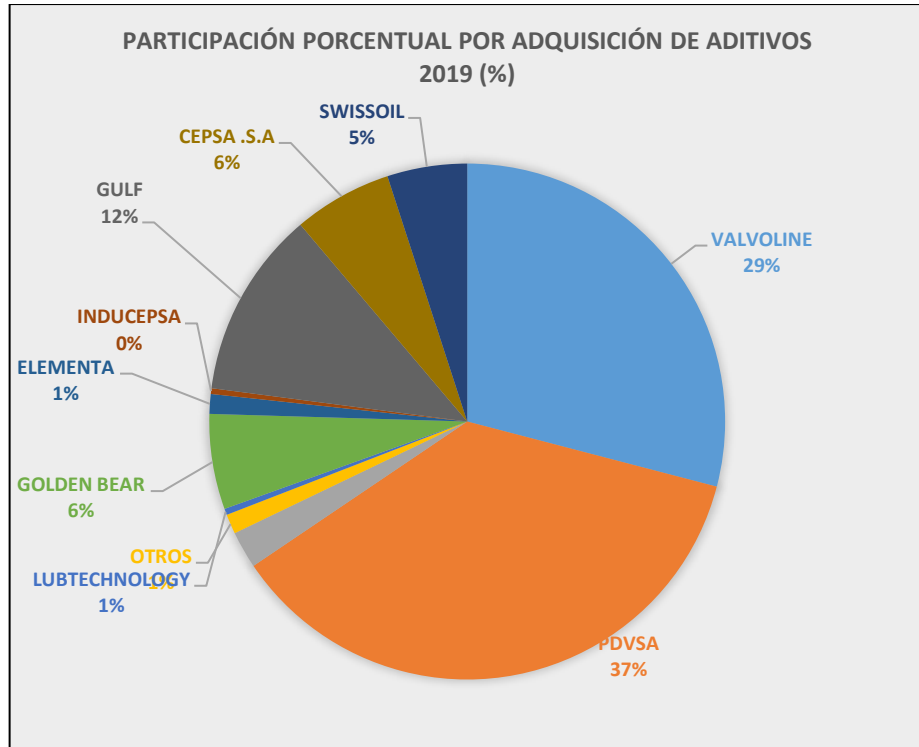
La empresa comercializadora e importadora de aditivos para lubricantes cuenta con una cartera de clientes, principalmente fábricas de lubricantes o comercializadoras de lubricantes, las cuales son descritas a continuación: Biofactor, Cepsa, Gulf, Golden Bear, Valvoline, Lubtechnology, Pdvsa, Swissoil, Elementa entre otros. En las siguientes figuras, observamos las participaciones porcentuales de los años 2018 y 2019 en base a las ventas anuales de aditivos a los clientes mencionados, en donde evidenciamos que el cliente "Swissoil" reduce su participación porcentual de 14.30% a 5.00% y ocasionando en consecuencia una disminución en las cifras de las ventas totales del año 2019.

Figura No.4.1 Participación porcentual por adquisición de aditivos 2018 (%)



Fuente: Análisis comparativo de ventas 2018 de la empresa comercializadora de aditivos.

Figura No.4.2 Participación porcentual por adquisición de aditivos 2019(%)



Fuente: Análisis comparativo de ventas 2018 de la empresa comercializadora de aditivos.

Tabla No. 4.1. Participación porcentual por adquisición de aditivos

Participación porcentual por adquisición de aditivos (Clientes)			
CLIENTES	2.018	2.019	Var (%)
VALVOLINE	20,56%	29,06%	8,51%
PDVSA	34,26%	36,48%	2,22%
BIOFACTOR	1,88%	2,33%	0,45%
OTROS	0,89%	1,24%	0,35%
LUBTECHNOLOGY	0,18%	0,39%	0,20%
GOLDEN BEAR	5,89%	5,97%	0,08%
ELEMENTA	1,17%	1,23%	0,07%
INDUCEPSA	0,34%	0,36%	0,02%
GULF	12,32%	11,74%	-0,57%
CEPSA .S.A	8,22%	6,19%	-2,03%
SWISSOIL	14,30%	5,00%	-9,29%
Total	100%	100%	

Fuente: Análisis comparativo de ventas 2018-2019 de la empresa comercializadora de aditivos.

Por lo anteriormente expuesto y de acuerdo a los datos proporcionados por la empresa comercializadora de aditivos durante los años 2018 y 2019 las ventas totales en expresadas en kilogramos fueron de 3.116.05 y 2.776.32 respectivamente.

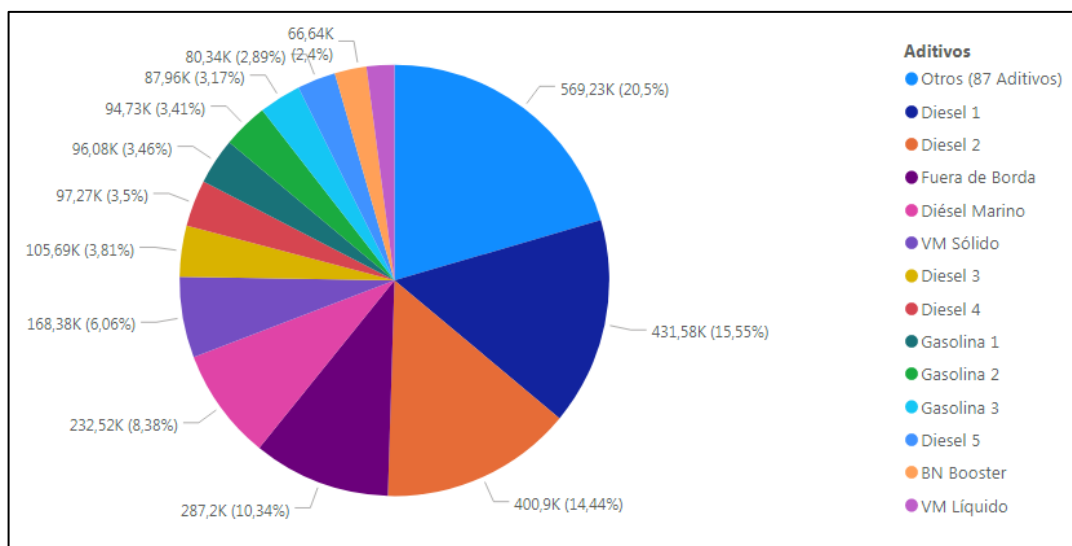
Tabla No. 4.2. Comparación de ventas de aditivos 2018-2019

Detalle de Aditivos	Aditivos	KG	Toneladas
Dejados de consumir	18	-	0
Bajaron consumo	41	(951.653)	-952
Nuevos consumos	28	135.498	135
Mantuvieron consumo	2	-	0
Incrementaron consumo	29	476.422	476
Descripción	2018	2019	Toneladas
Consumo total KG	3.116.05	2.776.32	
Aumento/Diminución KG		(339.73)	-340
% con respecto al año anterior		-10,90%	

Fuente: Análisis comparativo de ventas 2018-2019 de la empresa comercializadora de aditivos.

La empresa importadora y comercializadora de aditivos cuenta en la actualidad con alrededor de 100 aditivos en su portafolio de productos para la producción de lubricantes y grasas para los diferentes sectores productivos (construcción, minería, agricultura, etc.), y para efectos de comprender en qué aditivos se concentra el mayor volumen de ventas de aditivos de la empresa observamos a continuación las ventas totales de aditivos durante el año 2019 expresadas en kilogramos y en sus equivalencias porcentuales:

Figura No.4.3 Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019

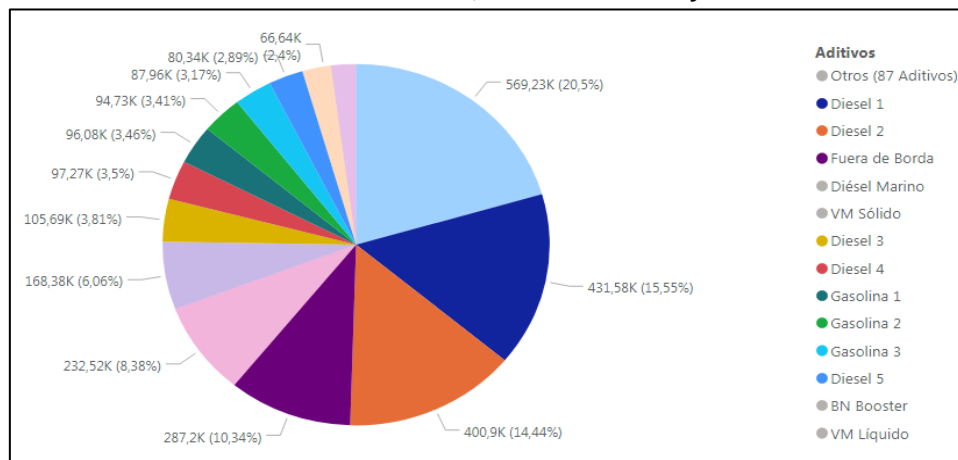


Elaborado por: Autor

En la figura anterior podemos observar que el 20,50% de las ventas de aditivos durante el 2019, se concentra en un total de 87 aditivos; mientras que el 79,50% restante se localiza en tan sólo 13 aditivos: Diésel 1, Diésel 2, Fuera de Borda, Diésel Marino, VM Sólido, Diésel 3, Diésel 4, Gasolina 1, Gasolina 2, Gasolina 3, Diésel 5, BN Booster y VM líquido.

Dentro del presente estudio, analizaremos estrictamente el comportamiento del 60,57% de las ventas de la empresa comercializadora de aditivos a través del estudio de tres segmentos de aditivos: Diésel, Fuera de Borda y Gasolina.

Figura No.4.4 Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019. Aditivos: Diésel, Fuera de Borda y Gasolina.

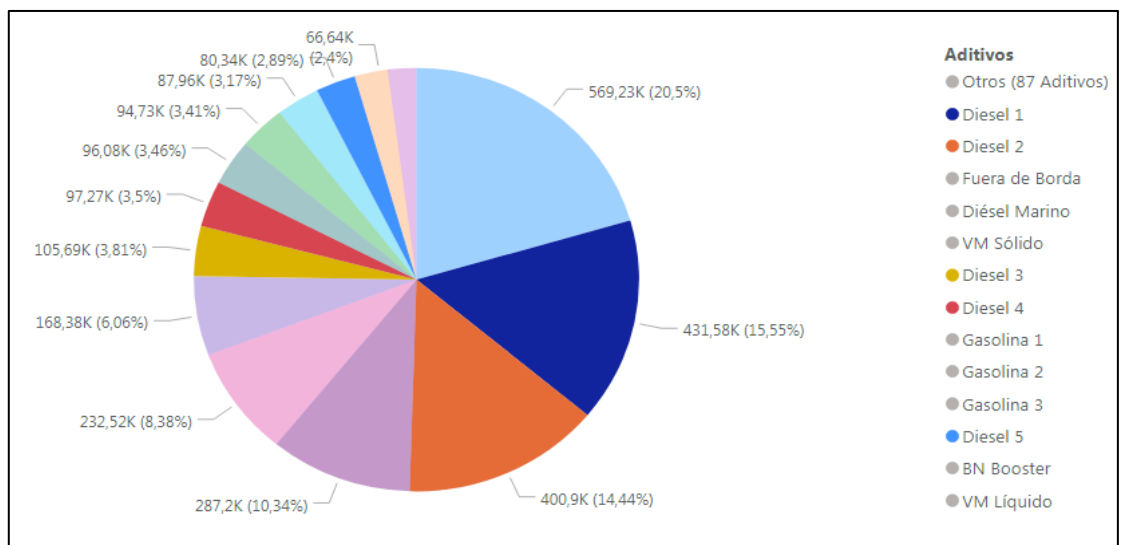


Elaborado por: Autor

(En esta figura se encuentran resaltadas los aditivos Diésel 1, Diésel 2, Fuera de Borda, Diésel 3, Diésel 4, Gasolina 1, Gasolina 2, Gasolina 3, Diésel 5).

- Diésel (40,19%): segmento conformado por los aditivos Diésel 1, Diésel 2, Diésel 3, Diésel 4 y Diésel 5.

Figura No.4.5 Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019. Aditivos: Diésel.

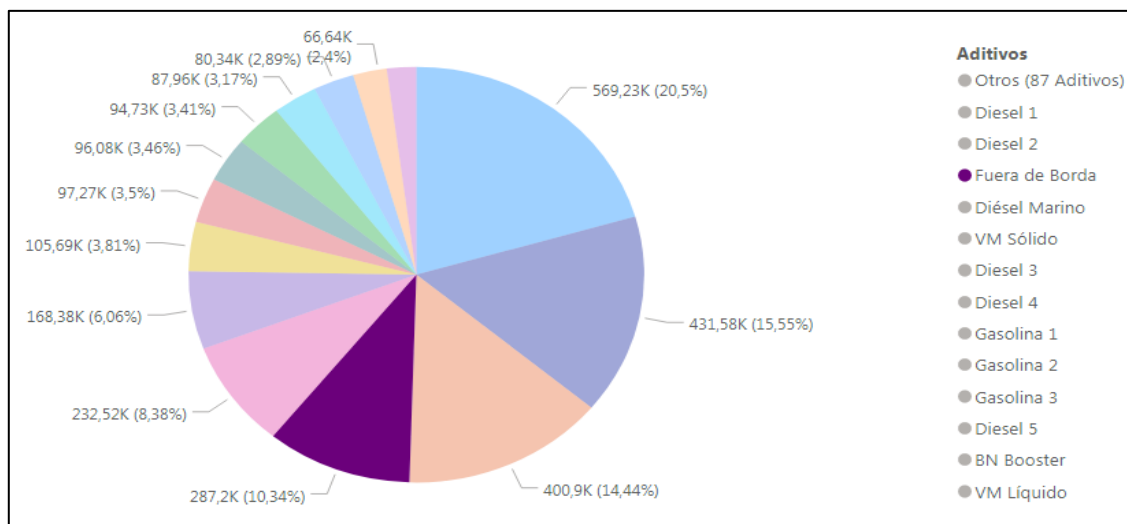


Elaborado por: Autor

(En esta figura se encuentran resaltadas los aditivos Diésel 1, Diésel 2, Diésel 3, Diésel 4, Diésel 5).

- Fuera de Borda (10,34%): segmento conformado por el aditivo Fuera de Borda.

Figura No.4.6. Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019. Aditivo: Fuera de Borda.

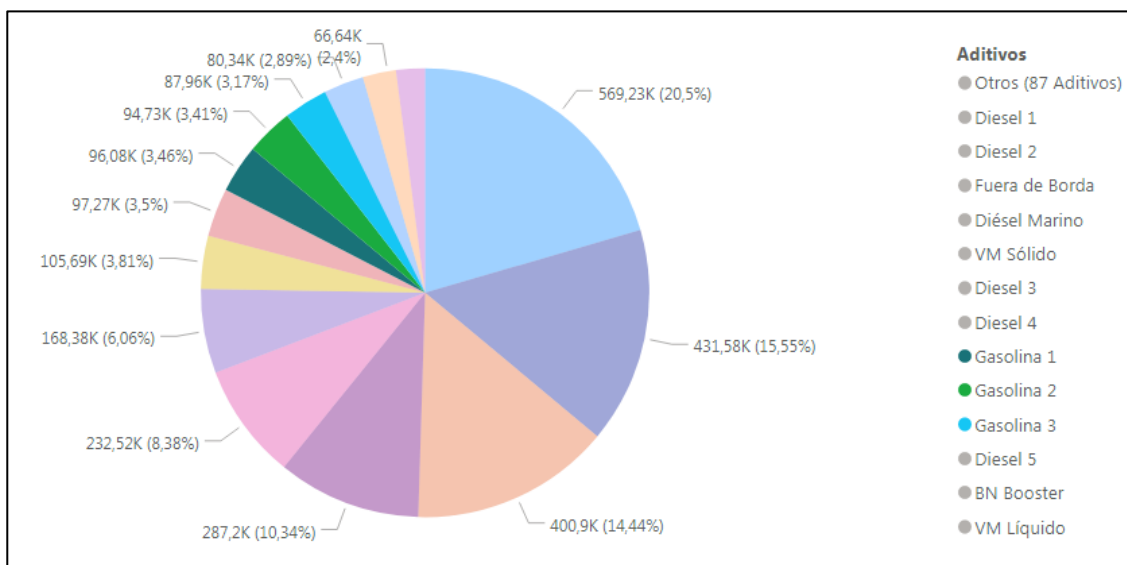


Elaborado por: Autor

(En esta figura se encuentra resaltado en aditivo Fuera de Borda).

- Gasolina (10,04%): segmento conformado por los aditivos Gasolina 1, Gasolina 2 y Gasolina 3.

Figura No.4.7. Segmentación de Ventas Totales Anuales de Aditivos en kilogramos, año 2019. Aditivos: Gasolina.



Elaborado por: Autor

(En esta figura se encuentran resaltadas los aditivos Gasolina 1, Gasolina 2, Gasolina 3).

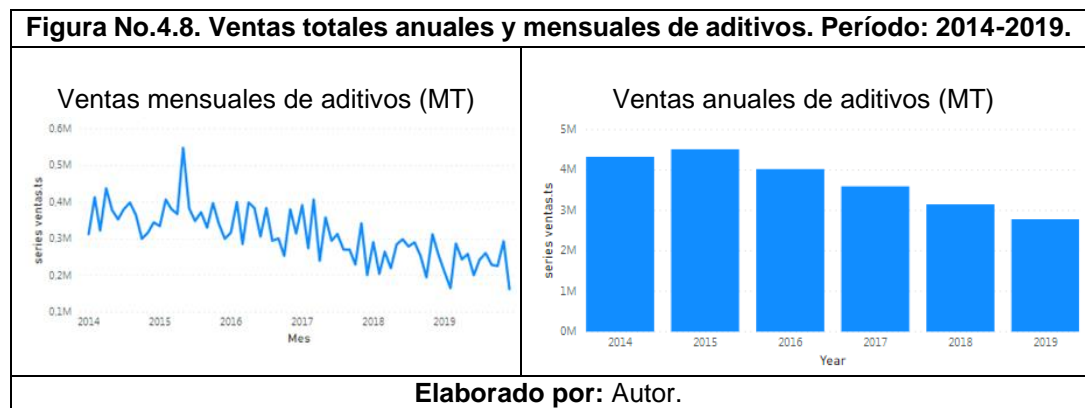
A fin de obtener un entendimiento más amplio tanto de las ventas totales de los aditivos como de los segmentos Diésel, Gasolina y Fuera de borda, analizaremos

el comportamiento de los grupos anteriormente mencionados durante los 6 últimos años (2014 a 2019).

4.1.1. Ventas Totales

Para el caso de las ventas totales observamos un decrecimiento continuo a partir del año 2015, de acuerdo a lo indicado por la empresa comercializadora e importadora de aditivos esto se debe principalmente a:

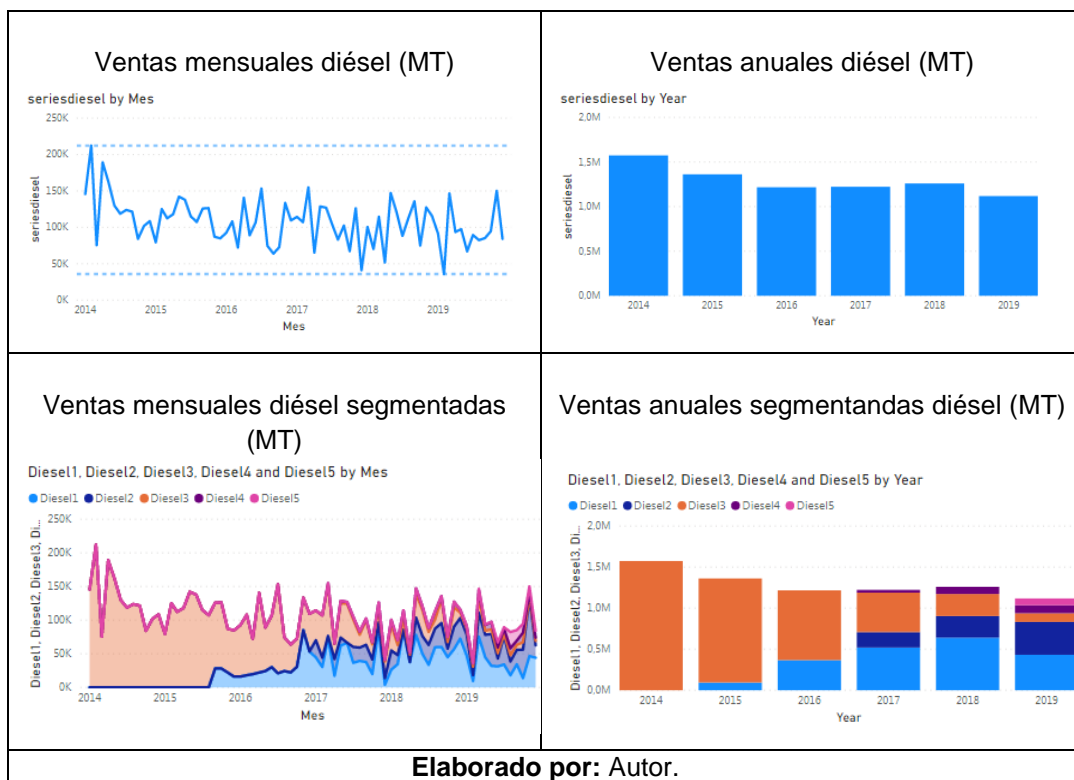
- Reducción de consumo de aditivos para la fabricación de lubricantes para motores de generación de energía eléctrica a partir de combustibles fósiles que fueron reemplazados plantas hidroeléctricas dentro del territorio ecuatoriano.
- Extensión de intervalos de cambio de aceites lubricantes en las flotas automotrices tanto de vehículos ligeros como pesados, a causa de mejores tecnologías de aditivos y aceites básicos que permiten vidas útiles prolongadas en los lubricantes; no obstante, la extensión de la vida útil del lubricante es equivalente a menor consumo de materias primas a largo plazo.
- Reducción de las ventas al cliente "Swissoil" por cambio de tecnología de aditivos hacia competidores.
- Incremento de las ventas de lubricantes terminados importados dentro del mercado de lubricantes ecuatorianos, que durante el año 2019 alcanzaron un 40% de las ventas totales de lubricantes en el país (APEL, 2019).



4.1.2. Segmento Diésel

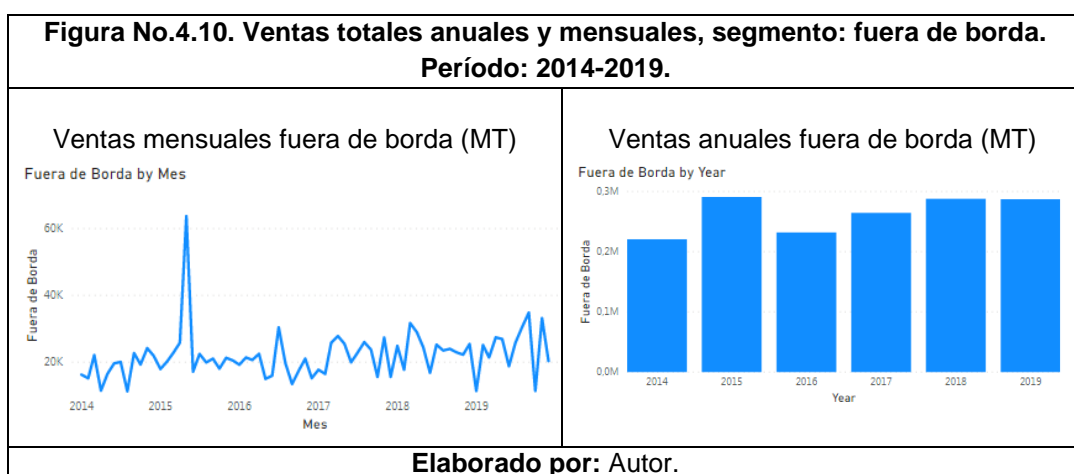
Para el caso de las ventas de los aditivos del segmento Diésel, observamos que el aditivo Diésel 3 ha sido paulatinamente reemplazado a partir del año 2015 por los aditivos Diésel 1, Diésel 2, Diésel 4 y Diésel 5, lo cual es congruente con lo observado en el comportamiento de las ventas totales de aditivos; puesto que estos últimos productos contienen tecnologías más avanzadas como consecuencia de los requerimientos del mercado de transporte pesado y de los motores alimentados por Diésel.

Figura No.4.9. Ventas totales anuales y mensuales, segmento: diésel. Período: 2014-2019.



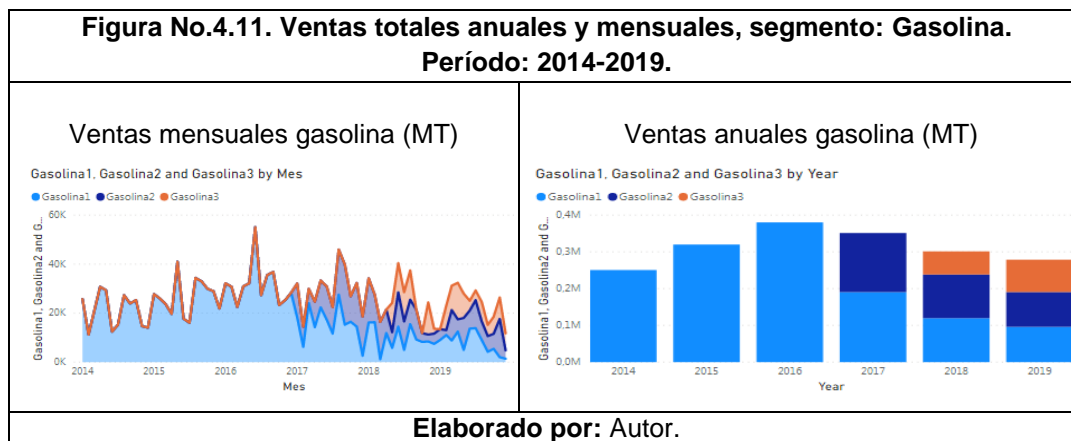
4.1.3. Segmento Fuera de Borda

De acuerdo al comportamiento en las ventas mensuales y anuales del aditivo “Fuera de borda”, estas han presentado crecimiento durante los últimos 4 años. De acuerdo a lo indicado por la empresa motivo de estudio, esto se debe principalmente por un negocio para la fabricación de lubricantes “Fuera de Borda” adjudicado al cliente PDVSA que constituye en la actualidad un 33% de las ventas de aditivos.



4.1.4. Segmento Gasolina

Acorde a lo observado en el comportamiento de los aditivos del segmento gasolina, podemos decir que se presenta un fenómeno similar al indicado dentro del segmento diésel en donde tecnologías con mejores rendimientos (Gasolina 2 y Gasolina 3) reemplazan a la más antigua (Gasolina 1). La reducción evidenciada a partir del año 2016 podría deberse a la extensión de periodos extendidos de cambio de lubricantes, pérdida de clientes cautivos y efecto del ingreso de lubricantes terminados importados al mercado de lubricantes ecuatorianos.



4.2. SEGMENTACIÓN DE PRODUCTOS

De acuerdo a lo verificado en el sistema XASS, utilizado actualmente en la empresa motivo de estudio, se mantienen activos 100 códigos; es decir, se planifica la importación, administración y comercialización de 100 aditivos. De estos 100 ítems, considerando la aplicación de Pareto como fue verificado en el numeral de estadística descriptiva de los datos, vemos que el 79,50% de las ventas se encuentran en los primeros 13 productos de la siguiente manera:

Tabla No. 4.3. Aplicación de Pareto a las ventas totales de aditivos del año 2019, expresadas en kilogramos.

No.	Descripción de Aditivo	Total (kg)	Ventas acum.	% acum.
1	Diésel 1	431.582	431.582	15,55%
2	Diésel 2	400.895	832.477	29,98%
3	Fuera de Borda	287.199	1.119.676	40,33%
4	Diésel Marino	232.522	1.352.198	48,70%
5	VM Sólido	168.375	1.520.573	54,77%
6	Diésel 3	105.688	1.626.261	58,58%
7	Diésel 4	97.271	1.723.532	62,08%
8	Gasolina 1	96.078	1.819.610	65,54%
9	Gasolina 2	94.731	1.914.341	68,95%
10	Gasolina 3	87.958	2.002.299	72,12%
11	Diésel 5	80.340	2.082.639	75,01%
12	BN Booster	66.640	2.149.279	77,41%
13	VM Líquido	57.807	2.207.086	79,50%
Ventas Totales (kg)			2.776.318 (*)	100,00%

Fuente: Empresa Comercializadora e Importadora de Aditivos.

Por lo anteriormente expuesto, dentro de este estudio se realizarán modelos específicos para las ventas totales mensuales de la empresa importadora y comercializadora de aditivos para lubricantes, así como de los principales segmentos especificados y de los cuales se han recopilado los valores históricos de ventas de los periodos 2014-2019, siendo estos: Fuera de Borda, Gasolina y Diésel.

(*) Para el desarrollo de la Tabla 4.3 se tomó en consideración el total anual de las ventas en kilogramos del año 2019 y se calculó la participación porcentual de cada ítem.

4.3. MODELO: VENTAS TOTALES

4.3.1. ANÁLISIS EXPLORATORIO DE DATOS

En esta sección se llevará a cabo la presentación de la serie mensual de las ventas totales de aditivos para la fabricación de lubricantes expresadas en kilogramos de una empresa importadora y comercializadora de materias primas, correspondientes a los periodos 2014 a 2019, obtenidos a través de los registros de ventas de la empresa importadora y comercializadora de materias primas.

Tabla No. 4.4. Serie histórica mensual de las ventas totales de aditivos para la fabricación de lubricantes expresadas en kilogramos, periodo 2014-2019.

Ventas de Aditivos Mensuales Totales (kg)						
Mes / Año	2014	2015	2016	2017	2018	2019
Enero	313112	335111	317305	391723	290933	210574
Febrero	413410	407030	400328	274736	204831	165847
Marzo	323048	381727	286297	407099	264653	287081
Abril	437745	367977	398936	240462	220089	244155
Mayo	378865	548072	384288	358111	284971	258600
Junio	353337	383300	306678	294899	298722	201017
Julio	381633	349017	383932	312784	279201	243786
Agosto	398667	372171	295066	271109	290364	260726
Septiembre	364594	330944	301077	270771	252761	228957
Octubre	300271	397476	253913	230289	194997	226055
Noviembre	316767	340178	380346	342062	312266	293123
Diciembre	344802	300271	314803	201535	257273	163112

Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

Considerando la tabla anterior, correspondiente a los registros de ventas mensuales (expresadas en kilogramos) de una empresa comercializadora de aditivos para lubricantes procedemos a ejecutar el análisis exploratorio de los datos con el fin de examinar los datos involucrados en el presente estudio antes de aplicar cualquier técnica estadística, de esta manera podremos conseguir un entendimiento básico de nuestros datos y de las posibles relaciones existentes entre las variables a ser analizadas. Para llevar a cabo esta exploración de nuestros datos utilizaremos el Software libre R, transformándolos a series temporales y los representaremos

gráficamente con el fin de verificar la posible presencia principalmente de tendencias o estacionalidades que den evidencia de la no-estacionariedad de la serie.

En primera instancia, procedemos con la importación de nuestros datos a R y la activación de las librerías correspondientes (1).

(1) *Nota de programación: Los códigos utilizados en el presente estudio se encuentra agregados en Anexos*

Puesto que, los datos son recibidos en kilogramos procedemos con la transformación de estos datos a toneladas métricas dividiendo la serie de tiempo para 1000, considerando la equivalencia: 1000 kg =1 MT.

Tabla No. 4.5. Serie histórica mensual de las ventas totales de aditivos para la fabricación de lubricantes expresadas en toneladas métricas, periodo 2014-2019.

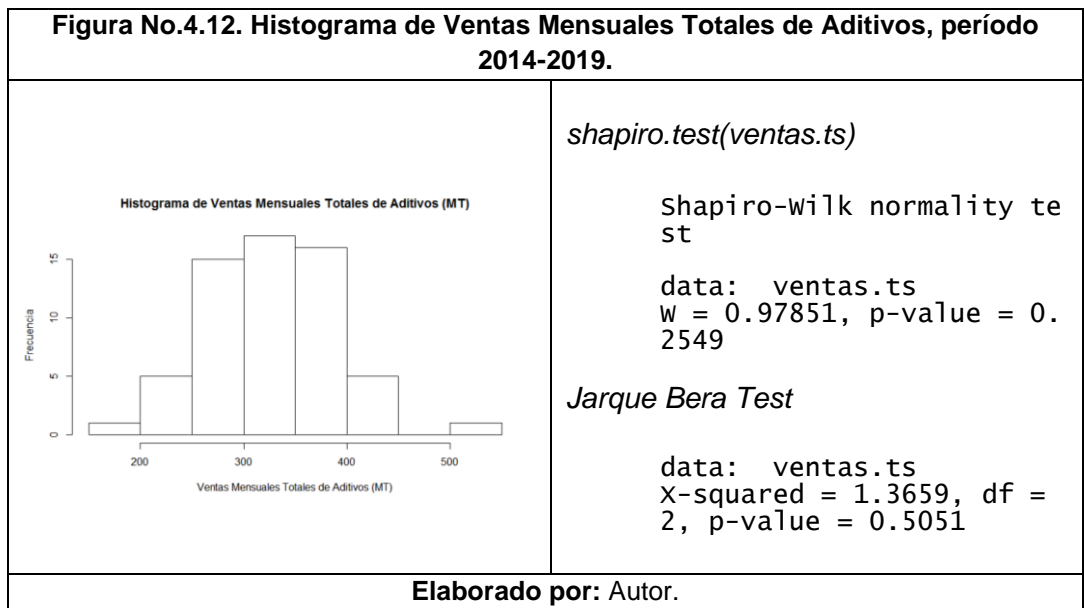
Ventas de Aditivos Mensuales Totales (MT)						
Mes / Año	2014	2015	2016	2017	2018	2019
Enero	313	335	317	392	291	211
Febrero	413	407	400	275	205	166
Marzo	323	382	286	407	265	287
Abril	438	368	399	240	220	244
Mayo	379	548	384	358	285	259
Junio	353	383	307	295	299	201
Julio	382	349	384	313	279	244
Agosto	399	372	295	271	290	261
Septiembre	365	331	301	271	253	229
Octubre	300	397	254	230	195	226
Noviembre	317	340	380	342	312	293
Diciembre	345	300	315	202	257	163

Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

Posteriormente generamos un histograma de frecuencias con la finalidad de representar la distribución de los datos por medio de barras, dándonos un vistazo general del comportamiento de las variables, su distribución, dispersión, aleatoriedad o tendencia.

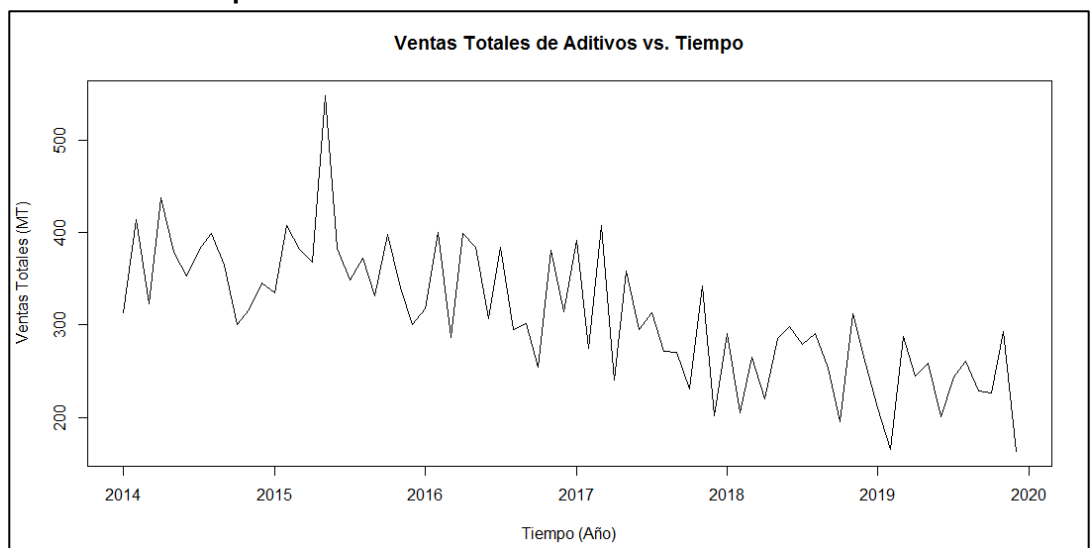
En el histograma que observamos en la siguiente tabla, podemos evidenciar que nuestros datos muy posiblemente tienen una distribución normal.



Por lo que adicionalmente llevamos a cabo una prueba de normalidad Shapiro-Wilk, obteniendo un valor p de la prueba de 0.26 y que es superior a 0.05, no rechazamos la hipótesis nula por lo que se puede concluir que los datos de las ventas siguen una distribución normal. Lo cual es congruente también con el valor p de 0.51 obtenido en la prueba jarque.bera, en donde tampoco rechazamos la hipótesis nula de que las ventas totales de aditivos son normales.

Continuando el análisis, procedemos a graficar los datos de nuestra serie respecto al tiempo, la cual nos permite observar la probable presencia de una tendencia descendente.

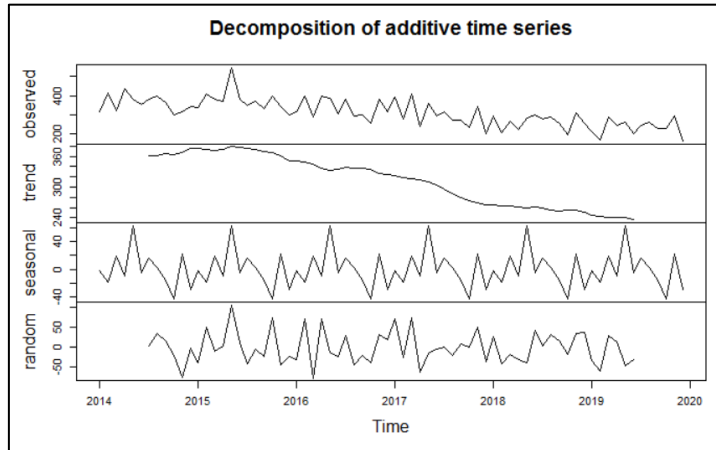
Figura No.4.13. Serie temporal de las ventas totales mensuales de aditivos expresada en toneladas métricas. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Con el fin de obtener más detalles de su comportamiento, descomponemos la serie temporal en todos sus posibles componentes (tendencia, estacional y aleatorio) a través de la descomposición estacional clásica de medias móviles.

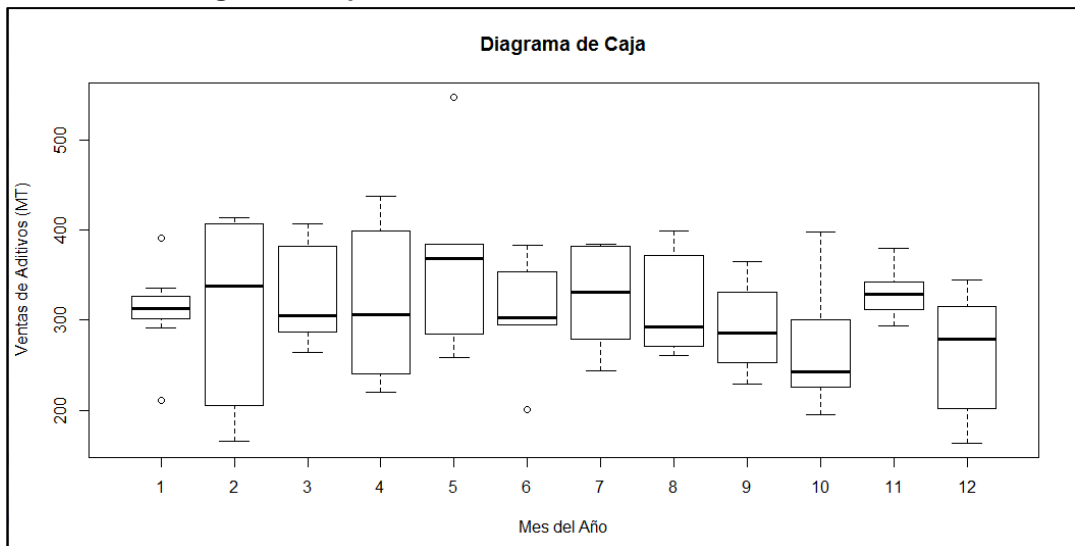
Figura No.4.14. Descomposición de la Serie Temporal de las ventas totales mensuales de aditivos. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Como resultado, en la gráfica anterior confirmamos la existencia de una tendencia descendente y la presencia del componente estacional y el aleatorio. Considerando que son datos de ventas mensuales y que se presentan cortes anuales, continuamos con la observación del comportamiento de los datos segmentándolos por meses siendo *Enero=1, Febrero=2....Diciembre=12*.

Figura No.4.15. Diagrama de caja de la serie temporal de las ventas totales de aditivos segmentada por mes, en toneladas métricas. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Como podemos ver en el diagrama de caja anterior los valores de las medias de las ventas totales mensuales tienen comportamientos diferentes entre sí.

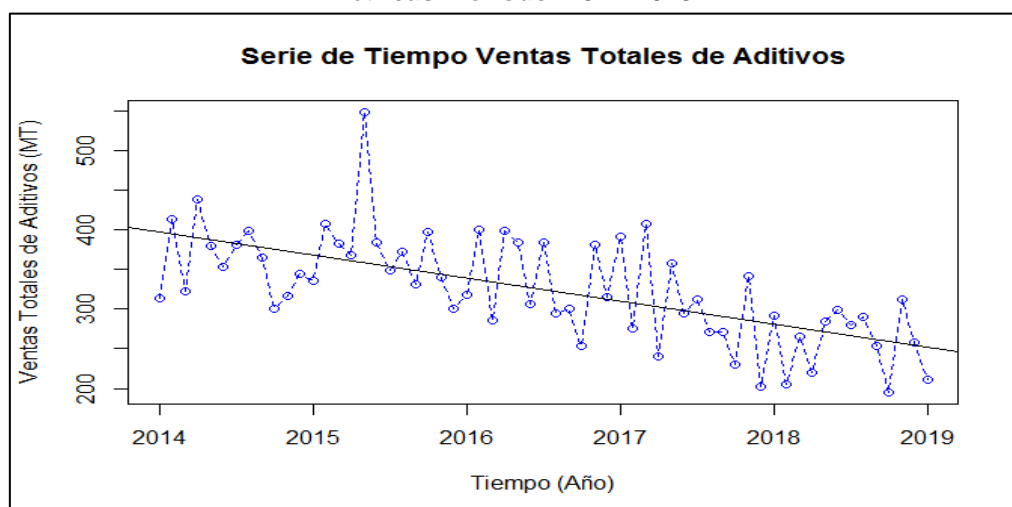
Adicionalmente observamos que los valores en términos de media más altos se presentan durante el mes de mayo; mientras que, durante el mes de febrero existe una variabilidad muy alta considerando que la dispersión observada es la más amplia de los valores recopilados. Adicionalmente podemos verificar que en el mes de noviembre se produce otro fenómeno, en donde el valor de media es el segundo más alto a lo largo de los años y su variabilidad es baja, esto se debe principalmente a que las empresas productoras de lubricantes abastecen sus bodegas tanto de materias primas como de productos terminado teniendo en cuenta que sólo producirán los primeros 15 días del mes de diciembre. Por otra parte, observamos ciertos datos anómalos en la gráfica que antecede, aquellos que se encuentran por encima del valor de la media de los meses enero y mayo fueron el resultado de un crecimiento abrupto en la generación termoeléctrica durante el año 2015 y las ventas de aditivos relacionados a este segmento, mientras que los que se encuentran por debajo de la media en los meses de enero y junio son por el contrario consecuencia de falta de abastecimiento de aditivos ocasionando una disminución directa en las ventas.

4.3.2. ANÁLISIS DE TENDENCIA

Una vez que concluimos el análisis exploratorio de nuestros datos, procedemos a aplicar la metodología Box-Jenkins; para esto es necesario que trabajemos únicamente con los datos de los primeros 5 años (2014 a 2018) para generar el modelo de pronóstico y utilicemos los datos correspondientes al año 2019 para comprobar qué tan cerca se encuentra nuestro modelo de la realidad.

Por lo anteriormente expuesto, procedemos a graficar únicamente el periodo 2014-2018 para verificar si estos se comportan de forma estacionaria:

Figura No.4.16. Serie temporal de las ventas totales de aditivos, en toneladas métricas. Período: 2014-2018.

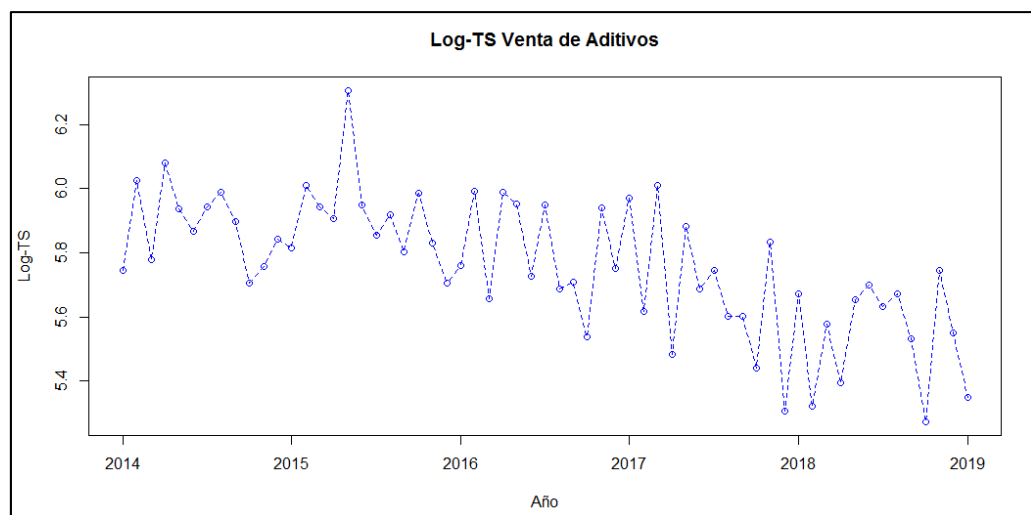


Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

En la gráfica anterior podemos evidenciar que nuestra serie de tiempo no es estacionaria respecto a su media lo cual es corroborado a través de la aplicación de la prueba Phillips-Perron que nos indica que debemos aplicar 1 diferencia a nuestra serie para alcanzar la estacionariedad; adicionalmente, como habíamos observado en la gráfica de descomposición existe una tendencia de datos decreciente, contiene también un componente estacional y sus fluctuaciones en el tiempo no son uniformes; es decir, tiene variaciones irregulares.

Por lo expresado en el párrafo anterior y dado que debemos eliminar las variaciones irregulares de nuestra serie de datos, aplicamos logaritmos a nuestra serie y los graficamos para verificar los cambios en su comportamiento:

**Figura No.4.17. Logaritmo de la serie temporal de las ventas totales de aditivos.
Período: 2014-2018.**



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

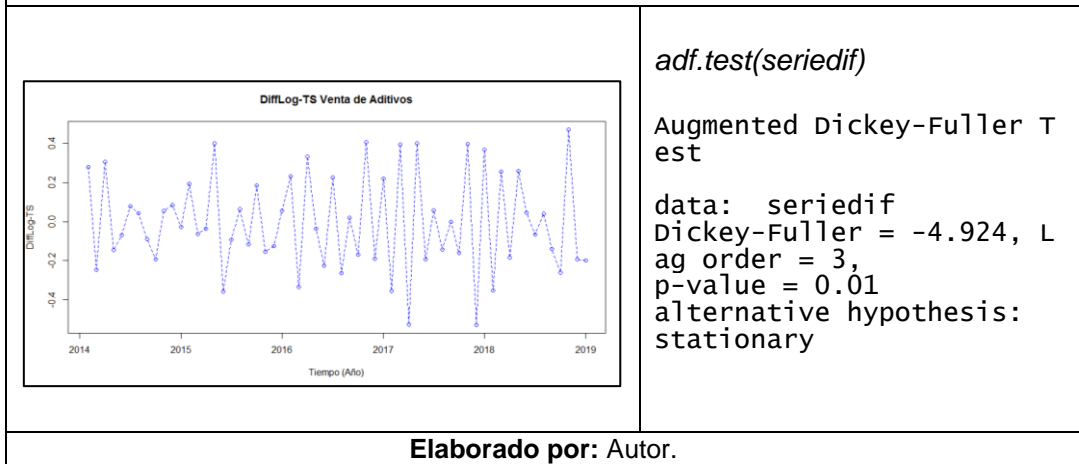
Elaborado por: Autor

Una vez que hemos aplicado los logaritmos a nuestra serie, creando una nueva serie llamada *serielog*, podemos ver que el rango de diferencia entre los valores mensuales es muy inferior a la serie inicial *ventas.ts*; y de esta manera conseguimos que esta nueva serie tenga una variabilidad regular.

De acuerdo a lo expresado por Box-Jenkins, con la finalidad de aplicar esta metodología de modelos autorregresivos, precisamos conseguir una serie de datos al menos débilmente estacionaria en términos media y para lograr esto necesitaríamos también remover la tendencia anteriormente observada; es por esto que, aplicamos una primera diferencia a nuestra serie.

Es recomendable también llevar a cabo un test que compruebe que lo efectuado es estadísticamente adecuado, por lo que procedemos a utilizar la Prueba Aumentada de Dickey-Fuller. Al ser el valor p obtenido menor que 0.05 rechazamos la hipótesis nula de que la serie temporal posee raíz unitaria, lo cual significaría que nuestra serie es estacionaria.

Figura No.4.18. Diferencia del logaritmo de la serie temporal de ventas totales de aditivos. Período: 2014-2018.

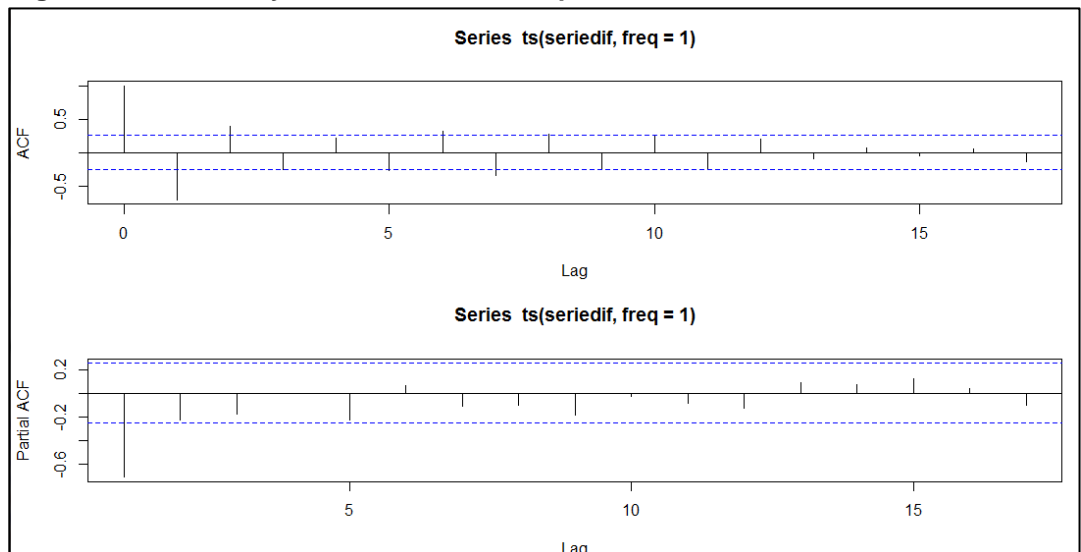


4.3.3. AJUSTE DEL MODELO

Considerando que nuestros datos se encuentran preparados y han cumplido todos los supuestos del modelo, con el objetivo de definir el orden del modelo que representará a los datos, son necesarias tres variables: p, d y q que representan el orden de las partes media autorregresivas, integradas y móviles.

Es por esto que procederemos con la visualización de las funciones de autocorrelación típicamente utilizadas dentro de la metodología Box-Jenkins: Función de Autocorrelación y Función de Autocorrelación Parcial.

Figura No.4.19. ACF y PACF de la serie temporal *seriedif*. Ventas Totales de Aditivos



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

Considerando lo expuesto en la figura anterior, en el gráfico de PACF (Función de autocorrelación parcial) observamos autocorrelaciones nulas a partir del primer rezago (lag); en base a esto es posible plantear un modelo AR de orden uno es decir el valor p será igual a 1; puesto que, el gráfico PACF nos indica el número de componentes que tendrá la parte AR de nuestro modelo. Por otra parte, comportamiento de observado en el gráfico ACF (Función de Autocorrelación) nos muestra detalles de la parte MA (Media Móvil), teniendo en cuenta que en este se evidencia un comportamiento de decaimiento exponencial muy probablemente tendrá un solo componente. En base a estos criterios procedemos a plantear los diferentes modelos del presente estudio.

4.3.3.1. Modelos ARIMA Estacionales (p,d,q)x(P,D,Q)s

Para plantear un modelo arima estacional o arima estacional multiplicativo, recorrimos los comportamientos de modelos previos de sus componentes autorregresivos y de medias móviles, probamos inicialmente un modelo AR(1) en donde el valor de p tomaba el valor de 1 conforme a la gráfica de PACF, en donde observamos que se supera el límite mínimo en el primer retardo. Debido a que el comportamiento de este no fue satisfactorio en la prueba de Box-Lung ($p = 0.26$), llevamos también acabo un análisis con AR (2), el cual presentó un valor mayor en la prueba de Box-Lung ($p = 0.45$). Para complementar el modelo arima con características estacionales, complementamos el planteamiento del modelo en base a lo expresado anteriormente, donde P tomaría el valor de 1, conforme a la gráfica PACF, D equivaldría a 1 por la aplicación de diferencias y Q sería igual a 1 conforme a la gráfica ACF y el decaimiento exponencial de la gráfica. Finalmente, agregamos a nuestro sistema de estudio las características estacionales de los datos de ventas de aditivos de la empresa comercializadora. Es decir, para este caso particular, al ser datos anuales, el periodo s será igual a 12; de tal forma planteamos el siguiente modelo:

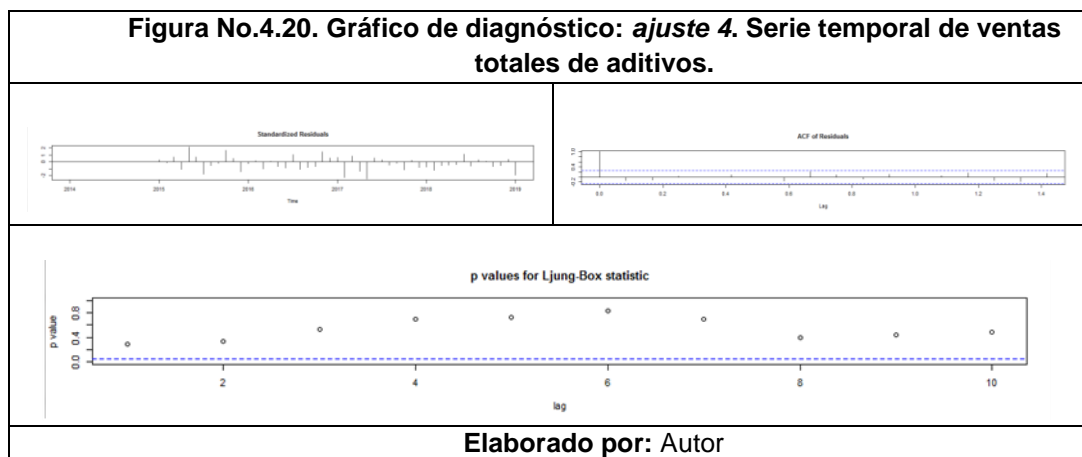
Tabla No. 4.6. Resumen Estadístico - Orden (2,0,0)x(1,1,1)₁₂, Serie Temporal Ventas Totales					
Orden: (2,0,0)x(1,1,1) ₁₂					
Coefficients:					
	ar1	ar2	sar1	sma1	
	0.1203	0.4883	-0.7163	0.2169	
s.e.	0.1247	0.1344	0.2652	0.4065	
sigma ² estimated as 3112: log likelihood = -264,					
aic = 538					
Training set error measures:					
	ME	RMSE	MAE	MPE	MAPE
MASE					
Training set	-10.29527	49.89295	35.97753	-4.816182	11.66981
	5557482				0.
		ACF1			
Training set	-0.1247321				
Elaborado por: Autor					

Tabla No. 4.7. Estadístico Z - Orden (2,0,0)x(1,1,1) ₁₂ , Serie Temporal Ventas Totales					
z test of coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
ar1	0.12032	0.12470	0.9649	0.3346084	
ar2	0.48829	0.13443	3.6323	0.0002809	***
sar1	-0.71634	0.26523	-2.7009	0.0069160	**
sma1	0.21686	0.40645	0.5335	0.5936659	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Elaborado por: Autor					

Contrario a lo esperado, vemos que el comportamiento del estadístico p (0.32) en la prueba de ruido blanco (Ljung-Box) es el de valor más bajo en comparación con todos los modelos planteados AR(2) y ARIMA(2,1,1); a pesar de ser el que cuenta con más características ingresadas al modelo a partir de los diferentes coeficientes.

Tabla No. 4.8. Prueba Box-Ljung- Orden (2,0,0)x(1,1,1) ₁₂ , Serie Temporal Ventas Totales	
Box-Ljung test	
data: residuals(ajuste4)	
X-squared = 0.98095, df = 1, p-value = 0.322	
Elaborado por: Autor	



4.3.4. ELECCIÓN DEL MEJOR MODELO

En base a los modelos planteados en el presente estudio realizaremos un análisis comparativo para seleccionar el modelo que mejor represente el comportamiento de nuestra variable de respuesta.

Para esto, utilizamos los resultados de los Errores Medios de Porcentaje Absoluto (MAPE) de cada modelo calculado a través de la aplicación de *summary* a cada uno; y por otra parte debemos calcular también los criterios de información de *Akaike* (AIC) y criterios de información bayesiano (BIC). Una vez finalizadas estas

estimaciones generamos el cuadro comparativo de los modelos resultantes del presente estudio:

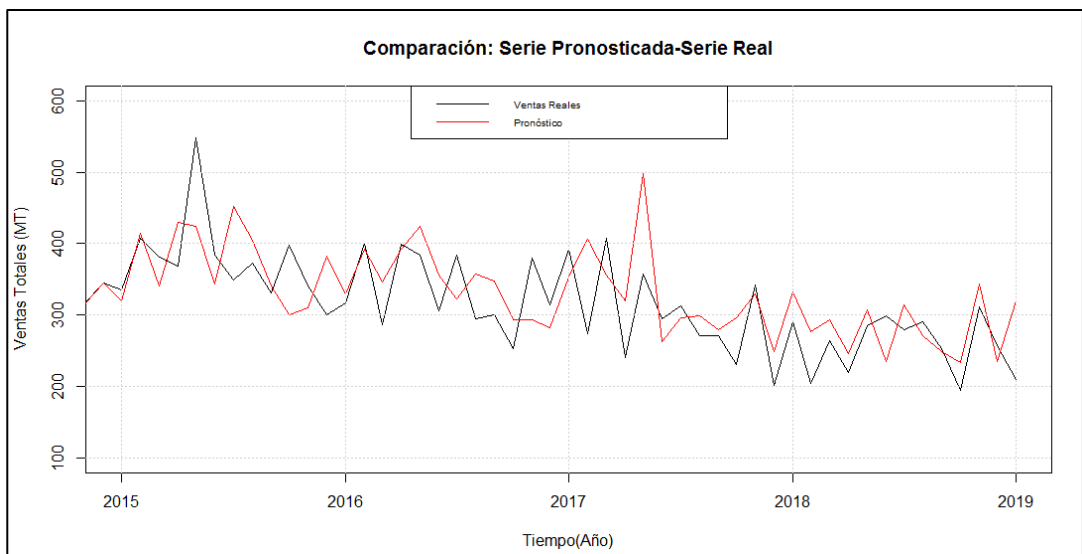
Tabla No. 4.9. Comparación de modelos resultantes. Serie temporal ventas totales de aditivos.

No.	MODELO	AIC	BIC	MAPE
1	ARIMA (1,0,0)	672,72	678,99	15,90
2	ARIMA (2,0,0)	654,58	662,96	13,28
3	ARIMA (2,1,1)	641,89	650,20	13,23
4	ARIMA (2,0,0)x(1,1,1) ₁₂	538	547,36	11,67

Elaborado por: Autor

De acuerdo con los criterios AIC y BIC, se selecciona el modelo ARIMA (2,0,0)x(1,1,1)₁₂, ya que presenta los menores valores en estos criterios. De forma similar, utilizando el criterio MAPE el modelo 4 volvió a ser seleccionado; puesto que también presenta el menor valor comparado con los demás modelos, lo cual indica que es el modelo que menos se desvía de los datos originales, por lo cual procedemos a graficar el modelo elegido en la siguiente figura:

Figura No.4.21. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de ventas totales de aditivos. Período: 2014-2018.



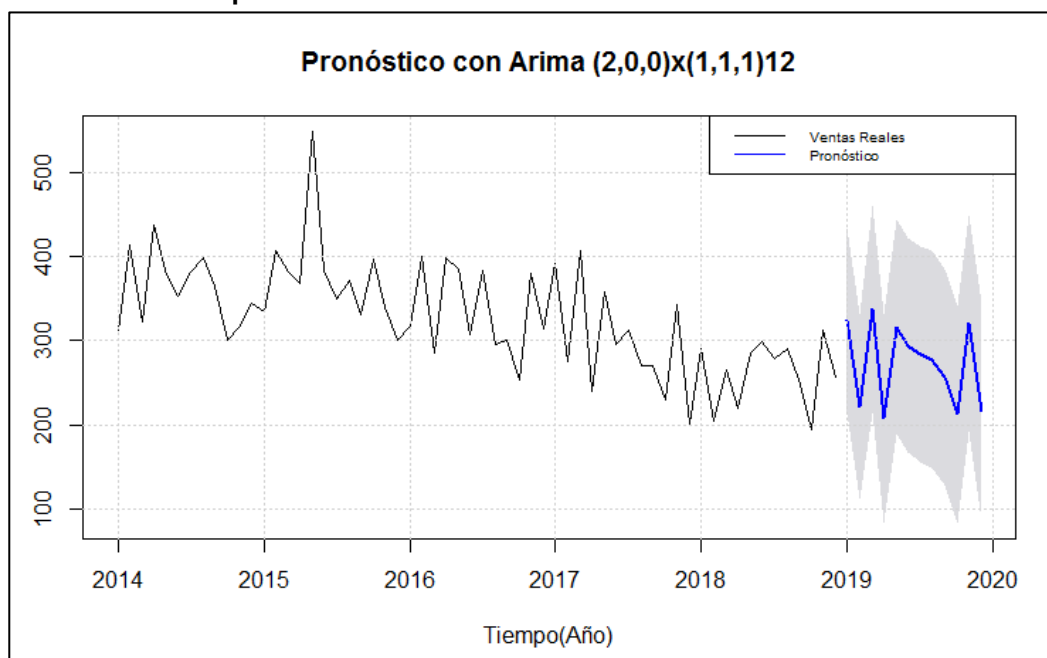
Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

4.3.5. PREDICCIÓN UTILIZANDO ARIMA

Una vez que hemos elegido el ajuste 4 como el más adecuado a nuestro conjunto de datos utilizaremos la función *forecast()* en R, con el fin de predecir o pronosticar los posibles valores futuros de nuestra serie temporal, determinando también intervalos de confianza:

Figura No.4.22. Pronóstico de ventas totales a través de Arima (2,0,0)x(1,1,1)12, expresado en toneladas métricas. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Adicionalmente, tabulamos los datos obtenidos dentro del presente pronóstico:

Tabla No. 4.10. Pronósticos resultantes del ajuste 4 (Periodo 2019). Serie temporal ventas totales

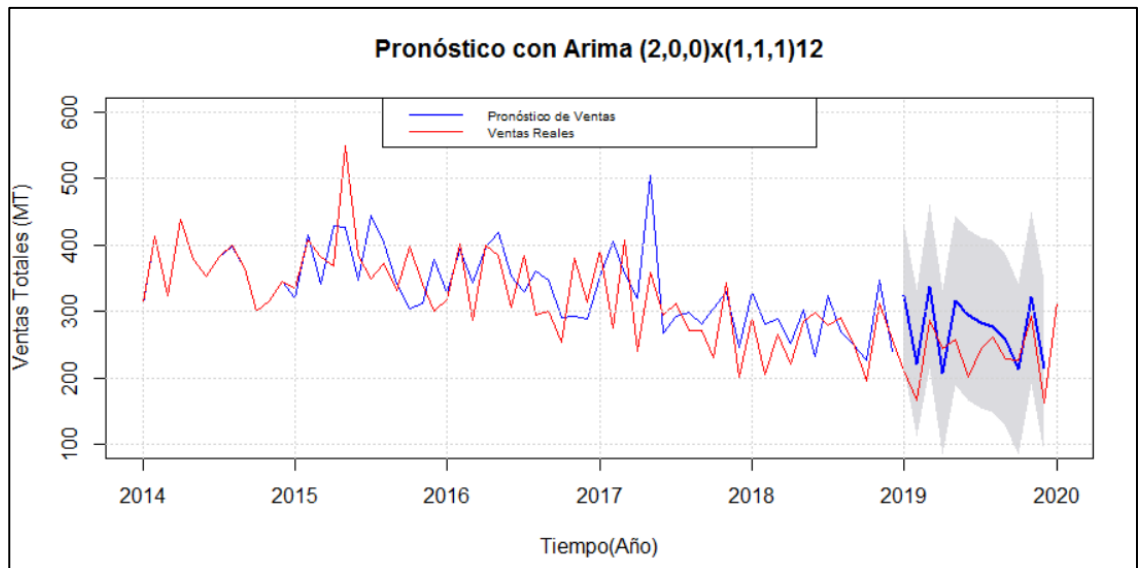
Mes	Pronóstico (MT)	Límite Inferior (MT)	Límite Superior (MT)
Enero	323,72	214,39	433,06
Febrero	221,34	111,22	331,46
Marzo	338,12	215,04	461,19
Abril	207,95	84,19	331,72
Mayo	316,5	189,52	443,48
Junio	294,72	167,36	422,08
Julio	283,49	155,25	411,74
Agosto	278	149,59	406,41
Septiembre	256,93	128,26	385,6
Octubre	212,88	84,14	341,62
Noviembre	321,05	192,23	449,86
Diciembre	215,87	87,02	344,71

Elaborado por: Autor

A continuación, también detallamos la estimación de ventas anual considerando la suma de los pronósticos mensuales para el año 2019, expresados en toneladas métricas: 3270,57 MT

Finalmente, con el objetivo de validar nuestro modelo debemos comparar los valores reales de las ventas del año 2019 que fueron separados del conjunto inicial de nuestra serie de tiempo y los datos obtenidos mediante nuestro modelo de mejor ajuste, es por esto que ilustramos ambas series de tiempo en la siguiente figura, en donde podemos observar que el modelo pronosticado *ajuste4* tiene un comportamiento similar a los valores reales de nuestra variable de respuesta; es decir, ventas reales de aditivos expresadas en toneladas métricas:

Figura No.4.23. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de ventas totales de aditivos. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

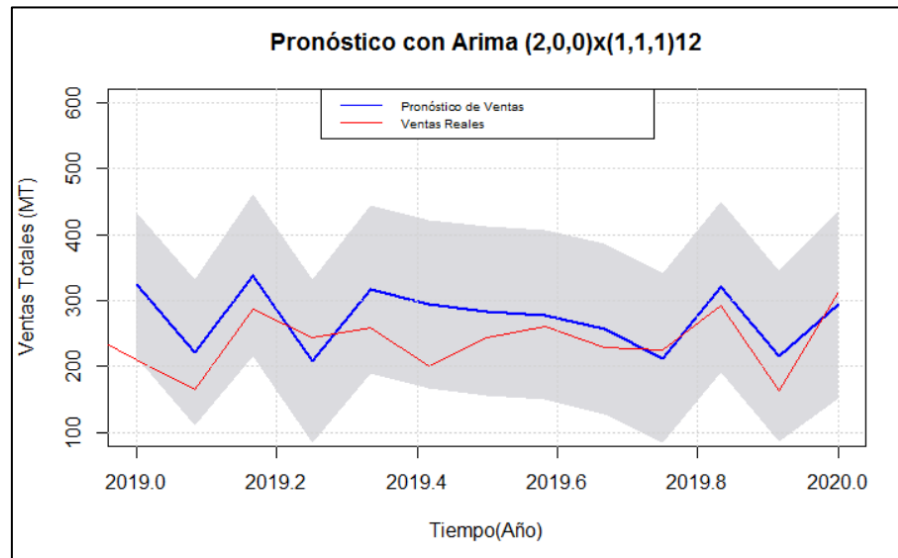
Así también, procedemos a comparar los resultados reales y pronosticados de las ventas de aditivos de la empresa comercializadora durante el año 2019 expresadas en toneladas métricas (MT):

Tabla No. 4.11. Comparación de ventas pronosticadas y reales durante el año 2019. Serie temporal ventas totales de aditivos.

Ventas Reales (MT)	Ventas Pronosticadas (MT)	Diferencia (MT)	Diferencia Porcentual
2783,03	3270,57	487,54	17.52%

Elaborado por: Autor

Figura No.4.24. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de ventas totales de aditivos. Período: 2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

4.4. MODELO: FUERA DE BORDA

4.4.1. ANÁLISIS EXPLORATORIO DE DATOS

En esta sección se llevará a cabo la presentación de la serie mensual de las ventas totales del aditivo "Fuera de Borda" para la fabricación de lubricantes de motores enfriados por agua expresadas en kilogramos de una empresa importadora y comercializadora de materias primas, correspondientes a los periodos 2014 a 2019, obtenidos a través de los registros de ventas de la empresa importadora y comercializadora de materias primas.

Tabla No. 4.12. Serie histórica mensual de las ventas del aditivo para la fabricación de lubricantes "Fuera de Borda" expresadas en kilogramos, periodo 2014-2019.

Ventas Mensuales de Aditivo "Fuera de Borda" (kg)						
Mes/Año	2014	2015	2016	2017	2018	2019
Enero	16257	17952	19244	17729	24872	11397
Febrero	15183	20205	21414	16496	17765	25128
Marzo	22113	22732	20663	25820	31694	21433
Abril	11491	25818	22488	27806	28985	27422
Mayo	16492	63732	14992	25533	24409	26918
Junio	19644	17167	15974	19966	16830	18796
Julio	20074	22453	30423	22862	25241	25806
Agosto	11285	19927	19709	25996	23505	30500
Septiembre	22728	21053	13464	23740	23984	34848
Octubre	19309	18066	17370	15596	22955	11405
Noviembre	24199	21293	21049	27376	22253	33184
Diciembre	21885	20544	15227	15622	25432	20362

Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Al igual que con las ventas mensuales totales (expresadas en kilogramos) detalladas en el numeral anterior se llevará a cabo el análisis exploratorio de los datos, utilizando el Software libre R, transformándolos a series temporales y los representaremos gráficamente con el fin de verificar la posible presencia principalmente de tendencias o estacionalidades que den evidencia de la no-estacionariedad de la serie.

Para esto importamos nuestros datos a R, aplicando las mismas librerías que con la serie de ventas totales mensuales del numeral anterior.

Puesto que, los datos son recibidos en kilogramos procedemos con la transformación de estos datos a toneladas métricas dividiendo la serie de tiempo para 1000, considerando la equivalencia: 1000 kg =1 MT.

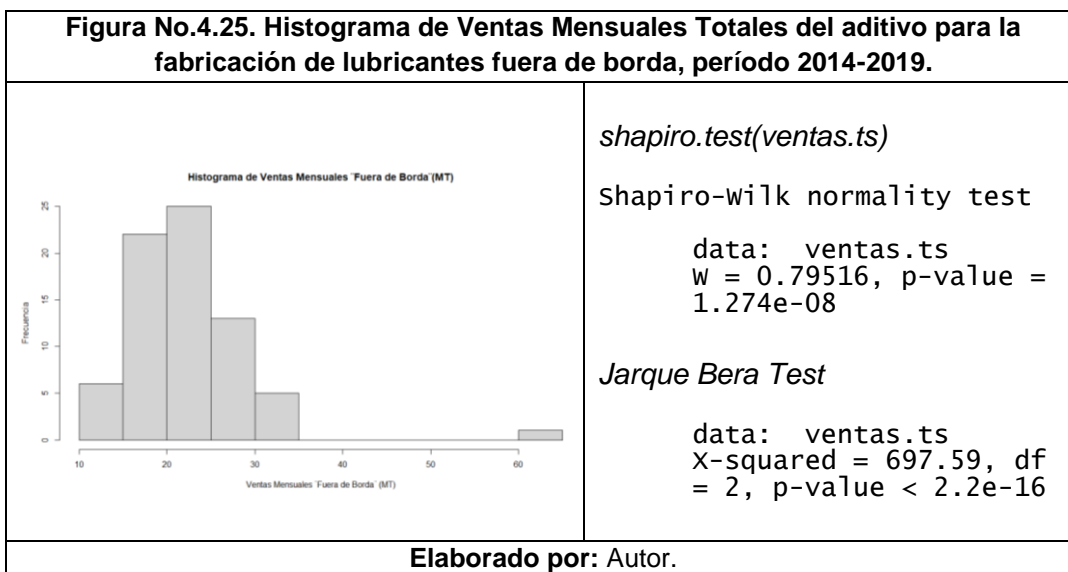
Tabla No. 4.13. Serie histórica mensual de las ventas del aditivo para la fabricación de lubricantes "Fuera de Borda" expresadas en toneladas métricas, periodo 2014-2019.

Ventas Mensuales de Aditivo "Fuera de Borda" (MT)						
Mes/Año	2014	2015	2016	2017	2018	2019
Enero	16	18	19	18	25	11
Febrero	15	20	21	16	18	25
Marzo	22	23	21	26	32	21
Abril	11	26	22	28	29	27
Mayo	16	64	15	26	24	27
Junio	20	17	16	20	17	19
Julio	20	22	30	23	25	26
Agosto	11	20	20	26	24	31
Septiembre	23	21	13	24	24	35
Octubre	19	18	17	16	23	11
Noviembre	24	21	21	27	22	33
Diciembre	22	21	15	16	25	20

Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

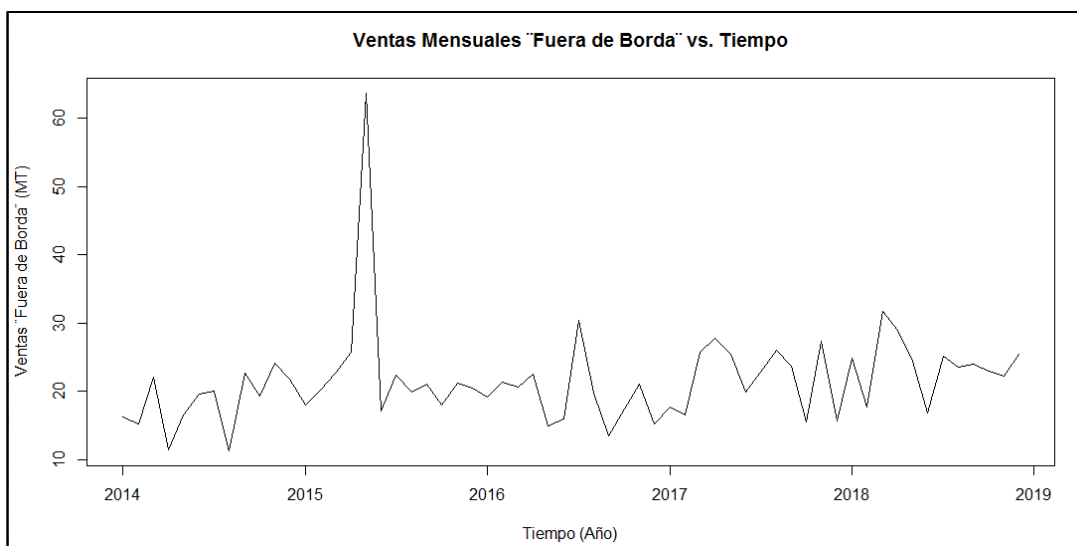
Elaborado por: Autor

A continuación, generamos un histograma de frecuencias con la finalidad de representar la distribución de los datos por medio de barras, en la siguiente figura podemos evidenciar que nuestros datos no tienen una distribución normal y para corroborar esta información llevamos a cabo las pruebas correspondientes de normalidad.



Puesto que este valor p obtenido en ambas pruebas es inferior a 0.05, rechazamos las hipótesis nulas de que la distribución de nuestros datos es normal. Continuando el análisis, procedemos a graficar los datos de nuestra serie respecto al tiempo, la cual nos permite observar posible tendencia positiva ascendente:

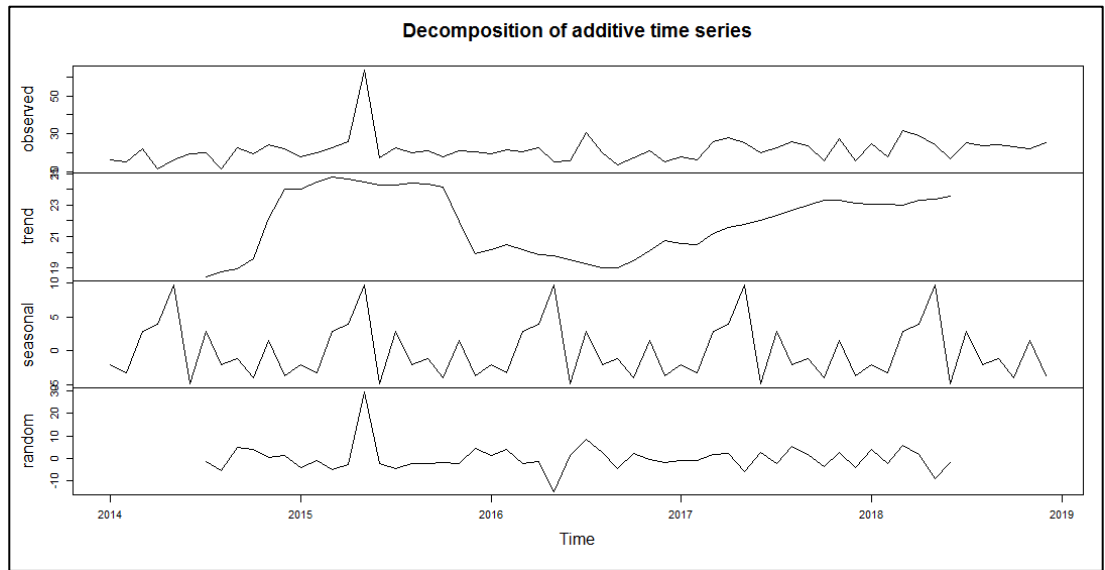
Figura No.4.26. Serie temporal de las ventas totales mensuales del aditivo para la fabricación de lubricantes fuera de borda, expresada en toneladas métricas. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Con el fin de obtener más detalles de su comportamiento, descomponemos la serie temporal en todos sus posibles componentes (tendencia, estacional y aleatorio) a través de la descomposición estacional clásica de medias móviles.

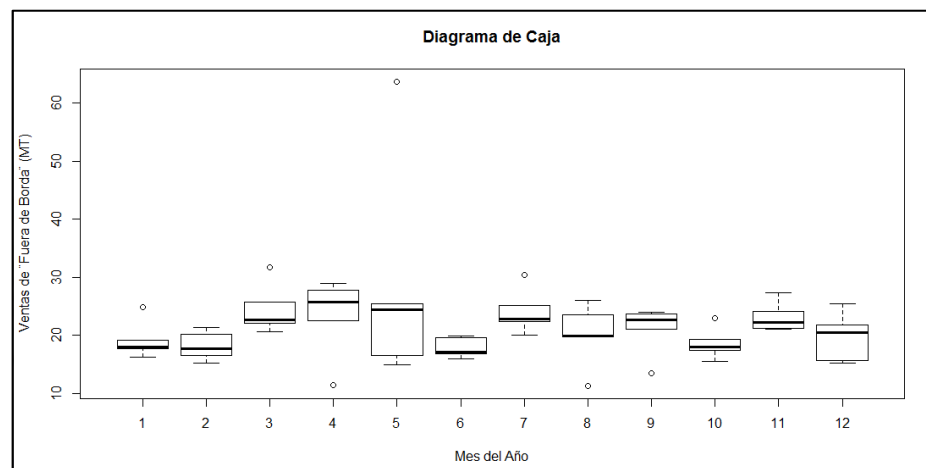
Figura No.4.27. Descomposición de la Serie Temporal de las ventas totales mensuales del aditivo para la fabricación lubricantes fuera de borda. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Como resultado, en la gráfica anterior no es posible evidenciar una tendencia ascendente que se mantenga constante respecto al tiempo. No obstante, si verificamos la del componente estacional y el aleatorio. Considerando que son datos de ventas mensuales y que se presentan cortes anuales, continuamos con la observación del comportamiento de los datos segmentándolos por meses siendo *Enero=1, Febrero=2....Diciembre=12*. En esta gráfica podemos observar que los valores de las medias presentan dispersiones más pequeñas, lo que nos puede dar una idea de que los consumos promedios de los meses de cada año tienen comportamientos similares.

Figura No.4.28. Diagrama de caja de la serie temporal de las ventas totales del aditivo fuera de borda segmentada por mes, en toneladas métricas. Período: 2014-2019.

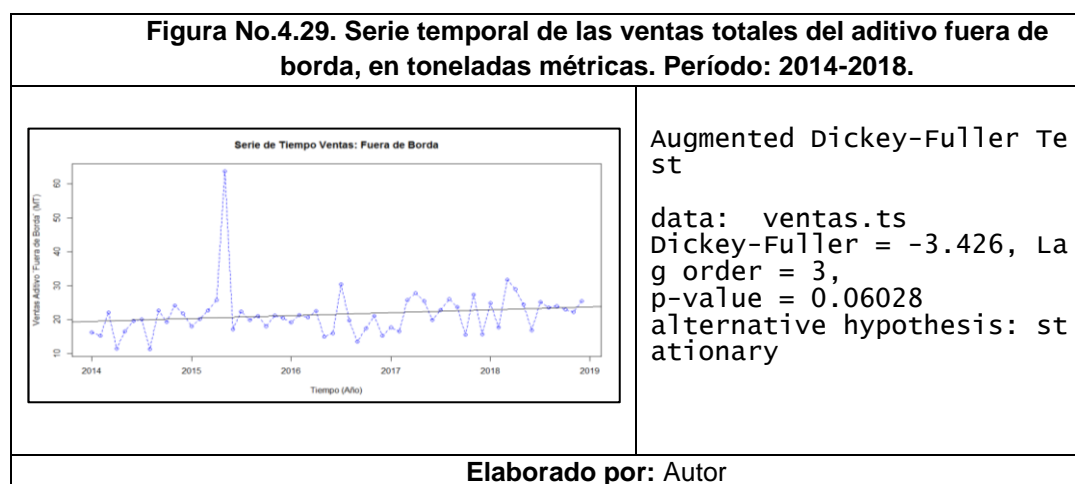


Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

4.4.2. ANÁLISIS DE TENDENCIA

Una vez que concluimos el análisis exploratorio de nuestros datos, procedemos a aplicar la metodología Box-Jenkins; para esto es necesario que trabajemos únicamente con los datos de los primeros 5 años (2014 a 2018) de las ventas del aditivo "Fuera de Borda" para generar el modelo de pronóstico y utilicemos los datos correspondientes al año 2019 para comprobar qué tan cerca se encuentra nuestro modelo de la realidad.

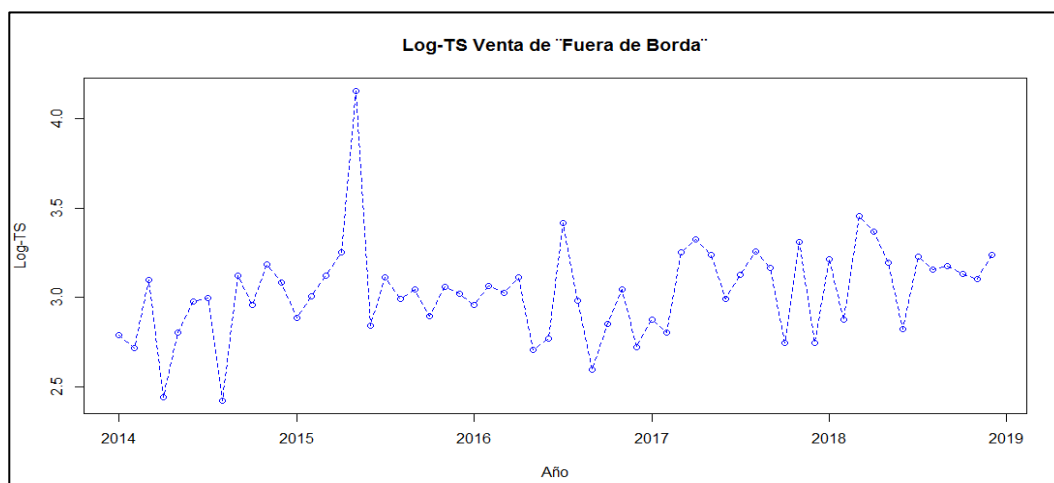
Por lo anteriormente expuesto, procedemos a graficar únicamente el periodo 2014-2018 para verificar si estos se comportan de forma estacionaria:



Adicionalmente a lo observado en la gráfica anteriormente expuesta, a través de la prueba Dickey-Fuller aumentada el valor p mostrado es superior a 0.05, es decir que aceptamos la hipótesis nula de que nuestra serie temporal tiene raíz unitaria y es estacionaria respecto a su media; adicionalmente, como habíamos observado en la gráfica de descomposición de nuestra serie de tiempo existe una tendencia de datos levemente ascendente, contiene también un componente estacional y sus fluctuaciones en el tiempo no son uniformes; es decir, tiene variaciones irregulares.

Por lo expresado en el párrafo anterior y dado que debemos eliminar las variaciones irregulares de nuestra serie de datos, aplicamos logaritmos a nuestra serie y los graficamos para verificar los cambios en su comportamiento.

Figura No.4.30. Logaritmo de la serie temporal de las ventas totales mensuales del aditivo fuera de borda. Período: 2014-2018.

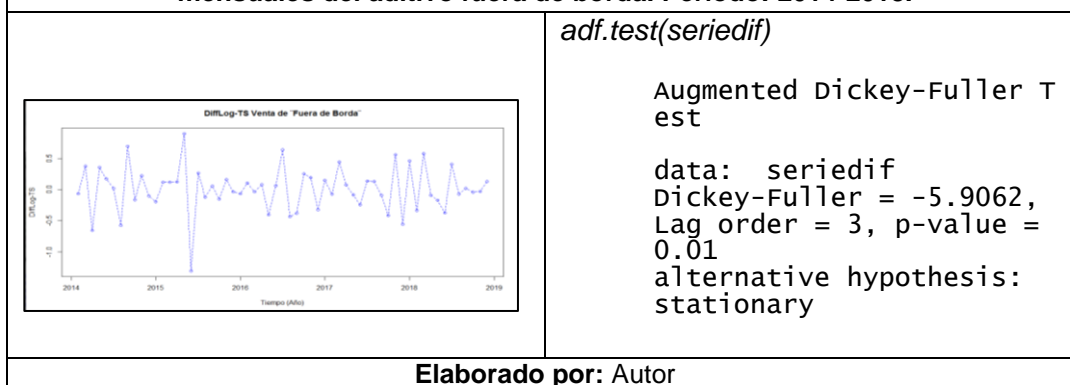


Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Una vez que hemos aplicado los logaritmos a nuestra serie, creando una nueva serie llamada *serielog*, podemos ver que el rango de diferencia entre los valores mensuales es muy inferior a la serie inicial *ventas.ts*; y de esta manera conseguimos que esta nueva serie tenga una variabilidad regular.

De acuerdo a lo expresado por Box-Jenkins, con la finalidad de aplicar esta metodología de modelos autorregresivos, precisamos conseguir una serie de datos al menos débilmente estacionaria en términos media y para lograr esto necesitaríamos también remover la tendencia anteriormente observada; es por esto que, aplicamos una primera diferencia a nuestra serie, a pesar de que al utilizar la prueba Phillip Perron nos haya indicado que no debíamos aplicar diferencias, dado que no era posible alcanzar un comportamiento estacionario en caso contrario:

Figura No.4.31. Diferencia del logaritmo de la serie temporal de ventas totales mensuales del aditivo fuera de borda. Período: 2014-2018.

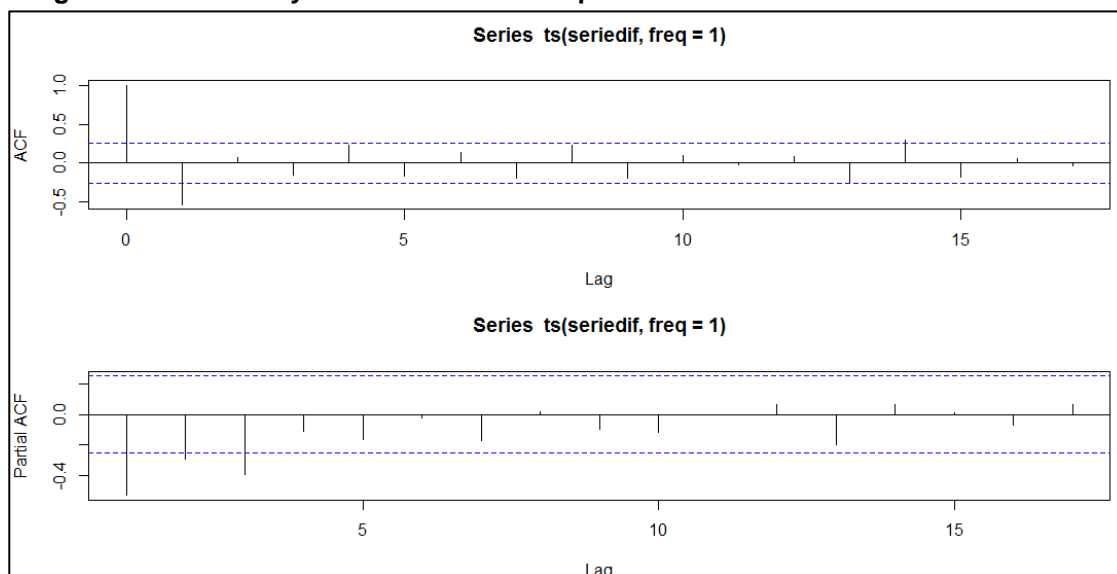


Como fue efectuado en el numeral anterior es recomendable también llevar a cabo un test que compruebe que lo efectuado es estadísticamente adecuado, por lo que procedemos a utilizar la Prueba Aumentada de Dickey-Fuller. Al ser el valor p obtenido menor que 0.05 rechazamos la hipótesis nula de que la serie temporal posee raíz unitaria, lo cual significaría que nuestra serie es estacionaria.

4.4.3. AJUSTE DEL MODELO

Tomando en cuenta los principios utilizados en el primer modelo planteado, procederemos con la visualización de las funciones de autocorrelación típicamente utilizadas dentro de la metodología Box-Jenkins: Función de Autocorrelación y Función de Autocorrelación Parcial.

Figura No.4.32. ACF y PACF de la serie temporal *seriedif*. Aditivo Fuera de Borda.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Considerando lo expuesto en la figura anterior, en el gráfico de PACF (Función de autocorrelación parcial) observamos autocorrelaciones nulas a partir del tercer rezago (lag); en base a esto es posible plantear un modelo AR de orden uno es decir el valor p será igual a 1 y un segundo modelo con p igual a 3; puesto que, el gráfico PACF nos indica el número de componentes que tendrá la parte AR de nuestro modelo. Por otra parte, comportamiento de observado en el gráfico ACF (Función de Autocorrelación) nos muestra detalles de la parte MA (Media Móvil), teniendo en cuenta que en este se evidencia un comportamiento de decaimiento exponencial muy probablemente tendrá un solo componente. En base a estos criterios procedemos a plantear los diferentes modelos del presente estudio.

4.4.3.1. Modelos ARIMA Estacionales $(p,d,q) \times (P,D,Q)_s$

Dentro de los criterios considerados para llevar a cabo el presente modelo son las características del componente de medias móviles en donde observamos nuevamente en la gráfica ACF un decaimiento exponencial, lo cual es indicativo de que este esté conformado por una sola unidad y por lo tanto Q sea igual a 1. Por otra parte, a pesar de que fueron probados modelos iniciales $AR(1)$ y $AR(3)$ los

resultados obtenidos en estos coeficientes daban evidencia de que no resultaban significativos para representar a la variable de respuesta. Por lo anteriormente expuesto, se plantea un modelo de características ARIMA estacionales, considerando únicamente las diferencias aplicadas a los datos siendo D igual a 1, y el componente de medias móviles Q equivalente a 1. Con el objetivo de plantear un modelo arima estacional o arima estacional multiplicativo para nuestro aditivo "Fuera de borda", agregaremos a nuestro sistema de estudio las características estacionales de los datos de ventas de aditivos de la empresa comercializadora. Es decir, para este caso particular, al ser datos anuales, el periodo s será igual a 12; de tal forma planteamos el siguiente modelo:

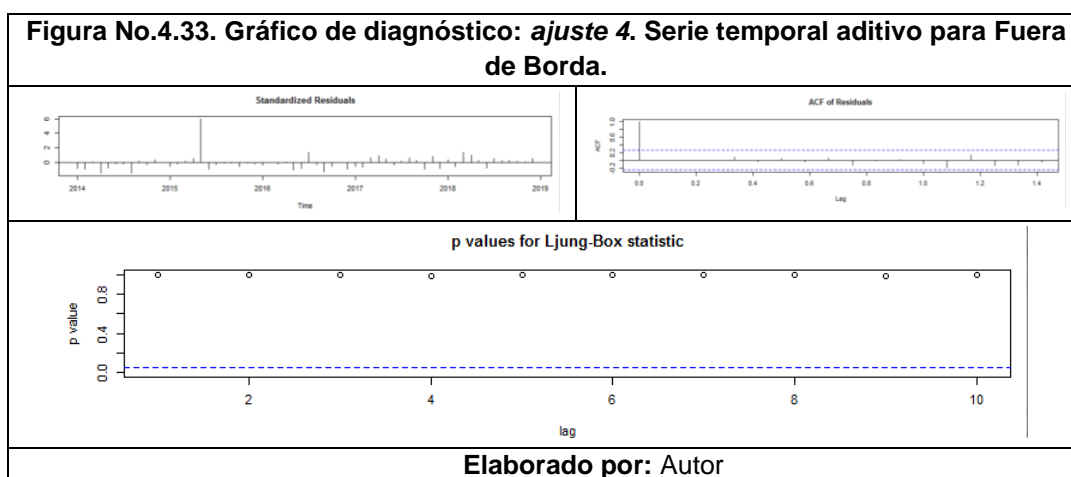
Tabla No. 4.14. Resumen Estadístico - Orden (0,0,0)x(0,1,1)₁₂, Serie Temporal Aditivo Fuera de Borda						
Orden: (0,0,0)x(0,1,1) ₁₂						
Coefficients:						
	sma1					
	-1.0000					
s.e.	0.2932					
sigma^2 estimated as 48.89: log likelihood = -171.12,						
aic = 346.23						
Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	
MASE						
Training set	1.283843	6.254196	3.523653	1.715159	14.97263	0.58
	87441					
	ACF1					
Training set	0.06580216					
Elaborado por: Autor						

Tabla No. 4.15. Estadístico Z - Orden (0,0,0)x(0,1,1)₁₂, Serie Temporal Aditivo Fuera de Borda						
z test of coefficients:						
	Estimate	Std. Error	z value	Pr(> z)		
sma1	-1.00000	0.29324	-3.4101	0.0006493 ***		

Signif. codes:	0	'***'	0.001	'**'	0.01	'*' 0.05 '.' 0.1 ' ' 1
Elaborado por: Autor						

Posteriormente, llevamos a cabo el análisis del comportamiento de nuestro modelo respecto a la prueba de ruido blanco (Ljung-Box) es el de valor más bajo en comparación con todos los modelos planteados AR(1), AR(3) y ARIMA(0,1,1); a pesar de ser el que cuenta con más características ingresadas al modelo a partir de los diferentes coeficientes.

Tabla No. 4.16. Prueba Box-Ljung- Orden $(0,0,0) \times (0,1,1)_{12}$, Serie Temporal Aditivo Fuera de Borda
Box-Ljung test data: residuals(ajuste4) x-squared = 0.27301, df = 1, p-value = 0.6013
Elaborado por: Autor



4.4.4. ELECCIÓN DEL MEJOR MODELO

Con el objetivo de comparar los modelos planteados realizaremos un análisis comparativo, utilizando los resultados de los Errores Medios de Porcentaje Absoluto (MAPE), los criterios de información de Akaike (AIC) y criterios de información bayesiano (BIC) según procede a continuación. Una vez finalizadas estas estimaciones generamos el cuadro comparativo de los modelos resultantes del presente estudio:

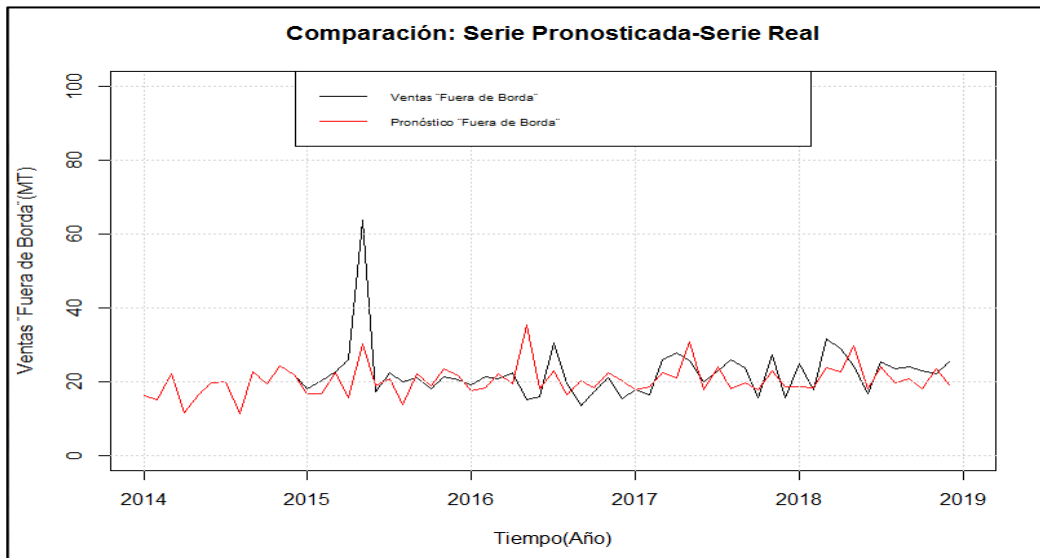
Tabla No. 4.17. Comparación de modelos resultantes. Serie temporal Aditivo Fuera de Borda.

No.	MODELO	AIC	BIC	MAPE
1	ARIMA (1,0,0)	410,05	416,34	20,33
2	ARIMA (3,0,0)	413,78	424,26	20,03
3	ARIMA (0,1,1)	406,42	410,57	18,62
4	ARIMA $(0,0,0) \times (1,1,1)_{12}$	346,23	349,98	14,97

Elaborado por: Autor

De acuerdo con los criterios AIC y BIC, se selecciona el modelo ARIMA $(0,0,0) \times (0,1,1)_{12}$, ya que presenta los menores valores en estos criterios. De forma similar, utilizando el criterio MAPE el modelo 4 es también el más idóneo puesto que presenta el menor valor comparado con los demás modelos, lo cual indica que es el modelo que menos se desvía de los datos originales, por lo cual procedemos a graficar el modelo elegido en la siguiente figura:

Figura No.4.34. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de las ventas del aditivo fuera de borda. Período: 2014-2018.

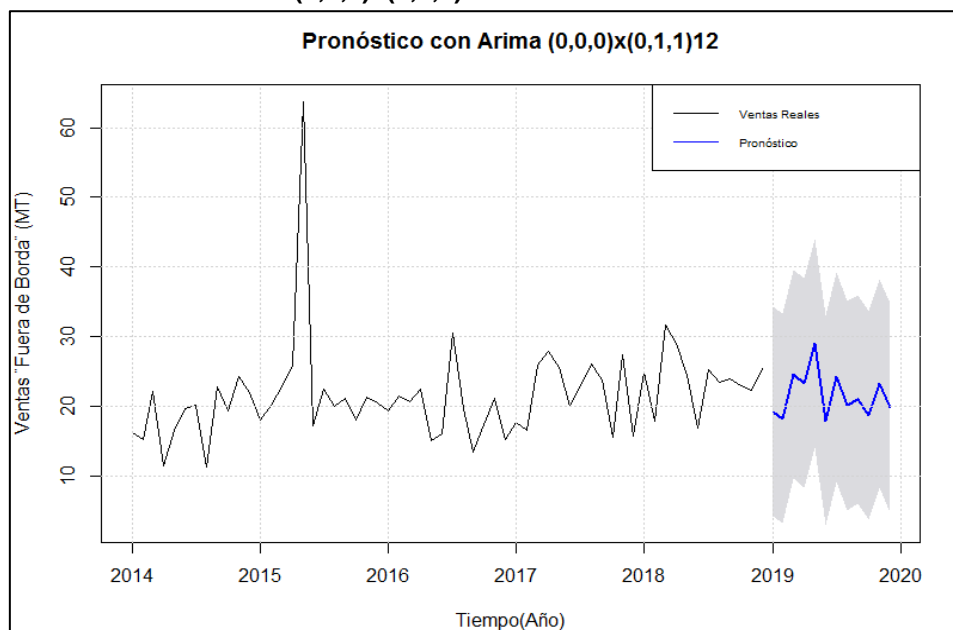


Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

4.4.5. PREDICCIÓN UTILIZANDO ARIMA

Finalizada la elección del modelo más adecuado a nuestro conjunto de datos utilizaremos la función *forecast()* en R, con el fin de predecir o pronosticar los posibles valores futuros de nuestra serie temporal, determinando también intervalos de confianza:

Figura No.4.35. Pronóstico de ventas totales del aditivo fuera de borda a través de Arima (0,0,0)x(0,1,1)12. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Adicionalmente, tabulamos los datos obtenidos dentro del presente pronóstico:

Tabla No. 4.18. Pronósticos resultantes del ajuste 4 (Periodo 2019). Serie temporal fuera de borda

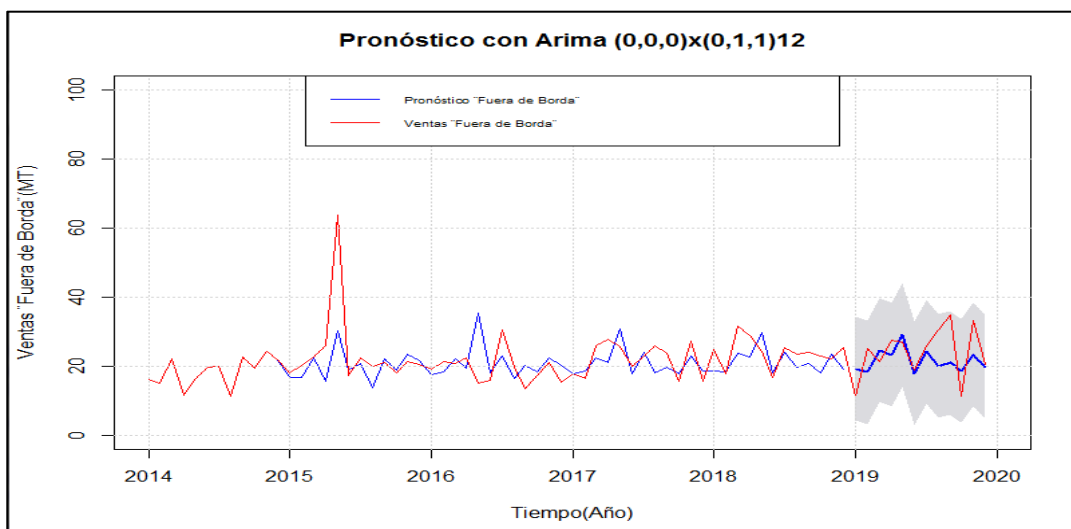
Mes	Pronóstico (MT)	Límite Inferior (MT)	Límite Superior (MT)
Enero	19,21	4,2	34,22
Febrero	18,21	3,2	33,23
Marzo	24,6	9,59	39,62
Abril	23,32	8,31	38,33
Mayo	29,03	14,02	44,04
Junio	17,92	2,9	32,93
Julio	24,21	9,2	39,22
Agosto	20,08	5,07	35,1
Septiembre	20,99	5,98	36,01
Octubre	18,66	3,65	33,67
Noviembre	23,23	8,22	38,25
Diciembre	19,74	4,73	34,76

Elaborado por: Autor

A continuación, también detallamos la estimación de ventas anual del aditivo "Fuera de Borda" considerando la suma de los pronósticos mensuales, expresados en toneladas métricas: 259.22

Finalmente, con el objetivo de validar nuestro modelo debemos comparar los valores reales de las ventas del aditivo "Fuera de Borda" del año 2019 que fueron separados del conjunto inicial de nuestra serie de tiempo y los datos obtenidos mediante nuestro modelo de mejor ajuste, es por esto que ilustramos ambas series de tiempo en la siguiente figura, en donde podemos observar que el modelo pronosticado *ajuste4* tiene un comportamiento similar a los valores reales de nuestra variable de respuesta; es decir, ventas reales de aditivos expresadas en toneladas métricas:

Figura No.4.36. Comparación de la serie temporal pronosticada y la serie temporal de datos reales del aditivo fuera de borda. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
laborado por: Autor

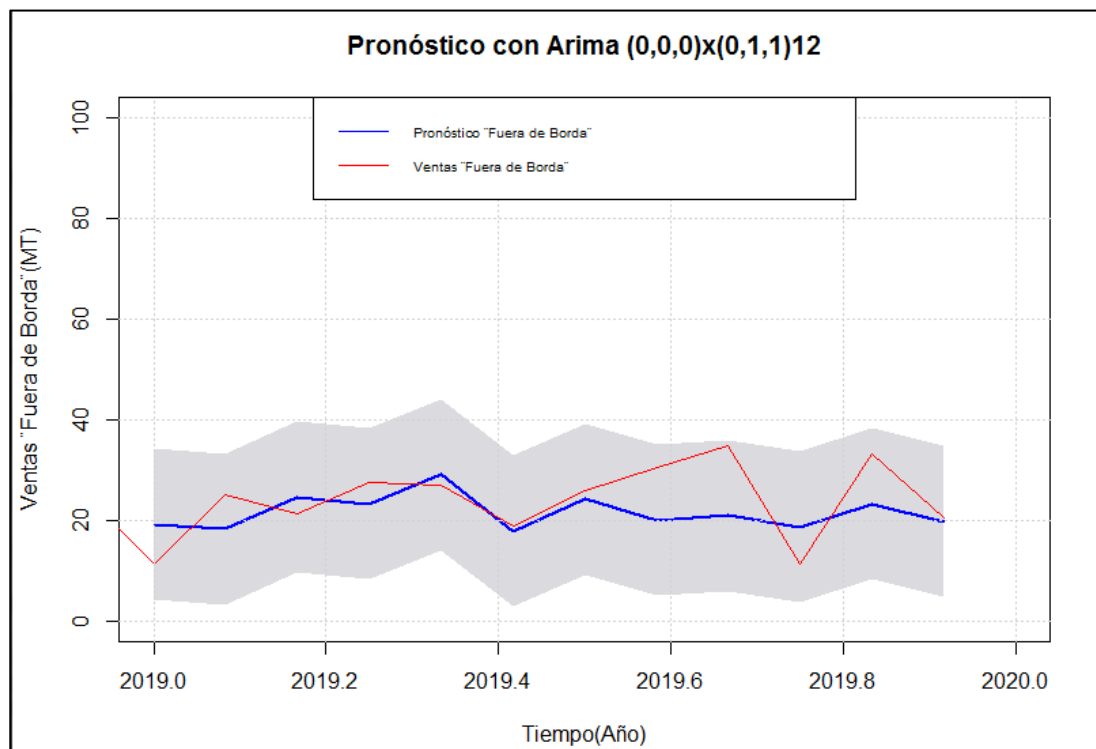
Así también, procedemos a comparar los resultados reales y pronosticados de las ventas del aditivo "Fuera de Borda" durante el año 2019 expresadas en toneladas métricas (MT): 287.20

Tabla No. 4.19. Comparación de ventas pronosticadas y reales durante el año 2019. Serie temporal aditivo fuera de borda.

Ventas Reales (MT)	Ventas Pronosticadas (MT)	Diferencia (MT)	Diferencia Porcentual
287,20	259.22	-27,98	-9,74%

Elaborado por: Autor

Figura No.4.37. Comparación de la serie temporal pronosticada y la serie temporal de datos reales del aditivo fuera de borda. Período: 2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

4.5. MODELO: GASOLINA

4.5.1. ANÁLISIS EXPLORATORIO DE DATOS

A continuación procederemos con el análisis de datos de los aditivos relacionados con el segmento de Gasolina a través de la sumatoria de los aditivos del segmento gasolina (Gasolina 1, Gasolina 2 y Gasolina 3) con el fin de formar la serie de los aditivos "Gasolina" para la fabricación de lubricantes de motores de automóviles alimentados por este tipo de combustibles, expresadas en kilogramos, correspondientes a los periodos 2014 a 2019, obtenidos a través de los registros de ventas de la empresa importadora y comercializadora de materias primas.

Tabla No. 4.20. Serie histórica mensual de las ventas de los aditivos para la fabricación de lubricantes para motores alimentados por "Gasolina", expresadas en kilogramos, periodo 2014-2019.

Ventas Mensuales de Aditivos "Gasolina" (kg)						
Mes/Año	2014	2015	2016	2017	2018	2019
Enero	25728	27738	32160	32091	34188	13445
Febrero	11256	25929	30753	14431	27481	23005
Marzo	20502	23718	22512	29920	16406	31194
Abril	30753	19698	30954	24671	21459	32387
Mayo	29346	41004	32160	33311	24169	27975
Junio	12462	17688	55074	30886	40412	25048
Julio	15276	16080	27336	22458	28565	29249
Agosto	27336	34371	35577	45937	37417	24605
Septiembre	23919	32964	36783	39876	21446	15198
Octubre	25326	29949	23316	26882	11841	18523
Noviembre	14673	28944	25326	32272	24376	26366
Diciembre	14070	21909	28341	18613	13627	11772

Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Tabla No. 4.21. Serie histórica mensual de las ventas de los aditivos para la fabricación de lubricantes para motores alimentados por "Gasolina" expresadas en toneladas métricas, periodo 2014-2019.

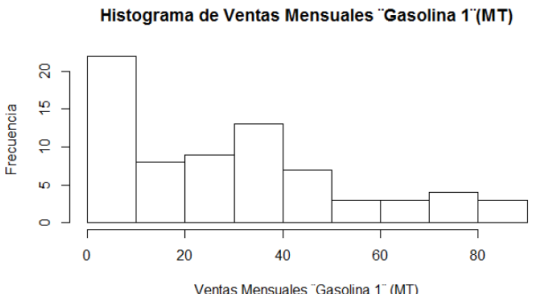
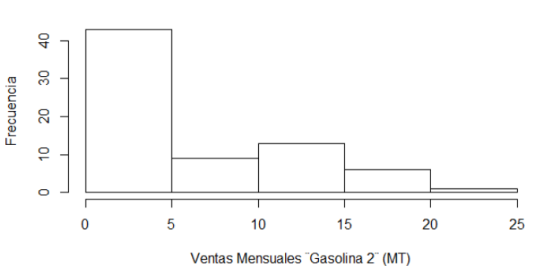
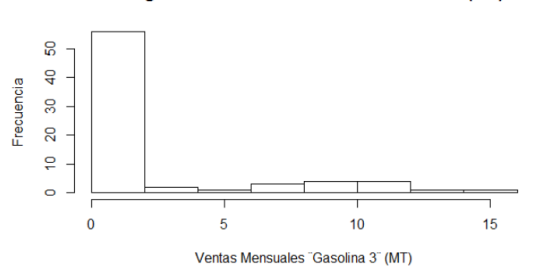
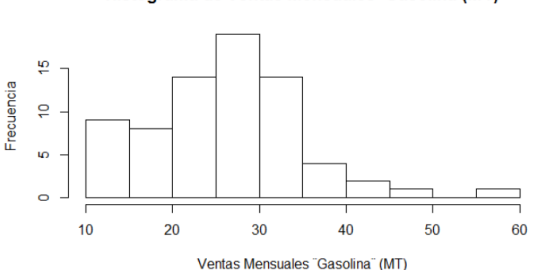
Ventas Mensuales de Aditivo "Gasolina" (MT)						
Mes/Año	2014	2015	2016	2017	2018	2019
Enero	26	28	32	32	34	13
Febrero	11	26	31	14	27	23
Marzo	21	24	23	30	16	31
Abril	31	20	31	25	21	32
Mayo	29	41	32	33	24	28
Junio	12	18	55	31	40	25
Julio	15	16	27	22	29	29
Agosto	27	34	36	46	37	25
Septiembre	24	33	37	40	21	15
Octubre	25	30	23	27	12	19
Noviembre	15	29	25	32	24	26
Diciembre	14	22	28	19	14	12

Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Procedemos entonces a transformar los datos a series temporales y a graficarlos con el fin de verificar sus comportamientos, importando los datos a R y activando las librerías a ser utilizadas.

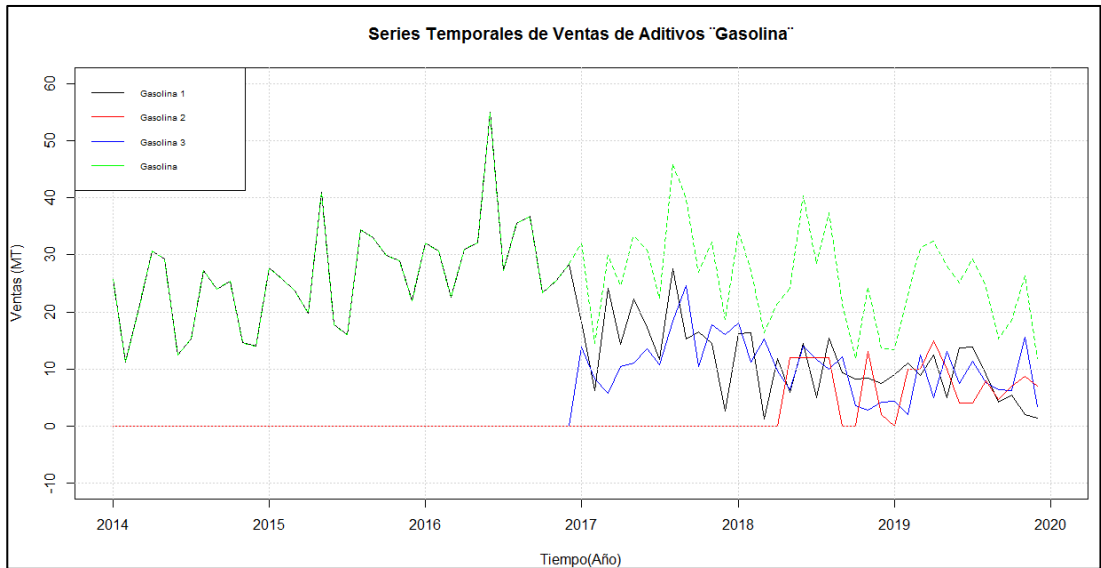
Considerando que la serie gasolina está compuesta por los datos de 3 aditivos diferentes, analizaremos inicialmente el comportamiento de cada uno de forma individual: "Gasolina 1", "Gasolina 2", "Gasolina 3" y su sumatoria.

Así también procedemos con la generación de los histogramas de frecuencias de cada una de las series ingresadas. Posteriormente llevamos a cabo pruebas de normalidad para cada serie a través de las pruebas de normalidad Shapiro-Wilk y Jarque Bera.

Figura No.4.38. Histograma de Ventas Mensuales Totales de los aditivos: Gasolina 1, Gasolina 2, Gasolina 3 y Gasolina, período 2014-2019.	
<p>Histograma de Ventas Mensuales "Gasolina 1"(MT)</p> 	<p><i>Shapiro-Wilk normality test</i></p> <p>data: ventas.tsg1 $w = 0.96537$, p-value = 0.04474</p> <p><i>Jarque Bera Test</i></p> <p>data: ventas.tsg1 X-squared = 3.9907, df = 2, p-value = 0.136</p>
<p>Histograma de Ventas Mensuales "Gasolina 2"(MT)</p> 	<p><i>Shapiro-Wilk normality test</i></p> <p>data: ventas.tsg2 $w = 0.80008$, p-value = 1.724e-08</p> <p><i>Jarque Bera Test</i></p> <p>data: ventas.tsg2 X-squared = 10.231, df = 2, p-value = 0.006002</p>
<p>Histograma de Ventas Mensuales "Gasolina 3"(MT)</p> 	<p><i>Shapiro-Wilk normality test</i></p> <p>data: ventas.tsg3 $w = 0.55888$, p-value = 2.231e-13</p> <p><i>Jarque Bera Test</i></p> <p>data: ventas.tsg3 X-squared = 43.366, df = 2, p-value = 3.829e-10</p>
<p>Histograma de Ventas Mensuales "Gasolina"(MT)</p> 	<p><i>Shapiro-Wilk normality test</i></p> <p>data: ventas.tsg $w = 0.96922$, p-value = 0.07471</p> <p><i>Jarque Bera Test</i></p> <p>data: ventas.tsg X-squared = 4.2395, df = 2, p-value = 0.1201</p>
Elaborado por: Autor.	

En base a los histogramas y a las pruebas de normalidad ejecutados para cada una de las series podemos observar que las series “gasolina 1” y la de la sumatorias de las ventas de los aditivos para gasolina *ventas.tsg* tienen una distribución normal. Para estos casos específicos los valores p obtenidos en estas pruebas para las series mencionadas son superiores a 0.05, por lo que decimos que aceptamos la hipótesis nula de que los datos siguen una distribución de los datos es normal. Procedemos entonces a graficar los datos de las series motivo de estudio respecto al tiempo.

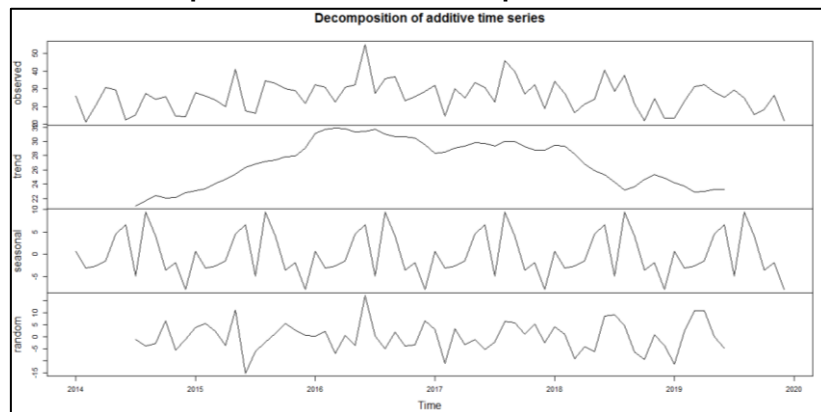
Figura No.4.39. Series temporales de las ventas de los aditivos para la fabricación de lubricantes alimentados por gasolina, expresadas en toneladas métricas. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Para efectos del análisis del comportamiento del segmento gasolina, analizaremos únicamente la serie *ventas.tsg* y continuamos con su descomposición respectiva:

Figura No.4.40. Descomposición de la Serie Temporal Gasolina. Período: 2014-2019.

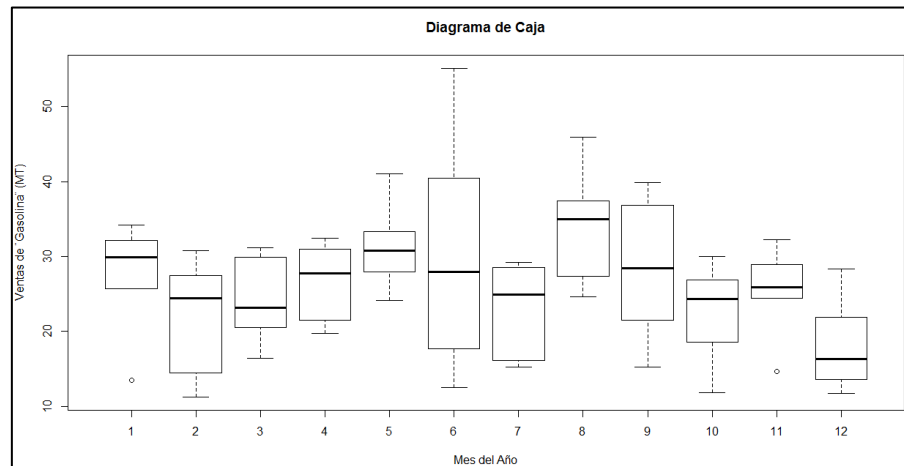


Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

En la gráfica anterior evidenciamos que no existe una tendencia que permanezca constante respecto al tiempo, adicionalmente observamos los componentes estacionales y aleatorios.

Con el fin de llevar a cabo un estudio más detallado de los datos, podemos encontrar la siguiente segmentación mensual a través de un diagrama de caja. En donde observamos que el mes de variabilidad más amplia es el mes de junio (6). Por otra parte, vemos también que el valor más alto en términos de media se encuentra en el mes de agosto (8) y se observa una tendencia descendente a partir de este mes hasta el mes de enero (12), comportamiento que al parecer se ha repetido durante los años 2014 a 2019. En esta gráfica podemos evidenciar también que durante el mes de noviembre el valor promedio aumenta por la demanda presentada por las empresas productoras de lubricantes debido a que durante el mes de diciembre efectúan actividades de mezcla de aceites hasta los primeros 15 días del mes de (12) de cada año:

Figura No.4.41. Diagrama de caja de la serie temporal gasolina, segmentada por mes, en toneladas métricas. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

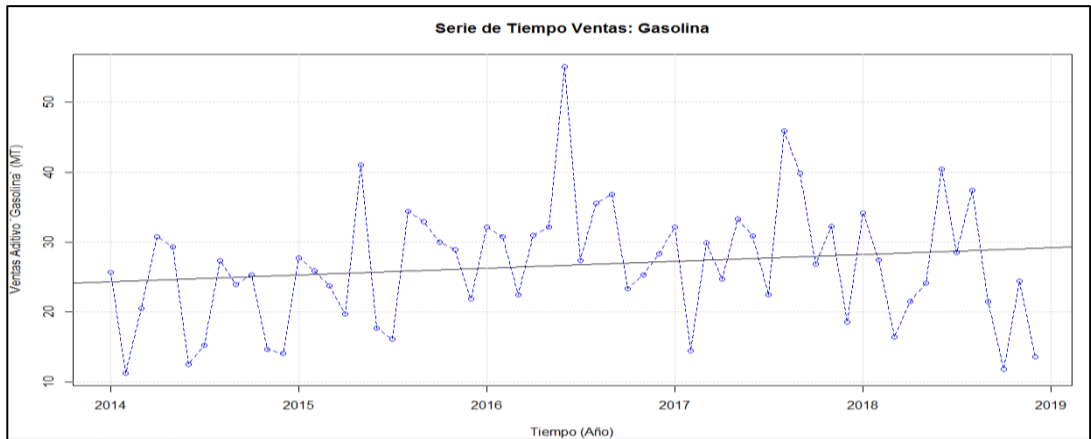
Elaborado por: Autor

4.5.2. ANÁLISIS DE TENDENCIA

Concluido el AED (Análisis Exploratorio de Datos), trabajaremos únicamente con los datos de los primeros 5 años (2014 a 2018) de las ventas de los aditivos "Gasolina", los datos correspondientes al año 2019 se utilizarán para verificar qué tan cercano se encuentra nuestro pronóstico de los datos reales.

Graficaremos únicamente el periodo 2014-2018 para verificar si estos se comportan de forma estacionaria y a través de la aplicación de la prueba Phillips-Perron, corroboramos que debemos aplicar 1 diferencia a nuestra serie para alcanzar la estacionariedad. Es decir, la serie de tiempo no es estacionaria respecto a su media y sus fluctuaciones en el tiempo no son uniformes, puesto que tiene variaciones irregulares.

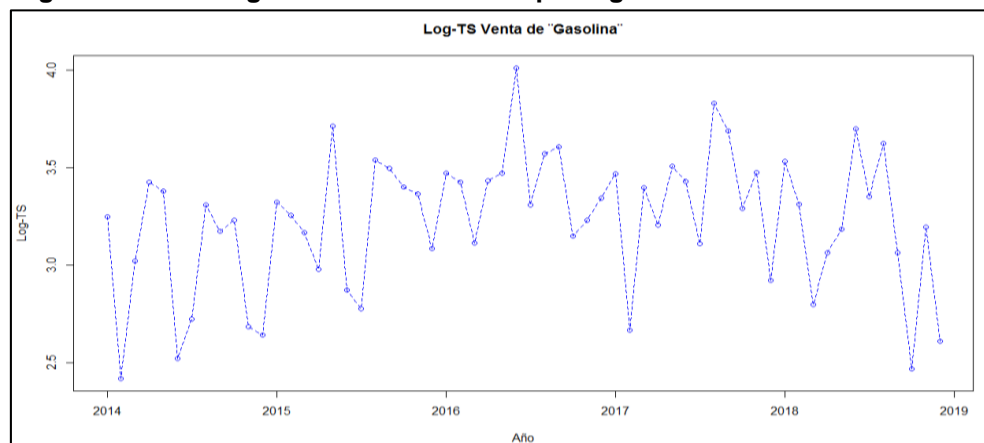
Figura No.4.42. Serie temporal gasolina, en toneladas métricas. Período: 2014-2018.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Con el fin de eliminar las variaciones irregulares de nuestra serie de datos, aplicamos logaritmos:

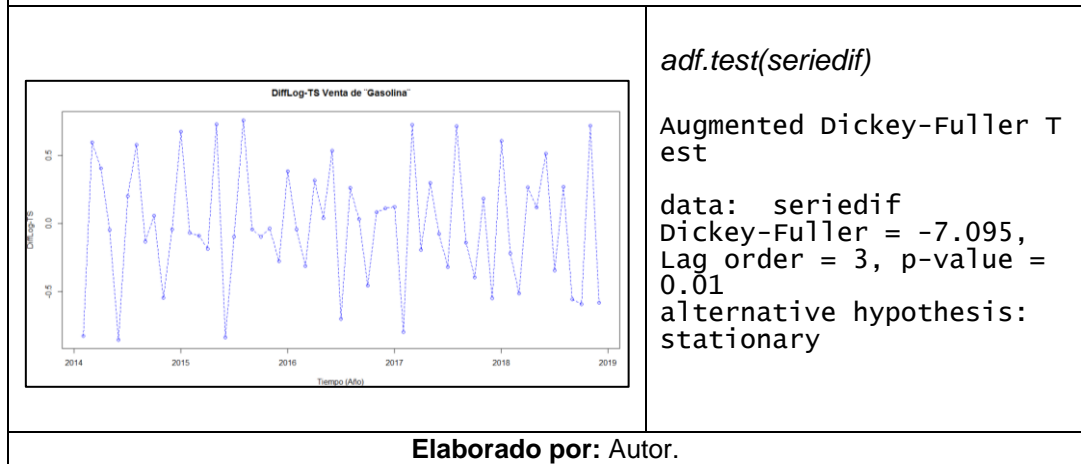
Figura No.4.43. Logaritmo de la serie temporal gasolina. Período: 2014-2018.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

De esta manera logramos que el rango de diferencia entre los valores mensuales sea muy inferior a la serie inicial *ventas.ts*; consiguiendo que esta nueva serie tenga una variabilidad regular. Posteriormente, y dado que es necesario trabajar con una serie de datos al menos débilmente estacionaria, aplicamos una primera diferencia a nuestra serie:

Figura No.4.44. Diferencia del logaritmo de la serie temporal gasolina. Período: 2014-2018.

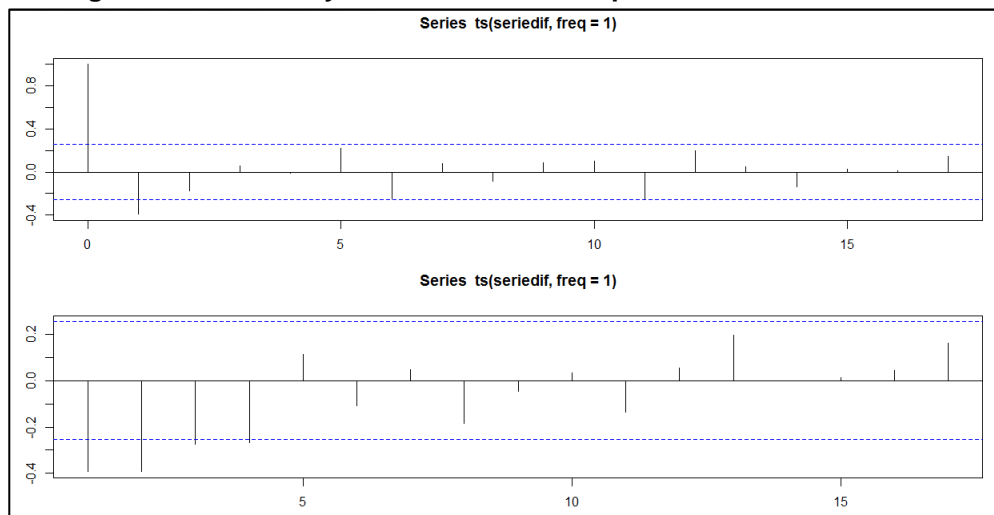


Al ser el valor p obtenido menor que 0.05 rechazamos la hipótesis nula de que la serie temporal posee raíz unitaria, lo cual significaría que nuestra serie es estacionaria.

4.5.3. AJUSTE DEL MODELO

A continuación, analizaremos las funciones de autocorrelación típicamente utilizadas dentro de la metodología Box-Jenkins: Función de Autocorrelación y Función de Autocorrelación Parcial:

Figura No.4.45. ACF y PACF de la serie temporal *seriedif*. Aditivos Gasolina.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

De acuerdo al gráfico de PACF vemos que existen autocorrelaciones nulas a partir del cuarto rezago (lag); no obstante, plantearemos un modelo AR de orden uno es decir el valor p será igual a 1 y un segundo modelo con p igual a 2, para verificar el nivel de significancia de los coeficientes ar1 y ar2 antes de evaluar rezagos superiores a estos niveles. Por otra parte, comportamiento de observado en el

gráfico ACF nos muestra un comportamiento de decaimiento exponencial muy probablemente tendrá un solo componente. En base a estos criterios procedemos a plantear los diferentes modelos del presente estudio.

4.5.3.1. Modelos ARIMA Estacionales (p,d,q)x(P,D,Q)s

Tomando en cuenta lo anteriormente expuesto y los modelos estudiados, AR(1) y AR(2), en donde se evidenció que sus coeficientes no tenían un nivel de significancia adecuado a través de la aplicación del estadístico z, se plantea un modelo de características ARIMA estacionales, considerando únicamente las diferencias aplicadas a los datos siendo D igual a 1, y el componente de medias móviles Q equivalente a 1. Con el objetivo de plantear un modelo arima estacional o arima estacional multiplicativo para nuestra serie temporal Gasolina, agregaremos a nuestro sistema de estudio las propiedades estacionales de los datos de ventas de aditivos, en otras palabras, al ser datos anuales, el periodo s será igual a 12. Por lo anteriormente expuesto, planteamos a continuación el siguiente modelo:

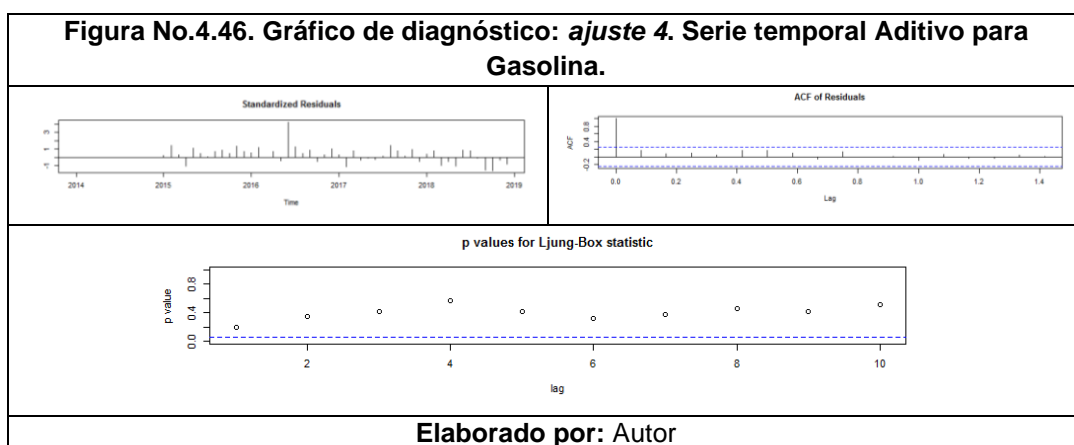
Tabla No. 4.22. Resumen Estadístico - Orden (0,0,0)x(0,1,1)₁₂, Serie Temporal Aditivo para Gasolina						
Orden: (0,0,0)x(0,1,1) ₁₂						
Coefficients:						
	sma1					
	-0.6043					
s.e.	0.2383					
sigma^2 estimated as 82.1: log likelihood = -176.59,						
aic = 357.17						
Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	M
ASE						
Training set	2.1651	8.104102	5.520786	2.288554	21.07896	0.6029
859						
	ACF1					
Training set	0.1617822					
Elaborado por: Autor						

Tabla No. 4.23. Estadístico Z - Orden (0,0,0)x(0,1,1)₁₂, Serie Temporal Aditivo para Gasolina				
z test of coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
sma1	-0.60434	0.23831	-2.536	0.01121 *

Signif. codes:	0	'***'	0.001	'**'
	0.01	'*'	0.05	'.'
	0.1	' '		1
Elaborado por: Autor				

Finalmente, analizamos el comportamiento de nuestro modelo respecto a la prueba de ruido blanco (Ljung-Box).

Tabla No. 4.24. Prueba Box-Ljung- Orden (0,0,0)x(0,1,1)₁₂, Serie Temporal Aditivo para Gasolina
Box-Ljung test data: residuals(ajuste4) X-squared = 1.6503, df = 1, p-value = 0.1989
Elaborado por: Autor



4.5.4. ELECCIÓN DEL MEJOR MODELO

A continuación, compararemos los modelos planteados utilizando los resultados de los Errores Medios de Porcentaje Absoluto (MAPE), los criterios de información de Akaike (AIC) y criterios de información bayesiano (BIC) según procede a continuación y finalmente, se muestra en breve el cuadro comparativo de los modelos resultantes en relación a la serie de los aditivos "Gasolina":

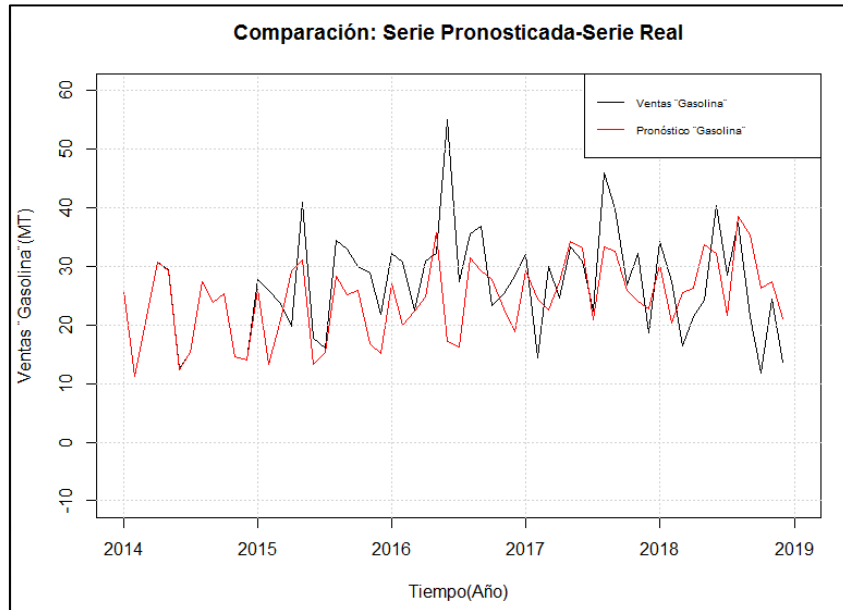
Tabla No. 4.25. Comparación de modelos resultantes. Serie temporal gasolina.

No.	MODELO	AIC	BIC	MAPE
1	ARIMA (1,0,0)	434,24	440,51	30,21
2	ARIMA (2,0,0)	436,21	444,59	30,24
3	ARIMA (0,1,1)	428,04	432,20	30,31
4	ARIMA (0,0,0)x(0,1,1) ₁₂	357,17	360,91	21,08

Elaborado por: Autor

De acuerdo con los criterios AIC y BIC, se selecciona el modelo ARIMA (0,0,0)x(0,1,1)₁₂, ya que presenta los menores valores en estos criterios. Por otra parte, utilizando el criterio MAPE el modelo 4 es también el más idóneo considerando que presenta el menor valor comparado con los demás modelos, esto significaría que es el modelo que menos se desvía de los datos originales:

Figura No.4.47. Comparación de la serie temporal pronosticada y la serie temporal de datos reales de las ventas del segmento gasolina. Período: 2014-2018.

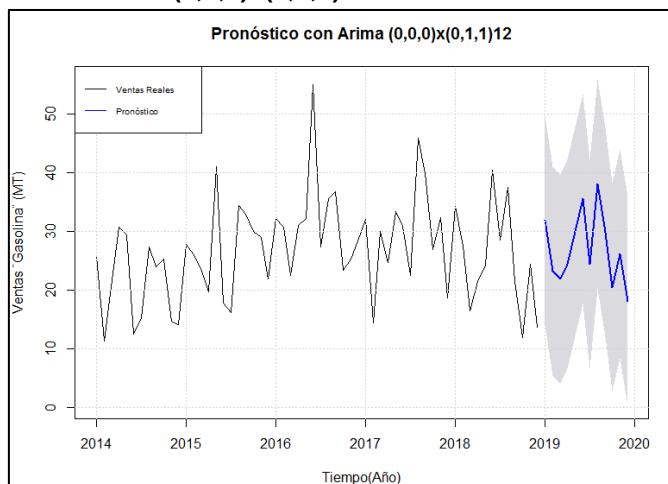


Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

4.5.5. PREDICCIÓN UTILIZANDO ARIMA

Con el fin de predecir o pronosticar los posibles valores futuros de nuestra serie temporal, determinando también intervalos de confianza utilizaremos la función *forecast()* en R,:

Figura No.4.48. Pronóstico de ventas de los aditivos del segmento gasolina a través de Arima (0,0,0)x(0,1,1)12. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Adicionalmente, tabulamos los datos obtenidos dentro del presente pronóstico:

Tabla No. 4.26. Pronósticos resultantes del ajuste 4 (Periodo 2019). Serie temporal gasolina

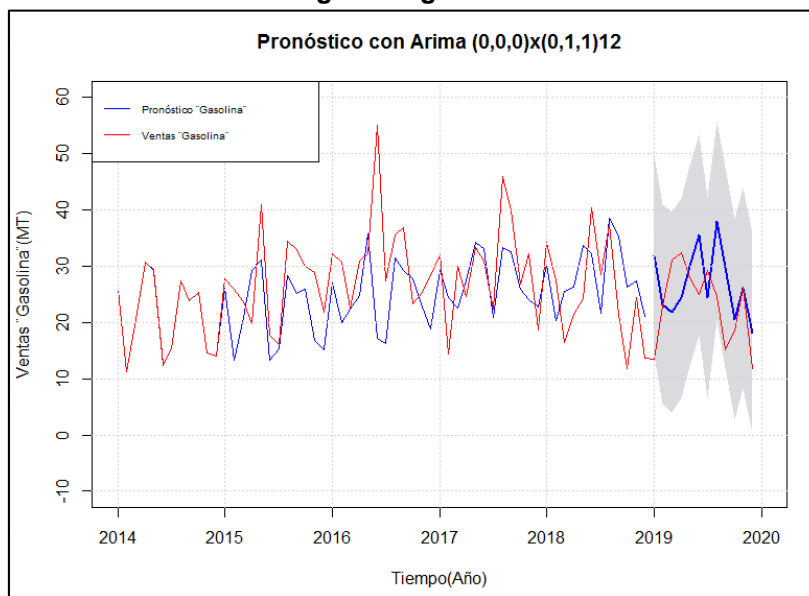
Mes	Pronóstico (MT)	Límite Inferior (MT)	Límite Superior (MT)
Enero	31,94	14,14	49,73
Febrero	23,21	5,42	41,01
Marzo	21,85	4,05	39,64
Abril	24,35	6,56	42,15
Mayo	29,94	12,14	47,73
Junio	35,54	17,74	53,33
Julio	24,39	6,59	42,18
Agosto	38,05	20,26	55,85
Septiembre	29,87	12,07	47,67
Octubre	20,46	2,67	38,26
Noviembre	26,15	8,35	43,94
Diciembre	18,1	0,31	35,9

Elaborado por: Autor

A través de la sumatoria de los pronósticos mensuales de los aditivos que conforman la serie "Gasolina", conseguimos las ventas totales estimadas del año 2019: 323.84MT

Para dar por finalizado el análisis de la presente serie, con la finalidad de validar nuestro modelo comparamos los valores reales de las ventas de la serie "Gasolina" del año 2019 y los datos pronosticados mediante la siguiente figura:

Figura No.4.49. Comparación de la serie temporal pronosticada y la serie temporal de datos reales del segmento gasolina. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

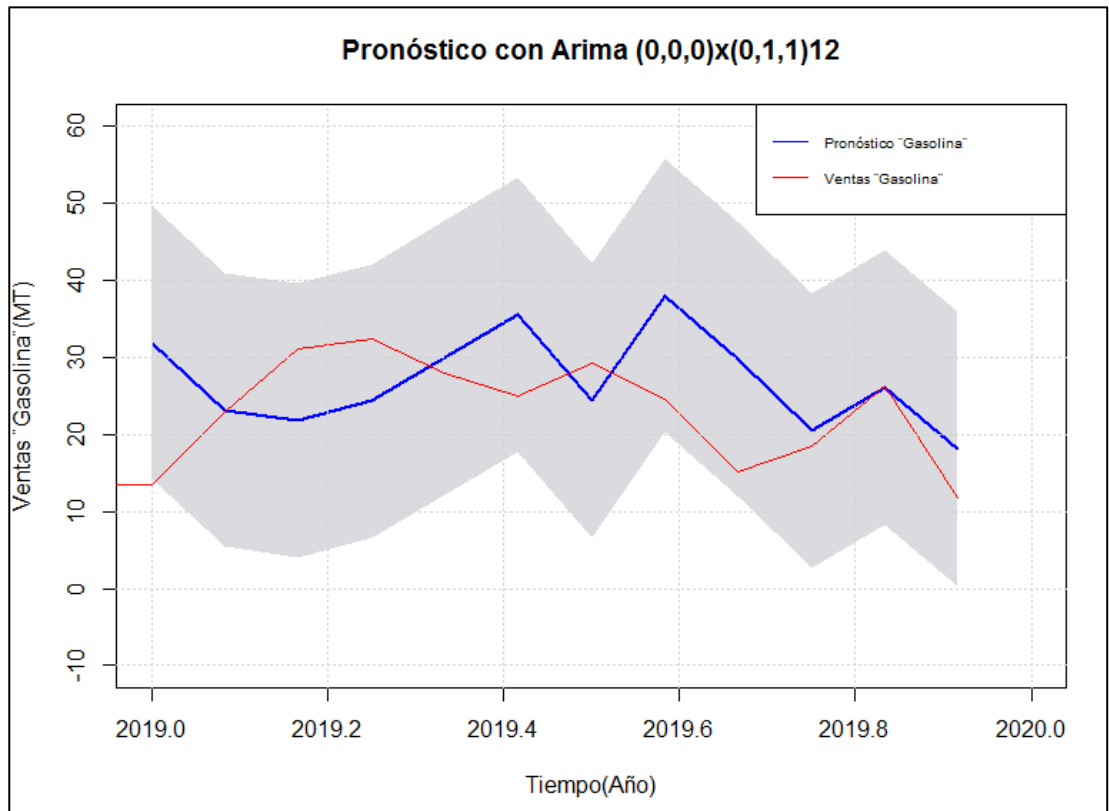
De esta forma logramos comparar los resultados reales y pronosticados de las ventas de la serie "Gasolina" durante el año 2019: 278.77

Tabla No. 4.27. Comparación de ventas pronosticadas y reales durante el año 2019. Serie temporal gasolina.

Ventas Reales (MT)	Ventas Pronosticadas (MT)	Diferencia (MT)	Diferencia Porcentual
278,77	323,84	45,07	16,17%

Elaborado por: Autor

Figura No.4.50. Comparación de la serie temporal pronosticada y la serie temporal de datos reales del segmento gasolina. Período: 2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

4.6. MODELO: DIESEL

4.6.1. ANÁLISIS EXPLORATORIO DE DATOS

Finalmente analizaremos los datos de las ventas de los aditivos relacionados con el segmento de Diésel a través de la sumatoria de los aditivos del segmento Diésel (Diésel 1, Diésel 2, Diésel 3, Diésel 4 y Diésel 5) con el fin de formar la serie de los aditivos "Diesel" para la fabricación de lubricantes de motores de automóviles alimentados por este tipo de combustibles, expresadas en kilogramos,

correspondientes a los periodos 2014 a 2019, obtenidos a través de los registros de ventas de la empresa importadora y comercializadora de materias primas.

Tabla No. 4.28. Serie histórica mensual de las ventas de los aditivos para la fabricación de lubricantes para motores alimentados por "Diésel", expresadas en kilogramos, periodo 2014-2019.

Ventas Mensuales de Aditivos "Diesel " (kg)						
Mes/ Año	2014	2015	2016	2017	2018	2019
Enero	145757	79386	92538	114397	100523	91610
Febrero	212345	125160	108104	107152	70256	35854
Marzo	75469	112414	72293	154910	114526	146582
Abril	189265	117982	140618	65300	51703	93626
Mayo	162442	142144	89095	128690	147209	97439
Junio	129852	137926	106824	126988	120328	67042
Julio	118706	115244	153322	104858	88325	89438
Agosto	123907	107296	74599	83221	112922	82435
Septiembre	121568	125876	63891	102440	135686	85184
Octubre	84158	126474	72621	67306	75068	94514
Noviembre	101880	87110	133508	126162	127344	150192
Diciembre	108588	84878	109487	41145	115607	84200

Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Tabla No. 4.29. Serie histórica mensual de las ventas de los aditivos para la fabricación de lubricantes para motores alimentados por "Diésel" expresadas en toneladas métricas, periodo 2014-2019.

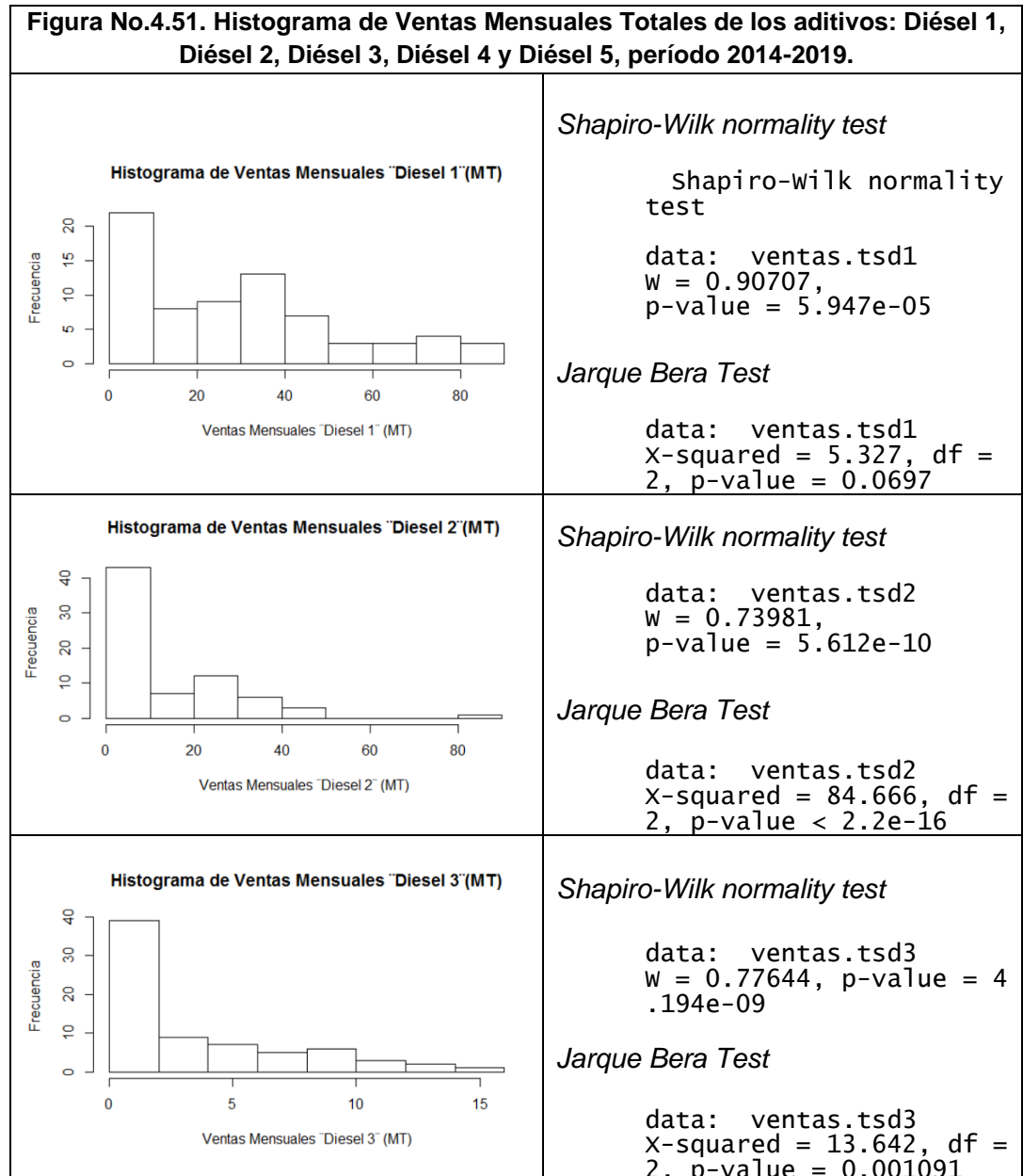
Ventas Mensuales de Aditivo "Diesel" (MT)						
Mes/Año	2014	2015	2016	2017	2018	2019
Enero	146	79	93	114	101	92
Febrero	212	125	108	107	70	36
Marzo	75	112	72	155	115	147
Abril	189	118	141	65	52	94
Mayo	162	142	89	129	147	97
Junio	130	138	107	127	120	67
Julio	119	115	153	105	88	89
Agosto	124	107	75	83	113	82
Septiembre	122	126	64	102	136	85
Octubre	84	126	73	67	75	95
Noviembre	102	87	134	126	127	150
Diciembre	109	85	109	41	116	84

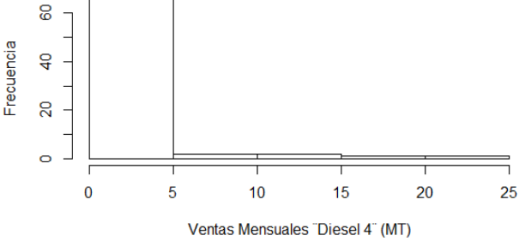
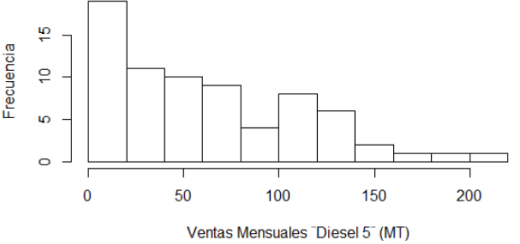
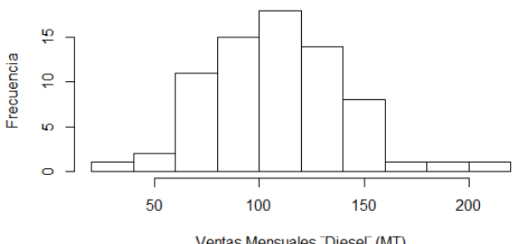
Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Procedemos entonces a transformar los datos a series temporales y a graficarlos con el fin de verificar sus comportamientos, importando los datos a R y activando

las librerías a ser utilizadas. Considerando que la serie diésel está compuesta por los datos de 5 aditivos diferentes, analizaremos inicialmente el comportamiento de cada uno de forma individual.

Así también procedemos con la generación de los histogramas de frecuencias de cada una de las series ingresadas. Posteriormente llevamos a cabo pruebas de normalidad para cada serie a través de las pruebas de normalidad Shapiro-Wilk y Jarque Bera:

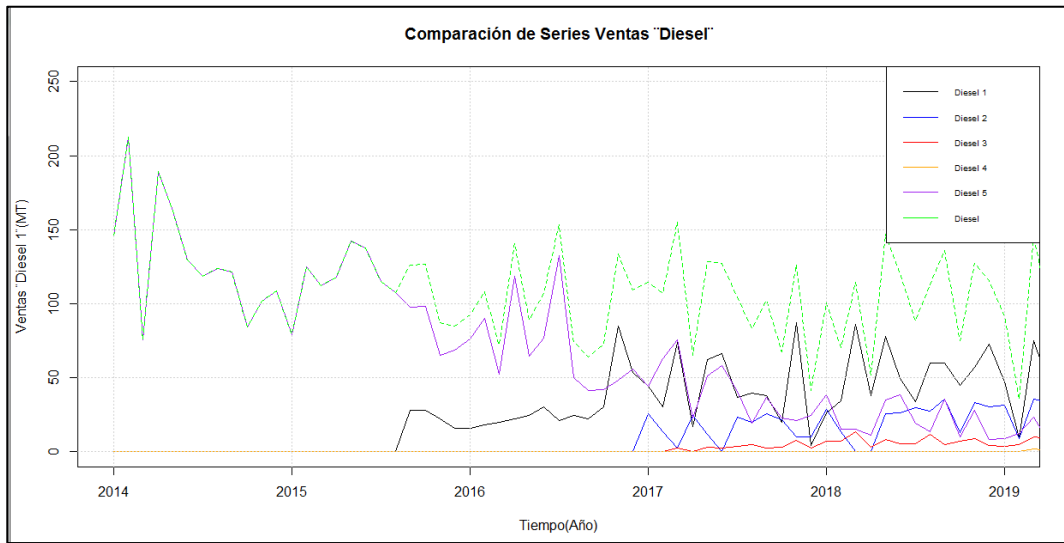


<p>Histograma de Ventas Mensuales "Diesel 4"(MT)</p> 	<p><i>Shapiro-Wilk normality test</i></p> <p>data: ventas.tsd4 $W = 0.32995$, $p\text{-value} < 2.2e-16$</p> <p><i>Jarque Bera Test</i></p> <p>data: ventas.tsd4 $X\text{-squared} = 1092.7$, $df = 2$, $p\text{-value} < 2.2e-16$</p>
<p>Histograma de Ventas Mensuales "Diesel 5"(MT)</p> 	<p><i>Shapiro-Wilk normality test</i></p> <p>data: ventas.tsd5 $W = 0.91573$, $p\text{-value} = 0.0001388$</p> <p><i>Jarque Bera Test</i></p> <p>data: ventas.tsd5 $X\text{-squared} = 6.9696$, $df = 2$, $p\text{-value} = 0.03066$</p>
<p>Histograma de Ventas Mensuales "Diesel"(MT)</p> 	<p><i>Shapiro-Wilk normality test</i></p> <p>data: ventas.tsd $W = 0.98252$, $p\text{-value} = 0.4174$</p> <p><i>Jarque Bera Test</i></p> <p>data: ventas.tsd $X\text{-squared} = 3.6768$, $df = 2$, $p\text{-value} = 0.1591$</p>
<p>Elaborado por: Autor.</p>	

En base a los histogramas y a las pruebas de normalidad ejecutados para cada una de las series podemos observar que sólo la serie de la sumatoria de las ventas de los aditivos para gasolina *ventas.tsd* tienen una distribución normal. Para este caso específico los valores p obtenidos en estas pruebas para la serie son 0.4174 (Shapiro-Wilk) y 0.1591 (Jarque Bera) siendo estos superiores a 0.05, decimos que no tenemos evidencia estadística suficiente para rechazar la hipótesis nula de que la distribución de nuestros datos es normal.

Procedemos entonces a graficar los datos de las series motivo de estudio respecto al tiempo:

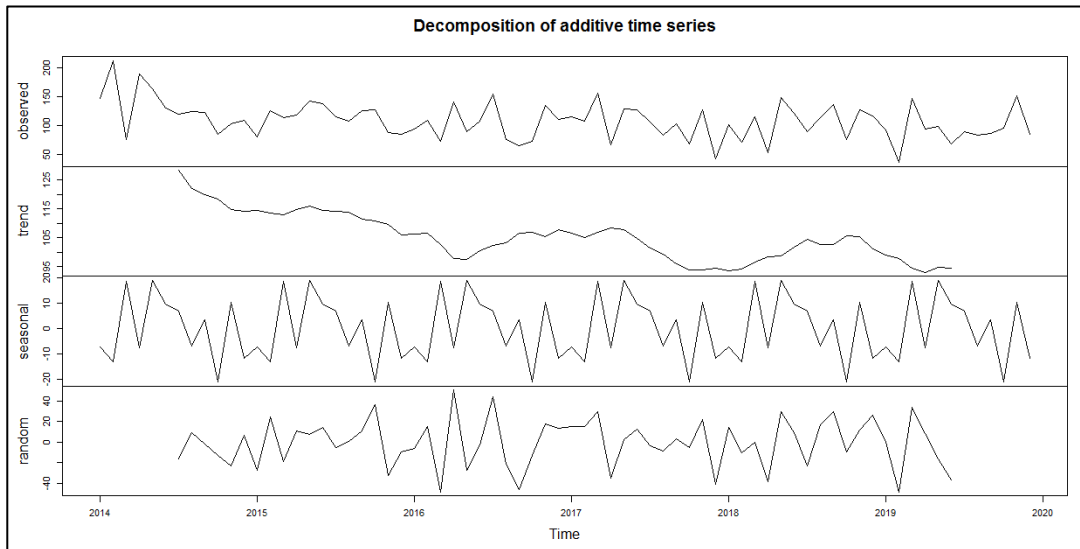
Figura No.4.52. Series temporales de las ventas de los aditivos para la fabricación de lubricantes alimentados por diésel, expresadas en toneladas métricas. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Para efectos del análisis del comportamiento del segmento diésel, utilizaremos únicamente la serie ventas.tsd y continuamos con su descomposición respectiva:

Figura No.4.53.Descomposición de la Serie Temporal Diesel. Período: 2014-2019.



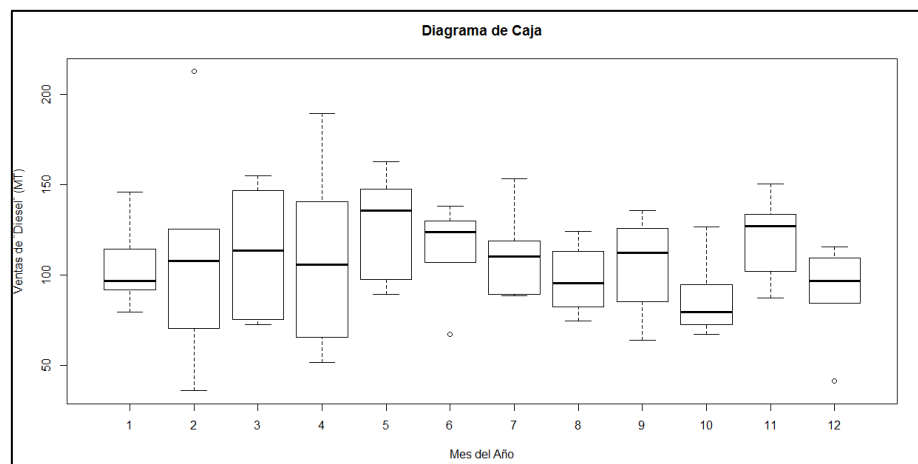
Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

En la figura anterior podemos evidenciar que existe una tendencia decreciente respecto al tiempo, adicionalmente observamos los componentes estacionales y aleatorios.

En el siguiente diagrama, en donde separamos los datos por meses observamos que el valor de media más alto se localiza en el mes de mayo (5), presentando una tendencia decreciente hasta el mes de agosto (8). Se repite el mismo fenómeno en

el mes de noviembre (11), que en las series de tiempo anteriormente estudiadas (ventas totales, gasolina y fuera de borda) el valor de media aumenta significativamente en relación al mes de octubre (11) debido a que las empresas se abastecen sus inventarios para la última producción del año, teniendo en cuenta que generalmente el mes de diciembre (12) tiene días de producción reducidos. Para este caso, también observamos 2 datos anómalos; el primero de ellos es un valor superior a la media en el mes de febrero asociado a consumos puntuales de la industria petrolera y el segundo en el mes de junio (6), que puede ser evidencia de falta de abastecimiento oportuno de aditivos trayendo como resultado reducción significativa en ventas.

Figura No.4.54. Diagrama de caja de la serie temporal diésel, segmentada por mes, en toneladas métricas. Período: 2014-2019.



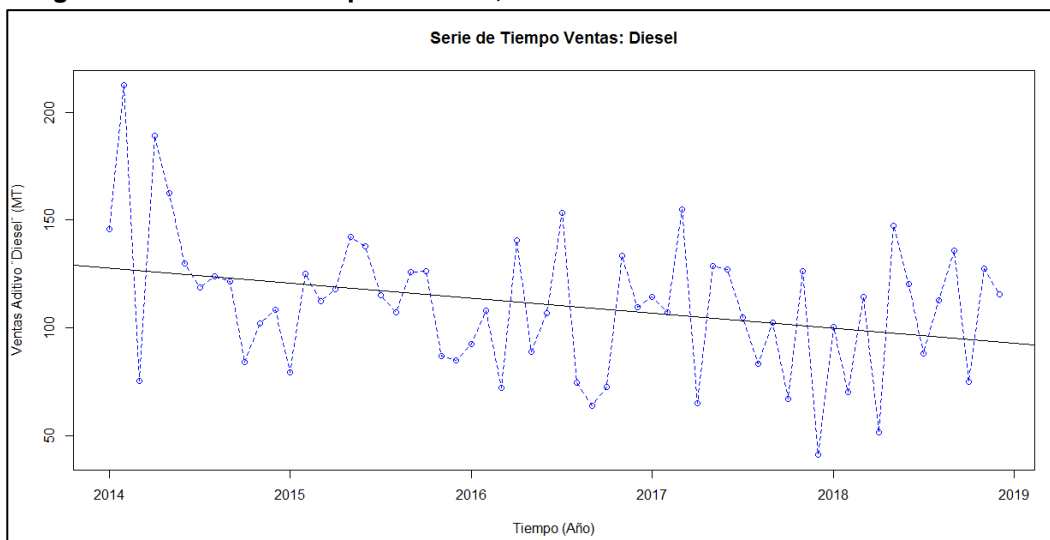
Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

4.6.2. ANÁLISIS DE TENDENCIA

Finalizado el AED de la serie temporal del segmento diésel, trabajemos únicamente con los datos de los primeros 5 años (2014 a 2018) de las ventas de los aditivos "Diésel", los datos correspondientes al año 2019 se utilizarán para verificar qué tan cercano se encuentra nuestro pronóstico de los datos reales.

Graficaremos únicamente el periodo 2014-2018 para verificar si estos se comportan de forma estacionaria:

Figura No.4.55. Serie temporal diesel, en toneladas métricas. Período: 2014-2018.

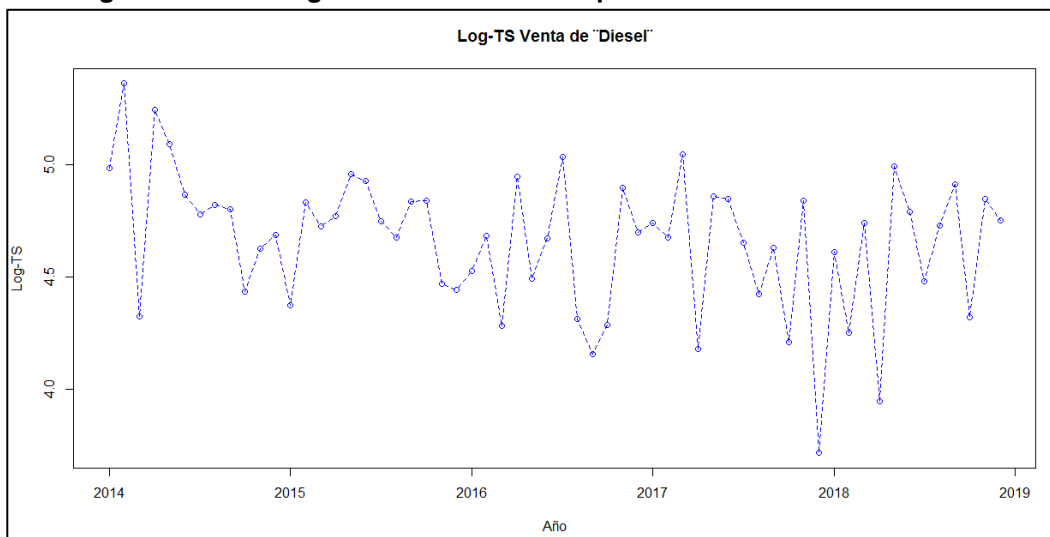


Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Evidenciamos en el gráfico anterior que la serie de tiempo no es estacionaria respecto a su media y sus fluctuaciones en el tiempo no son uniformes; es decir, tiene variaciones irregulares. Adicionalmente, corroboramos lo anteriormente expuesto a través de la aplicación de la prueba Phillips-Perron que nos indica que debemos aplicar 1 diferencia a nuestra serie para alcanzar la estacionariedad.

Por esta razón debemos eliminar las variaciones irregulares de nuestra serie de datos, aplicando logaritmos:

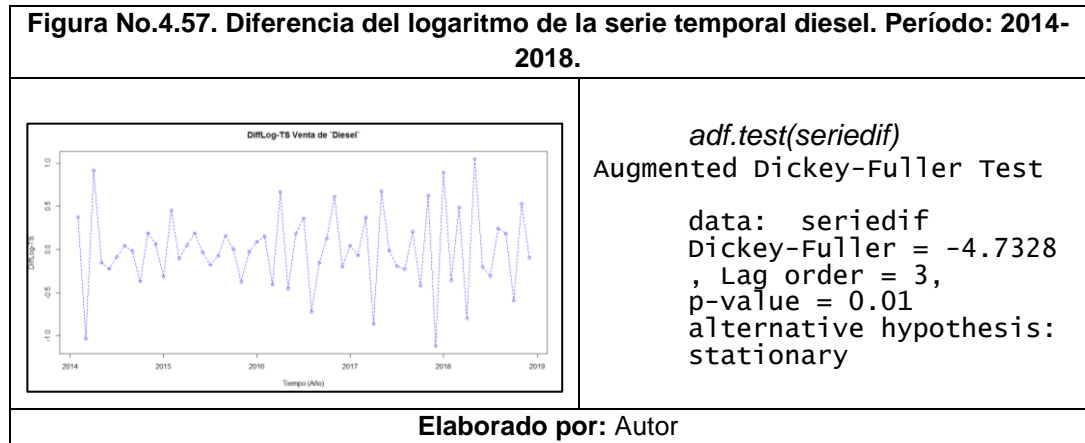
Figura No.4.56. Logaritmo de la serie temporal diesel. Período: 2014-2018.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Una vez que hemos reducido el rango de diferencia entre los valores mensuales, aplicamos una primera diferencia a nuestra serie. Con fin de comprobar si nuestra

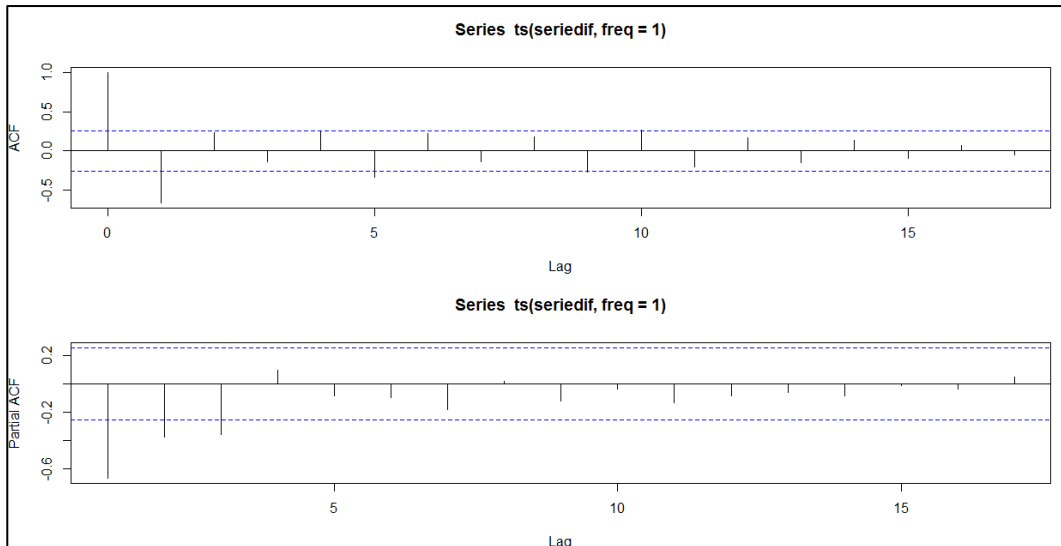
serie es estacionaria utilizamos la Prueba Aumentada de Dickey-Fuller. Al ser el valor p obtenido menor que 0.05 rechazamos la hipótesis nula de que la serie temporal posee raíz unitaria, lo cual significaría que nuestra serie es estacionaria.



4.6.3. AJUSTE DEL MODELO

Para este último modelo, también analizaremos las funciones de autocorrelación típicamente utilizadas dentro de la metodología Box-Jenkins: Función de Autocorrelación y Función de Autocorrelación Parcial:

Figura No.4.58. ACF y PACF de la serie temporal *seriedif*. Aditivos Diésel.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

De acuerdo al gráfico de PACF vemos que existen autocorrelaciones nulas a partir del tercer rezago (lag); para efectos de este modelo plantearemos modelos ARIMA, cuya parte AR serán de orden 3 y orden 2, para verificar el nivel de significancia de los coeficientes ar3 y ar2 antes de evaluar rezagos superiores a estos niveles. Por otra parte, y considerando que el comportamiento observado en el gráfico ACF nos muestra un decaimiento exponencial, la parte MA muy probablemente tendrá 1 o 2

componentes. En base a estos criterios procedemos a plantear los diferentes modelos del presente estudio.

4.6.3.1. Modelos ARIMA Estacionales (p,d,q)x(P,D,Q)s

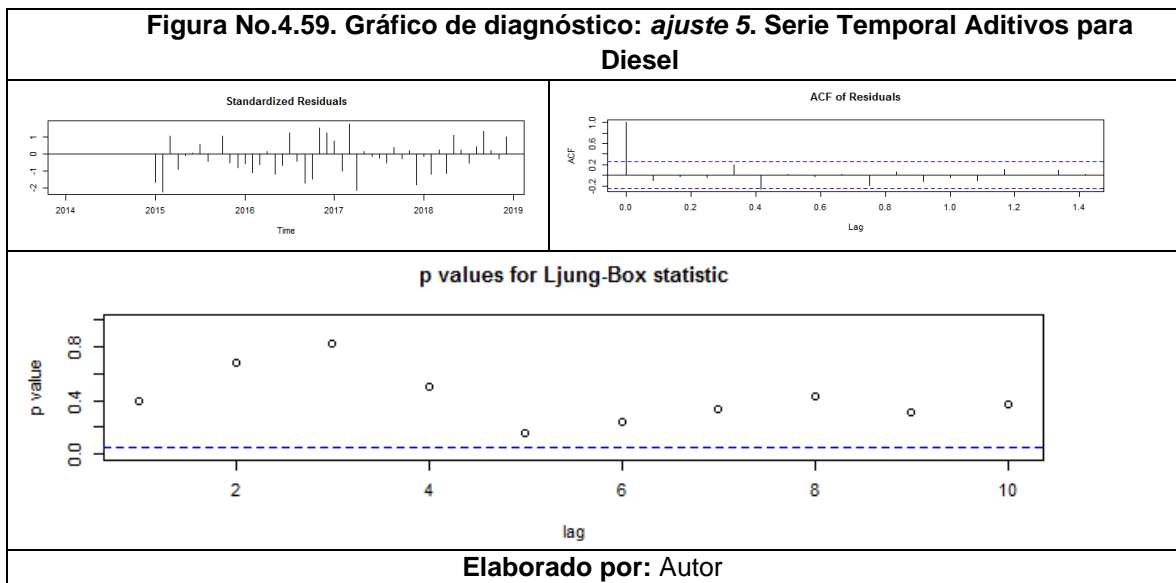
Como hemos visto a lo largo del presente estudio, tanto las ventas como los pronósticos de las diferentes líneas de productos pueden ser representadas por modelos de características muy similares. No obstante, para el planteamiento del modelo del aditivo diésel, al estudiar el modelo AR(3) su coeficiente ar3 resultó significativo dentro del resumen estadístico del modelo planteado; aprovechando este nivel de capacidad de representación de nuestra variable de respuesta, planteamos a continuación un modelo en donde el valor de p, asociado al componente autorregresivo tendrá un valor de 3 (lo cual es concordante con lo encontrado en el gráfico PACF y el retardo correspondiente), mientras que mantenemos los valores de D equivalente a 1 por la diferencia aplicada y Q por el componente de medias móviles que se ha mantenido constante a lo largo del análisis de nuestras variables. Adicionalmente, agregaremos a nuestro sistema de estudio las características estacionales a la serie, mediante el coeficiente 12 relacionados a los ciclos mensuales de ventas.

Tabla No. 4.30. Resumen Estadístico - Orden (3,0,0)x(0,1,1)₁₂, Serie Temporal Aditivo para Diesel						
Orden: (3,0,0)x(0,1,1) ₁₂						
Coefficients:						
	ar1	ar2	ar3	sma1		
	-0.0729	0.1373	0.3385	-0.4440		
s.e.	0.1461	0.1727	0.1546	0.2049		
sigma^2 estimated as 1180: log likelihood = -239.37,						
aic = 488.73						
Training set error measures:						
	MASE	ME	RMSE	MAE	MPE	MAPE
Training set	-5.038919	30.72224	22.24293	-11.08434	24.68768	0.6320895
		ACF1				
Training set	-0.1075354					
Elaborado por: Autor						

Tabla No. 4.31. Estadístico Z - Orden (3,0,0)x(0,1,1) ₁₂ , Serie Temporal Aditivo para Diesel				
z test of coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.072895	0.146113	-0.4989	0.61785
ar2	0.137260	0.172713	0.7947	0.42677
ar3	0.338493	0.154570	2.1899	0.02853 *
sma1	-0.443953	0.204884	-2.1669	0.03025 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Elaborado por: Autor				

Tabla No. 4.32. Prueba Box-Ljung- Orden (3,0,0)x(0,1,1) ₁₂ , Serie Temporal Aditivo para Diesel
Box-Ljung test
data: residuals(ajuste5)
X-squared = 0.72911, df = 1, p-value = 0.3932
Elaborado por: Autor



4.6.4. ELECCIÓN DEL MEJOR MODELO

Finalmente, compararemos los modelos planteados utilizando los resultados de los Errores Medios de Porcentaje Absoluto (MAPE), los criterios de información de *Akaike* (AIC) y criterios de información bayesiano (BIC):

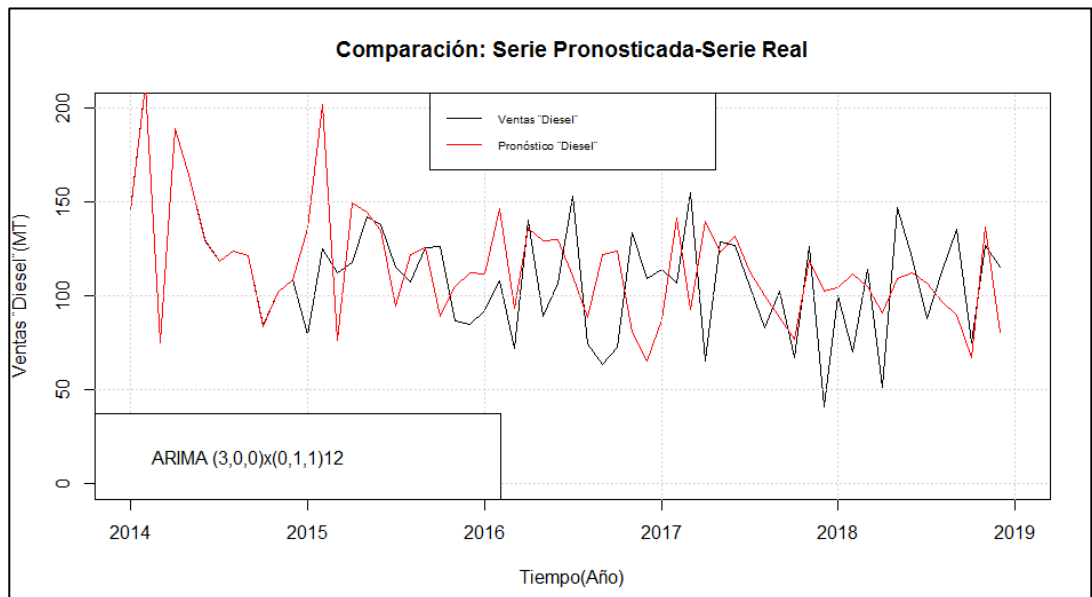
Tabla No. 4.33. Comparación de modelos resultantes. Serie temporal diesel.

No.	MODELO	AIC	BIC	MAPE
1	ARIMA (2,1,0)	580,47	586,70	26,46
2	ARIMA (3,1,0)	576,03	584,34	25,07
3	ARIMA (0,1,1)	580,13	584,29	27,73
4	ARIMA (0,1,2)	577,45	583,69	21,08
5	ARIMA (3,0,0)x(0,1,1) ₁₂	488,73	498,09	24,69
6	ARIMA (3,0,0)x(0,1,2) ₁₂	490,47	501,70	24,84

Elaborado por: Autor

De acuerdo con los criterios AIC y BIC, se selecciona el modelo ARIMA (3,0,0)x(0,1,1)₁₂, ya que presenta los menores valores en estos criterios. Cabe mencionar que, utilizando el criterio MAPE el ajuste 4, ARIMA (0,1,2) sería el más idóneo considerando que presenta el menor valor comparado con los demás modelos; no obstante, utilizaremos sólo el ajuste 5 para efectos de pronóstico; puesto que, representarían de forma más adecuada a nuestra variable de respuesta.

Figura No.4.60. Comparación de la serie temporal pronosticada arima (3,0,0)x(0,1,1)₁₂ y la serie temporal de datos reales de las ventas del segmento diésel. Período: 2014-2018.



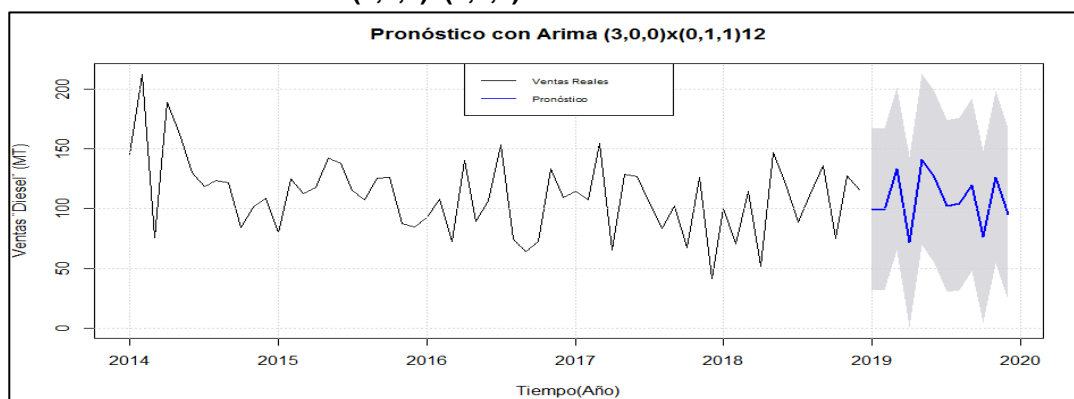
Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos

Elaborado por: Autor

4.6.5. PREDICCIÓN UTILIZANDO ARIMA

Con la finalidad de predecir o pronosticar los posibles valores futuros de nuestra serie temporal, utilizaremos los mejores ajustes seleccionados en el numeral anterior y adicionalmente estimaremos intervalos de confianza para los pronósticos obtenidos a través de la función *forecast()* en R:

Figura No.4.61. Pronóstico de ventas de los aditivos del segmento diésel a través de arima (3,0,0)x(0,1,1)12. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

Para efectos comparativos, procedemos con la tabulación los datos pronosticados a través del ajuste 5:

Tabla No. 4.34. Matriz de pronósticos Ajuste 5. Serie temporal diesel.

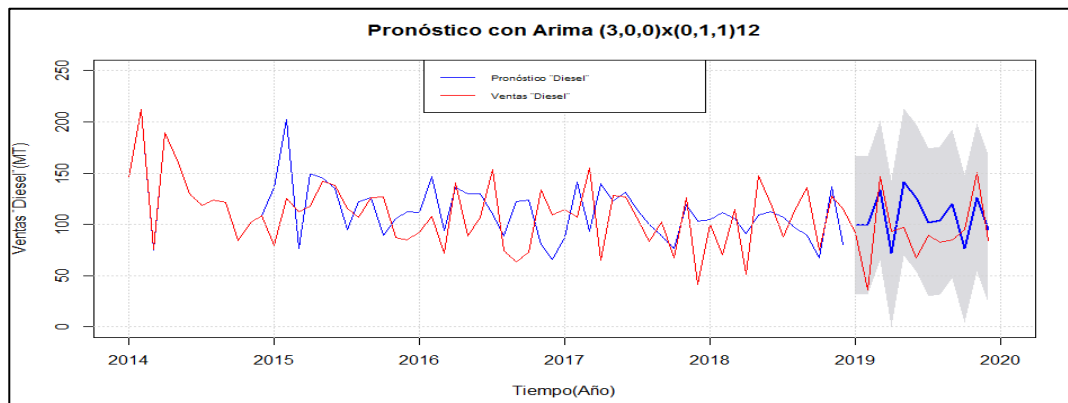
Mes	Pronóstico (MT)	Límite Inferior (MT)	Límite Superior (MT)
Enero	99,74	32,41	167,07
Febrero	99,24	31,73	166,75
Marzo	133,43	65,24	201,62
Abril	71,63	0,15	143,1
Mayo	141,35	69,86	212,85
Junio	126,27	54,49	198,05
Julio	102,06	29,98	174,13
Agosto	103,85	31,77	175,92
Septiembre	119,9	47,76	192,03
Octubre	76,3	4,14	148,46
Noviembre	126,33	54,17	198,5
Diciembre	95,03	22,85	167,2

Elaborado por: Autor

A través de la sumatoria de los pronósticos mensuales de los aditivos que conforman la serie "Diesel", conseguimos las ventas totales estimadas del año 2019 para el ajuste 5: 1348.48 MT

Para dar por finalizado el análisis de la presente serie, con el objetivo de validar nuestros modelos comparamos los valores reales de las ventas de la serie "Diesel" del año 2019 y los datos pronosticados mediante las siguientes figuras:

Figura No.4.62. Comparación de la serie temporal pronosticada con arima(3,0,0)x(0,1,1)₁₂ y la serie temporal de datos reales del segmento diésel. Período: 2014-2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

De esta forma conseguimos comparar los resultados reales y pronosticados de las ventas de la serie "Diésel" durante el año 2019:

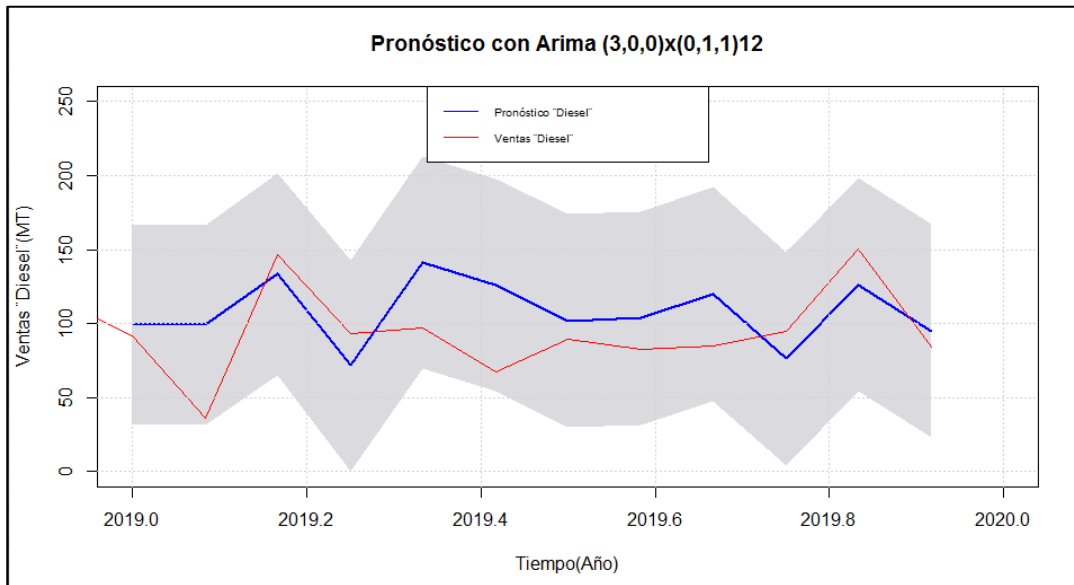
Tabla No. 4.35. Comparación de ventas pronosticadas con arima (3,0,0)x(0,1,1)₁₂ y ventas reales durante el año 2019. Serie temporal diesel.

Ventas Reales (MT)	Ventas Pronosticadas Ajuste 5 (MT)	Diferencia (MT)	Diferencia Porcentual
1118,12	1348,48	230,37	20,60%

Elaborado por: Autor

En relación a los resultados obtenidos, podemos decir que en el pronóstico generado con el ajuste 5, ARIMA (3,0,0)(0,1,1)₁₂, observamos que todos los datos reales se encuentran dentro del intervalo de confianza establecido para su pronóstico:

Figura No.4.63. Comparación de la serie temporal pronosticada con arima $(3,0,0) \times (0,1,1)_{12}$ y la serie temporal de datos reales del segmento diésel. Período: 2019.



Fuente: Registros de Ventas de la Empresa Comercializadora de Aditivos
Elaborado por: Autor

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- a) A través de la aplicación de la metodología expuesta en el presente estudio fue posible generar un modelo ARIMA con el objetivo de establecer pronósticos a corto plazo referentes a la demanda (ventas) mensuales totales de los aditivos de una empresa comercializadora de materias primas para la fabricación de aceites lubricantes, a pesar de que la variable de estudio presenta una variabilidad alta.
- b) El modelo de pronóstico ARIMA seleccionado que tuvo el mejor ajuste en relación al comportamiento de nuestra variable de respuesta, ventas totales de aditivos de la empresa comercializadora de materias primas para la producción de lubricantes es: $ARIMA(2,0,0) \times (1,1,1)_{12}$.
- c) Así también, fue posible observar que existió una diferencia anual de 487,54 MT (17,52%) entre las ventas totales pronosticadas y las ventas reales del año 2019. Esta variación, de acuerdo a lo investigado, se podría deber principalmente a una condición exógena que se presentó durante el año en mención, en la que uno de los principales clientes de la empresa comercializadora de aditivos, que constituía el 14.30% de las ventas de los aditivos, adquiriendo anualmente 445,51 MT de productos, migró hacia otra tecnología de estas especialidades químicas impactando a nuestra variable de respuesta.
- d) De forma general observamos que los modelos planteados para las series temporales: ventas totales de aditivos, fuera de borda, gasolina y diésel, son capaces de pronosticar los comportamientos dentro de los intervalos de confianza especificados para cada uno. No obstante, en relación a la toma de decisiones en base a estos pronósticos, se debe considerar que la generación de los modelos se fundamenta en los datos históricos (pasados) y de presentarse eventualidades significativas en el mercado en tiempo presente, estas deben ser tomadas en cuenta hasta que se tengan nuevos datos de ventas y puedan actualizarse los modelos generando nuevos pronósticos.

5.2. RECOMENDACIONES

- a) Con el objetivo de reducir la diferencia entre el valor pronosticado y el valor real podrían removerse de los datos, los registros de ventas del cliente que cambió de tecnología de aditivos, de esta forma dentro de la predicción se contemplarían únicamente los clientes actuales y se tendrían pronósticos más cercanos a la realidad actual de la empresa comercializadora de materias primas para la fabricación de lubricantes.
- b) Se recomienda también, validar los pronósticos con otras variables de control como son: solicitudes de consumo de aditivos periódicas a los clientes, verificación de inventarios periódica, reuniones mensuales para la comprobación y actualización de los datos.
- c) Se recomienda llevar a cabo la validación de forma periódica, y de ser necesario el replanteamiento de modelos establecidos para cada serie, con la finalidad de confirmar que estos continúan representando el comportamiento de la serie de forma adecuada.

6. Referencias

Munguía, L. & Estrada, L. (2018). Modelos de Pronóstico a corto plazo sobre el comportamiento inflacionario en países de América del Norte aplicando series de tiempo: 2006-2017, México: Universidad Autónoma del Estado de México.

Sánchez, D. (2018). Modelo ARIMA para el pronóstico de la producción de cacao en el Perú 2012-2018, Perú: Universidad Nacional de Trujillo.

Bedoya, S. (2018). Modelamiento univariado del número de defunciones infantiles producidas por infecciones respiratorias agudas, a través de la metodología Box-Jenkins, Puno 2008-2016, Perú: Universidad Nacional del Altiplano.

Box, G.E.P. y G.M. Jenkins (1970). Time series analysis: forecasting and control. San Francisco, Holden Day.

Zúñiga, L. (2018). Desarrollo de un modelo de planificación de demanda mediante el análisis de aditivo para la empresa Hidromecánica Andina Cía. Ltda, Ecuador: Universidad Católica Santiago de Guayaquil.

Tsay, R. (2010). Analysis of financial time series. New Jersey, John Wiley & Sons Inc.

Liu, L. & Hudak, G. (1992). Forecasting and time series analysis using the SCA statistical system. Chicago: Scientific Computing Associates Corp.

Flores, L. (2017). Pronóstico del índice nacional de precios al consumidor. *Revista Iberoamericana de contaduría, economía y administración*. Recuperado desde: [file:///C:/Users/HMA/Downloads/Dialnet-PronosticoDelIndiceNacionalDePreciosAlConsumidorFo-6089290%20\(1\).pdf](file:///C:/Users/HMA/Downloads/Dialnet-PronosticoDelIndiceNacionalDePreciosAlConsumidorFo-6089290%20(1).pdf)

APEL, (2019). Boletín del sector de lubricantes. Ecuador: Asociación de Productores y Exportadores de Lubricantes.

González, M. (2009). *Técnicas de predicción económica. Serie de documentación Sarriko-on*. Recuperado desde: <https://addi.ehu.es/bitstream/handle/10810/12493/05-09pil.pdf?sequence=1&isAllowed=y>

Lubrizol (2017). Lubrication Fundamentals. Wickliffe, Ohio: The Lubrizol Cooperation.

Huilca, A., (27 de abril de 2020). *Tipos de investigación científica y sus características*. Rastreador. <https://ulexion.com/tipos-de-investigacion-cientifica-y-sus-caracteristicas/>

Rodó, P., (2019). Contraste de Dickey Fuller. Rastreador: <https://economipedia.com/definiciones/contraste-de-dickey-fuller.html>

Quezada del Villara, A.V., Rodríguez Picón, L.A., Pérez Olguín I.JC., Rodríguez-Picón I. (2018). Procesos estocásticos aplicados en análisis de datos de degradación-revisión de literatura. *Mundo Fesc*, vol. 15, no. 1, pp. 85-94. Recuperado desde: <file:///C:/Users/HMA/Downloads/Dialnet-ProcesosEstocasticosAplicadosEnAnalisisDeDatosDeDe-6770919.pdf>

Sánchez, P., Velásquez, J. (2005). Modelado estructural no lineal de series temporales. Recuperado desde: [file:///C:/Users/HMA/Downloads/ModeladoEstructuralNoLinealdeSeriesTemporales%20\(1\).pdf](file:///C:/Users/HMA/Downloads/ModeladoEstructuralNoLinealdeSeriesTemporales%20(1).pdf)

Parra, F. (25 de enero de 2019) *Estadística y Machine Learning con R*. Rastreador: <https://bookdown.org/content/2274/portada.html>

Ramírez, A. (2019). Series de tiempo: Estacionariedad. Rastreador: <https://rpubs.com/Arthurus/492144#:~:text=%C2%BFPor%20qu%C3%A9%20es%20importante%20la,a%20la%20hora%20de%20predecir.>

Carvajal, A. (2014) *Series temporales: Modelos heterocedásticos condicionales. Una aplicación usando R*. España: Universidad de Granada.

Gujarati, D., Porter, D. (2010). *Econometría*. México, The McGraw-Hill Companies Inc.

Dietrichson, A. (2019). *Métodos Cuantitativos*. Rastreador: <https://bookdown.org/dietrichson/metodos-cuantitativos/edicion.html>

Dietrichson, A. (2019). *Métodos Cuantitativos*. Rastreador: <https://bookdown.org/dietrichson/metodos-cuantitativos/edicion.html>

Guamán, A. (2016) *Disposiciones técnicas de seguridad contra incendios*. Guayaquil, Benemérito Cuerpo de Bomberos.

León, E., Romo, T., (2018) Plan de marketing para la línea de lubricantes Amalie de la empresa Filtrocorp en la ciudad de Guayaquil. Ecuador: Universidad Católica Santiago de Guayaquil.

Flores, P., Muñoz. L., Sánchez, T., (2019) Estudio de potencia de pruebas de normalidad usando distribuciones desconocidas con distintos niveles de normalidad. Ecuador: Escuela Superior Politécnica de Chimborazo.

7. Apéndices y anexos

7.1. Base de datos utilizada

Los siguientes datos temporales son desde Enero del 2014 hasta Diciembre del 2019.

series ventas.ts	series424	series6446	series7067C	seriesPV1021
313112	16257	10776	16800	25728
413410	15183	7350	14400	11256
323048	22113	6513	18370	20502
437745	11491	8575	21475	30753
378865	16492	6860	15850	29346
353337	19644	5880	18250	12462
381633	20074	3675	22800	15276
398667	11285	4900	14400	27336
364594	22728	3542	18000	23919
300271	19309	2940	27600	25326
316767	24199	4655	16800	14673
344802	21885	3920	15600	14070
335111	17952	3920	19650	27738
407030	20205	4900	18700	25929
381727	22732	2450	28798	23718
367977	25818	4165	15850	19698
548072	63732	2450	17400	41004
383300	17167	3430	25650	17688
349017	22453	1960	14400	16080
372171	19927	3430	13425	34371
330944	21053	1960	25675	32964
397476	18066	3920	22800	29949
340178	21293	1715	13450	28944
300271	20544	735	21625	21909
317305	19244	2205	23050	32160
400328	21414	2450	14650	30753
286297	20663	2450	21350	22512
398936	22488	1470	13300	30954
384288	14992	3920	9600	32160
306678	15974	735	16950	55074
383932	30423	2695	16800	27336
295066	19709	735	16200	35577
301077	13464	4165	20400	36783
253913	17370	2450	14650	23316
380346	21049	1225	22925	25326
314803	15227	3185	4950	28341
391723	17729	2450	24675	18291

274736	16496	1960	8175	6231
407099	25820	3920	17916	24120
240462	27806	1960	13200	14271
358111	25533	490	12075	22311
294899	19966	2695	24000	17286
312784	22862	1960	24350	11658
271109	25996	4410	14400	27537
270771	23740	1715	14250	15276
230289	15596	2205	13200	16482
342062	27376	1715	18000	14472
201535	15622	2450	8925	2613
290933	24872	2205	15850	16188
204831	17765	2205	12000	16281
264653	31694	3185	13200	1206
220089	28985	1960	17000	11859
284971	24409	2695	10800	5829
298722	16830	2205	10800	14472
279201	25241	4410	19800	5025
290364	23505	2552	21600	15477
252761	23984	2205	10800	9246
194997	22955	2450	17600	8241
312266	22253	1960	15650	8442
257273	25432	3675	21600	7437
210574	11397	1470	19200	9045
165847	25128	2205	8400	11055
287081	21433	2450	12150	8844
244155	27422	6125	15600	12462
258600	26918	5390	18000	5025
201017	18796	7595	16000	13668
243786	25806	5390	13750	13869
260726	30500	6370	14400	9045
228957	34848	7350	13950	4221
226055	11405	6125	14050	5427
293123	33184	8085	14475	2010
163112	20362	8085	8400	1407

Los siguientes datos temporales son desde Enero del 2014 hasta Diciembre del 2019.

seriesPV7053	seriesPV1119	seriesgasolina	seriesCV2301	seriesCV2320
0	0	25728	0	0
0	0	11256	0	0
0	0	20502	0	0

0	0	30753	0	0
0	0	29346	0	0
0	0	12462	0	0
0	0	15276	0	0
0	0	27336	0	0
0	0	23919	0	0
0	0	25326	0	0
0	0	14673	0	0
0	0	14070	0	0
0	0	27738	0	0
0	0	25929	0	0
0	0	23718	0	0
0	0	19698	0	0
0	0	41004	0	0
0	0	17688	0	0
0	0	16080	0	0
0	0	34371	0	0
0	0	32964	28140	0
0	0	29949	28140	0
0	0	28944	22110	0
0	0	21909	16080	0
0	0	32160	16080	0
0	0	30753	18090	0
0	0	22512	19698	0
0	0	30954	22110	0
0	0	32160	24321	0
0	0	55074	30150	0
0	0	27336	21105	0
0	0	35577	24321	0
0	0	36783	22311	0
0	0	23316	30552	0
0	0	25326	85060	0
0	0	28341	53391	0
13800	0	32091	44507	25610
8200	0	14431	30398	13790
5800	0	29920	74074	2561
10400	0	24671	17286	24822
11000	0	33311	62190	11820
13600	0	30886	66373	122
10800	0	22458	36590	23640
18400	0	45937	39462	19700
24600	0	39876	37693	25610

10400	0	26882	19899	21670
17800	0	32272	87108	9850
16000	0	18613	3980	9850
18000	0	34188	26360	28565
11200	0	27481	34638	13396
15200	0	16406	85974	0
9600	0	21459	37788	0
6400	11940	24169	78101	25610
14000	11940	40412	49688	26595
11600	11940	28565	33674	29550
10000	11940	37417	59682	27749
12200	0	21446	60182	35460
3600	0	11841	44978	12805
2800	13134	24376	56865	33490
4200	1990	13627	72536	30535
4400	0	13445	47382	31520
2000	9950	23005	9648	8668
12400	9950	31194	75330	35854
5000	14925	32387	44765	33490
13000	9950	27975	32075	47280
7400	3980	25048	31241	11820
11400	3980	29249	34150	42749
7600	7960	24605	18135	20488
6400	4577	15198	34258	21670
6131	6965	18523	13869	42355
15600	8756	26366	46574	86483
3400	6965	11772	44155	18518

Los siguientes datos temporales son desde Enero del 2014 hasta Diciembre del 2019.

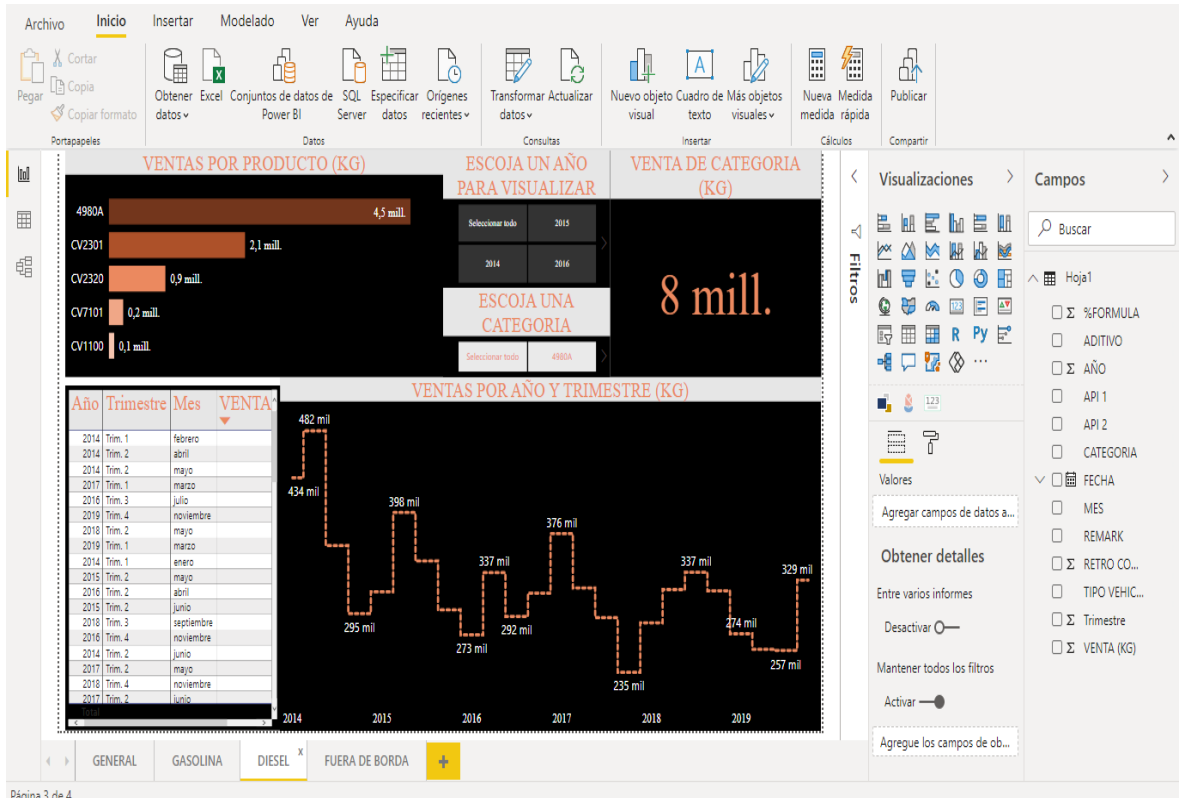
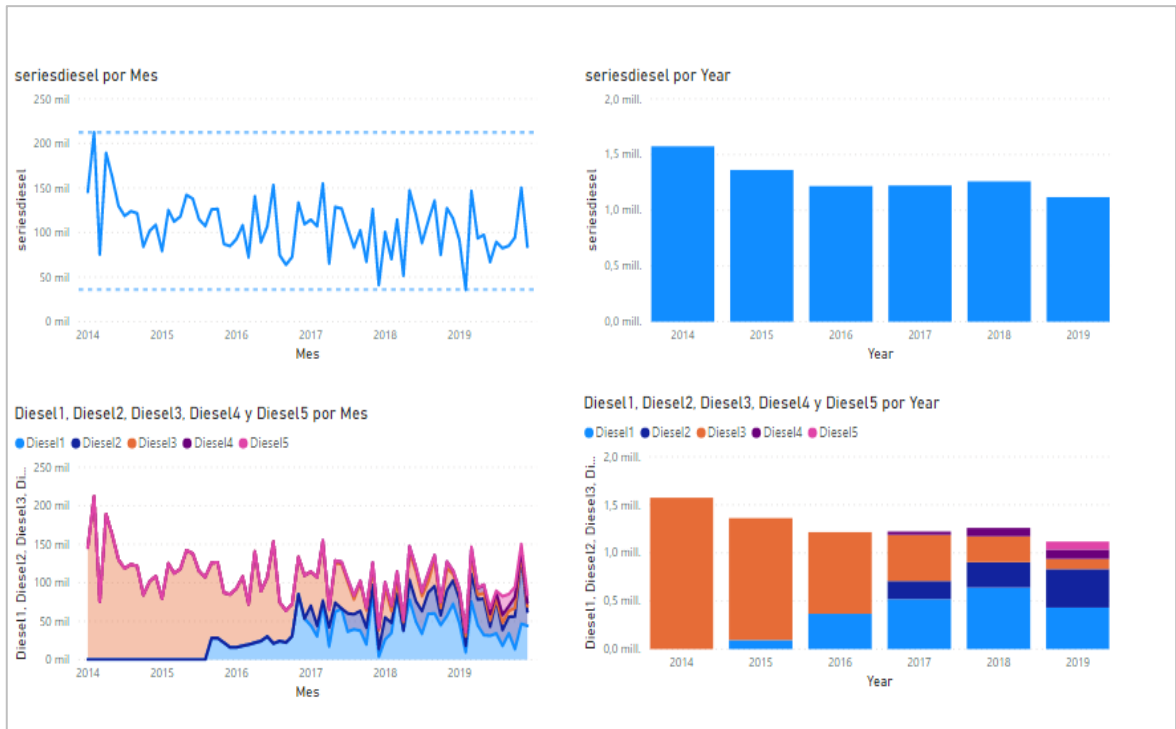
series4980A	seriesCV7101	seriesCV1100	seriesdiesel
145757	0	0	145757
212345	0	0	212345
75469	0	0	75469
189265	0	0	189265
162442	0	0	162442
129852	0	0	129852
118706	0	0	118706
123907	0	0	123907
121568	0	0	121568

84158	0	0	84158
101880	0	0	101880
108588	0	0	108588
79386	0	0	79386
125160	0	0	125160
112414	0	0	112414
117982	0	0	117982
142144	0	0	142144
137926	0	0	137926
115244	0	0	115244
107296	0	0	107296
97736	0	0	125876
98334	0	0	126474
65000	0	0	87110
68798	0	0	84878
76458	0	0	92538
90014	0	0	108104
52595	0	0	72293
118508	0	0	140618
64774	0	0	89095
76674	0	0	106824
132217	0	0	153322
50278	0	0	74599
41580	0	0	63891
42069	0	0	72621
48448	0	0	133508
56096	0	0	109487
44280	0	0	114397
62964	0	0	107152
75532	2743	0	154910
22770	422	0	65300
51304	3376	0	128690
57961	2532	0	126988
40830	3798	0	104858
19206	4853	0	83221
36816	2321	0	102440
22572	3165	0	67306
21186	8018	0	126162
24572	2743	0	41145
38424	7174	0	100523
15048	7174	0	70256
15048	13504	0	114526

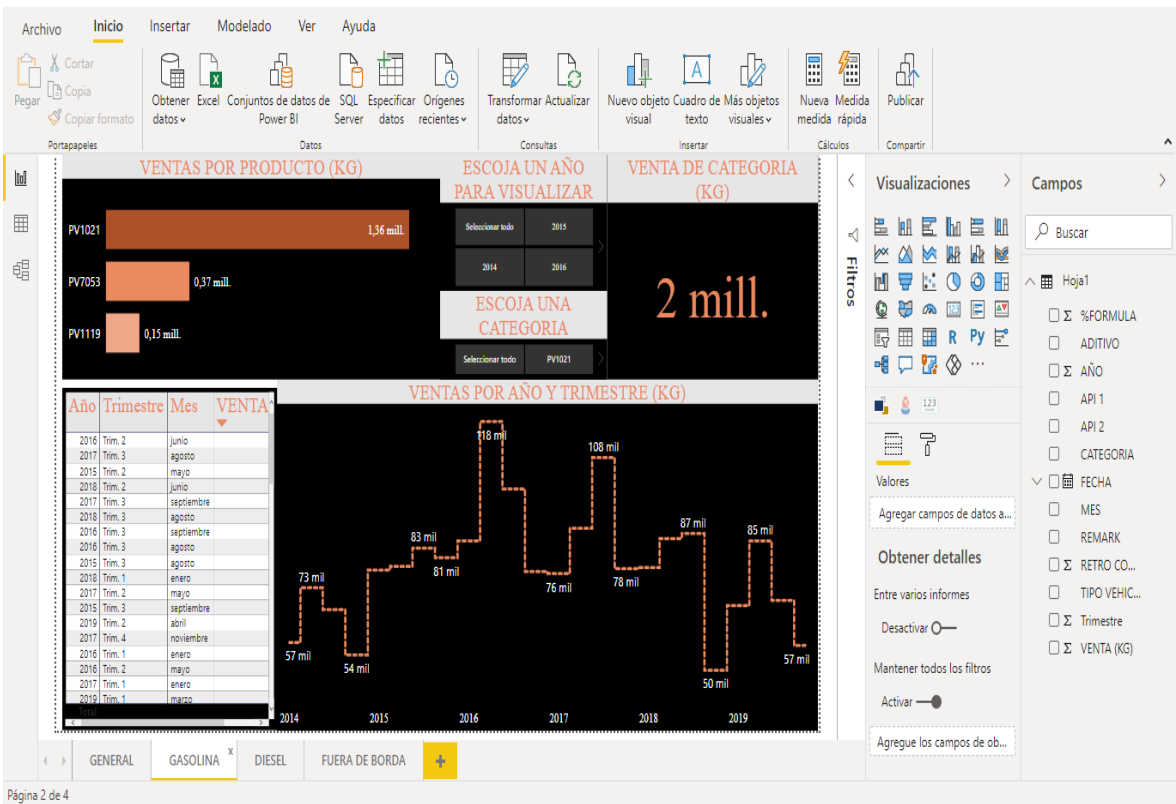
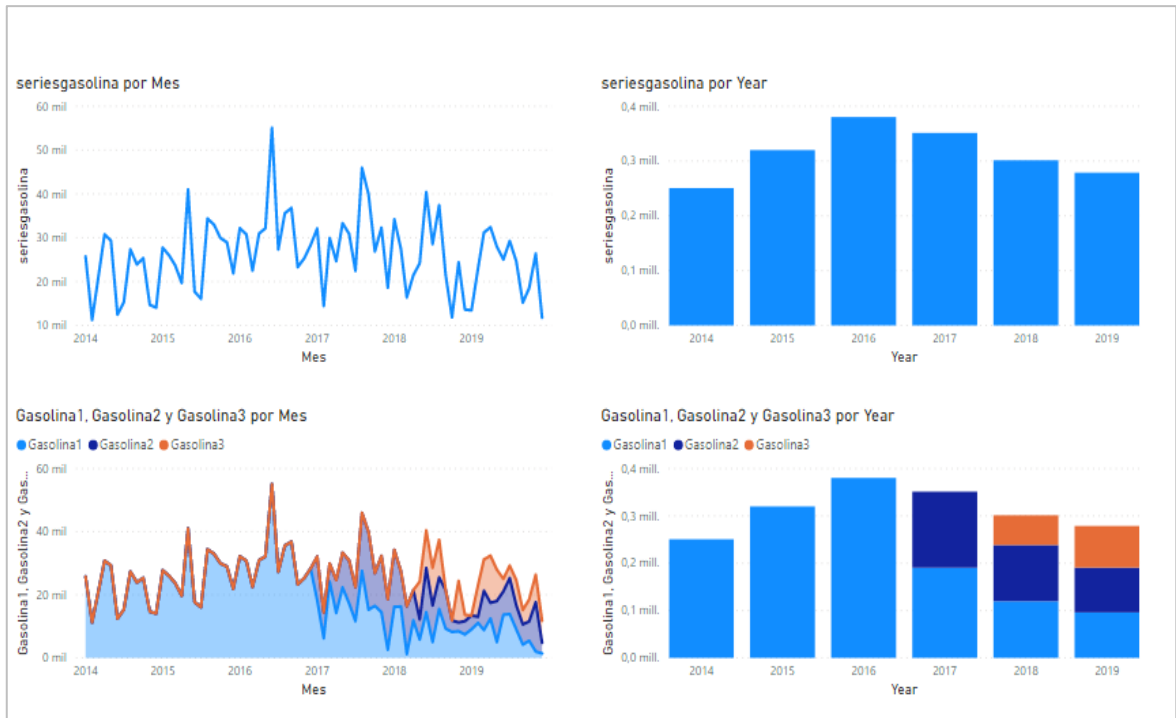
11088	2827	0	51703
35058	8440	0	147209
38770	5275	0	120328
19404	5697	0	88325
13464	12027	0	112922
35402	4642	0	135686
9900	7385	0	75068
28338	8651	0	127344
8316	4220	0	115607
8910	3798	0	91610
12474	5064	0	35854
23320	10128	1950	146582
5940	8651	780	93626
7128	10761	195	97439
7920	8651	7410	67042
2376	8018	2145	89438
9504	10128	24180	82435
7326	6330	15600	85184
11088	14137	13065	94514
3168	6752	7215	150192
6534	4853	10140	84200

7.2. Visualizaciones en Power Bi

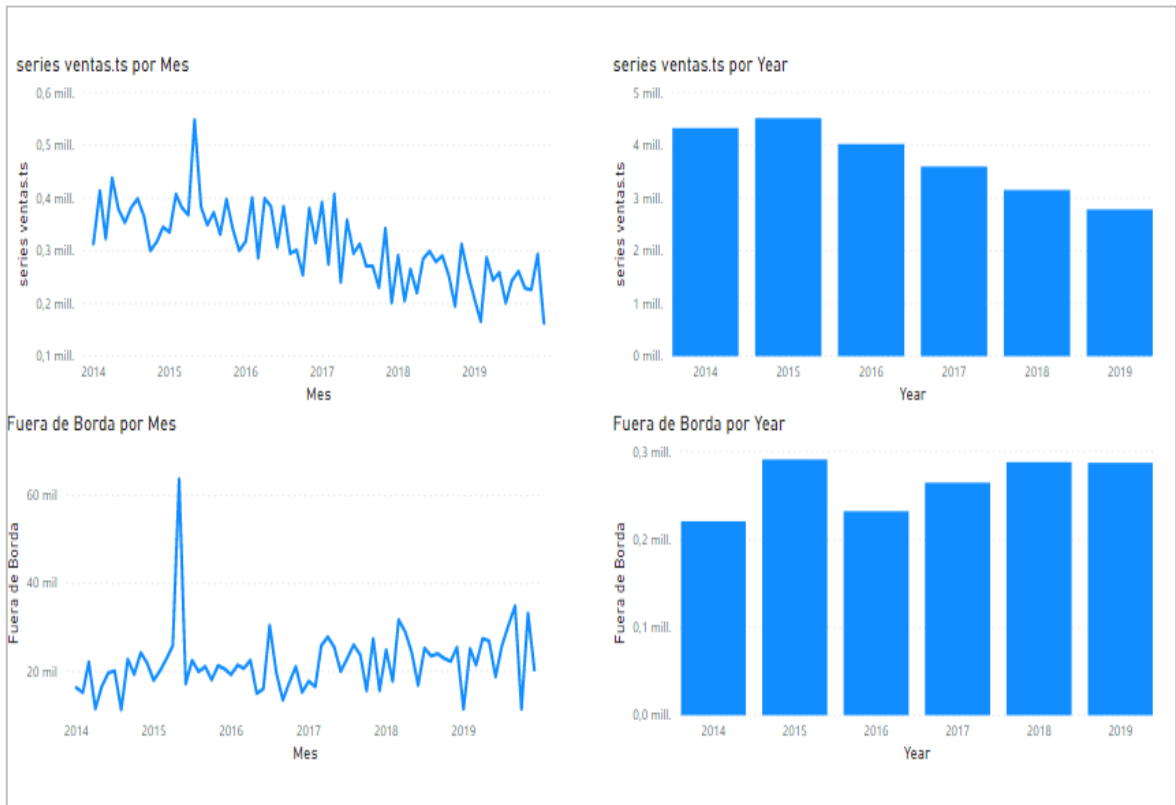
7.2.1. Diesel



7.2.2. Gasolina



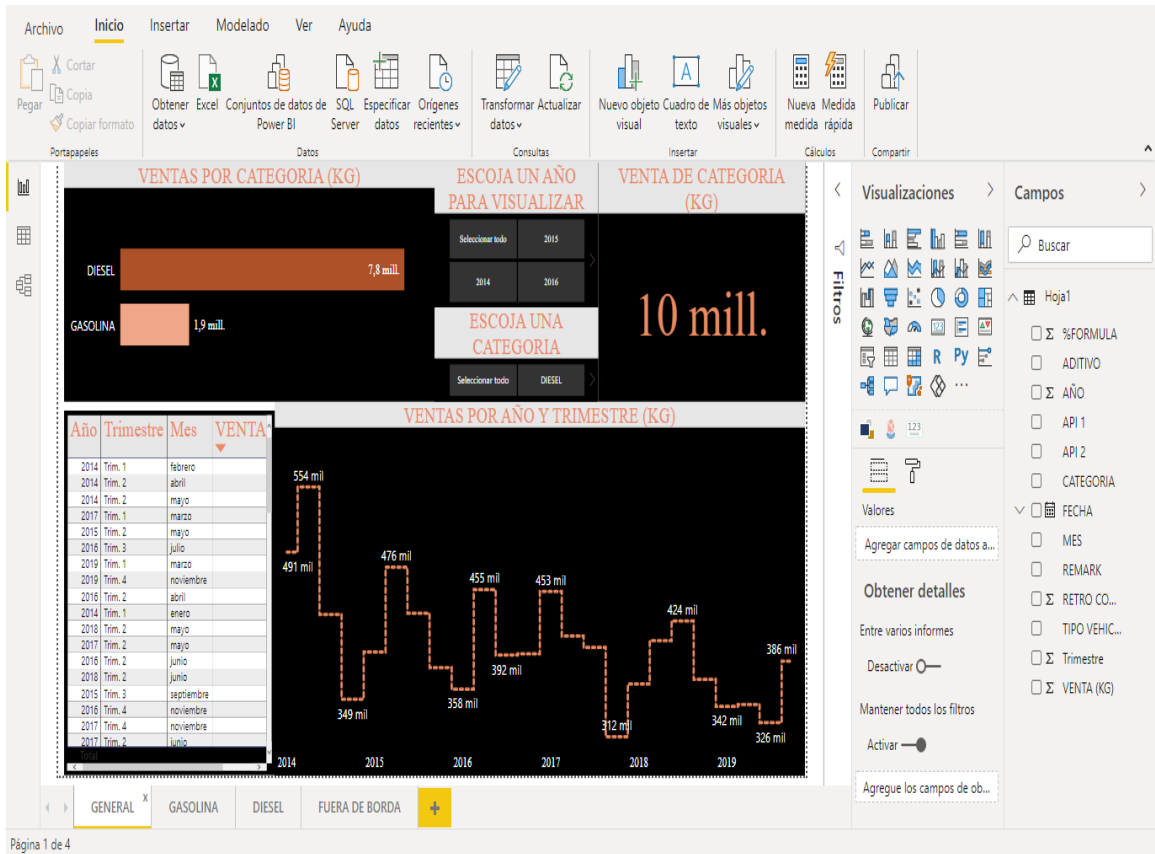
7.2.3. Totales y Fuera de Borda



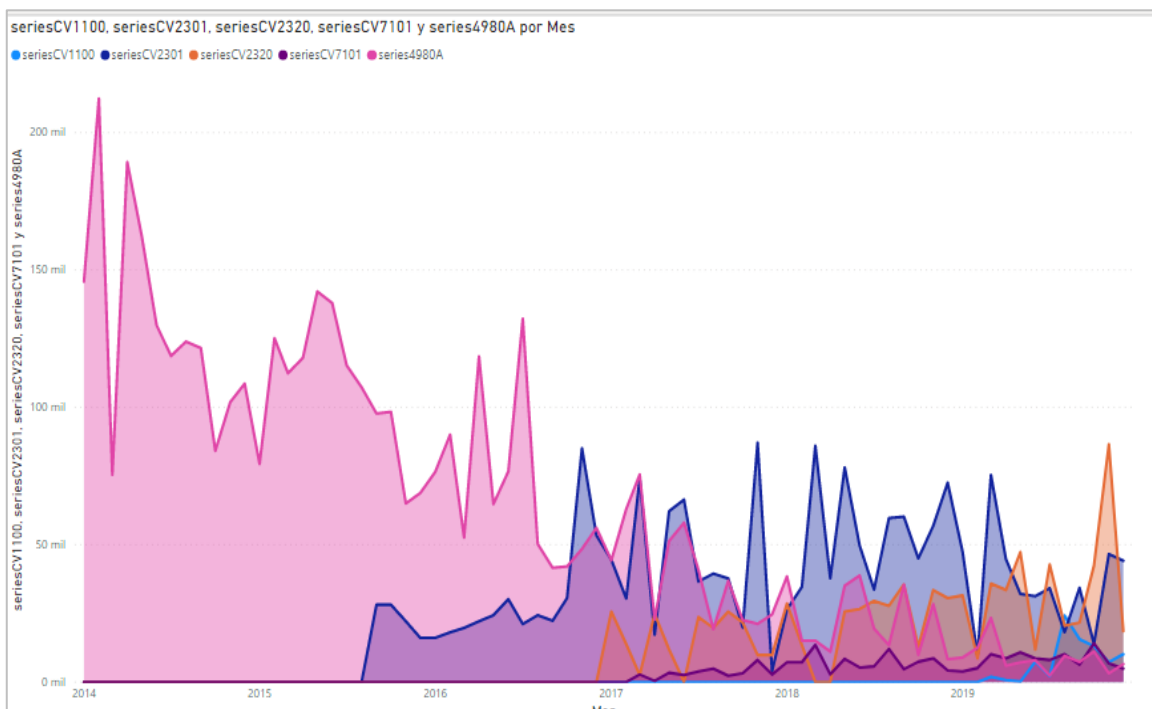
The screenshot shows the Microsoft Power BI interface with a dashboard titled 'Ventas por Mes y Año'. The dashboard includes:

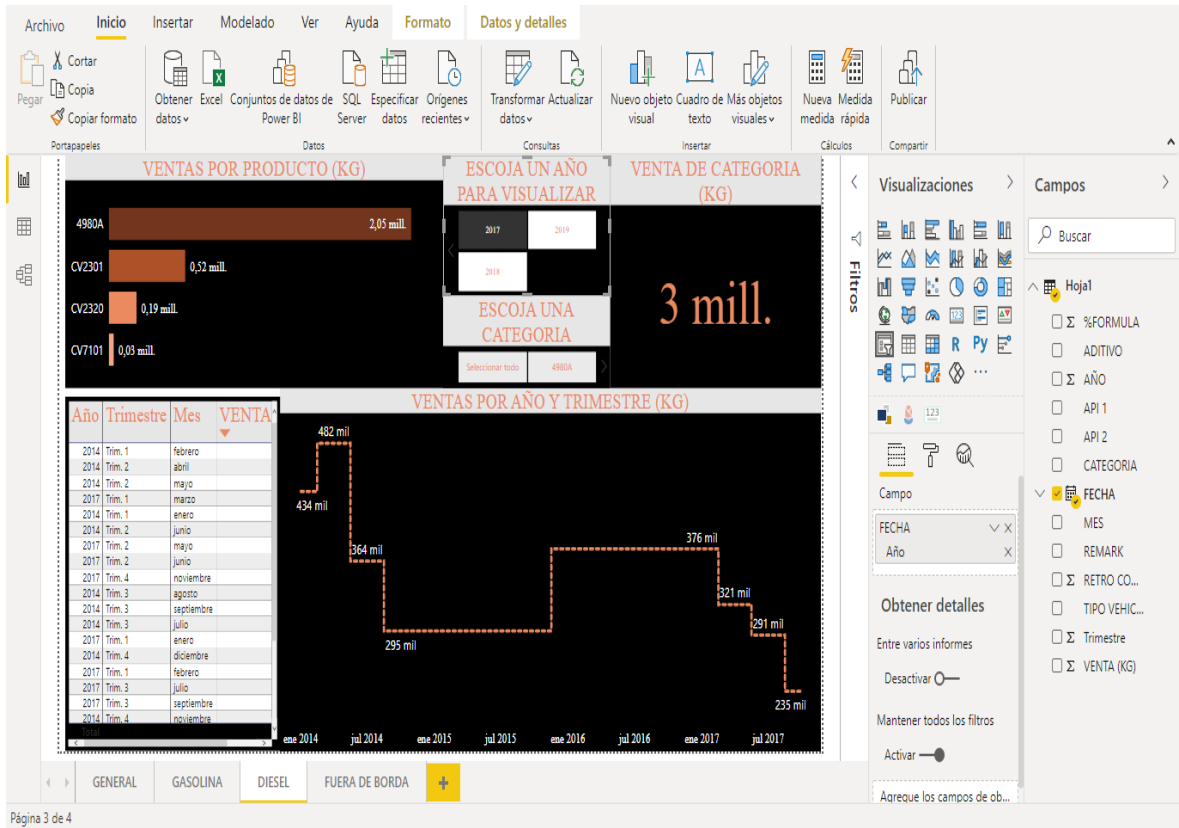
- Table:** A table showing sales by year from 2014 to 2020. The total for 2020 is 24 mill.
- Ventas por Mes y Año:** A stacked area chart showing monthly sales from May to December for each year from 2014 to 2020.
- Participación Por Producto:** A pie chart showing the distribution of sales by product: Gasolina (68.30%), Fuera de Borda (16.8%), and Diesel (13.84%).
- Ventas Por Productos:** A grouped bar chart showing sales for Diesel, Fuera de Borda, and Gasolina from 2014 to 2020.

The interface also shows the ribbon with various tools like 'Inicio', 'Insertar', and 'Modelado', and a right-hand pane with 'Filtros', 'Visualizaciones', and 'Campos'.



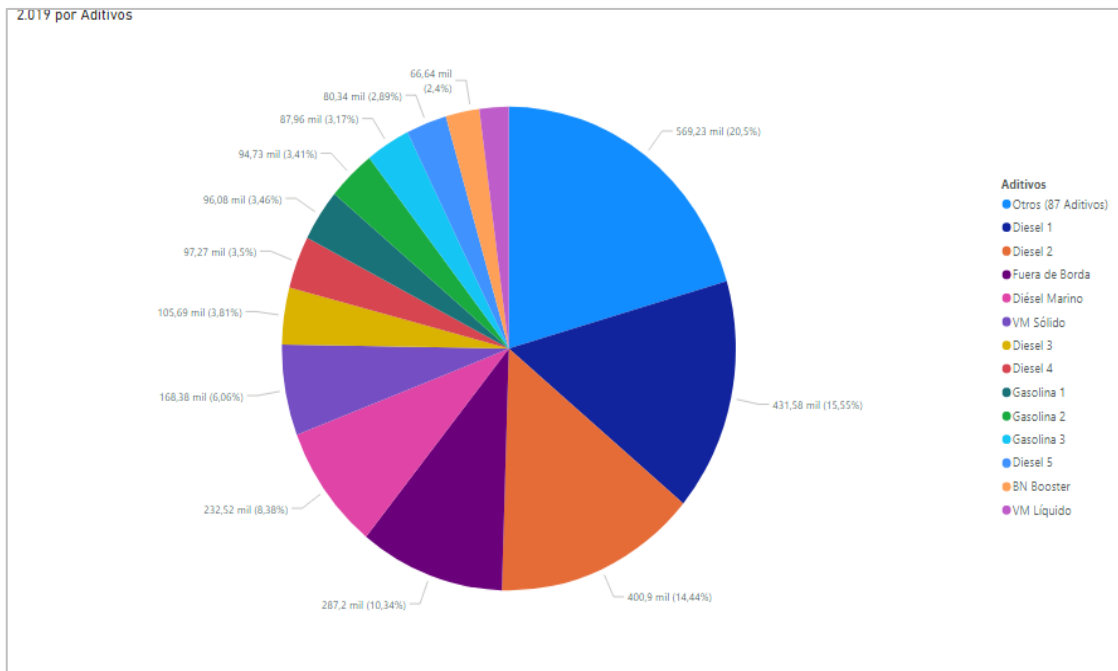
7.2.4. Diesel por Segmento





Página 3 de 4

7.2.5. Segmentación



7.3. Modelos estudiados

7.3.1. Serie Temporal: Ventas totales

7.3.1.1. Modelo AR(p)

7.3.1.1.1. Modelo AR(1)

Para efectos del presente estudio comparativo, ejecutaremos este primer modelo AR(1), teniendo en cuenta únicamente el valor p extraído de la interpretación de la gráfica PACF sin incluir su diferencia (d) o detalles de su componente estacional.

Tabla No. 7.1. Resumen Estadístico - Orden (1,0,0), Serie Temporal Ventas Totales					
Orden: (1,0,0)					
Coefficients:					
	ar1	intercept			
	0.2802	326.2849			
s.e.	0.1239	11.1548			
sigma^2 estimated as 3916: log likelihood = -333.36,					
aic = 672.72					
Training set error measures:					
	ME	RMSE	MAE	MPE	MAPE
MASE					
Training set	0.07100217	62.57408	48.8834	-3.855114	15.90012
					0.7551064
		ACF1			
Training set		-0.1433247			
Elaborado por: Autor					

Con el fin de verificar el nivel de significancia de los coeficientes obtenidos aplicamos la siguiente función y realizamos la comparación a través del estadístico Z:

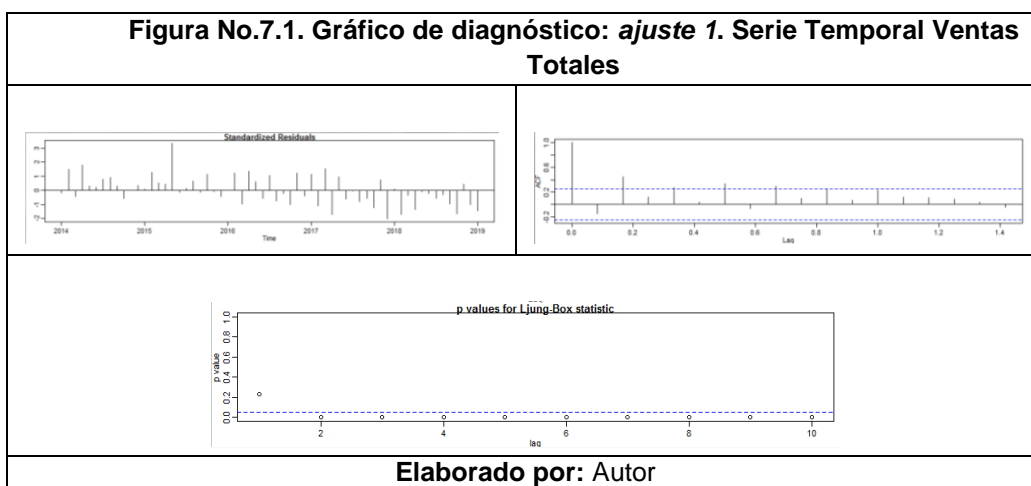
Tabla No. 7.2. Estadístico Z - Orden (1,0,0), Serie Temporal Ventas Totales				
z test of coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
ar1	0.28015	0.12387	2.2617	0.02372 *
intercept	326.28486	11.15479	29.2506	< 2e-16 ***

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.' 0.1 ' ' 1
Elaborado por: Autor				

Así también, constatamos que la ejecución del modelo se llevó a cabo de forma correcta ejecutando la siguiente prueba, en la que observamos que al obtener un p-valor superior a 0.05 no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.3. Prueba Box-Ljung- Orden (1,0,0), Serie Temporal Ventas Totales
Box-Ljung test
data: residuals(ajuste1) X-squared = 1.2952, df = 1, p-value = 0.2551
Elaborado por: Autor

Es recomendable llevar a cabo un diagnóstico general de los modelos de esta índole con el objetivo de que puedan ser comparados entre sí y pueda elegirse el modelo que mejor se ajuste al comportamiento de nuestros datos.



Como podemos notar en la figura anterior los valores de los estadísticos obtenidos a lo largo de la prueba Ljung-Box se encuentran en su mayoría por debajo de 0.05 por lo que podríamos considerar un segundo modelo AR para mejorar el comportamiento en relación al ruido blanco.

7.3.1.1.2. Modelo AR(2)

Plantaremos también un segundo modelo AR de orden 2, debido a que en el gráfico de autocorrelación parcial (PACF), observamos que aún en el segundo rezago dicho valor no es considerablemente próximo a cero.

Tabla No. 7.4. Resumen Estadístico - Orden (2,0,0), Serie Temporal Ventas Totales
Orden: (2,0,0)
Coefficients:
ar1 ar2 intercept
0.1242 0.5384 326.0964
s.e. 0.1069 0.1083 19.0415
sigma^2 estimated as 2768: log likelihood = -323.29,
aic = 654.58
Training set error measures:
MASE ME RMSE MAE MPE MAPE

Training set	-0.8063899	52.60915	41.99137	-2.90091	13.28284	0.6486445
						ACF1
Training set	-0.07242284					
Elaborado por: Autor						

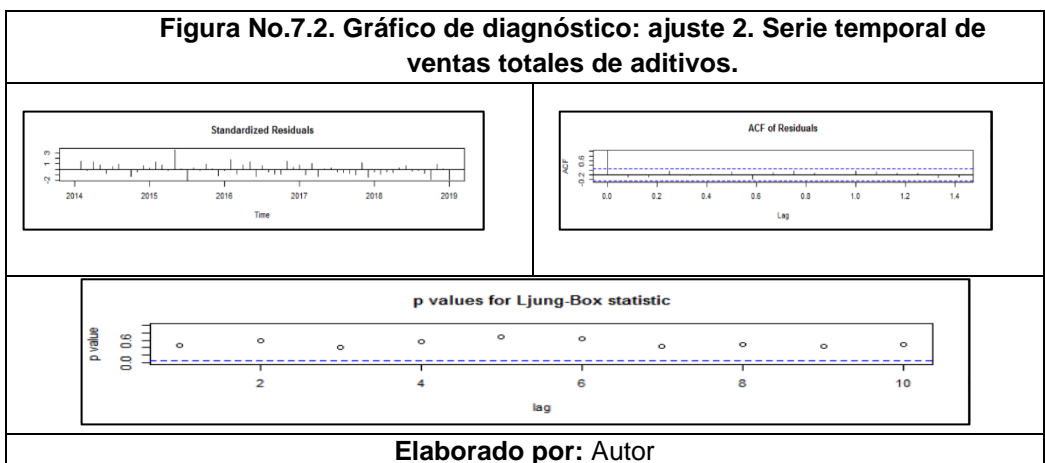
A continuación podemos observar que el coeficiente ar2 sería el de mayor significancia dentro del modelo planteado:

Tabla No. 7.5. Estadístico Z - Orden (2,0,0), Serie Temporal Ventas Totales						
z test of coefficients:						
	Estimate	Std. Error	z value	Pr(> z)		
ar1	0.12419	0.10692	1.1615	0.2454		
ar2	0.53842	0.10826	4.9733	6.582e-07	***	
intercept	326.09643	19.04150	17.1256	< 2.2e-16	***	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
Elaborado por: Autor						

Continuando el presente análisis, constatamos que la ejecución del modelo se llevó a cabo de forma correcta ejecutando la siguiente prueba, en la que observamos que al obtener un p-valor superior a 0.05 no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.6. Prueba Box-Ljung- Orden (2,0,0), Serie Temporal Ventas Totales	
Box-Ljung test	
data: residuals(ajuste2)	
X-squared = 0.56903, df = 1, p-value = 0.4506	
Elaborado por: Autor	



En efecto, este segundo modelo AR de orden 2, tiene un mejor comportamiento en relación al ruido blanco, podemos notar en la figura anterior que los valores de los estadísticos obtenidos a lo largo de la prueba Ljung-Box se encuentran en su mayoría por encima de 0.05.

7.3.1.2. Modelo ARIMA (p,d,q)

Una vez que hemos definido el coeficiente que mejor representa a la parte de regresión automática (AR) de nuestro modelo, procederemos a agregar la primera diferencia aplicada durante es análisis exploratorio de datos, siendo $d=1$ y el coeficiente de media móvil (MA); $q=1$, considerando que el gráfico de ACF de la figura 4.21 nos muestra un decaimiento exponencial. Por lo anteriormente expuesto plantearemos los siguientes modelos:

Tabla No. 7.7. Resumen Estadístico - Orden (2,1,1), Serie Temporal Ventas Totales						
Orden: (2,1,1)						
Coefficients:						
	ar1	ar2	ma1			
	-0.1641	0.2653	-0.8205			
s.e.	0.1699	0.1627	0.1008			
sigma^2 estimated as 2658: log likelihood = -316.94,						
aic = 641.89						
Training set error measures:						
	MASE	ME	RMSE	MAE	MPE	MAPE
Training set	-7.07034	51.12078	41.21277	-4.513409	13.23432	0.6
	366174					
		ACF1				
Training set	-0.02178556					
Elaborado por: Autor						

Tabla No. 7.8. Estadístico Z - Orden (2,1,1), Serie Temporal Ventas Totales				
z test of coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.16407	0.16993	-0.9655	0.3343
ar2	0.26531	0.16274	1.6302	0.1031
ma1	-0.82054	0.10075	-8.1441	3.821e-16 ***

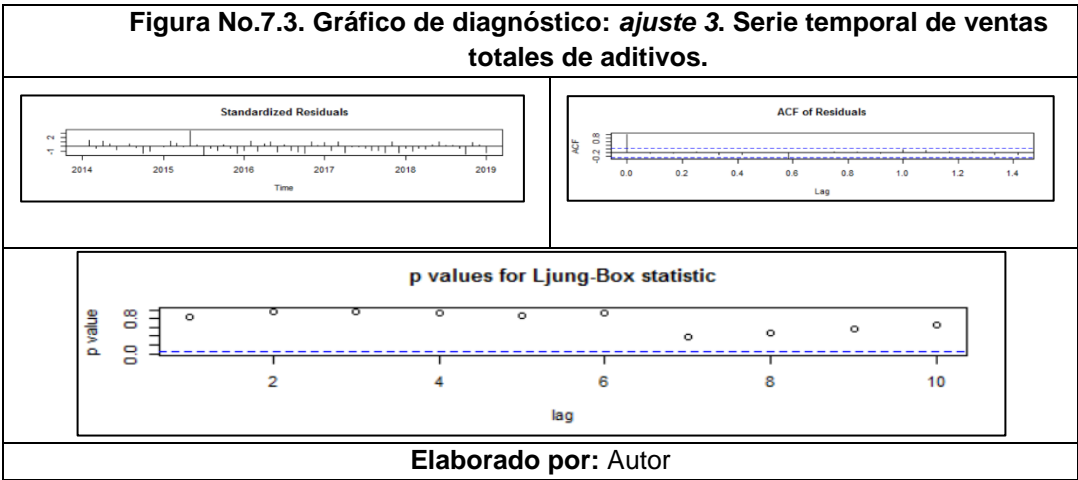
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Elaborado por: Autor				

Como fue ejecutado con los modelos anteriores, llevamos a cabo la prueba de Ljung-Box en la que verificamos que hemos realizado el modelo de forma adecuada; puesto que, obtenemos un p-valor de 0.86 superior a 0.05 y no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.9. Prueba Box-Ljung- Orden (2,1,1), Serie Temporal Ventas Totales
Box-Ljung test
data: residuals(ajuste3)

X-squared = 0.029925, df = 1, p-value = 0.8627

Elaborado por: Autor



7.3.2. Serie Temporal: Fuera de Borda

7.3.2.1. Modelo AR(p)

7.3.2.1.1. Modelo AR(1)

Para efectos del presente estudio comparativo, ejecutaremos este primer modelo AR(1), teniendo en cuenta únicamente el valor p extraído de la interpretación de la gráfica PACF sin incluir su diferencia (d) o detalles de su componente estacional.

Tabla No. 7.10. Resumen Estadístico - Orden (1,0,0), Serie Temporal Aditivos pa ra Fuera de Borda					
Orden: (1,0,0)					
Coefficients:					
	ar1	intercept			
	0.0393	21.6004			
s.e.	0.1289	0.9422			
sigma ² estimated as 49.22: log likelihood = -202.03,					
aic = 410.05					
Training set error measures:					
	ME	RMSE	MAE	MPE	MAPE
MASE					
Training set	0.003565227	7.015778	4.26677	-7.322126	20.32624
	.7129065				0
		ACF1			
Training set		-0.002226676			
Elaborado por: Autor					

Con el fin de verificar el nivel de significancia de los coeficientes obtenidos aplicamos la siguiente función y realizamos la comparación a través del estadístico z. En esta prueba podemos ver que el coeficiente ar1 con un valor p de 0.7602,

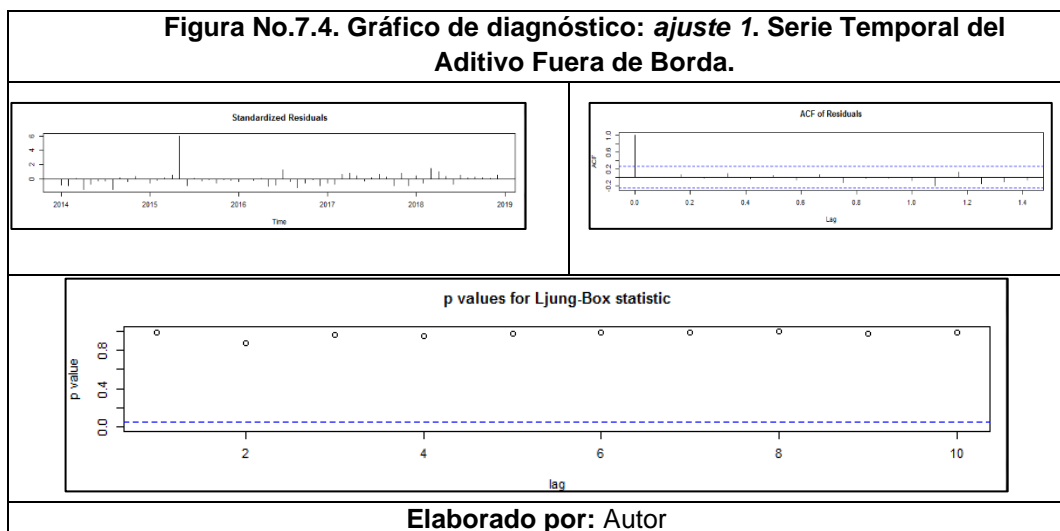
considerablemente superior a 0.05, no tiene mayor significancia respecto al presente modelo y nuestra variable de respuesta:

Tabla No. 7.11. Estadístico Z - Orden (1,0,0), Serie Temporal Aditivos para Fuera de Borda				
z test of coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
ar1	0.039324	0.128859	0.3052	0.7602
intercept	21.600408	0.942173	22.9262	<2e-16 ***
--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Elaborado por: Autor				

Así también, constatamos que la ejecución del modelo se llevó a cabo de forma correcta ejecutando la siguiente prueba, en la que observamos que al obtener un p-valor superior a 0.05 no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.12. Prueba Box-Ljung- Orden (1,0,0), Serie Temporal Aditivos para Fuera de Borda
Box-Ljung test
data: residuals(ajuste1) x-squared = 0.00031261, df = 1, p-value = 0.9859
Elaborado por: Autor

Es recomendable llevar a cabo un diagnóstico general de los modelos de esta índole con el objetivo de que puedan ser comparados entre sí y pueda elegirse el modelo que mejor se ajuste al comportamiento de nuestros datos.



Como podemos notar en la figura anterior los valores de los estadísticos obtenidos a lo largo de la prueba Ljung-Box se encuentran en su mayoría por encima de 0.05; no obstante el coeficiente ar1, no tiene mayor significancia respecto a nuestro

modelos por lo que podríamos considerar un segundo modelo AR para verificar si un tercer coeficiente mejoraría la representación de nuestra variable de respuesta.

7.3.2.1.2. Modelo AR(3)

Plantearémos también un segundo modelo AR de orden 3, debido a que en el gráfico de autocorrelación parcial (PACF), observamos que aún en el tercer rezago dicho valor no es considerablemente próximo a cero.

Tabla No. 7.13. Resumen Estadístico - Orden (3,0,0), Serie Temporal Aditivos para Fuera de Borda					
Orden: (3,0,0)					
Coefficients:					
	ar1	ar2	ar3	intercept	
	0.0389	0.0640	-0.0206	21.5939	
s.e.	0.1287	0.1285	0.1277	0.9831	
sigma ² estimated as 48.99: log likelihood = -201.89,					
aic = 413.78					
Training set error measures:					
	MASE	ME	RMSE	MAE	MPE MAPE
Training set	0.01255081	6.999497	4.208256	-7.235595	20.0312 0.7031298
		ACF1			
Training set	0.001208786				
Elaborado por: Autor					

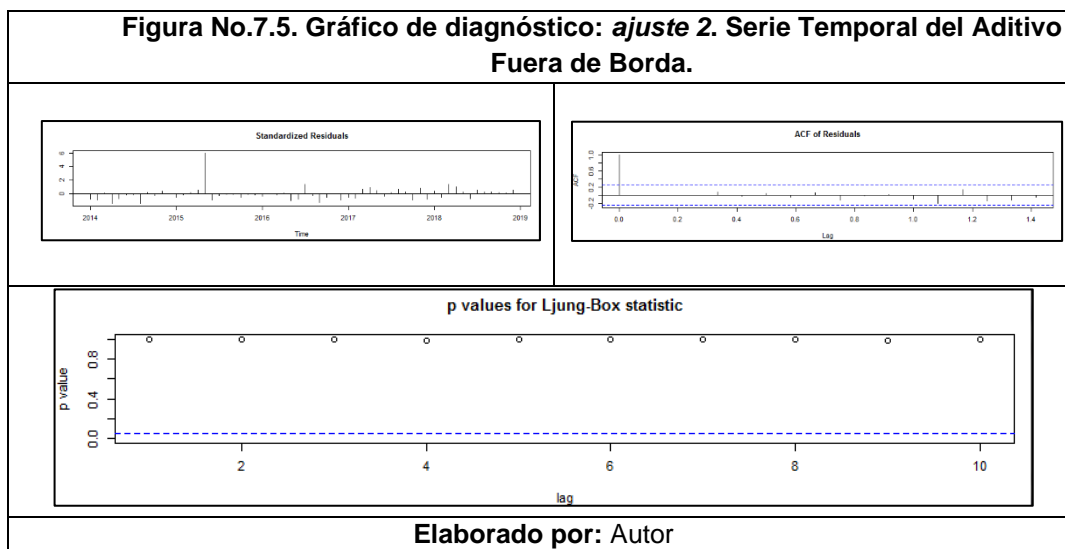
A continuación podemos observar que los coeficientes ar2 y ar3 tampoco serían de mayor significancia dentro del modelo planteado:

Tabla No. 7.14. Estadístico Z - Orden (3,0,0), Serie Temporal Aditivos para Fuera de Borda				
z test of coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
ar1	0.038897	0.128743	0.3021	0.7626
ar2	0.064046	0.128491	0.4984	0.6182
ar3	-0.020567	0.127723	-0.1610	0.8721
intercept	21.593942	0.983127	21.9646	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Elaborado por: Autor				

Continuando el presente análisis, constatamos que la ejecución del modelo se llevó a cabo de forma correcta ejecutando la siguiente prueba, en la que observamos que al obtener un p-valor superior a 0.05 no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.15. Prueba Box-Ljung- Orden (3,0,0), Serie Temporal Aditivos para Fuera de Borda
Box-Ljung test data: residuals(ajuste2) X-squared = 9.2128e-05, df = 1, p-value = 0.9923
Elaborado por: Autor



Ambos modelos AR(1) y AR(3) tienen un comportamiento adecuado de ruido blanco en base a los valores de los estadísticos obtenidos a lo largo de la prueba Ljung-Box se encuentran en su mayoría por encima de 0.05. No obstante, sus coeficientes no poseen una significancia considerable dentro del modelo y estos no deberían ser usados para representar el comportamiento de nuestra variable de respuesta.

7.3.2.2. Modelo ARIMA (p,d,q)

Procedemos entonces con el ingreso de la primera diferencia aplicada durante es análisis exploratorio de datos, siendo $d=1$ y el coeficiente de media móvil (MA); $q=1$, considerando que el gráfico de ACF de la figura 4.37 nos muestra un decaimiento exponencial. Por lo anteriormente expuesto plantearemos los siguientes modelos:

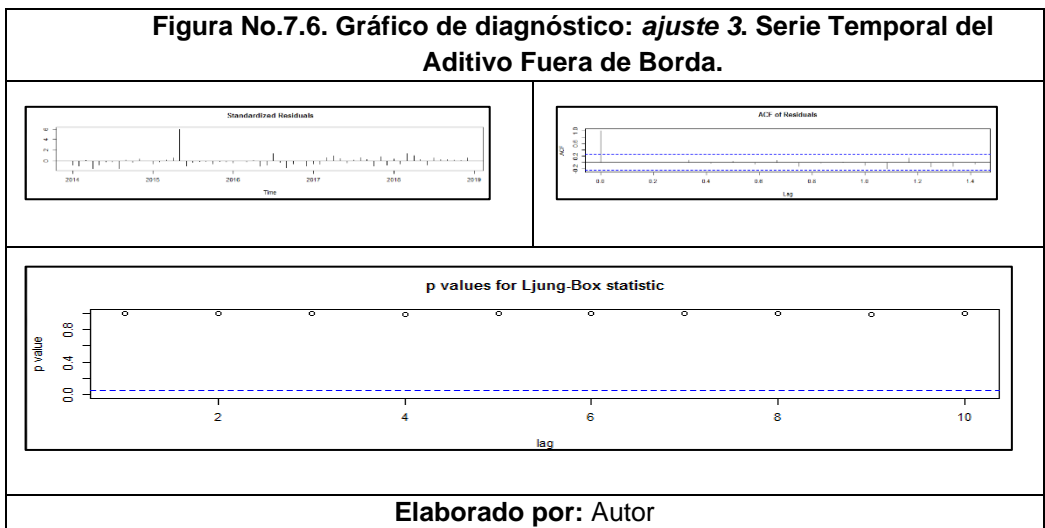
Tabla No. 7.16. Resumen Estadístico - Orden (0,1,1), Serie Temporal Aditivos para Fuera de Borda
Orden: (0,1,1) Coefficients: ma1 -0.9605 s.e. 0.0948 sigma^2 estimated as 51.39: log likelihood = -201.21, aic = 406.42 Training set error measures: ME RMSE MAE MPE MAPE MASE

Training set 1.330024 7.10895 4.222411 -0.08412284 18.62298 0.7054949
ACF1
Training set -0.001244824
Elaborado por: Autor

Tabla No. 7.17. Estadístico Z - Orden (0,1,1), Serie Temporal Aditivos para Fuera de Borda
z test of coefficients:
<pre> Estimate Std. Error z value Pr(> z) ma1 -0.960531 0.094823 -10.13 < 2.2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 </pre>
Elaborado por: Autor

Como fue ejecutado con los modelos anteriores, llevamos a cabo la prueba de Ljung-Box en la que verificamos que hemos realizado el modelo de forma adecuada; puesto que, obtenemos un p-valor de 0.99 superior a 0.05 y no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.18. Prueba Box-Ljung- Orden (0,1,1), Serie Temporal Aditivos para Fuera de Borda
Box-Ljung test
<pre> data: residuals(ajuste3) x-squared = 9.7703e-05, df = 1, p-value = 0.9921 </pre>
Elaborado por: Autor



7.3.3. GASOLINA

7.3.3.1. Modelo AR(p)

7.3.3.1.1. Modelo AR(1)

A continuación ejecutaremos este primer modelo AR(1), teniendo en cuenta únicamente el valor p extraído de la interpretación de la gráfica PACF sin incluir su diferencia (d) o detalles de su componente estacional.

Tabla No. 7.19. Resumen Estadístico - Orden (1,0,0), Serie Temporal Aditivos para Gasolina					
Orden: (1,0,0)					
Coefficients:					
	ar1	intercept			
	0.1341	26.6916			
s.e.	0.1293	1.2767			
sigma ² estimated as 73.63: log likelihood = -214.12,					
aic = 434.23					
Training set error measures:					
	ME	RMSE	MAE	MPE	MAPE
Training set	0.002298405	8.580768	6.712602	-12.2123	30.20617
	MASE	ACF1			
Training set	0.7331574	-0.0001510473			
Elaborado por: Autor					

Ahora verificamos el nivel de significancia de los coeficientes obtenidos aplicamos la siguiente función y realizamos la comparación a través del estadístico z. En esta prueba podemos ver que el coeficiente ar1 con un valor p de 0.2998, considerablemente superior a 0.05, no tiene mayor significancia respecto al presente modelo y nuestra variable de respuesta:

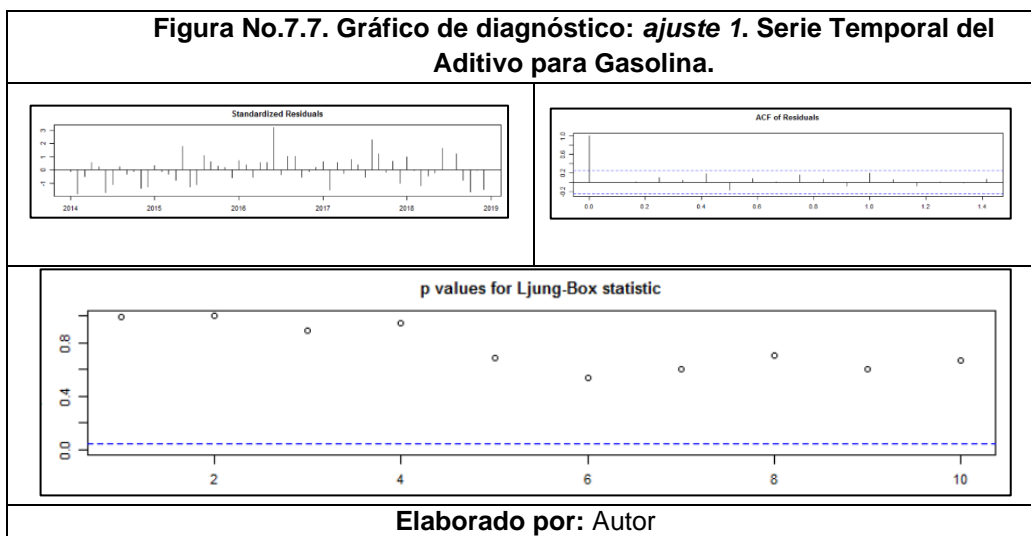
Tabla No. 7.20. Estadístico Z - Orden (1,0,0), Serie Temporal Aditivos para Gasolina					
z test of coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
ar1	0.13410	0.12934	1.0368	0.2998	
intercept	26.69156	1.27667	20.9071	<2e-16	***

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ' 1
Elaborado por: Autor					

Para revisar si la ejecución del modelo se llevó a cabo de forma correcta ejecutando la siguiente prueba, en la que observamos que al obtener un p-valor superior a 0.05 no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.21. Prueba Box-Ljung- Orden (1,0,0), Serie Temporal Aditivos para Gasolina
Box-Ljung test
data: residuals(ajuste1) X-squared = 1.4385e-06, df = 1, p-value = 0.999
Elaborado por: Autor

A continuación se detalla un diagnóstico general del modelo generado:



Como podemos notar en la figura anterior los valores de los estadísticos obtenidos a lo largo de la prueba Ljung-Box se encuentran en su mayoría por encima de 0.05; no obstante el coeficiente ar1, no tiene mayor significancia respecto a nuestro modelos por lo que podríamos considerar un segundo modelo AR para verificar si un tercer coeficiente mejoraría la representación de nuestra variable de respuesta.

7.3.3.1.2. Modelo AR(2)

Plantaremos también un segundo modelo AR de orden 2, debido a que en el gráfico de autocorrelación parcial (PACF), observamos que aún en rezagos posteriores dicho valor no es considerablemente próximo a cero.

Tabla No. 7.22. Resumen Estadístico - Orden (2,0,0), Serie Temporal Aditivos para Gasolina
Orden: (2,0,0)
Coefficients:
ar1 ar2 intercept
0.1317 0.0189 26.6798
s.e. 0.1304 0.1333 1.3031
sigma^2 estimated as 73.6: log likelihood = -214.11,
aic = 436.21

Training set error measures:					
	ME	RMSE	MAE	MPE	MAPE
MASE					
Training set	0.007485087	8.579273	6.71647	-12.20045	30.23974
	.7335798				0
	ACF1				
Training set	0.0002904848				
Elaborado por: Autor					

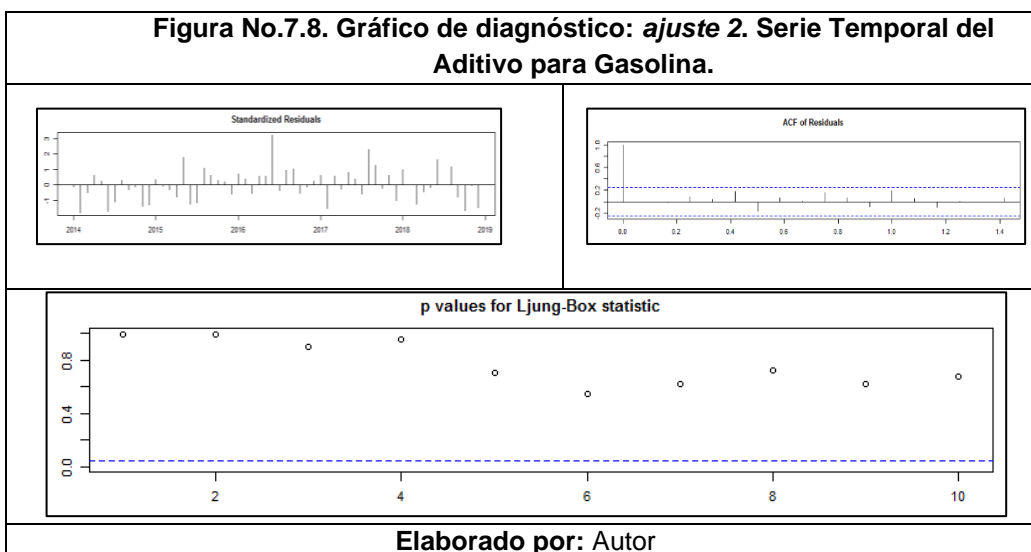
A continuación podemos observar que el coeficiente ar2 no tendría un nivel de significancia considerable respecto al modelo planteado:

Tabla No. 7.23. Estadístico Z - Orden (2,0,0), Serie Temporal Aditivos para Gasolina				
z test of coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
ar1	0.131706	0.130372	1.0102	0.3124
ar2	0.018919	0.133253	0.1420	0.8871
intercept	26.679794	1.303106	20.4740	<2e-16 ***

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.' 0.1 ' ' 1
Elaborado por: Autor				

Verificamos también que el modelo se llevó a cabo de forma correcta, obteniendo un p-valor superior a 0.05 no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.24. Prueba Box-Ljung- Orden (2,0,0), Serie Temporal Aditivos para Gasolina	
Box-Ljung test	
data: residuals(ajuste2)	
X-squared = 5.3203e-06, df = 1, p-value = 0.9982	
Elaborado por: Autor	



Ambos modelos AR(1) y AR(2) tienen un comportamiento adecuado de ruido blanco en base a los valores de los estadísticos obtenidos a lo largo de la prueba Ljung-Box se encuentran en su mayoría por encima de 0.05. No obstante, sus coeficientes no poseen una significancia considerable dentro del modelo y estos no deberían ser usados para representar el comportamiento de nuestra variable de respuesta.

7.3.3.2. Modelo ARIMA (p,d,q)

Para este modelo ingresamos la primera diferencia aplicada durante el análisis exploratorio de datos, siendo $d=1$ y el coeficiente de media móvil (MA); $q=1$, considerando que el gráfico de ACF de la figura anterior nos muestra un decaimiento exponencial característico de los modelos MA de orden 1. A continuación plantearemos los siguientes modelos:

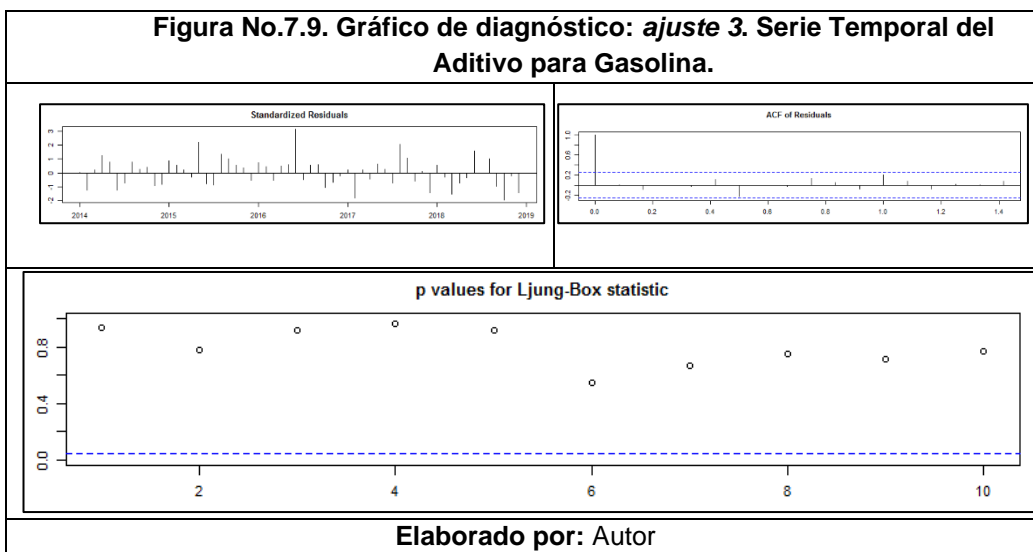
Tabla No. 7.25. Resumen Estadístico - Orden (0,1,1), Serie Temporal Aditivos para Gasolina						
Orden: (0,1,1)						
Coefficients:						
	ma1					
	-0.8634					
s.e.	0.0731					
sigma^2 estimated as 75.65: log likelihood = -212.02,						
aic = 428.04						
Training set error measures:						
	MASE	ME	RMSE	MAE	MPE	MAPE
Training set	0.2945652	8.624667	6.921802	-9.942894	30.30814	0.
	7560064					
		ACF1				
Training set	0.01039214					
Elaborado por: Autor						

Tabla No. 7.26. Estadístico Z - Orden (0,1,1), Serie Temporal Aditivos para Gasolina						
z test of coefficients:						
	Estimate	Std. Error	z value	Pr(> z)		
ma1	-0.863447	0.073051	-11.82	< 2.2e-16	***	

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '	1
Elaborado por: Autor						

Continuando el procedimiento, llevamos a cabo la prueba de Ljung-Box en la que verificamos que hemos realizado el modelo de forma adecuada; puesto que, obtenemos un p-valor de 0.93 superior a 0.05 y no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.27. Prueba Box-Ljung- Orden (0,1,1), Serie Temporal Aditivos para Gasolina
Box-Ljung test
data: residuals(ajuste3) X-squared = 0.0068093, df = 1, p-value = 0.9342
Elaborado por: Autor



7.4. DIESEL

7.4.1. Modelos ARIMA (p,d,q)

Para este modelo ingresamos la primera diferencia aplicada durante el análisis exploratorio de datos, siendo $d=1$ y el coeficiente de media móvil (MA); $q=0$, de esta forma buscamos observar el comportamiento autorregresivo (AR) de orden 2, con un $p=2$:

Tabla No. 7.28. Resumen Estadístico - Orden (2,1,0), Serie Temporal Aditivos para Diesel
Orden: (2,1,0)
Coefficients:
ar1 ar2
-0.9368 -0.4399
s.e. 0.1184 0.1284
sigma ² estimated as 974.8: log likelihood = -287.23,
aic = 580.47
Training set error measures:

	ME	RMSE	MAE	MPE	MAPE
MASE					
Training set	-1.817394	30.96076	25.17555	-9.056943	26.46358
	7154275				0.
	ACF1				
Training set	-0.1313941				

Elaborado por: Autor

Tabla No. 7.29. Estadístico Z - Orden (2,1,0), Serie Temporal Aditivos para Diesel					
z test of coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
ar1	-0.93677	0.11839	-7.9126	2.521e-15	***
ar2	-0.43994	0.12842	-3.4257	0.0006131	***

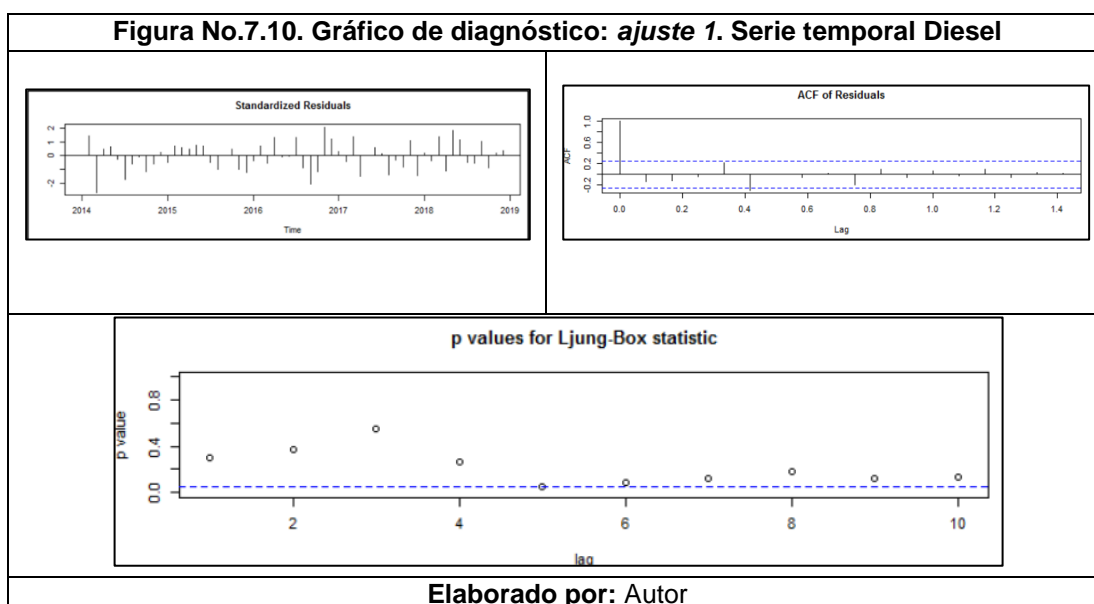
Signif. codes:	0	'***'	0.001	'**'	0.01
				'*'	0.05
				'.'	0.1
				' '	1

Elaborado por: Autor

A través de esta prueba podemos ver que tanto ar1 como ar2 son significativos para el presente modelo. Continuando el procedimiento, realizamos la prueba de Ljung-Box en la que constatamos que hemos realizado el modelo de forma adecuada; puesto que, obtenemos un p-valor de 0.30 superior a 0.05 y no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

Tabla No. 7.30. Prueba Box-Ljung- Orden (2,1,0), Serie Temporal Aditivos para Diesel	
Box-Ljung test	
data: residuals(ajuste1)	
X-squared = 1.0885, df = 1, p-value = 0.2968	

Elaborado por: Autor



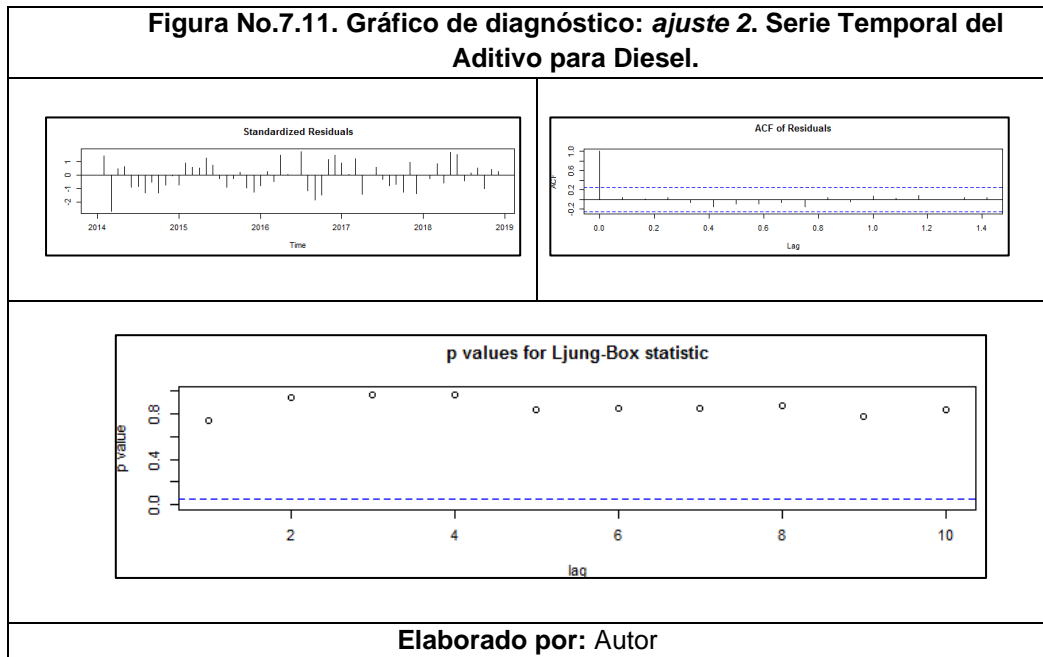
Efectuaremos un segundo modelo manteniendo la primera diferencia aplicada durante el análisis exploratorio de datos; $d=1$, y el coeficiente de media móvil (MA); $q=0$; mientras que para la parte (AR) utilizaremos un $p=3$, con la finalidad de identificar si un modelo de este orden podría representar de forma adecuada a nuestra variable de respuesta:

Tabla No. 7.31. Resumen Estadístico - Orden (3,1,0), Serie Temporal Aditivos para Diesel						
Orden: (3,1,0)						
Coefficients:						
	ar1	ar2	ar3			
	-1.0652	-0.7268	-0.3502			
s.e.	0.1222	0.1631	0.1331			
sigma ² estimated as 869.1: log likelihood = -284.02,						
aic = 576.03						
Training set error measures:						
	MASE	ME	RMSE	MAE	MPE	MAPE
Training set	-2.417664	29.23451	24.12434	-8.978125	25.07494	0.6855546
		ACF1				
Training set		0.04256803				
Elaborado por: Autor						

Tabla No. 7.32. Estadístico Z - Orden (3,1,0), Serie Temporal Aditivos para Diesel						
z test of coefficients:						
	Estimate	Std. Error	z value	Pr(> z)		
ar1	-1.06517	0.12225	-8.7131	< 2.2e-16	***	
ar2	-0.72676	0.16313	-4.4550	8.388e-06	***	
ar3	-0.35020	0.13315	-2.6301	0.008535	**	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
Elaborado por: Autor						

Tabla No. 7.33. Prueba Box-Ljung- Orden (3,1,0), Serie Temporal Aditivos para Diesel	
Box-Ljung test	
data: residuals(ajuste2)	
X-squared = 0.11425, df = 1, p-value = 0.7354	
Elaborado por: Autor	



En relación a los resultados obtenidos de este modelo, observamos que los coeficientes ar1, ar2 y ar3 tienen un nivel de significancia alta en relación al comportamiento de nuestra variable de respuesta. Por otra parte, al ejecutar la prueba de Ljung-Box constatamos que hemos realizado el modelo de forma adecuada; puesto que, obtenemos un p-valor de 0.74 superior a 0.05 y no rechazamos la hipótesis nula de que los residuos se comportan como ruido blanco.

A continuación analizaremos el comportamiento de la parte MA a través de los valores $q=1$ y $q=2$; manteniendo la diferencia aplicada los datos, $d=1$, y restringiendo el comportamiento AR, con $p=0$; como puede verse a continuación:

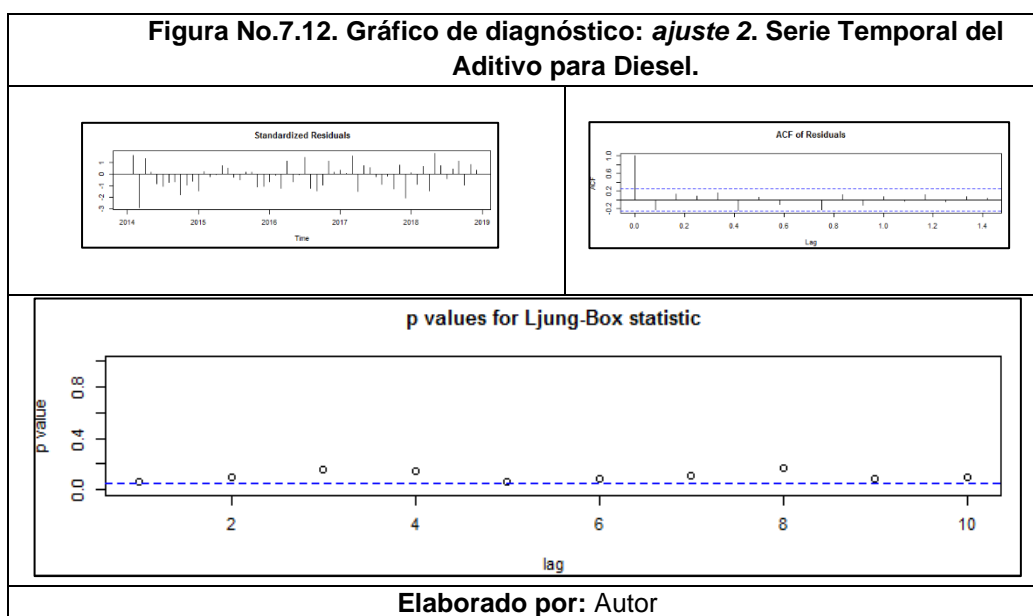
Tabla No. 7.34. Resumen Estadístico - Orden (0,1,1), Serie Temporal Aditivos para Diesel						
Orden: (0,1,1)						
Coefficients:						
	ma1					
	-0.8748					
s.e.	0.0760					
sigma ² estimated as 994.8: log likelihood = -288.07,						
aic = 580.13						
Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	
MASE						
Training set	-5.690875	31.27595	25.4614	-13.76963	27.73231	0.7
	235505					
		ACF1				
Training set		-0.2376422				

Elaborado por: Autor

Tabla No. 7.35. Estadístico Z - Orden (0,1,1), Serie Temporal Aditivos para Diesel	
z test of coefficients:	
Estimate	Std. Error z value Pr(> z)
ma1	-0.874769 0.075963 -11.516 < 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	
Elaborado por: Autor	

Tabla No. 7.36. Prueba Box-Ljung- Orden (0,1,1), Serie Temporal Aditivos para Diesel	
Box-Ljung test	
data: residuals(ajuste3)	
x-squared = 3.5607, df = 1, p-value = 0.05916	
Elaborado por: Autor	



Considerando los resultados obtenidos del *ajuste3*, observamos a pesar de que el coeficiente *ma1* tiene un nivel de significancia alta en relación al comportamiento de nuestra variable de respuesta, al ejecutar la prueba de Ljung-Box obtenemos un valor *p* cercano a 0.05; 0.06, lo cual es corroborado a través de la figura del diagnóstico del ajuste, y esto podría ser indicativo de que nuestros residuos podrían no comportarse como ruido blanco; por lo que continuamos con la generación de otro modelo que represente de mejor manera el comportamiento de nuestra variable de respuesta. Por lo anteriormente expuesto, analizaremos el comportamiento de la parte MA de orden 2, $q=2$; manteniendo la diferencia aplicada los datos, $d=1$, y restringiendo el comportamiento AR, con $p=0$; como puede verse a continuación:

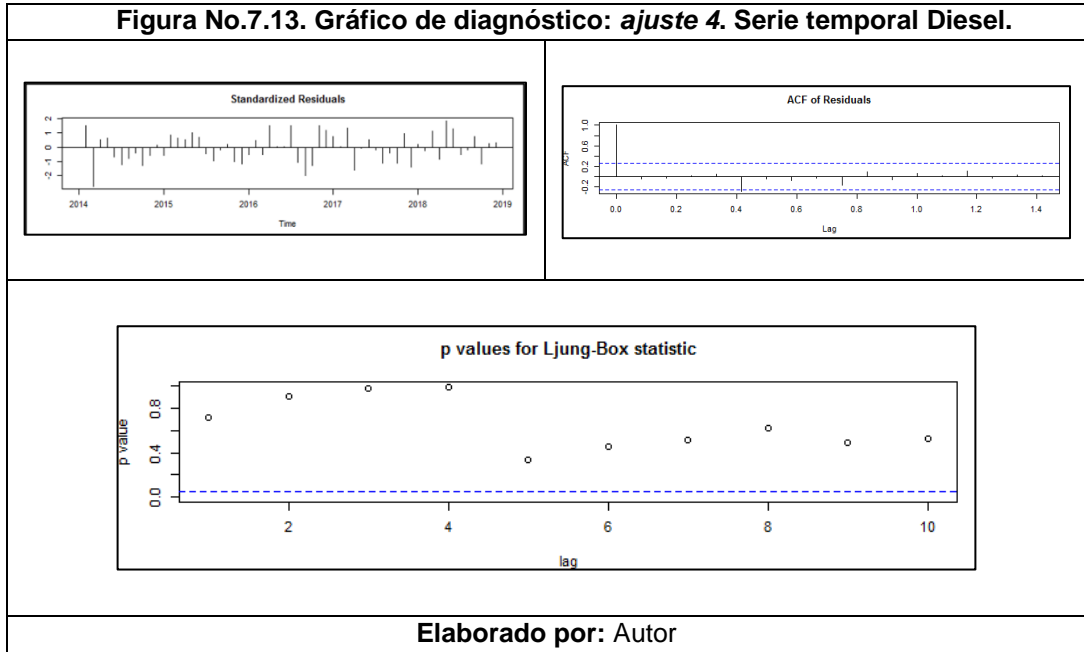
Tabla No. 7.37. Resumen Estadístico - Orden (0,1,2), Serie Temporal Aditivos para Diesel						
Orden: (0,1,2)						
Coefficients:						
	ma1		ma2			
	-0.9895		0.3842			
s.e.	0.1065		0.1497			
sigma^2 estimated as 925.4: log likelihood = -285.73,						
aic = 577.45						
Training set error measures:						
		ME	RMSE	MAE	MPE	MAPE
MASE						
Training set	-1.927927	30.1663	24.84543	-8.970743	25.99283	0.7060461
			ACF1			
Training set	-0.04616433					
Elaborado por: Autor						

Tabla No. 7.38. Estadístico Z - Orden (0,1,2), Serie Temporal Aditivos para Diesel						
z test of coefficients:						
	Estimate	Std. Error	z value	Pr(> z)		
ma1	-0.98952	0.10646	-9.2946	< 2e-16	***	
ma2	0.38419	0.14967	2.5669	0.01026	*	

Signif. codes:	0	'***'	0.001	'**'	0.01	'*' 0.05 '.' 0.1 ' ' 1
Elaborado por: Autor						

Tabla No. 7.39. Prueba Box-Ljung- Orden (0,1,2), Serie Temporal Aditivos para Diesel	
Box-Ljung test	
data: residuals(ajuste4)	
x-squared = 0.13437, df = 1, p-value = 0.7139	
Elaborado por: Autor	

Figura No.7.13. Gráfico de diagnóstico: ajuste 4. Serie temporal Diesel.



Elaborado por: Autor

Vemos para el *ajuste4* con MA de orden 2 y una vez diferenciado, los coeficientes *ma1* y *ma2* tienen un nivel de significancia aceptable y su comportamiento en relación al ruido blanco es mejor que el *ajuste3* considerando un valor p en la prueba de Ljung-Box de 0.7139.

Tabla No. 7.40. Resumen Estadístico - Orden (3,0,0), Serie Temporal Aditivos para Diesel

```

Orden: (3,0,0)
Coefficients:
      ar1      ar2      ar3      sma1      sma2
-0.0522  0.1061  0.3286 -0.4256  0.1151
s.e.    0.1532  0.1792  0.1542  0.1915  0.2251

sigma^2 estimated as 1185:  log likelihood = -239.24,
aic = 490.47

Training set error measures:
              ME   RMSE   MAE   MPE   MAPE
MASE
Training set -4.567525 30.786 22.55583 -10.54713 24.84176 0.64
09813
              ACF1
Training set -0.1053845
    
```

Elaborado por: Autor

Tabla No. 7.41. Estadístico Z - Orden (3,0,0), Serie Temporal Aditivos para Diesel

```

z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1  -0.052238  0.153244 -0.3409  0.73319
ar2   0.106110  0.179245  0.5920  0.55386
ar3   0.328572  0.154193  2.1309  0.03310 *
sma1 -0.425625  0.191467 -2.2230  0.02622 *
sma2  0.115099  0.225071  0.5114  0.60908
---
    
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Elaborado por: Autor

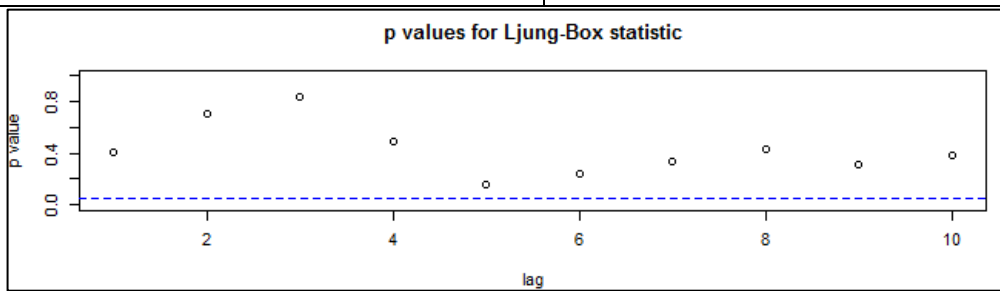
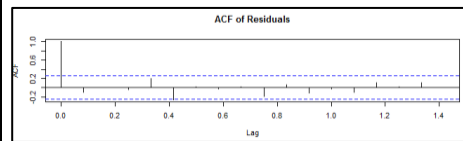
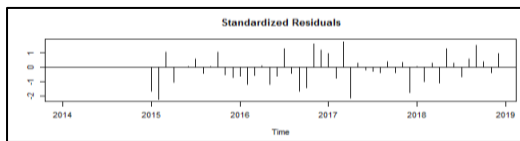
Tabla No. 7.42. Prueba Box-Ljung- Orden (3,0,0), Serie Temporal Aditivos para Diesel

Box-Ljung test

data: residuals(ajuste6)
x-squared = 0.70024, df = 1, p-value = 0.4027

Elaborado por: Autor

Figura No.7.14. Gráfico de diagnóstico: ajuste 6. Serie Temporal del Aditivo para Diesel.



Elaborado por: Autor

7.4. Nota de Programación

7.4.1. #Serie Temporal: Ventas Totales

```
#Importacion de datos y activación de librerías
ventas <- read.csv("C:/Users/HMA/Desktop/DATOS.csv",
                 sep=";")
attach(ventas)
str(ventas)
View(ventas)
library(forecast)
library(foreign)
library(astsa)
library(ggplot2)
library(tseries)
library(lmtest)
ventas.ts1=ts(ventas$series.ventas.ts,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.ts=ventas.ts1/1000
ventas.tsc

#MODELOS ARIMA (p,d,q)(P,D,Q)s
#Análisis exploratorio de datos
hist(ventas.ts,main="Histograma de Ventas Mensuales Totales de Aditivos
(MT)",xlab="Ventas Mensuales Totales de Aditivos (MT)",ylab="Frecuencia")
#Verificamos normalidad de datos
shapiro.test(ventas.ts)
jarque.bera.test(ventas.ts)
plot(ventas.ts,main="Ventas Totales de Aditivos vs. Tiempo",ylab="Ventas Totales (MT)",
xlab="Tiempo (Año)")
grid()
TScomp <- decompose(ventas.ts)
plot(TScomp)
cycle(ventas.ts)
boxplot(ventas.ts~cycle(ventas.ts),main="Diagrama de Caja",xlab="Mes del
Año",ylab="Ventas de Aditivos (MT)")

#Transformamos los datos en un objeto serie de tiempo
ventas.ts1=ts(ventas$series.ventas.ts,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.ts=ventas.ts1/1000
ventas.ts

#Determinación del número de diferencias aplicadas para una serie estacionaria
?ndiffs
ndiffs(ventas.ts,"pp")
```

```
plot(ventas.ts, type="o", col="blue", lty="dashed", main="Serie de Tiempo Ventas Totales
de Aditivos",
     ylab="Ventas Totales de Aditivos (MT)", xlab="Tiempo (Año)" )
abline(reg=lm(ventas.ts~time(ventas.ts)))
```

#Aqui observamos que nuestra serie de tiempo es no estacionaria en MEDIA
 #Al parecer tambien tiene un componente estacional (para eliminar esto aplicamos log) y
 variaciones irregulares

```
adf.test(ventas.ts)
```

#Al parecer tambien tiene un componente estacional y variaciones irregulares
 #Al no ser este valor mas pequeno que 0.05 no se rechaza hipotesis nula
 #de que la serie de tiempo posee raiz unitaria y esto significaria
 #que esta serie no es estacionaria.

```
acf(ventas.ts)
pacf(ventas.ts)
```

#Debemos buscar una la estacionariedad y remover la tendencia
 serielog<-log(ventas.ts)

```
plot(serielog, type="o", col="blue", lty="dashed", main="Log-TS Venta de
Aditivos", xlab="Año", ylab="Log-TS")
```

#Para eliminar la tendencia aplicamos diferencias
 seriedif<-diff(serielog)

```
plot(seriedif, type="o", col="blue", lty="dashed", main="DiffLog-TS Venta de
Aditivos", ylab="DiffLog-TS", xlab="Tiempo (Año)")
```

#Comprobación de estacionariedad.

```
adf.test(seriedif)
adf.test
```

#Preparacion de dimensiones para graficos
 x11()

```
par(mfrow=c(2,1), mar=c(4,4,4,1)+.1)
acf(ts(seriedif, freq=1))
pacf(ts(seriedif, freq=1))
```

#ANALISIS 1 AR(1)

```
ajuste1=arima(ventas.ts,order=c(1,0,0))
summary(ajuste1)
coeftest(ajuste1)
Box.test(residuals(ajuste1), type="Ljung-Box")
tsdiag(ajuste1)
```

#ANALISIS 2 AR(2)

```
ajuste2=arima(ventas.ts,order=c(2,0,0))
summary(ajuste2)
```

```

coefstest(ajuste2)
Box.test(residuals(ajuste2), type="Ljung-Box")
tsdiag(ajuste2)

#ANALISIS 3 ARIMA(2,1,1)
ajuste3=arima(ventas.ts,order=c(2,1,1))
summary(ajuste3)
coefstest(ajuste3)
Box.test(residuals(ajuste3), type="Ljung-Box")
tsdiag(ajuste3)

#Modelos Multiplicativos Estacionales

#La prueba 4 y 5 son las mismas con diferentes funciones ARIMA y SARIMA

ajuste4=arima(ventas.ts,order=c(2,0,0),seasonal=list(order=c(1,1,1),period=12))
summary(ajuste4)
coefstest(ajuste4)
Box.test(residuals(ajuste4), type="Ljung-Box")
tsdiag(ajuste4)

ajuste5=arima(ventas.ts,order=c(3,0,0),seasonal=list(order=c(1,1,1),period=12))
summary(ajuste5)
coefstest(ajuste5)
Box.test(residuals(ajuste5), type="Ljung-Box")
tsdiag(ajuste5)
#Eleccion del mejor modelo
AIC(ajuste1)
BIC(ajuste1)
AIC(ajuste2)
BIC(ajuste2)
AIC(ajuste3)
BIC(ajuste3)
AIC(ajuste4)
BIC(ajuste4)
AIC(ajuste5)
BIC(ajuste5)

#Pronósticos
pronostico<- forecast(ajuste4,12,level=95)
plot(pronostico, main="Pronóstico con Arima (2,0,0)x(1,1,1)12",
      xlab="Tiempo(Año)",
      ylab="Ventas Totales kg")
grid()
legend("topright", legend=c("Ventas Reales", "Pronóstico"),
      col=c("black", "blue"), lty=1:1, cex=0.6)

```

```

plot(ventas.ts, main="Comparación: Serie Pronosticada-Serie Real",
     xlab="Tiempo(Año)",
     ylab="Ventas Totales (MT)",ylim=c(0,600),xlim=c(2014,2021))
lines(fitted(ajuste4), col='red')
grid()
legend("top", legend=c("Ventas Reales", "Pronóstico"),
      col=c("black", "red"), lty=1:1, cex=0.6)

pronostico<- forecast(ajuste4,12,level=95)

plot(pronostico, main="Pronóstico con Arima (2,0,0)x(1,1,1)12",
     xlab="Tiempo(Año)",
     ylab="Ventas Totales (MT)",ylim=c(0,600),xlim=c(2019,2021))
lines(fitted(ajuste4), col='blue')
lines(ventas.tsc,col="red")
grid()
legend("topright", legend=c("Pronóstico de Ventas", "Ventas Reales"),
      col=c("blue", "red"), lty=1:1, cex=0.6)

matriz.pronosticos <-data.frame(pronostico$mean,pronostico$lower,pronostico$upper)
matriz.pronosticos

sum(pronostico$mean)

```

7.4.2. #Serie Temporal: Fuera de Borda

```

#Importacion de datos y activación de librerías
ventas <- read.csv("C:/Users/HMA/Desktop/DATOS.csv",
                  sep=";")
attach(ventas)
str(ventas)
View(ventas)
library(forecast)
library(foreign)
library(astsa)
library(ggplot2)
library(tseries)
library(lmtest)
ventas.ts1=ts(ventas$series424,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.ts=ventas.ts1/1000
ventas.ts

#MODELOS ARIMA (p,d,q)(P,D,Q)s

#Análisis exploratorio de datos

```

```

hist(ventas.ts,main="Histograma de Ventas Mensuales "Fuera de
Borda"(MT)",xlab="Ventas Mensuales "Fuera de Borda" (MT)",ylab="Frecuencia")
#Verificamos normalidad de datos
shapiro.test(ventas.ts)
jarque.bera.test(ventas.ts)
plot(ventas.ts,main="Ventas Mensuales "Fuera de Borda" vs. Tiempo",ylab="Ventas "Fuera
de Borda" (MT)", xlab="Tiempo (Año)")
TScomp <- decompose(ventas.ts)
plot(TScomp)
cycle(ventas.ts)
boxplot(ventas.ts~cycle(ventas.ts),main="Diagrama de Caja",xlab="Mes del
Año",ylab="Ventas de "Fuera de Borda" (MT)")

#Transformamos los datos en un objeto serie de tiempo
ventas.ts1=ts(ventas$series424,start=c(2014,1),end=c(2018,12),frequency=12)
ventas.ts=ventas.ts1/1000
ventas.ts

plot(ventas.ts, type="o", col="blue", lty="dashed", main="Serie de Tiempo Ventas: Fuera
de Borda",
      ylab="Ventas Aditivo "Fuera de Borda" (MT)",xlab="Tiempo (Año)" )
abline(reg=lm(ventas.ts~time(ventas.ts)))

#Estimacion de número de diferencias
ndiffs(ventas.ts,"pp")

#Pruebas de estacionariedad
adf.test(ventas.ts)

acf(ventas.ts)
pacf(ventas.ts)

#Aplicación de logaritmos
serielog<-log(ventas.ts)

plot(serielog, type="o", col="blue", lty="dashed", main="Log-TS Venta de "Fuera de
Borda"",xlab="Año",ylab="Log-TS")
#Para eliminar la tendencia aplicamos diferencias
seriedif<-diff(serielog)
plot(seriedif, type="o", col="blue", lty="dashed", main="DiffLog-TS Venta de "Fuera de
Borda"",ylab="DiffLog-TS",xlab="Tiempo (Año)")
#Comprobación de estacionariedad
adf.test(ventas.ts)

#Preparación de dimensiones de gráficos
x11()
par(mfrow=c(2,1), mar=c(4,4,4,1)+.1)

```

```

acf(ts(seriedif, freq=1))
pacf(ts(seriedif, freq=1))

#ANALISIS 1 AR(1)
ajuste1=arima(ventas.ts,order=c(1,0,0))
summary(ajuste1)
coefest(ajuste1)
Box.test(residuals(ajuste1), type="Ljung-Box")
tsdiag(ajuste1)

#ANALISIS 2 AR(2)
ajuste2=arima(ventas.ts,order=c(3,0,0))
summary(ajuste2)
coefest(ajuste2)
Box.test(residuals(ajuste2), type="Ljung-Box")
tsdiag(ajuste2)

#ANALISIS 3 ARIMA(0,1,1)
ajuste3=arima(ventas.ts,order=c(0,1,1))
summary(ajuste3)
coefest(ajuste3)
Box.test(residuals(ajuste3), type="Ljung-Box")
tsdiag(ajuste3)

#Modelos Multiplicativos Estacionales

#La prueba 4 y 5 son las mismas con diferentes funciones ARIMA y SARIMA

ajuste4=arima(ventas.ts,order=c(0,0,0),seasonal=list(order=c(0,1,1),period=12))
summary(ajuste4)
coefest(ajuste4)
Box.test(residuals(ajuste4), type="Ljung-Box")
tsdiag(ajuste4)

#Eleccion del mejor modelo
AIC(ajuste1)
BIC(ajuste1)
AIC(ajuste2)
BIC(ajuste2)
AIC(ajuste3)
BIC(ajuste3)
AIC(ajuste4)
BIC(ajuste4)

#Pronósticos
pronostico<- forecast(ajuste4,12,level=95)
plot(pronostico, main="Pronóstico con Arima (0,0,0)x(0,1,1)12",

```



```

      xlab="Tiempo(Año)",
      ylab="Ventas "Fuera de Borda" (MT)")
grid()
legend("topright", legend=c("Ventas Reales", "Pronóstico"),
      col=c("black", "blue"), lty=1:1, cex=0.6)

plot(ventas.ts, main="Comparación: Serie Pronosticada-Serie Real",
      xlab="Tiempo(Año)",
      ylab="Ventas "Fuera de Borda"(MT)",ylim=c(0,100),xlim=c(2014,2019))
lines(fitted(ajuste4), col='red')
grid()
legend("top", legend=c("Ventas "Fuera de Borda"", "Pronóstico "Fuera de Borda""),
      col=c("black", "red"), lty=1:1, cex=0.6)

pronostico<- forecast(ajuste4,12,level=95)

plot(pronostico, main="Pronóstico con Arima (0,0,0)x(0,1,1)12",
      xlab="Tiempo(Año)",
      ylab="Ventas "Fuera de Borda"(MT)",ylim=c(0,100),xlim=c(2014,2020))
lines(fitted(ajuste4), col='blue')
lines(ventas.tsc,col="red")
grid()
legend("top", legend=c("Pronóstico "Fuera de Borda"", "Ventas "Fuera de Borda""),
      col=c("blue", "red"), lty=1:1, cex=0.6)

matriz.pronosticos <-data.frame(pronostico$mean,pronostico$lower,pronostico$upper)
matriz.pronosticos

sum(pronostico$mean)
sum(ventas$series424[61:72])/1000

```

7.4.3. #Serie Temporal Gasolina

```

#Importación de datos y activación de librerías
ventas <- read.csv("C:/Users/HMA/Desktop/DATOS.csv",
      sep=";")

attach(ventas)
str(ventas)
View(ventas)
library(forecast)
library(foreign)
library(astsa)
library(ggplot2)
library(tseries)
library(lmtest)

```

```
ventas.ts1=ts(ventas$seriesPV1021,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsg1=ventas.ts1/1000
ventas.tsg1
```

```
ventas.ts2=ts(ventas$seriesPV7053,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsg2=ventas.ts2/1000
ventas.tsg2
```

```
ventas.ts3=ts(ventas$seriesPV1119,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsg3=ventas.ts3/1000
ventas.tsg3
```

```
ventas.ts4=ts(ventas$seriesgasolina,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsg=ventas.ts4/1000
ventas.tsg
```

```
#MODELOS ARIMA (p,d,q)(P,D,Q)s
```

```
#Análisis exploratorio de datos
```

```
hist(ventas.tsg1,main="Histograma de Ventas Mensuales "Gasolina 1"(MT)",xlab="Ventas Mensuales "Gasolina 1" (MT)",ylab="Frecuencia")
```

```
hist(ventas.tsg2,main="Histograma de Ventas Mensuales "Gasolina 2"(MT)",xlab="Ventas Mensuales "Gasolina 2" (MT)",ylab="Frecuencia")
```

```
hist(ventas.tsg3,main="Histograma de Ventas Mensuales "Gasolina 3"(MT)",xlab="Ventas Mensuales "Gasolina 3" (MT)",ylab="Frecuencia")
```

```
hist(ventas.tsg,main="Histograma de Ventas Mensuales "Gasolina"(MT)",xlab="Ventas Mensuales "Gasolina" (MT)",ylab="Frecuencia")
```

```
#Verificamos normalidad de datos
```

```
shapiro.test(ventas.tsg1)
```

```
shapiro.test(ventas.tsg2)
```

```
shapiro.test(ventas.tsg3)
```

```
shapiro.test(ventas.tsg)
```

```
jarque.bera.test(ventas.tsg1)
```

```
jarque.bera.test(ventas.tsg2)
```

```
jarque.bera.test(ventas.tsg3)
```

```
jarque.bera.test(ventas.tsg)
```

```
plot(ventas.tsg1,main="Ventas Mensuales "Gasolina 1" vs. Tiempo",ylab="Ventas "Gasolina 1" (MT)", xlab="Tiempo (Año)")
```

```
plot(ventas.tsg2,main="Ventas Mensuales "Gasolina 2" vs. Tiempo",ylab="Ventas "Gasolina 2" (MT)", xlab="Tiempo (Año)")
```

```
plot(ventas.tsg3,main="Ventas Mensuales "Gasolina 3" vs. Tiempo",ylab="Ventas "Gasolina 3" (MT)", xlab="Tiempo (Año)")
```

```
plot(ventas.tsg,main="Ventas Mensuales "Gasolina" vs. Tiempo",ylab="Ventas "Gasolina 3" (MT)", xlab="Tiempo (Año)")
```

```

abline(reg=lm(ventas.ts~time(ventas.ts)))

plot(ventas.tsg1, main="Series Temporales de Ventas de Aditivos "Gasolina",
      xlab="Tiempo(Año)",
      ylab="Ventas (MT)",ylim=c(-10,60),xlim=c(2014,2020))
lines(ventas.tsg2, col='blue')
lines(ventas.tsg3, col='red')
lines(ventas.tsg, col="green", lty="dashed")
grid()
legend("topleft", legend=c("Gasolina 1", "Gasolina 2", "Gasolina 3", "Gasolina"),
      col=c("black", "red", "blue", "green"), lty=1:1, cex=0.6)

#Descomposición
TScomp <- decompose(ventas.tsg)
plot(TScomp)

cycle(ventas.tsg)
boxplot(ventas.tsg~cycle(ventas.tsg),main="Diagrama de Caja",xlab="Mes del
Año",ylab="Ventas de "Gasolina" (MT)")

#Determinación del número de diferencias aplicadas para una serie estacionaria
ndiffs(ventas.ts,"pp")

#Transformamos los datos en un objeto serie de tiempo
ventas.ts1=ts(ventas$seriesgasolina,start=c(2014,1),end=c(2018,12),frequency=12)
ventas.ts=ventas.ts1/1000
ventas.ts

plot(ventas.ts, type="o", col="blue", lty="dashed", main="Serie de Tiempo Ventas:
Gasolina",
      ylab="Ventas Aditivo "Gasolina" (MT)",xlab="Tiempo (Año)" )
abline(reg=lm(ventas.ts~time(ventas.ts)))
grid()

#Comprobación de estacionariedad
adf.test(ventas.ts)
acf(ventas.ts)
pacf(ventas.ts)

#Debemos buscar una la estacionariedad y remover la tendencia
serielog<-log(ventas.ts)

plot(serielog, type="o", col="blue", lty="dashed", main="Log-TS Venta de
"Gasolina",xlab="Año",ylab="Log-TS")
#Para eliminar la tendencia aplicamos diferencias

```

```

seriedif<-diff(serieIog)
plot(seriedif, type="o", col="blue", lty="dashed", main="DiffLog-TS Venta de
"Gasolina",ylab="DiffLog-TS",xlab="Tiempo (Año)")
#Comprobación de estacionariedad
adf.test(seriedif)

#Preparación de dimensiones de gráficos
x11()
par(mfrow=c(2,1), mar=c(3,3,3,1)+.1)
acf(ts(seriedif, freq=1))
pacf(ts(seriedif, freq=1))

#ANALISIS 1 AR(1)
ajuste1=arima(ventas.ts,order=c(1,0,0))
summary(ajuste1)
coeftest(ajuste1)
Box.test(residuals(ajuste1), type="Ljung-Box")
tsdiag(ajuste1)

#ANALISIS 2 AR(2)
ajuste2=arima(ventas.ts,order=c(2,0,0))
summary(ajuste2)
coeftest(ajuste2)
Box.test(residuals(ajuste2), type="Ljung-Box")
tsdiag(ajuste2)

#ANALISIS 3 ARIMA(0,1,1)
ajuste3=arima(ventas.ts,order=c(0,1,1))
summary(ajuste3)
coeftest(ajuste3)
Box.test(residuals(ajuste3), type="Ljung-Box")
tsdiag(ajuste3)

#Modelos Multiplicativos Estacionales

#ANALISIS 4 SARIMA

ajuste4=arima(ventas.ts,order=c(0,0,0),seasonal=list(order=c(0,1,1),period=12))
summary(ajuste4)
coeftest(ajuste4)
Box.test(residuals(ajuste4), type="Ljung-Box")
tsdiag(ajuste4)

#Eleccion del mejor modelo
AIC(ajuste1)
BIC(ajuste1)
AIC(ajuste2)

```

```

BIC(ajuste2)
AIC(ajuste3)
BIC(ajuste3)
AIC(ajuste4)
BIC(ajuste4)

#Pronósticos
pronostico<- forecast(ajuste4,12,level=95)
plot(pronostico, main="Pronóstico con Arima (0,0,0)x(0,1,1)12",
      xlab="Tiempo(Año)",
      ylab="Ventas "Gasolina" (MT)")
grid()
legend("topleft", legend=c("Ventas Reales", "Pronóstico"),
      col=c("black", "blue"), lty=1:1, cex=0.6)

plot(ventas.ts, main="Comparación: Serie Pronosticada-Serie Real",
      xlab="Tiempo(Año)",
      ylab="Ventas "Gasolina"(MT)",ylim=c(-10,60),xlim=c(2014,2019))
lines(fitted(ajuste4), col='red')
grid()
legend("topright", legend=c("Ventas "Gasolina"", "Pronóstico "Gasolina""),
      col=c("black", "red"), lty=1:1, cex=0.6)

pronostico<- forecast(ajuste4,12,level=95)

plot(pronostico, main="Pronóstico con Arima (0,0,0)x(0,1,1)12",
      xlab="Tiempo(Año)",
      ylab="Ventas "Gasolina"(MT)",ylim=c(-10,60),xlim=c(2019,2020))
lines(fitted(ajuste4), col='blue')
lines(ventas.ts, col="red")
grid()
legend("topright", legend=c("Pronóstico "Gasolina"", "Ventas "Gasolina""),
      col=c("blue", "red"), lty=1:1, cex=0.6)

matriz.pronosticos <-data.frame(pronostico$mean,pronostico$lower,pronostico$upper)
matriz.pronosticos

sum(pronostico$mean)
sum(ventas$seriesgasolina[61:72])/1000

```

7.4.4. #Serie Temporal: Diesel

```

#Importación de datos y activación de librerías
ventas <- read.csv("C:/Users/HMA/Desktop/DATOS.csv",
                  sep=";")
attach(ventas)
str(ventas)

```

```

View(ventas)
library(forecast)
library(foreign)
library(astsa)
library(ggplot2)
library(tseries)
library(lmtest)
ventas.ts1=ts(ventas$seriesCV2301,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsd1=ventas.ts1/1000
ventas.tsd1

```

```

ventas.ts2=ts(ventas$seriesCV2320,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsd2=ventas.ts2/1000
ventas.tsd2

```

```

ventas.ts3=ts(ventas$seriesCV7101,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsd3=ventas.ts3/1000
ventas.tsd3

```

```

ventas.ts4=ts(ventas$seriesCV1100,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsd4=ventas.ts4/1000
ventas.tsd4

```

```

ventas.ts5=ts(ventas$series4980A,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsd5=ventas.ts5/1000
ventas.tsd5

```

```

ventas.ts=ts(ventas$seriesdiesel,start=c(2014,1),end=c(2019,12),frequency=12)
ventas.tsd=ventas.ts/1000
ventas.tsdc

```

```

#MODELOS ARIMA (p,d,q)(P,D,Q)s

```

```

#Análisis exploratorio de datos

```

```

hist(ventas.tsd1,main="Histograma de Ventas Mensuales "Diesel 1"(MT)",xlab="Ventas Mensuales "Diesel 1" (MT)",ylab="Frecuencia")
hist(ventas.tsd2,main="Histograma de Ventas Mensuales "Diesel 2"(MT)",xlab="Ventas Mensuales "Diesel 2" (MT)",ylab="Frecuencia")
hist(ventas.tsd3,main="Histograma de Ventas Mensuales "Diesel 3"(MT)",xlab="Ventas Mensuales "Diesel 3" (MT)",ylab="Frecuencia")
hist(ventas.tsd4,main="Histograma de Ventas Mensuales "Diesel 4"(MT)",xlab="Ventas Mensuales "Diesel 4" (MT)",ylab="Frecuencia")
hist(ventas.tsd5,main="Histograma de Ventas Mensuales "Diesel 5"(MT)",xlab="Ventas Mensuales "Diesel 5" (MT)",ylab="Frecuencia")

```

```
hist(ventas.tsd,main="Histograma de Ventas Mensuales "Diesel"(MT)",xlab="Ventas Mensuales "Diesel" (MT)",ylab="Frecuencia")
```

```
#Verificamos normalidad de datos
```

```
shapiro.test(ventas.tsd1)
```

```
shapiro.test(ventas.tsd2)
```

```
shapiro.test(ventas.tsd3)
```

```
shapiro.test(ventas.tsd4)
```

```
shapiro.test(ventas.tsd5)
```

```
shapiro.test(ventas.tsd)
```

```
jarque.bera.test(ventas.tsd1)
```

```
jarque.bera.test(ventas.tsd2)
```

```
jarque.bera.test(ventas.tsd3)
```

```
jarque.bera.test(ventas.tsd4)
```

```
jarque.bera.test(ventas.tsd5)
```

```
jarque.bera.test(ventas.tsd)
```

```
#Verificacion de numero de diferencias
```

```
ndiffs(ventas.tsd)
```

```
plot(ventas.tsd1,main="Ventas Mensuales "Diesel 1" vs. Tiempo",ylab="Ventas "Gasolina 1" (MT)", xlab="Tiempo (Año)")
```

```
plot(ventas.tsd2,main="Ventas Mensuales "Diesel 2" vs. Tiempo",ylab="Ventas "Gasolina 2" (MT)", xlab="Tiempo (Año)")
```

```
plot(ventas.tsd3,main="Ventas Mensuales "Diesel 3" vs. Tiempo",ylab="Ventas "Gasolina 3" (MT)", xlab="Tiempo (Año)")
```

```
plot(ventas.tsd4,main="Ventas Mensuales "Diesel 4" vs. Tiempo",ylab="Ventas "Gasolina 4" (MT)", xlab="Tiempo (Año)")
```

```
plot(ventas.tsd5,main="Ventas Mensuales "Diesel 5" vs. Tiempo",ylab="Ventas "Gasolina 5" (MT)", xlab="Tiempo (Año)")
```

```
plot(ventas.tsd,main="Ventas Mensuales "Diesel" vs. Tiempo",ylab="Ventas "Diesel" (MT)", xlab="Tiempo (Año)")
```

```
abline(reg=lm(ventas.tsd~time(ventas.tsd)))
```

```
plot(ventas.tsd1, main="Comparación de Series Ventas "Diesel"",
```

```
  xlab="Tiempo(Año)",
```

```
  ylab="Ventas "Diesel 1"(MT)",ylim=c(0,250),xlim=c(2014,2019))
```

```
lines(ventas.tsd2, col='blue')
```

```
lines(ventas.tsd3, col='red')
```

```
lines(ventas.tsd4, col='orange')
```

```
lines(ventas.tsd5, col='purple')
```

```
lines(ventas.tsd, col="green", lty="dashed")
```

```
grid()
```

```
legend("topright", legend=c("Diesel 1", "Diesel 2", "Diesel 3", "Diesel 4","Diesel 5", "Diesel"),
```

```

col=c("black", "blue", "red", "orange", "purple", "green"), lty=1:1, cex=0.6)

#Descomposición
TScomp <- decompose(ventas.tsd)
plot(TScomp)

cycle(ventas.tsd)
boxplot(ventas.tsd~cycle(ventas.tsd),main="Diagrama de Caja",xlab="Mes del
Año",ylab="Ventas de "Diesel" (MT)")

#Transformamos los datos en un objeto serie de tiempo
ventas.ts1=ts(ventas$seriesdiesel,start=c(2014,1),end=c(2018,12),frequency=12)
ventas.ts=ventas.ts1/1000
ventas.ts

plot(ventas.ts, type="o", col="blue", lty="dashed", main="Serie de Tiempo Ventas: Diesel",
ylab="Ventas Aditivo "Diesel" (MT)",xlab="Tiempo (Año)" )
abline(reg=lm(ventas.ts~time(ventas.ts)))

#Comprobación de estacionariedad
adf.test(ventas.ts)

acf(ventas.ts)
pacf(ventas.ts)

#Aplicación de logaritmos
serielog<-log(ventas.ts)

plot(serielog, type="o", col="blue", lty="dashed", main="Log-TS Venta de
"Diesel",xlab="Año",ylab="Log-TS")
#Para eliminar la tendencia aplicamos diferencias
seriedif<-diff(serielog)
plot(seriedif, type="o", col="blue", lty="dashed", main="DiffLog-TS Venta de
"Diesel",ylab="DiffLog-TS",xlab="Tiempo (Año)")
#Comprobación de estacionariedad a la serie diferenciada
adf.test(seriedif)

#Preparación de las dimensiones de los gráficos
x11()
par(mfrow=c(2,1), mar=c(4,4,4,1)+.1)
acf(ts(seriedif, freq=1))
pacf(ts(seriedif, freq=1))

#ANALISIS 1 AR(2)
ajuste1=arima(ventas.ts,order=c(2,1,0))
summary(ajuste1)
coeftest(ajuste1)

```



```
Box.test(residuals(ajuste1), type="Ljung-Box")
tsdiag(ajuste1)
```

```
#ANALISIS 2 AR(3)
ajuste2=arima(ventas.ts,order=c(3,1,0))
summary(ajuste2)
coefstest(ajuste2)
Box.test(residuals(ajuste2), type="Ljung-Box")
tsdiag(ajuste2)
```

```
#ANALISIS 3 ARIMA(0,1,1)
ajuste3=arima(ventas.ts,order=c(0,1,1))
summary(ajuste3)
coefstest(ajuste3)
Box.test(residuals(ajuste3), type="Ljung-Box")
tsdiag(ajuste3)
```

```
#ANALISIS 4 ARIMA(0,1,2)
ajuste4=arima(ventas.ts,order=c(0,1,2))
summary(ajuste4)
coefstest(ajuste4)
Box.test(residuals(ajuste4), type="Ljung-Box")
tsdiag(ajuste4)
```

```
#Modelos Multiplicativos Estacionales
```

```
#ANALISIS 5 SARIMA
```

```
ajuste5=arima(ventas.ts,order=c(3,0,0),seasonal=list(order=c(0,1,1),period=12))
summary(ajuste5)
coefstest(ajuste5)
Box.test(residuals(ajuste5), type="Ljung-Box")
tsdiag(ajuste5)
```

```
#ANALISIS 6 SARIMA
```

```
ajuste6=arima(ventas.ts,order=c(3,0,0),seasonal=list(order=c(0,1,2),period=12))
summary(ajuste6)
coefstest(ajuste6)
Box.test(residuals(ajuste6), type="Ljung-Box")
tsdiag(ajuste6)
```

```
#Eleccion del mejor modelo
```

```
AIC(ajuste1)
BIC(ajuste1)
AIC(ajuste2)
BIC(ajuste2)
```

```

AIC(ajuste3)
BIC(ajuste3)
AIC(ajuste4)
BIC(ajuste4)
AIC(ajuste5)
BIC(ajuste5)
AIC(ajuste6)
BIC(ajuste6)

```

```
#Pronosticos
```

```

plot(ventas.ts, main="Comparación: Serie Pronosticada-Serie Real",
     xlab="Tiempo(Año)",
     ylab="Ventas Diesel (MT)",ylim=c(0,200),xlim=c(2014,2019))
lines(fitted(ajuste5), col='red')
grid()
legend("top", legend=c("Ventas Diesel", "Pronóstico Diesel"),
       col=c("black", "red"), lty=1:1, cex=0.6)
legend("bottomleft", legend=c("ARIMA (3,0,0)x(0,1,1)12"))

```

```

plot(ventas.ts, main="Comparación: Serie Pronosticada-Serie Real",
     xlab="Tiempo(Año)",
     ylab="Ventas Diesel (MT)",ylim=c(0,200),xlim=c(2014,2019))
lines(fitted(ajuste4), col='red')
grid()
legend("top", legend=c("Ventas Diesel", "Pronóstico Diesel"),
       col=c("black", "red"), lty=1:1, cex=0.6)
legend("bottomleft", legend=c("ARIMA (0,1,2)"))

```

```

pronostico<- forecast(ajuste5,12,level=95)
plot(pronostico, main="Pronóstico con Arima (3,0,0)x(0,1,1)12",
     xlab="Tiempo(Año)",
     ylab="Ventas Diesel (MT)")
grid()
legend("top", legend=c("Ventas Reales", "Pronóstico"),
       col=c("black", "blue"), lty=1:1, cex=0.6)

```

```

pronostico<- forecast(ajuste4,12,level=95)
plot(pronostico, main="Pronóstico con Arima (0,1,2)",
     xlab="Tiempo(Año)",
     ylab="Ventas Diesel (MT)")
grid()
legend("top", legend=c("Ventas Reales", "Pronóstico"),
       col=c("black", "blue"), lty=1:1, cex=0.6)

```

```
pronostico<- forecast(ajuste5,12,level=95)
```

```

plot(pronostico, main="Pronóstico con Arima (3,0,0)x(0,1,1)12",
     xlab="Tiempo(Año)",
     ylab="Ventas "Diesel"(MT)",ylim=c(0,250),xlim=c(2019,2020))
lines(fitted(ajuste5), col='blue')
lines(ventas.tsdc,col="red")
grid()
legend("top", legend=c("Pronóstico "Diesel"", "Ventas "Diesel""),
      col=c("blue", "red"), lty=1:1, cex=0.6)

pronostico<- forecast(ajuste4,12,level=95)

plot(pronostico, main="Pronóstico con Arima (0,1,2)",
     xlab="Tiempo(Año)",
     ylab="Ventas "Diesel"(MT)",ylim=c(0,300),xlim=c(2014,2020))
lines(fitted(ajuste4), col='blue')
lines(ventas.tsdc,col="red")
grid()
legend("top", legend=c("Pronóstico "Diesel"", "Ventas "Diesel""),
      col=c("blue", "red"), lty=1:1, cex=0.6)

matriz.pronosticos <-data.frame(pronostico$mean,pronostico$lower,pronostico$upper)
matriz.pronosticos

sum(pronostico$mean)
sum(ventas$seriesdiesel[61:72])/1000

```

7.5. Resumen de Resultados

No.	Serie Temporal	Descripción	AIC	BIC	MAPE	Modelo Seleccionado
1	Ventas Totales	ARIMA (1,0,0)	672,72	679,00	15,90	No
2	Ventas Totales	ARIMA (2,0,0)	654,58	662,96	13,28	No
3	Ventas Totales	ARIMA (2,1,1)	641,89	650,20	13,23	No
4	Ventas Totales	ARIMA (2,0,0)x(1,1,1) ₁₂	538,00	547,36	11,67	Si
5	Fuera de Borda	ARIMA (1,0,0)	410,05	416,34	20,33	No
6	Fuera de Borda	ARIMA (3,0,0)	413,78	424,26	20,03	No
7	Fuera de Borda	ARIMA (0,1,1)	406,42	410,57	18,62	No
8	Fuera de Borda	ARIMA (0,0,0)x(1,1,1) ₁₂	346,23	349,98	14,97	Si
9	Gasolina	ARIMA (1,0,0)	434,24	440,51	30,21	No
10	Gasolina	ARIMA (2,0,0)	436,21	444,59	30,24	No
11	Gasolina	ARIMA (0,1,1)	428,04	432,20	30,31	No
12	Gasolina	ARIMA (0,0,0)x(0,1,1) ₁₂	357,17	360,91	21,08	Si
13	Diésel	ARIMA (2,1,0)	580,47	586,70	26,46	No
14	Diésel	ARIMA (3,1,0)	576,03	584,34	25,07	No
15	Diésel	ARIMA (0,1,1)	580,13	584,29	27,73	No
16	Diésel	ARIMA (0,1,2)	577,45	583,69	21,08	No
17	Diésel	ARIMA (3,0,0)x(0,1,1) ₁₂	488,73	498,09	24,69	Si
18	Diésel	ARIMA (3,0,0)x(0,1,2) ₁₂	490,47	501,70	24,84	No