

**Universidad de Valparaíso
Facultad de Ingeniería
Escuela de Ingeniería Civil Industrial**



**Pronóstico y análisis la demanda del Retail de las
bombas hidráulicas: Caso empresa Distribuidora
Koslan**

Por

Patricio Andrés Viveros Avilés

Trabajo de Título Para Optar Al Grado de Licenciado en
Ciencias de La Ingeniería y Título de la Ingeniería Civil
Industrial

Prof. Guía Boris Carikeo Aguilera

Julio 2017

Agradecimientos

A dios por darme la fortaleza espiritual y la salud necesaria para poder seguir adelante y permitirme cumplir con mis objetivos.

A mi familia que estuvieron en todo momento conmigo ofreciéndome su apoyo y que inclusive en los momentos más difíciles me supieron subir el ánimo, logrando darme todas fuerzas para salir adelante. Muy agradecido por el apoyo constante que me dieron y de convencerme que todas las metas que uno se imponga las puede lograr, que nada es imposible.

A mis padres que en todo momento se esmeraron en buscarme estudiar en los mejores institutos educaciones, pero no solamente en donde existieran mejores educadores, sino en donde se inculcarán valores, en ellos aprendí a apreciar más las cosas simples, en respetar a la familia, al compañero de curso independiente de que piense distinto o no, por todo aquello les estaré eternamente agradecido.

A mis amigos más cercanos que siempre nos apoyamos mutuamente al estar viviendo lo mismo en nuestra formación profesional, el compartir con ellos noches de estudios sin a veces incluso no dormir nada con tal de llegar a la meta final. Ellos me enseñaron a que siempre es bueno compartir conocimientos, de ser por sobretodo siempre solidarios, en definitiva aprender a trabajar en equipo, eso me ayudó mucho a ser mejor persona y motivarme a seguir siempre adelante.

A los profesores por la formación que me otorgaron, en especial a mi profesor guía que me ayudó a resolver el presente caso de la mejor manera posible y tuvo harta paciencia conmigo.

Indice

Lista de abreviaturas y siglas.....	6
Índice de figuras	7
Lista de tablas	10
Resumen	12
Introducción.....	14
Capítulo I: Antecedentes de la Industria	16
1. Motivación del estudio	17
2. Aspectos importantes del mercado de bombas hidráulicas en Chile	18
2.1 Definición y clasificación de bombas hidráulicas.....	18
2.2 Industria	19
2.3 Ventas de Koslan	23
2.4 Proveedores de Koslan.....	23
2.5 Detalle de ventas del principal producto.....	25
2.6 Definición de las áreas de negocios	26
2.7 Detalle de contribución por área de negocios	28
Capítulo II: Planteamiento del Problema	30
1. Antecedentes generales e identificación del problema	31
2. Objetivos.....	33
2.1 Objetivo general	33
2.2 Objetivos específicos	33
3. Alcance	34
Capítulo III: Marco Teórico	35
1. Pronósticos.....	36
1.1 Series de tiempo	36
2. Estadístico Durbin Watson (DW).....	43
3. Medición de rendimiento de un pronóstico	45
3.1 Tipos de medidas de error del pronóstico	45
4. Modelos auto regresivos integrados de promedio móvil (ARIMA).....	46
4.1 Definición de los conceptos básicos de los modelos ARIMA	48
4.2 Transformaciones según Box – Cox	48
4.3 Tipos de modelos básicos ARIMA	51

4.4 Metodología Box – Jenkins.....	55
Capitulo IV: Análisis de la Situación Actual	61
1. Situación actual para centrifugas	63
2. Situación actual para agua sucia	64
3. Situación actual para multietapas	65
Capitulo V: Descripción de la Metodologia del Pronostico.....	66
1. Introducción de la metodología	67
Selección de datos por grupo de Koslan	67
2. Metodología Box – Jenkins modificada	68
Capitulo VI: Análisis de Resultados	73
1. Metodología Box – Jenkins modificada	74
1.1 Análisis de datos e identificación del pronóstico centrifugas	74
1.2 Estimación del pronóstico para centrifugas	83
1.3 Validación de pronósticos para centrifugas	85
1.4 Predicción del modelo para centrifugas	88
1.5 Comparación de la situación actual y la propuesta seleccionada.....	91
1.6 Análisis de datos e identificación del pronóstico para agua sucia	92
1.7 Estimación del modelo para agua sucia	99
1.8 Validación del modelo	100
1.9 Predicción del modelo para agua sucia	103
1.10 Comparación situación actual y propuesta seleccionada	105
1.11 Análisis de datos para multietapas	106
1.12 Estimación del modelo para multietapas	113
1.13 Validación de pronósticos para multietapas.....	115
1.14 Predicción del modelo para multietapas	117
1.15 Comparación de la situación actual y propuesta seleccionada para multietapas	120
2. Conclusiones del análisis de resultados	121
Capitulo VII: Evaluación Economica del Modelo Elegido.....	122
1. Inversión	123
2. Costos	124
2.1 Costos FOB y CIF	124
3. Ingresos	125
4. Flujos de caja	126

5. Propuestas	126
5.1 Propuesta 1: Electrobombas centrífugas	126
5.2 Propuesta 2: Electrobombas agua sucia	129
5.3 Propuesta 3: Electrobombas multietapas	131
6. Resumen	133
Capítulo VIII: Conclusiones y Recomendaciones.....	134
Conclusiones	135
Recomendaciones	136
Bibliografía.....	138
Anexo	140
A Distintos análisis para electrobombas Centrífugas	140
A1 Análisis de distribución normal	140
A2 Análisis de correlogramas	141
A3 Análisis de correlogramas residuales	142
B Electrobombas agua sucia	144
B1 Análisis de normalidad de datos.....	144
B2 Análisis de los correlogramas.....	148
B3 Análisis de gráficos residuales	149
C Producto electrobombas multietapas	151
C1 Análisis de los correlogramas.....	151
C2 Análisis de gráficos residuales	152
D Diagrama causa – efecto.....	153
D1 Ishikawa caso: problemas de stock empresa Koslan.....	153

Lista de abreviaturas y siglas

ACF = Valor de función de autocorrelación
 AR = Autorregresivo
 ARIMA = Autorregresivos integrados medias móviles
 ARMA = Autorregresivos medias móviles
 CIF = Cost Insurance and Freight (Costo seguro y flete)
 CLP = Peso Chileno
 CTE = Constante
 DIST = Distribución
 DW = Durbin Watson
 EUR = Euro
 EXP = Exponencial
 FAC = Función de autocorrelación simple
 FACP = Función de autocorrelación parcial
 FOB = Free on bord (Libre a bordo)
 Hr = Hora
 IVA = Impuesto al valor agregado
 LBQ = Estadístico Ljung – Box
 LEAD TIME = Tiempo de llegada de un producto
 LIC = Límite de control inferior
 LSC = Límite de control superior
 MA = Media móvil
 MAE = Mean Absolute Error (Error Absoluto Medio)
 MAPE = Mean Absolute Percentage Error (Error Porcentual Absoluto Medio)
 RB = Ruido blanco
 RETAIL = Venta al detalle
 RMSE = Root Mean Squared Error (raíz error media cuadrática)
 SKU = Stock-keeping unit (unidad de mantenimientos de existencias)
 SSE = Sum of Squared Errors (Suma del Error Cuadrático)
 TIR = Tasa interna de retorno
 USD = Dólar
 VAN = Valor actual neto
 VAR = Varianza

Índice de figuras

Figura 1. 1 Funcionamiento de una bomba de pozo.....	18
Figura 1. 2 Clasificación general de bombas hidráulicas.....	19
Figura 1. 3 Industria de bombas hidráulicas en Chile.....	20
Figura 1. 4 Principales distribuidores de bombas hidráulicas en Chile.....	20
Figura 1. 5 Industria de bombas hidráulicas Mercado directo de Koslan.....	21
Figura 1. 6 Importaciones Competencia directa de Koslan.....	22
Figura 1. 7 Ventas anuales desde el 2015 hasta el 2014 de Koslan.....	23
Figura 1. 8 Contribución de los principales proveedores durante 2014.....	24
Figura 1. 9 Ventas anuales de Pedrollo.....	25
Figura 1. 10 PKM 60.....	26
Figura 1. 11 Esquema de las áreas de negocios de Koslan.....	27
Figura 1. 12 Grafico de la contribución por división de negocios 2014.....	28
Figura 2.1 Diagrama de procesos de abastecimiento desde Europa a Chile.....	32
Figura 3.1 Diagramas de dispersión por patrón.....	39
Figura 3.2 Ejemplos de gráficos de FAC y FACP.....	42
Figura 3.3 Ejemplo de una no correlación o independencia.....	43
Figura 3.4 Ejemplo de una correlación negativa o positiva.....	44
Figura 3.5 Ejemplo de una correlación negativa o positiva indefinida.....	44
Figura 3.6 Ejemplos de graficas de dispersión de datos con homocedasticidad y heterocedasticidad.....	50
Figura 3.7 Demostración de una serie con varianza constante.....	51
Figura 3.8 Ejemplos de correlogramas de modelos AR(1).....	55
Figura 3.9 Ejemplos de correlogramas de modelos AR(2).....	56
Figura 3.10 Ejemplos de correlogramas del modelo MA(1).....	56
Figura 3.11 Ejemplos de correlogramas del modelo MA(2).....	57
Figura 3.12 Ejemplos de correlogramas del modelo ARMA(1,1).....	58
Figura 3.13 tabla resumen de los correlogramas de los pronósticos ARIMA más comunes.....	59
Figura 4.1 Comparación serie suavizada y la situación actual para centrifugas (2011 – 2014).....	63
Figura 4.2 Comparación serie suavizada y la situación actual para agua sucia (2011 – 2014).....	64
Figura 4.3 Comparación serie suavizada y la situación actual para multietapa (2011 – 2014).....	65
Figura 5.1 Grafico de la contribución por grupo de electrobombas de distribución.....	68
Figura 5.2 Diagrama de flujos de la metodología utilizada para pronosticar.....	71
Figura 5.3 Diagrama de flujos de la metodología de análisis de datos.....	72
Figura 6.1 Diagrama de control para centrifugas.....	74

Figura 6.2 Diagrama comparativo entre la serie original y la regresionada.....	76
Figura 6.3 Comparación de la serie original y la suavizada.....	76
Figura 6.4 Comparación de rectas de regresiones lineales.....	77
Figura 6.5 Histograma de distribución de la normal de la serie suavizada aplicando primera diferencia para centrifugas.....	78
Figura 6.6 Grafico rango – media serie suavizada para centrifugas.....	79
Figura 6.7 FAC serie suavizada aplicada primera diferencia estacional para centrifugas.....	81
Figura 6.8 FACP serie suavizada aplicada primera diferencia estacional para centrifugas.....	81
Figura 6.9 FAC residual SARIMA (0,0,0)x(0,1,1) ₁₂ para centrifugas.....	85
Figura 6.10 FACP residual SARIMA (0,0,0)x(1,1,0) ₁₂ para centrifugas.....	85
Figura 6.11 Grafico comparativo entre la serie suavizada y pronostico SARIMA (0,0,0)X(0,1,1) ₁₂ para centrifugas.....	89
Figura 6.12 Grafico comparativo del promedio mensual de la serie y la proyección de los últimos cuatro años para centrifugas.....	90
Figura 6.13 Grafico comparativo del promedio anual de la serie suavizada y el pronóstico SARIMA (0,0,0)x(0,1,1) ₁₂ para centrifugas.....	90
Figura 6.14 Diagrama de control para agua sucia.....	92
Figura 6.15 Diagrama comparativo entre la serie original y regresionada para agua sucia.....	93
Figura 6.16 Comparación de la serie original con la serie suavizada para agua sucia.....	94
Figura 6.17 Comparación de rectas aplicadas a la serie original y suavizada para agua sucia.....	94
Figura 6.18 Histograma de distribución normal de la serie suavizada para agua sucia.....	95
Figura 6.19 Grafico rango – media de la serie suavizada para agua sucia.....	96
Figura 6.20 FAC de la serie suavizada y transformada con raíz cuadrada y primera diferencia para agua sucia.....	97
Figura 6.21 FACP de la serie suavizada y transformada con raíz cuadrada y primera diferencia para agua sucia.....	97
Figura 6.22 FAC residual de ARIMA (0,1,1) para agua sucia.....	101
Figura 6.23 FACP residual de ARIMA (0,1,1) para agua sucia.....	101
Figura 6.24 Grafica comparativa entre la serie suavizada y el pronóstico ARIMA (0,1,1) para agua sucia.....	104

Figura 6.25 Grafico comparativo del promedio mensual de la serie y la proyección en los últimos cuatro años para agua sucia.....	104
Figura 6.26 Grafico comparativo del promedio anual de la serie suavizada y el pronóstico ARIMA (0,1,1) para agua sucia.....	105
Figura 6.27 Diagrama de control para multietapas.....	106
Figura 6.28 Diagrama comparativo entre la serie original y regresionada para multietapa.....	107
Figura 6.29 Comparación de la serie original con la serie suavizada para multietapa.....	107
Figura 6.30 Comparación de rectas aplicadas a la serie original y suavizada para multietapa.....	108
Figura 6.31 Histograma de distribución normal de la serie suavizada para multietapa.....	109
Figura 6.32 Grafico rango – media de la serie suavizada para multietapa.....	110
Figura 6.33 FAC de la serie suavizada y aplicando primera diferencia para multietapa.....	111
Figura 6.34 FACP de la serie suavizada y aplicando primera diferencia para multietapa.....	111
Figura 6.35 FAC residual de SARIMA (0,0,0)x(0,1,1) ₁₂ para multietapa.....	115
Figura 6.36 FACP residual de SARIMA (0,0,0)x(0,1,1) ₁₀ para multietapa.....	115
Figura 6.37 Grafica comparativa entre la serie suavizada y el pronóstico SARIMA (0,0,0)x(0,1,1) ₁₂ para multietapa.....	118
Figura 6.38 Grafico comparativo del promedio mensual de la serie y la proyección en los últimos cuatro años para multietapa.....	119
Figura 6.39 Grafico comparativo del promedio anual de la serie suavizada y el pronóstico SARIMA (0,0,0)x(0,1,1) ₁₂ para multietapa.....	119

Lista de tablas

Tabla 1.1 Detalle de las ventas del producto más vendido de Koslan.....	26
Tabla 3.1 Comparación de las distintas herramientas predictivas.....	47
Tabla 3.2 Tipos de transformaciones Box – Cox.....	49
Tabla 4.1 Estadísticos de error de la situación actual de la empresa para centrifugas.....	61
Tabla 4.2 Estadísticos de error de la situación actual de la empresa para agua sucia.....	62
Tabla 4.3 Estadísticos de error de la situación actual de la empresa para multietapa.....	63
Tabla 6.1 límites de control máximo - mínimo y estimadores.....	73
Tabla 6.2 Funciones de regresión lineal recta 1 y 2.....	75
Tabla 6.3 Prueba Shapiro – Wilk Electrobombas centrifugas.....	76
Tabla 6.4 Estadísticos del FAC aplicando primera diferencia estacional.....	80
Tabla 6.5 Estadísticos estimados SARIMA (0,0,0)x(1,1,1) ₁₂ para centrifugas.....	81
Tabla 6.6 Estadísticos estimados SARIMA (0,0,0)x(0,1,1) ₁₂ para centrifugas.....	82
Tabla 6.7 Estadísticos estimados SARIMA (0,0,0)x(1,1,0) ₁₂ para centrifugas.....	82
Tabla 6.8 Comparación de los distintos estadísticos de error por pronóstico en etapa de estimación para centrifugas.....	84
Tabla 6.9 Comparación de los distintos estadísticos de error por pronóstico en etapa de validación para centrifugas.....	85
Tabla 6.10 Comparación de estadísticos entre la situación actual y la propuesta para centrifugas.....	88
Tabla 6.11 Funciones de regresión lineal para serie original y suavizada para agua sucia.....	91
Tabla 6.12 Prueba Shapiro – Wilk para agua sucia.....	92
Tabla 6.13 Estadísticos del FAC aplicando transformación raíz cuadrada y primera diferencia.....	95
Tabla 6.14 Estadísticos estimativos ARIMA (0,1,1) para agua sucia.....	96
Tabla 6.15 Estadísticos estimativos ARIMA (3,1,0) para agua sucia.....	96
Tabla 6.16 Estadísticos estimativos ARIMA (1,1,1) para agua sucia.....	97
Tabla 6.17 Comparación de los distintos estadísticos de error por pronóstico en la etapa de estimación para agua sucia.....	99
Tabla 6.18 Comparación de los distintos estadísticos de error por pronóstico en la etapa de validación para agua sucia.....	99
Tabla 6.19 Comparación de estadísticos entre la situación actual y la propuesta para agua sucia.....	102
Tabla 6.20 Funciones de regresión lineal para serie original y suavizada para multietapa.....	105
Tabla 6.21 Prueba Shapiro – Wilk para multietapa.....	106
Tabla 6.22 Estadísticos del FAC aplicando primera diferencia estacional.....	109

Tabla 6.23	Estadísticos	estimativos	SARIMA	$(0,0,0)x(1,1,0)_{12}$	para multietapa.....	110
Tabla 6.24	Estadísticos	estimativos	SARIMA	$(0,0,0)x(2,1,0)_{12}$	para multietapa.....	110
Tabla 6.25	Estadísticos	estimativos	SARIMA	$(0,0,0)x(3,1,0)_{12}$	para multietapa.....	111
Tabla 6.26	Estadísticos	estimativos	SARIMA	$(0,0,0)x(0,1,1)_{12}$	para multietapa.....	111
Tabla 6.27	Estadísticos	estimativos	SARIMA	$(0,0,0)x(1,1,1)_{12}$	para multietapa.....	111
Tabla 6.28	Comparación de los distintos estadísticos de error por pronóstico en la etapa de estimación para multietapa.....					113
Tabla 6.29	Comparación de los distintos estadísticos de error por pronóstico en la etapa de validación para multietapa.....					114
Tabla 6.30	Comparación de estadísticos entre la situación actual y la propuesta para centrifugas.....					117
Tabla 7.1	Inversión de la herramienta predictiva propuesta para todos los grupos.....					121
Tabla 7.2	Unidades de compra por trimestre para centrifugas.....					123
Tabla 7.3	Costos	por	propuesta	promedio 2011 – 2014	para centrifugas.....	124
Tabla 7.4	Ingresos	por	quiebres	de stock	para centrifugas.....	124
Tabla 7.5	Flujo de caja	promedio	entre el 2011 – 2014	para centrifugas.....		125
Tabla 7.6	Unidades de compra por trimestre para centrifugas.....					126
tabla 7.7	Costos	por	propuesta	promedio 2011 – 2014	para agua sucia.....	126
Tabla 7.8	Ingresos por quiebres de stock para agua sucia.....					127
Tabla 7.9	Flujo de caja	promedio	entre el 2011 – 2014	para agua sucia.....		127
Tabla 7.10	Unidades	de	compra	por trimestre	para multietapas.....	128
Tabla 7.11	Costos	por	propuesta	promedio 2011 – 2014	para multietapas.....	128
Tabla 7.12	Ingresos	por	quiebres	de stock	para multietapas.....	129
Tabla 7.13	Flujo de caja	promedio	entre el 2011 – 2014	para multietapa.....		129
Tabla 7.14	Flujo de caja promedio entre el 2011 – 2014 resumen.....					130

Resumen

La presente tesis tiene como base principal la reducción del problema presentado en la empresa Koslan Ltda, cuya organización está orientada en la distribución de bombas hidráulicas y accesorios por todo Chile. Es líder en la comercialización de bombas orientadas a un nicho de mercado particularmente caracterizado por el área doméstica, civil y agrícola.

Este proyecto se enfocó en el área de gerencia de ventas, específicamente en la sección de negocios de distribución de la empresa Koslan. Sus principales clientes son pertenecientes a pequeñas y grandes ferreterías, más conocido como el sector del *Retail*. La empresa Koslan siempre se ha caracterizado por importar su mercadería, esta cifra representa aproximadamente el 90% del total de sus productos, gran parte de aquellos provenientes desde Italia (sector reconocido a nivel mundial por poseer los mejores estándares de trabajo y calidad), lo que la posiciona como una empresa que oferta productos de calidad al mejor precio, acá se manifiesta un problema que no se ha podido solucionar, debido a que no existe una respuesta inmediata cuando la demanda es mayor al stock, lo que conlleva a importantes aumentos en los costos asociados, que provocan pérdidas de negocios vitales.

Koslan no posee una metodología estricta de análisis de la demanda, su condición actual es basarse solo en el juicio intuitivo del experto generando reportes simples a través de promedios de los años anteriores, lo que implica el aumento en los tiempos de respuesta a los clientes del *Retail*, mayores costos por sobre-abastecimiento innecesarios e inclusive la pérdida en algunos casos de sus clientes.

Para resolver el actual problema la técnica predictiva tiene que cumplir los siguientes requisitos: primero; ser un modelo que integre todas las variables del entorno, que elimine aquellas que distorsionan la serie de tiempo y considere las que sirvan, ósea, que analice los datos, segundo; al momento de simular lo más aconsejable es presentar un modelo multifuncional, es decir, que presente la opción de aplicar varios parámetros según el caso. En consecuencia el objetivo es seleccionar aquel que simule mejor la realidad y a su vez tenga resultados más precisos y exactos en comparación a los otros.

En definitiva se decide optar por seleccionar el modelo ARIMA bajo metodología Box – Jenkins para cada grupo de productos. Por lo tanto el objetivo general de este proyecto es pronosticar y analizar la demanda de los grupos de modelos más importantes del *Retail*, entendiéndose como los más importantes los que alcancen en conjunto el 60% del peso total por contribución.

Previo a la implementación de la metodología se seleccionaron los productos más rentables de la empresa a través de un análisis ABC, posteriormente se aplicó la metodología que consistió primero en analizar los datos e identificar los posibles modelos predictivos, a continuación se estimaron los distintos modelos identificados con el fin de reducir su complejidad en caso de requerirlo, para luego finalizar con la validación los distintos modelos de pronósticos estimados con el objetivo de comprobar cuál o cuáles eran los adecuados para

cada caso y elegir el que se adapte mejor al comportamiento general de la serie de tiempo, ósea, el que obtenga menores resultados en sus estadísticos de error. Al momento de elegir el mejor pronóstico se procedió a predecir los datos con un horizonte de 12 meses.

Al identificar la demanda futura, se pudo planificar de mejor manera la gestión de inventario de acuerdo a la necesidades de la empresa, a través de una mejor coordinación con los proveedores, una mayor retroalimentación de la información ya disponible con el departamento de abastecimiento y logística, lo que conllevará a mejorar los tiempos de respuesta de la demanda, disminuir los costos derivados por conceptos de transporte, controlar el abastecimiento por tipo de bomba hidráulica y disminuir los quiebres de stock.

Además con el modelo encontrado se realizará una inversión en la implementación con el fin de justificar su rentabilidad frente a la situación actual.

Introducción

A lo largo de la historia los negocios orientados a la comercialización de algún producto o servicio, al momento de tomar decisiones, se respaldaban en la sola intuición del experto, dejando de lado las variables o factores que provocaban cambios constantemente, esta incertidumbre impactaba directamente en todas las operaciones de las empresas. Durante el transcurso del tiempo las empresas adquirieron una vasta experiencia, permitiendo desarrollar mecanismos que de cierta manera lograron disminuir el factor de incertidumbre, es así como aprendieron a respaldar sus decisiones en algo más concreto que una simple corazonada o intuición, siendo la gran respuesta para aquello el desarrollo de pronósticos concretos y precisos, que no garantizan la eliminación total de la incertidumbre, pero si disminuirla en un gran porcentaje.

En la actualidad cada día existe una mayor presión por parte de las empresas por utilizar de mejor manera las herramientas de pronósticos. Para ello, las empresas deben saber utilizar la información con la que disponen, mediante la correcta selección de datos y revisión del pronóstico más adecuado, con el fin de poder determinar de forma más precisa las ventas que se generarán y tomar las mejores decisiones.

En el caso particular de la industria de bombas hidráulicas, los factores que provocan cambios en la demanda son impredecibles, por lo que se justifica mayor aun de la utilización de un método predictivo que encajé de la mejor forma con la planificación de la demanda. La empresa de distribución Koslan es un claro ejemplo de lo expresado, (que actualmente no posee una planificación adecuada de ventas), necesita producir algún tipo de pronóstico, que le permita obtener mayor beneficios.

En el capítulo uno se describirán los antecedentes de la industria, permitiendo otorgar información relevante de la empresa como: definición y clasificación de las bombas hidráulicas, principales competidores, ventas históricas de los productos de Koslan, principales proveedores y clientes por área de negocios, seleccionando en este último a los clientes de distribución (*Retail*) como el área en donde presente el problema de abastecimiento. Además se facilitan cifras importantes en porcentajes como: de mercado directo e indirecto de la empresa, de contribución por área de negocios (perfil del cliente) y de proporción por grupos de productos de distribución, llegando a seleccionar el conjunto de grupos que represente el 60% de la totalidad de los clientes del *Retail*.

El capítulo dos se enfocará en la formulación del problema de la empresa, exponiendo su condición actual de trabajo, demostrando cuál es su causa raíz, manifestando en donde se origina esta falencia y que impacto provoca. Además en base a lo anterior se enunciará el objetivo principal y los secundarios.

En el capítulo tres se definirá el marco teórico, detallando conceptos básicos como: que es un pronóstico, que es una serie de tiempo y cuáles son los componentes que intervienen en una serie. Para posteriormente describir las diferentes herramientas predictivas, definir cuáles serán los distintos estadísticos para analizar los datos, formular en qué parámetros de cálculos de error se enfocará cada pronóstico, para finalizar justificando cual será el mejor método de pronóstico para la condición actual (este proyecto se enfocará en métodos ARIMA).

En el capítulo cuatro se dará a conocer la situación actual de la empresa, exponiendo la herramienta utilizada para planificar las ventas anuales.

El capítulo cinco describirá detalladamente la metodología en la que se enfocará el cálculo de la herramienta predictiva, básicamente esta se basará en la metodología Box – Jenkins [Hanke&Wichern10], pero con una pequeña modificación con el objetivo de adaptarla mejor a la condición actual del caso.

En el capítulo seis se detallará un análisis completo de cada uno de los resultados por grupo de bombas hidráulicas, abarcando desde el análisis de la serie de tiempo original por grupo hasta la selección de la herramienta de pronósticos más adecuada para cada perfil. Además se desarrollará una evaluación económica a todo el trabajo propuesto en esta tesis, para demostrar que la metodología impuesta es la mejor opción para Koslan.

En el capítulo siete se presentará las evaluaciones económicas de la propuesta ARIMA seleccionada por grupo frente a la situación actual aplicada por la empresa, con el objetivo de demostrar que tan rentable es la proyección utilizando metodología Box – Jenkins.

En el capítulo ocho se enfocará en el desarrollo de conclusiones y recomendaciones, con el fin de comprobar lo eficiente y eficaz que será implementar este tipo de pronósticos a la organización estudiada.

Capítulo I: Antecedentes de la Industria

1. Motivación del estudio

La empresa Koslan es una organización con más de 20 años de vida, se ha dedicado siempre a la distribución de bombas hidráulicas y accesorios, abarcando distintos industrias de mercado, como el rubro del área doméstica, civil y agrícola, con posibilidad de entrar a mediano plazo al mercado industrial, para aquello necesitarán de grandes cambios logísticos, ya que las dimensiones volumétricas que poseen estas bombas son bastante mayores, además de ser mayormente rigurosos con los estándares de calidad, ya que los clientes industriales son principalmente provenientes de las mineras.

Aproximadamente el 90% de sus productos los importa, dentro de las importaciones la gran mayoría son provenientes desde Europa y en un porcentaje menor desde China, entre las marcas más importantes están: Pedrollo (Italia), Zenit (Italia), Rovatti (Italia) y Daishin (China).

Además la organización posee distintas áreas de negocios, estas se distribuyen en; distribución, grandes tiendas, especialistas y servicio técnico.

Las bombas hidráulicas en particular cumplen la condición de ser un producto muy diverso, Koslan tiene la respuesta a una amplia gama de modelos que cumplen con la mayoría de los requerimientos de los clientes, el problema está cuando la demanda sobrepasa el stock disponible, es ahí en donde hay que encontrar un método de pronósticos de demanda, que permita planificar las futuras compras.

2. Aspectos importantes del mercado de bombas hidráulicas en Chile

2.1 Definición y clasificación de bombas hidráulicas

Antes de describir el mercado de este producto, hay que definir a las “bombas hidráulicas como aquel que transforma la energía mecánica en energía fluida o hidráulica” (Zubicarag79), cuya función principal es el transporte del fluido que puede ser agua o mezcla del líquido con otros sólidos, desde una zona de menor presión o altura a una que le proporcione una mayor presión o altura.

El transporte del fluido se inicia desde una fuente de origen en donde aspira la bomba, hasta la descarga del fluido en un estanque hidráulico, que almacena el agua y la descarga para el uso del cliente. Su aplicación puede ser por ejemplo para para regadío, para el consumo de agua a nivel doméstico, para la industria minera, entre las principales. La figura 1.1 es un claro ejemplo lo anteriormente expresado.

Figura 1.1 Funcionamiento de una bomba de pozo



Fuente: Imagen extraída de un video demostrativo, Ministerio de la producción, Calama, 2011

A modo general las bombas se pueden dividir en dos grandes grupos dependiendo de la aspiración en su fuente origen, estas pueden ser bombas superficiales, sumergible y bombas subterráneas de pozo.

Las bombas superficiales, como su nombre lo dice solo sirven para aspirar agua desde zonas provenientes de la superficie, cuya fuente de succión es en ríos, vertientes, lagos, mientras que las bombas sumergibles, aspiran agua bajo tierra con una profundidad máxima desde la superficie de 7 metros y las bombas subterráneas de pozo, aspiran agua bajo profundidades que pueden sobrepasar los 100 metros de profundidad.

En la figura 1.2 se demuestran unos ejemplos de la clasificación de bombas antes mencionadas, son de la marca Pedrollo, de procedencia Italia, el principal proveedor de Koslan.

Figura 1.2 Clasificación general de bombas hidráulicas



Fuente: Elaboración propia

2.2 Industria

La industria de bombas hidráulicas se basa en la comercialización de una gran variedad de bombas hidráulicas, en términos generales este mercado se puede dividir en las siguientes industrias: domestico, civil, agrícola y minero.

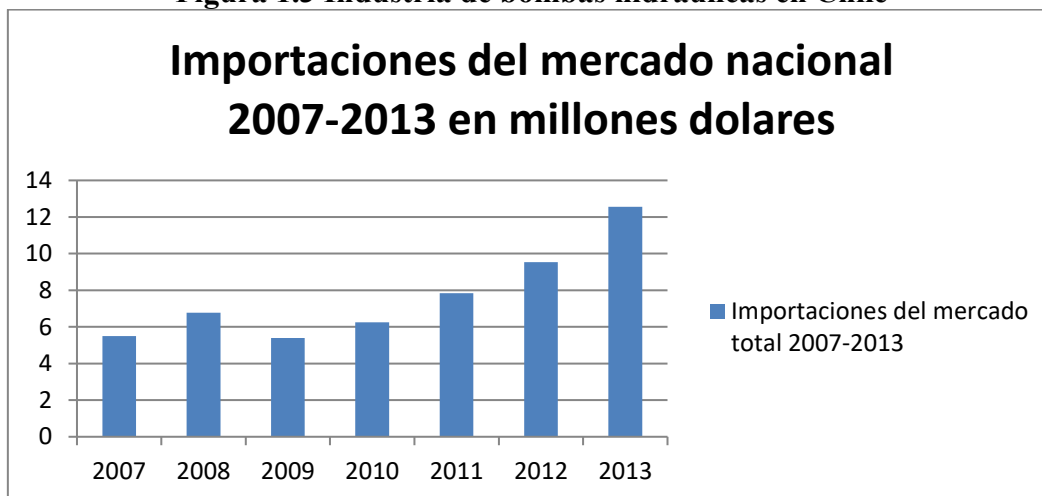
Koslan hoy en día está inmerso en la venta de bombas a nivel doméstico, civil y agrícola, con cierto interés en entrar en la industria minera.

2.2.1 Comportamiento del mercado de bombas hidráulicas a nivel nacional

El mercado a describir abarca desde las bombas más básicas consideradas dentro del mercado doméstico, hasta las más complejas consideradas dentro del mercado industrial, es decir, la totalidad de bombas hidráulicas de Chile. Este análisis abarca el comportamiento del mercado de los últimos 7 años.

En la figura 1.3 se registran los datos de todas las importaciones anuales del país. Se puede apreciar que el negocio de bombas hidráulicas en los últimos 7 años ha tenido un crecimiento de un 43%, llegando a obtener importaciones de 12, 5 millones de dólares, por lo que se concluye que en el mercado ha reflejado alzas sostenibles en el tiempo.

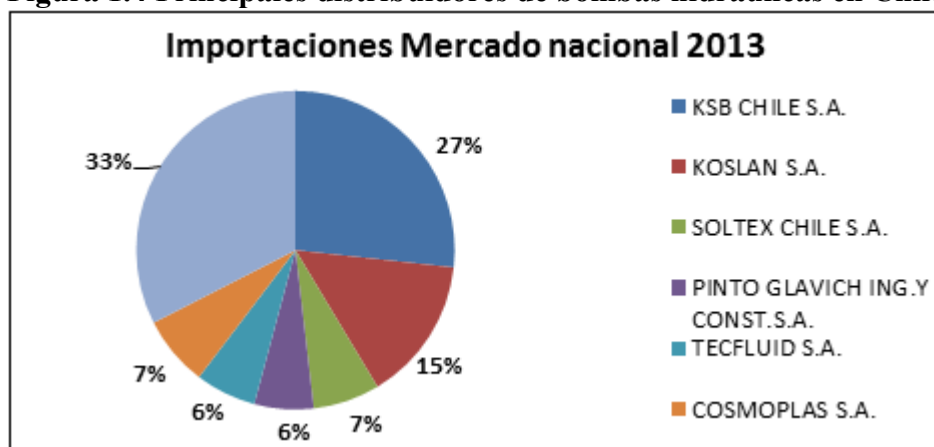
Figura 1.3 Industria de bombas hidráulicas en Chile



Fuente: Elaboración propia, información obtenida del portal de comercio exterior (www.portalcomexccs.cl/sitio/)

A continuación se presenta el listado de competidores a nivel nacional de Koslan, entre ellos están: KSB Chile S.A, Soltex Chile S.A, Cosmoplas, Tecfluid S.A, PGIC y otros.

Figura 1.4 Principales distribuidores de bombas hidráulicas en Chile.



Fuente: Elaboración propia, información obtenida del portal de comercio exterior (www.portalcomexccs.cl/sitio/)

En este gráfico (figura 1.4) se puede ver que KSB tiene la mayor participación del mercado con un 27%, seguido por Koslan con un 15%, esta diferencia de porcentaje radica en que KSB es una empresa que ofrece productos con capacidades y volúmenes mucho mayores, además abarca un mayor mercado que el Koslan, principalmente el industrial, dejándolo fuera de la clasificación como competidor directo.

Como competidores directos se encuentran principalmente PGIC Y Cosmoplas, estos tienen un porcentaje de participación mucho menor, un 6% y 7% respectivamente, lo cual demuestra el liderazgo obtenido por parte de Koslan en su mercado, esto quedará más claro en la figura 1.5.

La participación de mercado de Koslan no deja de ser importante, ya que con respecto al año anterior 2012, su participación era de un 8%, ósea, prácticamente se duplicó.

2.2.2 Comportamiento del mercado directo de Koslan

El mercado de bombas hidráulicas directo para Koslan es el de venta de productos que abarcan principalmente el sector doméstico, civil y agropecuario. El sector doméstico y civil están enfocados en clientes ubicados en zonas urbanizadas, el primero atiende clientes de condominios o villas, mientras que el segundo se enfoca en clientes mayoritariamente de pequeñas y grandes edificaciones u obras civiles y el sector agropecuario se focaliza en atender a pequeños y grandes clientes agricultores y ganaderos ubicados en zonas rurales. En el siguiente gráfico (figura 1.5) se demuestra el tamaño de este mercado.

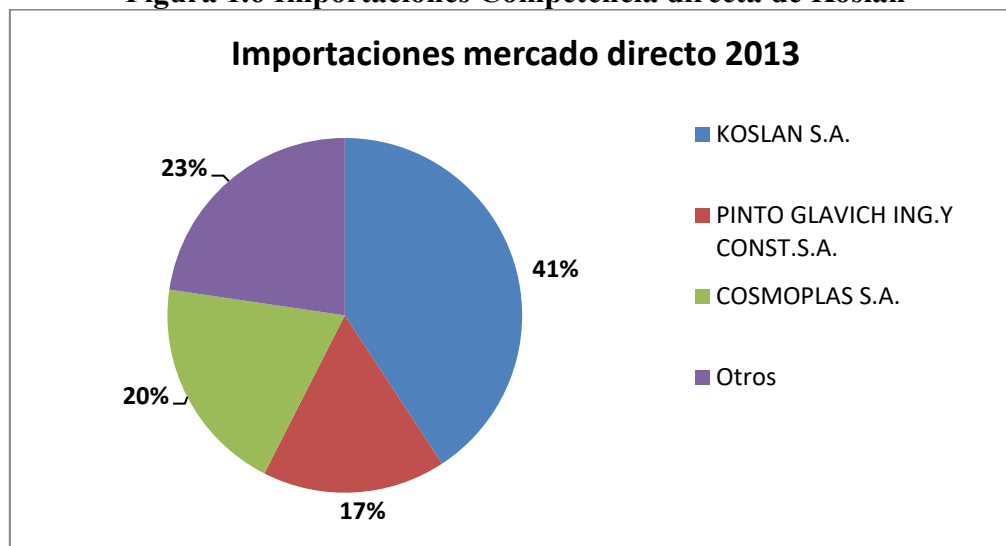


Fuente: Elaboración propia, información obtenida del portal de comercio exterior (www.portalcomexccs.cl/sitio/)

En el gráfico anterior se pudo describir el crecimiento significativo que hubo en el año 2013 llegando a importaciones de casi 5 millones de dólares, prácticamente duplicando el año anterior, para los próximos años se pronostican crecimientos muchos mayores lo que se permite concluir que este mercado está en pleno desarrollo.

Con respecto a sus competidores directos como se describió anteriormente estos son dos: Cosmoplas y PGIC, sumando a los dos junto con Koslan alcanzan el 77% de todas las importaciones del país, siempre enfocadas en el sector del perfil de clientes domésticos, civiles y agropecuarios. En la siguiente gráfica se presenta las importaciones del mercado directo.

Figura 1.6 Importaciones Competencia directa de Koslan



Fuente: Elaboración propia, información obtenida del portal de comercio exterior

En la gráfica anteriormente descrita (figura 1.6) queda demostrado el dominio total que posee Koslan en este mercado, alcanzando un 41% de participación, además hay que agregar su notable crecimiento, ya que el año anterior alcanzaba una participación en este mercado del 32%.

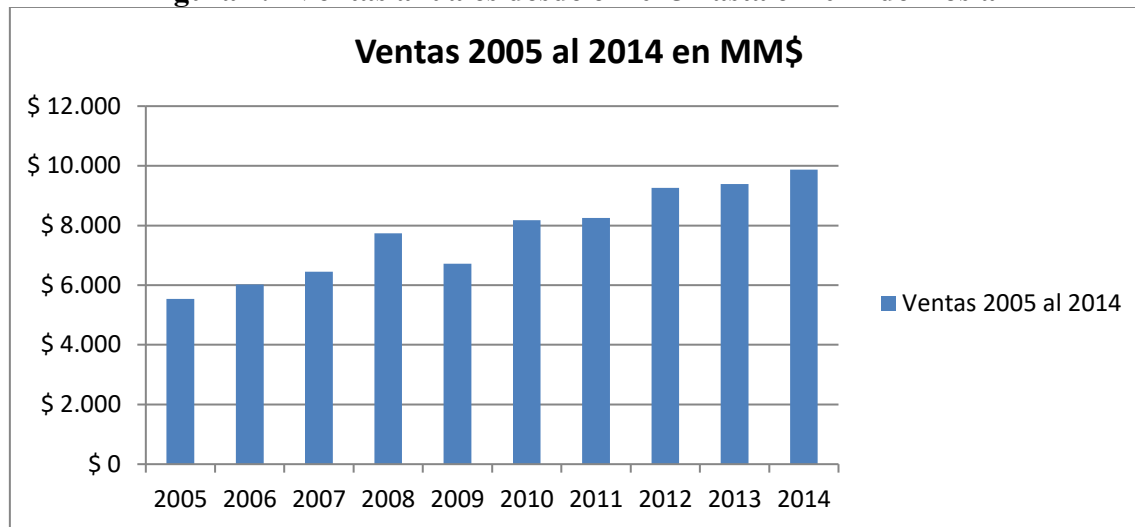
Con respecto a sus competidores si bien tienen participaciones menores no dejan de ser peligrosas sobre todo en el caso de Cosmoplas, ya que el año 2012 tenía una participación del 15% experimento un gran crecimiento que le permitió subir 5 puntos porcentuales.

Estas 3 empresas al tener el dominio de mercado comparten por ende la misma estrategia, la cual es repartir el 23% restante de clientes no abarcados por ellos, esto porque el año 2012 entre las 3 sumaban un porcentaje de mercado del 69% obteniendo un alza de 8 puntos porcentuales, por lo tanto es ahí en donde se centra la principal disputa.

2.3 Ventas de Koslan

En la figura 1.7 se describe el comportamiento de Koslan en relación a sus ventas anuales registradas desde el 2005 en adelante, estas consideran las ventas de todos los sku's de la empresa, divididas principalmente en: bombas hidráulicas, accesorios y repuestos. Cerca del 80% de las ventas las cubren los distintos tipos de bombas hidráulicas (es en estos productos en donde aplicaremos el pronóstico).

Figura 1.7 Ventas anuales desde el 2005 hasta el 2014 de Koslan



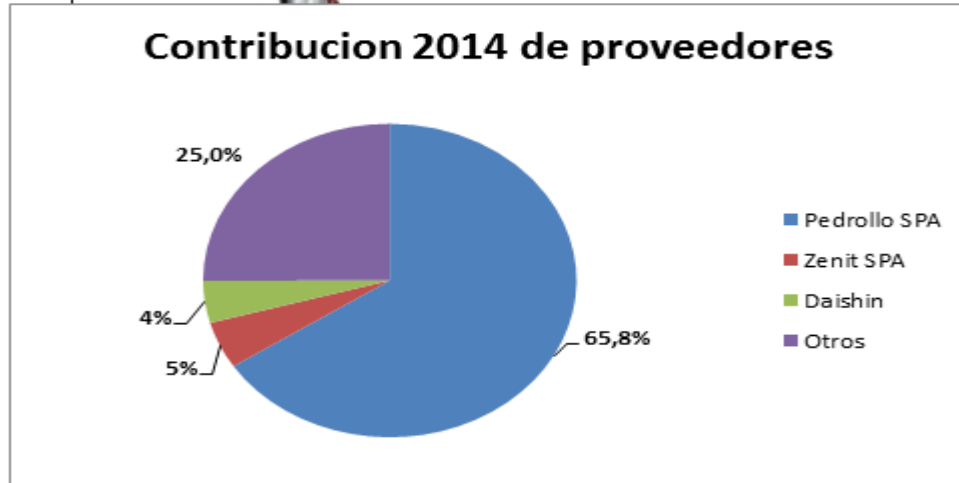
Fuente: Planilla ventas y detalles, Koslan, 2005

Como se puede apreciar las ventas han tenido una tendencia al crecimiento, llegando a obtener ventas en el último año de casi 10 mil millones de pesos, ósea, casi 16 millones de dólares.

2.4 Proveedores de Koslan

Koslan trabaja con una serie de proveedores, su objetivo siempre es y ha sido distribuir productos con los mejores estándares de calidad y servicio (los mejores estándares son los Italianos), por lo mismo su mayor contacto son proveedores Europeos, de procedencia netamente Italiana, como lo son por ejemplo, Pedrollo, Zenit, Rovatti, Varem, Mac 3, estos representan aproximadamente el 80% de sus importaciones, el otro resto del porcentaje lo tiene en gran parte Daishin, siendo un producto Chino sale de la tendencia u objetivo normal, pero esto se debe solamente porque Koslan ha sabido posicionar esta línea de productos independiente de su procedencia.

A continuación se expresa un detalle completo de los principales proveedores, justificando porque son los más importados y su división porcentual mediante un gráfico.

Figura 1.8 Contribución de los principales proveedores durante 2014

Fuente: Planilla ventas y detalles Koslan 2015

Los proveedores a considerar dentro de este pronóstico serán los especificados en la imagen 1.8, esto porque representan el 75% de la contribución total de la empresa durante el 2014 y también porque son estos los proveedores que distribuyen gran parte de las bombas hidráulicas a Koslan.

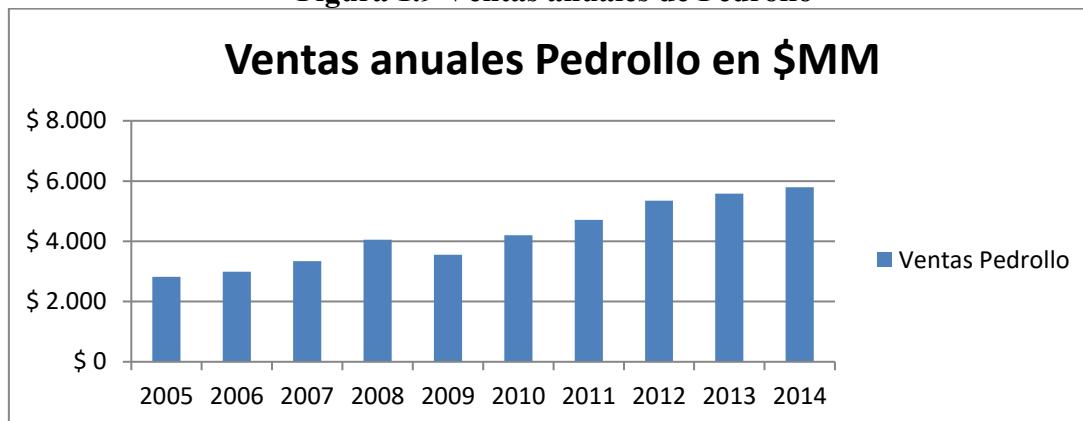
Koslan al ser el único representante exclusivo de Pedrollo por más de 15 años se convierte como el distribuidor estrella dentro de Chile representando el 65% de las ventas totales anuales.

2.4.1 Principal proveedor

Tal como se comentó antes el proveedor principal de Koslan es Pedrollo, empresa de gran prestigio en la fabricación de bombas tanto de superficie como de pozo profundo, de procedencia Italiana, posee lo mayores estándares de calidad a nivel europeo.

Las ventas solamente de Pedrollo llegaron a alcanzar el 2014 casi seis mil millones de pesos, aproximadamente 10 millones de dólares, esto demuestra el peso que es para la empresa la venta de esta marca de productos, en la figura 1.9 queda todo más claro.

Figura 1.9 Ventas anuales de Pedrollo



Fuente: Planilla ventas y destalles Koslan 2015

Este éxito se debe a que Koslan ha posicionado esta marca hace más de 15 años, siendo representante exclusivo de esta marca en Chile, reconocida por su buena calidad y rendimiento, su principal estrategia es que Pedrollo cuenta con personal altamente capacitado que está en constante contacto con sus clientes, además posee la ventaja de conocer el mercado sudamericano, esto debido a que también cuenta con distribuidores en Perú, Uruguay y Brasil, siempre con el objetivo de perfeccionar su producto año a año, lo que le ha permitido obtener estos grandes frutos.

2.5 Detalle de ventas del principal producto

El producto más vendido de Koslan es la bomba hidráulica de superficie PKM 60, es de la marca Pedrollo. La característica de esta bomba es que alcanza una altura máxima de 40 metros columna de agua y un caudal máximo de 40 l/min, utilizando 0,5 hp de potencia, lo cual la hace muy eficiente por el consumo de energía que ocupa, además tiene otra característica que la hace distinta a los demás, ya que es muy pequeña alcanzando una dimensión volumétrica de $0,008 m^3$ y pesando 5 kilos.

Según la siguiente tabla, las ventas netas de este producto siempre han ido en ascenso, sin embargo en estos 2 últimos años el mercado ha experimentado un alza en la competitividad, lo que ha provocado sacrificar los márgenes normales, esto para poder mantener el nivel de crecimiento. En el último año esta electrobomba llegó a experimentar ventas de aproximadamente 1 millón de dólares, lo cual lo hace ser el producto más importante de la compañía.

Tabla 1.1 Detalle de las ventas del producto más vendido de Koslan

Año	Venta neta	Unidades vendidas
2005	\$ 337.319.536	14260
2006	\$ 299.587.422	12763
2007	\$ 316.581.093	13052
2008	\$ 469.834.022	17964
2009	\$ 393.060.383	12782
2010	\$ 463.920.738	15111
2011	\$ 490.281.627	16007
2012	\$ 600.623.065	19680
2013	\$ 588.173.351	18846
2014	\$ 634.265.429	18789

Fuente: Ventas y detalles de Koslan 2015

Figura 1.10 PKM 60

Fuente: Ficha técnica Koslan – Pedrollo 2012

2.6 Definición de las áreas de negocios

Koslan distribución: son todas las ventas generadas hacia los clientes del *Retail*, ósea, todas las ventas generadas a pequeñas y grandes ferreterías, esta abarca el mayor porcentaje de las ventas de Koslan (en la figura 1.12 se detalla), por lo que es el área de negocios más importante de esta empresa.

En esta área las ventas son muy sensibles, ya que estas se concretan en un periodo a corto plazo, con un horizonte de tiempo mes a mes, por lo que existen negociaciones de manera constante y al no tener stock de ciertos productos han provocado graves problemas para la compañía, no solo de perder ventas puntuales sino de incluso perder a clientes valiosos, es en esta área donde se pretende pronosticar la demanda.

Koslan especialistas: son todas las ventas generadas a proyectos específicos como por ejemplo a constructoras, a empresas encargadas de hacer las instalaciones civiles o sanitarias, en definitiva son todas las ventas generadas bajo condiciones específicas.

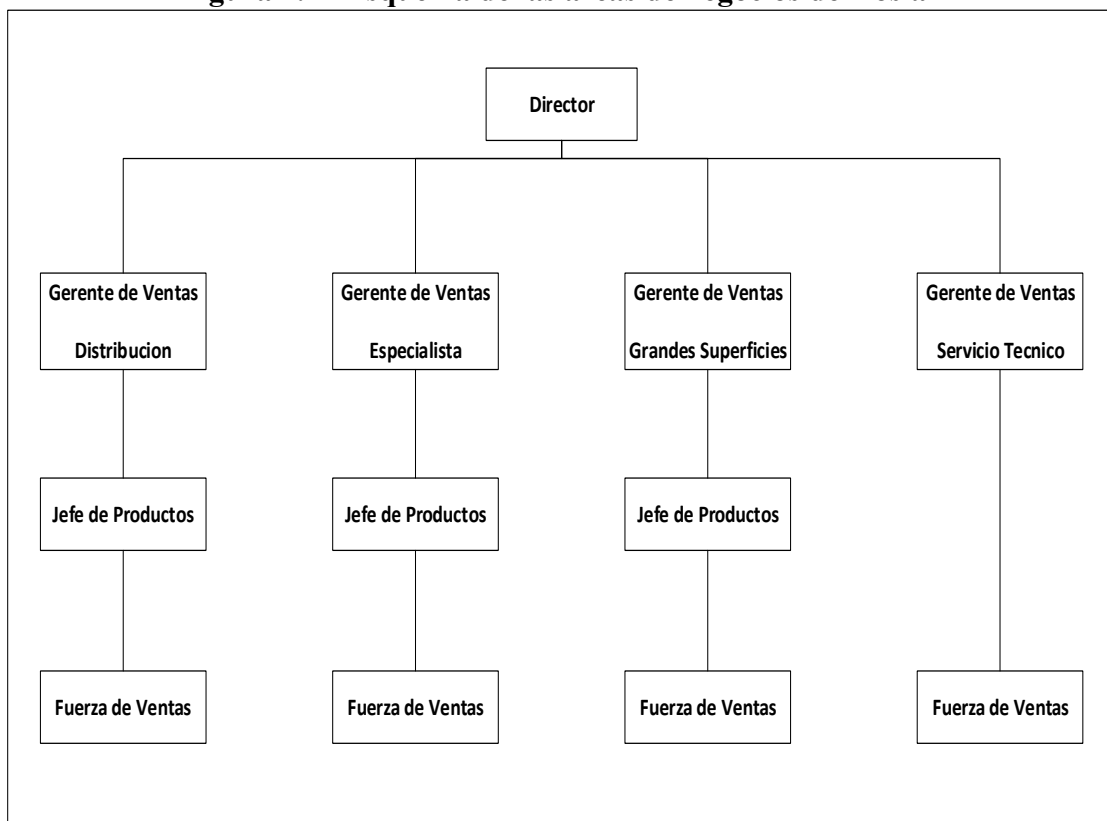
En especialistas el horizonte de negociaciones es a mediano y largo plazo, son muy pocos los casos en donde se concretan negocios a corto plazo, debido a esto la empresa prácticamente no tiene problemas de stock, por lo que ha podido planificar correctamente las compras de abastecimiento.

Koslan grandes superficies: Son todas las ventas relacionados a las grandes empresas como: *Easy, Homecenter* y *Sodimac*, entre las más importantes. En grandes superficies la política de abastecimiento se repite generalmente en el mismo periodo, las ventas se pactan con un horizonte trimestral, son muy pocos los casos en donde se logran mensualmente, esto debido a que el cliente (*Easy, Homecenter y Sodimac*) cuenta con sólidos métodos de pronósticos y una política de compra demarcada y clara.

Servicio técnico: Es todo el servicio pos venta de Koslan, ellos ven todas las mantenciones de los productos, si es que se necesita cambiar o reponer alguna pieza. Al momento de reponer alguna pieza ellos si no cuentan con cantidad suficiente, optan con la solución de desarmar equipos viejos o en el peor (siendo en muy pocas ocasiones) de los casos equipos nuevos.

A continuación mediante un mapa conceptual se presentará la composición de cada una de estas áreas y su estructura jerárquica.

Figura 1.11 Esquema de las áreas de negocios de Koslan



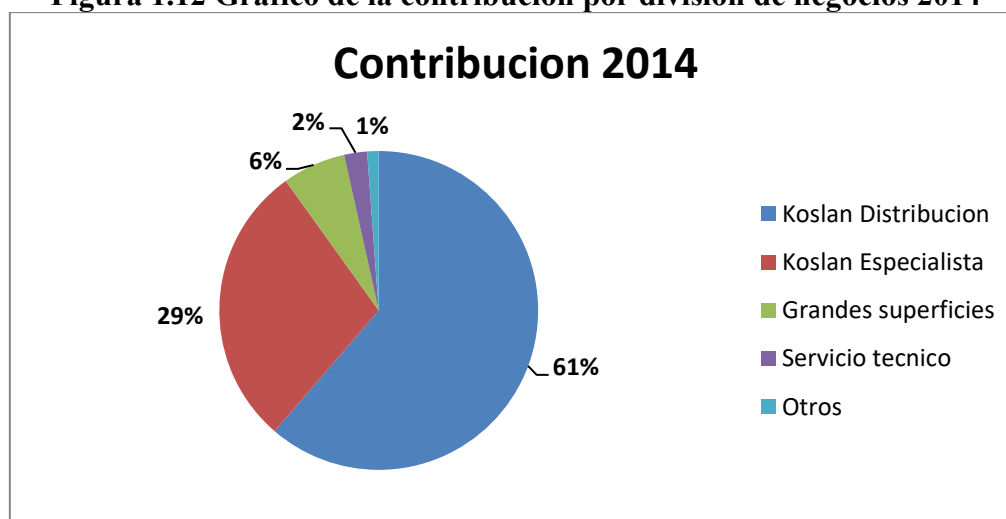
Fuente: Administración de Koslan 2010

Como se describe en la figura 1.11 el área de distribución consta de un gerente de ventas, de un jefe de productos y de 9 vendedores, todos estos están distribuidos en distintas zonas del país, otorgando servicios desde Arica hasta Punta arenas. El área de especialistas consta de un gerente de ventas, de un jefe de productos y de 5 vendedores, estos están distribuidos en la zona central y parte de la quinta región. El área de grandes superficies consta de un gerente de ventas, de un jefe de productos y de un vendedor, que abarca gran parte de la zona central y parte de algunas regiones del país tanto en la zona norte como zona sur de Chile. Y el área de servicio técnico consta de un gerente de ventas y de distintos vendedores de mesón, estos ofrecen ventas de repuestos como también ayuda en servicios post-ventas, distribuidos en todo Chile.

2.7 Detalle de contribución por área de negocios

En la siguiente grafica se describe de manera más clara el peso en contribución real, ósea, lo que realmente queda como utilidad de cada una de las áreas de negocios de Koslan.

Figura 1.12 Grafico de la contribución por división de negocios 2014



Fuente: Ventas y detalles de Koslan 2014

En el grafico anterior (figura 1.12) se puede apreciar la importancia en las ventas del área de negocios de “Koslan distribución” alcanzando un 61% en el peso, además cuenta con los mayores márgenes en promedio en comparación con sus otras áreas. Las demás entidades no dejan de ser importantes, sobre todo “Koslan especialista” que alcanza el 29% de las ventas totales de la compañía, esta como las demás áreas (“Grandes superficies y “Servicio técnico”) tal como se explica anteriormente no cuentan con este problema de falta de stock.

2.7.1 Contribución del área de distribución

En relación a los movimientos del año 2014, en distribución hubo una contribución aproximada de 2,4 mil millones de pesos, de los cuales el 75% corresponde solo a electrobombas el resto se reparte entre accesorios y repuestos. Tal como se describió en los capítulos anteriores, las electrobombas se pueden clasificar de acuerdo a su aplicación o uso en distintos grupos, entre los más importantes se encuentran: Electrobombas centrifugas, sumergibles de agua sucia, y multietapas, la suma de todas ellas abarcan aproximadamente el 65% de la rentabilidad total del área de distribución de la empresa, es por esta causa que se analizarán estos grupos, para poder proyectarlos mediante el mejor modelo de predicción.

Capitulo II:

Planteamiento del Problema

1. Antecedentes generales e identificación del problema

La empresa Koslan Ltda tal como se describió en el capítulo dos es una empresa líder a nivel de país, que en cuanto a ventas por volumen se ha mantenido con una leve alza en los últimos años, esta cifra podría ser mayor si existiera una real planificación de abastecimiento.

La organización importa aproximadamente el 90% de sus productos (principalmente provenientes del continente Europeo de preferencia Italia y otro resto pero en menor cantidad desde China), dentro de este porcentaje cerca del 80% son bombas hidráulicas.

La planificación de abastecimiento comienza desde el momento que se emiten las órdenes de compra, existe un *Lead time* (tiempo de llegada del pedido) que depende de distintos factores, primero puede variar dependiendo del lugar de origen (Europa, Estados Unidos y China), segundo por el tipo de transporte, este puede ser marítimo, aéreo o *currier*, (hay que considerar que la vía marítima es la ideal y recomendado por los costos asociados) y por ultimo por el grado de producción o disponibilidad del proveedor.

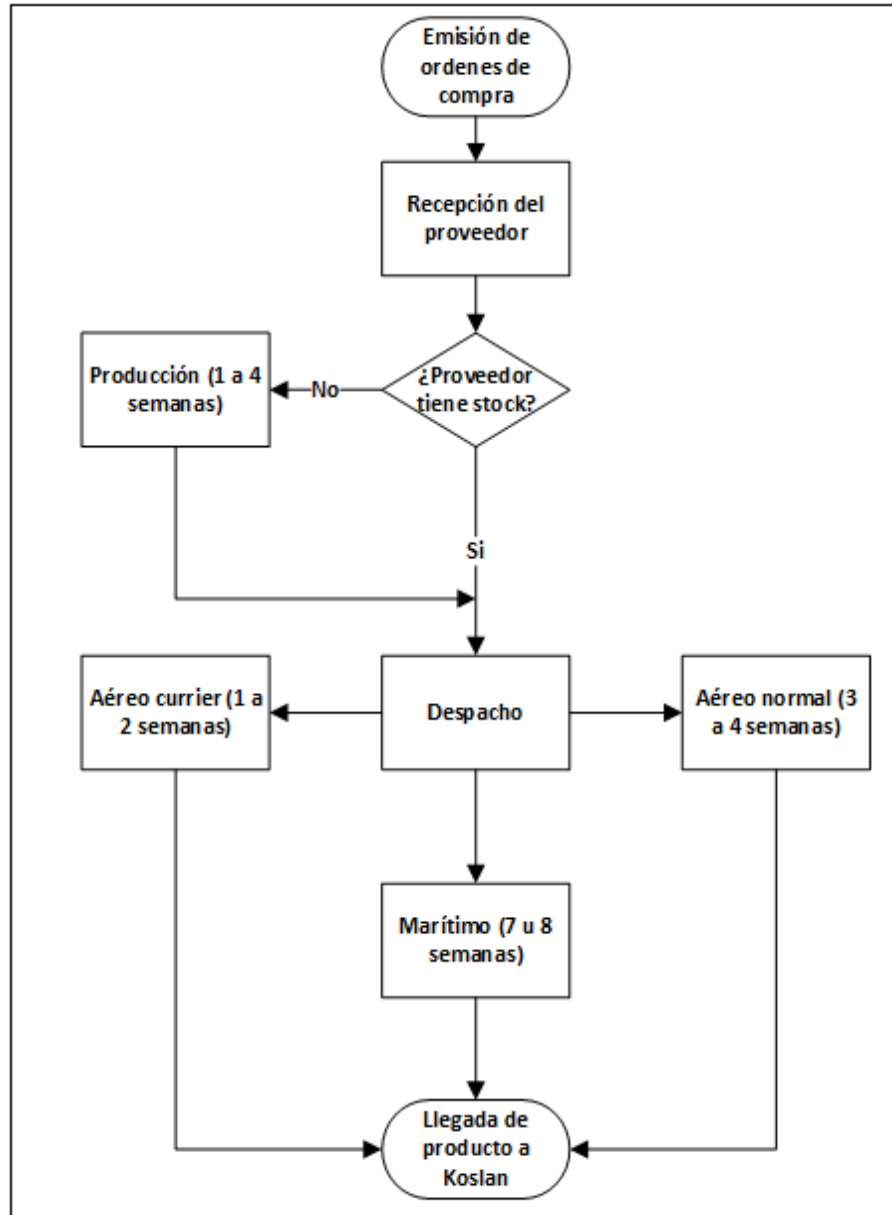
Considerando cualquier origen y transporte vía marítima en este caso la mercadería tiene un *Lead time* de siete (para Europa y Estados Unidos) a diez semanas (China), en caso de existir riesgos de quiebres de stock se opta por vía aérea o *currier*, si es por vía aérea se considera un rango de tres a cuatro semanas y si es por vía *currier* (se hace en menor de los casos por su alto costo) siete a diez días, mientras que si existe un caso desfavorable en donde el proveedor no posee stock con entrega inmediata, este tiempo dependiendo de la disponibilidad del mismo y puede variar entre una a incluso cuatro semanas más, llegando a un *Lead time* total de hasta cuatro meses.

En síntesis considerando que la condición normal y por ende más favorable de la empresa es transportar su mercadería por vía marítima, el *Lead time* de cada producto experimenta un rango entre tres a cuatro meses, es en este punto en donde existen problemas de planificación de compra.

Las compras las realiza el gerente de control y planificación de abastecimiento, estas se materializan una vez al mes, mediante estrategias de demanda que se desarrollan de forma anual planificadas por el área de gerencia de ventas, estas se basan mediante los registros históricos de la demanda del año anterior y calculadas meramente por su juicio intuitivo, el problema radica en que no se consideran ciertas variables importantes como, quiebres de stock históricos, campañas de marketing, alza de precios, comportamiento general del mercado y condiciones de país, lo que se traduce en un plan de demanda muy poco sólido.

En definitiva, en la empresa Koslan específicamente en el área de ventas, no existe ninguna técnica predictiva de demanda que se inspire en una metodología y que sirva como parámetro para planificaciones de ventas. Su estrategia se basa en el juicio intuitivo de parte de la gerencia de ventas, que luego lo discute con la gerencia de control y planificación de abastecimiento, todo con el objetivo de planificar la demanda junto con las ventas que tendrán año a año.

Figura 2.1 Diagrama de procesos de abastecimiento desde Europa a Chile



Fuente: Elaboración propia

En la imagen anterior se describe el proceso completo de compra desde el momento que se libera una orden de compra, este caso es simulando compras desde Europa, que para el presente caso será origen Italia.

Este problema produce:

- Complicaciones de abastecimiento, en situaciones en donde la demanda es mayor a las unidades que existen en stock, causado por las demoras (antes mencionadas) en traer los productos, provocando quiebres de stock y la pérdida de ventas valiosas.
- Aumento en los costos asociados de transporte, esto debido a la mala planificación de compra al solicitar volúmenes de cantidades inadecuadas al proveedor, incurriendo a gastos adicionales por transporte aéreo o inclusive vía *currier* (caso más costoso).
- Nula respuesta adecuada ante eventuales cambios inesperados, esto causado por la mala coordinación de compra, poca solidez de los datos disponibles en los sistemas de información y escasa inversión en el análisis de los productos más rentables.

La metodología aplicada necesaria para detectar el presente problema fue Ishikawa o diagrama causa y efecto, se utiliza este diagrama porque expresa claramente la causa raíz del problema (presentando en el anexo D1).

2. Objetivos

2.1 Objetivo general

- Desarrollar un modelo predictivo de la demanda de los grupos de bombas hidráulicas más preponderantes, mediante el pronóstico más adecuado.

2.2 Objetivos específicos

- Analizar la demanda de las bombas hidráulicas y seleccionar las más relevantes para el negocio.
- Identificar cuáles son los parámetros que influyen en estas bombas hidráulicas elegidas.
- Evaluar la precisión de los distintos métodos de pronósticos y seleccionar el que arroje mejores resultados.
- Evaluar económicamente los resultados del modelo de predicción de demanda desarrollado.

3. Alcance

Es por todo aquello que se pretende implementar en la gerencia de ventas y abastecimiento, particularmente en clientes del área de distribución (*Retail*), un método de pronósticos que pueda integrar todas las variables que afecten en la industria de bombas hidráulicas y poder adaptarlo particularmente a las condiciones de esta organización, el pronóstico a seleccionar del tipo ARIMA inspirado en la metodología Box – Jenkins. Con este método predictivo se busca agilizar los tiempos de respuesta a sus clientes, mediante la generación de un plan de demanda más robusto que permita mejorar los procesos de generación de compra, considerando además un stock de seguridad, que permita en definitiva disminuir quiebres de stock y además optimizar los costos logísticos.

Capítulo III: **Marco Teórico**

1. Pronósticos

Los pronósticos son predicciones del futuro en un tiempo determinado, basados en una serie de datos recopilados en el pasado. Estos datos se proyectan mediante una serie de cálculos estadísticos y matemáticos que representan, lo que podría ocurrir en un futuro dudoso, dándole mayor énfasis al pronóstico que arroje resultados menos erróneos con respecto a la realidad, ósea, el más exacto. Además, el pronóstico es una de las herramientas que ayuda en la toma de decisiones de una empresa.

Existen distintos tipos de pronósticos según el grado de complejidad de los objetivos que plantea la empresa. El analista tendrá la difícil labor de revisar cuál será la mejor alternativa, para ello, tendrá que considerar todos los factores que puedan influir en una demanda, desde políticas internas de la empresa hasta factores externos que conlleven a otros (alza en el dólar, inflación, etc.). Por ende, la elección de un pronóstico es importante, y a su vez, requiere de un trabajo muy riguroso. A grandes rasgos, los pronósticos se pueden dividir en dos tipos: cuantitativos y cualitativos.

Cualitativos: Las predicciones se realizan bajo bases subjetivas usando solamente el criterio o percepción del analista. Estos pronósticos se utilizan cuando existe escasez de datos numéricos y algo confusos.

Cuantitativas: Dentro de estos existen: series de tiempo y causales. Utilizan la proyección en base a los datos provenientes de series de tiempo.

1.1 Series de tiempo

Las series de tiempo, son una secuencia de datos medidos durante un intervalo de tiempo, en donde, cada dato está ordenado cronológicamente, además entre cada dato existe el mismo intervalo de tiempo (ejemplo: registro de días, semanas, meses, etc.).

“Para construir un modelo de series de tiempo, lo único que se necesita es la información muestral de la variable a analizar” (Fernández0?).

Una serie de tiempo se define como una cadena de vectores, dependientes del tiempo transcurrido.

$$t: \dot{x}(t) \\ (1)$$

Donde $t = 0, 1, 2, \dots, n$

1.1.1 Componentes de una serie de tiempo

En una serie de tiempo es muy importante identificar cual es el comportamiento que se está originando en el análisis de los datos transcurridos en un cierto periodo, esto se conoce como patrón de datos. Entre los componentes de datos mayormente registrados se encuentran cuatro: horizontales, tendencias, estacionales y cíclicos.

Horizontales: Los datos observados se mantienen sin cambios durante el tiempo, ósea, constantes.

Tendencias: Existe una tendencia cuando los datos registrados crecen o bajan en periodos sucesivos, estos datos quedan mejor registrados trazando una recta, resumiendo cuanto es lo que se ha mantenido al alza o a la baja. Su comportamiento puede ser tanto lineal como exponencial y se reflejan por lo general a largo plazo. En definitiva los datos que han bajado o subido son menores o mayores que los antecesores (ejemplo: la disminución del precio del cobre, el alza en el dólar, etc.).

Estacionarias: Los datos tienden a repetirse en el mismo periodo año tras año, (ejemplo: las ventas de los helados tienen un comportamiento estacional, ya que sus ventas son mayores en verano).

Componente cíclico: Es de comportamiento irregular, o sea, los datos aumentan y bajan en periodos no fijos, por lo tanto sus fluctuaciones son en forma de onda alrededor de la tendencia, lo que lo hace ser muy difícil de pronosticar. Estos son sucesos que muy rara vez pueden volver a ocurrir, como por ejemplo, guerras, terremotos, tornados, etc.

1.1.2 Métodos básicos de series de tiempo

En las series de tiempo se pueden medir mediante distintos tipos de proyecciones, en esta primera parte se describirán los tres enfoques más sencillos, estos son: métodos de promedio, métodos de suavización y métodos de regresión.

- **Métodos por promedios**

Estos se caracterizan por ser un tipo de pronóstico que pondera los datos registrados del pasado mediante un promedio de ellos, con el objetivo de minimizar o suavizar cambios inesperados que ocurren en el corto plazo, provocados por factores internos y externos (Hanke&Wichern10), en la siguiente ecuación se demuestra este método:

$$\hat{Y}_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i$$

(2)

Donde \hat{Y}_{t+1} es el pronóstico en el momento inicial t, calculado por el promedio de i observaciones pasadas.

Este método se puede dividir en: promedios simples y promedios móviles dobles. Los primeros utilizan todas las observaciones históricas y las promedian, mientras que los promedios móviles dobles son más complejos y utilizan dentro de su cálculo los datos más recientes, otorgando a estos últimos un mayor valor agregado. Esta última herramienta es la que la empresa utiliza para proyectar sus datos, esto se describirá más adelante en el capítulo 5.

- **Método de suavización exponencial**

La particularidad de este método es que en vez de considerar solo los datos más recientes, como es el caso de los promedios móviles, utiliza el mismo concepto para calcular la media, pero a cada dato le otorga un cierto peso exponencial, este peso es otorgado a cada dato de forma decreciente, ósea, el que se obtenga mayor peso será el dato más reciente y así sucesivamente hasta llegar al resultado (Hanke&Wichern2010).

- **Métodos causales**

Son aquellos en donde una o varias variables dependientes cambian su comportamiento a medida que se mueve una variable independiente, por ejemplo una variable dependiente pueden ser la cantidad de ventas de bombas hidráulicas y una variable independiente el tiempo en meses, en donde a medida que pasan los meses esta variable puede impactar tanto negativa como positivamente el volumen en las ventas de este insumo.

Entre los modelos causales se encuentran los de: regresión simple y regresión múltiple.

Regresión lineal simple

El primero es un método que permite analizar las relaciones entre dos variables, con el objetivo de predecir la variable dependiente a través de la comprensión de la variable independiente (Moreno2002), definido como correlaciones entre variables.

La forma más didáctica de expresar la relación entre dos variables es a través de un gráfico de coordenadas bidimensionales llamado diagrama de dispersión, donde un dato podría ser “Y” y el “X”.

En la siguiente ecuación queda mejor expresada la relación lineal.

$$y = b_0 + b_1 * x + u$$

(3)

Donde Y es la variable dependiente y X la variable independiente o explicativa, los valores b_0 y b_1 son coeficientes que definen la recta, donde el primero es el punto en que la recta corta el eje vertical y el segundo es la pendiente de la recta, u se le asigna como la perturbación o error (Moreno2002).

La pendiente de la recta se puede calcular bajo la siguiente ecuación:

$$b_1 = \frac{\sum x \sum y - n \sum xy}{(\sum x)^2 - n \sum x^2}$$

(4)

Donde “n” es la cantidad total de datos.

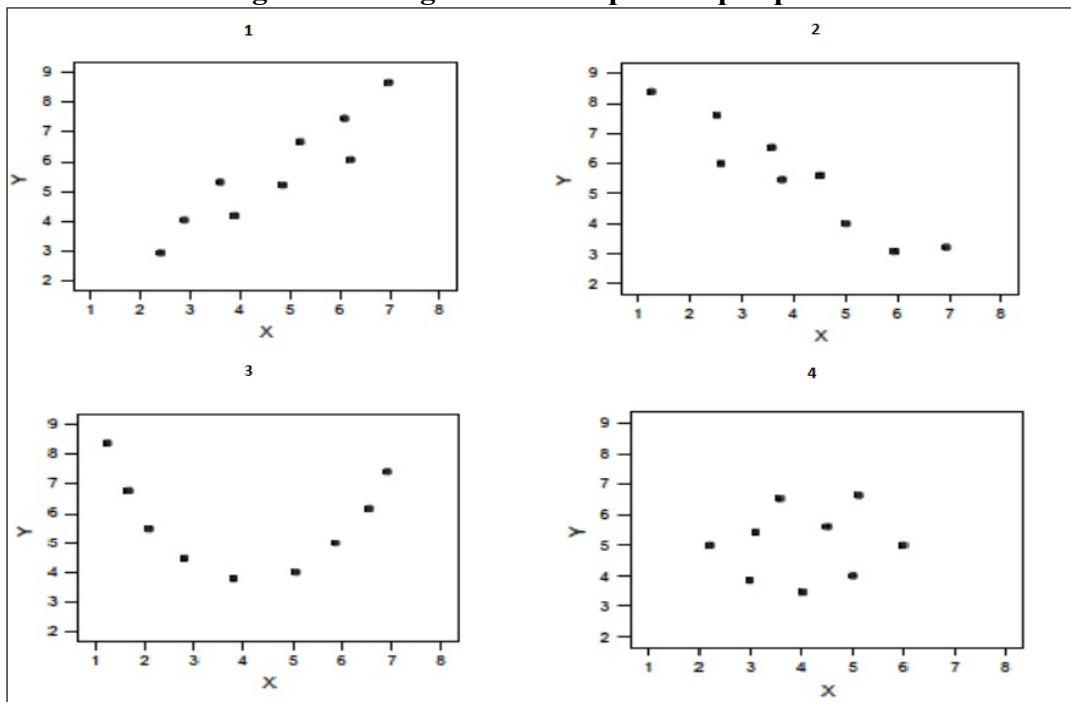
Y la ecuación para calcular el punto en que la recta corta al eje vertical se expresa a continuación:

$$b_0 = \frac{\sum y - b_1 \sum x}{n}$$

(5)

En las siguientes imágenes se expresan los tipos de relaciones entre dos variables anteriormente definidos, mediante diagramas de dispersión.

Figura 3.1 Diagramas de dispersión por patrón



Fuente: Pronostico en los negocios, John Hanke, 2010

El mejor indicador que puede traducir que tan fuerte es el nivel de relación entre dos fenómenos es el coeficiente de correlación. Este indicador se define con la letra r, dependiendo del nivel de relación adquiere los siguientes valores, correlación $r = -1$ esto ocurre cuando dos variables tienen una relación negativa perfecta, como se describe en la figura 3.1 en el diagrama 2, si la correlación $r = 1$ las dos variables tienen una relación positiva perfecta, como se describe en el diagrama 1 y si la correlación es baja, es decir, r está cercano al 0 la relación entre las

variables no es lineal, como se demuestra en los diagramas 3 y 4, por lo tanto el rango del valor de la coeficiente varía entre -1 a 1.

En definitiva el cálculo del coeficiente de correlación queda registrado en la siguiente ecuación (Sierra0?):

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \quad (6)$$

Donde las dos variables a analizar si es que existe alguna relación son “X”, “Y”.

Regresión múltiple

Este tipo de regresión implica la relación existente entre una variable dependiente con más de una variable independiente (Rojo2007).

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + \dots b_k * x_k + u \quad (7)$$

La auto correlación se define como *“la correlación que existe entre observaciones que se registran en diferentes periodos, en donde la variable analizada puede estar retrasada uno o más periodos consigo misma, este número de retrasos se expresará con la letra K”* (Hanke&Wichern2010).

El indicador que calcula esta expresión de relaciones en distintos retrasos de tiempo, se denomina coeficiente de auto correlación y queda expresada en la siguiente ecuación (Sierra 0?):

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (8)$$

Donde \bar{Y} es la media de los valores de la serie de tiempo, Y_t es la observación en el tiempo t y el dato Y_{t-k} son los datos ocurridos en “K” periodos anteriores expresados en t-k.

“Una serie es aleatoria cuando las auto correlaciones entre Y_t y Y_{t-k} para cualquier retraso de tiempo “K” son cercanas de 0, lo que se concluye que los valores sucesivos de tiempo no están relacionados entre sí” (Hanke&Wichern2010).

Una serie puede mostrar tendencia cuando los coeficientes de correlación sean distintos de 0, para los primeros retrasos y después disminuirá de forma lenta o gradual.

“El error estándar ayuda a saber si los datos de una serie son aleatorios, asumiendo los coeficientes de auto correlación de datos aleatorios como una distribución muestral que

puede aproximarse mediante una curva normal, con media 0 y distribución estándar $\frac{1}{\sqrt{n}}$ bajo este parámetro se puede trabajar con datos muestrales, el error se calcula bajo la siguiente ecuación” (Hanke&Wichern2010):

$$SE(r_k) = \sqrt{\frac{1+2 \sum_{i=1}^{k-1} r_i^2}{n}} \quad (9)$$

Donde r_i es la autocorrelación en el retraso i, k retraso de tiempo y n número de observaciones en la serie de tiempo.

Hay que recalcar que este error calcula de forma separada lo datos si son o no aleatorios, lo cual no lo hace muy exacto para grupos de datos.

Para calcular con mayor certeza si una serie es aleatoria independiente de los retrasos, se tiene que utilizar el estadístico Q de Ljung-Box, este tiene una distribución chi cuadrado con m grados de libertad.

En la siguiente ecuación queda demostrado el cálculo del estadístico Q

$$\varphi = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k} \quad (10)$$

Donde n es el número de observaciones en la serie de tiempo, k retraso de tiempo, m número de retrasos de tiempo que se van a probar y r_k la función de auto correlación muestral de los residuos atrasados k periodos.

Los valores de que $Q < \text{distribución Chi}$, implica que se puede aceptar que los datos son aleatorios, esto implica que ninguno de los retrasos analizados está fuera de un determinado intervalo de confianza. De lo contrario si el valor $Q > \text{distribución Chi}$, se concluye que los datos son auto correlacionados, en que de que el valor fuese muy alto, se dice que los datos están fuertemente correlaciones, lo que se traduciría en una serie con componente tendencial (Hanke&Wichern2010).

También se puede analizar la prueba de Box – Pierce, bajo el mismo concepto de la prueba Ljung - Box, pero esta prueba servirá para analizar los correlogramas residuales, (estos se definirán en el siguiente punto) la idea es demostrar la aleatoriedad en sus datos, ósea, que la serie obtenga RB, principal objetivo a la hora de validar una herramienta predictiva.

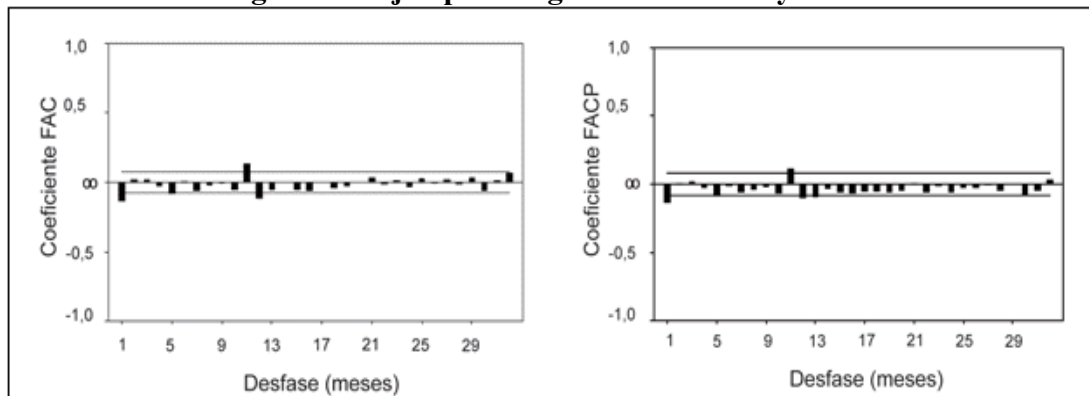
Para presentar RB la serie tiene que describir un comportamiento estable en sus correlogramas, es decir, que tenga desfases insignificantes y además mediante la prueba Box – Pierce demuestre que su estadístico sea mayor o igual a 0,05 (valor calculado y considerado con 95% de libertad).

1.1.3 Funciones de auto correlación

La auto correlación se expresa a través dos tipos de funciones, aquellas que se expresan en gráficos que son; FAC y FACP. El primero estima el coeficiente de auto correlación a través de la relación entre los valores de la serie en “K” retrasos, mientras que la segunda no considera las correlaciones acumuladas hasta el retardo “K”, ósea, entre el FAC y FACP son complementos.

En la siguiente imagen se expresa un ejemplo de los dos correlogramas antes definidos:

Figura 3.2 Ejemplos de gráficos de FAC y FACP



Fuente: Elaboración propia

En la imagen anterior se puede describir en el gráfico izquierdo un auto correlograma simple y en el lado derecho uno parcial, en el eje “y” está el cálculo del coeficiente que oscilan entre 1 a -1, en el eje “x” está el número de desfases en meses, las barritas ennegrecidas demuestran el valor de cada desfase, estos están delimitados por una región de confianza (generalmente se encuentra entre un rango de 0,25 a – 0,25), para demostrar si un valor es significativo debe estar fuera de la región delimitada, por lo tanto en este ejemplo se puede concluir que los dos gráficos en este caso se comportan de manera similar y ambos presentan tres desfases significativos (retardo 1, 11 y 12), se podría decir que es una serie estacionaria ya que sus desfases decaen a 0 de forma rápida.

Adicionalmente se pueden analizar y comprobar la aleatoriedad de datos en correlogramas residuales, estos cumplen la misma lógica que los FAC y FACP, pero con la particularidad que describen los residuos del pronóstico ya calculado, siempre considerado por periodo o desfase.

Por lo tanto la auto correlación de los residuos en un retraso k mide la fuerza de la correlación entre residuos de k periodos distantes (StatPoint2006).

La auto correlación de las diferencias en un retardo k es calculado de la siguiente forma:

$$r_k = \frac{\sum_{t=1}^{n-k} (e_t - \bar{e})(e_{t+k} - \bar{e})}{\sum_{t=1}^n (e_t - \bar{e})^2} \quad (11)$$

Donde: e_t es el error en el periodo t, \bar{e} es el error promedio, e_{t+k} es el error en el tiempo t más k retrasos y r_k es el residuo en el retraso k.

2. Estadístico Durbin Watson (DW)

El estadístico DW comprueba si existe o no una correlación entre residuos, a través de la aplicación de una regresión lineal entre ellos, el estadístico se calcula mediante la siguiente formula:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}, \quad 0 \leq DW \leq 4 \quad (12)$$

Donde:

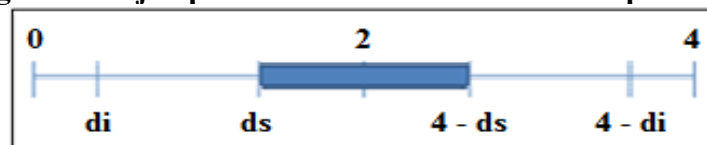
$$e_t = Y_t - \hat{Y}_t = \text{residuo para el periodo } t$$

$$e_{t-1} = Y_{t-1} - \hat{Y}_{t-1} = \text{residuo para el periodo } t - 1$$

Para un determinado nivel de confianza, las tablas del DW ofrecerán siempre dos valores, uno inferior “di” y otro superior “ds”, cada caso se describe en las siguientes tablas (Maddala1996):

Caso 1: Si los valores DW son iguales o cercanos a 2 (1.5 – 2.5) no se demuestra relación entre las variables, ósea, los valores que estén dentro de la franja azul en la siguiente imagen.

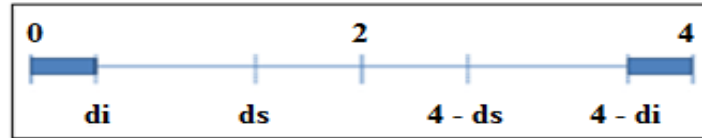
Figura 3.3 Ejemplo de una no correlación o independencia



Fuente: Informe “Conceptos básicos sobre la auto correlación en el modelo básico de regresión lineal”: pagina 11, Ramón Mahía, 2010.

Caso 2: Si los valores DW están en los extremos, ósea, cercanos o iguales a 0, estos demuestran una autocorrelación positiva, mientras que si son cercanos o iguales a 4 experimentan autocorrelación negativa, tal como se demuestra en la siguiente figura:

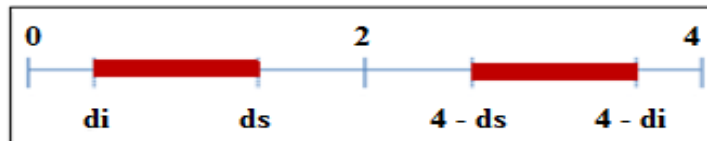
Figura 3.4 Ejemplo de una correlación negativa o positiva



Fuente: Informe “Conceptos básicos sobre la autocorrelación en el modelo básico de regresión lineal”: página 11, Ramón Mahía, 2010.

Caso 3: Existen zonas de “duda” en las que el estadístico DW no permite rechazar o aceptar la hipótesis nula, son las zonas intermedias entre d_i y d_s y entre $4 - d_i$ y $4 - d_s$, en ellas se optará como criterio que existen autocorrelación.

Figura 3.5 Ejemplo de una correlación negativa o positiva indefinida



Fuente: Informe “Conceptos básicos sobre la autocorrelación en el modelo básico de regresión lineal”: página 11, Ramón Mahía, 2010.

En definitiva la independencia de una variable aleatoria entre los residuos se calculará cuando el estadístico de Durbin-Watson tome un valor de 2 o cercano a él, más específicamente dentro de la región (1.5 - 2.5), en caso de tomar un valor $DW < 2$ indicaría autocorrelación positiva y si $DW > 2$ negativa.

Este estadístico se aplicará cuando se confeccione una gráfica rango – medio, para comprobar si existe o no dependencia entre ambas variables (Arce&Mahía2009).

3. Medición de rendimiento de un pronóstico

El objetivo principal es demostrar cual es la mejor herramienta predictiva para un caso en particular, ósea, se define como la capacidad de error que puede tener el mismo, considerando en este resultado todos los componentes que distorsionan el comportamiento normal de una serie de tiempo original.

En la siguiente formula se describe este indicador:

$$e_t = Y_t - \dot{Y}_t \quad (13)$$

Donde:

e_t = Error del pronóstico en el periodo t

Y_t = Ventas reales en el periodo t

\dot{Y}_t = Ventas estimadas o pronosticados en el periodo t

Los resultados de este indicador no son lo mejor, esto debido a que si los datos proyectados son mayores a los reales se producirán resultados negativos que distorsionarán el cálculo final, sin llegar a alguna conclusión, para ello será necesario buscar distintos tipos de medidas de error, estos se explican en el siguiente punto.

3.1 Tipos de medidas de error del pronóstico

Desviación media absoluta calcula los errores absolutos promedios del pronóstico en unidades, proporciona un resultado respecto a la media del error de un pronóstico, se calcula bajo la siguiente ecuación:

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \dot{Y}_t| \quad (14)$$

Error cuadrático medio, las diferencias de los valores reales y pronosticados, se elevan al cuadrado y luego se dividen por el número de observaciones, el resultado final se le aplica raíz cuadrada, este método otorga mayor ponderación a errores grandes, lo que lo transforma en el indicador más fiable, es por este motivo que para efecto de conclusiones finales este será el mejor indicador a considerar, este se expresa en la siguiente ecuación:

$$RMSE = \sqrt{\left(\frac{1}{n} \sum_{t=1}^n (Y_t - \dot{Y}_t)^2\right)} \quad (15)$$

Error porcentual absoluto se define como el cálculo de los errores de cada periodo divididos entre el valor real observado de este periodo y promediando estos errores porcentuales

absolutos. Sirve para cuando los valores de reales son muy altos siempre de forma porcentual, lo que permite una interpretación más simple y amigable, este error se expresa en la siguiente ecuación:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{|Y_t|}$$

(16)

En resumen, al comparar cada uno de los distintos pronósticos, el que mejor se acomode a la serie original será el que tenga los valores más pequeños en cuanto a MAD, RMSE y MAPE, en donde, los dos primeros serán interpretados en unidades y el último en porcentaje.

4. Modelos auto regresivos integrados de promedio móvil (ARIMA)

Box y Jenkins (1978) propusieron un modelo novedoso para la predicción de series de tiempo. Popularmente es conocida como metodología Box-Jenkins aunque técnicamente se conoce como metodología ARIMA, en la que se analizan las propiedades probabilísticas o estocásticas de las series de tiempo bajo la filosofía de “permitir que la información hable por sí misma” (Hanke&Wichern2010).

Este modelo trabaja con la asociación de sus variables de forma de enfoque lineal, permitiendo llegar a pronósticos a través de la iteración constante. Además los modelos ARIMA permiten obtener buenas aproximaciones cuando se aplica a series de tiempo estacionarias, si en caso que sean no estacionarias, también pueden convertirse en estacionarias mediante transformaciones adecuadas.

A continuación se describen las ventajas de estos modelos ARIMA frente a los otros modelos ya descritos anteriormente:

- Los conceptos y métodos asociados al ajuste de modelos ARIMA están respaldados por teorías sólidas de Probabilidad.
- ARIMA integra una familia de modelos (no solamente un modelo) lo que lo posiciona en ser un modelo mucho más versátil.
- Este modelo se sustenta bajo la metodología llamada Box Jenkins para el ajuste, permitiendo analizar y descifrar variables externas que pueden afectar directa o indirectamente cada serie de tiempo.
- El modelo ARIMA, es quizás el método más completo para pronosticar series de tiempos, ya sea que trata series estacionales y no estacionales (mediante la transformación).

A continuación se presentará una tabla comparativa de las herramientas predictivas más comunes.

Tabla 3.1 Comparación de las distintas herramientas predictivas

Métodos por Series de Tiempo	Descripción del Método	Usos	Exactitud (en un plazo)			Exactitud (en cantidad de datos)			Costo Relativo
			Corto	Mediano	Largo	Pocos	Medianos	Varios	
			(De 1 mes, máximo de 6 meses)	(De 6 meses a 2 años)	(Más de 2 de años)	12 a 24	24 a 60	60 a más	
Promedio Móvil	Se basa en un promedio aritmético calculado de un número de puntos de datos en el pasado.	Planeación de corto plazo para inventarios, niveles de producción y programación. Es útil cuando hay muchos productos.	Buena a muy buena	Regular	Muy mala	Bueno	Regular	Malo	Bajo
Suavización Exponencial	Similar al promedio móvil. Da un mayor peso exponencial a los datos más recientes.	Igual que el Promedio Móvil. Adaptado para cuando es necesario pronosticar un gran número de artículos.	Buena a muy buena	Regular	Muy mala	Regular	Bueno	Bueno	Medio
Modelos Matemáticos	Un modelo lineal o no lineal ajustado con los datos de series de tiempo, normalmente mediante regresión. Incluye las líneas de tendencias, polinomios, logaritmos lineales, series de Fourier, etc.	Igual que el promedio móvil pero con limitaciones debido al costo y a su uso con pocos productos.	Muy buena	Regular a buena	Muy mala	Bueno	Regular	Malo	Bajo a medio
Box - Jenkins	Métodos de auto correlación que se usan para identificar las series de tiempo subyacentes y para ajustar el modelo.	Para series de datos muy complejas que son difíciles de simular, pronósticos muy exactos a mediano plazo.	Regular a buena	Buena a muy buena	Regular	Regular	Bueno	Excelente	Medio a alto

Fuente: Elaboración propia

En la tabla anterior se resume la descripción de los distintos pronósticos aplicados bajo distintas circunstancias, demostrando que el pronóstico ARIMA puede ser la mejor opción siempre y cuando: los objetivos de predicción sean a mediano plazo, si se aplican a varios grupos de productos, si se cuenta con una serie de tiempo con varios datos (60 a lo menos) y en donde exista además una situación con una serie de datos más compleja (todos los antecedentes son los que presenta nuestro caso).

Por lo tanto se optará en este proyecto por aplicar a cada grupo el tipo de modelos predictivos ARIMA, demostrando que es la mejor opción bajo esta condición.

4.1 Definición de los conceptos básicos de los modelos ARIMA

Los pronósticos ARIMA, para su comprensión, se necesita tener conocimiento de los procesos estocásticos. Un proceso estocástico es una sucesión de variables aleatorias ordenadas en el tiempo que puede ser definido como Y_t , es decir, es una clara explicación de una serie de tiempo. Cada variable que compone un proceso estocástico tendrá su propia función de distribución con sus correspondientes momentos, esto mismo pasará con un conjunto de variables que esta vez tendrá una función de distribución conjunta, es aquí en donde se llegan a procesos estocásticos estacionarios y no estacionarios.

Procesos estocásticos estacionarios: Es una serie estacionaria cuando sus variables aleatorias se mantienen estables en el tiempo, es decir, cuando su distribución es normal y sus medias y varianzas son constantes por el tiempo, este tipo de procesos estocásticos son los ideales para poder aplicar estos pronósticos ARIMA.

Procesos estocásticos no estacionarios: Es una serie que presenta valores que no son constantes en el tiempo, en donde presentan patrones tendenciales y variabilidades que cambian en el futuro, es en estas series de tiempo es donde es necesario hacer transformaciones para poder llegar a producir procesos estocásticos estacionarios.

4.2 Transformaciones según Box – Cox

Las transformaciones se aplicarán en series de tiempo donde no se compruebe estacionalidad, para demostrarlo hay que comprobar que existe una distribución normal en la serie. Para eso, tiene que existir una varianza y una media que se mantengan constantes, en caso contrario, habrá que aplicar transformaciones en los datos originales (StatPoint2006).

Esta clase de transformaciones consideradas son el tipo potencia, las cuales son calculadas bajo la siguiente formula:

$$\dot{Y} = (Y)^{\lambda_1} \quad (17)$$

Donde \dot{Y} será el valor transformado, Y el valor original correspondiente a la serie, λ_1 será el factor de potencia ocupado a la hora de la conversión de la serie, en la siguiente tabla se describen los valores que puede tomar este factor.

Tabla 3.2 Tipos de transformaciones Box – Cox

Potencia	Transformación	Descripción
$\lambda_1 = 2$	$Y' = Y^2$	Cuadrado
$\lambda_1 = 1$	$Y' = Y$	Datos sin transformar
$\lambda_1 = 0.5$	$Y' = Y^{(1/2)}$	Raíz cuadrada
$\lambda_1 = 0.333$	$Y' = Y^{(1/3)}$	Raíz cúbica
$\lambda_1 = 0$	$Y' = \ln(Y)$	Logaritmo
$\lambda_1 = -0.5$	$Y' = 1/(Y^{(1/2)})$	Raíz cuadrada inversa
$\lambda_1 = -1$	$Y' = 1/Y$	Recíproco

Fuente: Instructivo Statgraphics, 2006

Si λ_1 es cero indica que la transformación correspondiente es un logaritmo natural, esta es la más recomendada debido a los valores similares que arroja, siendo la más exacta en sus intervenciones, es por esto que en todas las conversiones si lo necesitan, primero se empezará por aplicar un logaritmo natural. Si ninguna de las transformaciones anteriores ayuda a explicar la estacionalidad de los datos, será necesario aplicar una diferencia entre los valores, esta transformación se realiza según las siguientes ecuaciones (Hanke&Wichern2010):

$$\Delta Y_t = Y_t - Y_{t-1} \quad \text{Diferencia de primer orden} \quad (18)$$

Si persiste la no estacionalidad será necesario aplicar otra diferencia, descrita como:

$$\Delta^2 Y_t = \Delta (\Delta Y_t) = \Delta (Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + 2Y_{t-2} \quad (19) \text{ Diferencia de segundo orden}$$

Esta transformación se utilizará si o si cuando la media no demuestre ser constante en el tiempo, más adelante se explicará.

Distribución normal: Para demostrar si la serie experimenta una distribución en forma de campana, habrá que verificarlo mediante un histograma y luego corroborarlo bajo la prueba de Shapiro Wilk.

Shapiro Wilk es una de las pruebas más famosas que analizan la normalidad de los datos, para efectuarla se calcula la media y varianza muestral, expresada como " S^2 ", también se calcula la suma de diferencias entre todas las observaciones al cuadrado como " D^2 ", estas se ordenarán de menor a mayor, en donde el primer valor se restará con el último, el segundo con el penúltimo y así sucesivamente, n será el número de observaciones, en la siguiente ecuación queda demostrado:

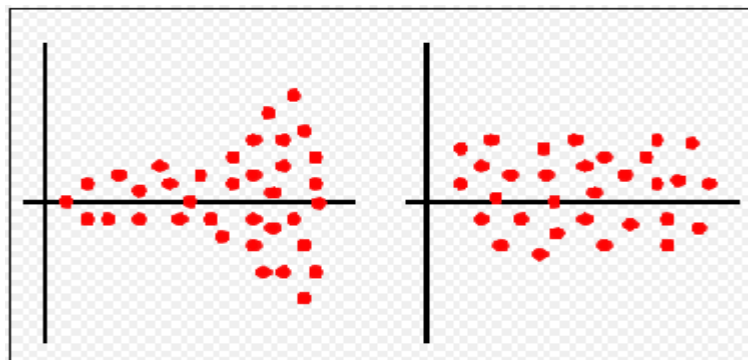
$$W = \frac{D^2}{nS^2} \quad (20)$$

Dónde: el estadístico “W” será el valor – P de la prueba, que rechazará la hipótesis de normalidad si este valor es menor a 0,05 (valor estándar considerando un 95% de confianza). En caso contrario será necesario aplicar alguna transformación Box – Cox demostrada en la ecuación 17 y si no resulta en última opción aplicar alguna diferencia según las ecuaciones 18 y 19.

Varianza constante: Para poder demostrar si una serie tiene una varianza constante o que no crece o decrece en el tiempo, será necesario definir el concepto de homocedasticidad y heterocedasticidad. La primera se define como un conjunto de datos que mantiene su varianza inalterable en el tiempo y la segunda se expresa cuando existen distintas varianzas en distintos tiempos, por lo tanto, para que una varianza sea constante será necesario que presente homocedasticidad en sus datos.

En la siguiente grafica se demuestra un ejemplo de lo anteriormente explicado.

Figura 3.6 Ejemplos de graficas de dispersión de datos con homocedasticidad y heterocedasticidad



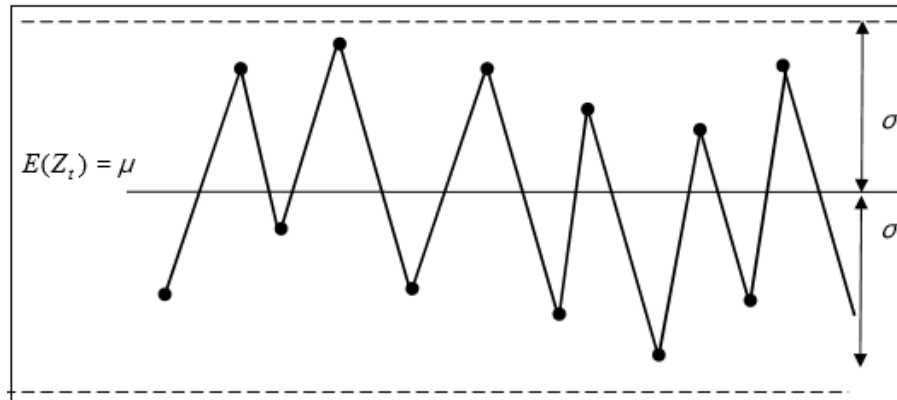
Fuente: Imagen Wikipedia, Chile, 2009.

En la figura 3.6 el grafico de dispersión izquierdo es un ejemplo heterocedasticidad y el otro de homocedasticidad.

Para poder analizar las varianzas se fabricarán gráficos con dos variables, rango (valores máximos – valores mínimos) y media (el promedio de la serie), cuyo objetivo es demostrar que no existe una relación lineal en ambas variables, para ello habrá que describir una dispersión de puntos parecido al grafico derecho de la figura 3.6 y además demostrar el principio de Durbin Watson, con esto se podrá comprobar la independencia de datos y por ende la homocedasticidad de los mismos, en caso contrario habría que aplicar una transformación Box – Cox (ecuación 17) o por ultimo alguna diferencia (ecuaciones 18 y 19), según la metodología Box – Jenkins.

En la siguiente figura 3.7 se describe un claro ejemplo de una varianza que es constante en el tiempo, en donde obtiene experimenta el mismo cambio a medida que pasa el tiempo, ósea, se mantiene dentro de la misma región de confianza acotada por un sigma.

Figura 3.7 Demostración de una serie con varianza constante



Fuente: Tesis “Pronósticos y series de tiempo, Gilberto Macías Calvario, 2007”

Media constante: Una serie de tiempo experimenta media constante cuando su promedio de datos es el mismo, es decir, no presenta tendencia, esto se puede comprobar a través de la lectura de los correlogramas, si los desfases durante el transcurso del tiempo caen rápidamente a 0, se puede decir que la serie estacionaria en media, en caso contrario habría que aplicar una transformación, esta se tiene que aplicar una vez efectuada la transformación logarítmica si se requería en el punto anterior (comprobación de varianza constante).

La transformación recomendada es la aplicación de diferencias en la serie de tiempo suponiendo que la serie original es Y_t , según las ecuaciones 18 y 19.

La diferenciación se hace hasta que una gráfica de los datos indica que la serie varía alrededor de un nivel fijo y sus correlaciones desaparecen con rapidez, por lo general momento de aplicar la segunda diferencia se debería estar en condiciones de comprobar la estacionariedad, con esta última transformación se podría aplicar el pronóstico ARIMA.

4.3 Tipos de modelos básicos ARIMA

4.3.1 Modelos auto regresivos

Este modelo auto regresivos (AR) al poseer una parte auto regresiva, donde una variable endógena correspondiente a un período t es explicada por las observaciones de ella misma en períodos anteriores añadiéndose un término de error, además posee un subíndice, este se abrevia con la letra “p”.

Así AR(1) pronóstico auto regresivo de orden 1, proyecta una observación en el tiempo “ Y_t ” como el valor observado en un tiempo anterior más un error aleatorio “ e_t ”, esto se expresa en la siguiente ecuación:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + e_t \quad (21)$$

Donde Y_t la variable dependiente en el tiempo t , Y_{t-1} la variable independiente, ϕ_0 , ϕ_1 son los coeficientes auto regresivos que serán estimados y e_t el error estimado el cual representa los efectos en variables no explicadas por el modelo, para todos los tipos de pronósticos este concepto es el mismo.

Mientras que AR(2) pronóstico auto regresivo de orden 2, es parecido al AR(1), pero con la diferencia que explica el valor del pronóstico mediante los valores observados en los dos tiempos anteriores más la componente de error, según la siguiente ecuación:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \quad (22)$$

Por lo tanto AR(p) pronóstico auto regresivo de orden k , se expresaría de la siguiente forma:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (23)$$

La forma abreviada de escribir lo anterior sería:

$$\phi_p(L) Y_t = e_t \quad (24)$$

Donde el $\phi_p(L)$ se denomina operador polinomial de retardos.

4.3.2 Modelos de promedio móvil

Modelo designado con la abreviatura MA, posee un parte de media móvil, esta expresa una respuesta de la media “ $Y_t - \mu$ ” en el periodo t mediante una sucesión de errores correspondientes a periodos anteriores, su orden se designa con la letra “ q ”.

Así MA(1), MA(2) y MA(q) quedan expresados en las ecuaciones (25), (26) y (27) respectivamente:

$$Y_t = \mu - \omega_1 e_{t-1} \quad (25)$$

$$Y_t = \mu - \omega_1 e_{t-1} - \omega_2 e_{t-2} \quad (26)$$

$$Y_t = \mu - \omega_1 e_{t-1} - \omega_2 e_{t-2} - \dots - \omega_q e_{t-q} \quad (27)$$

Donde μ es el promedio constante en el proceso, $\omega_1, \omega_2, \omega_q$ son los coeficientes estimativos de promedio móvil y $e_{t-1}, e_{t-2}, e_{t-q}$ son los errores en periodos anteriores del tiempo t .

Para abreviar la última ecuación se le agrega el operador polinomial de retardos $\omega_q(L)$, quedando conformado con la siguiente ecuación:

$$Y_t = \mu - \omega_q(L)e_t \quad (28)$$

4.3.3 Modelos mixtos de promedio móvil auto regresivos

Este modelo se puede expresar con la notación ARMA (p,q), donde su subíndice “p” es el orden de la parte auto regresiva y q es el orden de la parte del promedio móvil, permita describir una mayor variedad de comportamientos de las series de tiempo estacionarias, este modelo estaría considerando tanto los valores del pasado y presente de la variable dependiente y de sus errores, queda mejor expresado en la siguiente ecuación:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \omega_1 e_{t-1} - \omega_2 e_{t-2} - \dots - \omega_q e_{t-q} \quad (29)$$

El orden del modelo puede describir que tipo de pronóstico explicará la variable de respuesta, estos como anteriormente se describe dependerán de los valores y errores finitos de las observaciones anteriores. Por ejemplo un modelo ARMA (2,2) explicará el pronóstico mediante los valores y errores observados en los dos periodos anteriores.

4.3.4 Modelos mixtos auto regresivos e integrados de promedio móvil

Cuando las series no son estacionarias existen formas para volverlas estacionarias, como por ejemplo reemplazar la serie original por una diferenciada (tal como se explicó en el punto 4.2), para lograr llegar a estos resultados estacionarios es necesario usar el modelo ARIMA(p,d,q), donde d indica el número total de diferenciaciones necesarias para que el modelo pase de no estacionario a estacionario, donde las siglas p y q son los indicadores auto regresivos y promedios móviles respectivamente, en el caso que originalmente el modelo arroje datos estacionarios la sigla $d = 0$, ósea, es un modelo ARMA.

Por lo tanto la fórmula para ARIMA (p,1,q) y ARIMA (p,d,q) respectivamente son:

$$\Delta Y_t = \phi_0 + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \dots + \phi_p \Delta Y_{t-p} + e_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q} \quad (30)$$

$$\Delta^d Y_t = \phi_0 + \phi_1 \Delta^d Y_{t-1} + \phi_2 \Delta^d Y_{t-2} + \dots + \phi_p \Delta^d Y_{t-p} + e_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q} \quad (31)$$

4.3.5 Modelos mixtos con una componente estacional

Es un tipo de modelo ARIMA que posee una componente estacional, se manifiesta en series de corto plazo que presentan patrones repetitivos menores a un año, estos pueden ser principalmente mensuales o inclusive trimestrales, si se cumple el primero obtendría datos que se repetirían cada 12 meses en el tiempo y si fuesen semestrales a cada 4.

Por lo tanto su estructura estará conformada por dos componentes una regular y otra estacional, la primera identificada como un modelo ARIMA (p,d,q) anteriormente definido y la segunda por un ARIMA (P,D,Q)_{12 o 4}, donde las letras “P”, “D” y “Q” serían los indicadores auto regresivos, integrativos y media móvil respectivamente de la parte estacional y los sub – índices 12 o 4 los patrones repetitivos mensuales o trimestrales dependiendo del caso, conformando en definitiva un modelo mixto estacional identificado como un SARIMA (p,d,q)x(P,D,Q)_{12 o 4}, cuya ecuación final sería:

$$\Delta_{12} Y_t = Y_t - Y_{t-12} \quad \text{Si existe una diferencia estacional} \quad (32)$$

$$Y_t = \text{Ecuacion (30)} + \theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + e_t - W_1 e_{t-1} - W_2 e_{t-2} - \dots - W_Q e_{t-Q} \quad (33)$$

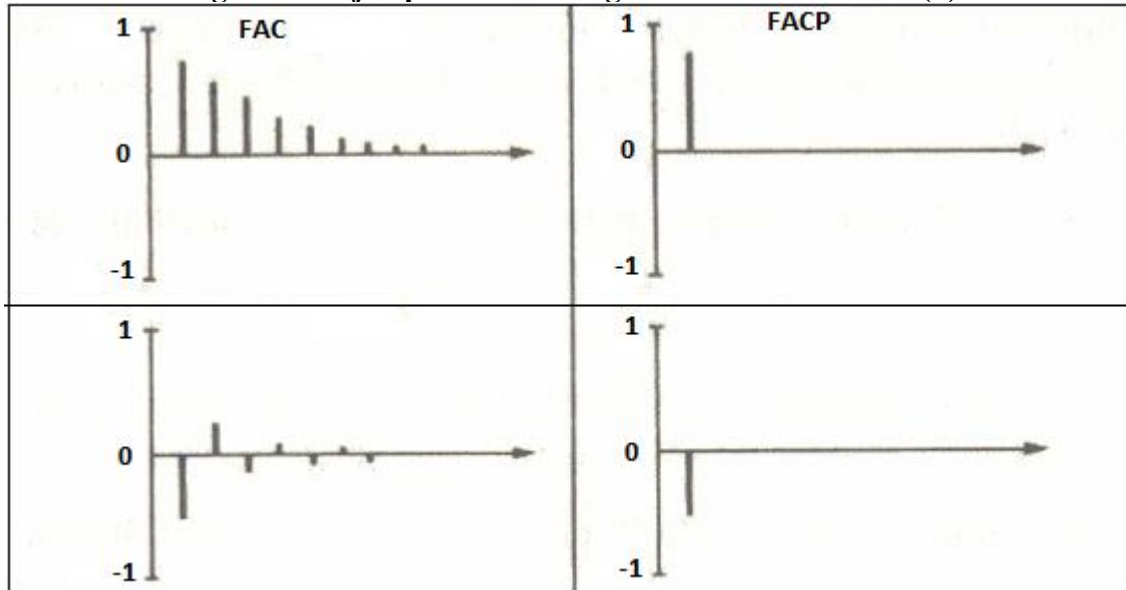
$$\Delta^{d,D} Y_t = \text{Ecuacion (31)} + \theta_0 + \theta_1 \Delta^D Y_{t-1} + \theta_2 \Delta^D Y_{t-2} + \dots + \theta_p \Delta^D Y_{t-p} + e_t - W_1 e_{t-1} - W_2 e_{t-2} - \dots - W_Q e_{t-Q} \quad (34)$$

Donde $\theta_0, \theta_1, \theta_2, \theta_p, W_1, W_2, W_Q$ son los coeficientes auto regresivos y de media móvil respectivamente que serán estimados en la parte estacional.

4.4 Metodología Box – Jenkins

Identificación: Para poder identificar un pronóstico primero es necesario comprobar que la serie es estacionaria (definido en el punto 4.2), al comprobarla habrá luego que leer los correlogramas FAC y FACP con el objetivo de extraer desde allí los distintos tipos de pronósticos posibles, a continuación se describirán los coeficientes de auto correlación teóricos de algunos de los modelos ARIMA más comunes (Hanke&Wichern2010).

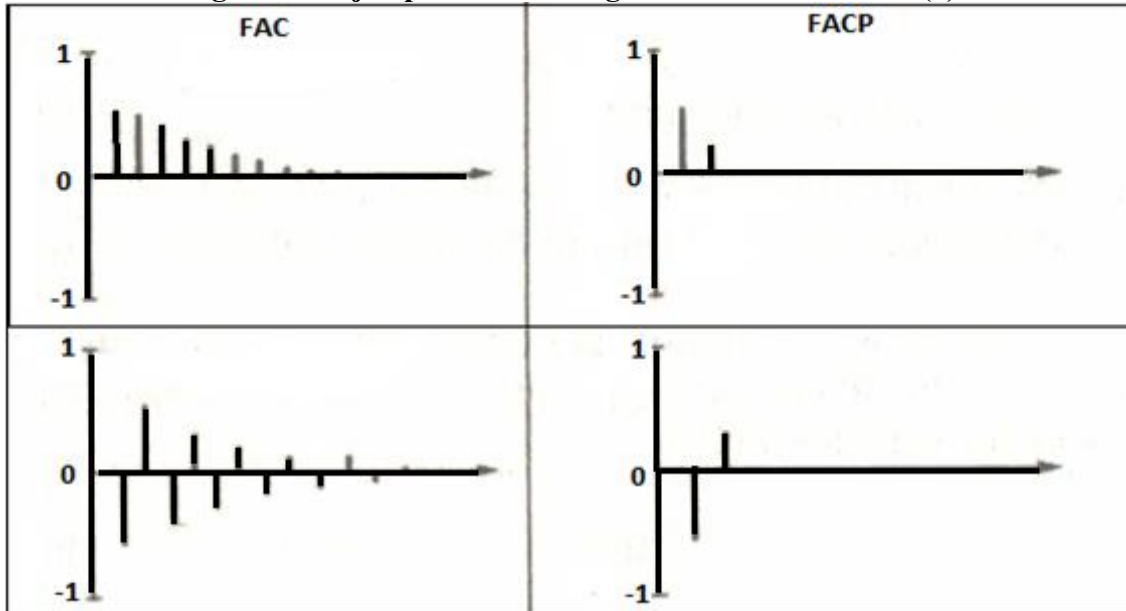
Figura 3.8 Ejemplos de correlogramas de modelos AR(1)



Fuente: Libro “Pronósticos en los negocios”: La metodología de Box – Jenkins, John E. Hanke y Dean W. Wichern, 9º edición

En la imagen anterior las dos parejas de FAC y FACP adquieren un comportamiento típico de un modelo auto regresivo de orden 1, en la primera los desfases caen rápidamente de manera exponencial a 0 en el FAC y describe 1 solo desfase significativo en el FACP, en el segundo en el FAC sus retrasos decrecen a 0 rápidamente pero de forma zigzagueante y en el FACP tiene un solo desfase fuera de la región pero negativo.

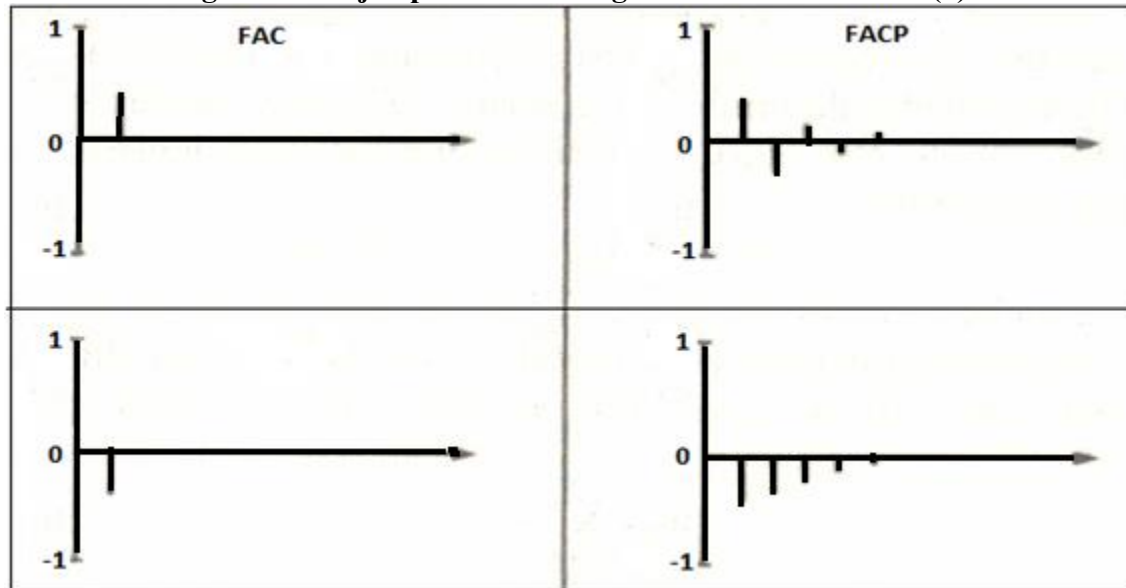
Figura 3.9 Ejemplos de correlogramas de modelos AR(2)



Fuente: Libro “Pronósticos en los negocios”: La metodología de Box – Jenkins, John E. Hanke y Dean W. Wichern, 9º edición

Mientras que en la figura 3.9 las dos parejas son un ejemplo claro de un modelo AR(2), adquiriendo comportamientos muy similares a las parejas de la figura 3.8, pero con la única diferencia que en los FACP se obtienen dos desfases significativos.

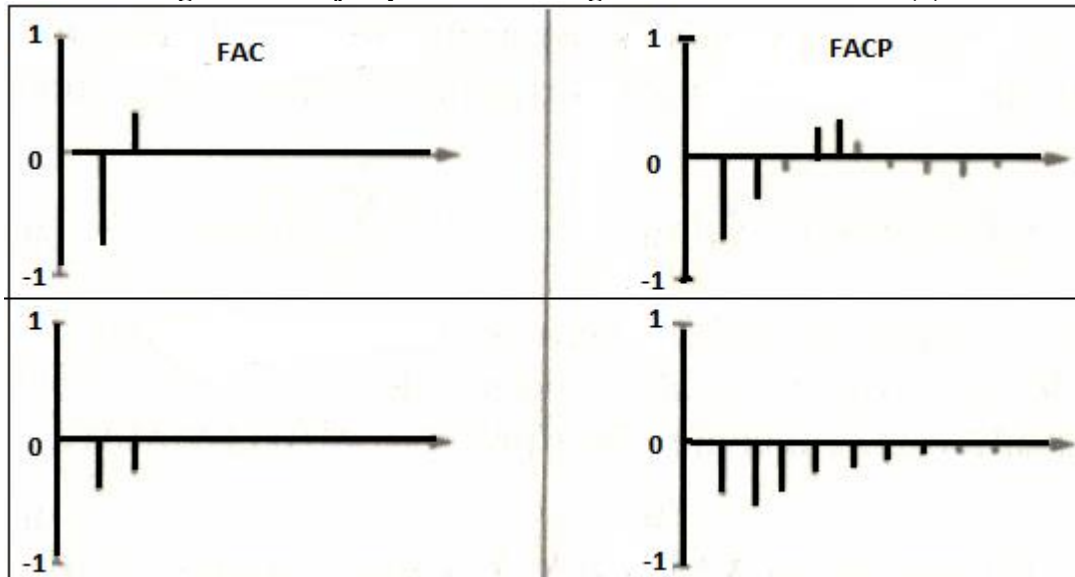
Figura 3.10 Ejemplos de correlogramas del modelo MA(1)



Fuente: Libro “Pronósticos en los negocios”: La metodología de Box – Jenkins, John E. Hanke y Dean W. Wichern, 9º edición

En la figura 3.10 las dos parejas son ejemplos de modelos con media móvil de orden 1, en donde poseen patrones de desfases repetitivos muy similares a los del AR(1), la diferencia radica en que el FAC adquiere el comportamiento del FACP y el FACP replica al FAC.

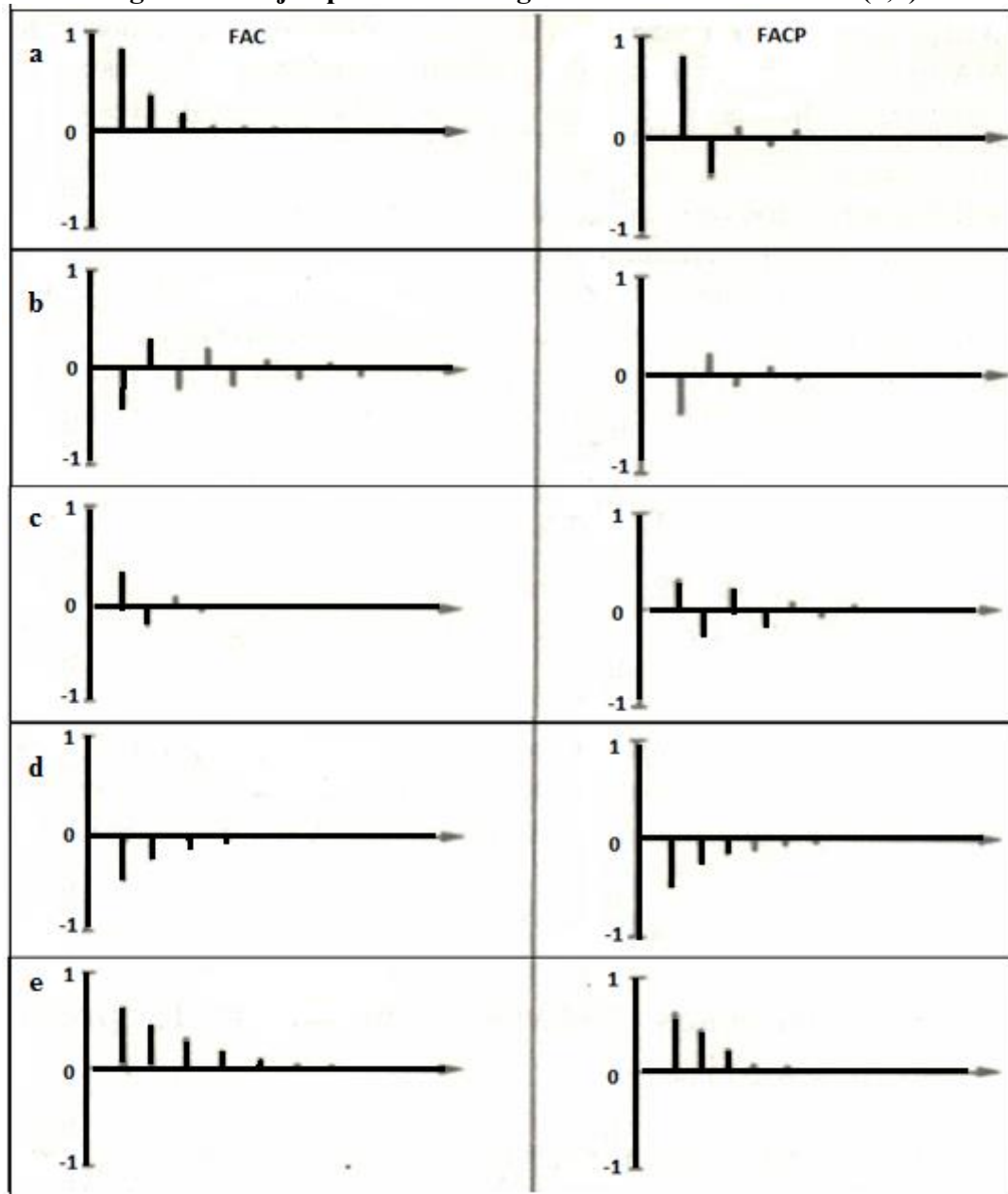
Figura 3.11 Ejemplos de correlogramas del modelo MA(2)



Fuente: Libro “Pronósticos en los negocios”: La metodología de Box – Jenkins, John E. Hanke y Dean W. Wichern, 9º edición

En la imagen se describe que las dos parejas resultan ser demostraciones de modelos MA(2), replicando comportamientos similares al AR(2) pero de forma al revés entre los FAC y FACP.

Figura 3.12 Ejemplos de correlogramas del modelo ARMA(1,1)



Fuente: Libro “Pronósticos en los negocios”: La metodología de Box – Jenkins, John E. Hanke y Dean W. Wichern, 9º edición

En la imagen anterior se describen ejemplos de comportamientos de modelos mixtos con componentes auto regresivos y de media móvil. En las parejas b, c, d y e se obtienen comportamientos similares, con FAC y FACP con desfases que de caen de manera zigzagueante y exponencialmente, mientras que en el primer ejemplo (letra a) sus correlogramas obtienen combinación de ambas.

Lo anteriormente expresado se podría resumir en la siguiente figura:

Figura 3.13 tabla resumen de los correlogramas de los pronósticos ARIMA más comunes

	FAC	FAP
MA(q)	Se anula para retardos superiores a q	Decrecimiento rápido sin llegar a anularse
AR(p)	Decrecimiento rápido sin llegar a anularse	Se anula para retardos superiores a p
ARMA(p,q)	Decrecimiento rápido sin llegar a anularse	Decrecimiento rápido sin llegar a anularse

Fuente: Libro “Pronósticos en los negocios”: La metodología de Box – Jenkins, John E. Hanke y Dean W. Wichern, 9ª edición

Estimación: Luego de haber identificado a los modelos tentativos, se debe identificar cada uno de sus parámetros, estos se estiman usando mínimos cuadrados, a través del cálculo proveniente de la función mínima de la suma de sus errores al cuadrado. La estimación de los parámetros de un pronóstico está formado por los siguientes indicadores: los coeficientes estimados (son los mínimos cuadrados calculados por un tipo de pronóstico), sus respectivos errores estándar (errores provenientes de los mínimos cuadrados), estadístico t (se calcula dividiendo el coeficiente estimado por su error estándar) y valor P calculado de la distribución t de Student. Hay que considerar que los valores P pequeños (menores que 0.05 si opera a un nivel de confianza del 95%) corresponden a coeficientes estadísticamente significativos. Si cualquier Valor-P es mayor que 0.05, se debe considerar reducir la complejidad del modelo, cumpliendo el objetivo de este punto.

Validación: Antes de elegir el modelo será necesario ver que tan adecuado es, para ello hay que comprobar si la proyección tiene ruido blanco (RB).

RB: Si un pronóstico describe ruido blanco, manifiesta que su proyección de datos es válida, esto al comprobar que sus residuos son aleatorios, ósea, que obtiene frecuencias aproximadamente iguales. Los pronósticos que en sus correlogramas residuales no presentan ningún desfase realmente significativo demuestran que sus proyecciones tienen RB. Para corroborar todo aquello será necesario analizarlo con la prueba Box – Pierce, principal indicador que demuestra si un conjunto de residuos presentan aleatoriedad.

Por lo tanto, para que un pronóstico tenga validez en una situación dada, tiene indicar que el conjunto de sus residuos deben ser aleatorios, de lo contrario se sugiere que el modelo es inadecuado y que por lo tanto debe ser modificado o cambiado por otro modelo. Además, en este punto se deben calcular sus errores o residuos con el objetivo de comprar los distintos modelos identificados y seleccionar los que se desempeñan de mejor manera, para ello se evaluarán los que tengan los estadísticos de error menores.

Predicción: Una vez validado los modelos y seleccionado el mejor, se debe predecir a través del cálculo del modelo.

Capítulo IV: Análisis de la Situación Actual

Tal como se describió anteriormente en el capítulo tres, el encargado de emitir las órdenes de compra es el gerente de ventas y que lo canaliza al gerente de abastecimiento, su método se basa netamente en su juicio intuitivo enfocado en las ventas históricas de los dos años anteriores agregando un factor de crecimiento por concepto de metas, este método se viene realizando desde el 2011.

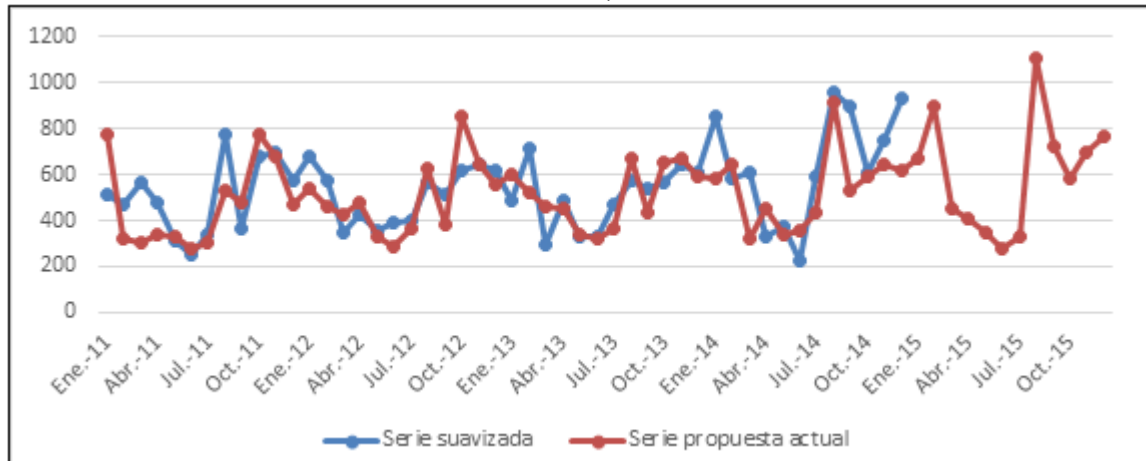
La situación actual de la empresa por lo tanto se enfocará principalmente en proyectar las ventas de un año, cuyo calculo básicamente se basará en el promedio mensual de los dos años anteriores, adicionando un 5% por concepto de meta de crecimiento anual, esto es aplicado a todos los grupos de productos, sin discriminar cual es la más importante de la empresa, ósea, no se implementa un análisis ABC, para efectos de análisis se dividirán los resultados por grupo de productos que serán revisados en el capítulo siete, con el objetivo final de ser comparados con la propuesta elegida del pronóstico ARIMA.

Estos resultados se compararán con los datos de la serie ya suavizada por producto, se optó por esta serie ya que es la que mejor representa el comportamiento de la demanda histórico del producto y además servirá de parámetro a la hora de comparar la situación actual y la propuesta ARIMA. Hay que considerar que la base de datos de los cálculos para la situación actual proviene de serie real, por lo tanto existe otra falencia de la situación actual porque incorpora en su proyección datos cíclicos.

1. Situación actual para centrifugas

Se presentará a continuación la situación actual de la empresa para electrobombas centrifugas, esta será comparada con la serie suavizada (esta se obtendrá en el capítulo seis específicamente en el punto 1.1.2), con el objetivo de exponer los datos con los que se planifica actualmente la empresa y analizarlos.

Figura 4.1 Comparación serie suavizada y la situación actual para centrifugas (2011 – 2014)



Fuente: Elaboración propia

En la gráfica anterior se observa que la situación actual obtiene un comportamiento que se intenta estacionar en la media de la serie suavizada, pero con datos en general insuficientes, al no poder alcanzar a igualar a la serie, esto debido a que se describe un gran cantidad de datos bajo la misma, a continuación se describe los estadísticos de error para estos datos:

Tabla 4.1 Estadísticos de error de la situación actual de la empresa para centrifugas

Estadístico	Situación actual
RMSE	145,44
MAE	112,82
MAPE	22,65%

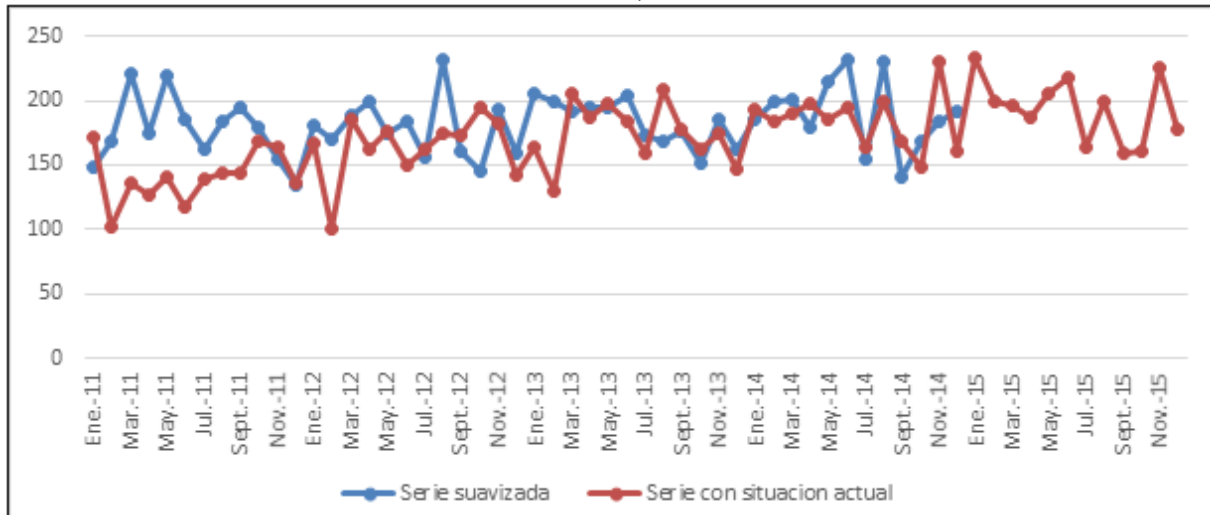
Fuente: Elaboración propia

En la tabla anterior se puede describir los estadísticos de error, estos serán comparados con la mejor propuesta ARIMA desarrollada en el próximo capítulo.

2. Situación actual para agua sucia

Para este tipo de productos tal como se describió anteriormente se ocuparon las mismas fórmulas para proceder con el cálculo de la situación actual, esta será comparada con la serie suavizada que se calculará en el capítulo seis en el punto 2.1.2, a continuación se presenta la serie de tiempo:

Figura 4.2 Comparación serie suavizada y la situación actual para agua sucia (2011 – 2014)



Fuente: Elaboración propia

Se puede apreciar que en la gráfica anterior la situación actual se encuentra a nivel general muy por debajo de la serie suavizada, sobre todo durante el año 2011, adquiriendo en el año 2015 (dato simulado) una pequeña alza que permite acercarse un poco a la serie histórica suavizada.

A continuación se presentan los estadísticos de error de la situación actual propuesta por la empresa para este tipo de productos:

Tabla 4.2 Estadísticos de error de la situación actual de la empresa para agua sucia

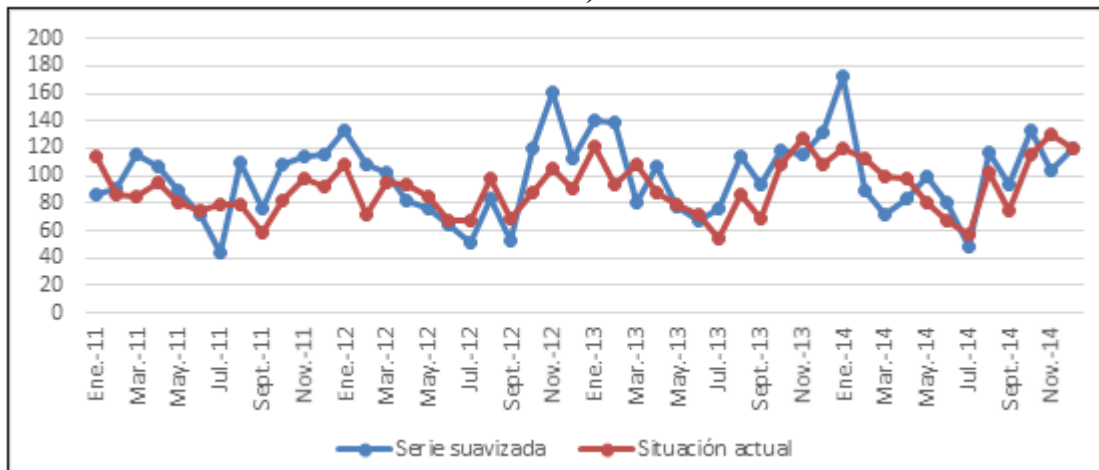
Estadísticos	Situación actual
RMSE	35,77
MAE	27,86
MAPE	14,98%

Fuente: Elaboración propia

3. Situación actual para multietapas

Los datos calculados en la siguiente grafica consideran la serie suavizada que se describirá en el punto 3.1.2 del capítulo seis y la situación actual efectuada por la empresa:

Figura 4.3 Comparación serie suavizada y la situación actual para multietapa (2011 – 2014)



Fuente: Elaboración propia

En la figura anterior se puede comprobar que se mantiene dentro de la media de la serie, pero con algunos datos que se sitúan sobre o por debajo de la serie suavizada.

Además se describirá sus estadísticos de error:

Tabla 4.3 Estadísticos de error de la situación actual de la empresa para multietapa

Estadístico	Situación actual
RMSE	23,32
MAE	19,85
MAPE	20,52%

Fuente: Elaboración propia

Capitulo V: Descripción de la Metodologia del Pronostico

1. Introducción de la metodología

Se implementará una metodología general que será aplicada a todos los datos, sin discriminación por tipo de productos, con el objetivo de describir un análisis previo al que será sometido cada grupo de productos, este se basa en los siguientes puntos:

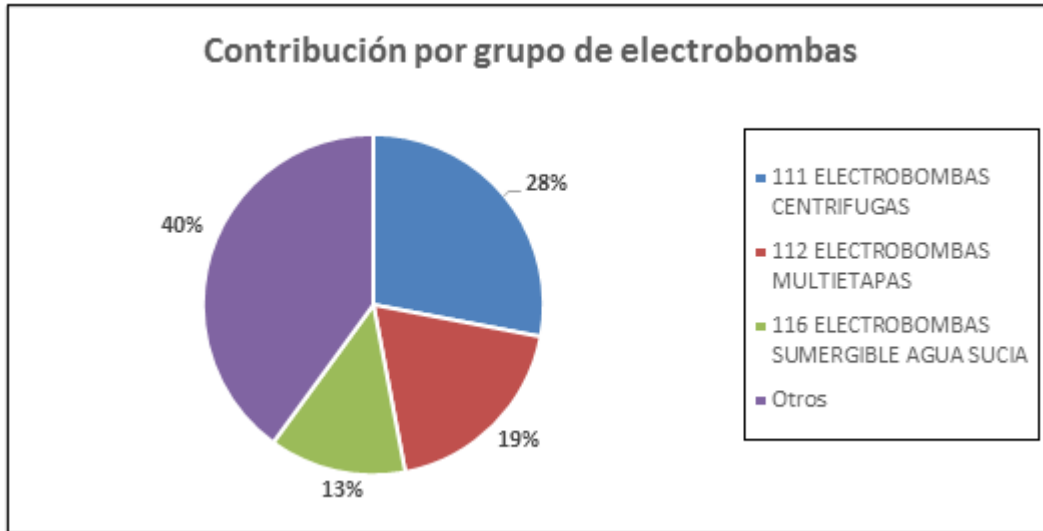
- Recolectar una base de datos con las ventas del 2014 de las bombas hidráulicas pertenecientes al área de distribución de Koslan.
- Identificar y clasificar por grupos las bombas más vendidas al Retail (con el sistema ABC).
- Seleccionar los grupos más rentables, que representen el 60% de la contribución total de distribución.
- Analizar y Registrar la información de ventas de los últimos diez años de cada grupo seleccionado.

Selección de datos por grupo de Koslan

En este punto se tomarán en cuenta todos los enunciados anteriores pertenecientes a la introducción.

En el área de distribución actualmente se comercializan alrededor de 1515 distintos tipos de modelos de bombas hidráulicas, estos están divididos en 16 grupos, el pronóstico se enfocará en los seis grupos más importantes (tal como se describió en el punto 2.3.2.1), esta selección se utilizará bajo el método de Pareto, más conocido como análisis ABC cuyo criterio de decisión será, los productos que ofrezcan la mayor contribución durante el 2014 para la organización por grupo.

Figura 5.1 Grafico de la contribución por grupo de electrobombas de distribución



Fuente: Ventas y detalles de Koslan 2014

En la gráfica (figura 5.1) anterior se demuestran los 3 grupos de electrobombas más importantes, entre todos ellos suman el 60% de la contribución total del área de distribución. Es por este motivo que los pronósticos se van a centralizar en analizar los datos de estos grupos.

Por lo tanto se analizará la información de estos seis grupos, considerando las ventas históricas de los últimos diez años (2005 – 2014), poniendo principal énfasis a los datos más recientes, ósea, al momento de tomar la decisión de que pronostico elegir será el que replique mejor la demanda de los últimos cuatro años (2011 – 2014).

2. Metodología Box – Jenkins

Para el desarrollo del pronóstico se utilizará la metodología Box – Jenkins, agregando un análisis estricto de los datos, el detalle se describe en los siguientes puntos.

- **Análisis de los datos e identificación del pronóstico:**

El objetivo de este punto es primero analizar los datos y comprobar que se cumple el principio de estacionalidad (datos con distribución normal, con media y varianza), una vez verificado lo anterior proceder a la lectura de correlogramas para poder identificar que pronóstico es el recomendado.

- **Identificación de componentes:**

Los componentes se podrán identificar a lo largo de la serie de tiempo, estos pueden ser: cíclicos, tendenciales, estacionarios o aleatorios.

- Suavizamiento de los datos:

Lo primero será identificar los datos cíclicos, estos serán revisados por la información recopilada de la empresa y corroborados mediante un diagrama de control, en paralelo se aplicará a la serie original una regresión simple, luego los datos identificados como cíclicos serán reemplazarlos por los datos regresionados con el objetivo de suavizar los datos que no cumplan con el real comportamiento de la serie.

- Comprobación de distribución normal:

Para verificar una distribución normal será necesario revisar su histograma y el estadístico Shapiro Wilk, si el conjunto de barras del histograma forma una campana y el estadístico es mayor o igual a 0,05 se podría comprobar que la serie está normalizada, en caso contrario habrá que someterla a una transformación.

- Análisis de varianza:

Para analizar la varianza será necesario confeccionar una gráfica con variables rango – media y revisar el estadístico Durbin-Watson, si la gráfica demuestra que ambas variables no están relacionados de manera lineal y DW está dentro del intervalo $[1.5 - 2.5]$, se demuestra varianza constante, en caso contrario será necesario transformar la serie a logaritmo natural.

- Identificación de patrones:

Mediante el análisis de los estadísticos de los correlogramas (FAC y FACP) se podrán identificar los distintos patrones como: comprobación de aleatoriedad, de datos correlacionados, si existe una componente estacional y si la serie tiene estacionariedad, este último punto es muy importante ya que si se comprueba aquello los datos estarán listos para poder ser pronosticados.

- Lectura de correlogramas:

Al analizar los correlogramas se podrán descifrar que tipo de componentes y órdenes son las seleccionadas para identificar el perfil del pronóstico ARIMA.

- Estimación del Pronóstico:

Una vez identificados los distintos pronósticos será necesario estimar los parámetros de cada pronóstico identificado, con el fin de reducir su complejidad.

- Selección de periodos de predicción y validación:

La cantidad de datos de predicción queda a juicio del analista, en este caso se ocuparán por 12 con el fin de llegar a mayores conclusiones de factibilidad del pronóstico mismo y con respecto a los datos de validación se optarán por 48.

- Descripción del modelo estimado:

La idea es describir los distintos parámetros del modelo y ver cuáles son significativos, si es que los hay se recomienda reducir la complejidad del pronóstico.

- **Validación del pronóstico:**

El objetivo es elegir la herramienta predictiva pre - seleccionada que mejor se asimila a la serie original y proyectar su modelamiento matemático.

- Evaluación:

En este punto será necesario revisar sus correlogramas residuales, para comprobar si poseen “RB”, si no se demuestra lo anterior será necesario descartar aquel pronóstico.

- Comparación de los pronósticos:

Los pronósticos validados serán comparados, con el objetivo de elegir el que más se acomode a la serie, a través del análisis de los parámetros de cada estadístico de error y varianza calculados.

- Elección del pronóstico:

Al momento de validar o demostrar que son adecuados los pronósticos pre – seleccionados se procede a elegir el mejor, esto se obtiene mediante el cálculo del RMSE, el que obtiene el menor valor será el seleccionado.

- **Predicción del mejor modelo validado:**

Describir el modelo predictivo y la serie proyectada vs la original.

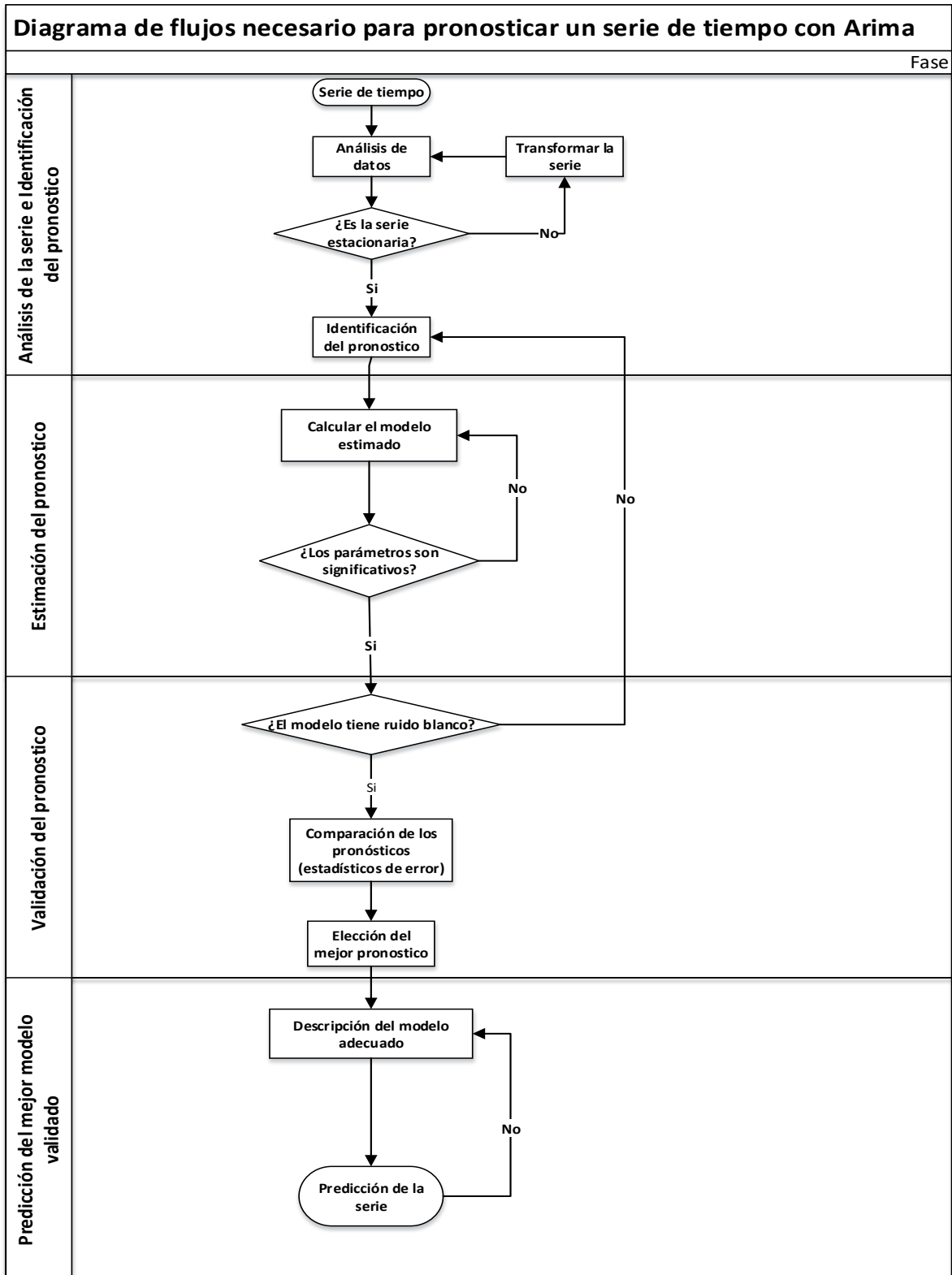
- Descripción del modelo adecuado:

Al obtener este modelamiento se podrá calcular cualquier dato futuro, por lo que se estaría llegando al objetivo inicial de este proyecto.

- **Comparación de la condición actual y el modelo seleccionado:**

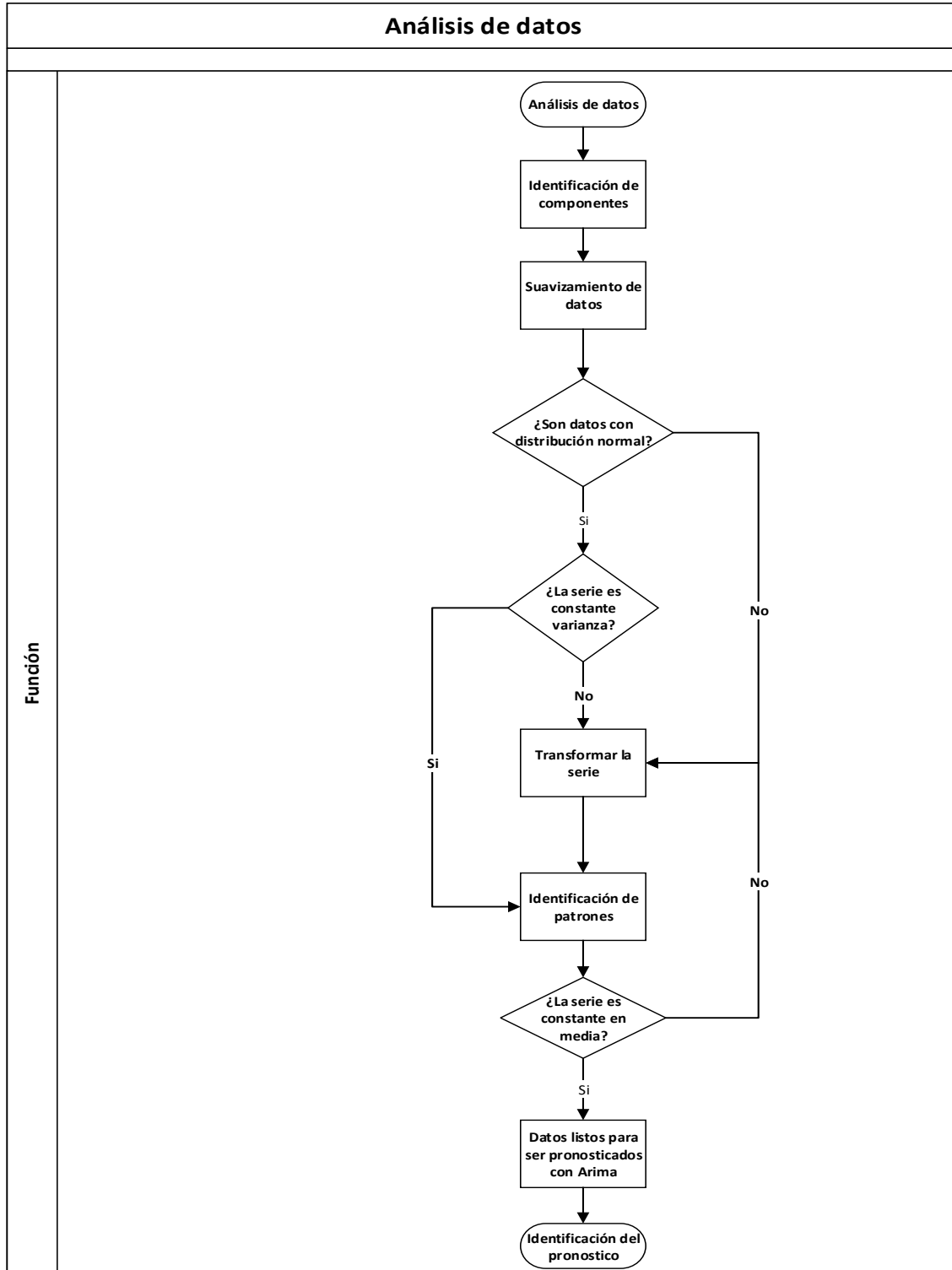
Se verificarán los resultados de la condición actual y se compararán con los estadísticos de error y posteriormente a través de una factibilidad económica, esto permitirá obtener conclusiones al respecto.

Figura 5.2 Diagrama de flujos de la metodología utilizada para pronosticar



Fuente: Elaboración propia

Figura 5.3 Diagrama de flujos de la metodología de análisis de datos



Fuente: Elaboración propia

Capítulo VI: Análisis de Resultados

En este capítulo se desarrollará el análisis completo de la metodología que se implementará por grupo de productos, se analizarán sus resultados y se llegará a obtener la mejor propuesta ARIMA, con el fin de identificar cual es la opción más óptima al momento de compararla con la situación actual expuesta en el capítulo cuatro.

1. Metodología Box – Jenkins

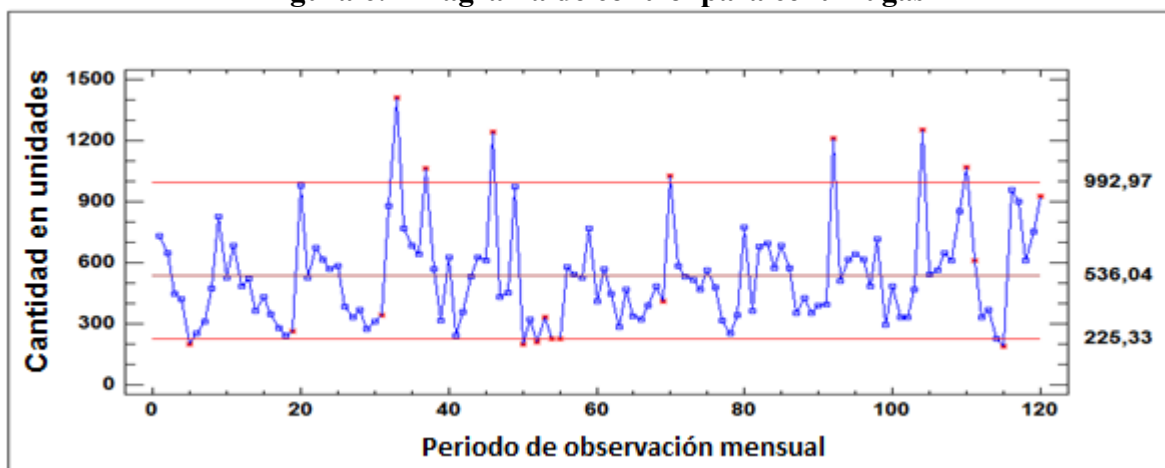
Esta metodología Box – Jenkins será la misma demostrada en el capítulo anterior, cada metodología será aplicada para cada grupo de productos en particular.

1.1 Análisis de datos e identificación del pronóstico centrifugas

1.1.1 Identificación de componentes de la serie centrifugas

En este procedimiento se estima crear una gráfico de valores individuales para electrobombas centrifugas, demarcados por límites de control, aquellos que se encuentren fuera de la región serán categorizados como datos no representativos del real comportamiento de la serie, ósea, datos clasificados cíclicos. Para el límite superior será calculado considerando un sigma de 2,5 (se opta por aquel valor porque es el más idóneo para este caso, debido a la cantidad que a priori se consideraban como datos fuera de lo común de la región de control) y para el límite inferior un sigma de 2 (esto se debe por la menor cantidad de datos bajo la media).

Figura 6.1 Diagrama de control para centrifugas



Fuente: Elaboración propia

En la gráfica (figura 6.1) anterior se puede identificar algunos componentes de datos, la serie en general tiene un comportamiento a priori constante en su media, no se aprecia algún indicio de tendencia y experimenta datos cíclicos que se clasifican como datos fuera de control.

Al calcular los límites de control superior e inferior, se puede observar que existen doce puntos fuera de la región, para los primeros se puede decir que se produjeron por causa de ventas anómalas ocasionados por ofertas y campañas de marketing, estos son: Sep – 07, Ene y oct – 08, oct – 10, ago – 12 y 13 y feb – 14, mientras que para los datos fuera de la región inferior se puede indicar que fueron generados por quiebres de stock, ocasionando pérdida en sus ventas, estos son: mayo – 05, feb, abr y jun – 09 y jul – 14, para normalizar estos datos categorizados como cíclicos, se normalizarán mediante el suavizamiento de los mismos aplicando una regresión lineal a la serie. Por lo tanto estos serán suavizados con el objetivo de ingresar los datos más representativos al método predictivo a utilizar.

En la siguiente tabla se describen los límites de control máximo y mínimo de la gráfica “X” junto con su media y su desviación estándar que es 182,773, este último cálculo servirá como parámetro comparativo, con el objetivo de discriminar cual será la herramienta predictiva que se acerca más al comportamiento general de la serie original.

Tabla 6.1 límites de control máximo - mínimo y estimadores

Gráfico X		Estimados	
Período	#1-120	Período	#1-120
LSC: +2,5 sigma	992,974	Media de proceso	536,042
Línea Central	536,042	Sigma de proceso	182,773
LIC: -2 sigma	225,327		

Fuente: Elaboración propia

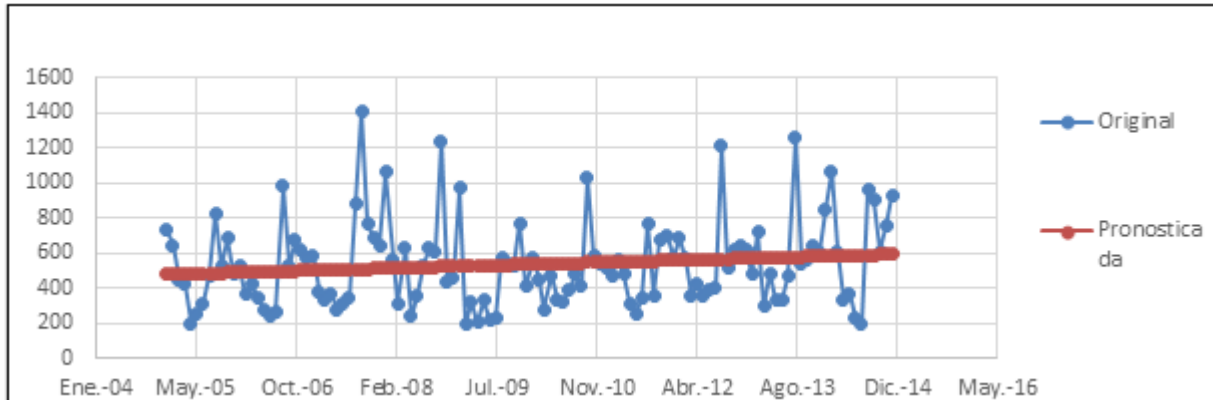
1.1.2 Suavizamiento de datos fuera de control para centrifugas

Serie utilizando la regresión lineal

Se utilizará una regresión lineal simple con el objetivo de suavizar los datos que fueron clasificados como datos fuera de control, ocupando la siguiente formulas (9) y (10).

En la siguiente figura 6.2 se describe mediante una gráfica el comportamiento de la serie original vs la serie aplicando una regresión lineal simple.

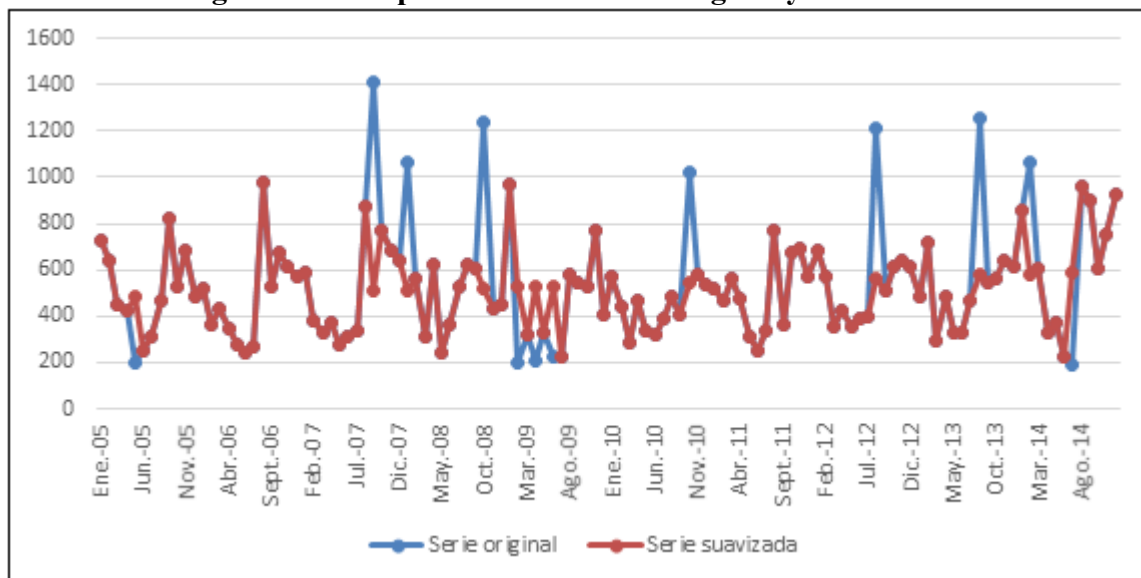
Figura 6.2 Diagrama comparativo entre la serie original y la regresionada



Fuente: Elaboración propia

En la siguiente grafica se demuestra la serie original y la suavizada, esta última será desde ahora la serie que se ocupará en los próximos análisis.

Figura 6.3 Comparación de la serie original y la suavizada



Fuente: Elaboración propia

En la comparación de ambas series de tiempo queda demostrado los datos que fueron reemplazados (en este caso doce), logrando una serie más normalizada.

Comparación de rectas aplicadas con regresiones lineales para centrifugas

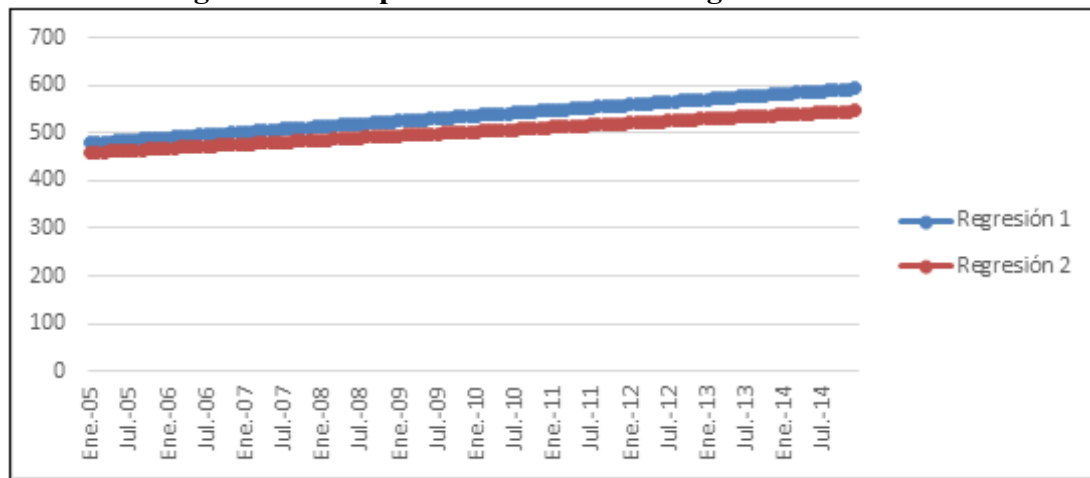
El objetivo de la comparación de ambas pendientes obtenidas en las regresiones lineales es demostrar que el suavizamiento o reemplazo de los puntos que estaban fuera de la región de límite, representen el comportamiento real de las ventas históricas y así representar una serie más sólida que permita dar resultados de pronósticos más óptimos.

Tabla 6.2 Funciones de regresión lineal recta 1 y 2

Recta 1	Recta 2
$Y = 478,05 + 0,96x$	$Y = 458,59 + 0,73x$

Fuente: Elaboración propia

Figura 6.4 Comparación de rectas de regresiones lineales



Fuente: Elaboración propia

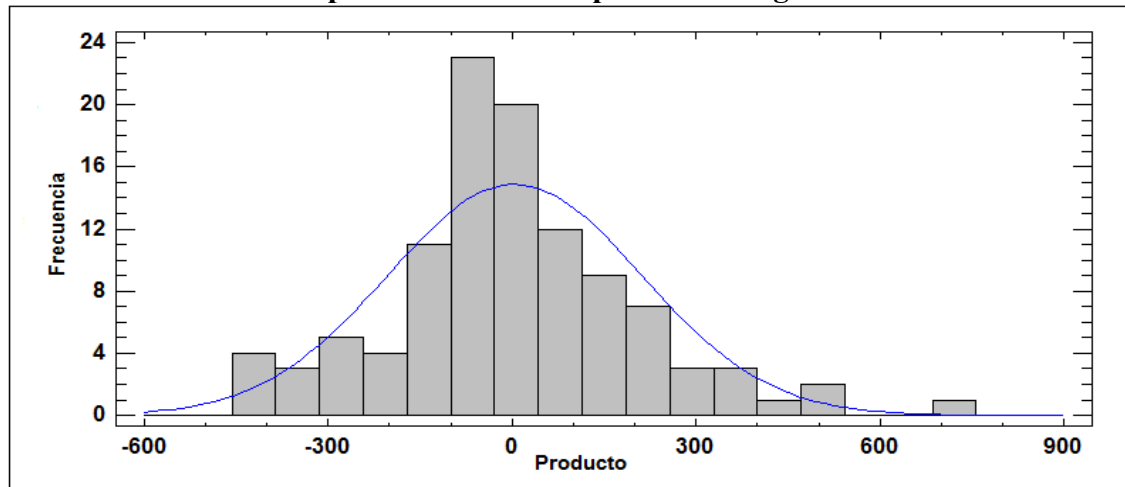
En definitiva al normalizar doce de los 120 datos, se puede observar que las rectas están levemente separadas, debido a una menor pendiente en la segunda regresión de verificación, concluyendo que el comportamiento de ambas rectas son prácticamente similares y que representan de buena forma la serie original.

Con lo anterior nos permite demostrar que con los datos suavizados se puede ahora recién analizar mediante variables estadísticas los datos y parámetros para poder aplicar algún método predictivo.

1.1.3 Comprobar normalidad de la serie para centrifugas

Uno de los requisitos para que la serie se demuestre estacionariedad es que cumple con la condición de que la serie posea un comportamiento de distribución normal. Para poder llegar a demostrar la normalidad se tuvo que transformar la serie suavizada aplicando una primera diferencia (los demás análisis se presentan en el anexo A1), en la siguiente imagen se describe.

Figura 6.5 Histograma de distribución de la normal de la serie suavizada aplicando primera diferencia para centrifugas



Fuente: Elaboración propia

En la figura 6.5 se demuestra que el conjunto de barras del histograma adquiere un comportamiento en forma de campana, demostrando a priori que existe normalidad.

En la siguiente tabla queda más claro si se comprueba o no normalidad en la serie.

Tabla 6.3 Prueba Shapiro – Wilk Electrobombas centrifugas

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,973204	0,203974

Fuente: Elaboración propia

La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos, este estadístico permite demostrar si existe o no una distribución normal.

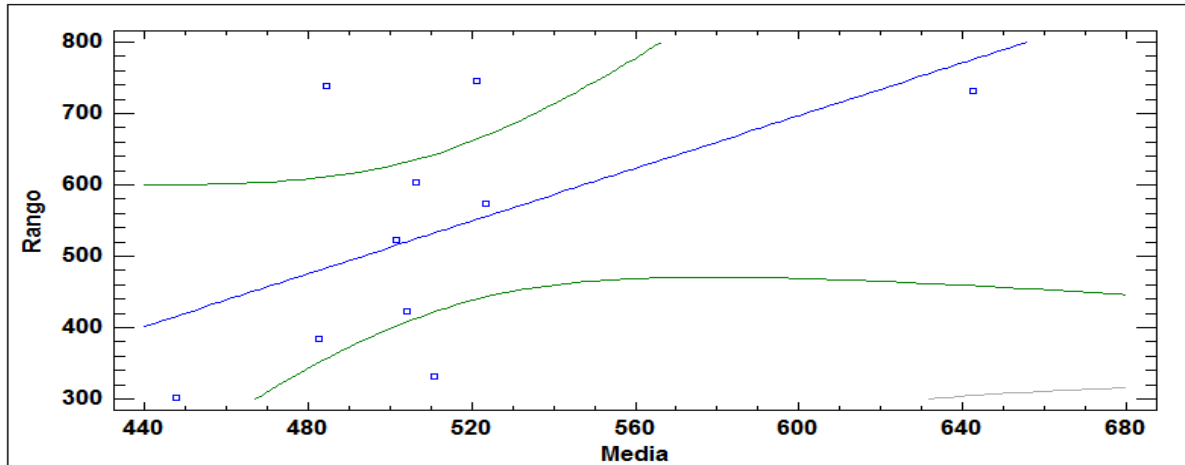
Debido a que el valor-P más pequeño de las pruebas realizadas es mayor a 0,05, no se puede rechazar la idea de que la serie transformada proviene de una distribución normal con 95% de confianza.

1.1.4 Análisis de varianza para centrifugas

Otros de los requisitos para demostrar la estacionariedad es verificar si la serie presenta crecimientos iguales de varianza a medida que pasa el tiempo, ósea, una varianza constante, para ello es necesario confeccionar un gráfico rango - medio, con la finalidad de demostrar que no exista una relación lineal entre ambas variables.

Para comprobar el concepto de homocedasticidad no fue necesario transformar la serie, en la siguiente grafica queda detallado el comportamiento de las dos variables.

Figura 6.6 Grafico rango – media serie suavizada para centrifugas



Fuente: Elaboración propia

En la figura 6.6 se puede concluir que no se aprecia una dispersión de puntos que estén entorno a una recta lineal, comprobando que a priori no existe heterocedasticidad de datos.

La ecuación del modelo ajustado es:

$$\text{Lineal: } Y = b \cdot X = \text{Rango} = -411,44 + 1,84754 \cdot \text{Media}$$

Como el valor del estadístico Durbin-Watson es = 2,01, demuestra estar cercano a 2 y su valor $-P$ es 0,45 (siendo mayor a 0,05, lo que implica rechazar la hipótesis nula que dice que la serie no tiene varianza constante), por lo que se comprueba la homocedasticidad de los datos, ósea, que la serie mantiene su varianza constante.

1.1.5 Identificación de patrones para centrifugas

El objetivo es identificar los siguientes patrones:

- Comprobar si la serie es aleatoria.
- Verificar si los datos están correlacionados, en caso de obtener un número muy elevado podría ser indicio de que exista algún patrón de tendencia.
- Si la serie tiene algún patrón estacionario.
- Comprobar estacionariedad de datos.

Para comenzar a justificar los correlogramas, primero hay que comprobar la aleatoriedad de los datos, en este caso se contrarrestará el 1 er desfase, esto porque sobrepasa la región límite llegando a un valor de 0,28 (lo máximo 0,2) por medio de las siguientes hipótesis.

$$\begin{aligned} H_0 : p_1 &= 0 \\ H_1 : p_1 &\neq 0 \end{aligned}$$

Los límites de decisión de la prueba utilizando un intervalo de un 95% de confianza y por medio de la tabla de distribución normal es:

$$0 \pm 1,96 \left(\frac{1}{120} \right)$$

Y el estadístico de prueba es:

$$t = \left(\frac{r_1 - p_1}{SE_{r_1}} \right) = \left(\frac{r_1 - 0}{SE_{r_1}} \right) = \left(\frac{r_1}{SE_{r_1}} \right)$$

SE es la desviación estándar, para el segundo retraso en adelante esta se obtiene con la formula (14).

En este caso el valor considerado será del primer retraso, por lo que el valor calculado será, $SE = \frac{1}{\sqrt{n}}$, $SE = 0,91$

Al revisar el primer desfase en la tabla 6.4 el valor $t = 1,24$ por lo tanto $1,24 < 1.96$, en consecuencia no se rechaza la hipótesis nula, la serie es aleatoria, comprobando la insignificancia del primer desfase.

Luego por medio del estadístico Q de Ljung – Box (LBQ), se sabrá si los primeros datos están correlacionados, ósea, si realmente se comprueba la aleatoriedad de los datos en conjunto. Esto se calcula con la formula (15), se recomienda utilizar como base los primeros 10 retrasos.

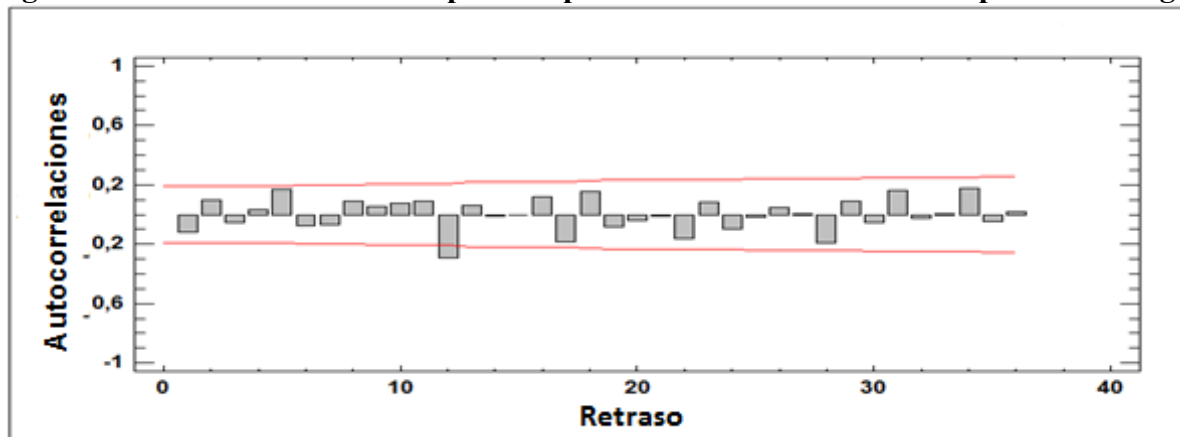
Por lo tanto en este caso, se puede decir que en los primeros 10 retrasos un valor de: $LQB = 9,79 < 18,31$ (valor obtenido de la tabla Chi – cuadrada con 10 gados de libertad y una significancia de un 5%, presente en la tabla 6.4), lo que demuestra que en las auto correlaciones para los primeros 10 retrasos como grupo no son significativamente distintos de 0, comprobando que los datos no están correlacionados y demostrando que la serie al aplicar la primera diferencia estacional se estabiliza.

El otro punto a considerar es confirmar si existe algún patrón estacional en la serie, efectivamente si existe una parte estacional descrita en las figura 6.7 y 6.8, en ella se repiten valores significativos en los desfases múltiplos de 12, en este caso es un valor el 12 (valores mensuales).

Y el último punto es comprobar si la serie es estacionaria, se puede apreciar en los FAC y FACP siguientes (figura 6.7 y 6.8 respectivamente), en ellos los datos al transformarlos con una diferencia son más estables (comparados con los correlogramas sin aplicar primera diferencia, esta se encuentra en el anexo A2), con todos los valores dentro de la región en la parte regular y con un valor significativo en el primer desfase de la parte estacional, lo cual no repercute de ninguna manera, esto debido a que baja rápidamente a 0.

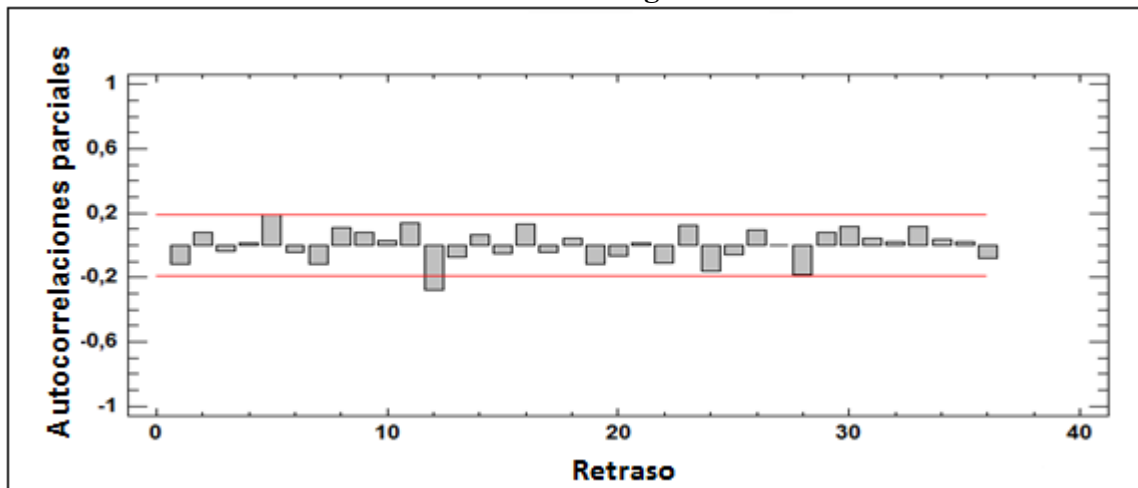
En definitiva se demuestra la estacionariedad en la parte regular y en la parte estacional, se puede decir además que no debería ser necesario aplicar otra diferencia.

Figura 6.7 FAC serie suavizada aplicada primera diferencia estacional para centrifugas



Fuente: Elaboración propia

Figura 6.8 FACP serie suavizada aplicada primera diferencia estacional para centrifugas



Fuente: Elaboración propia

En la siguiente tabla presenta los patrones del FAC de la serie suavizada transformada en logaritmo natural y primera diferencia.

Tabla 6.4 Estadísticos del FAC aplicando primera diferencia estacional

Desfase	ACF	T	LBQ
1	-0,12	-1,24	1,59
2	0,10	0,99	2,64
3	-0,06	-0,58	3,01
4	0,03	0,33	3,13
5	0,17	1,75	6,59
6	-0,07	-0,72	7,22
7	-0,07	-0,65	7,74
8	0,09	0,89	8,73
9	0,06	0,55	9,12
10	0,07	0,71	9,79

Fuente: Elaboración propia

En la tabla anterior se describe el ACF (valor referencial por cada desfase, si es mayor a 0,2 aquel es un desfase significativo), t (estadístico de prueba para demostrar si un desfase es aleatorio) y LBQ (estadístico Q de Ljung – Box, comprueba si los desfases o retardos como grupo están correlacionados).

Revisado y analizado todos los puntos anteriormente expuestos, se comprueba que los datos transformados (aplicando primera diferencia estacional) están aptos para poder ser pronosticados mediante un método del tipo ARIMA.

1.1.6 Lectura de correlogramas para centrifugas

El objetivo es identificar o descifrar los posibles pronósticos a través de la lectura de los correlogramas (FAC, FACP), utilizando el material de apoyo de los distintos ejemplos de posibles comportamientos de pronósticos ARIMA, hay que considerar que estos se utilizan como referencia, ya que dependerá mucho de la interpretación.

Al revisar los correlogramas aplicando la primera diferencia, se puede leer desde las imágenes 6.7 y 6.8, que su comportamiento en la parte regular es estable, tanto así que no posee ningún valor significativo, por lo tanto se concluye que no es necesario aplicar ningún parámetro estadístico, las posibles interpretaciones se harán en la parte estacional, según los siguientes puntos:

- Al existir un comportamiento similar en ambos correlogramas (FAC y FACP) se puede decir que en la parte estacional primero fue sometida a una diferencia, por lo que tiene una componente integrativa $D(1)$, luego demuestra que tanto el en FAC como en el FACP existe un valor fuera de la región, este es el primer retraso (12) y cae rápidamente a 0, por lo que se podría describir según la figura (3.12) que su comportamiento

demuestra una componente AR(1) más un MA(1) . En definitiva el pronóstico mixto describe un SARIMA (0,0, 0) x (1,1,1)₁₂.

- Aquí se puede interpretar que en la FAC sus desfases caen a cero de manera rápida, mientras que en la FACP cae a la región luego del primer retardo, otorgando un parámetro auto regresivo de orden 1. Por lo que en síntesis la lectura final es un SARIMA (0,0,0) x (1,1,0)₁₂.
- Se puede decir que los correlogramas tienen un comportamiento que posee un término de media móvil del término 1, esto debido a que en la FAC sus retrasos caen de golpe a la región límite luego del primer desfase (retraso 12) y en la FACP llega a obtener valores insignificantes luego del primer retraso estacional. Por lo que en síntesis la lectura final es un SARIMA (0,0,0) x (1,1,0)₁₂.

1.2 Estimación del pronóstico para centrifugas

1.2.1 Selección de periodos de predicción y validación para centrifugas

El número de datos considerados para la predicción será de 12 meses, con el objetivo de obtener una proyección que considere un año y así planificar de mejor forma el abastecimiento, la idea es ir monitoreando anualmente los resultados del pronóstico elegido, para ir validando los resultados constantemente. En cuanto a los datos para validación siempre se recomienda utilizar por lo menos 1/3 de los datos totales, es por aquello que se optaron por 48 datos, además se logró verificar que con esta cantidad al momento de iterar (en el *statgraphics*) se llegaban a mejores resultados, permitiendo así interpretar los datos de validación y estimación de manera más precisa, esta cantidad se replicará para todos los demás grupos de productos próximos a analizar.

1.2.2 Descripción de los modelos estimados para centrifugas

En esta etapa como se explicó en la metodología Box – Jenkins la idea es revisar los parámetros de cada pronóstico, con el fin de verificar si hay que reducir o no la complejidad del mismo. A continuación se presentan los distintos modelos identificados y calculados por el programa “*Statgraphics*”.

SARIMA (0,0,0)x(1,1,1)12

Al modelar este pronóstico se logaron calcular los siguientes parámetros:

Tabla 6.5 Estadísticos estimados SARIMA (0,0,0)x(1,1,1)12 para centrifugas

Parámetro	Estimado	Error Estd.	t	Valor-P
SAR(1)	-0,128336	0,14597	-0,879193	0,382925
SMA(1)	0,846058	0,0556032	15,216	0

Fuente: Elaboración propia

En la tabla anterior se puede apreciar que el parámetro auto regresivo tiene un valor – P que es mayor a 0,05, por lo que se recomienda reducir este a cero, mientras que el otro parámetro demuestra que es significativo, obteniendo en definitiva un pronóstico estimado SARIMA (0,0,0)x(0,1,1)12, este se describe a continuación.

SARIMA (0,0,0)x(0,1,1)12

El presente pronóstico está compuesto por un solo parámetro que es de media móvil estacionario, por lo que se cataloga como un modelo simple, en la siguiente tabla se demuestra todos sus indicadores.

Tabla 6.6 Estadísticos estimados SARIMA (0,0,0)x(0,1,1)12 para centrifugas

Parámetro	Estimado	Error Estd.	t	Valor-P
SMA(1)	0,867068	0,0470591	18,4251	0

Fuente: Elaboración propia

Como se puede apreciar el presente modelo (tabla 6.6) demuestra ser el ideal para este caso, sin necesidad de reducir su complejidad.

SARIMA (0,0,0)x(1,1,0)12

Tabla 6.7 Estadísticos estimados SARIMA (0,0,0)x(1,1,0)12 para centrifugas

Parámetro	Estimado	Error Estd.	t	Valor-P
SAR(1)	-0,704623	0,0943656	-7,46695	0

Fuente: Elaboración propia

En la presente tabla 6.7 se comprueba que la composición del modelo actual es el recomendable, esto debido a que todos sus parámetros demostraron ser significativos.

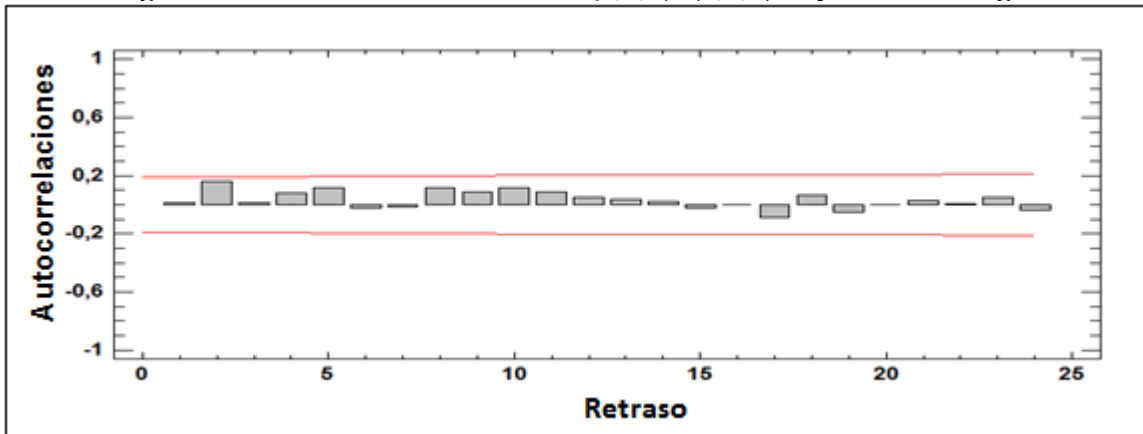
1.3 Validación de pronósticos para centrifugas

1.3.1 Verificación e interpretación de los correlogramas residuales

En las siguientes figuras 6.9 y 6.10 se presentarán los correlogramas del pronóstico a seleccionar el SARIMA (0,0,0)x(0,1,1)₁₂, los demás se presentarán en los anexos (A3).

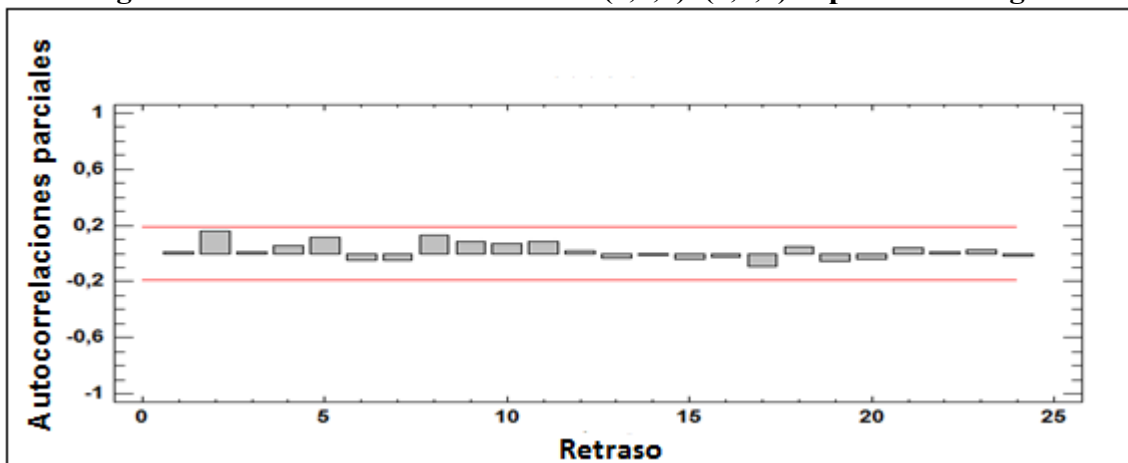
En las siguientes imágenes se proponen los correlogramas residuales tanto para el FAC Y FACP, el objetivo es demostrar si los resultados del pronóstico propuesto describen ruido blanco, para que aquella condición se cumpla no tiene que existir ningún desfase que sea realmente significativo, ósea, que esté fuera de la región trazada con rectas rojas, normalmente se consideran como realmente significativos valores mayores a 0,3 tanto en negativo como en positivo.

Figura 6.9 FAC residual SARIMA (0,0,0)x(0,1,1)₁₂ para centrifugas



Fuente: Elaboración propia

Figura 6.10 FACP residual SARIMA (0,0,0)x(0,1,1)₁₂ para centrifugas



Fuente: Elaboración propia

En las gráficas anteriores 6.9 y 6.10 se puede decir que todos sus desfases en ambos correlogramas se encuentran dentro de la región límite, comprobando que posee aleatoriedad, esto al rectificar que a través de la prueba de Box-Pierce el valor $-P$ es 0,96 demostrando ser mayor a 0,05, (considerando un nivel de significancia del 95%) por que se comprueba que la serie tiene ruido blanco y se verifica que este pronóstico puede proyectar datos futuros de manera sólida.

Con respecto a los demás pronósticos todos demuestran un comportamiento similar comprobando RB, por lo tanto todos los pronósticos estimados son adecuados para proyectar este tipo de serie de tiempo.

1.3.2 Comparación de modelos estimados y validados para centrifugas

Para elegir el modelo que más se acerca a la serie original, se van a calcular los parámetros estadísticos considerados en la etapa de estimación y validación estos son: RMSE, MAD y MAPE, dando una mayor importancia al parámetro estadístico RMSE ya que otorga una mayor ponderación a los residuos grandes, al elevar al cuadrado el error de un periodo, permitiendo discriminar de manera más precisa cual es el mejor pronóstico, además los datos considerados con mayor importancia serán los obtenidos en la etapa de validación, con el objetivo de elegir la herramienta predictiva que se acerque más al comportamiento de los datos más recientes.

En la comparación el pronóstico que arroje mejores resultados en la etapa de validación tal como se dijo anteriormente será el elegido, primero se analizarán los estadísticos por etapas, estos son descritos en las siguientes tablas 6.8 y 6.9.

Etapas de estimación

En esta etapa tal como se describió en el punto 1.2.1 de este capítulo los datos a utilizar serán 60, correspondientes a los meses más antiguos de la serie, estos abarcan desde enero del 2006 hasta diciembre del 2010, a continuación se describe la tabla con los cálculos de sus respectivos estadísticos.

Tabla 6.8 Comparación de los distintos estadísticos de error por pronóstico en etapa de estimación para centrifugas

Comparación de los distintos pronósticos en la etapa de estimación		
Estadístico	SARIMA (0,0,0)x(0,1,1) ₁₂	SARIMA (0,0,0)x(1,1,0) ₁₂
RMSE	131,108	148,031
MAE	107,659	119,242
MAPE	23,3958	24,5314

Fuente: Elaboración propia

Considerando los estadísticos de niveles de error de la etapa de estimación el que mejor se acomoda a nivel general y considerando el RMSE estadístico más importante para predecir con datos históricos es el SARIMA $(0,0,0) \times (0,1,1)_{12}$.

Etapas de validación

En aquella etapa se consideran los datos más recientes, estos abarcan los últimos 48 meses, desde enero del 2011 hasta diciembre del 2014.

Tabla 6.9 Comparación de los distintos estadísticos de error por pronóstico en etapa de validación para centrifugas

Comparación de los distintos pronósticos en la etapa de validación		
Estadístico	SARIMA $(0,0,0) \times (0,1,1)_{12}$	SARIMA $(0,0,0) \times (1,1,0)_{12}$
RMSE	127,801	155,274
MAE	91,9136	116,511
MAPE	16,9658	22,0389

Fuente: Elaboración propia

Considerando los estadísticos de niveles de error de la etapa de validación el que mejor se comporta es SARIMA $(0,0,0) \times (0,1,1)_{12}$, rectificando lo expuesto en la etapa de estimación, además hay que considerar que sus estadísticos decrecen, obteniendo comportamiento mejor en esta etapa.

Por lo tanto existe un único pronóstico que representa de mejor forma esta serie de datos, obteniendo mejores resultados en los datos más próximos, por lo que cumple con los objetivos planteados por la empresa (resaltar las ventas de los últimos años).

1.4 Predicción del modelo para centrifugas

1.4.1 Descripción del modelo para centrifugas

Los parámetros estadísticos estimados por el software “*Statgraphics*” fueron indicados en la tabla 7.6 en la etapa de estimación, en base a ella se procederán a los cálculos del modelo:

Donde el coeficiente estimativo es:

$$W_1 = -0,8671$$

Hay que recordar la serie fue transformada para aplicar el pronóstico, por lo que para llegar a los resultados con la misma unidad original, será necesario reflejar la transformación, aplicando las primeras diferencias.

Utilizando las ecuaciones (33) y (32) se procederá a calcular el modelo:

$$Y_t = W_1 e_{t-12} \quad (33)$$

$$\Delta_{12} Y_t = Y_t - Y_{t-12} \quad (32)$$

Por lo tanto reemplazando ambas se tiene:

$$Y_t - Y_{t-12} = W_1 e_{t-12}$$

$$Y_t = 0,8671 e_{t-12} + Y_{t-12}$$

Con esta fórmula se podrá calcular los datos pronosticados de cada periodo, el siguiente punto se demuestra toda la serie pronosticada.

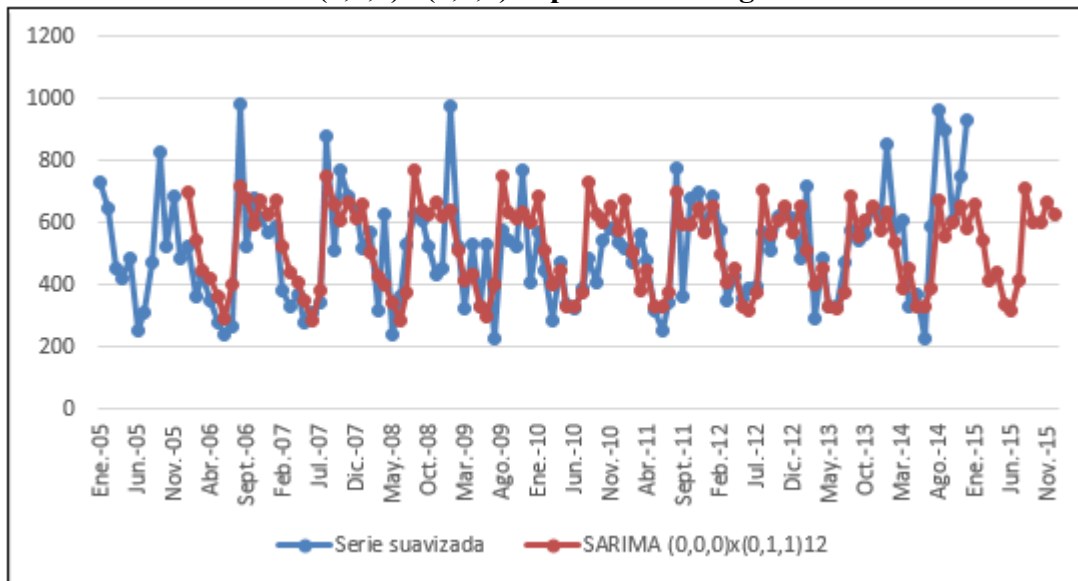
El pronóstico converge luego de correr 6 iteraciones.

La idea es describir los resultados del pronóstico actual y demostrar su comportamiento, a través de la comparación con las distintas formas de expresión de la venta historia de este tipo de productos.

Pronostico Sarima (0,0,0)x(0,1,1)12 seleccionado con 12 datos de predicción y 48 de validación.

En la siguiente grafica se describe el cálculo del pronóstico efectuado por el Software “Statgraphics” ingresando los datos transformados.

Figura 6.11 Grafico comparativo entre la serie suavizada y pronostico SARIMA (0,0,0)X(0,1,1)12 para centrifugas

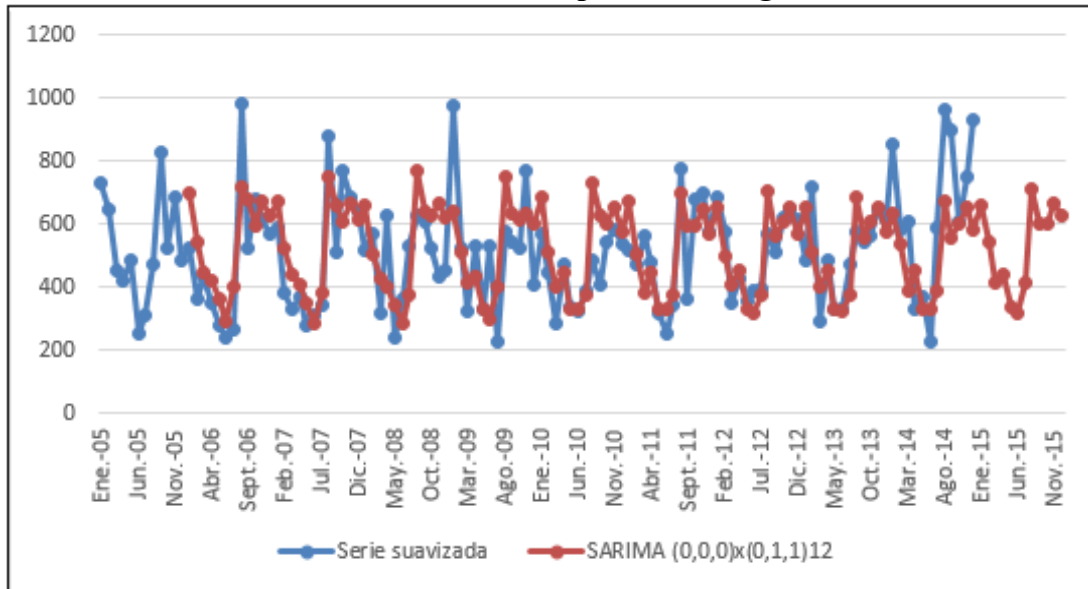


Fuente: Elaboración propia

En la gráfica(figura 6.11) anterior se puede apreciar que el resultado de la predicción replica muy bien la serie suavizada, obteniendo datos que se mantienen constantes en su media, al suavizar algunos datos muy altos y muy bajos, reflejando en definitiva un pronóstico que se estaciona en el promedio general de la serie de electrobombas centrifugas.

En la siguiente grafica se describe el pronóstico promedio por mes de los cuatro años tanto de la serie suavizada como la proyectada (2011 – 2014), esto permitirá visualizar de mejor manera el comportamiento real que toma mensualmente el pronóstico elegido y que grado de asertividad cumple.

Figura 6.12 Grafico comparativo del promedio mensual de la serie y la proyección de los últimos cuatro años para centrifugas

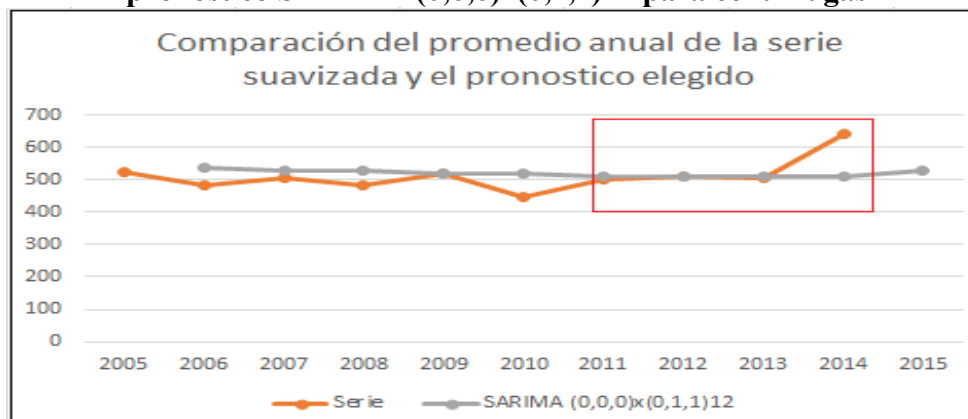


Fuente: Elaboración propia

En la gráfica 6.12 queda claro que los datos mensuales proyectados adquieren un patrón de datos similar al promedio histórico de los años más recientes, existe una leve diferencia en los meses de febrero, marzo, julio y diciembre, pero que no repercute a nivel general.

En la siguiente grafica (figura 6.13) se comprueba el comportamiento promedio anual de la serie de tiempo suavizada y el pronóstico elegido.

Figura 6.13 Grafico comparativo del promedio anual de la serie suavizada y el pronóstico SARIMA (0,0,0)x(0,1,1)12 para centrifugas



Fuente: Elaboración propia

En la figura 6.13 se puede concluir que la herramienta predictiva siempre otorga como respuesta de la media de los datos, suavizando los datos máximos y mínimos de los distintos años, demostrando que su comportamiento es mucho más cercano en los años más recientes, 2011, 2012, 2013, escapándose en el 2014 por ser un año con cifras de ventas record, esto se demuestra en el área del cuadrado de rojo.

1.5 Comparación de la situación actual y la propuesta seleccionada

En base a lo descrito anteriormente se puede resumir en la siguiente tabla comparativa:

Tabla 6.10 Comparación de estadísticos entre la situación actual y la propuesta para centrifugas

Estadístico	Situación actual	SARIMA (0,0,0)x(0,1,1) ¹²
RMSE	144,72	127,8
MAE	108,59	91,91
MAPE	19,71%	16,97%

Fuente: Elaboración propia

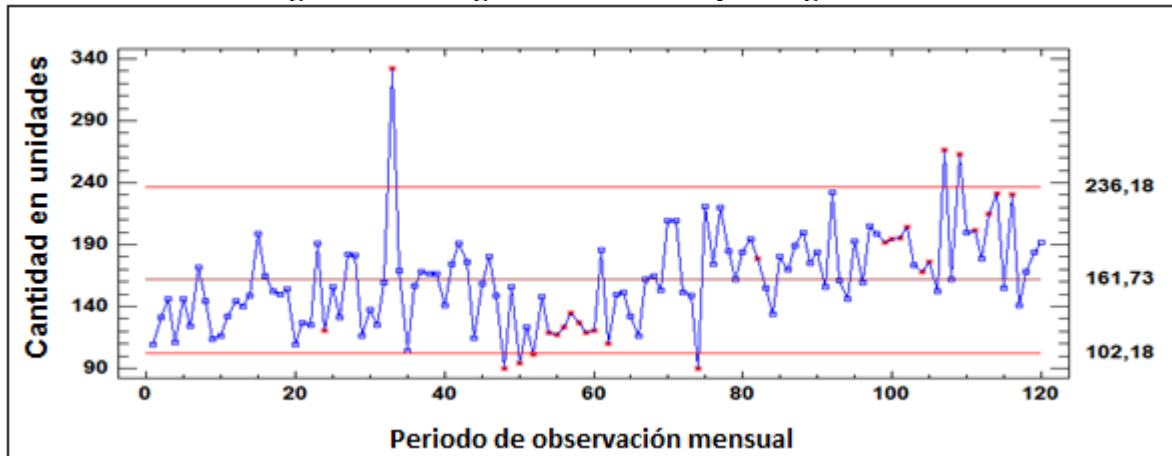
En la tabla anterior por lo tanto se puede comprobar que la propuesta ARIMA elegida, obtiene los tres estadísticos de error con valores menores en comparación con la situación actual, demostrando que es una opción viable para poder implementarse en la empresa.

1.6 Análisis de datos e identificación del pronóstico para agua sucia

1.6.1 Identificación de componentes de la serie para agua sucia

A continuación se describirá la serie de tiempo del grupo de productos clasificados como agua sucia, el registro presenta la base de datos históricos experimenta desde el 2005 hasta el 2014, con el objetivo de identificar los componentes que a priori afectarán en esta serie de datos.

Figura 6.14 Diagrama de control para agua sucia



Fuente: Elaboración propia

Se puede observar que la serie de tiempo en los primeros cuatro años se mantiene generalmente entre la media y el límite inferior, luego en el 2009 y primera parte del 2010 desarrolla una tendencia a la baja, es ahí donde experimenta algunos quiebres, para finalmente describir una tendencia a la alta en los siguientes años cuatro años (2011 – 2014). Por lo que a priori se podría decir que la media no es constante y habrá que corregirla en los siguientes puntos.

En relación a los datos cíclicos, se pueden observar que hay varios tanto en la parte alto como baja que sobresalen de los limites, causados en el primer caso por campañas de marketing y en el segundo por quiebres de stock en donde las ventas fueron muy bajas, en este último punto también se puede agregar que la ventas en general de la empresa durante el año 2009 fueron muy bajas, también a causa de cambios internos de la misma y del mercado propiamente tal.

El cálculo de la media del proceso fue de 161,7 y de su desviación estándar 28,78 experimentando una variabilidad baja explicada por la gran mayoría de los datos que se mantienen dentro de los límites de confianza.

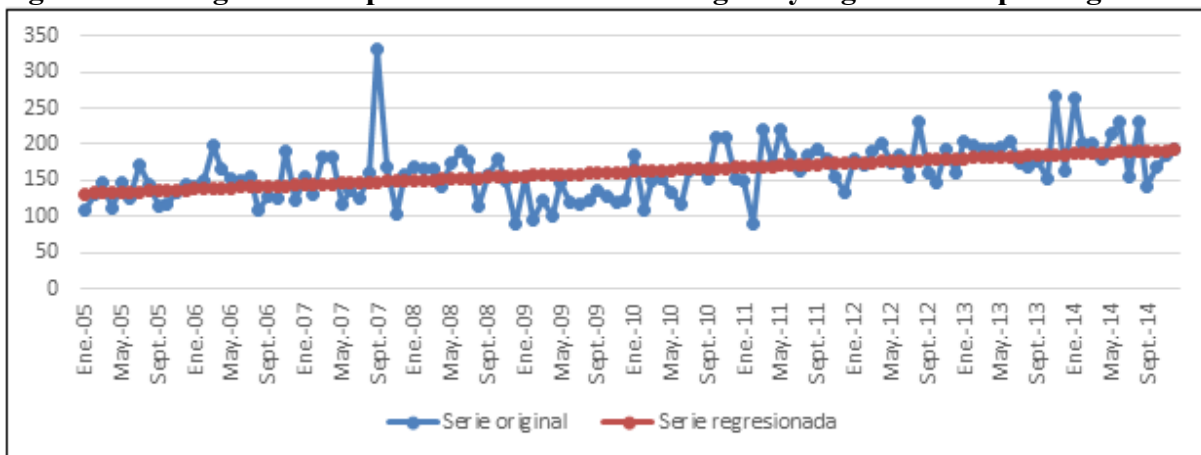
1.6.2 Suavizamiento de datos fuera de control para agua sucia

Serie utilizando la regresión lineal simple para agua sucia

El objetivo de este análisis es describir el comportamiento que se obtiene al aplicar una regresión lineal a los datos y comprobar que los datos cíclicos suavizados no afectan en nada el comportamiento general de serie de tiempo.

Los datos a suavizar serán: sep – 07, dic – 08, feb y abril – 09, feb – 11, nov – 13 y ene – 14, en las figuras 6.15 y 6.16 se presentarán tanto el comportamiento de la serie original como la regresionada y la comparación entre la serie original y suavizada.

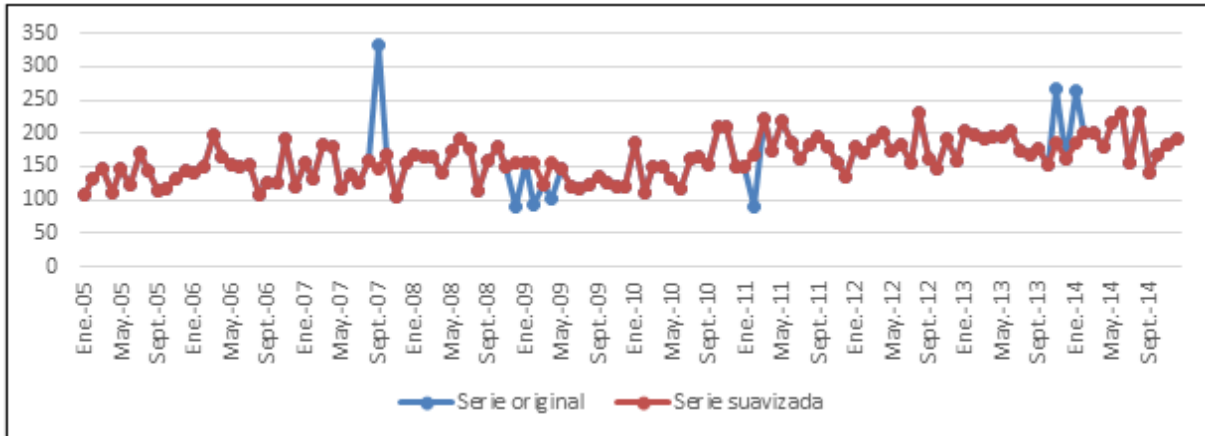
Figura 6.15 Diagrama comparativo entre la serie original y regresionada para agua sucia



Fuente: Elaboración propia

Se puede observar en la gráfica anterior que la serie regresionada demuestra un incremento a medida que pasan los años, demostrando ser justo el comportamiento que se apreció la gráfica de diagrama de control.

A continuación se describe la comparación de la serie original y la suavizada luego de aplicar la regresión lineal en los datos cíclicos.

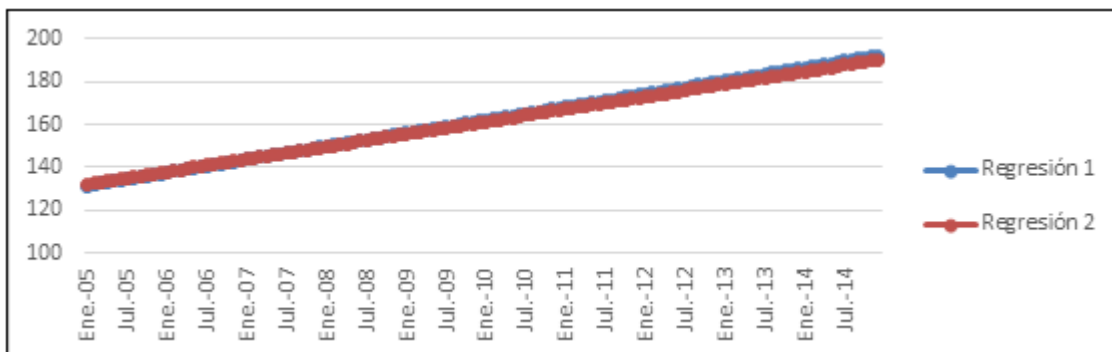
Figura 6.16 Comparación de la serie original con la serie suavizada para agua sucia

Fuente: Elaboración propia

Comparación de rectas aplicadas con regresiones lineales**Tabla 6.11 Funciones de regresión lineal para serie original y suavizada para agua sucia**

Recta 1	Recta 2
$Y = 130,95 + 0,51x$	$Y = 131,67 + 0,49x$

Fuente: Elaboración propia

Figura 6.17 Comparación de rectas aplicadas a la serie original y suavizada para agua sucia

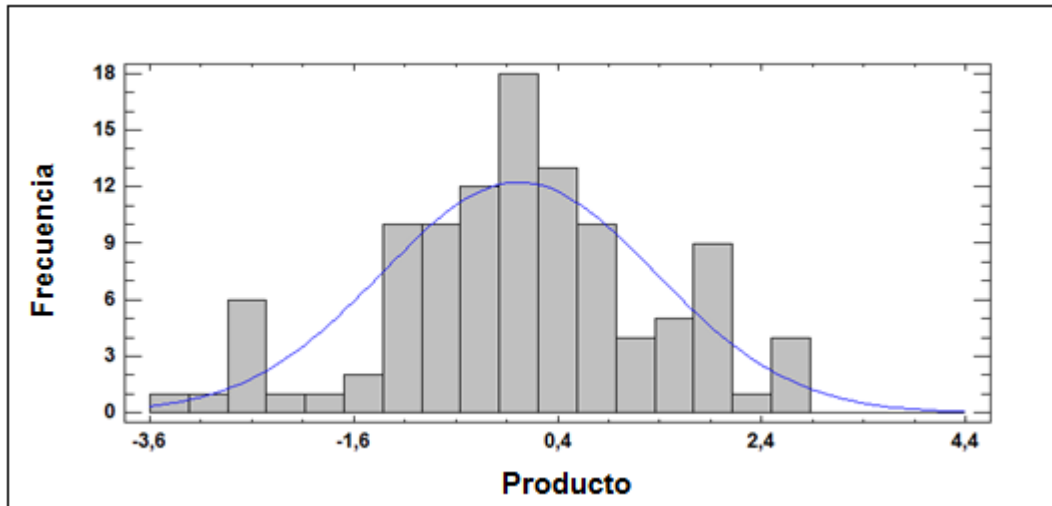
Fuente: Elaboración propia

Por lo tanto se puede concluir que la segunda recta se encuentra muy próxima a la serie original, demostrando que los datos cíclicos reemplazados son los adecuados al demostrar el verdadero comportamiento de la serie de tiempo, concluyendo que la serie suavizada está apta para poder ser analizada por los siguientes patrones.

1.6.3 Análisis de la distribución normal para Agua sucia

Para poder demostrar una distribución normal, fue necesario aplicar a la serie un par de transformaciones Cox – Box (las demás análisis de distribución normal se describen en el anexo B1), primero se estableció aplicar una raíz cuadrada y después una primera diferencia, en la siguiente figura 6.18 se demuestra el histograma de la serie intervenida.

Figura 6.18 Histograma de distribución normal de la serie suavizada para agua sucia



Fuente: Elaboración propia

Se puede decir que la figura 6.18 demuestra un comportamiento en forma de campana, ósea, cercano a una distribución normal, por lo que a priori se podría indicar que la serie tiene normalidad en sus datos, para terminar de corroborar aquello se calculará el estadístico Shapiro – Wilk, bajo la siguiente tabla:

Tabla 6.12 Prueba Shapiro – Wilk para agua sucia

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,965802	0,0545667

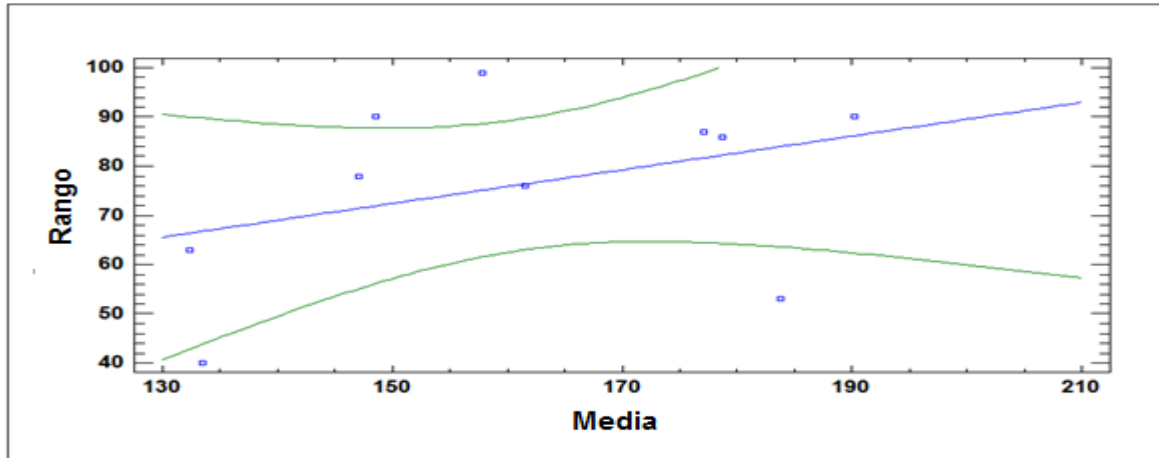
Fuente: Elaboración propia

Debido a que el valor-P más pequeño de las pruebas realizadas es mayor ó igual a 0,05, no se puede rechazar la idea de que este serie de datos proviene de una distribución normal con 95% de confianza.

1.6.4 Análisis de varianza

El objetivo de este análisis es demostrar si la serie presenta homocedasticidad y por ende demuestre una varianza constante, en la siguiente grafica se demuestra el comportamiento del rango y media de este tipo de productos.

Figura 6.19 Grafico rango – media de la serie suavizada para agua sucia



Fuente: Elaboración propia

Para este caso particular en la figura 6.19 no fue necesario aplicar alguna transformación, esto porque la gráfica describe un comportamiento que a priori presenta homocedasticidad, para justificar lo anterior hay que revisar además si el valor de la prueba Dubin Watson es cercano a 2, efectivamente el valor calculado llega a ser 2,49 y su estadístico valor – P es igual a 0,7 demostrando ser mayor a 0,05, por lo tanto se comprueba que no hay indicio de que las variables tengan algún relación lineal y se concluye que la serie tiene una varianza que se mantiene constante.

La configuración de los componentes de la recta queda de la siguiente manera:

Lineal: $Y = a + b \cdot X = \text{Rango} = 21,0616 + 0,342332 \cdot \text{Media}$.

1.6.5 Identificación de patrones

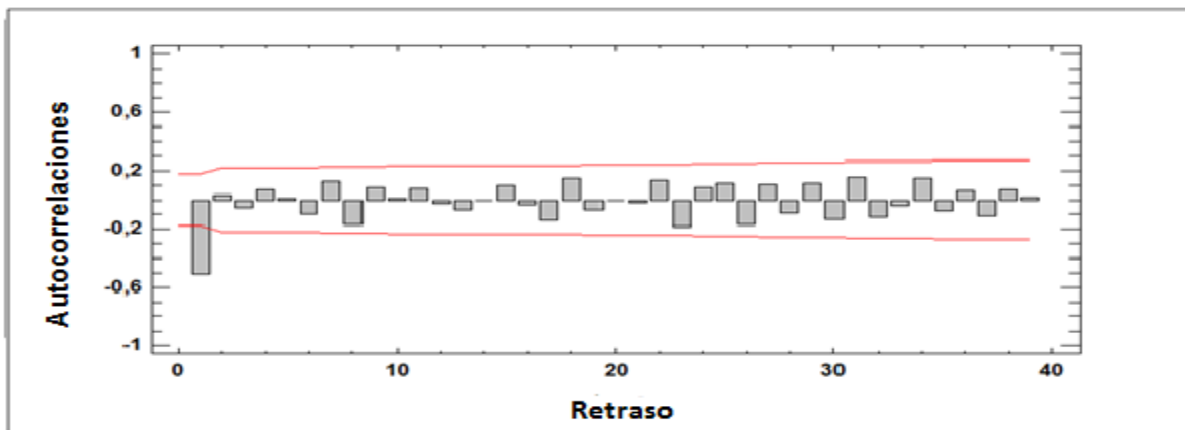
Tal como se describió en el producto anterior, primero hay que demostrar si la serie presenta aleatoriedad en el primer desfase, para ello hay que revisar la tabla el valor $t = 5,54$ negativo por lo tanto $5,54 > 1.96$, en consecuencia se rechaza la hipótesis nula, la serie no es aleatoria, comprobando que el primer desfase es significativo.

Luego por medio del estadístico Q de Ljung – Box (LBQ), hay que verificar si la serie está correlacionada, al revisar el valor de: $LQB = 40,6 > 18,31$ (valor obtenido de la tabla Chi – cuadrada con 10 gados de libertad y una significancia de un 5%, presente en la tabla 6.13), se demuestra que en las auto correlaciones para los primeros 10 retrasos como grupo son significativamente distintos de 0, principalmente se comprueba la correlación de datos, debido al gran nivel de significancia obtenido por el primer retraso, lo que hace que arrastre a los 10 primeros de forma conjunta, esto no debería repercutir en la decisión si la serie es o no estacionaria, considerando que experimenta un solo retraso significativo.

Además se puede apreciar que en los correlogramas aplicando la transformación de raíz cuadrada (ubicados en el anexo B2) se estabilizan, demostrando que no existe un patrón estacional, por lo que sólo se manifiesta una parte regular, aquella experimenta desfases que no caen rápidamente a la región de confianza, por lo que se confirma que existe una tendencia demarcada y por ende una media que no se demuestra constante en el tiempo, concluyendo que la serie no presenta estacionariedad y que por lo tanto se le recomienda aplicar una primera diferencia.

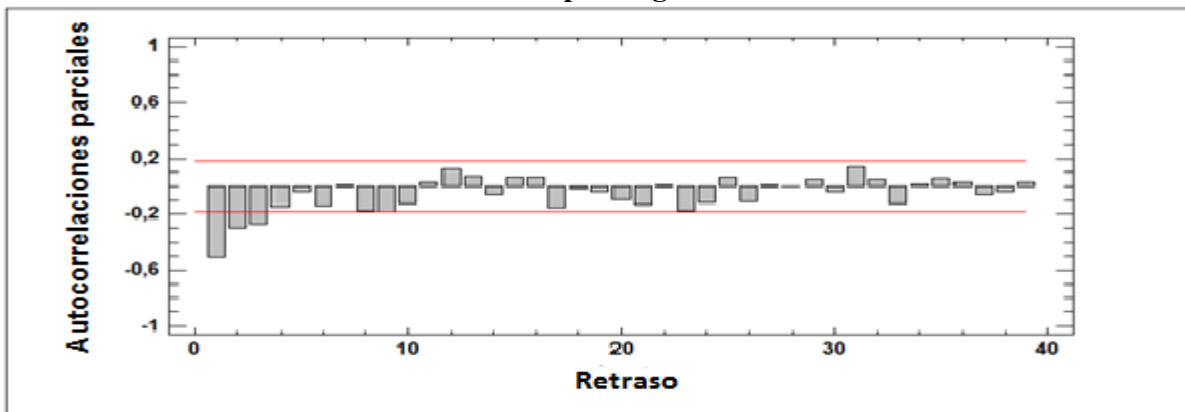
A continuación se presentan los correlogramas sometidos adicionalmente a una primera diferencia:

Figura 6.20 FAC de la serie suavizada y transformada con raíz cuadrada y primera diferencia para agua sucia



Fuente: Elaboración propia

Figura 6.21 FACP (Función de Autocorrelación Parcial) de la serie suavizada y transformada con raíz cuadrada y primera diferencia para agua sucia



Fuente: Elaboración propia

Al aplicar la primera diferencia en la figura se puede decir que la serie demuestra un comportamiento en sus retrasos más estable, debido a que la gran mayoría de sus desfases no son significativos, comprobando que la serie presenta estacionariedad y por ende se encuentra lista para poder identificar y aplicar alguna herramienta predictiva ARIMA.

Tabla 6.13 Estadísticos del FAC aplicando transformación raíz cuadrada y primera diferencia

Desfase	ACF	T	LBQ
1	-0,51	-5,54	31,42
2	0,03	0,31	31,56
3	-0,06	-0,49	31,94
4	0,07	0,63	32,58
5	0,01	0,10	32,60
6	-0,10	-0,87	33,85
7	0,13	1,10	35,86
8	-0,17	-1,50	39,73
9	0,08	0,68	40,58
10	0,01	0,11	40,60

Fuente: Elaboración propia.

En la tabla anterior se describe el ACF (valor referencial por cada desfase, si es mayor a 0,2 aquel es un desfase significativo), t (estadístico de prueba para demostrar si un desfase es aleatorio) y LBQ (estadístico Q de Ljung – Box, comprueba si los desfases o retardos como grupo están correlacionados).

1.6.6 Lectura de correlogramas

Tal como se describió anteriormente sólo existe una parte regular, para aquella todas las lecturas tendrán adicionalmente una integración del orden uno, esto a causa de la transformación de la primera diferencia aplicada para estabilizar los desfases tanto de la FAC como la FACP (figuras 6.20 y 6.21):

- Se puede indicar que en la FAC luego del primer desfase significativo los retrasos se encuentran estables, mientras que en la FACP los primeros tres desfases caen a cero de manera rápida, comprobando que es un claro comportamiento de media móvil, demostrando que el pronóstico leído es un ARIMA (0,1,1).
- En este punto se puede identificar que los correlogramas presentan una componente auto regresiva, esto debido a que experimenta una caída brusca hacia cero en la FAC y que luego de su tercer desfase llega a la región de confianza en la FACP, obteniendo en definitiva un término AR(3). Obteniendo un ARIMA (3,1,0).

- En este punto se podría asumir un comportamiento similar en ambos correlogramas, independiente en que desfase llegó a cero, manifestando que todos los retardos fuera de la región llegan a ser insignificantes rápidamente. Por lo tanto el comportamiento identificado es un ARIMA (1,1,1).

1.7 Estimación del modelo para agua sucia

1.7.1 Descripción de los modelos estimados para agua sucia

ARIMA (0,1,1)

En la siguiente tabla se demuestra la composición del modelo estimado en esta ocasión:

Tabla 6.14 Estadísticos estimativos ARIMA (0,1,1) para agua sucia

Parámetro	Estimado	Error Estd.	t	Valor-P
MA(1)	0,833908	0,0730064	11,4224	0

Fuente: Elaboración propia

En la tabla 6.14 se puede decir que el parámetro de media móvil presenta un valor – P menos a 0,05 por lo que demuestra ser significativamente diferente de cero. Permitiendo mantener el modelo actual sin modificaciones.

ARIMA (3,1,0)

Tabla 6.15 Estadísticos estimativos ARIMA (3,1,0) para agua sucia

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	-0,736573	0,11392	-6,46571	0
AR(2)	-0,555579	0,129655	-4,28507	0,000059
AR(3)	-0,377709	0,115726	-3,26383	0,001722

Fuente: Elaboración propia

Los parámetros auto regresivos de la tabla 6.15 que posee este modelo mantienen un valor – P menor que 0,05, comprobando que el modelo actual no necesita ser reducido.

ARIMA (1,1,1)

En la siguiente tabla se podrá demostrar que el único parámetro presente en este modelo es estadísticamente distinto de cero.

Tabla 6.16 Estadísticos estimativos ARIMA (1,1,1) para agua sucia

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	0,184669	0,118438	1,5592	0,123525
MA(1)	0,993569	0,00871389	114,021	0

Fuente: Elaboración propia

El termino MA(1) describe un valor – P mayor que 0,05, por lo que se recomienda reducir su sub – índice a cero, mientras que el termino AR(1) comprueba que su valor – P tiene significancia, por lo que se concluye que el presente modelo tiene que reducir su complejidad, llegando a un modelo ARIMA (0,1,1), este ya fue analizado y comprobado que es el adecuado para esta etapa.

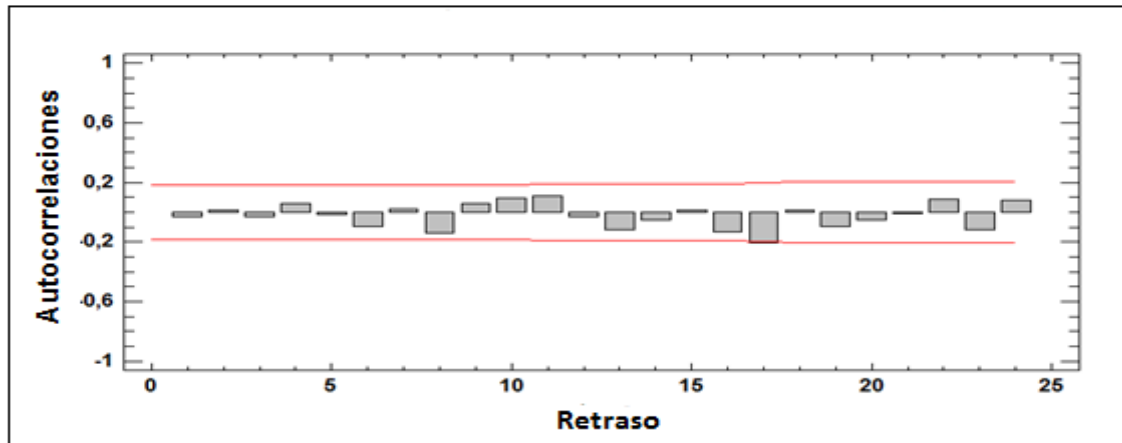
1.8 Validación del modelo

1.8.1 Verificación e interpretación de correlogramas residuales para agua sucia

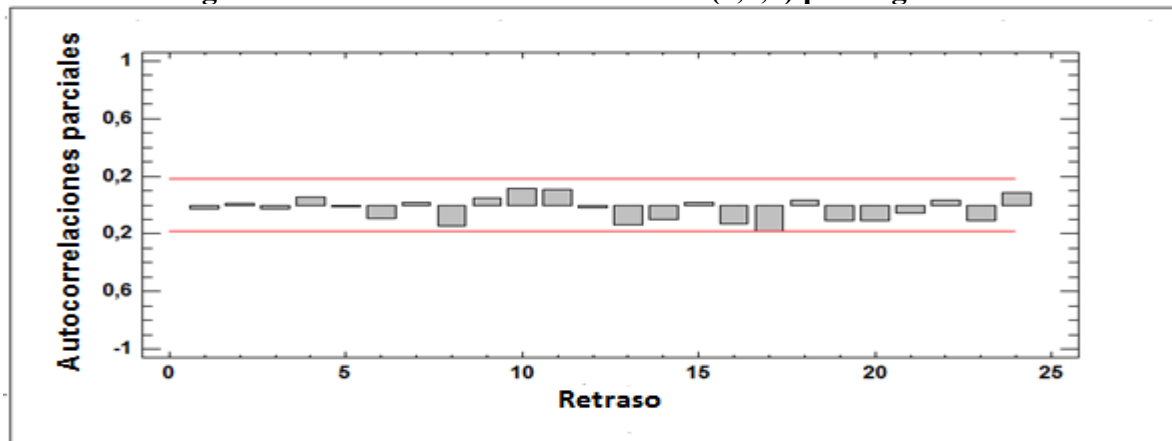
El objetivo de la revisión de este punto es revisar los correlogramas residuales, identificar si existe algún desfase realmente significativo para poder demostrar si la serie proyectada presenta o no RB.

En las siguientes figuras se presentarán a continuación, se describirán los correlogramas del pronóstico que se elegirá más adelante el ARIMA (0,1,1), los demás se presentarán en los anexos.

En las siguientes gráficos se presentan los comportamientos que demuestran RB del pronóstico que se elegirá al término de este punto (los demás serán presentados en los anexos B3).

Figura 6.22 FAC residual de ARIMA (0,1,1) para agua sucia

Fuente: Elaboración propia

Figura 6.23 FACP residual de ARIMA (0,1,1) para agua sucia

Fuente: Elaboración propia

Se puede verificar que los correlogramas no presentan ningún desfase realmente significativo, lo que demuestra que a nivel general están muy estables, a priori se podría decir que la serie demuestra en sus residuos un comportamiento con aleatoriedad, para rectificar aquello se realizó además la prueba Box – Pierce con el objetivo de contrarrestar la hipótesis nula que dice que la serie no tiene RB, al obtener un valor de $0,57 > 0,05$, se puede concluir que se rechaza la hipótesis nula y la serie presenta por ende RB.

Este mismo análisis se hizo en los demás pronósticos estimados, logrando comprobar que el modelo ARIMA (3,1,0) también presenta RB, concluyendo que de los tres pronósticos identificados y estimados, dos de ellos son válidos ARIMA (0,1,1) y ARIMA (3,1,0).

1.8.2 Comparación de modelos estimados y validados para agua sucia

Etapas de estimación

En esta etapa se verificarán los datos más antiguos, demostrando resultados parciales de la herramienta predictiva frente a la serie de tiempo, es por aquello que el pronóstico seleccionado acá será considerado pre – seleccionado y deberá comprobar sus resultados si son los menores o no en la etapa de validación.

Tabla 6.17 Comparación de los distintos estadísticos de error por pronóstico en la etapa de estimación para agua sucia

Comparación de los distintos pronósticos en la etapa de estimación		
Estadístico	ARIMA (0,1,1)	ARIMA (3,1,0)
RMSE	25,77	26,1363
MAE	20,3681	20,8498
MAPE	13,9052	14,2932

Fuente: Elaboración propia

Se puede apreciar que de los dos modelos presentados, el que obtiene menores resultados en los tres estadísticos de error es el ARIMA (0,1,1) por lo que se concluye que en esta etapa aquel modelo es el que más se acerca a la serie.

Etapas de validación

En la etapa de validación al ser la más importante permite describir el comportamiento del pronóstico que se acerque más a la serie, pero concentrándose en los datos más recientes, es por aquello que el modelo elegido en esta etapa será el que definitivamente cumpla con los objetivos iniciales.

Tabla 6.18 Comparación de los distintos estadísticos de error por pronóstico en la etapa de validación para agua sucia

Comparación de los distintos pronósticos en la etapa de validación		
Estadístico	ARIMA (0,1,1)	ARIMA (3,1,0)
RMSE	25,501	27,9247
MAE	20,5105	22,6946
MAPE	11,323	12,7855

Fuente: Elaboración propia

En la tabla 6.18 se puede indicar que el modelo que arroja mejores resultados es el mismo que fue seleccionado en la etapa de estimación, además se puede ver que todos sus estadísticos decrecen, a diferencia del otro modelo (en donde algunos decaen), comprobando que el presente modelo cumple con ser el que obtiene mejores resultados, tanto para datos recientes como para anteriores, logrando ser el más equilibrado en todos los periodos, además cumple también bajo el principio de parsimonia.

1.9 Predicción del modelo para agua sucia

Descripción del modelo SARIMA (0,0,0)x(0,1,1)10

Los parámetros estadísticos estimados por el software “*Statgraphics*” fueron indicados en la tabla en la etapa de estimación, en base a ella se procederán a los cálculos del modelo:

Donde el coeficiente estimativo es:

$$\omega_1 = 0,8339$$

Para poder hacer el respectivo modelamiento y calcular el pronóstico, será necesario invertir las transformaciones antes aplicadas (raíz cuadrada y primera diferencia), tal como se hizo con el producto anterior (electrobombas centrifugas).

Utilizando las ecuaciones (30) y (18) se procederá a calcular el modelo:

$$Y_t = \omega_1 e_{t-1} \quad (30)$$

$$\Delta Y_t = Y_t - Y_{t-1} \quad (18)$$

Por lo tanto reemplazando ambas se tiene:

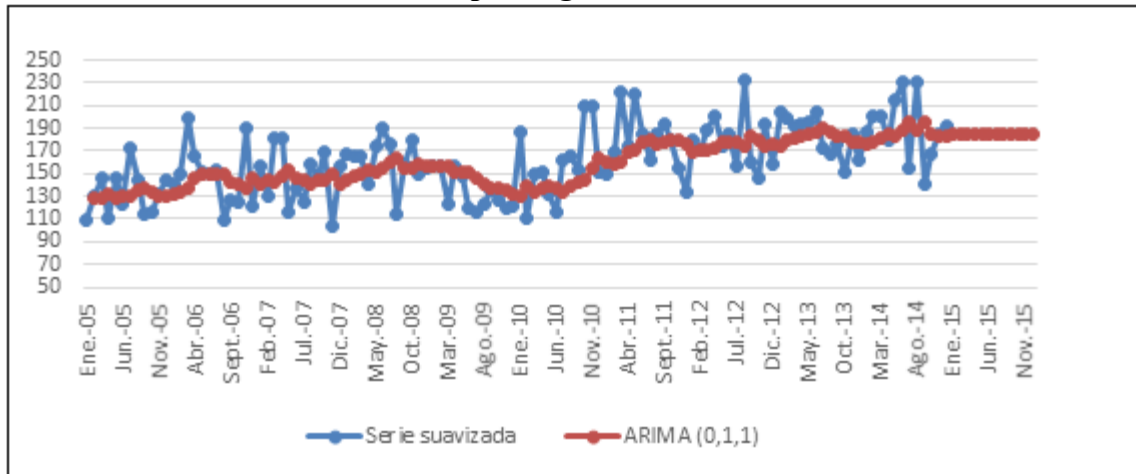
$$\begin{aligned} Y_t - Y_{t-1} &= \omega_1 e_{t-1} \\ Y_t &= 0,83396 e_{t-1} + Y_{t-1} \\ (Y_t)^2 &= [0,83396 e_{t-1} + Y_{t-1}]^2 \end{aligned}$$

Con esta fórmula se podrá calcular los datos pronosticados de cada periodo, el siguiente punto se demuestra toda la serie pronosticada.

El pronóstico converge luego de correr 4 iteraciones, demostrando ser un pronóstico simple, al llegar a los resultados de forma más rápida, lo que lo hace ser una herramienta predictiva simple y menos riesgosa al presentar menos estadísticos de error.

Pronostico ARIMA (0,1,1) seleccionado con 12 datos de predicción y validación

Figura 6.24 Grafica comparativa entre la serie suavizada y el pronóstico ARIMA (0,1,1) para agua sucia

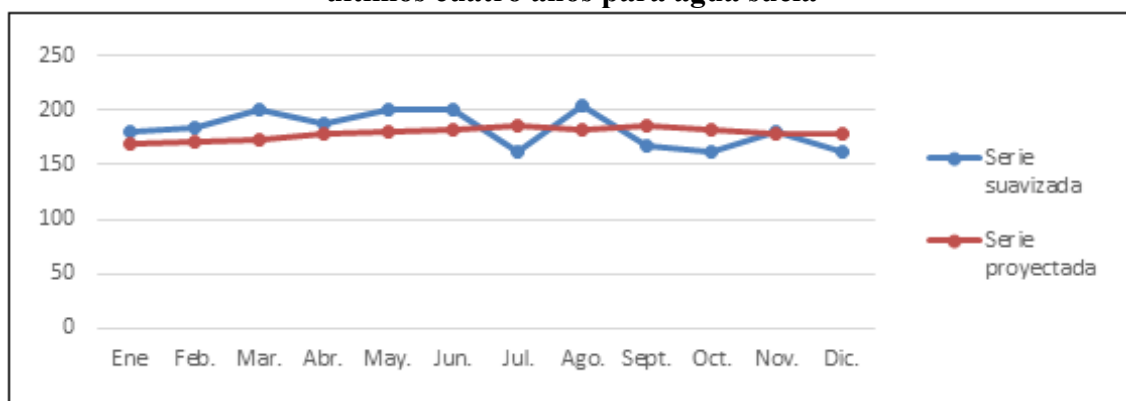


Fuente: Elaboración propia

En la gráfica anterior se puede comprobar que la serie proyectada demuestra estacionarse en los puntos medios de la serie original, suavizando los datos en donde las ventas históricas fueron las más bajas y altas, obteniendo resultados que emulan el promedio general de la serie de tiempo, quedando demostrado en la proyección para el año 2015 en donde se pronostican ventas prácticamente iguales para todos los meses del año.

En la gráfica siguiente se demuestra más claramente el comportamiento promedio mensual de la serie suavizada y la proyectada, considerando los datos más recientes, ósea, los datos históricos de los últimos cuatro años (2011 – 2014).

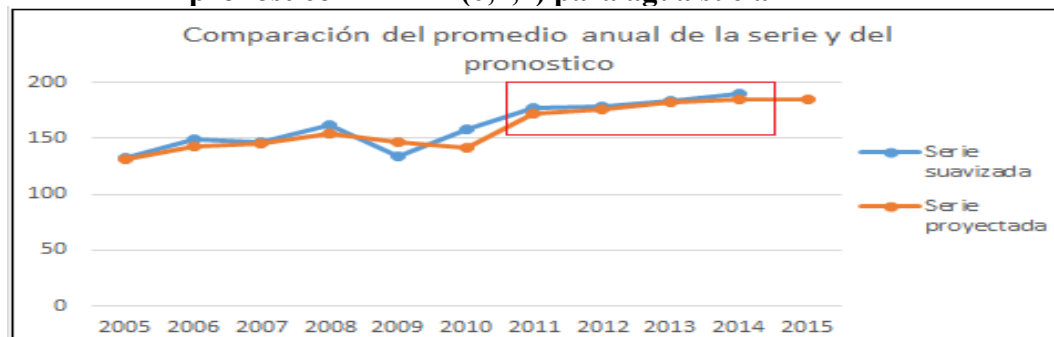
Figura 6.25 Grafico comparativo del promedio mensual de la serie y la proyección en los últimos cuatro años para agua sucia



Fuente: Elaboración propia

Con respecto al promedio mensual arrojado por la herramienta de proyección, demuestra mantenerse en la parte media de la serie, en donde gran parte de los primeros meses experimenta ventas por debajo de la serie suavizada, pero que luego repuntan y se sitúan sobre ella, las diferencias radican en que la herramienta predictiva optó por simular el promedio general de la serie, considerando las ventas de la primera parte (2005 – 2010) aquellas que fueron más bajas con respecto a las ventas experimentadas en los cuatro últimos años (2011 – 2014).

Figura 6.26 Grafico comparativo del promedio anual de la serie suavizada y el pronóstico ARIMA (0,1,1) para agua sucia



Fuente: Elaboración propia

Se puede decir que en el grafica anterior la serie proyectada simula un comportamiento anual promedio similar a la de la serie suavizada, estando más cercana a todos los datos a nivel general, demostrando que los datos más próximos ubicados en el área que se encuentra encerrada en el cuadrado de color rojo, adquieren proyecciones ideales, cumpliendo con los objetivos antes expuestos.

1.10 Comparación situación actual y propuesta seleccionada

Esta comparación se puede describir en la siguiente tabla resumen:

Tabla 6.19 Comparación de estadísticos entre la situación actual y la propuesta para agua sucia

Estadísticos	Situación actual	ARIMA (0,1,1)
RMSE	35,77	25,5
MAE	27,86	20,51
MAPE	14,98%	11,32%

Fuente: Elaboración propia

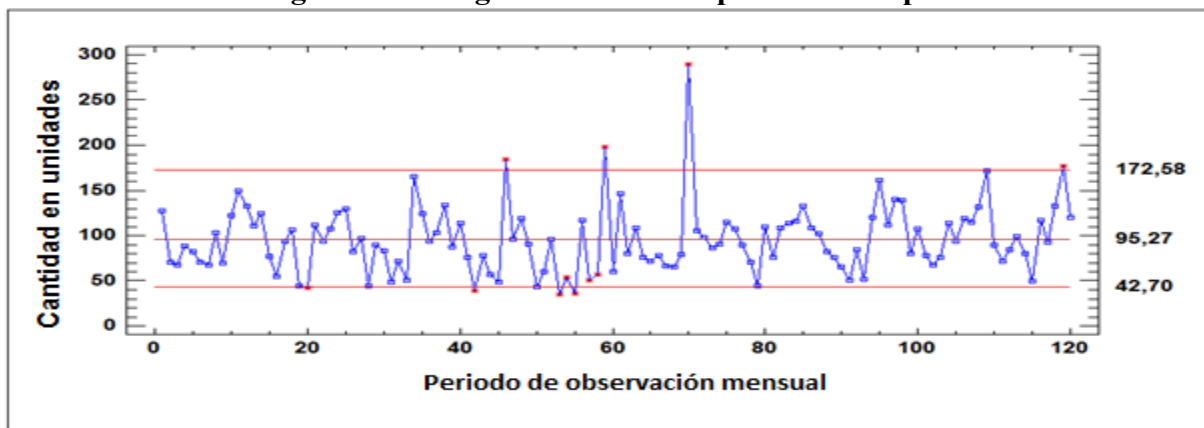
Por lo tanto se puede comprobar que el pronóstico seleccionado en este análisis obtiene un comportamiento que se acerca más a la serie suavizada.

1.11 Análisis de datos para multietapas

1.11.1 Identificación de componentes en la serie para multietapas

Se realiza el mismo procedimiento que el análisis de los límites de control de los otros productos ya analizados, con el fin de verificar que componentes interfieren en la serie de tiempo, a continuación se presenta el diagrama de control para el grupo de multietapas.

Figura 6.27 Diagrama de control para multietapas



Fuente: Elaboración propia

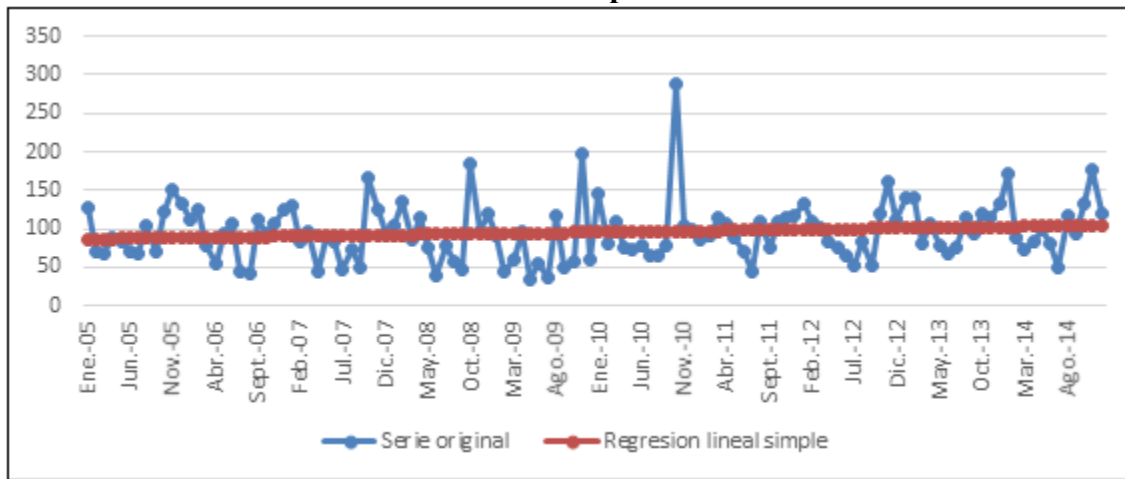
En el diagrama se puede observar que describe un comportamiento en general que se mantiene en torno a su media, sin apreciar una tendencia demarcada, con una varianza que se mantiene con un crecimiento similar a medida que pasa el tiempo, por lo que se concluiría a priori que la serie de tiempo presentaría una clara homocedasticidad en sus datos. Como dato adicional la media del proceso es de 95,27 y su desviación estándar es de 30,92.

Además existen datos que sobresalen de los límites de control, tanto en la parte superior como inferior, en la primera se justifica por campañas de marketing y ofertas varias que impulsaron las altas ventas, mientras que en la segunda fue a causa de problemas de falta de stock y en algunos casos quiebres que provocaron ventas bajísimas. Por lo que estos ocho datos serán suavizados mediante la aplicación de una regresión lineal simple, esta se demuestra en el siguiente punto.

1.11.2 Suavizamiento de datos aplicando regresión lineal para multietapas

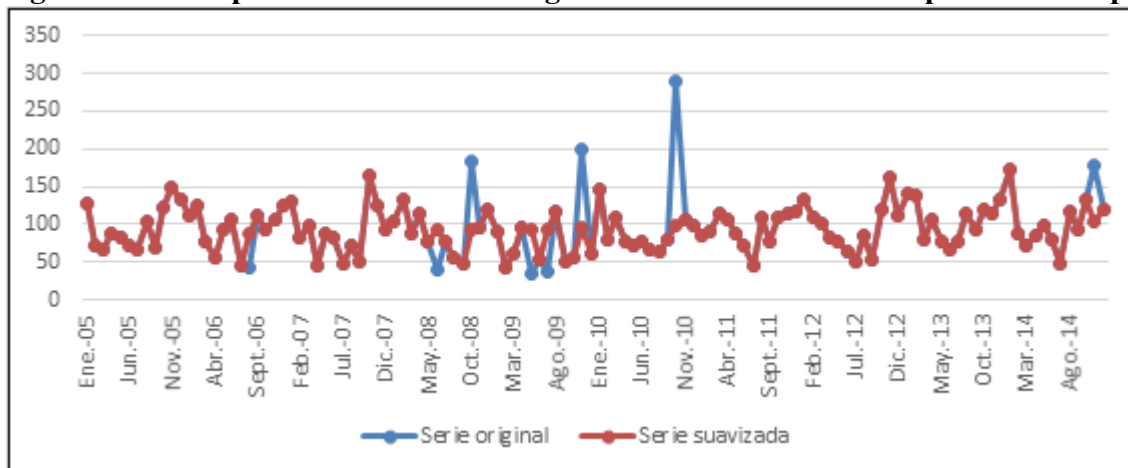
Los datos a suavizar según lo expuesto anteriormente son: Ago – 06, jun y oct – 08, may, jul y nov – 09, oct – 10 y nov – 14, para ello primero se someterá a aplicar la regresión lineal a la serie completa, en la siguiente grafica se demuestran un par de comparaciones:

Figura 6.28 Diagrama comparativo entre la serie original y regresionada para multietapa



Fuente: Elaboración propia

Figura 6.29 Comparación de la serie original con la serie suavizada para multietapa



Fuente: Elaboración propia

En ambas figuras (6.28 y 6.29) se puede decir a priori que los datos suavizados cumplen con el comportamiento original de la serie, para mayor reafirmar todo aquello se compararán a continuación sus respectivas rectas, la primera regresión corresponderá a la serie original y la segunda a la serie ya suavizada.

Comparación de rectas aplicadas con regresiones lineales para multietapas

Tabla 6.20 Funciones de regresión lineal para serie original y suavizada para multietapa

Recta 1	Recta 2
$Y = 86,4 + 0,15x$	$Y = 87,9 + 0,09x$

Fuente: Elaboración propia

Figura 6.30 Comparación de rectas aplicadas a la serie original y suavizada para multietapa

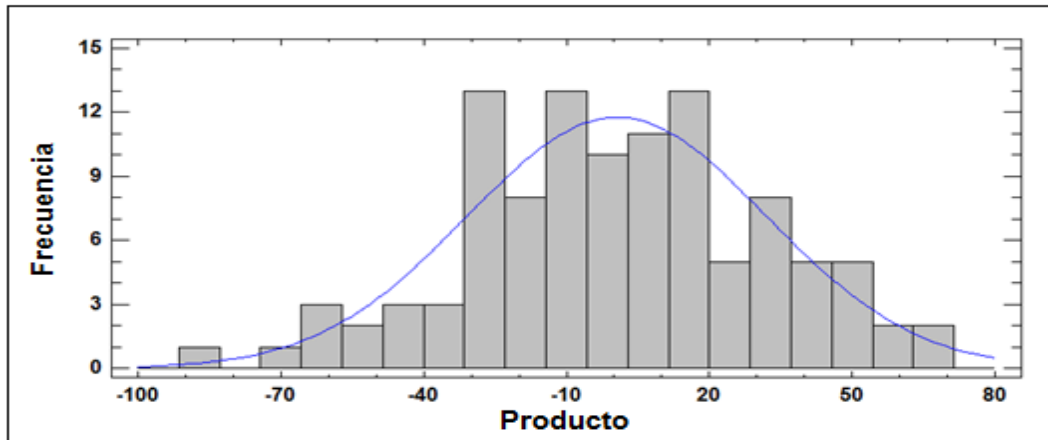


Fuente: Elaboración propia

Se describe en la gráfica 6.30 que al momento de aplicar la segunda regresión lineal esta es muy similar a la primera, con una pendiente muy parecida, demostrando que los cambios efectuados al suavizar los ocho puntos cíclicos no inciden en el comportamiento de la serie regresionada. Por lo tanto todo indica que no hay inconvenientes en poder trabajar con esta serie ya suavizada.

1.11.3 Comprobar la normalidad de la serie para multietapas

Figura 6.31 Histograma de distribución normal de la serie suavizada para multietapa



Fuente: Elaboración propia

En el histograma descrito en la figura 6.31 se puede apreciar que la serie de datos experimenta una distribución que se asimila a una normal, se podría comprobar que existe normalidad a priori, esto se terminará de corroborar bajo el estadístico siguiente:

Tabla 6.21 Prueba Shapiro – Wilk para multietapa

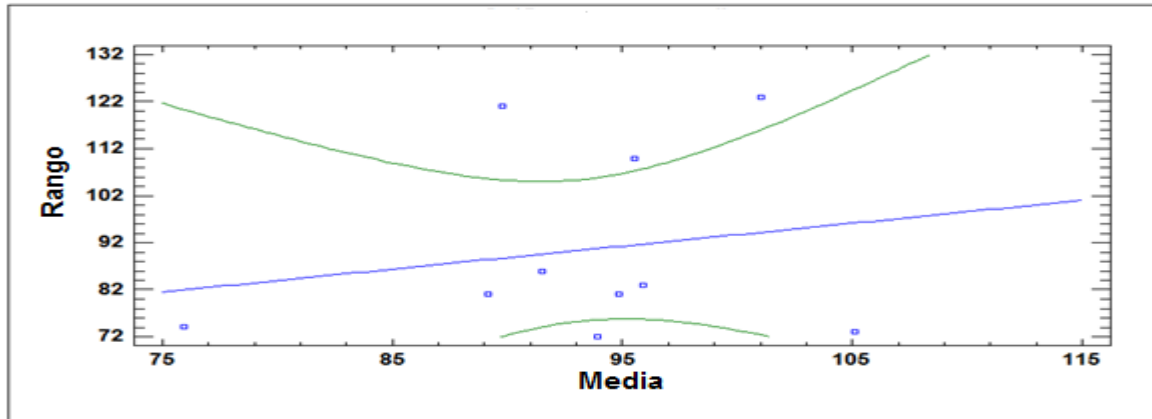
Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,986214	0,829619

Fuente: Elaboración propia

Debido a que el valor-P más pequeño de las pruebas realizadas es mayor ó igual a 0,05, no se puede rechazar la idea de que la serie proviene de una distribución normal con 95% de confianza.

1.11.4 Análisis de la varianza

Para poder analizar y verificar si la varianza se mantiene constante en el tiempo, tal como en los puntos anteriores, se confeccionará una gráfica con dos variables, una de rango y otro de la media, cada valor calculado anualmente, es por aquello que se describirán 10 puntos o datos, todo esto se demuestra en la siguiente gráfica:

Figura 6.32 Grafico rango – media de la serie suavizada para multietapa

Fuente: Elaboración propia

La grafica demuestra claramente una homocedasticidad, al no describir un comportamiento lineal entre el rango y la media, además se puede decir que el estadístico DW calculado obtiene un valor de 2,66 comprobando que la serie de datos presenta una varianza constante y por ende no es necesario aplicar alguna transformación Box – Cox.

La recta queda compuesta por: Lineal:

$$Y = a + b \cdot X = \text{Rango} = 44,8958 + 0,487938 \cdot \text{Media}$$

1.11.5 Identificación de parámetros para multietapa

Primero hay que comprobar si la serie presenta aleatoriedad en el primer desfase, para ello hay que revisar la tabla el valor $t = 0,47$ negativo por lo tanto $0,47 < 1.96$, en consecuencia no se rechaza la hipótesis nula y la serie es aleatoria, comprobando que el primer desfase se encuentra dentro de la región de confianza.

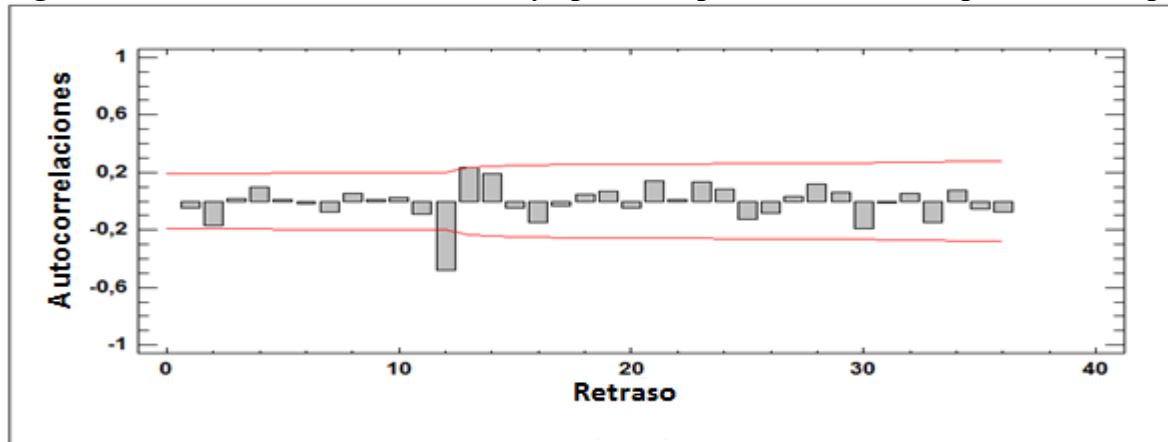
Al revisar si la serie presenta correlación, el valor $LQB = 5,82 < 18,31$ (valor obtenido de la tabla Chi – cuadrada con 10 grados de libertad y una significancia de un 5%, presente en la tabla 6.22), se demuestra que en las auto correlaciones para los primeros 10 retrasos como grupo no son diferentes a 0, demostrando la estabilidad de los datos al aplicar una primera diferencia estacional, comprobando que los datos no están correlacionados.

Se puede indicar que en los correlogramas de la serie suavizada (presentados en el anexo C1), estos demuestran dos partes una regular y otra estacional de orden 12 (al repetir el mismo patrón a cada 12 retardos), en la primera se puede decir que sus retardos caen rápidamente a cero comprobando que experimentan una media constante y por ende presentan estacionariedad, mientras que en la parte estacional sus retardos caen a cero de forma lenta.

Por lo tanto se comprueba que es necesario aplicar una diferencia estacional, al no dejar del todo claro su comportamiento de estacionariedad.

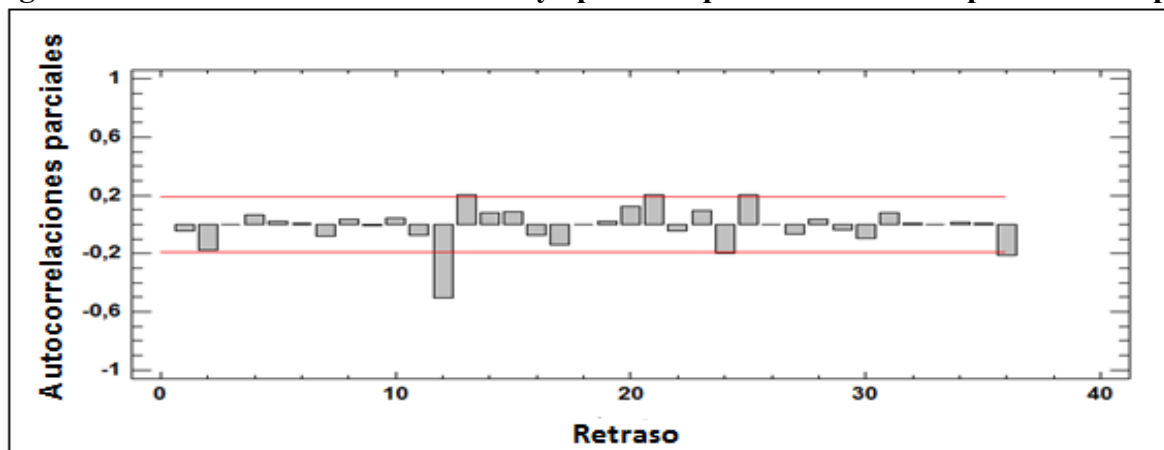
En las siguientes graficas se describen los comportamientos de los correlogramas aplicando primera raíz en la parte estacional.

Figura 6.33 FAC de la serie suavizada y aplicando primera diferencia para multietapa



Fuente: Elaboración propia

Figura 6.34 FACP de la serie suavizada y aplicando primera diferencia para multietapa



Fuente: Elaboración propia

En las figuras 6.33 y 6.34 se puede demostrar que al aplicar una primera diferencia la serie se corrige, obteniendo retrasos más estables permitiendo a los mismos caer a la región de confianza de manera más rápida, concluyendo que los datos están aptos para poder ser leídos en los correlogramas e identificar los posibles pronósticos, esto se expondrá en el siguiente punto.

Tabla 6.22 Estadísticos del FAC aplicando primera diferencia estacional

Desfase	ACF	T	LBQ
1	-0,05	-0,47	0,23
2	-0,17	-1,79	3,58
3	0,02	0,17	3,61
4	0,09	0,95	4,62
5	0,01	0,09	4,62
6	-0,02	-0,20	4,67
7	-0,08	-0,78	5,39
8	0,05	0,52	5,72
9	0,01	0,13	5,74
10	0,03	0,26	5,82

Fuente: Elaboración propia

En la tabla anterior se describe el ACF (valor referencial por cada desfase, si es mayor a 0,2 aquel es un desfase significativo), t (estadístico de prueba para demostrar si un desfase es aleatorio) y LBQ (estadístico Q de Ljung – Box, comprueba si los desfases o retardos como grupo están correlacionados).

1.11.6 Lectura de correlogramas

En los correlogramas anteriores se puede describir un comportamiento en general estable en la parte regular, sin la necesidad de aplicar algún parámetro predictivo, por lo que se concluye que no será inducido a ningún tratamiento, por lo que todos los análisis o interpretaciones siguientes serán referidos a la parte estacional, recordando que todos tendrán una integración del sub – índice uno.

- En este punto la parte estacional se observa un término SAR(1), esto porque en la FAC sus retrasos demuestran su significancia de forma rápida y en la FACP cae luego del primer desfase (12), asumiendo que los desfases 24 y 36 no son preponderantes al considerarse que están dentro de la región. Por lo tanto se traduce un SARIMA (0,0,0)x(1,1,0)₁₂.
- Acá se interpreta lo mismo que el punto anterior, la diferencia radica que si se considerando como relevante el segundo desfase estacional, por lo que adquiere un comportamiento final del SARIMA (0,0,0)x(2,1,0)₁₂.
- Se manifiesta lo mismo que en los dos puntos anteriores, pero considerando como primordial que todos los desfases se encuentren dentro de la región de confianza aunque sea lo forma mínimo, bajo ese supuesto se podría decir que en la FACP caen sus desfases a cero luego del tercer retardo, por lo que en la parte estacional se experimenta un término SAR(3). Identificando un SARIMA (0,0,0)x(3,1,0)₁₂.

- En referencia a este punto se puede describir que la FAC luego de su primer desfase este llega a la significancia y que en la FACP sus desfases caen de golpe a cero, por lo que se traduce un pronóstico SARIMA $(0,0,0) \times (0,1,1)_{12}$.
- Se asume que ambos correlogramas se comportan de manera similar, llegando rápidamente a la región de confianza, por lo que se experimenta una componente auto regresiva y de media móvil en la parte estacional. Obteniendo un SARIMA $(0,0,0) \times (1,1,1)_{12}$.

1.12 Estimación del modelo para multietapas

1.12.1 Descripción de los modelos estimados para multietapas

SARIMA $(0,0,0) \times (1,1,0)_{12}$

Tabla 6.23 Estadísticos estimativos SARIMA $(0,0,0) \times (1,1,0)_{12}$ para multietapa

Parámetro	Estimado	Error Estd.	t	Valor-P
SAR(1)	-0,743339	0,0976027	-7,61597	0

Fuente: Elaboración propia

Se puede demostrar que el modelo identificado es adecuado para esta etapa de estimación, sin necesidad de tener que ser intervenido.

SARIMA $(0,0,0) \times (2,1,0)_{12}$

Tabla 6.24 Estadísticos estimativos SARIMA $(0,0,0) \times (2,1,0)_{12}$ para multietapa

Parámetro	Estimado	Error Estd.	t	Valor-P
SAR(1)	-0,942876	0,118337	-7,96772	0
SAR(2)	-0,773292	0,0986365	-7,83982	0

Fuente: Elaboración propia

Se puede decir que todos los valores – P de los parámetros que compone este modelo son significativos, demostrando que el modelo presente funciona sin necesidad de reducir su complejidad.

SARIMA (0,0,0)x(3,1,0)₁₂**Tabla 6.25 Estadísticos estimativos SARIMA (0,0,0)x(3,1,0)₁₂ para multietapa**

Parámetro	Estimado	Error Estd.	t	Valor-P
SAR(1)	-0,848543	0,109976	-7,71571	0
SAR(2)	-0,848982	0,114335	-7,42537	0
SAR(3)	-0,960124	0,0497597	-19,2952	0

Fuente: Elaboración propia

En la tabla 6.25 se puede comprobar que cada uno de los valores – P de los parámetros auto regresivos de la parte estacional son diferentes a cero, por lo que no se recomienda reducir su composición.

SARIMA (0,0,0)x(0,1,1)₁₂**Tabla 6.26 Estadísticos estimativos SARIMA (0,0,0)x(0,1,1)₁₂ para multietapa**

Parámetro	Estimado	Error Estd.	t	Valor-P
SMA(1)	0,882741	0,0496071	17,7946	0

Fuente: Elaboración propia

El presente modelo no es necesario reducir su único término SMA(1) debido a que demuestra tener un valor – P mayor a 0,05.

SARIMA (0,0,0)x(1,1,1)₁₂**Tabla 6.27 Estadísticos estimativos SARIMA (0,0,0)x(1,1,1)₁₂ para multietapa**

Parámetro	Estimado	Error Estd.	t	Valor-P
SAR(1)	-0,210989	0,145291	-1,45218	0,151841
SMA(1)	0,847141	0,0615903	13,7545	0

Fuente: Elaboración propia

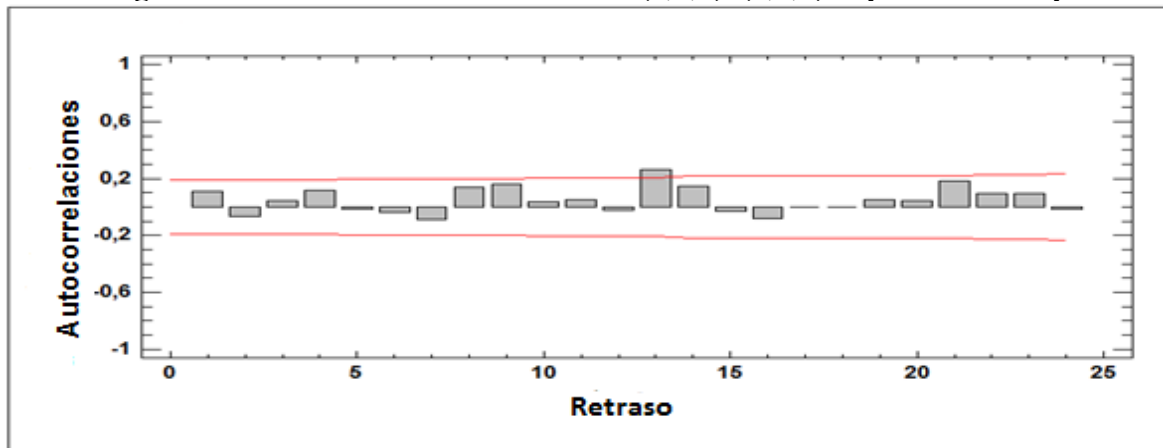
En la tabla anterior se puede apreciar que el valor – P del termino SAR(1) es mayor a 0,05, no demostrando ser estadísticamente distinto de cero, por lo que será necesario reducir su sub – índice a cero, logrando mantener solo el parámetro SMA(1) ya que demuestra ser significativo, obteniendo un modelo reducido SARIMA (0,0,0)x(0,1,1)₁₂ (modelo ya analizado anteriormente).

1.13 Validación de pronósticos para multietapas

1.13.1 Verificación e interpretación de correlogramas residuales para multietapas

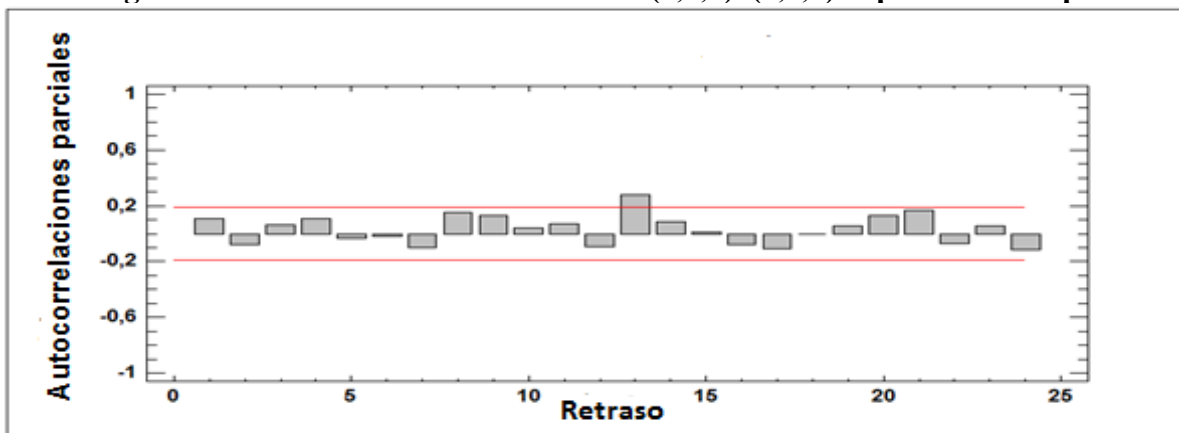
En las siguientes figuras se presentarán los correlogramas del pronóstico a elegir el SARIMA (0,0,0)x(0,1,1)₁₂, los demás se presentarán en los anexos (C2).

Figura 6.35 FAC residual de SARIMA (0,0,0)x(0,1,1)₁₂ para multietapa



Fuente: Elaboración propia

Figura 6.36 FACP residual de SARIMA (0,0,0)x(0,1,1)₁₀ para multietapa



Fuente: Elaboración propia

En los correlogramas anteriores se puede apreciar un comportamiento a nivel general estable, con desfases casi en su totalidad dentro de la región de confianza, salvo el retardo 13 que se encuentra levemente fuera, clasificándose como estadísticamente significativo con un nivel de confianza del 95,0%, implicando que los residuos pueden no ser completamente aleatorios o no presentar RB.

Para comprobar si la serie presenta RB se analiza el estadístico de la prueba Box – Pierce como el valor – P es 0,27, esto implica ser mayor a 0,05 demostrando que la serie definitivamente presenta RB.

Bajo estos análisis el único que acompaña al pronóstico anterior es el SARIMA $(0,0,0) \times (2,1,0)_{12}$, concluyendo que el resto de modelos estimados no son válidos para este tipo de serie de tiempo.

1.13.2 Comparación de los pronósticos y elección del mejor para multietapas

Para poder elegir el mejor modelo se comparará a continuación para estadístico de error tanto en las etapas de estimación como validación, el que obtenga menores valores de tasas de error será la herramienta predictiva que se acomode de menor manera a los datos históricos del producto multietapas.

Etapas de estimación

Tabla 6.28 Comparación de los distintos estadísticos de error por pronóstico en la etapa de estimación para multietapa

Comparación de los distintos pronósticos en la etapa de estimación		
Estadístico	SARIMA $(0,0,0) \times (2,1,0)_{12}$	SARIMA $(0,0,0) \times (0,1,1)_{12}$
RMSE	26,4119	25,7329
MAE	19,5208	21,1052
MAPE	26,1215	27,9708

Fuente: Elaboración propia

Al resultar tener solo dos modelos válidos, se puede decir que sus estadísticos tienen valores muy parecidos, en donde el SARIMA $(0,0,0) \times (0,1,1)_{12}$ tiene un menor margen de error cuadrático medio, a priori se podría elegir este modelo al considerar el RMSE como el mejor estadístico, pero también el SARIMA $(0,0,0) \times (2,1,0)_{12}$ tiene menores valores en los otros dos estadísticos (de desviación media y error porcentual absoluto), por lo que no queda claro cuál pronóstico pre – seleccionar, en la siguiente etapa se comprobará cual será la herramienta predictiva que se acopla mejor a la serie de tiempo analizada.

Etapas de validación

Tabla 6.29 Comparación de los distintos estadísticos de error por pronóstico en la etapa de validación para multietapa

Comparación de los distintos pronósticos en la etapa de validación		
Estadístico	SARIMA (0,0,0)x(2,1,0)12	SARIMA (0,0,0)x(0,1,1)12
RMSE	28,8814	19,4582
MAE	23,2802	15,4546
MAPE	23,8128	15,9153

Fuente: Elaboración propia

Al revisar la tabla anterior, se puede concluir que definitivamente el mejor modelo predictivo es el SARIMA (0,0,0)x(0,1,1)12, al comprobar que sus valores son menores que el otro modelo y además demostrar que en todos sus estadísticos sus valores decrecen en comparación con la etapa anterior, verificando que es el mejor modelo y que también cumple con el principio de parsimonia.

1.14 Predicción del modelo para multietapas

1.14.1 Descripción del modelo para multietapas

Al aplicar una primera diferencia será necesario invertir la transformación, aplicando una diferencia al modelo expresado.

Donde el coeficiente estimativo es:

$$W_1 = -0,883$$

Utilizando las ecuaciones (33) y (32) se procederá a calcular el modelo:

$$Y_t = W_1 e_{t-12} \quad (33)$$

$$\Delta_{12} Y_t = Y_t - Y_{t-12} \quad (32)$$

Por lo tanto reemplazando ambas se tiene:

$$Y_t - Y_{t-12} = W_1 e_{t-12}$$

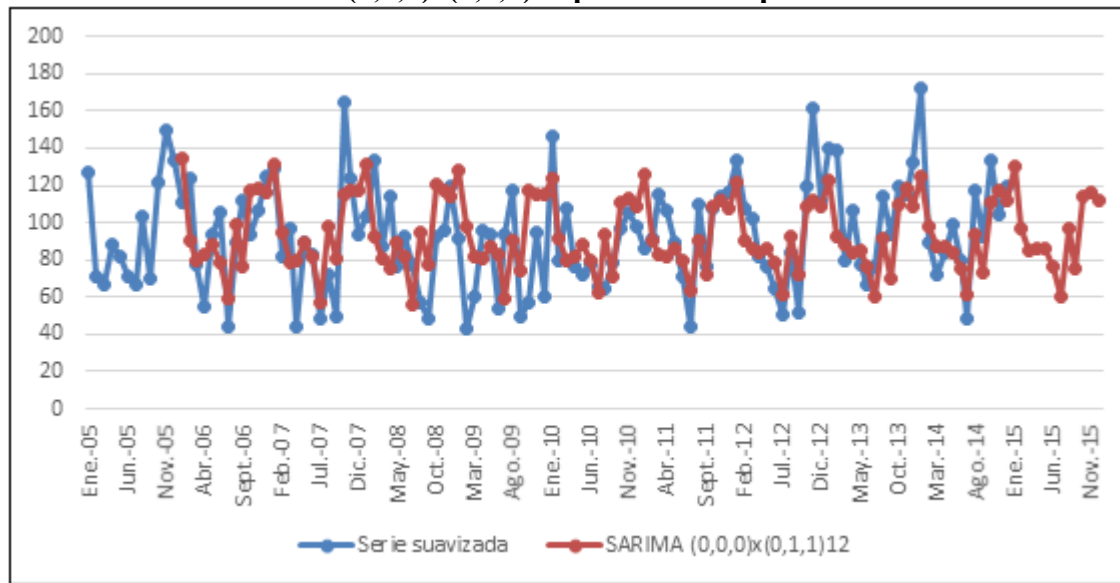
$$Y_t = 0,883 e_{t-12} + Y_{t-12}$$

Con esta fórmula se podrá calcular los datos pronosticados de cada periodo, el siguiente punto se demuestra toda la serie pronosticada.

El pronóstico converge luego de correr 5 iteraciones.

Pronostico SARIMA (0,0,0)x(0,1,1)12 seleccionado con 12 datos de predicción y 48 validación

Figura 6.37 Grafica comparativa entre la serie suavizada y el pronóstico SARIMA (0,0,0)x(0,1,1)12 para multietapa

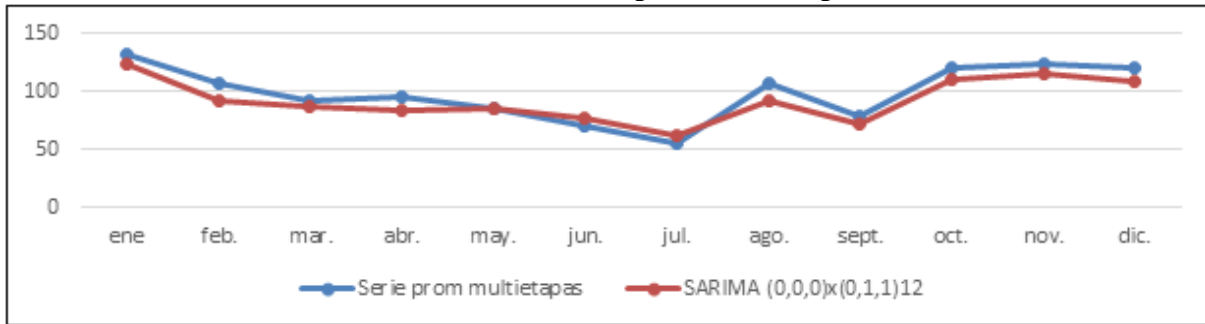


Fuente: Elaboración propia

En la gráfica de dispersión anterior se puede comprobar que la serie obtenida en la predicción replica prácticamente el mismo comportamiento año tras año, emulando a la serie histórica a nivel general, suavizando los datos más altos y bajos estacionándolos en su media.

En la siguiente grafica se describe el promedio mensual de los últimos cuatro años (2011 – 2014) del comportamiento de la serie predictiva y serie de tiempo suavizada, con el objetivo de demostrar que tan cerca el pronóstico replica el promedio mensual de la serie de tiempo.

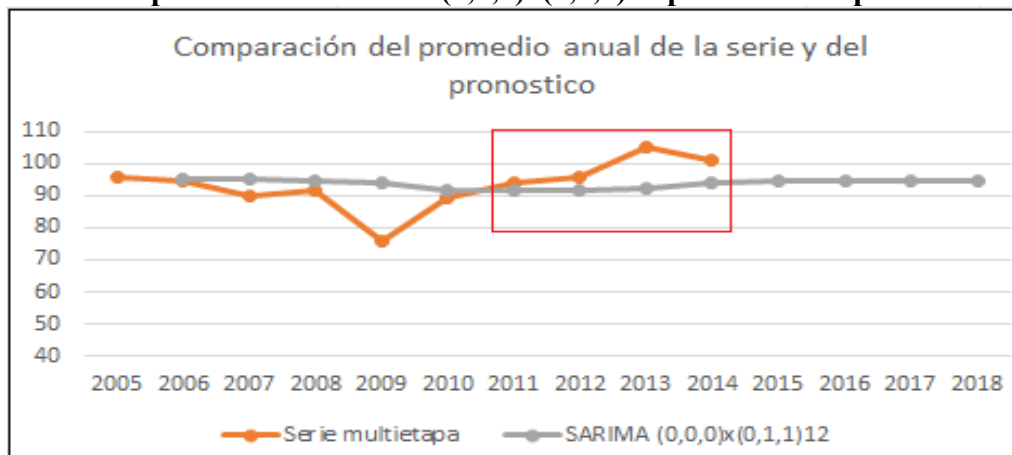
Figura 6.38 Grafico comparativo del promedio mensual de la serie y la proyección en los últimos cuatro años para multietapa



Fuente: Elaboración propia

Se puede demostrar que en la imagen 6.38 el pronóstico elegido predice el promedio mensual de la serie suavizada de manera casi perfecta.

Figura 6.39 Grafico comparativo del promedio anual de la serie suavizada y el pronóstico SARIMA (0,0,0)x(0,1,1)12 para multietapa



Fuente: Elaboración propia

Se puede decir que en la gráfica la herramienta predictiva obtiene promedios anuales que se instalan en la media de la serie suavizada, comprobando que los datos en donde se acerca más son los años más recientes, ubicados en el cuadrado de color rojo, donde tiene un comportamiento más cercano en los años 2011 y 2012, alejándose un poco durante el 2013, año donde este producto experimentó las mejores ventas.

1.15 Comparación de la situación actual y propuesta seleccionada para multietapas

Tabla 6.30 Comparación de estadísticos entre la situación actual y la propuesta para centrifugas

Estadístico	Situación actual	SARIMA (0,0,0)x(0,1,1)₁₂
RMSE	23,32	19,46
MAE	19,85	15,45
MAPE	20,52%	15,92%

Fuente: Elaboración propia

Se puede por lo tanto comprobar que la propuesta elegida anteriormente se acomoda mejor a la serie suavizada, lo que permite asegurar que es una herramienta óptima para ser recomendada en este tipo de productos.

2. Conclusiones del análisis de resultados

Gracias a las técnicas de pronósticos ARIMA calculados a los grupos más relevantes encontrados por el sistema ABC, se pudo llegar a resultados concretos y validar su efectividad, demostrando en definitiva que adquirirían estadísticos de error más bajos que la situación actual de la empresa, en los siguientes comentarios se describe lo anterior:

- **Electrobombas centrifugas:** Se lograron identificar componentes con datos cíclicos que se corrigieron a través de un suavizamiento aplicando regresión lineal, además su varianza y media demostraron ser no constantes, estas se sometieron a transformaciones Box – Cox (primer logaritmo natural y luego primera diferencia estacional) para ser estabilizadas. Luego se lograron identificar tres tipos de modelos predictivos y que al momento de estimarlos y validarlos solo pasaron dos la prueba, llegando a elegir el SARIMA (0,0,0)x(0,1,1)₁₂ por ser el que tuviera mejores resultados en la etapa de validación con un RMSE de 127,801.
- **Electrobombas de agua sucia:** En estos datos se manifestaron componentes con datos cíclicos que fueron corregidos mediante el suavizamiento de aquellos, también se encontró una distribución que no demostró ser normal, por lo que fue sometida a un par de transformaciones Box – Cox (raíz cuadrada y primera diferencia), además su varianza presentó homocedasticidad y su media demostró ser no constante por lo que recomendó aplicar una primera diferencia. Luego al momento de leer los correlogramas se llegaron a identificar tres modelos, de los cuales dos pasaron las pruebas de estimación y validación, eligiendo en definitiva el ARIMA (0,1,1) por demostrar mejores resultados en las etapas de estimación y validación, llegando a un valor RMSE de 25,5, resultado muy bajo y más que aceptable.
- **Electrobombas multietapa:** Al analizar los datos se eliminaron inmediatamente los componentes que influían en la demanda, identificando datos cíclicos y una media no constante, aplicando suavizamiento y primera diferencia estacional respectivamente, obteniendo un pronóstico SARIMA (0,0,0)x(0,1,1)₁₂ como el mejor de todos al demostrar RB en sus datos y llegando a un resultado RMSE de 19,46, el más bajo de todos los grupos analizados, demostrando ser el que se acerque más al comportamiento general de la serie de tiempo analizada por caso.

Capítulo VII:

Evaluación Económica del Modelo Elegido

Al comprobar los resultados efectivos de cada propuesta elegida por producto, ahora habrá que justificar su implementación mediante un análisis económico, con el objetivo de demostrar si son herramientas viables para la empresa.

Para ello habrá que tener claro la inversión total necesaria de cada propuesta, sus costos relacionados una vez implementada (reflejado en las órdenes de compra total durante un año y sus costos de transporte) y sus ingresos. Todo esto se reflejará en un flujo de caja junto con los cálculos de sus tasas de retorno.

Hay que aclarar que todos los costos por conceptos de importación se adquieren en moneda europea, ósea, en EUR, estos fueron calculados según el promedio de conversión a peso Chileno (CLP) de los últimos cuatro años (2011 – 2014), este fue \$669,58 (datos adquiridos en de la página de investing).

1. Inversión

Para evitar hacer capacitaciones y disminuir los costos, se venderá la exclusividad a la empresa por medio de una consultoría, que garantizará a la empresa la entrega de todos los reportes de proyección de los distintos grupos de productos.

Por conceptos de costos por consultoría exclusiva se optará por capacitar al área de productos en vez de contratar a otro profesional, esto debido a que un profesional en el mercado está cobrando alrededor de \$800.000 a \$1.000.000 mensual, mientras que el curso completo de capacitación será de \$500.000.

Además se contratará un ingeniero capacitado que estará encargado de calcular todos estos pronósticos y de retroalimentar la información al gerente de ventas y el de control de abastecimiento, con la finalidad de generar las órdenes de compra respectivas. Todo de forma mensual y anual.

La licencia del *Statgraphics* a adquirir se instalará en dos computadores, con el objetivo de dejar un *backup* en caso de cualquier inconveniente del titular

En la siguiente tabla se demuestra el costo de inversión necesaria para implementar las propuestas elegidas, estas serán las mismas para cualquier grupo de productos:

Tabla 7.1 Inversión de la herramienta predictiva propuesta para todos los grupos

Ítem	Cantidad	Precio con IVA	Total
Contrato consultoría	1	\$4.536.000	\$4.536.000
Licencia SW (permanente)	2	\$2.063.550	\$4.127.100
Computador	1	\$400.000	\$400.000
Total con IVA			\$9.063.100

Fuente: Elaboración propia

En la tabla 8.1 se describen los ítems de la adquisición de la licencia del *Software Statgraphics* y la capacitación.

En la primera se obtuvo el precio promedio de mercado por costos por consultoría, en la segunda se adquirió desde la página oficial de *Statgraphics*, justifica este precio porque se obtiene la licencia permanente sin la necesidad de ir renovando año a año, considerando que la licencia anual vale más de la mitad del valor que la permanente, por lo que se paga la inversión luego de dos años. Se consideró un computador adicional en la inversión para el profesional contratado, para el backup no será necesario uno nuevo, basta con el que posee, ya que demuestra ser un sistema operativo apto para la programación del presente *Software* y estar actualizado con todas las firmas de office y antivirus.

2. Costos

Como se trata de productos que sólo se importan y no se manufacturan, los costos a considerar serán los costos FOB y costos CIF, los primeros consideran solo los costos de productos terminados y los segundos consideran los costos del producto terminado y los de transporte, divididos en costos por el flete y seguros varios.

Además se describirán los costos por stock de seguridad, estos serán como regla general un 15% sobre el costo CIF de cada producto por un periodo de un trimestre, este valor se obtiene como un costo promedio por espacio ocupado de inventario no rotado (dato adquirido del área de importaciones de la empresa).

2.1 Costos FOB y CIF

Con respecto al caso particular los costos pactados con el proveedor son en FOB por lo tanto estos estarán sujetos a cambio monetario, para este caso se utilizarán Euros (todos los grupos analizados tendrán origen Italia). Para transformarlos a CIF a los anteriores se les agregarán los costos incurridos en transporte, estos serán otorgados por moneda local, para conceptos de optimización de estos costos, se recomienda como regla general llenar

contenedores de 20 o 40 pies, mientras más grande el contenedor menores serán los costos proporcionados, es por aquello y gracias a las cantidades sugeridas por las propuestas ARIMA elegidas, se podrán programar estas compras y considerar en aquellas el abastecimiento necesario para tres o cuatro meses, esta magnitud de tiempo dependerá de la optimización en los espacios de contenedores.

Actualmente los costos promedio de la empresa por transporte llegan a un 20% del costo total del producto, este porcentaje llega a ser muy alto impactando directamente al producto, causado por una serie de costos indebidos, el primero por el no llenado de contenedores y el segundo por la vía de transporte, que para mitigar posibles quiebres de stock se realizan compras urgentes de forma aérea.

Es por todo aquello que con la implementación de la propuesta ARIMA sugerida, se reducirán significativamente estos costos, llegando por conceptos de costos por transporte a un 10% del total, logrando optimizar los costos totales CIF.

2.1.1 Costos por stock de seguridad

La planificación de compra de productos se llevará a cabo cada tres meses, ósea, una vez que entre la mercadería del pedido anterior automáticamente se activará el siguiente pedido, considerando cada pedido un abastecimiento para tres meses, así con esta estrategia se espera disminuir los quiebres de stock abasteciendo un año de compras, por lo tanto se creará un stock de seguridad que antes no existía, para ello se incurrirán a costos de este tipo de stock por mantener la mercadería en inventario, esta estrategia será utilizada para todos los grupos antes analizados.

Con este plan de abastecimiento generado por la propuesta, la idea será considerar en definitiva componentes que afecten la demanda histórica (consideraras por la pronostico ARIMA y la demanda actual (considerada por el stock de seguridad).

3. Ingresos

Los ingresos se definirán como las ganancias obtenidas por los quiebres mitigados por la propuesta ARIMA y los ahorros por costos de transporte, esto se comparará con la condición actual de la empresa que no consideró los quiebres, todo en relación al promedio de los últimos cuatro años, por grupo.

4. Flujos de caja

La idea es describir todos los puntos contables importantes para justificar monetariamente la propuesta a implementar y comprobar lo viable que será para cada caso en la empresa.

En el flujo de caja siguiente se describirán los ítems de inversión necesaria para este proyecto, los egresos que vendrían siendo todos los tipos de costos y los ingresos obtenidos por las ventas de quiebres y normales.

5. Propuestas

5.1 Propuesta 1: Electrobombas centrífugas

Sugerido de compra

Tabla 7.2 Unidades de compra por trimestre para centrífugas

Sugerido de compra por trimestre		
Trimestre	Situación actual	Propuesta proyecto
T1	1469	1558
T2	1041	1105
T3	1458	1635
T4	1931	1823
Total	5899	6122

Fuente: Elaboración propia

En la tabla anterior se describe el sugerido de compra calculados por cada propuesta, estas se dividirán en cuatro compras al año, cada uno de magnitud trimestral.

Análisis de costos para centrífugas

En la siguiente tabla se describen los distintos costos definidos anteriormente, comparados por propuesta y de magnitud trimestral.

Tabla 7.3 Costos por propuesta promedio 2011 – 2014 para centrifugas

Costos por propuesta trimestral								
	Trimestre 1		Trimestre 2		Trimestre 3		Trimestre 4	
Item	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto
Costos FOB	\$ 254.647.844	\$ 270.112.094	\$ 180.432.431	\$ 191.526.812	\$ 252.755.625	\$ 283.498.645	\$ 334.821.578	\$ 316.025.579
Costos por transporte	\$ 50.930.156	\$ 29.468.247	\$ 36.086.903	\$ 20.894.879	\$ 50.551.708	\$ 30.928.671	\$ 66.965.088	\$ 34.477.241
Costo de stock de seguridad	\$ -	\$ -	\$ -	\$ 895.290	\$ -	\$ -	\$ -	\$ -
Total	\$ 305.578.000	\$ 299.580.341	\$ 216.519.333	\$ 213.316.980	\$ 303.307.333	\$ 314.427.316	\$ 401.786.667	\$ 350.502.820
Ahorro en costos	\$ 5.997.659		\$ 3.202.353		\$ -11.119.983		\$ 51.283.846	

Fuente: Elaboración propia

En el flujo de caja anterior se puede decir que frente a la situación actual de la empresa, en la propuesta ARIMA se invirtieron \$ 38,5 millones más en costos FOB en total, además se optimizaron los costos por conceptos de transporte en un 43%. Logrando obtener beneficios en ahorros por costos totales de \$ 49,36 millones.

Ingresos por quiebres para centrifugas

En la siguiente tabla se describen los beneficios adquiridos por quiebres de stock por aplicar la propuesta ARIMA elegida.

Tabla 7.4 Ingresos por quiebres de stock para centrifugas

Contribución por ventas de quiebres de stock								
	Trimestre 1		Trimestre 2		Trimestre 3		Trimestre 4	
Item	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto
Ventas de quiebres de stock	\$ -	\$13.713.743	\$ -	\$ 4.776.697	\$ -	\$ 27.273.399	\$16.641.396	\$ -

Fuente: Elaboración propia

Con esta mayor compra descrita en la tabla 7.2 simulando los datos promedio entre el 2011 – 2014 se hubieran logrado disminuir lo quiebres de stock en unos \$29,12 millones, cifra más que óptima.

Flujo de caja para centrifugas

En la siguiente tabla se demuestra el flujo de caja para los cuatro trimestres pertenecientes al promedio de los últimos cuatro años (2011 – 2014), especificando los ingresos como los ahorros de esta propuesta frente a la situación actual de la empresa, extraído del detalle de las tablas 7.3 y 7.4.

Tabla 7.5 Flujo de caja promedio entre el 2011 – 2014 para centrifugas

Flujo de caja 2011 - 2014 para centrifugas						
		Inicio	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
Inversión	Total inversión	\$ -3.021.033	\$ -	\$ -	\$ -	\$ -
Egresos	Capital humano	\$ -	\$ 900.000	\$ 900.000	\$ 900.000	\$ 900.000
	Total egresos	\$ -	\$ 900.000	\$ 900.000	\$ 900.000	\$ 900.000
Ingresos	Ahorro en costos	\$ -	\$ 5.997.659	\$ 3.202.353	\$ -11.119.983	\$ 51.283.846
	Ahorro en quiebres de stock	\$ -	\$ 13.713.743	\$ 4.776.697	\$ 27.273.399	\$ -16.641.396
	Total ingresos	\$ -	\$ 19.711.402	\$ 7.979.050	\$ 16.153.416	\$ 34.642.450
	Total	\$ -3.021.033	\$ 18.811.402	\$ 7.079.050	\$ 15.253.416	\$ 33.742.450
	VAN	\$64.924.907				
	TIR	572%				

Fuente: Elaboración propia

En la tabla anterior se describe la inversión requerida para implementar esta propuesta (especificada en la tabla 7.1), los egresos aplicados por el pago mensual al profesional que realizará los pronósticos y los ingresos provenientes por los ahorros por reducción de costos y disminución de quiebres de stock expresados en las tablas 7.3 y 7.4 respectivamente.

Se puede concluir que considerando una tasa de descuento interbancario del 3,5% mensual (dato extraído de la página del banco central y calculado como el promedio de los últimos cuatro años 2011 – 2014) el VAN calculado total significa que la inversión de la propuesta ARIMA frente a la situación actual de la empresa genera riquezas de \$6,74 millones anuales, describiendo ingresos positivos a partir del primer trimestre.

La tasa interna de retorno calculada muestra que la propuesta aplicando la metodología Box – Jenkins genera una rentabilidad del 572%, notablemente superior a una tasa de rentabilidad bancaria convencional, cubriendo la inversión inicial gracias al retorno proveniente de los flujos futuros positivos (diferencia de los ingresos frente a los egresos).

5.2 Propuesta 2: Electrobombas agua sucia

Sugerido de compra para agua sucia

Tabla 7.6 Unidades de compra por trimestre para centrifugas

Sugerido de compra por trimestre		
Trimestre	Situación actual	Propuesta proyecto
T1	482	515
T2	504	540
T3	504	552
T4	502	539
Total	1992	2146

Fuente: Elaboración propia

Análisis de costos para agua sucia

Tabla 7.7 Costos por propuesta promedio 2011 – 2014 para agua sucia

Costos por propuesta trimestral								
	Trimestre 1		Trimestre 2		Trimestre 3		Trimestre 4	
Item	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto
Costos FOB	\$ 62.005.447	\$ 66.148.774	\$64.833.371	\$ 69.393.527	\$64.753.033	\$ 70.980.828	\$64.544.152	\$ 69.260.358
Costos transporte	\$ 12.400.897	\$ 7.216.351	\$12.966.473	\$ 7.570.331	\$12.950.405	\$ 7.743.494	\$12.908.630	\$ 7.555.803
Costo de stock de seguridad	\$ -	\$ -	\$ -	\$ -	\$ -	\$ 386.294	\$ -	\$ 735.688
Total	\$ 74.406.344	\$ 73.365.125	\$77.799.844	\$ 76.963.858	\$77.703.438	\$ 79.110.616	\$77.452.781	\$ 77.551.848
Ahorro en costos	\$ 1.041.219		\$ 835.986		\$ -1.407.178		\$ -99.067	

Fuente: Elaboración propia

En el flujo de caja anterior se puede decir que frente a la situación actual de la empresa, en la propuesta ARIMA se invirtieron 19,65 millones más en costos FOB, además se hubieran optimizado los costos por conceptos de transporte en \$21 millones traducidos un 41% menos. Logrando obtener beneficios en ahorros por costos totales de \$ 370 mil.

Ingresos por quiebres para agua sucia

Tabla 7.8 Ingresos por quiebres de stock para agua sucia

Contribución por ventas de quiebres de stock								
	Trimestre 1		Trimestre 2		Trimestre 3		Trimestre 4	
Item	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto
Ventas de quiebres de stock	\$ -	\$ 5.595.621	\$ -	\$ 6.104.313	\$ -	\$ 1.695.643	\$ -	\$ -

Fuente: Elaboración propia

Con esta mayor compra se hubieran logrado disminuir en quiebres de stock unos \$13,4 millones.

Flujo de caja para agua sucia

Tabla 7.9 Flujo de caja promedio entre el 2011 – 2014 para agua sucia

Flujo de caja 2011 - 2014 para agua sucia						
Inversión	Total inversión	Inicio	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
		\$-3.021.033	\$ -	\$ -	\$ -	\$ -
Egresos	Capital humano		\$ 900.000	\$ 900.000	\$ 900.000	\$ 900.000
	Total egresos		\$ 900.000	\$ 900.000	\$ 900.000	\$ 900.000
Ingresos	Ahorro en costos		\$ 1.041.219	\$ 835.986	\$-1.407.178	\$ -99.067
	Ahorro en quiebres de stock		\$ 5.595.621	\$ 6.104.313	\$ 1.695.643	\$ -
	Total ingresos		\$ 6.636.839	\$ 6.940.299	\$ 288.464	\$ -99.067
	Total	\$-3.021.033	\$ 5.736.839	\$ 6.040.299	\$ -611.536	\$ -999.067
	VAN	\$ 6.738.291				
	TIR	162%				

Fuente: Elaboración propia

El VAN calculado total significa que la inversión de la propuesta ARIMA frente a la situación actual de la empresa genera riquezas de \$6,74 millones anuales, describiendo ingresos positivos a partir del primer trimestre.

La tasa interna de retorno calculada muestra que la propuesta aplicando la metodología Box – Jenkins genera una rentabilidad del 162%, notablemente superior a una tasa de rentabilidad bancaria convencional.

5.3 Propuesta 3: Electrobombas multietapas

Sugerido de compra para multietapas

Tabla 7.10 Unidades de compra por trimestre para multietapas

Sugerido de compra por trimestre		
Trimestre	Situación actual	Propuesta proyecto
T1	303	303
T2	245	247
T3	224	225
T4	317	334
Total	1089	1109

Fuente: Elaboración propia

Análisis de costos para multietapas

Tabla 7.11 Costos por propuesta promedio 2011 – 2014 para multietapas

Costos por propuesta trimestral								
	Trimestre 1		Trimestre 2		Trimestre 3		Trimestre 4	
Item	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto
Costos FOB	\$44.257.089	\$44.257.089	\$35.785.435	\$36.077.561	\$32.718.112	\$32.864.175	\$46.301.971	\$48.785.042
Costos por transporte y seguros	\$ 8.851.539	\$ 4.828.002	\$ 7.157.185	\$ 3.935.698	\$ 6.543.712	\$ 3.585.150	\$ 9.260.521	\$ 5.321.956
Costo de stock de seguridad	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -
Total	\$53.108.628	\$49.085.091	\$42.942.620	\$40.013.259	\$39.261.824	\$36.449.325	\$55.562.492	\$54.106.998
Ahorro en costos	\$ 4.023.537		\$ 2.929.361		\$ 2.812.499		\$ 1.455.494	

Fuente: Elaboración propia

En el flujo de caja anterior se puede decir que frente a la situación actual de la empresa, en la propuesta ARIMA se invirtieron 2,9 millones más en costos FOB, además se hubieran optimizado los costos por conceptos de transporte en \$14,14 millones traducidos un 44,45% menos. Logrando obtener beneficios en ahorros por costos totales de \$11,22

Ingresos por quiebres para multietapas

Tabla 7.12 Ingresos por quiebres de stock para multietapas

Contribución por ventas de quiebres de stock								
	Trimestre 1		Trimestre 2		Trimestre 3		Trimestre 4	
Item	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto	Situación actual	Propuesta proyecto
Ventas de quiebres de stock	\$ -	\$ -	\$ -	\$ 291.906	\$ -	\$ 145.953	\$ -	\$ 2.481.201

Fuente: Elaboración propia

Con esta mayor compra se hubieran logrado disminuir en quiebres de stock unos \$2,9 millones.

Flujo de caja para multietapas

Tabla 7.13 Flujo de caja promedio entre el 2011 – 2014 para multietapa

Flujo de caja 2011 - 2014 para multietapas						
		Inicio	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
Inversión	Total inversión	\$-3.021.033	\$ -	\$ -	\$ -	\$ -
Egresos	Capital humano	\$ -	\$ 900.000	\$ 900.000	\$ 900.000	\$ 900.000
	Total egresos	\$ -	\$ 900.000	\$ 900.000	\$ 900.000	\$ 900.000
Ingresos	Ahorro en costos	\$ -	\$ 4.023.537	\$ 2.929.361	\$ 2.812.499	\$ 1.455.494
	Ahorro en quiebres de stock	\$ -	\$ -	\$ 291.906	\$ 145.953	\$ 2.481.201
	Total ingresos	\$ -	\$ 4.023.537	\$ 3.221.267	\$ 2.958.452	\$ 3.936.695
	Total	\$-3.021.033	\$ 3.123.537	\$ 2.321.267	\$ 2.058.452	\$ 3.036.695
	VAN	\$ 6.666.714				
	TIR	82%				

Fuente: Elaboración propia

El VAN calculado total significa que la inversión de la propuesta ARIMA frente a la situación actual de la empresa genera riquezas de \$6,7 millones anuales, describiendo ingresos positivos a partir del primer trimestre.

La tasa interna de retorno calculada muestra que la propuesta aplicando la metodología Box – Jenkins genera una rentabilidad del 82%, notablemente superior a una tasa de rentabilidad bancaria convencional.

6. Resumen

Para que se demuestre de manera más genérica de la herramienta seleccionada ARIMA, se sumaron todos los egresos, ingresos e inversiones de todos los grupos de electrobombas seleccionados, con el fin de demostrar su VAN y TIR, en la siguiente tabla se demuestra además el flujo de caja resumen:

Tabla 7.14 Flujo de caja promedio entre el 2011 – 2014 resumen

Flujo de caja 2011 - 2014 total						
		Inicio	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
Inversión	Total inversión	\$ -9.063.100	\$ -	\$ -	\$ -	\$ -
Egresos	Capital humano	\$ -	\$ 2.700.000	\$ 2.700.000	\$ 2.700.000	\$ 2.700.000
	Total egresos	\$ -	\$ 2.700.000	\$ 2.700.000	\$ 2.700.000	\$ 2.700.000
Ingresos	Ahorro en costos	\$ -	\$ 11.062.414	\$ 6.967.700	\$ -9.714.662	\$ 52.640.273
	Ahorro en quiebres de stock	\$ -	\$ 19.309.364	\$ 11.172.916	\$ 29.114.995	\$ -14.160.195
	Total ingresos	\$ -	\$ 30.371.778	\$ 18.140.616	\$ 19.400.333	\$ 38.480.078
Total		\$ -9.063.100	\$ 27.671.778	\$ 15.440.616	\$ 16.700.333	\$ 35.780.078
VAN		\$ 78.329.912				
TIR		272%				

Fuente: Elaboración propia

El VAN calculado total significa que la inversión de la propuesta ARIMA frente a la situación actual de la empresa genera riquezas de \$78,33 millones anuales, describiendo ingresos positivos a partir del primer trimestre, manteniendo la misma tendencia en los demás periodos, logrando en definitiva cubrir con la inversión inicial y absorber los gastos mensuales incurridos por capital humano.

La tasa interna de retorno calculada muestra que la propuesta aplicando la metodología Box – Jenkins genera una rentabilidad del 272%, notablemente superior a una tasa de rentabilidad bancaria convencional.

Por lo tanto aquella propuesta seleccionada ARIMA para proyectar y determinar la estructura de compra es capaz de generar ahorros importantes en comparación a la condición actual de la empresa Koslan.

Capítulo VIII:

Conclusiones y Recomendaciones

Conclusiones

Una vez desarrollado y validado todos los objetivos antes propuestos se pueden llegar a descifrar ciertas conclusiones.

Se descubrió la necesidad en la área de negocios de distribución en desarrollar metodologías traducidas en modelos de pronósticos que no existían, produciendo demoras en respuesta de la demanda que en varios casos se traducían en quiebres de stock, lo que conllevaba a la pérdidas de varios clientes y al aumento de costos de logística que se incurrían en transporte, por traer productos de urgencia por vía marítima sin llenar contenedores o por vía aérea, sin poder optimizar los costos finales.

Al momento de analizar la demanda histórica de cada grupo, en el área de distribución se logró apreciar que el comportamiento del volumen de cada producto era muy disperso, identificando una gran cantidad de parámetros o componentes que afectaban la demanda y que en la empresa antes en ningún momento se consideraban, obteniendo nulo trabajo en el análisis de datos.

Como se pudo apreciar en la parte final del capítulo 6, todos los pronósticos elegidos obtuvieron mejores resultados con parámetros de media móvil y además siempre fueron seleccionados los que cumplieron con el principio de parsimonia (los más simples son los más asertivos). Además demostraron en todos los casos describir comportamientos más próximos a la serie suavizada y con menores valores en sus errores estadísticos, frente a la situación actual de la empresa, pudiendo rectificar que la propuesta sugerida es la más óptima para mitigar y solucionar el problema raíz.

Además a través de la evaluación económica se pudo demostrar que gracias a las propuestas ARIMA seleccionadas se optimizaban los costos de transporte, mediante el aumento de las compras calzadas (contenedores completos) y la disminución de las compras aéreas que se activaban como emergencia para suplir posibles quiebres, además al obtener pronósticos que se acercaban más a la realidad, al obtener menos tasa de error, se comprobó una disminución significativa en los quiebres de stock experimentados históricamente por la empresa.

Por otra parte *Statgraphics* demostró ser una herramienta bastante amigable y completa al momento de generar reportes, optimizando los procesos, lo que provocó disminuir los tiempos de trabajo y facilitó la exportación de información clara y metódica.

Recomendaciones

Al comprobar que los pronósticos del tipo ARIMA seleccionados por grupo fueron los adecuados para el presente caso, se recomienda generar un plan logístico adecuado, ya que no sirve de mucho tener la información adecuada sino se manipula o aplica de buena forma.

Considerando lo anterior es recomendable seguir las siguientes recomendaciones:

- **Tener una coordinación oportuna y de forma constante con el proveedor:** Para ello considerando que los *Lead Time* por producto son de tres a cuatro meses, habrá que enviar un informe de las cantidades necesarias a comprar (otorgadas por la propuesta), que sirvan para abastecer los distintos de grupos más rentables (ya categorizados) de los próximos cuatro meses (esto estará sujeto a la cuadratura del llenado de contenedores, para disminuir los costos CIF). La primera solicitud de requerimiento de compra se recomendará que se realice cuatro meses antes, esto porque si la mercadería normalmente tiene un Lead time de tres meses, habrá que considerar además el caso supuesto que el proveedor no posee el requerimiento con entrega inmediata y tendrá que producirlo (para aquello tiene un plazo máximo de tres semanas), en caso favorable que tenga entrega inmediata será necesario gestionar la compra tres meses antes, considerando el previo acuerdo con el proveedor. Las próximas solicitudes de compra normalmente se recomendarán que se efectúen luego de tres meses concretada la primera, ósea, una vez que hayan ingresado a la fábrica central de Koslan, igual es necesario estar constantemente monitoreando y revisando la disponibilidad de stock del proveedor, con el objetivo final de optimizar los lead time del producto y que estos ingresen además con el menor costo posible.
- **Disponer de información adecuada en los sistemas:** La idea es registrar de manera adecuada todo la información de las cantidades en stock tanto disponibles como en tránsito por producto, esto ya que actualmente hay muchas dudas al respecto, en donde la información ingresada es más aproximada que exacta, con la ayuda de estos pronósticos ARIMA se permitirá obtener resultados más cercanos a la realidad y mientras se disponga de información exacta en el sistema, se logrará hacer un seguimiento a la mercadería de manera más confiable. Se recomienda además ingresar al sistema la información de las compras que a realizarán durante el año, con el objetivo de integrar al departamento de importaciones para que pueda generar las comprar de manera automática y se permita así disminuir los tiempos internos producidos en la empresa por integración entre áreas.
- **Disponibilidad de información accesible para todas las áreas involucradas:** Si bien es cierto esta propuesta de pronósticos es para mitigar el problema que existe en la génesis de compra en el área ventas y abastecimiento, también es importante y necesario liberarla a todas las áreas que influyen en la cadena de valor del producto, con el

objetivo de aumentar la coordinación entre las mismas y optimizar los procesos internos.

- **Beneficios obtenidos:** Al momento de obtener proyecciones robustas se pueden tomar decisiones más asertivas considerando un entorno cambiante, mediante la planificación de los procesos de forma más eficiente, lo que conlleva una mayor optimización de los tiempos de llegada del producto y menores costos asociados de logística.

En una empresa en donde su principal forma de abastecimiento es mediante la importación de sus insumos, estas herramientas ARIMA permiten obtener información valiosa, ya que gracias a ella se puede planificar de mejor forma una estrategia de compra con los proveedores, logrando obtener una mejor y mayor retroalimentación con ellos. Además beneficia a todas las áreas involucradas en el proceso, a través de la disponibilidad de esta información en el sistema, logrando que las áreas trabajen en paralelo, provocando una mayor rapidez de respuesta.

- **Costos asociados:** Al momento de saber cuántas son las unidades necesarias para satisfacer la demanda, se puede optimizar los costos por importación, mediante la planificación de compras por contenedores, esto porque está comprobado que si se logran llenar contenedores, por tema de volumen y peso los costos CIF (costos, seguros y fletes) son menores, otorgando mayores beneficios, traduciendo el ahorro en costos de un 10% sobre el costo normal efectuado en transporte.

Al planificar las compras anuales siempre en teoría quedará algunas unidades inmovilizadas, este plan de prevención de quiebres, otorgará costos por stock de seguridad que impactarán a la empresa entre un 10% a un 15% sobre el costo total original del producto. Si bien es cierto esto implica un aumento en los costos de la empresa, estos se justificarán a medida que se implemente esta propuesta, ya que se esperará aumento en sus ventas, por conceptos de mitigación de quiebres de stock, mayores clientes conformes y mayores ahorros en otros tipos de costos.

- **Trabajos futuros:** Con estos resultados la idea es ir monitoreando año a año los pronósticos, agregando siempre información más actualizada, que en definitiva demuestra ser para la empresa en cuestión la más relevante.

Bibliografía

(Adam&Ebert1991) Adam, E., & Ebert, R. (1991). Administración de la producción y operaciones. Conceptos modelos y funcionamiento. México: Prentice – Hall.

(Arce&Mahia2009) Arce, R., & Mahía, R. (2009). Conceptos básico sobre la heterocedasticidad en el modelo básico de regresión lineal tratamiento con E – views. España: Universidad autónoma de Madrid.

(Fernandez0?) Fernández, A: Análisis de series de tiempo, texto guía, pp. 1 – 68. (Disponible vía web en http://www.ccee.edu.uy/ensenian/catectr/material/arima_1.pdf. Visitado en octubre 2014).

(Hanke&Wichern2010) Hanke, J., & Wichern, D. (2010). Pronósticos en los negocios. México: Pearson.

(Hillier&Lieberman2012) Hillier, F., & Lieberman, G. (2012). Introducción a la investigación de operaciones. Mcgraw- Hill.

(Jimenez2011) Jiménez Lillo Daniel Loreto (2011). Análisis y pronósticos de demanda para telefonía móvil. Tesis para optar al grado de ingeniería civil industrial, Universidad de Chile, Chile. (Disponible vía web en http://www.tesis.uchile.cl/tesis/uchile/2011/cf-jimenez_dl/pdfAmont/cf-jimenez_dl.pdf. Visitado en noviembre 2014)

(Macias2007) Macías Calvario Gilberto German (2007). Metodología para calcular el pronóstico de la demanda y una medición de su precisión, en una empresa de autopartes, caso estudio. Tesis para optar al grado de maestro en ciencias con la especialidad de administración, Instituto politécnico nacional. México. (Disponible vía web en <http://148.204.210.201/tesis/332.pdf>. Visitada en noviembre 2014).

(Maddala1996) Maddala, G. S. (1996). Introducción a la Econometría, PRENTICE-HALL HISPANOAMERICANA, S.A. México, texto guía.

(Makridakis&Wheelwright1998) Makridakis & Wheelwright. (1998). Métodos de pronósticos. México: Noriega Limusa.

(Moreno2002) Moreno, V. (2002). Regresión Lineal, texto guía, 1 – 47 (Disponible vía web en <http://lbe.uab.es/vm/upf/bio2/vm/regresion-lineal.pdf>. Visitado en noviembre 2014).

(Nahmias2007) Nahmias. (2007). Análisis de la producción y las operaciones. México: Mcgraw- Hill.

(Rojo2007) José Manuel Rojo Abuin. (2007). Regresión lineal múltiple, texto guía, España, pp. 1 – 32. (Disponible vía web en

http://portal.uned.es/pls/portal/docs/PAGE/UNED_MAIN/LAUNIVERSIDAD/VICERRECTORADOS/INVESTIGACION/O.T.R.I/OFERTAS%20TECNOLOGICAS/DMAC/DOCUMENTOS%20Y%20TUTORIALES/REGRESION LINEAL MULTIPLE 3.PDF. Visitado en noviembre 2014).

(Sierra0?) Sierra Guerrero Adela Marisol. Pronostico de la demanda, texto guía, pp. 1 – 19. (Disponible vía web en <http://pronosticos-de-la-demanda.wikispaces.com/file/view/pronosticodelademanda.pdf>. Visitado en noviembre 2014).

(Statpoint2006) Statpoint, inc. (2006). Manual introductorio Statgraphics, texto guía, pp. 1 – 42

(UAEM2006) Universidad autónoma del estado de México. (2006). Estadística. Series de tiempo, texto guía, México. (Disponible vía web en http://www.seduca2.uaemex.mx/ckfinder/uploads/files/u3tema_3_series_de_t.pdf. Visitado en octubre 2014).

(Villavicencio0?) Villavicencio John: Introducción a las series de tiempo, documento guía, Puerto Rico. (Disponible vía web en http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D. Visitado en noviembre 2014.

(Zubicarag1979) Zubicarag viejo. (1979): Bombas, diseño, teoría y aplicaciones. Texto guía: Editorial Limusa. <http://fluidos.eia.edu.co/hidraulica/articulos/maquinashidraulicas/clasificacionbombashidraulicas/clasificaciondelasbombashidraulicas.html>. Visitado en noviembre 2014.

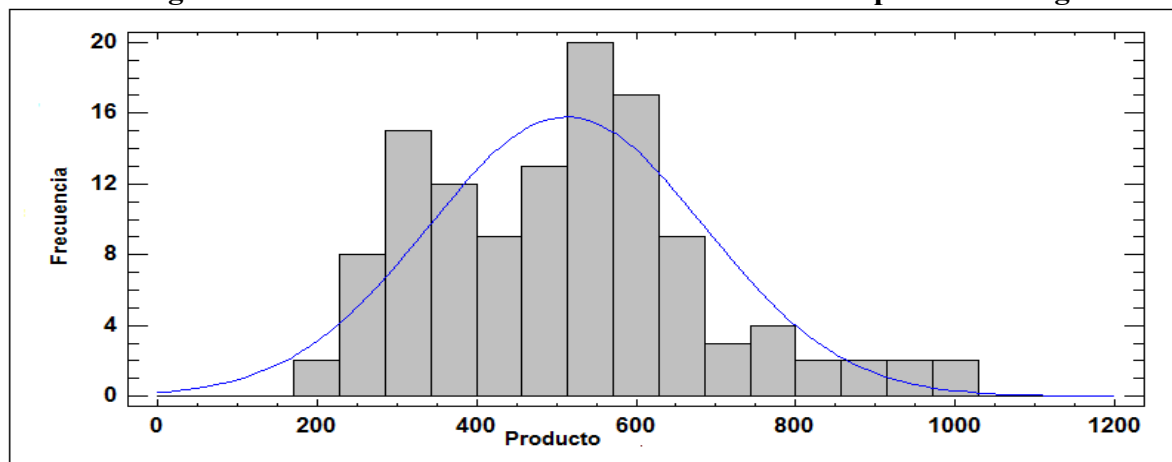
Anexo

A Distintos análisis para electrobombas Centrífugas

A1 Análisis de distribución normal

Caso 1: condición normal

Histograma de distribución normal de la serie suavizada para centrifugas



Fuente: Elaboración propia

Tabla Prueba Shapiro – Wilk para serie suavizada centrifugas

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,946679	0,00025708

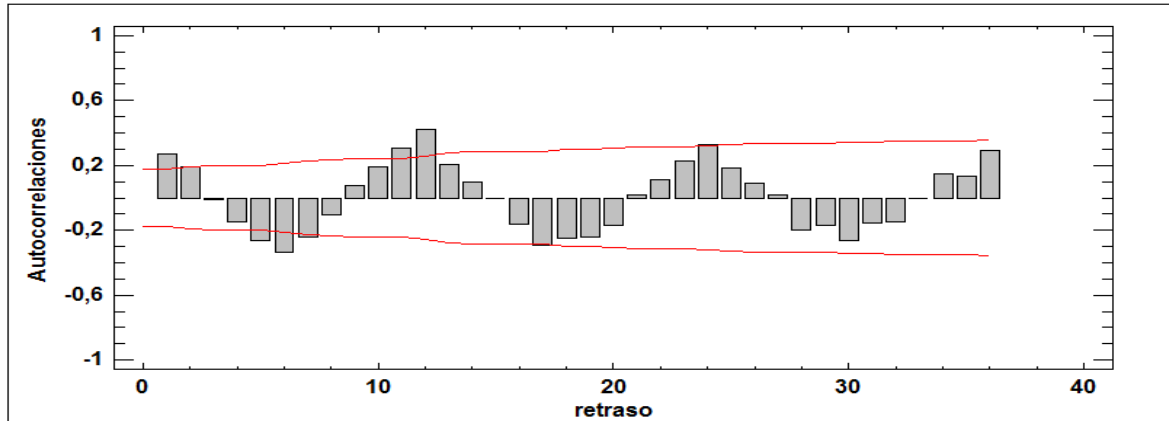
Fuente: Elaboración propia

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0,05, se puede rechazar la idea de que las electrobombas centrifugas provienen de una distribución normal con 95% de confianza.

A2 Análisis de correlogramas

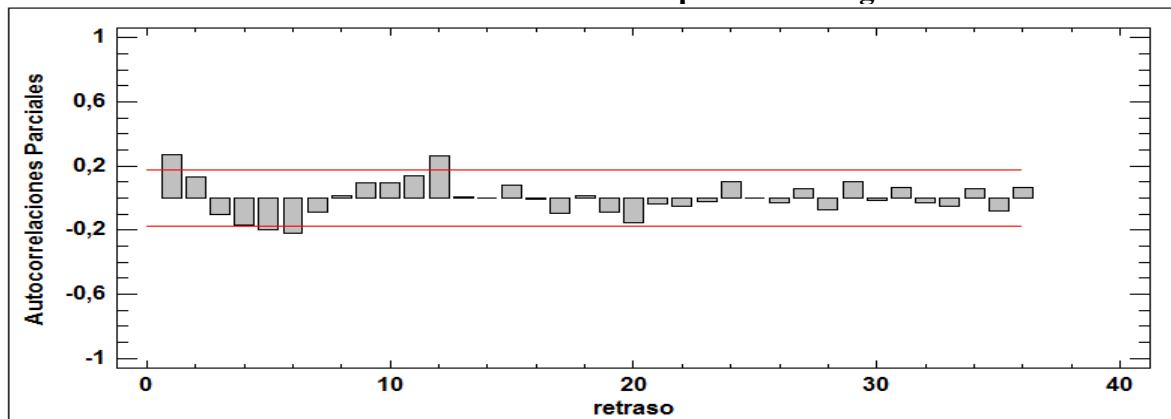
Caso 1: Serie normal sin aplicar alguna transformación

FAC de la serie suavizada para centrifugas



Fuente: Elaboración propia

FACP de la serie suavizada para centrifugas



Fuente: Elaboración propia

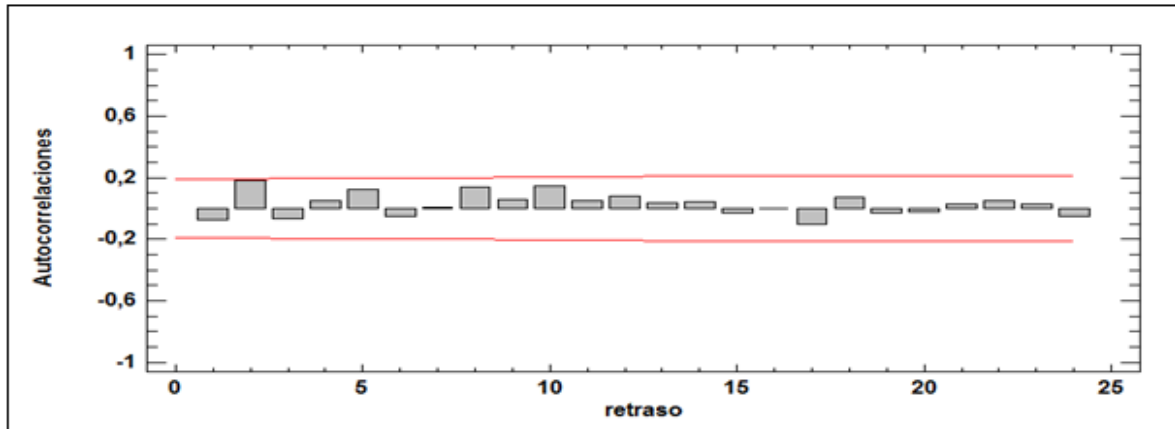
El comportamiento de los correlogramas demuestran dos partes una regular y otra estacionaria, en la primera se puede indicar que luego del primer desfase cae rápidamente a cero, por lo que se concluye que tiene estacionariedad en la parte regular, mientras que en la otra parte describe desfases significativos en los múltiplos de 12, como lo son los desfases 12 y 24, esto caen muy lentamente a cero, por lo que se concluye que esta parte no tiene estacionariedad.

Por lo que todo indica que la mejor solución para corregir esta serie con media no constante es aplicando una primera diferencia.

A3 Análisis de correlogramas residuales

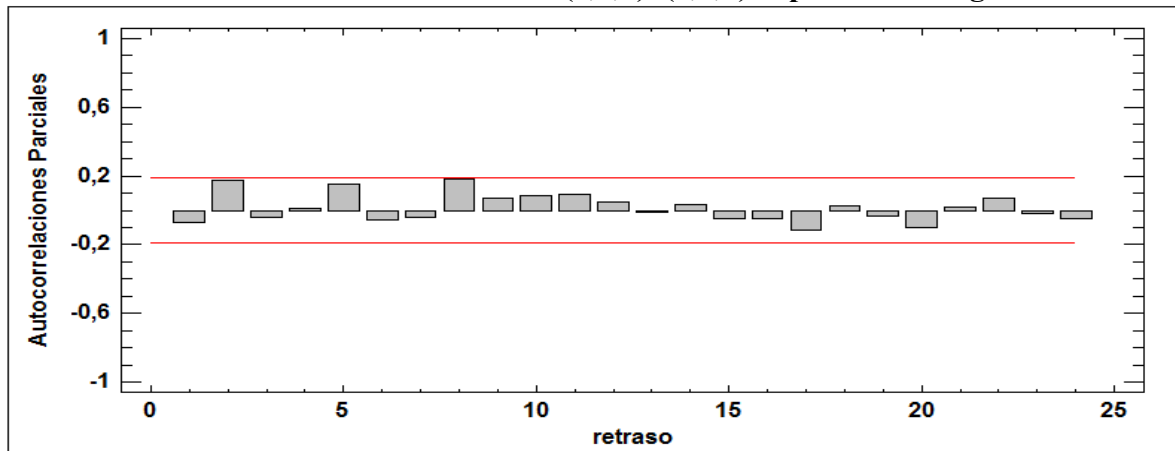
SARIMA (0,0,0)x(1,1,0)12

FAC residual de SARIMA (0,0,0)x(1,1,0)12 para centrifugas



Fuente: Elaboración propia

FACP residual de SARIMA (0,0,0)x(1,1,0)12 para centrifugas

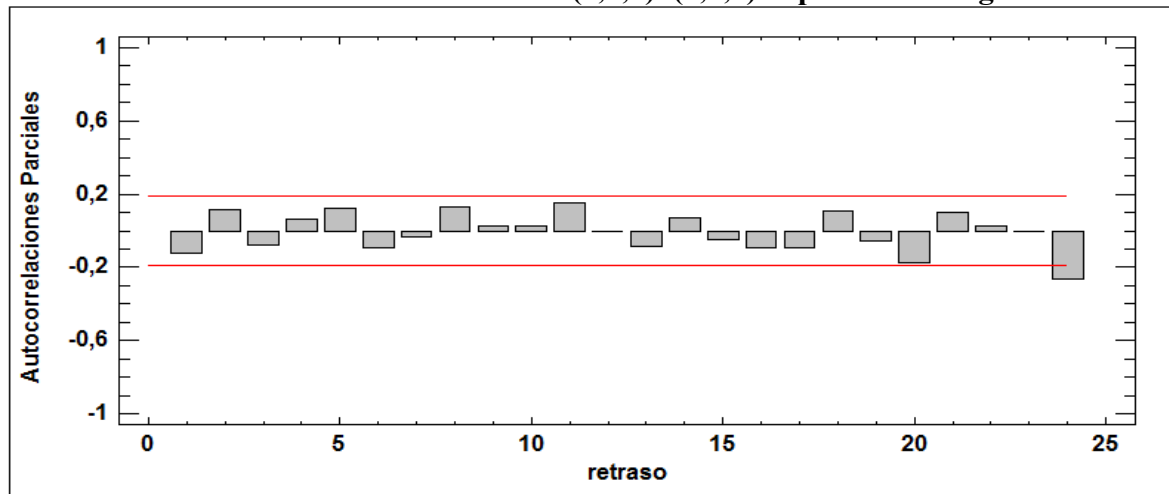


Fuente: Elaboración propia

Se puede decir que los correlogramas anteriores ambos tienen un comportamiento estable, además se comprueba mediante el estadístico de prueba Box – Pierce que su valor – P es mayor a 0,05, por lo tanto se comprueba que no posee RB, permitiendo que la herramienta predictiva actual es adecuada para este tipo de datos.

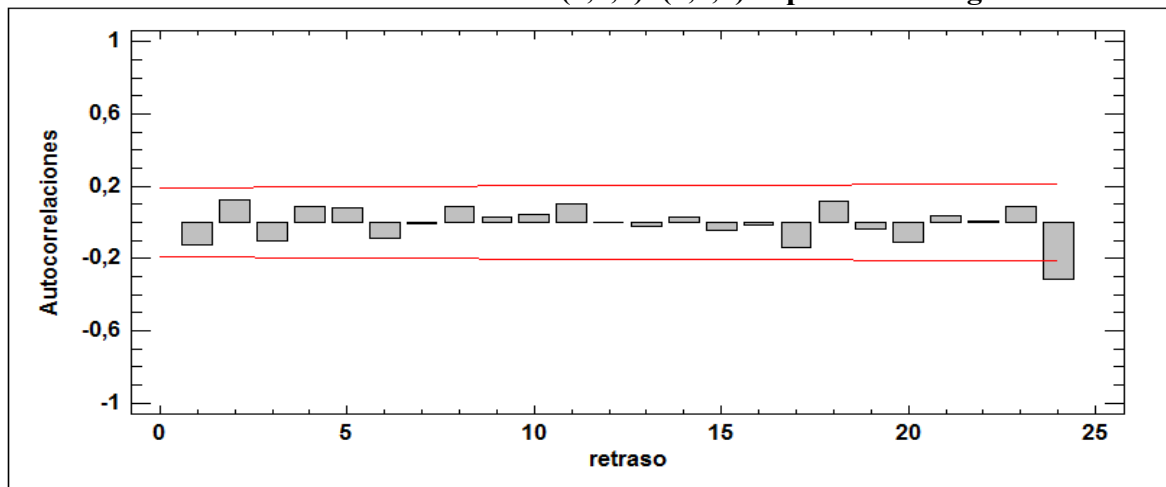
SARIMA (0,0,0)X(1,1,1)12

FAC residual de SARIMA (0,0,0)x(1,1,1)12 para centrifugas



Fuente: Elaboración propia

FACP residual de SARIMA (0,0,0)x(1,1,1)12 para centrifugas



Fuente: Elaboración propia

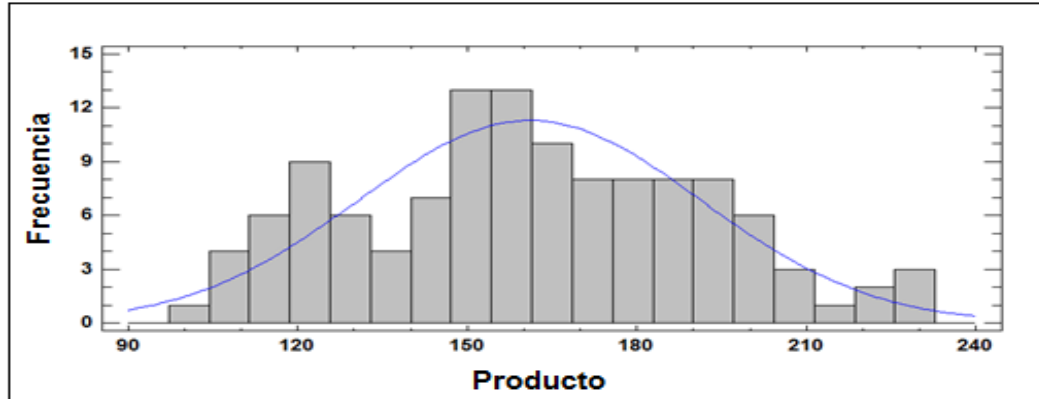
Se puede decir que los correlogramas anteriores ambos tienen un comportamiento estable, además se comprueba mediante el estadístico de prueba Box – Pierce que su valor – P es mayor a 0,05, por lo tanto se comprueba que no posee RB, permitiendo que la herramienta predictiva actual es adecuada para este tipo de datos.

B Electrobombas agua sucia

B1 Análisis de normalidad de datos

Caso 1: Serie original sin interpretar alguna transformación Box – Cox

Histograma de distribución normal de la serie suavizada para agua sucia



Fuente: Elaboración propia

Tabla Prueba Shapiro – Wilk para serie suavizada para agua sucia

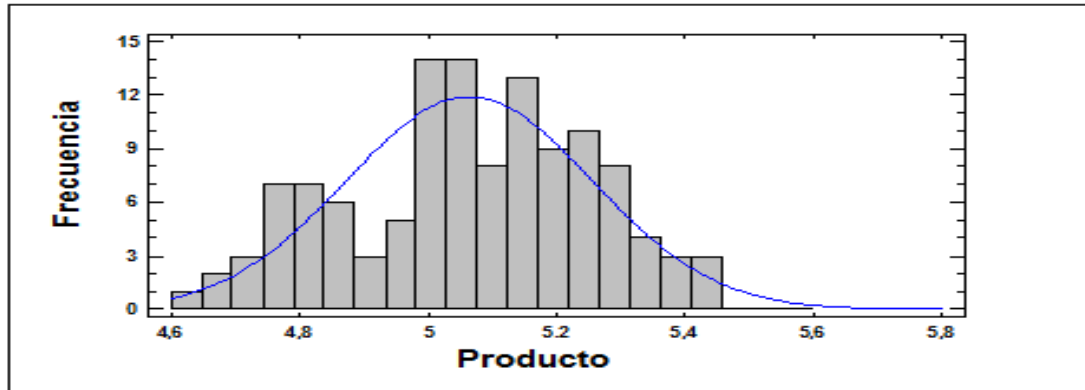
Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,964166	0,0271585

Fuente: Elaboración propia

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0,05, se puede rechazar la idea de que Agua sucia proviene de una distribución normal con 95% de confianza.

Caso 2: Serie aplicando transformación de logaritmo natural

Histograma de distribución normal de la serie suavizada transformada a logaritmo natural para agua sucia



Fuente: Elaboración propia

Tabla Prueba Shapiro – Wilk para serie suavizada transformada a logaritmo natural agua sucia

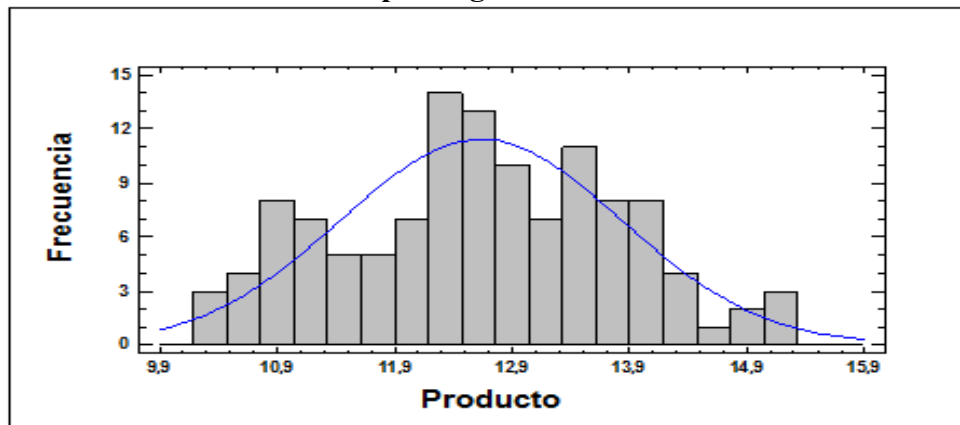
Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,961994	0,0162609

Fuente: Elaboración propia

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0,05, se puede rechazar la idea de que Agua sucia proviene de una distribución normal con 95% de confianza.

Caso 3: Aplicando raíz cuadrada

Histograma de distribución normal de la serie suavizada transformada a raíz cuadrada para agua sucia



Fuente: Elaboración propia

Tabla Prueba Shapiro – Wilk para serie suavizada transformada a raíz cuadrada agua sucia

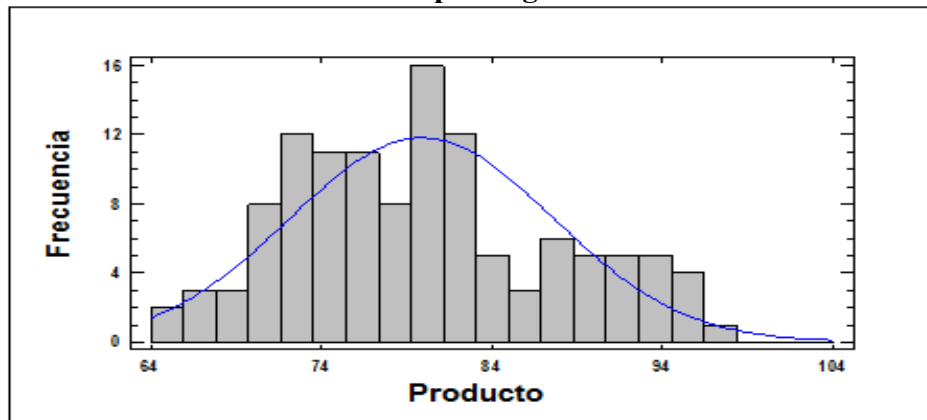
Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,965492	0,0367207

Fuente: Elaboración propia

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0,05, se puede rechazar la idea de que Agua sucia proviene de una distribución normal con 95% de confianza.

Caso 4: Aplicando raíz cuadrada inversa

Histograma de distribución normal de la serie suavizada transformada a raíz cuadrada inversa para agua sucia



Fuente: Elaboración propia

Tabla Prueba Shapiro – Wilk para serie suavizada transformada a raíz cuadrada inversa agua sucia

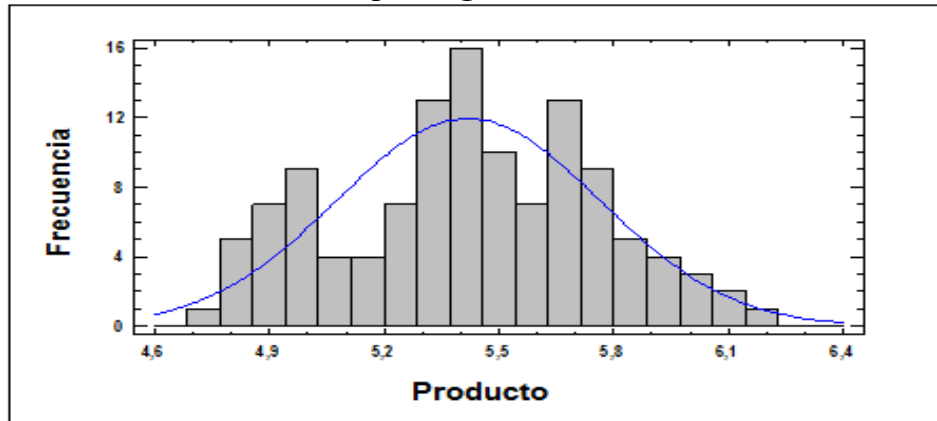
Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,953892	0,00200688

Fuente: Elaboración propia

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0,05, se puede rechazar la idea de que Agua sucia proviene de una distribución normal con 95% de confianza.

Caso 5: Aplicando transformación de raíz cubica

Histograma de distribución normal de la serie suavizada transformada en raíz cubica para agua sucia



Fuente: Elaboración propia

Tabla Prueba Shapiro – Wilk para serie suavizada transformada a raíz cubica agua sucia

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,964855	0,0318034

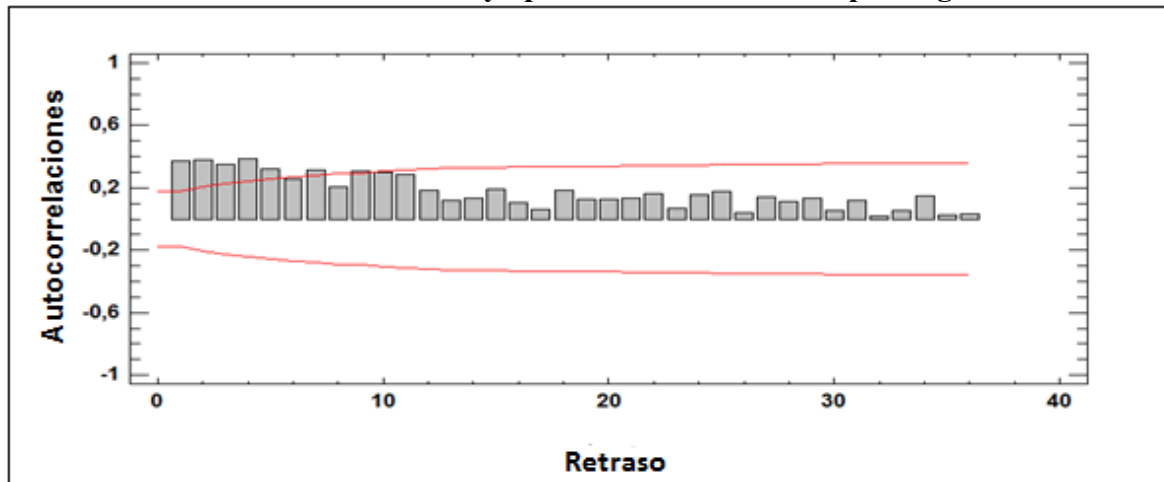
Fuente: Elaboración propia

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0,05, se puede rechazar la idea de que Agua sucia proviene de una distribución normal con 95% de confianza.

B2 Análisis de los correlogramas

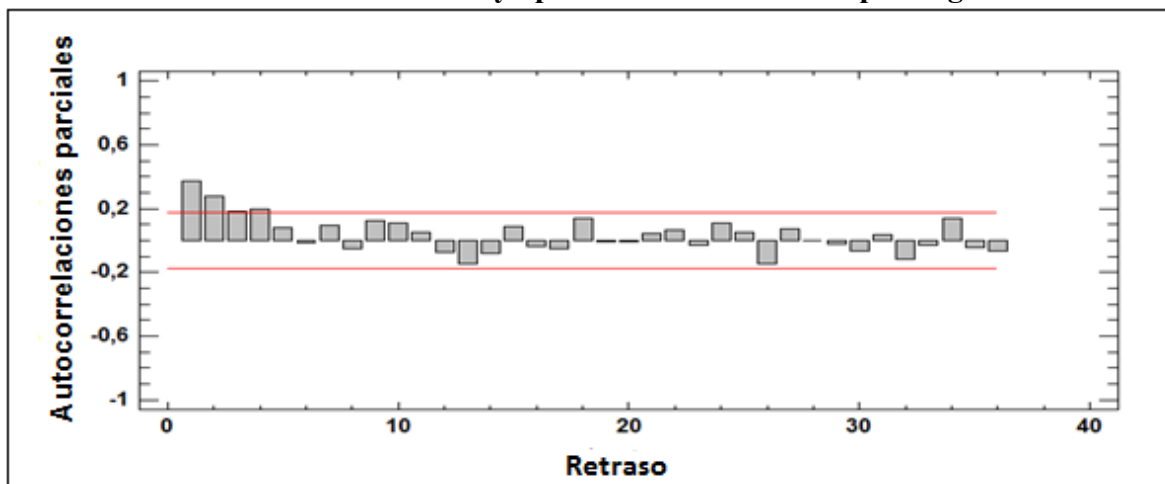
Opción 1: Serie aplicando raíz cuadrada

FAC de la serie suavizada y aplicando raíz cuadrada para agua sucia



Fuente: Elaboración propia

FACP de la serie suavizada y aplicando raíz cuadrada para agua sucia



Fuente: Elaboración propia

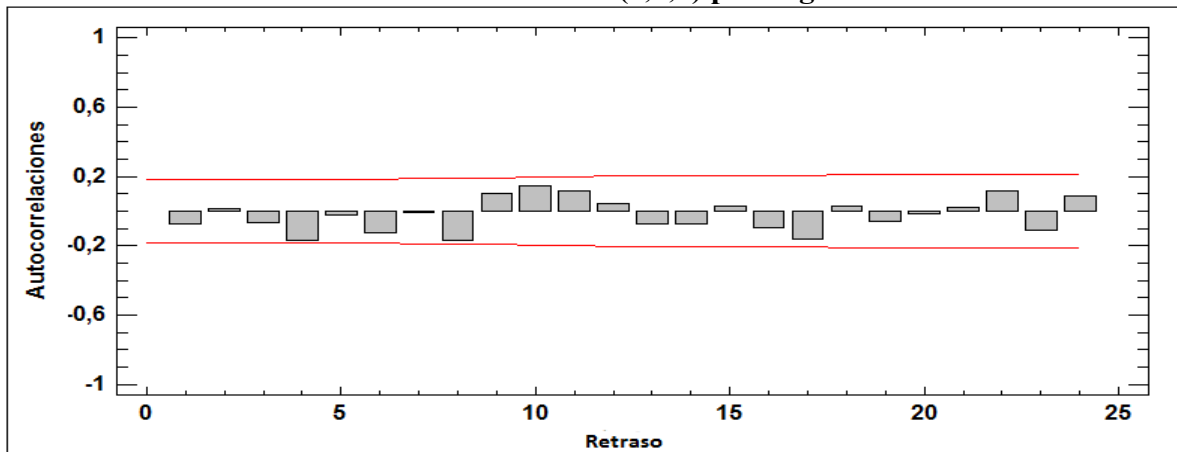
Se puede apreciar que en los correlogramas anteriores no existen un patrón estacional, por lo que sólo existe una parte que es la regular, esta experimenta desfases que no caen rápidamente a la región de confianza, por lo que se confirma que existe una tendencia demarcada y por ende una media que no se demuestra constante en el tiempo, concluyendo que la serie no presenta estacionariedad y que por lo tanto se le recomienda aplicar una primera diferencia.

B3 Análisis de gráficos residuales

El objetivo de este análisis es demostrar si el pronóstico presenta o no ruido blanco.

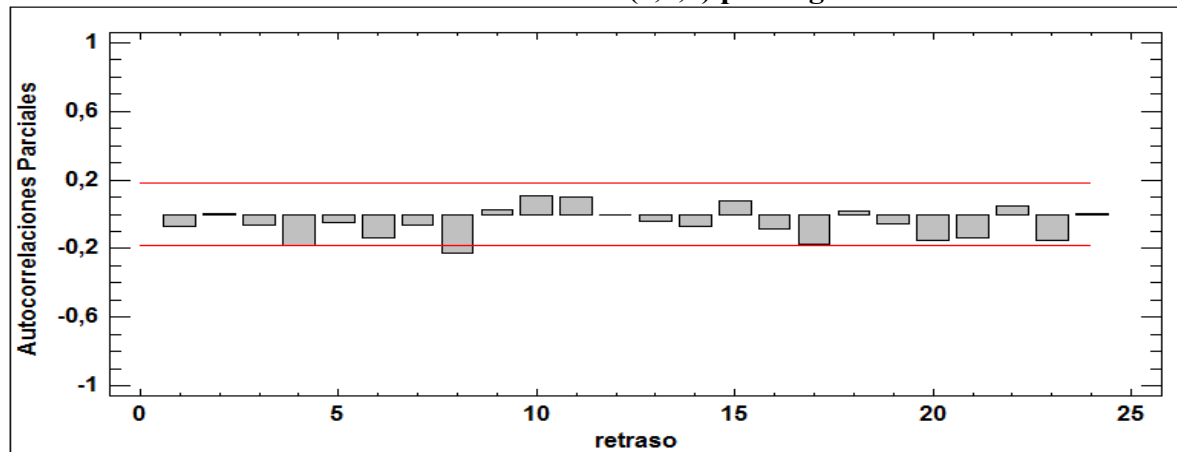
ARIMA (3,1,0)

FAC residual de ARIMA (3,1,0) para agua sucia



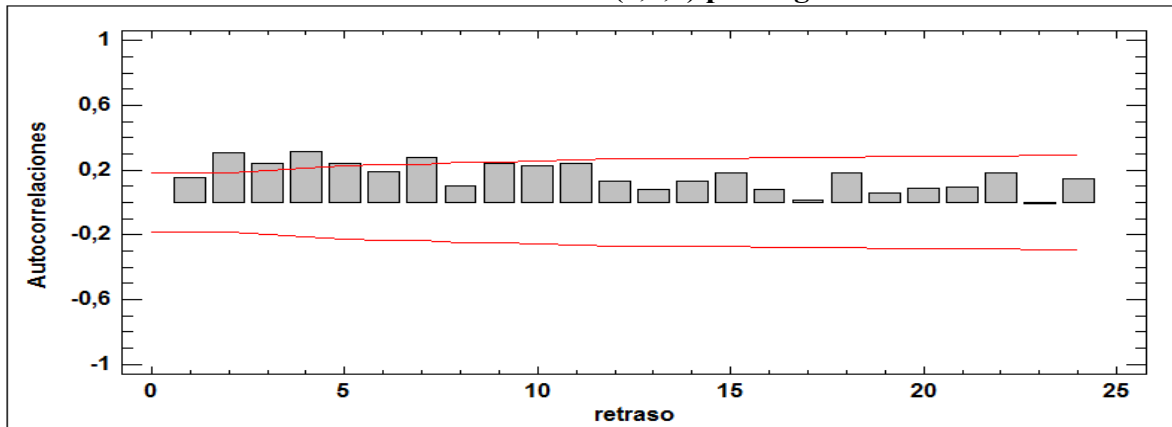
Fuente: Elaboración propia

FACP residual de ARIMA (3,1,0) para agua sucia

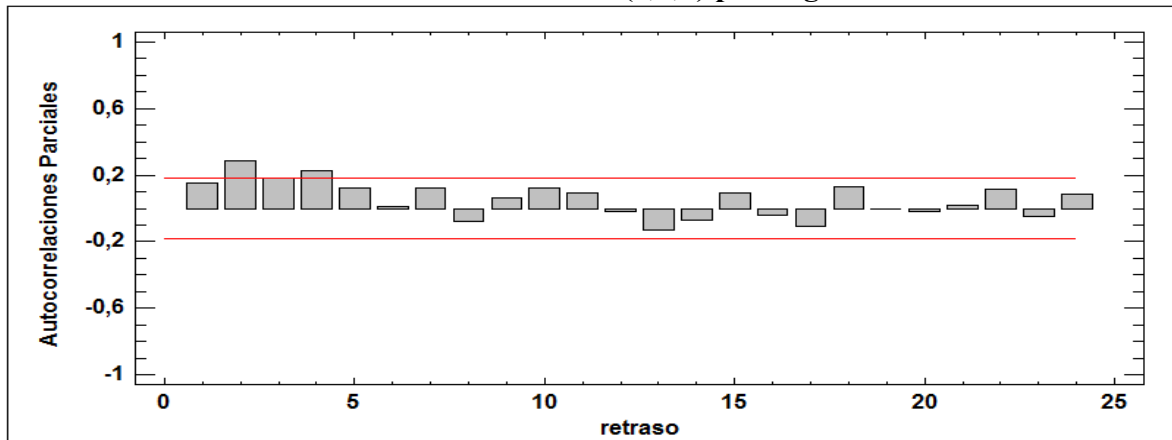


Fuente: Elaboración propia

Se puede verificar que los correlogramas presentan un desfase significativo, pero en general demuestran ser muy estables, a priori se podría decir que la serie demuestra en sus residuos un comportamiento con aleatoriedad, para rectificar aquello se realizó además la prueba Box – Pierce con el objetivo de contrarrestar la hipótesis nula que dice que la serie no tiene RB, al obtener un valor de $0,21 > 0,05$, se puede concluir que se rechaza la hipótesis nula y la serie presenta por ende RB.

ARIMA (1,1,1)**FAC residual de ARIMA (1,1,1) para agua sucia**

Fuente: Elaboración propia

FACP residual de ARIMA (1,1,1) para agua sucia

Fuente: Elaboración propia

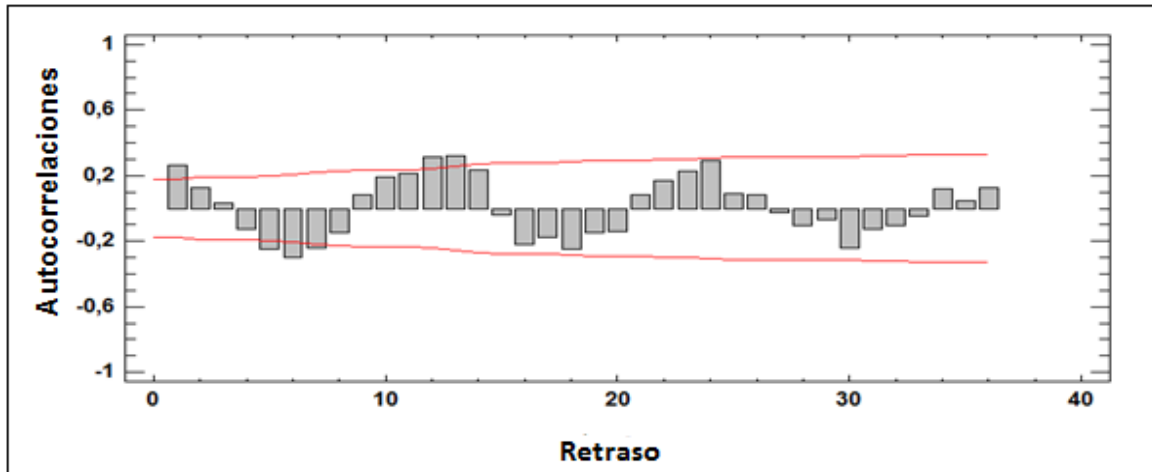
Se puede decir en los gráficos anteriores que en ambos existen muchos desfases significativos, por lo que a priori se podría decir que esta serie presenta algún RB, con la siguiente prueba se rectificará o rechazará lo expuesto.

C Producto electrobombas multietapas

C1 Análisis de los correlogramas

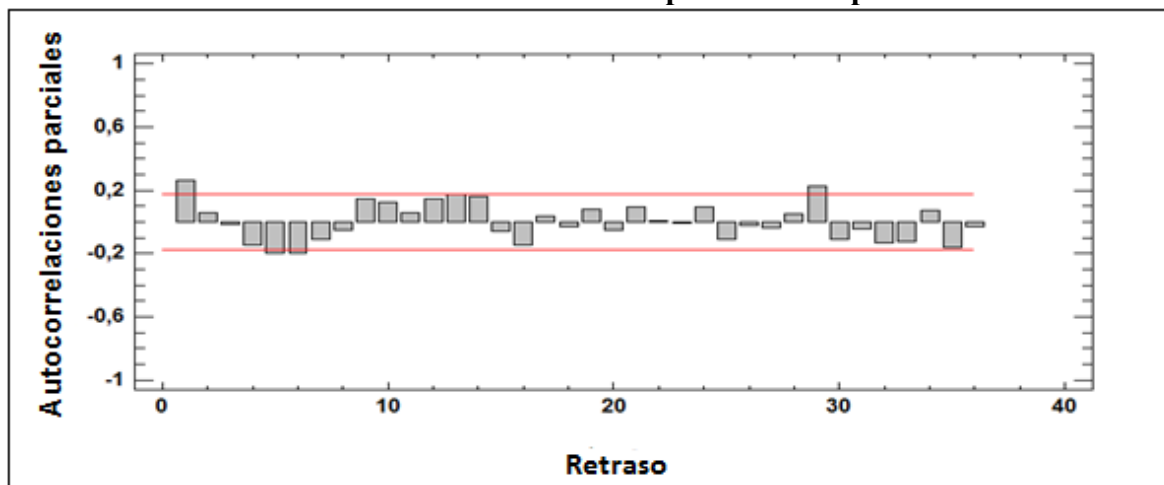
Opción 1: Asumiendo que se ocupará la lectura de correlogramas de la serie suavizada.

FAC de la serie suavizada para multietapas



Fuente: Elaboración propia

FACP de la serie suavizada para multietapas



Fuente: Elaboración propia

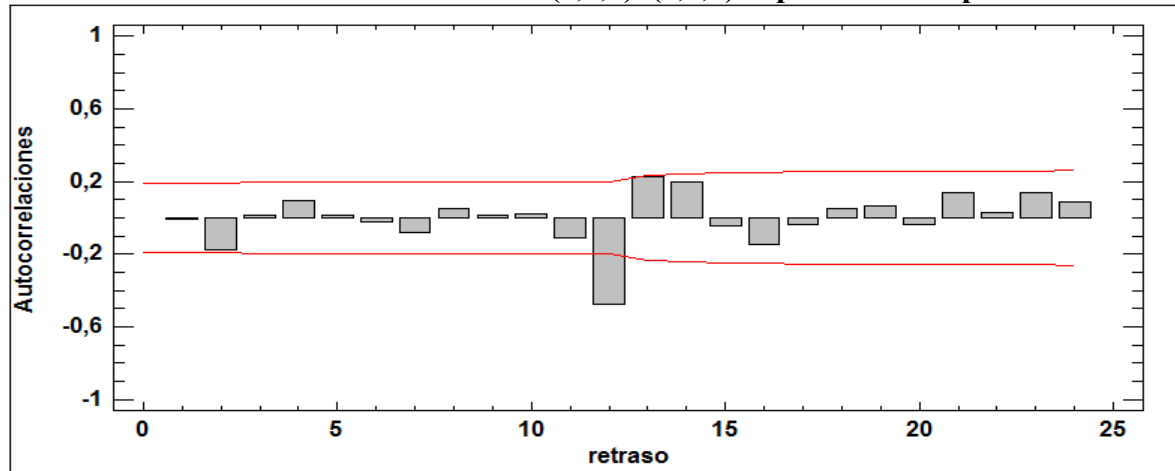
Se puede indicar que los correlogramas presentan una parte regular y además una parte estacional, esto debido a que esta última presenta un comportamiento similar que se repite cada 12 meses, la primera parte experimenta una media constante, pero en la segunda demuestra

desfases significativos y por ende no experimenta un decrecimiento rápido, por lo que será necesario aplicar una primera diferencia en esta parte.

C2 Análisis de gráficos residuales

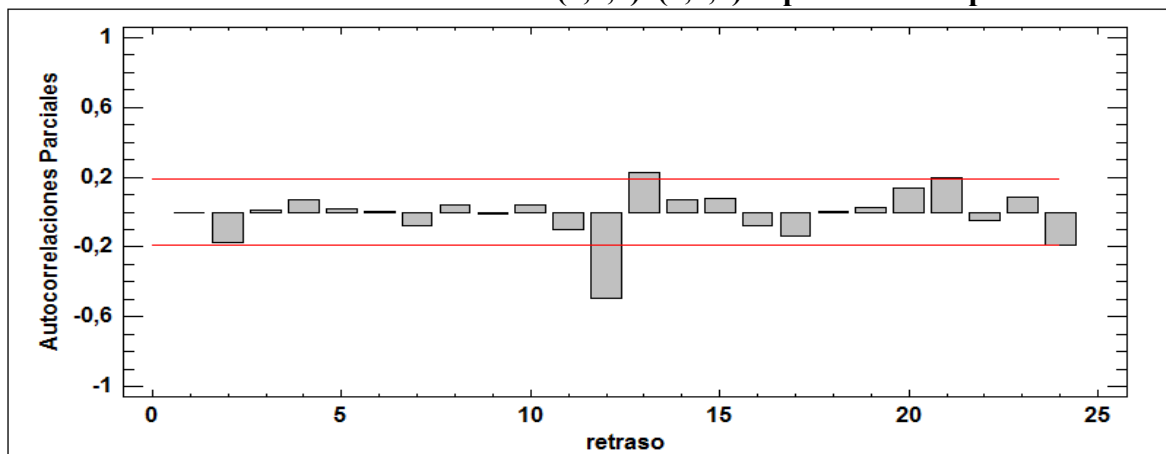
SARIMA (0,0,0)x(1,1,0)12

FAC residual de SARIMA (0,0,0)x(0,1,1)12 para multietapas



Fuente: Elaboración propia

FACP residual de SARIMA (0,0,0)x(0,1,1)12 para multietapas

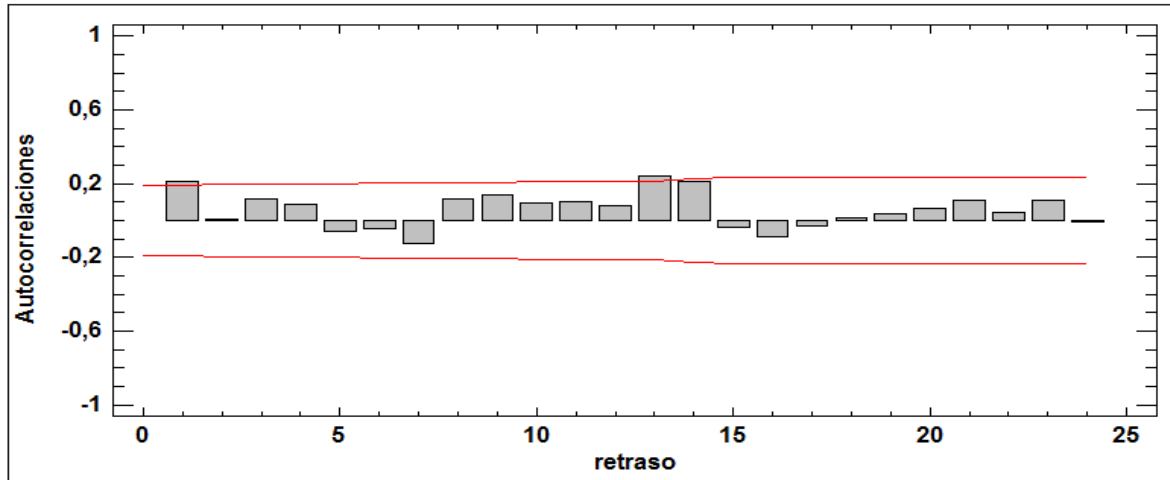


Fuente: Elaboración propia

Se comprueba que presenta RB luego de someter la serie a la prueba Box – Pierce y presentar un valor – P menor a 0,05, por lo que se concluye que no es recomendable utilizar este pronóstico en este tipo de productos.

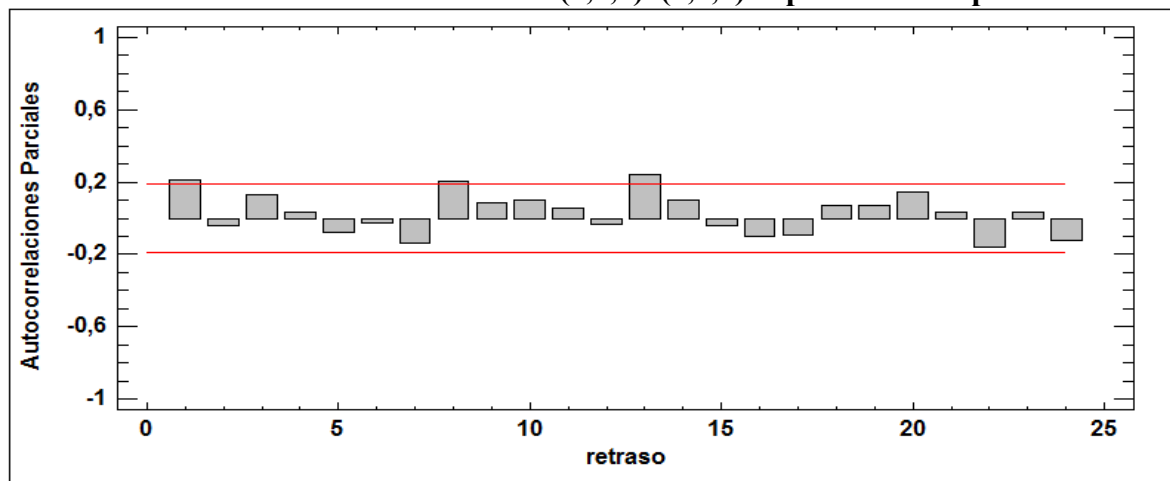
SARIMA (0,0,0)x(2,1,0)12

FAC residual de SARIMA (0,0,0)x(2,1,0)12 para multietapas



Fuente: Elaboración propia

FACP residual de SARIMA (0,0,0)x(2,1,0)12 para multietapas

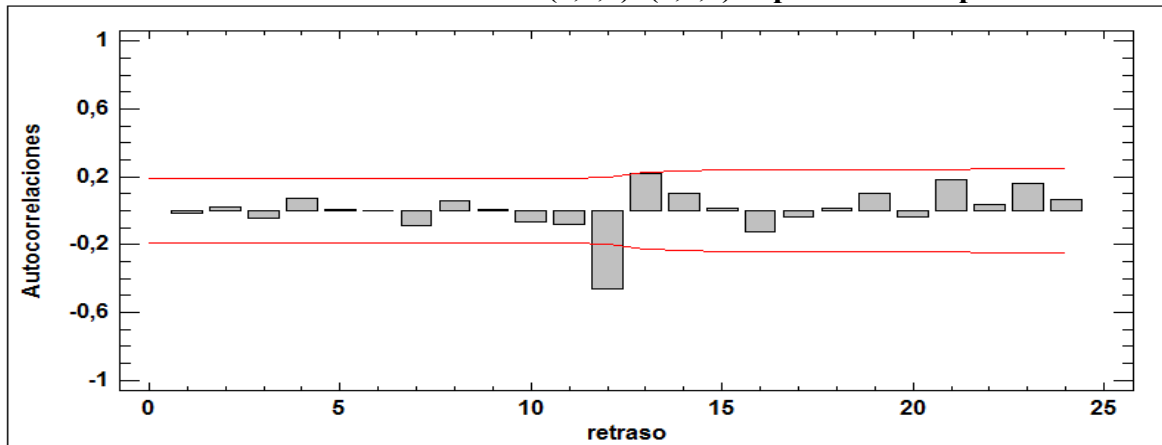


Fuente: Elaboración propia

Se puede verificar que los correlogramas presentan algunos desfases levemente significativos, pero en general demuestran ser muy estables, a priori se podría decir que la serie demuestra en sus residuos un comportamiento con aleatoriedad, para rectificar aquello se realizó además la prueba Box – Pierce con el objetivo de contrarrestar la hipótesis nula que dice que la serie no tiene RB, al obtener un valor de $0,092 > 0,05$, se puede concluir que se rechaza la hipótesis nula y la serie presenta por ende RB.

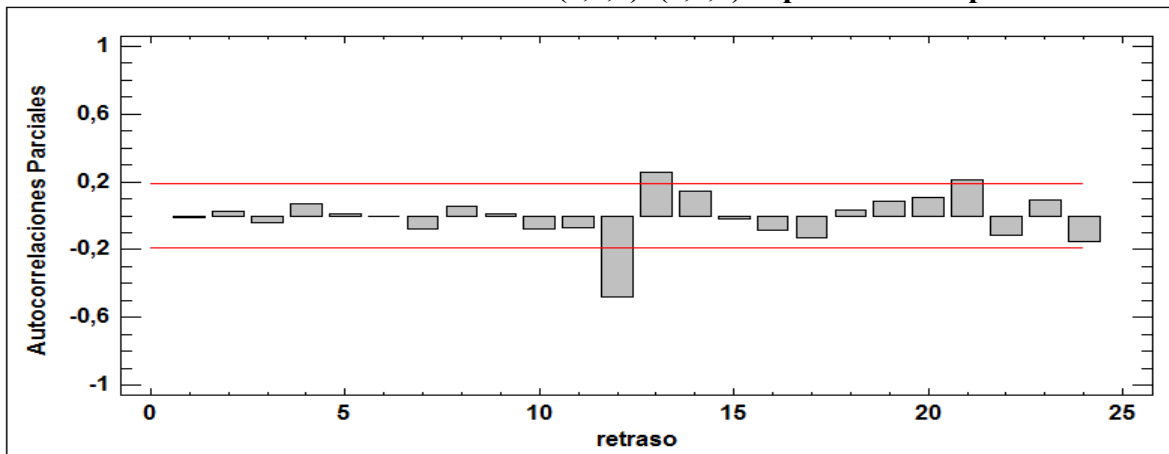
SARIMA (0,0,0)x(3,1,0)12

FAC residual de SARIMA (0,0,0)x(3,1,0)12 para multietapas



Fuente: Elaboración propia

FACP residual de SARIMA (0,0,0)x(3,1,0)12 para multietapas

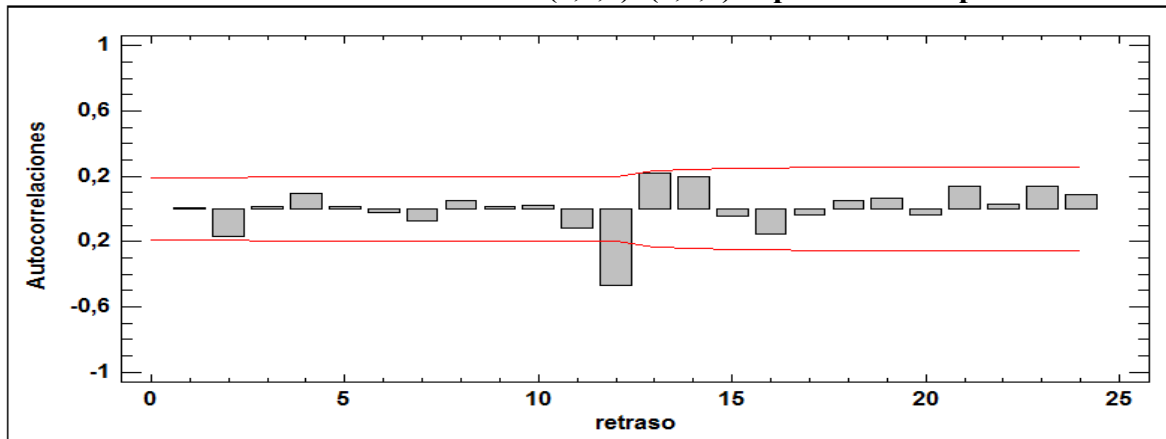


Fuente: Elaboración propia

Se comprueba que presenta RB luego de someter la serie a la prueba Box – Pierce y presentar un valor – P menor a 0,05, por lo que se concluye que no es recomendable utilizar este pronóstico en este tipo de productos.

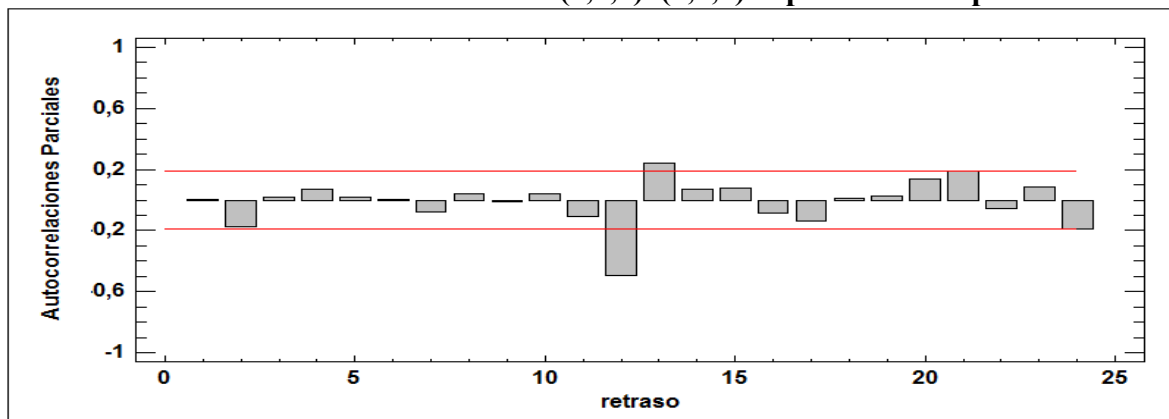
SARIMA (0,0,0)X(1,1,1)12

FAC residual de SARIMA (0,0,0)x(1,1,1)12 para multietapas



Fuente: Elaboración propia

FACP residual de SARIMA (0,0,0)x(1,1,1)12 para multietapas



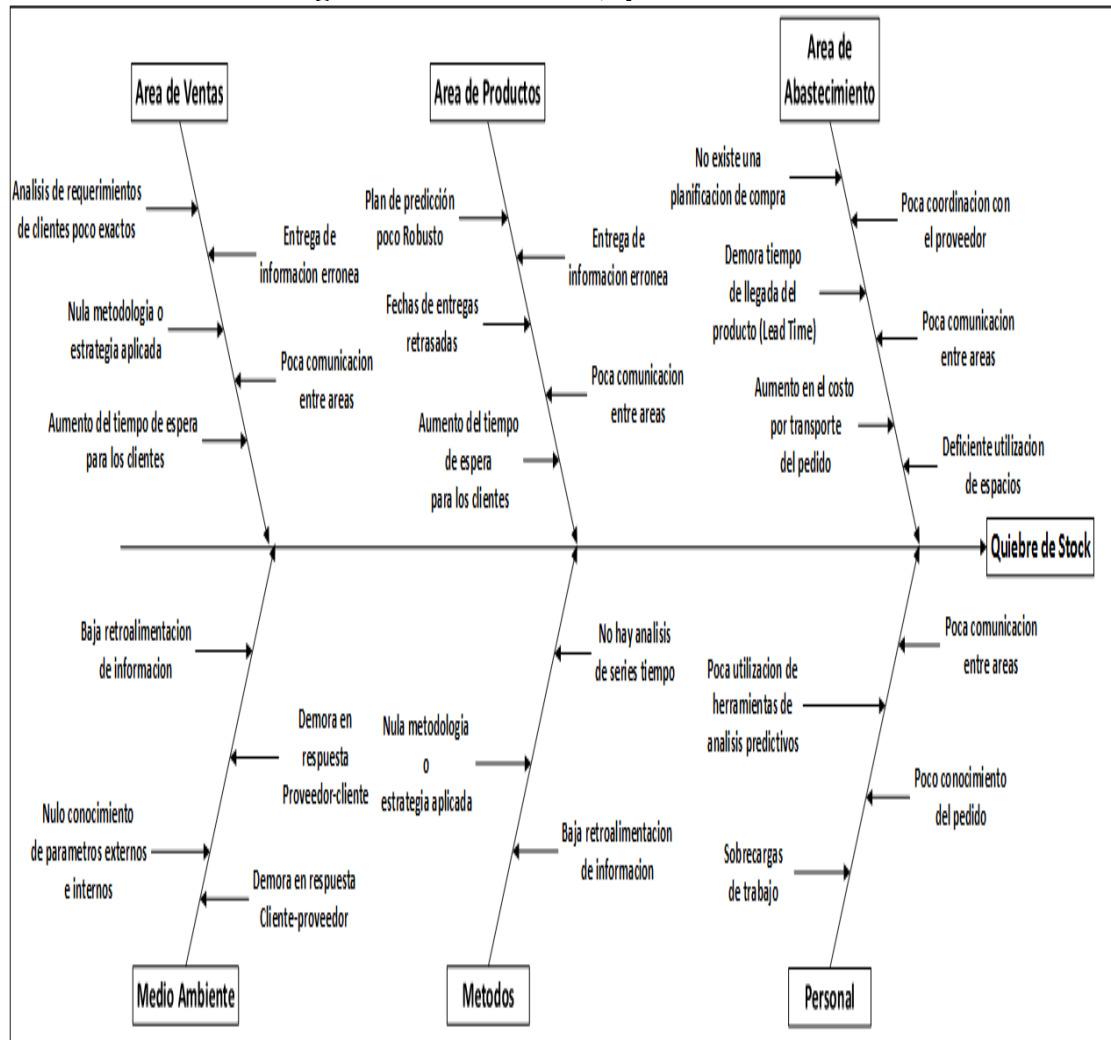
Fuente: Elaboración propia

Se comprueba que presenta RB luego de someter la serie a la prueba Box – Pierce y presentar un valor – P menor a 0,05, por lo que se concluye que no es recomendable utilizar este pronóstico en este tipo de productos.

D Diagrama causa – efecto

D1 Ishikawa caso: problemas de stock empresa Koslan

Diagrama causa – efecto, quiebres de stock



Fuente: Elaboración propia