



UNIVERSIDAD NACIONAL DEL ALTIPLANO DE PUNO

FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA

ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA



**MODELO PARA LA PREDICCIÓN DE LA DEMANDA
MENSUAL DE HORAS DE MAQUINARIA AGRÍCOLA EN LA
MUNICIPALIDAD DISTRITAL DE CARACOTO – 2020**

TESIS

PRESENTADA POR:

Bach. MAMANI CONDORI BETHO

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2020



DEDICATORIA

El presente trabajo de investigación es dedicado en primer lugar a Dios, por ser el inspirador y brindarme las fuerzas necesarias para continuar en este proceso de obtener uno de mis anhelos más deseados.

A mi padre mi madre ERCILIA CONDORI QUISPE, por su amor, trabajo y sacrificio en todos estos años, he logrado llegar hasta donde me encuentro y convertirme una persona de bien. Me inunda el orgullo y el privilegio de ser vuestro hijo, una gran madre. Y a mi amigo EDY MAYTA por sus consejos y apoyo moralmente, que me proporciono a lo largo de esta etapa de mi vida.

A todos mis compañeros de la facultad, docentes que me han apoyado y me han compartido sus conocimientos, en el transcurso del desarrollo de mi perfil profesional.

BETHO MAMANI CONDORI



AGRADECIMIENTOS

- ❖ A Dios, por permitir llegar a este momento tan especial de mi vida. Por los triunfos y los momentos difíciles que me han enseñado a valorarlo cada día más.
- ❖ A mi madre quien a lo largo de toda la vida me ha apoyado y motivado mi formación académica, creyeron en mí en todo momento y no dudaron de mis habilidades.
- ❖ A la Universidad Nacional del Altiplano de Puno, a la Facultad de Ingeniería Estadística e Informática por haber compartido sus enseñanzas y hacer realidad mi sueño de ser un profesional para el servicio de la sociedad.
- ❖ A los miembros de mi jurado de tesis, quienes con sus aportes y correcciones han hecho posible la realización de este trabajo de investigación. A mi presidente de jurado, así como a mi asesor D. Sc. Percy Huata Panca, con sus correcciones y asesoramientos han hecho posible la culminación de esta tesis.
- ❖ A los docentes de la Escuela Profesional de Ingeniería Estadística e Informática, por todo el conocimiento esparcido durante los años de mi formación académica.

BETHO MAMANI CONDORI



ÍNDICE GENERAL

DEDICATORIA

AGRADECIMIENTOS

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

ÍNDICE DE ACRÓNIMOS

RESUMEN..... 13

ABSTRACT..... 14

CAPÍTULO I

INTRODUCCIÓN

1.1. DESCRIPCIÓN DEL PROBLEMA 16

1.2. FORMULACIÓN DEL PROBLEMA 17

1.3. OBJETIVOS..... 18

1.3.1. Objetivo General 18

1.3.2. Objetivos Específicos 18

1.4. HIPOTESIS 18

1.4.1. Hipótesis General 18

1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN..... 19

1.6. LIMITACIÓN DE LA INVESTIGACIÓN..... 19

CAPÍTULO II

REVISIÓN DE LITERATURA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN 20

2.1.1. A Nivel Internacional 20

2.1.2. A Nivel Nacional..... 21

2.1.3. A nivel Local 22

2.2. DEFINICIÓN DE TÉRMINOS BÁSICOS 23

2.2.1. Maquinarias Agrícolas 23

2.2.2. Municipio 24

2.2.3. Modelo 24

2.2.4. Predicción..... 24

2.3. MARCO TEÓRICO 25

2.3.1. Metodologías de diseño de investigación..... 25



2.3.2.	Series de tiempo	27
2.3.3.	Clasificación Descriptiva de la Series de Temporales	29
2.3.4.	Utilización de las Series de Tiempo	31
2.3.5.	Modelo de Series de Tiempo.....	32
2.3.6.	Modelo Clásico de una Serie de Tiempo.....	33
2.3.7.	Modelo Univariante.....	34
2.3.8.	Procesos Estocásticos	34
2.3.9.	Ruido blanco (“White noise”).....	35
2.3.10.	Autocorrelación	36
2.3.11.	Proceso Lineal Estacionarios	37
2.3.12.	Procesos Lineales no Estacionarios.....	40
2.3.13.	Transformación de la Seria de Tiempo	41
2.3.14.	Metodología Box - Jenkins	41
2.3.15.	Prueba de Dickey – Fuller Ampliado (ADF)	45
2.3.16.	Diagrama de caja.....	46
2.3.17.	Prueba chi cuadrado de Pearson.....	46
2.3.18.	Medición del error en pronósticos de series.....	47

CAPITULO III

MATERIALES Y METODOS

3.1.	LOCALIZACIÓN.....	50
3.2.	OPERACIONALIZACIÓN DE VARIABLES	50
3.3.	MATERIALES.....	51
	3.3.1. Software	51
	3.3.2. Hardware	51
3.4.	TIPO Y DISEÑO DE INVESTIGACION	52
	3.4.1. Tipo de Investigación	52
	3.4.2. Diseño de la investigación.....	52
3.5.	POBLACION Y MUESTRA DE INVESTIGACION.....	53
	3.5.1. Población.....	53
	3.5.2. Muestra.....	53
3.6.	TÉCNICA E INSTRUMENTO DE RECOLECCIÓN DE DATOS	53
3.7.	METODO DE RECOLECCIÓN DE DATOS	53
3.8.	METODOLOGIA DE ANALISIS DE DATOS	54
3.9.	METODOS DE TRATAMIENTO DE DATOS	54



3.9.1. Metodología de Box - Jenkins.....	55
CAPITULO IV	
RESULTADOS Y DISCUCIÓN	
4.1. RESULTADOS DE RECOPIACIÓN DE LA SERIE DE DATOS	59
4.1.1. Fase de análisis del comportamiento de la serie	60
4.2. TRANSFORMACIÓN DE LA SERIE DE DATOS.....	67
4.2.1. Aplicando la Primera Diferencia de la serie de datos	67
4.3. APLICACIÓN DE LA METODOLOGIA BOX JENKINS PARA	
TIEMPO DE TRABAJO DE LAS MAQUINARIAS AGRICOLAS	
.....	71
4.3.1. Identificación del Modelo	71
4.3.2. Estimación del modelo	73
4.3.3. Validación del Modelo	81
4.3.4. Pronostico con el Modelo Estimado.....	85
V. CONCLUSIONES.....	88
VI. RECOMENDACIONES	89
VII. REFERENCIAS BIBLIOGRÁFICAS.....	90
ANEXOS.....	94
ANEXO N° 01 Comandos en R Exploración de serie de datos.	94
ANEXO N° 02 Comando R Descomposición de la serie de datos.....	95
ANEXO N° 03 Comando R Grafica de estacionalidad de la serie de	
datos	95
ANEXO N° 04 Comando R Análisis de la serie de datos	96
ANEXO N° 05 Comando R Calibración de la serie de datos.....	98
ANEXO N° 06 Comando R Aplicamos el test de Ljung box Ruido	
blanco para los 24 coeficiente	101

Área : Estadística
Tema : Series de Tiempo

FECHA DE SUSTENTACION: 30 de octubre 2020



ÍNDICE DE FIGURAS

Figura 1	Tendencia de las desviaciones de temperatura global.	27
Figura 2	Serie de muertes por neumonía mensual	28
Figura 3	Pasajeros de líneas aéreas internacionales para el periodo comprendido entre enero de 1949 y diciembre de 1962	28
Figura 4	Ventas semanales por arriba de 100,000 rollos de toallas de papel absorbente.....	29
Figura 5	Leguas diarias recorridas por la flota de Cristóbal Colón en su primer viaje a América desde la isla de la Gomera a la de San Salvador.....	30
Figura 6	Nacimientos anuales en España.....	31
Figura 7	Grafico de serie de tiempo y autocorrelación simple.	35
Figura 8	Ejemplo de autocorrelación simple y autocorrelación parcial.	38
Figura 9	Etapas de la metodología box – Jenkins	44
Figura 10	Explicación de diagrama de cajas.....	46
Figura 11	Serie original del número de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.....	60
Figura 12	Diagrama de descomposición de la serie de datos horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 - 2019.	61
Figura 13	Grafica de la tendencia de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 - 2019.....	62



Figura 14 Gráfico de estacionalidad de la serie de horas de trabajo de las maquinarias agrícolas de la municipalidad distrital de Caracoto 2010 - 2019.....	62
Figura 15 Gráfico de cajas de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 - 2019.....	63
Figura 16 Histograma y distribución normal de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.....	64
Figura 17 Grafico de normalidad Q – Q Plot de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.....	65
Figura 18 Resultados del test de Pearson de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.....	66
Figura 19 Resultado del comando Auto-arima de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.	66
Figura 20 Grafico de caja con primera diferencia de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 al 2019.....	68
Figura 21 Histograma y distribución normal con primera diferencia de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 al 2019.....	69



Figura 22 Grafico de normalidad Q – Q Plot de la serie con primera diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.....	70
Figura 23 Resultados del test de Pearson de la serie con primera diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.	71
Figura 24 Grafico de la serie con una diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.....	72
Figura 25 Resultados del test de Dickey – Fuller de la serie con segunda diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.....	73
Figura 26 Modelos propuestos mediante <i>Auto– Arima</i> a la serie con una diferencia de las Horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.....	75
Figura 27 Coeficientes del mejor modelo propuesto con Auto – Arima de la serie dhora con un $MA(1)$, una diferencia en la parte regular y una en la parte estacionaria de las Horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.....	75
Figura 28 Correlación simple y parcial del modelo recomendado por <i>AUTO – ARIMA</i> de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.	76
Figura 29 Grafico de radar comparación de Medición de errores (ER, ECM, TFA y PTFA) de los modelos propuestos	81



Figura 30 Autocorrelación y Box - Ljung para el modelo ARIMA (2, 1, 0) (1, 1, 1) (12) de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.....	84
Figura 31 Grafico del Pronostico de la serie de las horas de trabajo de las maquinarias agrícolas de la municipalidad de Caracoto 2020.	86



ÍNDICE DE TABLAS

Tabla 1 Operacionalización de variables.....	50
Tabla 2 Serie original del total de horas de trabajo mensual de las maquinarias de la municipalidad de Caracoto 2010 – 2019. ...	59
Tabla 3: Datos con primera diferencia de la serie horas de trabajo de la municipalidad distrital de Caracoto 2010 al 2019.	68
Tabla 4 Primer Grupo ARENDIZAJE serie de horas de trabajo de la municipalidad distrital de Caracoto del 2010 al 2018 y Grupo de Testing del 2019.....	77
Tabla 5 Cuadro de los Errores Cuadráticos Medios de los modelos calibrados de la función programada.	78
Tabla 6 Cuadro de Medición de errores en predicción de series de los modelos propuestos para la serie de horas de trabajo de maquinarias de la municipalidad distrital de Caracoto.....	80
Tabla 7 Tabla de prueba Box – Ljung para los 24 retardos.	83
Tabla 8 Pronostico de la serie de las horas de trabajo de las maquinarias agrícolas de la municipalidad distrital de Caracoto 2020.	86



ÍNDICE DE ACRÓNIMOS

ARIMA: Acrónimo del inglés autoregressive integrated moving average.

ARADO: Causante en muchos casos del crecimiento y desarrollo irregular de los cultivos.

ACF: Función de Auto – Correlación.

PACF: Función de Auto – Correlación Parcial.

ADF: Dickey – Fuller Ampliado



RESUMEN

La Municipalidad Distrital de Caracoto - San Román, es una institución pública que cuenta con una superficie territorial de 285,87 Km², y según su población está compuesta por grandes agricultores que demandan de maquinaria agrícola para el arado de sus tierras; y las Municipalidad reciben solicitudes de maquinaria agrícola y según a las demandas recibidas se programa las maquinas agrícola, los operarios y la cantidad de combustible que se necesita, es por ello que se necesita realizar un estudio y saber las futuras demandas. El objetivo fue hacer un estudio que se ajustó al modelo de series de tiempo SARIMA. La metodología que se utilizó para el estudio fue la de Box-Jenkins, la información fue archivada por la Municipalidad y los datos se tomaron la cantidad de horas mensuales de demanda de maquinarias de las distintas actividades agrícolas entre los años 2010 y 2019; el cual fue proveído por la municipalidad para digitalizarlo, el trabajo de investigación para el procesado de los datos se realizó con el lenguaje de programación R, que nos proporciona un amplio abanico de herramientas estadísticas (modelos lineales y no lineales, test estadísticos, análisis de series temporales, algoritmos de clasificación y agrupamiento, etc.) que nos ayudó a encontrar el modelo para anticiparnos a las futuras demandas y que la Municipalidad tome decisiones. Como resultado obtuvimos el modelo de pronóstico identificado SARIMA (2,1, 0) (1, 1, 1) (12) que más explica la serie. Se concluye que el modelo obtenido realiza la predicción para el año 2020 con un nivel de confianza del 95%.

Palabras Clave: Demanda, Estimación, Modelo, Parámetro, Predicción.



ABSTRACT

The District Municipality of Caracoto - San Román, is a public institution that has a territorial area of 285.87 km², and according to its population it is composed of large farmers who demand agricultural machinery to plow their land; and the Municipalities receive requests for agricultural machinery and according to the received demands, the agricultural machines, the operators and the amount of fuel needed are programmed, which is why a study is needed and to know the future demands. The objective was to carry out a study that was adjusted to the SARIMA time series model. The methodology used for the study was that of Box-Jenkins, the information was filed by the Municipality and the data was taken on the number of monthly hours of demand for machinery from the different agricultural activities between 2010 and 2019; which was provided by the municipality to digitize it, the research work for data processing was carried out with the R programming language, which provides us with a wide range of statistical tools (linear and non-linear models, statistical tests, analysis of time series, classification and grouping algorithms, etc.) that helped us to find the model to anticipate future demands and for the municipality to make decisions. As a result, we obtained the SARIMA (2,1, 0) (1, 1, 1) (12) identified forecast model that further explains the series. It is concluded that the model obtained makes the prediction for the year 2020 with a confidence level of 95%.

Keywords: Demand, Estimate, Model, Parameter, Prediction.



CAPÍTULO I

INTRODUCCIÓN

Muchos analistas nacionales y extranjeros inmersos en la problemática agraria, consideran que el aumento de la producción agropecuaria es vital; pero este logro sería relativo si paralelo a ello no se mejora el aspecto social, económico y medio ambiental del poblador rural, a través de innovaciones tecnológicas que modernicen el proceso de la agricultura, de tal manera que se logre aliviar el esfuerzo físico que requiere las diversas labores agrícolas. Nuestro país cuenta con valles costeros e interandinos.

El tractor agrícola se ha convertido en un factor importante para transformar y desarrollar energía en la producción agropecuaria, contribuyendo decididamente a mejorar las condiciones de vida del productor rural; hay que tener en cuenta el elevado costo que significa la adquisición de la maquinaria, que limita su compra por un solo productor, para ello la participación de la municipalidad facilita a este sector y puedan brindar los servicios de maquinaria a los agricultores de la zona, a todos los pobladores que soliciten con anticipación detallando el tiempo, el lugar y para que operación solicita la maquinaria, después de un seguimiento se da el servicio. Como se puede apreciarse, los servicios que se prestan en el altiplano son aún deficiente por la poca organización anticipada y generando gastos a los agricultores para la contratación de maquinarias particulares.

Por ello se realizó este trabajo de investigación que busca obtener un modelo de predicción para saber un estimado de horas de demanda de maquinarias agrícolas de próximos meses, de esta manera la municipalidad podrá organizarse anticipadamente en cuanto a las maquinarias, combustible y operarios para dar



servicio en tiempo optimizar y dar oportunidad de realizar labores agrícolas a tiempo.

En el presente trabajo se estudia la demanda en horas por mes de maquinarias en la municipalidad distrital de Caracoto, del 2010 hasta el 2019.

El presente trabajo de investigación está dividido en cuatro capítulos:

En el primer capítulo se detallan la descripción y formulación del problema, así como los objetivos, hipótesis, justificación y limitaciones del trabajo de investigación.

En el segundo capítulo se adjuntó los antecedentes y el sustento teórico de la investigación.

En el tercer capítulo se puntualizaron el lugar, operacionalización de variables, los métodos del trabajo de investigación, detallando el diseño metodológico y las técnicas estadísticas para el análisis de la serie de tiempo.

En el cuarto capítulo se realizó el análisis de los datos obtenidos, usando la metodología de Box Jenkins la identificación, Estimación, Verificación y Pronostico.

Finalmente presentamos las conclusiones y recomendaciones.

1.1. DESCRIPCIÓN DEL PROBLEMA

La Municipalidad Distrital de Caracoto, cuenta con una gran cantidad pobladores de área rural, por lo cual tiene una gran cantidad de agricultores, que necesitan el apoyo de la municipalidad brindándole el servicio a tiempo y costo accesible para el arado y distintas operaciones que permite la maquinaria agraria, de esta manera aportar la producción de alimentos andinos y economía de los



agricultores. Evitando el gasto en alquiler de maquinarias de particulares que también brindan el servicio en un precio mayor.

La Municipalidad Distrital de Caracoto, brinda servicio por años a su población agricultora, aumentando la cantidad de maquinarias para abarcar a una gran cantidad de agricultores, sin embargo, todavía es deficiente el servicio por la poca organización, generando espera por parte de los agricultores, postergación de días citados y más.

La organización del tiempo de trabajo lo realiza el encargado de las maquinarias agrarias, no es tan eficiente ya que no se previene la distancia a la que va, de esta manera se retrasa para los siguientes usuarios que solicitaron.

El uso de las maquinarias agrícolas es muy necesario para poder desarrollar la agricultura y es una de los factores de producción importante debido al tiempo y oportunidad de realizar las labores como: arado de tierra, restras, trilladoras, empacadores y segadoras.

Finalmente, contar con un modelo de predicción para saber un estimado de horas de demanda de maquinarias agrícolas de próximos meses, de esta manera la municipalidad podrá organizarse anticipadamente en cuanto a las maquinarias, combustible y operarios para dar servicio en tiempo optimizar y dar oportunidad de realizar labores agrícolas a tiempo.

1.2. FORMULACIÓN DEL PROBLEMA

¿Cuál es el modelo que mejor se ajusta para predecir la demanda mensual en horas de maquinarias para los agricultores del Distrito de Caracoto?



1.3. OBJETIVOS

1.3.1. Objetivo General

Determinar el mejor modelo que se ajuste para predecir la demanda mensual en horas de maquinarias, a través de la metodología Box-Jenkins para los agricultores del Distrito de Caracoto – 2020.

1.3.2. Objetivos Específicos

- Explorar la serie de tiempo de horas de trabajo mensual de las maquinarias agrícolas de la Municipalidad Distrital de Caracoto.
- Encontrar y validar el modelo identificado que mejor se ajusta para la serie de las horas de trabajo mensual de las maquinarias agrícola de la Municipalidad Distrital de Caracoto.
- Determinar el pronóstico con el modelo alcanzado para la serie de tiempo de la demanda mensuales de horas de maquinaria agrícola en la Municipalidad Distrital de Caracoto.

1.4. HIPOTESIS

1.4.1. Hipótesis General

El modelo usando la metodología Box – Jenkins proporciona un mejor ajuste para predecir la demanda mensual de horas de maquinarias agrícola en la Municipalidad Distrito de Caracoto.



1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN

El proyecto se consideró muy necesario, ya que la Municipalidad cuenta con gran cantidad de agricultores que se dedican al sembrío, En grandes cantidades por lo cual es necesario el arado de sus tierras. Actualmente la atención a los pedidos es regularmente eficiente. El modelo es necesario para predecir la demanda de los próximos meses y de esta manera la municipalidad distrital podrá atender a tiempo las solicitudes de maquinarias para el arado, se tendrá operarios a tiempo y combustible. El trabajo se realizó debido a los diferentes tipos de incidentes que se presentaron en los últimos años con la manera en la que se atiende la solicitud de maquinaria, usualmente la municipalidad recibe las solicitudes y de acuerdo a la demanda organiza la maquinaria, operarios y combustible por lo cual atendía las solicitudes con retraso.

1.6. LIMITACIÓN DE LA INVESTIGACIÓN

No se encontró datos de los tres trimestres del año 2020 por las medidas tomadas por el gobierno frente a la situación generada por el COVID-19 y se entró en cuarentena, otra limitación fue la dificultad de la recopilación de información histórica por meses, ya que se encontraron en diferentes documentos físicos, se tuvo que digitalizar de los diversos documentos que nos facilitó la Municipalidad Distrital de Caracoto.



CAPÍTULO II

REVISIÓN DE LITERATURA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

Las investigaciones de proyecciones a partir de series de tiempo, se ha realizado en variedad de disciplinas, en la parte financiera, es conveniente mencionar las investigaciones que mucho se asemeja, aplicando herramientas estadísticas para la toma de decisiones.

2.1.1. A Nivel Internacional

(Morales Zapata, 2015) En la presente investigación “diseño de un sistema de arado para la preparación del suelo en cultivos orgánicos dirigido a pequeños agricultores” presentó la labranza cero o siembra directa como una técnica conservadora de producción agrícola que ha surgido como respuesta a la degradación de los recursos, principalmente del suelo y del medio ambiente, dando como resultado ciertos beneficios que serán expuestos para demostrar la productividad y sostenibilidad de este sistema.

(Bernejo Mancera, 2011) En su trabajo de investigación “Métodos estadísticos en series temporales no lineales, con aplicación en energía eólica” cuyo objetivo fue desarrollar técnicas para mejorar la implementación de modelos de series temporales no lineales, con especial énfasis en la predicción de potencia eólica. Llegando a la conclusión que el modelo SETARTVCD Presupuesto es una buena alternativa para predecir la producción eólica.



(Villazón Bustillos, Rubio Arias, & Ortega Gutierrez, 2015) En la presente investigación “Análisis en series de tiempo para el pronóstico de sequía en la región noroeste del estado de Chihuahua” Concluye que se empleó el modelo ARIMA con estacionalidad de nueve años, se simuló la precipitación del 2000 al 2012 y se pronosticó del 2013 al 2024. Se concluye que entre el 2018 y 2019, ocurrirá una sequía de moderada a severa en la región noroeste del estado de Chihuahua.

2.1.2. A Nivel Nacional

(Huillca Vargas, 2017) El desarrollo del presente trabajo de investigación; cuyo objetivo principal es modelar el comportamiento de los ingresos diarios en nuevos soles y la cantidad de bicicletas que ingresan a diario al taller de bicicletas SAN LUIS utilizando los modelos ARIMA. En el segundo capítulo se presenta los conceptos teóricos de Series de Tiempo, los modelos clásicos de descomposición de una serie de tiempo, los modelos AR, MA, ARMA, ARIMA, SARMA, SARIMA, así como la metodología de ajuste de estos modelos, que consiste en 4 etapas que son: identificación, estimación, validación y predicción. Utilizando los paquetes estadístico SPSS versión 22, MINITAB versión 17 y EVIEWS versión 9.

(Modesto Arias, 2017) “Comportamiento de la demanda de depósitos a plazo fijo en moneda nacional en las cajas municipales periodo 2004 - 2016” En su trabajo de investigación. Se empleó método econométrico tradicional y serie de tiempo con interés de contrastar los resultados. Se evaluó la capacidad de predictiva mediante criterios de información, evaluación de residuos y error de pronóstico. se concluye que la mejor alternativa es plantear un modelo ARIMA, ya que de considerar otra variable explicativa se obtiene una regresión espuria. El



modelo elegido es un modelo SARIMA, debido a que la serie presentó la estacionalidad de cada seis meses. Los coeficientes resultaron estadísticamente significativos y haciendo una evaluación de los residuos estos fueron aleatorios por lo que se concluye que el modelo elegido es apropiado para realizar pronóstico.

(Rugel Barreto, 2011) “Modelos de Pronostico de Estadísticas Vitales en la Provincia de Huaraz a Nivel de Centros Poblados 2011” El objetivo de la presente investigación fue encontrar los modelos de pronósticos que mejor se ajustan a los datos. Se aplicó Metodologías de Box-Jenkins, Modelos de Regresión Múltiple, Modelos de Ajuste Estacional con Censur XII y Ajuste de datos a la Función Logística. El modelo de pronóstico que mejor se ajusta a los datos de nacimientos, que posee una serie histórica estable a través del tiempo, es con el Ajuste a una Función Logística; el modelo de pronóstico que mejor se ajusta a los datos de las defunciones, que presenta un patrón ciclo en su serie histórica, es mediante los modelos de Ajuste Estacional con Censur XII y el modelo de pronóstico que mejor se ajusta a los datos de matrimonios, que presenta un patrón estacional e irregular en su serie histórica, es a través de los Modelos de Ajuste Estacional con Censur XII

2.1.3. A nivel Local

(Melo Mayta, 2016) “Modelo de predicción mensual de Mortalidad General Intrahospitalaria en el Hospital Regional Manuel Núñez Butron – Puno, 2008 – 2016 - I” Usando la metodología usada fue Box – Jenkins para el modelamiento de la serie del número de mortalidad general intrahospitalaria. El resultado de la estimación del mejor modelo univariante para la predicción de la



serie original de mortalidad intrahospitalaria, en el Hospital Regional Manuel Núñez Butron – Puno es un modelo SARIMA (2, 1, 0) (0, 1, 1) (12)

(Bedoya Luza, 2018) “Modelamiento univariado del número de defunciones infantiles producidas por infecciones respiratorias agudas, a través de la metodología box-jenkins, puno 2008-2016” El resultado obtenido de esta investigación, el mejor modelo univariante para pronosticar la serie del número de defunciones infantiles causadas por las infecciones respiratorias agudas (I.R.AS), en niños menores de 05 años de la Dirección Regional de Salud de Puno es el modelo ARIMA (0, 1,1,)x(1, 1,0)

2.2. DEFINICIÓN DE TÉRMINOS BÁSICOS

2.2.1. Maquinarias Agrícolas

(Gilardi, 1977) Indica que, es una maquina automotriz compuesta de muchas partes y mecanismos que están sometidos a continuo desgaste y es utilizada para operar otras máquinas e implementos agrícolas.

(Frank, 1977) Señala que, una maquinaria agrícola es cualquier dispositivo que consiste en un cuerpo o sistema de cuerpos apoyados en puntos o ejes destinados a suministrar, transformar, transferir, aplicar o regular energía para realizar trabajos agropecuarios o coadyuvar a ellos.

(Murillo, 1987) Sostiene que, un tractor es una maquina dotada de motor para su desplazamiento, puede ser usada para tirar de equipos como arados, restras, etc. O para accionar mecanismos de máquinas estacionarias trilladoras, etc. Y de máquinas móviles como empacadores y segadoras.



2.2.2. Municipio

Es la entidad administrativa que se puede agrupar una sola localidad o varias y que puede ser un hacer referencia a una ciudad, un pueblo o una aldea.

Un municipio es, al mismo tiempo, una división territorial y una entidad administrativo de nivel local, constituida por territorio, población y poderes públicos. En este sentido, es un ente organizativo dentro del Estado que goza autonomía gubernamental y administrativa, cuya función es gestionar los intereses de una comunidad y dar solución a sus problemas. (Coelho, 2018)

2.2.3. Modelo

Según (Kishurim, Tecnice, Tecnimat, Griduc, & Gidsaw, 2013) el modelo de un fenómeno un conjunto de representaciones formales, que incorpora sin ambigüedad los conocimientos adquiridos mediante todas las fuentes pertinentes sobre el fenómeno de interés para el estudio. De esta forma, el modelo consiste en la especificación formal de los elementos de un sistema, las relaciones entre mismos y los parámetros que permiten contextualizar el desempeño del sistema de acuerdo con las características del entorno y las relaciones del sistema del mundo real con el mismo.

2.2.4. Predicción

Las ideas anteriores pueden servir de ayuda en el importante problema de la predicción en las series de tiempo. Sin embargo, se debe tener en cuenta que un tratamiento matemático de los datos no resuelve por si solo todos los problemas.

Justamente con el sentido común, experiencia, habilidad y buen juicio del investigador, tales análisis matemáticos pueden, no obstante, ser de valor para la predicción a largo y corto plazo. (Murray Ralph & Larry J, 2009)



2.3. MARCO TEÓRICO

2.3.1. Metodologías de diseño de investigación

(Valadez Silvas, 2019) Dentro de las partes del diseño de información se encuentra el análisis y la reflexión de la teoría y metodología del diseño. Que lo que nos ayuda a organizar nuestras ideas para finalmente tomar la mejor que se adecue a lo que queremos. El diseño surge para resolver necesidades y problemas.

En el presente trabajo se revisaron las siguientes metodologías.

Metodologías de diseño

- **Metodología de Bruno Munari**

Características (Munari, 1985):

- Es un método lineal y presenta orden lógico
- Todo problema es susceptible de descomponerse en pequeños problemas o subproblemas.
- Su finalidad es conseguir un máximo resultado con mínimo esfuerzo.

- **Metodología de Hans Gugelot**

Características (mclaranarvaez, 2017):

- Presentación del problema
- Análisis del estado actual
- Definición del problema y las metas
- Creaciones alternativas
- Evaluación y selección
- Planificación de la producción metodológica.

- **Metodología de Gui Bonsiepe**

Características (Mendez Lima, 2018):

- Análisis funcional



- Análisis morfológico
- Sinéctica
- Síntesis formal
- Optimizar características de uso
- Visualización de ideas
- **Metodología de Christopher Alexander**

Características (Sarli, 2017):

- Reducción de tiempos
- Reducir esfuerzo de mantención
- Aumentar la eficiencia
- Aumentar la fiabilidad
- Proteger la inversión
- **Metodología de Hernández Sampieri Roberto**

Características del enfoque cuantitativo (Tesis de Investigadores, 2011):

- Las hipótesis se generan antes de recolectar y analizar los datos
- La recolección de los datos se fundamenta en la medición
- La recolección de los datos se fundamenta en la medición
- Debido a que los datos son productos de mediciones, se representan mediante números (cantidades) y se deben analizar a través de métodos estadísticos
- En el proceso se busca el máximo control para lograr que otras explicaciones posibles distintas a la propuesta del estudio (hipótesis) sean desechadas y se excluya la incertidumbre y minimice el error
- Los análisis cuantitativos se interpretan a la luz de las predicciones iniciales (hipótesis) y de estudios previos (teoría)

- La investigación cuantitativa debe ser lo más objetiva posibles

2.3.2. Series de tiempo

Una serie tiempo es una secuencia de observaciones, medidos en determinados momentos del tiempo, ordenados cronológicamente y, espaciados entre sí de manera uniforme, así los datos usualmente son dependientes entre sí. El principal objetivo de una serie de tiempo X_t , donde $t = 1, 2, \dots, n$ es su análisis para hacer pronóstico. (Villavicencio, 2010)

Clasificación de Movimientos de Series de Tiempo

- **Movimiento de Tendencia**

Se puede definir como un cambio a largo plazo que se produce en la relación al nivel medio, o el cambio a largo plazo de la media. La tendencia se identifica con un movimiento suave de la serie a largo plazo. (Villavicencio, 2010)

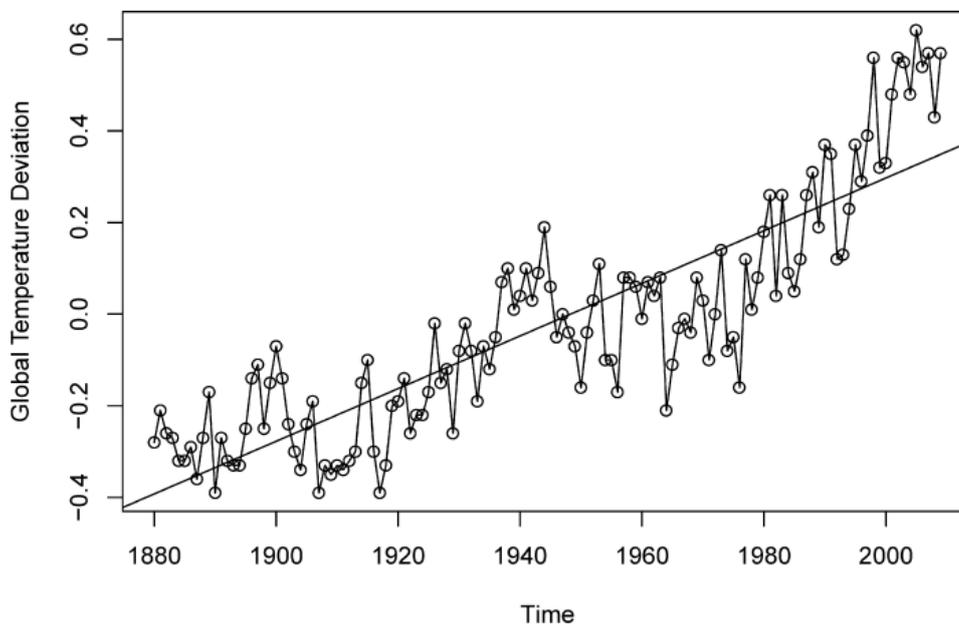


Figura 1 Tendencia de las desviaciones de temperatura global.

Fuente: (Robert H. & David S., September 2010, pág. 49)

- **Movimiento Cíclicos**

Se refieren a las oscilaciones que se presenta en la serie de datos que se repiten aproximadamente después de transcurrir un ciclo. Estos ciclos pueden ser o no periódicos. (Spiegel, 1983).

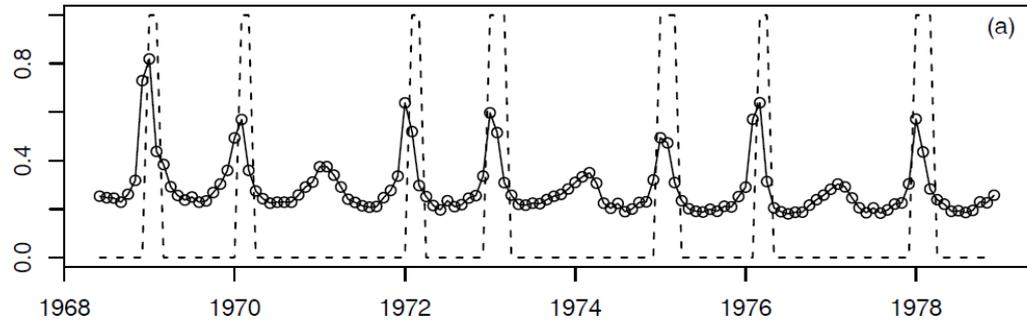


Figura 2 *Serie de muertes por neumonía mensual*

Fuente: (Robert H. & David S., September 2010, pág. 375)

- **Movimiento Estacional**

Son las series de tiempo que presentan un patrón permanente en lapsos de un año, mes, trimestre, etc. (Anderson, Sweeney, & Wilians, 2008)

Pasajeros en Líneas Aéreas

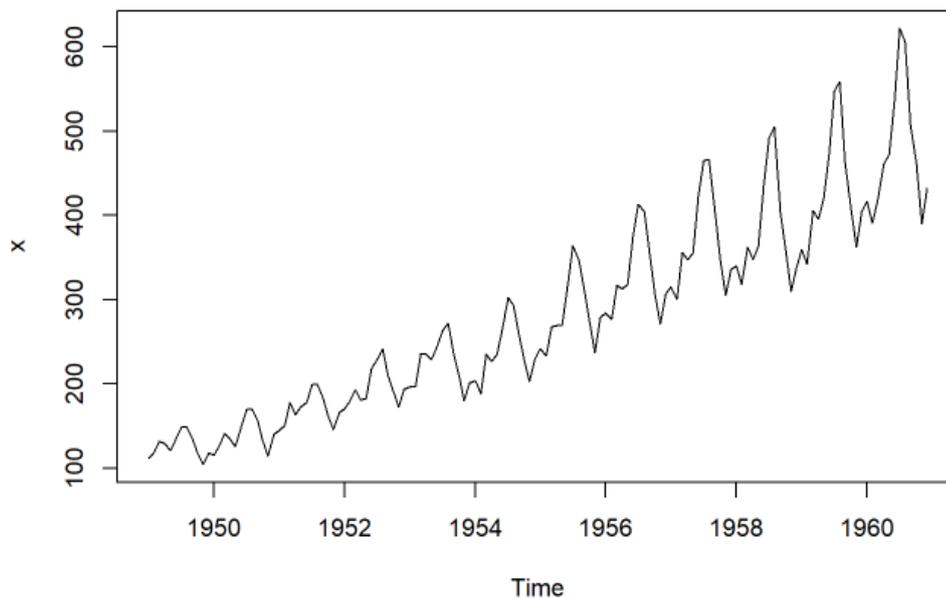


Figura 3 *Pasajeros de líneas aéreas internacionales para el periodo comprendido entre enero de 1949 y diciembre de 1962*

Fuente: (Parra, Estadística y Machine Learning con R, 25 de Enero de 2019)

- **Movimiento Irregulares**

Este componente irregular es producido por factores a corto plazo, imprevistos y no recurrentes que afectan a la serie de tiempo. Dado que este componente da cuenta de la variabilidad aleatoria en una serie de tiempo, es un componente impredecible. No es posible predecir su efecto sobre la serie de tiempo. (Anderson, Sweeney, & Wilians, 2008)

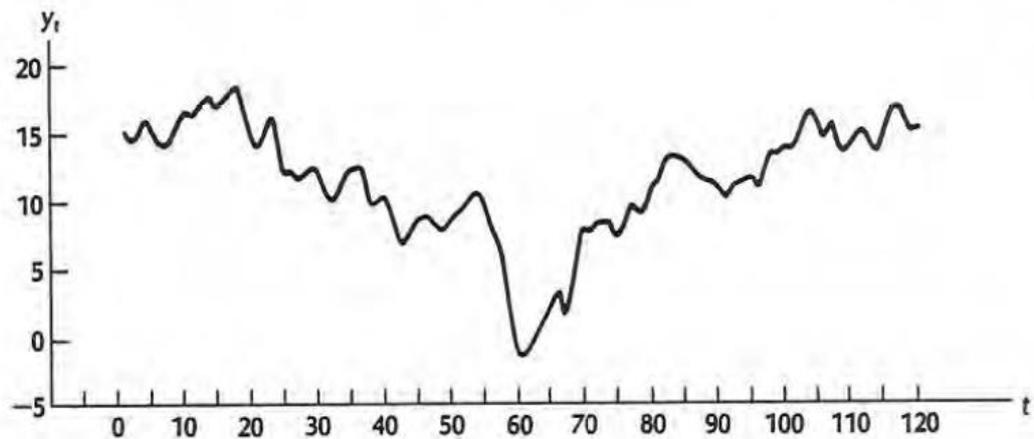


Figura 4 Ventas semanales por arriba de 100,000 rollos de toallas de papel absorbente.

Fuente: (Anderson, Sweeney, & Wilians, 2008)

2.3.3. Clasificación Descriptiva de la Series de Temporales

Pero, ¿qué es una serie temporal? Por definición, una serie temporal es una sucesión de observaciones de una variable realizadas a intervalos regulares de tiempo. (Parra, Estadística y Machine Learning con R, 2019)

El objetivo fundamental del estudio de las series temporales es el conocimiento del comportamiento de una variable a través del tiempo para, a partir de dicho conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales, poder realizar predicciones, es decir, determinar qué valor tomará la variable objeto de estudio en uno o más períodos de tiempo situados en

el futuro, mediante la aplicación de un determinado modelo calculado previamente. (Parra, Estadística y Machine Learning con R, 2019)

Según (Bruce L, Richard T, & Anne B, 2007) las series temporales pueden clasificarse en:

a) Estacionarias

Es cuando la serie de tiempo es estable a lo largo del tiempo, quiere decir que cuando la media y la varianza son constantes en el tiempo.

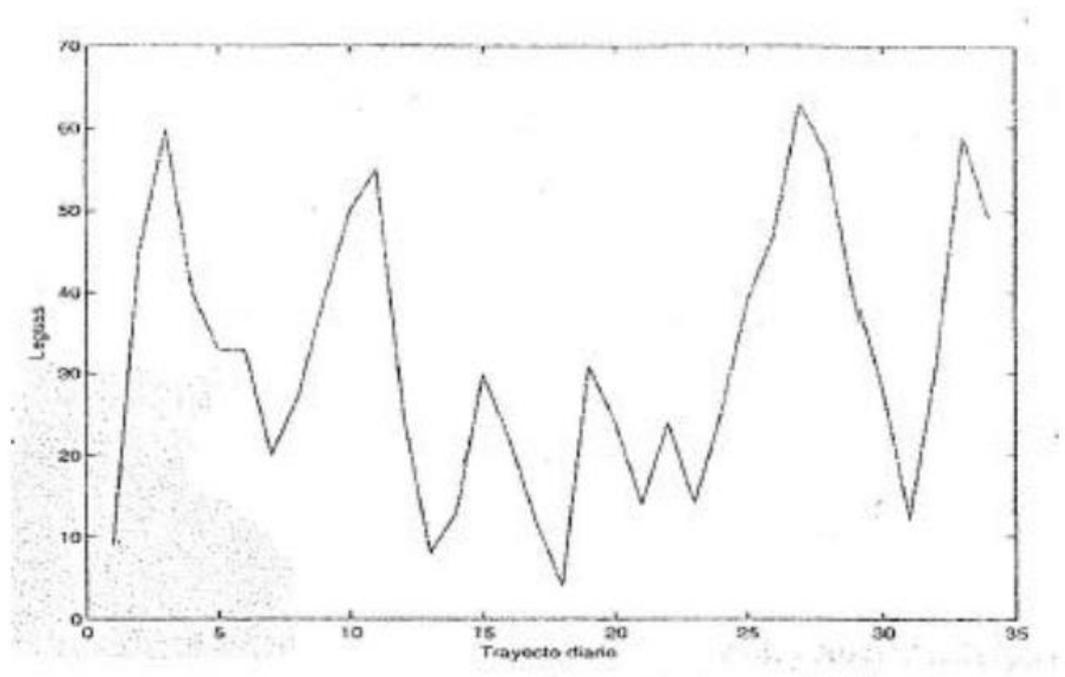


Figura 5 Leguas diarias recorridas por la flota de Cristóbal Colón en su primer viaje a América desde la isla de la Gomera a la de San Salvador.

Fuente: (Rivera, 2010, pág. 22)

b) No estacionarias

Una serie no estacionaria es cuando la tendencia y la variabilidad cambian en el tiempo. Los cambios en la media establecen una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante. (Pérez Porto, 2008)

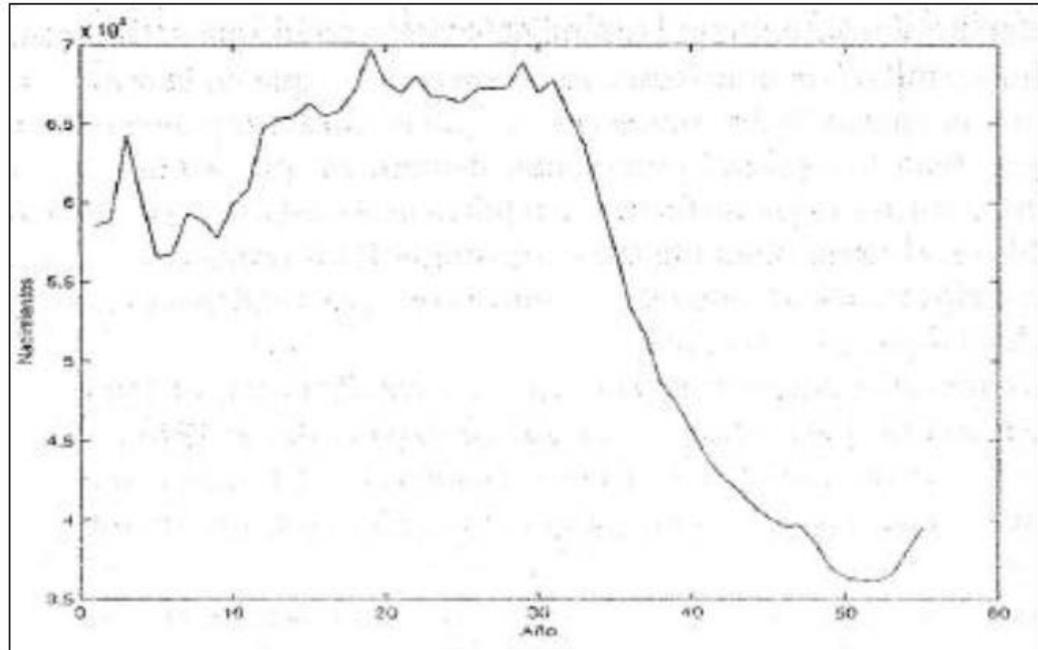


Figura 6 *Nacimientos anuales en España.*

Fuente: (Rivera, 2010, pág. 22)

2.3.4. Utilización de las Series de Tiempo

Según (Rios, 2008) hoy en día diversas organizaciones requieren conocer el comportamiento futuro de ciertos fenómenos con el fin de planificar, prevenir, quiere decir se utilizan para predecir lo que ocurrirá con una variable en el futuro a partir del comportamiento de esa variable en el pasado. En las organizaciones, instituciones, empresas, etc. Es de mucha utilidad en predicciones a corto y mediano plazo, por ejemplo, ver qué ocurrirá con la demanda de un cierto producto, las ventas a futuro, decisiones sobre inventario, insumo, entre otros.

Algunas de las áreas de aplicación de Series de Tiempo son:

- **Economía:** Precios de un artículo, tasas de desempleo, tasa de inflación, índice de precios, precio del dólar, precio del cobre, precios de acciones, ingreso nacional bruto, etc.
- **Meteorología:** Cantidad de agua caída, temperatura máxima diaria, Velocidad del viento (energía eólica), energía solar, etc.



- **Geofísica:** Series
- **Química:** Viscosidad de un proceso, temperatura de un proceso.
- **Demografía:** Cantidad de agua caída, temperatura máxima diaria, Velocidad del viento (energía eólica), energía solar, etc.
- **Medicina:** Electrocardiograma, electroencefalograma.
- **Marketing:** Series de demanda, gastos, utilidades, ventas, ofertas.
- **Telecomunicaciones:** Análisis de señales
- **Transporte:** Series de tráfico.

2.3.5. Modelo de Series de Tiempo

Según (Rios, 2008), señala que todos los modelos clásicos de series de tiempo presentados parten del supuesto que uno conoce una historia prolongada de la serie, sigue algún comportamiento estacionario (u obtenerlo al diferenciar la serie), y se tiene una sola serie de tiempo. En muchos casos, estos supuestos no se cumplen, por lo que los modelos presentados anteriormente fallan.

Los modelos de series de tiempo tienen un enfoque netamente predictivo y en ellos los pronósticos se elaborarán sólo con base al comportamiento pasado de la variable de interés.

Para que una serie temporal tenga un buen modelo de pronóstico se debe cumplir con las siguientes condiciones:

- El ruido blanco debe estar fluctuando por arriba y por debajo de la media cero. Esto garantizara que los ruidos blancos son independientes.
- En el gráfico para la ACF y para las PACF las barras de las correlaciones deben estar dentro de las bandas o en todo caso solo el 5% de ellas pueden estar superados por las bandas. Para probar esta condición se debe utilizar la prueba Box-Pierce Lung.

- El ruido blanco debe satisfacer una distribución normal.

2.3.6. Modelo Clásico de una Serie de Tiempo

Los cambios en los datos de la serie pueden estar afectados por factores que hacen que la serie resultante pueda tener uno de los cuatro patrones o movimientos subyacentes siguientes: El movimiento tendencial T_t , el movimiento estacional E_t , el movimiento cíclico C_t , el movimiento irregular I_t .

En la forma matemática, si cada dato de la serie está afectado por los cuatro patrones, el modelo que le corresponda podrá ser alguno de las dos formas siguientes según si la serie tenga o no una fuerte volatilidad en la varianza. (Rivera, 2010).

- **Modelo aditivo:** Si la varianza anual es estable para todos los años de observación.

$$y_t = T_t + E_t + C_t + I_t$$

- **Modelo Multiplicativo:** Si la varianza anual no es estable (volátil o creciente) para todos los años de observación.

$$y_t = T_t * E_t * C_t * I_t$$

- **Modelo Mixto:** Si la serie está afectada de forma multiplicativa por la estacionalidad y la tendencia y de forma aditiva por la irregularidad.

$$y_t = T_t * E_t + C_t * I_t$$

Donde:

y_t : Componente tendencia

E_t : Componente tendencia

C_t : Componente tendencia

I_t : Componente tendencia

2.3.7. Modelo Univariante

Una serie temporal univariante consiste en un conjunto de observaciones de una variable, quiere decir que se consideran todos aquellos que solamente tiene una sola variable observada en el tiempo Y . Si hay T observaciones, se denota por:

$$y_t, t \in \mathbb{Z} \quad Y_t, t = 1, \dots, T$$

Estos tipos de modelos se enuncian en forma polinomial. Entre las metodologías univariantes están algunas muy sencillas, así como el modelo autorregresivo de primer orden, el modelo de tendencia lineal o exponencial, entre otros. Las técnicas más rigurosas para la predicción univariante son las denominadas técnicas o modelos de Box – Jenkins, o más concretamente modelos ARIMA, pues las técnicas de Box – Jenkins constituyen un conjunto más amplio, dentro del cual los modelos ARIMA univariantes son solo una parte. (Ljung & Box, 1979)

2.3.8. Procesos Estocásticos

Desde un punto de vista intuitivo, un proceso estocástico se describe como una secuencia de datos que evolucionan en el tiempo. Las series temporales se definen como un caso articular de los procesos estocásticos. (Hernández, 2015)

- **Proceso Estocástico Estacionario**

Un proceso estocástico se dice que es estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende solamente de la distancia o rezago entre estos dos periodos de tiempo y no del tiempo en el cual se ha calculado la covarianza. (Villavicencio, 2010)

Sea X_t una serie de tiempo entonces con estas propiedades:

$$\text{Media } E(X_t) = E(X_{t+k}) = \mu$$

$$\text{Varianza } V(X_t) = V(X_{t+k}) = \sigma^2$$

$$\text{Covarianza } \gamma_k = E[(X_t - \mu)(X_{t+k} - \mu)]$$

Dende γ_k , la covarianza (o auto covarianza) al rezago k , es la covarianza entre los valores de X_t y X_{t+k} , que están separados k periodos.

En resumen, si una serie de tiempo es estacionaria, su media, su varianza y su auto covarianza (en diferentes rezagos) permanecen iguales sin importar el momento en el cual se midan; es decir, son invariantes respecto al tiempo.

2.3.9. Ruido blanco (“White noise”)

Un ruido blanco es un caso simple de los procesos estocásticos, donde los valores son independientes e idénticamente distribuidos a lo largo del tiempo con media cero e igual varianza, se denota por ε_t . (Villavicencio, 2010)

$$\varepsilon_t \sim N(0, \sigma^2) \text{ cov}(\varepsilon_{t_i}, \varepsilon_{t_j}) = 0 \quad \forall t_i \neq t_j$$

la gráfica muestra un ruido blanco con media cero y varianza constante e igual a uno

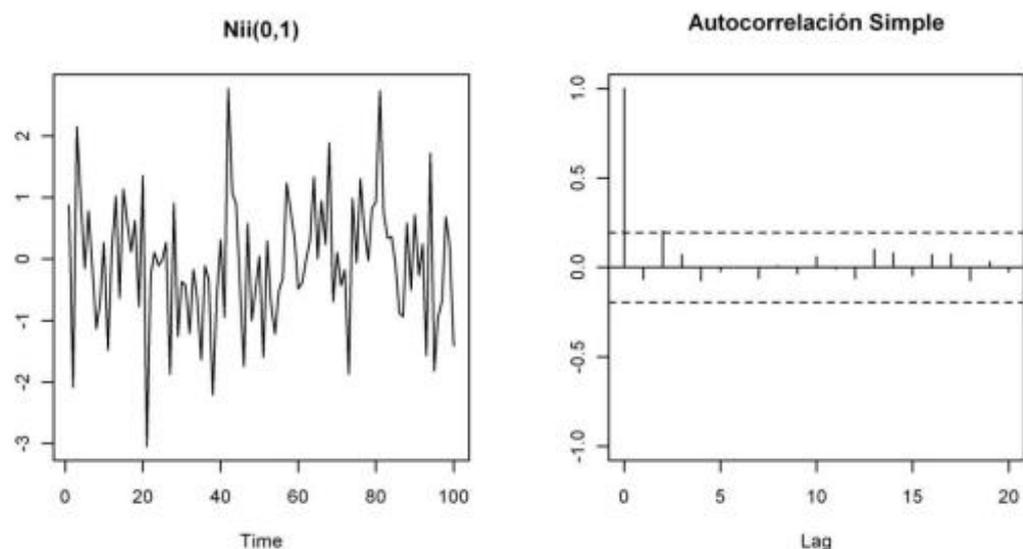


Figura 7 Grafico de serie de tiempo y autocorrelación simple.
Fuente: (Villavicencio, 2010)

- **Prueba de Ljung – Box**

(Peña, 2005) Esta prueba permite probar en forma conjunta de que todos los coeficientes de autocorrelación son simultáneamente iguales a cero, esto es que son independientes, está definida como:

$$LB = n(n + 2) \sum_{k=1}^m \left(\frac{\hat{\rho}_{(m)}^2}{n - k} \right) \sim \chi_{(m)}^2$$

Donde n tamaño de la muestra, m longitud del rezago.

H_0 : Las autocorrelaciones son independientes (es un ruido blanco $\rho_k = 0$).

H_a : Las autocorrelaciones no son independientes (no es un ruido blanco $\rho_k \neq 0$).

En una aplicación, si el Q calculada excede el valor Q crítico de la tabla ji cuadrada al nivel de significancia seleccionado, no se acepta hipótesis nula de que todos los “coeficientes” de autocorrelación son iguales a cero; por lo menos algunos de ellos deben ser diferentes de cero.

2.3.10. Autocorrelación

En ocasiones en una serie de tiempo acontece, que los valores que toma una variable en el tiempo no son independientes entre sí, si no que un valor determinado depende de los valores anteriores, existen dos formas de medir esta dependencia de las variables. (Garcia Bellido, González Such, & Jornet Meliá, 2010)

- **Función de autocorrelación (ACF)**

La autocorrelación mide la correlación entre dos variables separadas por k periodos.

$$\rho_j = \text{corr}(X_j, X_{j-k}) = \frac{\text{cov}(X_j, X_{j-k})}{\sqrt{V(X_j)}\sqrt{V(X_{j-k})}}$$

La función de autocorrelación simple tiene siguientes propiedades:

$$\rho_j = 1$$

$$-1 \leq \rho_j \leq 1$$

$$\text{Simetria } \rho_j = \rho_{-j}$$

- **Función de autocorrelación Parcial (PACF)**

La autocorrelación parcial mide la correlación entre dos variables separadas por k periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas.

$$\pi_j = \text{corr}(X_j, X_{j-k} / X_{j-1} X_{j-2} \dots X_{j-k+1})$$

$$\pi_j = \frac{\text{cov}(X_j, \hat{X}_j X_{j-k}, \hat{X}_{j-k})}{\sqrt{V(X_j, \hat{X}_j)}\sqrt{V(X_{j-k}, \hat{X}_{j-k})}}$$

2.3.11. Proceso Lineal Estacionarios

- **Proceso autorregresivos AR(p)**

(Bruce L, Richard T, & Anne B, 2007) Los modelos autorregresivos se basan en la idea de que el valor actual de la serie, puede explicarse en función de valores pasados $X_{t-1}, X_{t-2}, \dots, X_{t+p}$, donde determina el número de rezagos necesarios para pronosticar un valor actual.

El modelo autorregresivo de orden está dado por:

$$X_{t-1} = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t+p} + \varepsilon_t \dots (1)$$

Expresado en términos del operador de retardos

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

$$\varphi_p(L) X_t = \varepsilon_t$$

donde ε_t es un proceso de ruido blanco y $\varphi_1, \theta_1, \theta_2$ son los parámetros del modelo.

- **Modelo de medias móviles MA(q)**

Modelo “determinados por una fuente externa”. Estos modelos suponen linealidad, el valor actual de la serie, X_t , está influenciado por los valores de la fuente externa.

El modelo de promedio móviles de orden p está dado por:

$$X_t = \theta_0 - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_p\varepsilon_{t-p} + \varepsilon_t \dots (2)$$

Expresado en términos del polinomio operador de retardos se tiene:

$$X_t = (1 - \theta_1L - \theta_2L^2 - \dots - \theta_pL^p)\varepsilon_t$$

$$X_t = \theta_p(L)\varepsilon_t$$

donde ε_t es un proceso de ruido blanco, y $\mu, \theta_1, \theta_2, \dots, \theta_q$ son los parámetros del modelo.

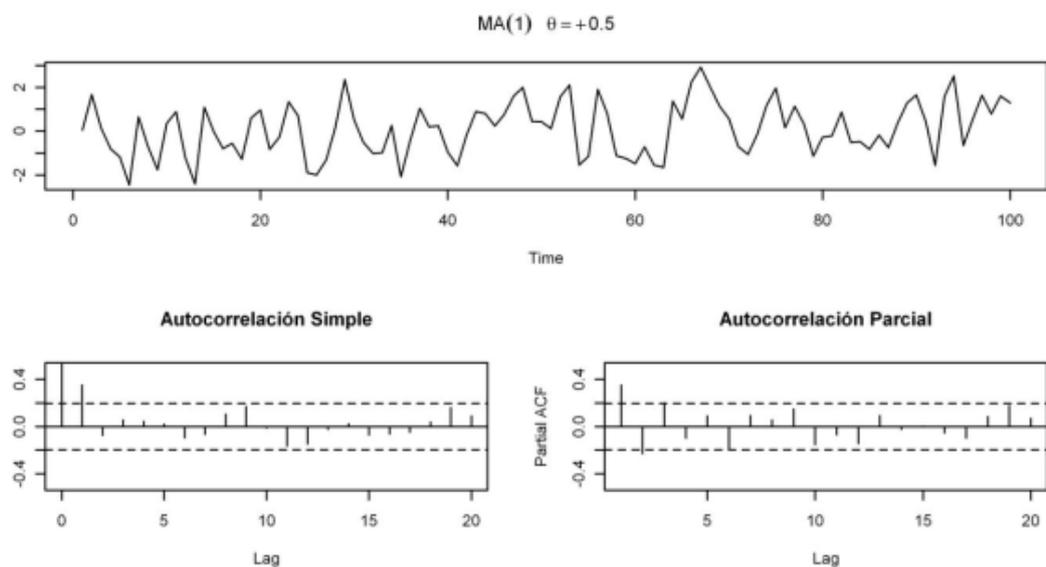


Figura 8 Ejemplo de autocorrelación simple y autocorrelación parcial.

Fuente: (Bruce L, Richard T, & Anne B, 2007)

La figura muestra un proceso de medias móviles de orden uno, esto es $X_t = \varepsilon_t + 0.5\varepsilon_{t-1}$ con sus respectivas graficas de autocorrelación simple y parcial.

- **Proceso Autorregresivo de Medias Móviles ARMA(p,q)**

Es muy probable que una serie de tiempo, X_t , tenga características de AR y de MA a la vez y, por consiguiente, sea ARMA. Así, X_t sigue un proceso ARMA (p, q), en este proceso habrá términos autorregresivos y términos de media móvil.

$$X_t = c + \underbrace{\varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p}}_{AR(p)} + \underbrace{\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}}_{MA(q)} + \varepsilon_t$$

donde es un proceso de ruido blanco, y son los parámetros del modelo.

Para un proceso ARMA (p, q) una condición de estacionariedad es la misma que para un proceso AR(p), del mismo modo una condición de invertibilidad es la misma que para el proceso MA(q).

El modelo ARMA (p, q) se puede escribir en términos del operador de retardos como sigue:

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

Donde:

$\varphi_p(L)$: es el polinomio autorregresivo.

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

La figura muestra la serie del cambio porcentual mes a mes de la serie desestacionalizada del consumo de energía eléctrica en Puerto Rico con sus respectivas graficas de autocorrelación simple y parcial.

2.3.12. Procesos Lineales no Estacionarios

- **Proceso Autorregresivo integrado y de Medias Móvil ARIMA (p, d, q)**

Los modelos de series de tiempo analizados hasta ahora se basan en el supuesto de estacionariedad, esto es, la media y la varianza para una serie de tiempo son constantes en el tiempo y la covarianza es invariante en el tiempo. Pero se sabe que muchas series de tiempo y en especial las series económicas no son estacionarias, porque pueden ir cambiando de nivel en el tiempo o sencillamente la varianza no es constante en el tiempo, a este tipo de proceso se les considera procesos integrados. Por consiguiente, se debe diferenciar una serie de tiempo veces para hacerla estacionaria y luego aplicarla a esta serie diferenciada un modelo ARMA (p,q), se dice que la serie original es ARIMA(p,q,d), es decir, una serie de tiempo autorregresiva integrada de media móvil. Donde p denota el número de términos autorregresivos, el número de veces que la serie debe ser diferenciada para hacerla estacionaria y q el número de términos de la media móvil invertible. (Wilfredo Toledo, 2013)

Su expresión algebraica es:

$$X_t^d = c + \underbrace{\varphi_1 X_{t-1}^d + \varphi_2 X_{t-2}^d + \dots + \varphi_p X_{t-p}^d}_{AR(p)} + \underbrace{\theta_1 \varepsilon_{t-1}^d + \theta_2 \varepsilon_{t-2}^d + \dots + \theta_q \varepsilon_{t-q}^d}_{MA(q)} + \varepsilon_t^d$$

Expresado en forma del polinomio operador de retardos el modelo es:

$$\varphi(L)(1-L)^d X_t = c + \theta(L)\varepsilon_t$$

donde X_t^d es al serie de las diferencia de orden d, ε_t^d , es un proceso de ruido blanco,

$c, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ son los parámetros del modelo

2.3.13. Transformación de la Seria de Tiempo

El trabajo de exploración consiste en construir la gráfica de la serie de tiempo para visualizar el patrón de comportamiento y adecuarla para el uso de la metodología Box – Jenkins mediante los patrones Raíz unitaria, tendencia estocástica, tendencia deterministas y estacionariedad. (Wilfredo Toledo, 2013)

Cuando la serie no lo es, es preciso transformarla. Este proceso es muy sencillo. Para eliminar la tendencia se toman una o varias diferencias en la serie. Una serie se diferencia restando a cada observación la observación anterior: (Wilfredo Toledo, 2013)

$$w_t = z_t - z_{t-1}$$

Es decir, la serie w_t se obtiene restando de cada observación de z_t la observación anterior. Evidentemente la serie diferenciada w tiene una observación menos que la serie original z , ya que la primera observación se pierde. Si el gráfico de una serie muestra tendencia, se diferencia la serie y se comprueba si ha perdido la tendencia. En caso de que no la haya perdido se diferencia una segunda vez. Es muy raro necesitar más de una o dos diferencias para eliminar la tendencia de una serie.

Normalmente se utiliza la nomenclatura $\nabla z_t = w_t$ para representar la serie con una diferencia. (Wilfredo Toledo, 2013)

2.3.14. Metodología Box - Jenkins

Protocolo para identificación de los modelos ARIMA, teniendo en consideración que la serie se ajuste a una distribución normal.

Etapas de la metodología box – Jenkins

1. Identificación

Representar gráficamente la serie, además de su función de autocorrelación

simple (ACF) y función de autocorrelación parcial (PACF). La gráfica de la serie nos indica si la serie es estacionaria o no. Según los motivos por los que la serie no es estacionaria, tendremos que aplicar los siguientes procedimientos hasta hacerla estacionaria. (Villavicencio, 2010)

- Si tiene tendencia: Tomaremos diferencias regulares hasta que desaparezca. Normalmente el orden de la diferencia es 1, y raramente será mayor a 3.
- Si la serie tiene estacionalidad: Tomaremos diferencias estacionales hasta que desaparezca el patrón estacional. En la práctica es muy raro tener que aplicar más de una diferencia estacional.
- Si es heterocedástica, es decir, no tiene varianza constante, habrá que transformar la serie. Con tomar el logaritmo en muchos casos es suficiente, aunque existen algunas transformaciones más sofisticadas, como las de Box-Cox.

Una vez que el gráfico de la nueva serie (transformación de la original) indica que es estacionaria, podemos intentar deducir la estructura de la serie (¡no la de la serie original!) observando su ACF y PACF. (Villavicencio, 2010)

2. Estimación y verificación

Observando las dos gráficas del ACF y PACF de la serie transformada podemos hacernos una idea del modelo que describe nuestra serie, o al menos de cuáles son los primeros candidatos que debemos probar. Para comprobar analíticamente (no visualmente) un modelo frecuentemente se ajusta varios modelos candidatos ARIMA(p,d,q) y escogeremos como un buen modelo aquel que tenga los residuales semejantes al de un ruido



blanco, además que tenga los valores del AIC (Criterio de Información de Akaike) y BIC (Criterio de Información Bayesiana) menores con relación al resto de los modelos candidatos. (Villavicencio, 2010)

3. Predicción

Una de las razones de la popularidad del proceso de construcción de modelos es su éxito en la predicción. Los modelos son buenos para realizar predicciones a corto plazo. (Villavicencio, 2010)

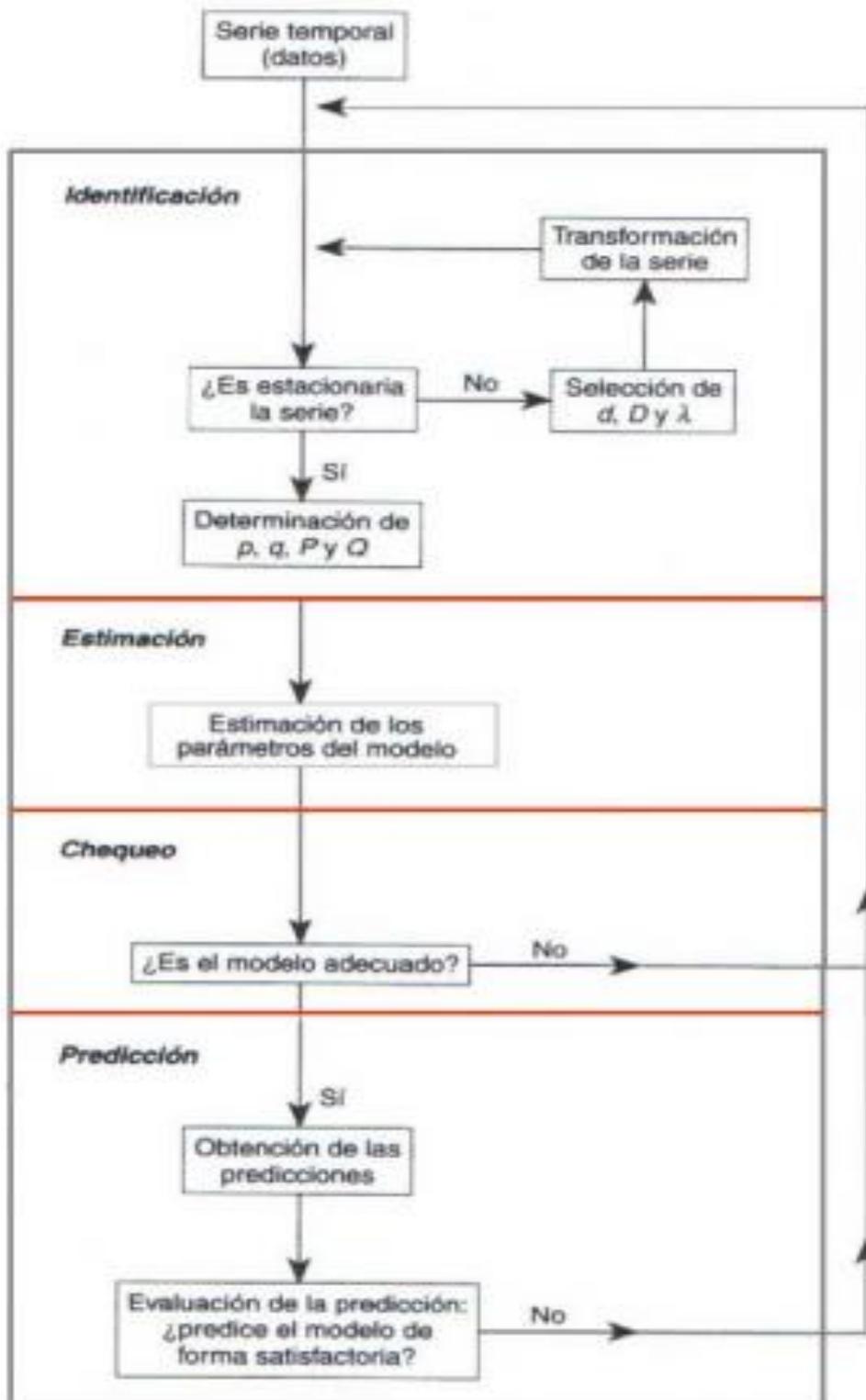


Figura 9 Etapas de la metodología box – Jenkins

Fuente: (Fernandez, 2011)

2.3.15. Prueba de Dickey – Fuller Ampliado (ADF)

Sin duda alguna, el test más habitual a la hora de determinar la estacionariedad de una serie temporal, consiste en la aplicación del conocido como test de Dickey – Fuller (Test DF) o Dickey – Fuller ampliado (Test ADF). (Robert H. & David S., September 2010)

Este es un contraste de “no estacionariedad” ya que la hipótesis nula es precisamente la presencia de una raíz unitaria en el proceso generador de datos de la serie analizada.

Como modelo de partida para el análisis de una determinada serie X_t el de un proceso generador de datos de la serie analizada.

Como modelo de partida el análisis de una determinada serie X_t , el de un proceso estacionario autorregresivo de orden uno:

$$X_t = \varphi_t X_{t-1} + \varepsilon_t$$

Como hipótesis nula H_0 , el modelo alternativo de un paseo aleatorio no estacionario del tipo 1.

(Robert H. & David S., September 2010) El test de Dickey – Fuller Ampliado (DFA) contrasta la presencia de una raíz unitaria en una serie que sigue un proceso AR(p), deberá aplicarse el procedimiento expuesto para el caso simple AR (1), pero suponiendo ahora el modelo:

$$\Delta X_t = a_0 + \rho X_{t-1} + \sum B_t \Delta X_{t-i+1} + \varepsilon_t$$

H_0 : > 0.5 raíz unitaria (proceso no estacionario de los residuales)

H_1 : < 0.5 no es de raíz unitaria (proceso estacionario de los residuales)

2.3.16. Diagrama de caja

El diagrama de caja son una representación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría.

Para su realización se representan los tres cuartiles y los valores mínimos y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente.

(Hernández, 2015)

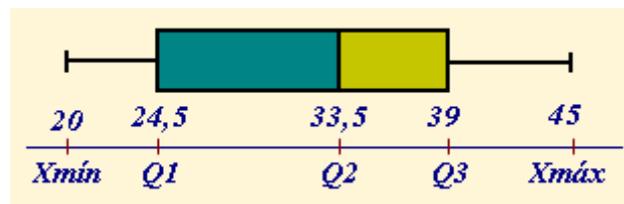


Figura 10 Explicación de diagrama de cajas.

Fuente: (Hernández, 2015)

El bigote de la izquierda representa al colectivo de edades (X_{\min} , Q1)

La primera parte de la caja a (Q1, Q2),

La segunda parte de la caja a (Q2, Q3)

El bigote de la derecha viene dado por (Q3, X_{\max}).

2.3.17. Prueba chi cuadrado de Pearson

(Plackett, 1983) La prueba X^2 de Pearson se considera una prueba no paramétrica que mide la discrepancia entre una distribución observada y otra teórica (bondad de ajuste), indicando en qué medida as diferencias existentes entre ambas, de haberlas, se deben al azar en el contraste de hipótesis. También se utiliza para probar a independencia de dos variables entre sí, mediante la presentación de los datos en tabas de contingencia.

La fórmula que da el estadístico es la siguiente:

$$X^2 = \sum_i \frac{(\text{observada}_i - \text{teórica}_i)^2}{\text{teórica}_i}$$

Cuanto mayor sea el valor de X^2 , menor verosímil es que la hipótesis nula (que asume la igualdad entre ambas distribuciones) sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de chi – cuadrado, más ajustadas están ambas distribuciones.

Los grados de libertad gl vienen dados por:

$$gl = (r - 1) (k - 1)$$

Donde r es el número de filas y K el de columnas.

Criterio de decisión

No se rechaza H_0 cuando $x^2 < x^2_{\alpha}((r - 1)(k - 1))$ En caso contrario si se rechaza.

Donde t representa el valor proporcionado por las tablas, según el nivel de significación estadística elegido.

2.3.18. Medición del error en pronósticos de series

(Betancourt, 2016) Define error de medición como la diferencia entre un valor que se mide y un valor verdadero. Si esto lo transportamos al ámbito empresarial, en nuestros pronósticos de demanda, y en el sentido más general, podemos definir error de pronóstico como la comparación entre el valor pronosticado y el valor real.

- **Error Relativo**

(Betancourt, 2016) El error relativo, porcentualiza la diferencia de lo pronosticado sobre lo ocurrido en la realidad, es decir, en qué porcentaje relativo a lo real fue en lo que se equivocó el pronóstico:

$$Error\ Relativo = \sum_{i=1}^N |P_i - R_i| / \sum_{i=1}^N |R_i|$$

N : Numero de datos a pronosticar

P_i : Valor pronosticado en el tiempo i

R_i : Valor real ocurrido en el tiempo i

- **Error Cuadrático Medio (ECM)**

(Betancourt, 2016) El ECM es una medida de dispersión del error de pronóstico, sin embargo, esta medida máxima el error al elevar al cuadrado, castigando aquellos periodos donde la diferencia fue más alta a comparación de otros. En consecuencia, se recomienda el uso del ECM para periodos con desviaciones pequeñas.

$$ECM = \frac{1}{N} \sum_{i=1}^N (P_i - R_i)^2$$

N : Numero e datos a pronosticar

P_i : Valor pronosticado en el tiempo i

R_i : Valor real ocurrido en el tiempo i

- **Porcentaje de fallos hacia arriba**

(Betancourt, 2016) Este porcentaje indica en cuantos, de los datos pronosticados, el valor real fue mayor o igual al pronóstico del modelo:

$$\text{porcentaje de fallos hacia arriba} = \sum_{I=1}^E 1/N$$

N : Numero de pronósticos

E : Numero datos en los que el real sobre paso al pronostico

- **Porcentaje de fallos hacia abajo**

(Betancourt, 2016) Este porcentaje indica en cuántos, de los datos pronosticados, el valor real fue menor al pronóstico del modelo:

$$\text{porcentaje de fallos hacia abajo} = 1 - \sum_{I=1}^E 1/N$$



N : Numero de pronósticos

E : Numero datos en los que el real sobrepaso al pronostico

- **Porcentaje del total en fallos hacia arriba**

(Betancourt, 2016) Este índice indica el porcentaje de error del total que corresponde a la diferencia entre lo real y el pronóstico, pero solo en los días en los que lo real sobrepaso al pronóstico:

$$\text{Porcentaje del total en fallos hacia arriba} = \frac{\sum_{i=1}^E (R_i - P_i)}{\sum_{i=1}^E (R_i)}$$

P_i : Total, pronosticado en el tiempo i

R_i : Total, real en el tiempo i

E : Numero datos en los que el real sobrepaso al pronostico

CAPITULO III

MATERIALES Y METODOS

En este capítulo de la tesis describimos la localización, operacionalización de variables, materiales utilizado para la investigación, métodos y técnicas que se emplearon para cada uno de los objetivos propuestos anteriormente.

3.1. LOCALIZACIÓN

El presente trabajo de investigación se realizó en la Municipalidad Distrital de Caracoto, tiene una superficie territorial de $285,87 \text{ Km}^2$. este Distrito se encuentra situado al norte con el Distrito de Juliaca, al este con el Distrito de Huatta y Coata Provincia de Puno, al oeste con el Distrito de Cabana San Román y en la parte sur con la provincia puno. En la actualidad el Distrito de Caracoto cuenta con 53 centros poblados entre comunidades y parcialidades. Según INEI, cuenta con 6058 habitantes.

3.2. OPERACIONALIZACIÓN DE VARIABLES

Por ser una variable directa en el siguiente cuadro se detalla:

Tabla 1 Operacionalización de variables.

VARIABLE	INDICADOR	NIVEL DE MEDICIÓN	INSTRUMENTO
VARIABLE DEPENDIENTE: Horas mensuales de trabajo de las maquinarias agrícolas de la Municipalidad de Caracoto. 2010 al 2019.	Expresado en horas	Horas	Datos registrados por la Municipalidad Distrital de Caracoto.
VARIABLE INDEPENDIENTE: Horas mensuales de trabajo de las maquinarias agrícolas de la Municipalidad de Caracoto. 2010 al 2019.	Periodo de registro de datos por mes	Meses	Datos registrados por la Municipalidad Distrital de Caracoto.



Fuente: Elaboración propia

3.3. MATERIALES

3.3.1. Software

Para el procesamiento de los datos se usó los siguientes softwares estadísticos:

- El lenguaje de programación R versión 3.6.2 (2019-12-12).

R es un software libre y viene sin GARANTIA ALGUNA.

Usted puede redistribuirlo bajo ciertas circunstancias.

Escriba 'license()' o 'licence()' para detalles de distribución.

- La interfaz de desarrollo RStudio Desktop 1.3.1093.

RStudio es un conjunto de herramientas integradas diseñadas para ayudarlo a ser más productivo con R. Incluye una consola, un editor de resaltado de sintaxis que admite la ejecución directa de código y una variedad de herramientas sólidas para trazar, ver el historial, depurar y administrar su espacio de trabajo:

<https://rstudio.com/products/rstudio/download/>.

- Excel 2019.

3.3.2. Hardware

Para el análisis de los datos del presente trabajo de investigación, se usó una Laptop con las siguientes características:

- Procesador AMD Ryzen 7 3700U
- RAM 12.00 GB



3.4. TIPO Y DISEÑO DE INVESTIGACION

3.4.1. Tipo de Investigación

(Hernández Sampieri, Fernández Collado, & baptista Lucio, 1999) Los estudios descriptivos tienen el propósito del investigador es describir situaciones y eventos. Esto es, decir como es y cómo se manifiesta determinado fenómeno. los estudios descriptivos buscan especificar las propiedades importantes de personas, grupos, comunidades o cualquier otro fenómeno que sea sometido a análisis.

El presente trabajo de investigación es de enfoque cuantitativo con diseño descriptivo simple según (Hernández Sampieri, Fernández Collado, & baptista Lucio, 1999) se realizó con datos cuantitativos por que los datos mostrados con una variable discreta, el cual se estudiaran las horas acumuladas de trabajo en meses obtenidas por la Municipalidad Distrital de Caracoto, para determinar el comportamiento con el fin de responder al planteamiento del problema, y analizar con seguridad la hipótesis formulada. El alcance del presente trabajo es describir la tendencia de los datos mensuales de los 10 últimos años.

3.4.2. Diseño de la investigación

Investigación no experimental según (Hernández Sampieri, Fernández Collado, & baptista Lucio, 1999) nos dice que el interés del investigador es analizar cambios a través del tiempo en determinadas variables o las relaciones entre estas. Entonces se dispone de los diseños longitudinales, los cuales recolectan datos a través del tiempo en puntos o periodos especificados, para hacer interferencia respecto al cambio, sus determinantes y consecuencias.



El diseño a utilizar será “no experimental” – Longitudinal de tendencia, porque se observó el fenómeno tal como fueron registrados en los documentos mensuales durante los 10 últimos años, sus determinantes y consecuencias, los datos mensuales son las horas de trabajo acumuladas.

3.5. POBLACION Y MUESTRA DE INVESTIGACION

3.5.1. Población

La población para el presente estudio fue constituida por horas totales de trabajo acumuladas durante el mes de los últimos 10 años, que fueron registrados por la Municipalidad Distrital de Caracoto durante el periodo del 2010 al 2019.

3.5.2. Muestra

La muestra es la misma de la población, constituida por horas de trabajo acumuladas durante el mes de los últimos 10 años, que fueron registrados por la Municipalidad Distrital de Caracoto durante el periodo 2010 - 2019.

3.6. TÉCNICA E INSTRUMENTO DE RECOLECCIÓN DE DATOS

Con el fin de asegurar la efectividad y originalidad de los datos, estos datos fueron recopilados directamente de los registros de horas de trabajo de las maquinarias agrícolas de la Municipalidad Distrital de Caracoto.

3.7. METODO DE RECOLECCIÓN DE DATOS

El procedimiento de recolección de datos se realizó de la siguiente forma:

1. Se solicitó la autorización de la Municipalidad Distrital de Caracoto.



2. Con la finalidad de mantener la confianza y originalidad de los datos, estos datos fueron recopilados directamente de la oficina a cargo de las maquinarias agrícolas.
3. La información fue obtenida en horas mensuales de trabajo de las maquinarias.

3.8. METODOLOGIA DE ANALISIS DE DATOS

Para el presente trabajo de investigación se analizó los datos mediante los siguientes gráficos:

- Representación de grafica de la serie de tiempo.
- Descomposición de datos
- Diagrama de estacionalidad
- Diagrama de cajas
- Histograma
- Normalidad
- Test de Pearson

Luego del análisis se trasformo de la serie de datos mediante la aplicación de diferencias de la serie de datos, hasta cumplir con los requisitos necesarios para la aplicación de la metodología de Box – Jenkins.

3.9. METODOS DE TRATAMIENTO DE DATOS

Para el presente trabajo de investigación se hará uso de la metodología de Box – Jenkins.

3.9.1. Metodología de Box - Jenkins

La metodología de Box – Jenkins es uno de los métodos predictivos que se fundamenta en la estimación de los parámetros por medio de procesos iterativos, consta de cuatro fases, los cuales son:

Paso 1. Identificación del modelo

Para determinar si nuestra serie es estacionaria analizaremos primeramente el grafico de la serie en un ploteo de dispersión. Si la serie no es estacionaria se puede convertir a una serie estacionaria mediante el método de diferenciación, para lo cual especificaremos el grado de diferenciación y el algoritmo de BOX – JENKINS que nos ayudaran a convertir los datos en una serie estacionaria y realiza los cálculos subsecuentes utilizando los datos ya convertidos.

El orden de la integración define el parámetro “d” del modelo ARIMA. En la práctica es suficiente tomar una diferencia (d=1), o dos diferencias (d=2) para obtener una serie estacionaria en media, a partir de la tercera diferencia la varianza se deforma, es decir la varianza crece.

Un aspecto importante en la modelación ARIMA de una serie de tiempo simple es el número de veces que necesita una diferencia antes de fijar el modelo. Para esto se utilizaremos la prueba de Dickey Fuller o de raíces unitarias en la cual la hipótesis a verificar es:

TEST DE DICKEY – FULLER

a. Formulación de la hipótesis Estadística

$H_0: \rho = 1$ La serie tiene raíz unitaria (serie no estacionario)

$H_1: \rho < 1$ La serie no tiene raíz unitaria (serie estacionario)

b. Regla de Decisión

- Aceptamos H_0 si el valor p – valor es igual a 1 que el nivel de significancia elegido (0.05)
- Rechazamos H_0 si el valor p – valor es menor a 1 que el nivel de significancia elegido (0.05)

c. Estadístico de prueba

Para ver la estacionariedad de la serie de tiempo se utilizó la prueba de Dickey - Fuller

Paso 2. Estimación y prueba de adecuación

Para la estimación del modelo es necesario la observación visual del correlograma (*f.a.c.m.* y *f.a.c.p.*) de la variable diferencia y en función del comportamiento de los picos se sugieren los modelos y se determinan los valores (p , d y q)

Una vez obtenida una serie estacionaria con un valor para diferencia “ d ”, se identificó la forma del modelo a utilizar encontrado los valores apropiados de p y q , mediante la comparación de los coeficientes de autocorrelación parcial de los datos.

calibración del modelo con los datos

Aplicamos *Auto - arima* a la serie con una diferencia y graficamos la autocorrelación simple y autocorrelación parcial, del mejor modelo recomendado por *Auto - arima*.

Comparamos los errores de los modelos obtenidos mediante tres formas, de las cuales obtendremos el mejor modelo con menor error.

Habiendo identificado los valores apropiados de p , d y q , la siguiente etapa es estimar los parámetros. Algunas veces, este cálculo puede hacerse mediante

mínimos cuadrados simples, pero otras se tendrán que recurrir a métodos de estimación no lineales (en parámetros). Puesto que esta labor es llevada a cabo a través de rutinas de diversos paquetes estadísticos, en la práctica no es preciso preocuparse por los desarrollos matemáticos de la estimación.

Paso 3. Validación de diagnóstico

Después de seleccionar un modelo *ARIMA* o *SARIMA* particular y de estimar sus parámetros, se trata de ver si el modelo seleccionado ajusta los datos en forma razonable buena, ya que es posible que exista otro modelo *ARIMA* o *SARIMA* que también lo haga, es por eso que el diseño *ARIMA* de BOX – JENKINS es un arte más que una ciencia; se requiere gran habilidad para seleccionar el modelo *ARIMA* o *SARIMA* correcto.

La finalidad de esta fase consiste en analizar la adecuación entre el modelo y los datos, o dicho de otra forma en qué medida se cumple lo siguiente.

Prueba de Ljung – Box Análisis de los residuales

Los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco.

1. Planteamiento de la hipótesis para el modelo

H_0 : Los residuales son ruido blanco (es decir, las correlaciones en la serie de los valores residuales son 0)

H_1 : Los residuales no son ruido blanco

2. Nivel de significancia:

$$\alpha < 0.05 = 5\%$$

3. Regla de Decisión

- Aceptamos H_0 si el *p - valor* > 0.05 quiere decir, que si el valor *p* es mayor que el nivel de significancia elegido (0.05)



- Rechazamos H_0 si el p - *valor* < 0.05 quiere decir, que si el valor p es menor que el nivel de significancia elegido (0.05)

4. Estadístico de prueba

Para probar la hipótesis de que todos los coeficientes de autocorrelación son simultáneamente iguales a cero se utilizó el estadístico Q Ljung – Box Pierce. Será más favorable al ruido blanco cuanto mayor sea la probabilidad p valor.

En esta etapa de verificación la existencia de ruido blanco, para luego pasar a la predicción.

Paso 4 Predicción

Una de las razones de la popularidad del proceso de modelación ARIMA o SARIMA es su éxito en la predicción. En muchos casos las predicciones obtenidas por este método son más confiables que aquellas obtenidas de la elaboración tradicional de modelos particularmente para predicciones de corto plazo. Y teniendo el modelo definido se procedió a realizar los pronósticos a futuro haciendo uso del modelo obtenido en el procedimiento.

CAPITULO IV

RESULTADOS Y DISCUSIÓN

En este capítulo presentamos los resultados del análisis de datos obtenidos de la Municipalidad Distrital de Caracoto. Los resultados de investigación realizada se presentan con su respectivo análisis e interpretación, la misma que está organizando según los objetivos planteados, como: encontrar el modelo que mejor se ajuste al trabajo en horas de las maquinarias agrícolas.

4.1. RESULTADOS DE RECOPIACIÓN DE LA SERIE DE DATOS

La información de las horas de trabajo de las maquinarias agrícolas, fueran obtenidas de la municipalidad distrital de Caracoto.

Tenemos la siguiente tabla que refleja la serie de tiempo de horas de trabajo de las maquinarias agrícolas de la Municipalidad Distrital de Caracoto.

Tabla 2 *Serie original del total de horas de trabajo mensual de las maquinarias de la municipalidad de Caracoto 2010 – 2019.*

Años /Meses	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2010	14	25	94	81	198	155	168	20	84	181	190	70
2011	8	21	96	89	197	94	156	44	98	224	200	66
2012	10	23	78	90	179	90	156	46	119	290	206	66
2013	10	23	76	108	183	100	168	46	125	308	240	70
2014	20	31	88	123	194	110	168	48	256	343	270	70
2015	14	27	70	150	204	116	194	54	230	350	315	80
2016	16	27	86	154	234	122	234	64	250	364	350	70
2017	48	31	92	126	296	134	228	78	255	351	396	62
2018	32	33	126	148	313	150	252	84	275	381	376	92
2019	52	37	156	161	333	167	276	102	310	420	386	135
promedios mensuales	22.4	27.8	96.2	123	233.1	124	200	58.6	200.2	321.2	292.9	78.1

Fuente: Elaboración propia en base a los datos de la municipalidad distrital de Caracoto.

4.1.1. Fase de análisis del comportamiento de la serie

GRAFICA DE LA SERIE ORIGINAL HORA

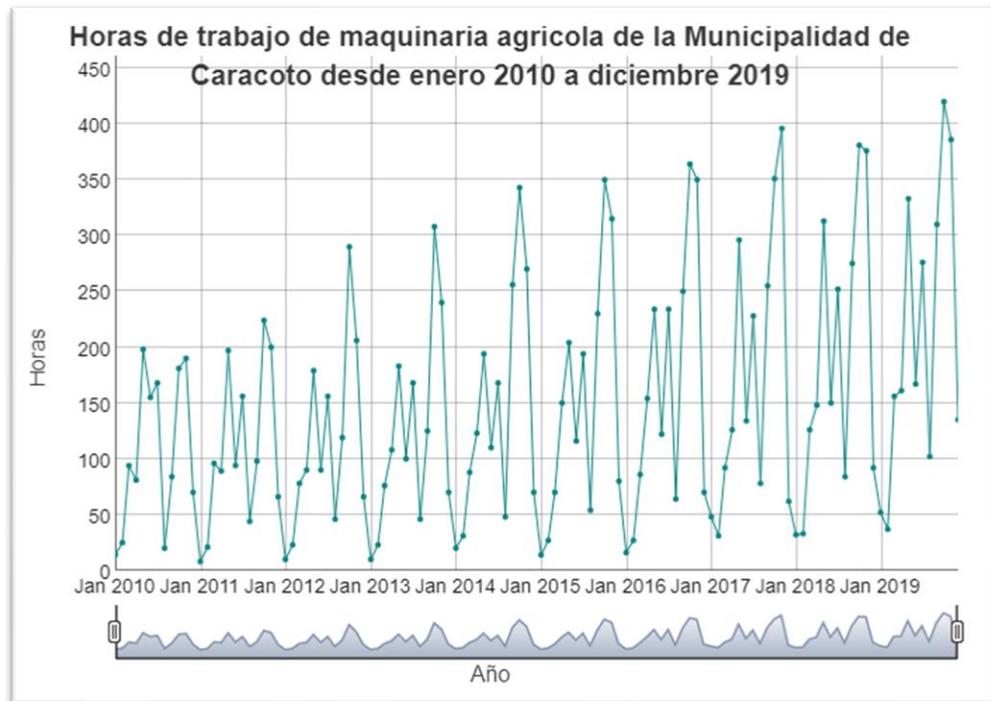


Figura 11 Serie original del número de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad Distrital de Caracoto.

Interpretación:

En la tabla y figura N° 11 se muestran los datos originales de las horas de trabajo de las maquinarias, registrados en meses y años de la Municipalidad Distrital de Caracoto, periodo 2010 – 2019. Muestra una tendencia creciente, con un efecto estacional que produce picos en el mes de octubre. Así como también se observa un comportamiento estacional cada 12 meses durante el periodo en estudio, esto nos indica que la serie tiene estacionalidad. Además, conlleva a plantear un modelo; es decir que para la aplicación de la metodología BOX – JENKINS, la serie tiene que ser transformada.

GRAFICA DE LA SERIE DESCONPUESTA

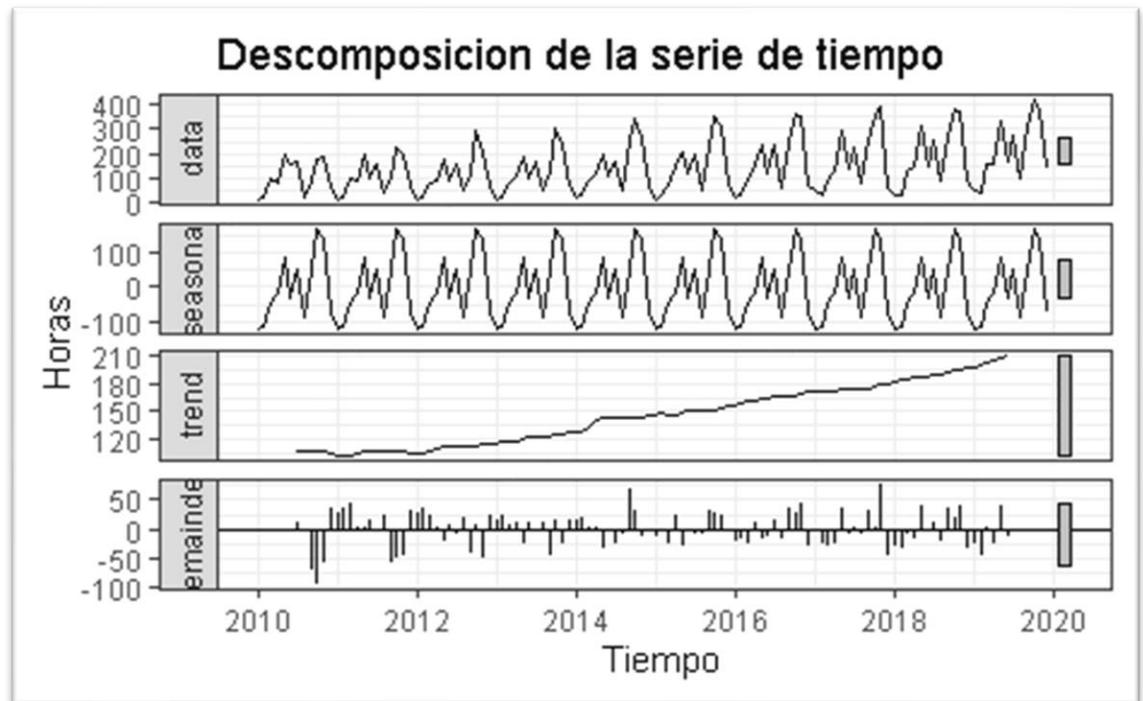


Figura 12 Diagrama de descomposición de la serie de datos horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 - 2019.

Fuente: Elaboración propia en base a los datos de la Municipalidad distrital de Caracoto.

Interpretación:

En este grafico podemos observar 4 sub gráficos:

- **Data:** Permite apreciar el comportamiento de los datos originales de las horas de trabajo de las maquinarias sin ninguna transformación a través del tiempo, su tendencia va en aumento moderado.
- **Seasonal:** la estacionalidad de los datos en patrón mensual / anual de los puntos de los datos.
- **Trend:** La tendencia de los datos el cual se observa que es positiva.
- **Remainder:** El ruido blanco la descomposición espectral

GRAFICA DE LA SERIE CON LINEA DE TENDENCIA

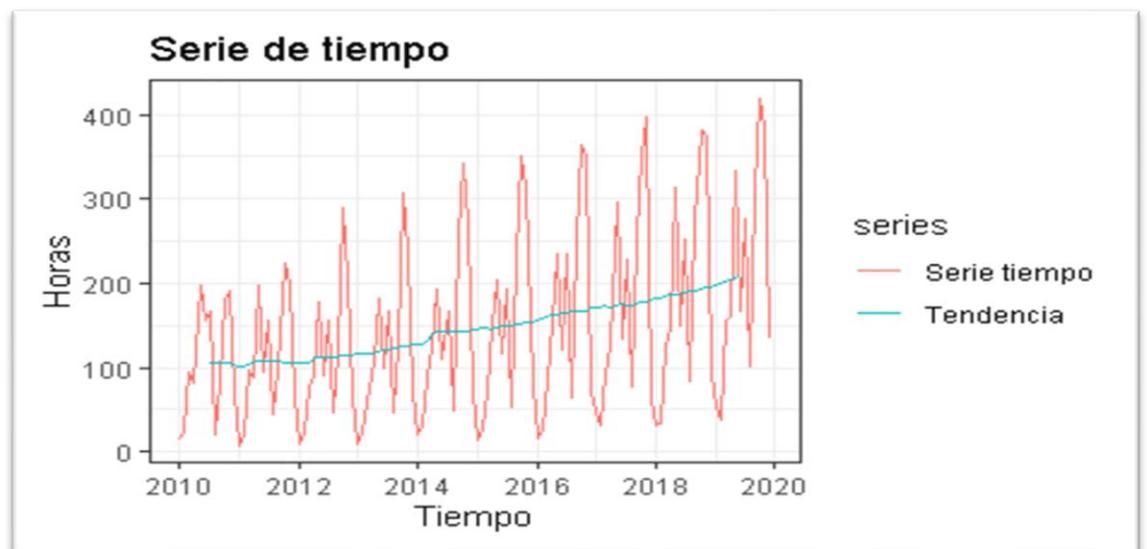


Figura 13 Grafica de la tendencia de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 - 2019.
Fuente: Elaboración propia en base a los datos de la Municipalidad distrital de Caracoto.

Interpretación:

En el grafico se observa que la serie de horas de trabajo de las maquinarias de la Municipalidad Distrital de Caracoto, tiene tendencia creciente (positiva).

GRAFICA DE ESTACIONALIDAD DE LA SERIE HORA

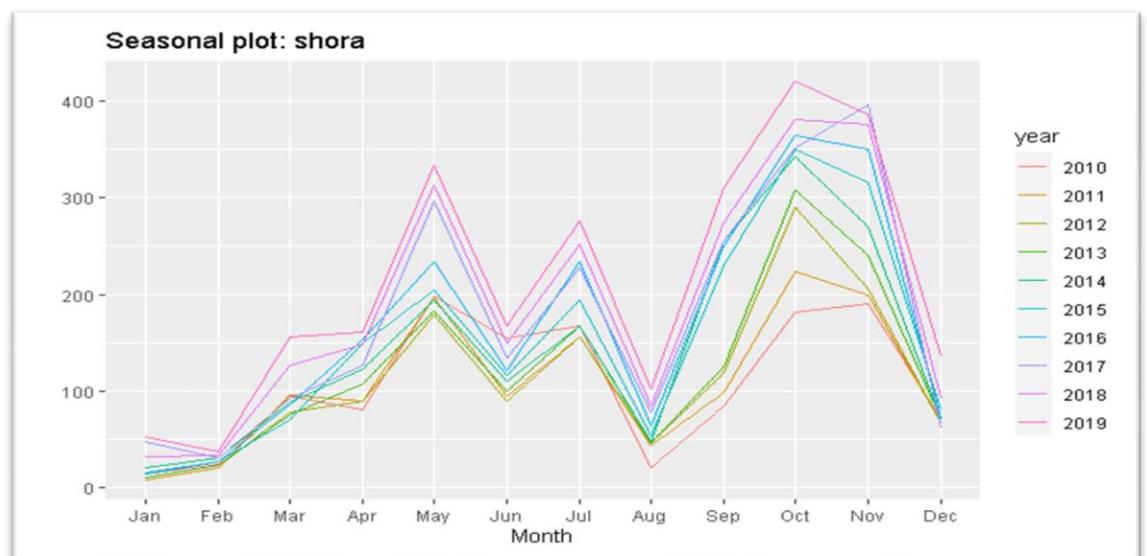


Figura 14 Gráfico de estacionalidad de la serie de horas de trabajo de las maquinarias agrícolas de la municipalidad distrital de Caracoto 2010 - 2019.
Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

Con esta gráfica se puede observar mejor la estacionalidad, mostrando claramente que en enero, abril, agosto y diciembre se tiene un menor número de horas y en los meses de mayo y octubre se tiene un mayor número de horas de trabajo de las maquinarias agrícola de la Municipalidad Distrital de Caracoto.

GRAFICA DE CAJAS DE LA SERIE HORA

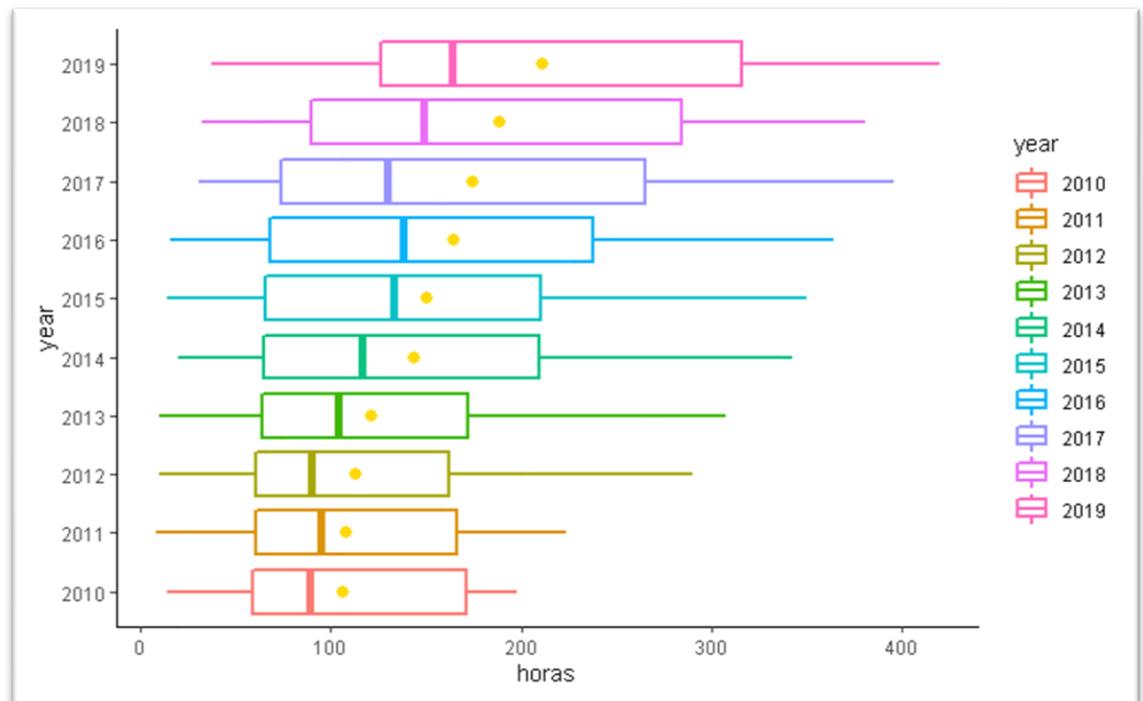


Figura 15 Gráfico de cajas de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 - 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En el gráfico se muestra que en la serie de horas de trabajo de las maquinarias agrícolas de la Municipalidad Distrital de Caracoto hay tendencia en varianza y en media año tras año en la distribución de los datos.

GRAFICO DE HISTOGRAMA DE LA SERIE HORA

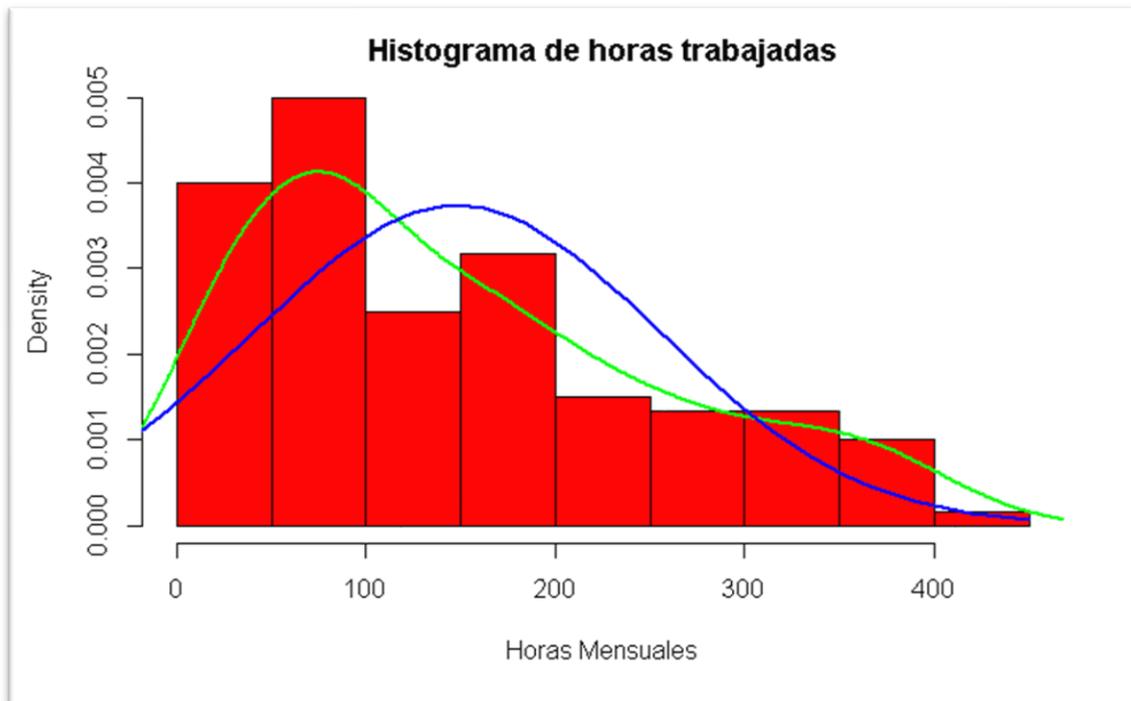


Figura 16 Histograma y distribución normal de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la figura 16, se observa el histograma de la serie de horas de trabajo de las maquinarias agrícola de la Municipalidad Distrital de Caracoto que no se ajusta a una distribución normal, con curva de asimetría Positiva; los valores individuales de la serie no giran en torno a su valor medio y no tiene variación.

GRAFICO DE Q-Q PLOT DE LA SERIE HORA

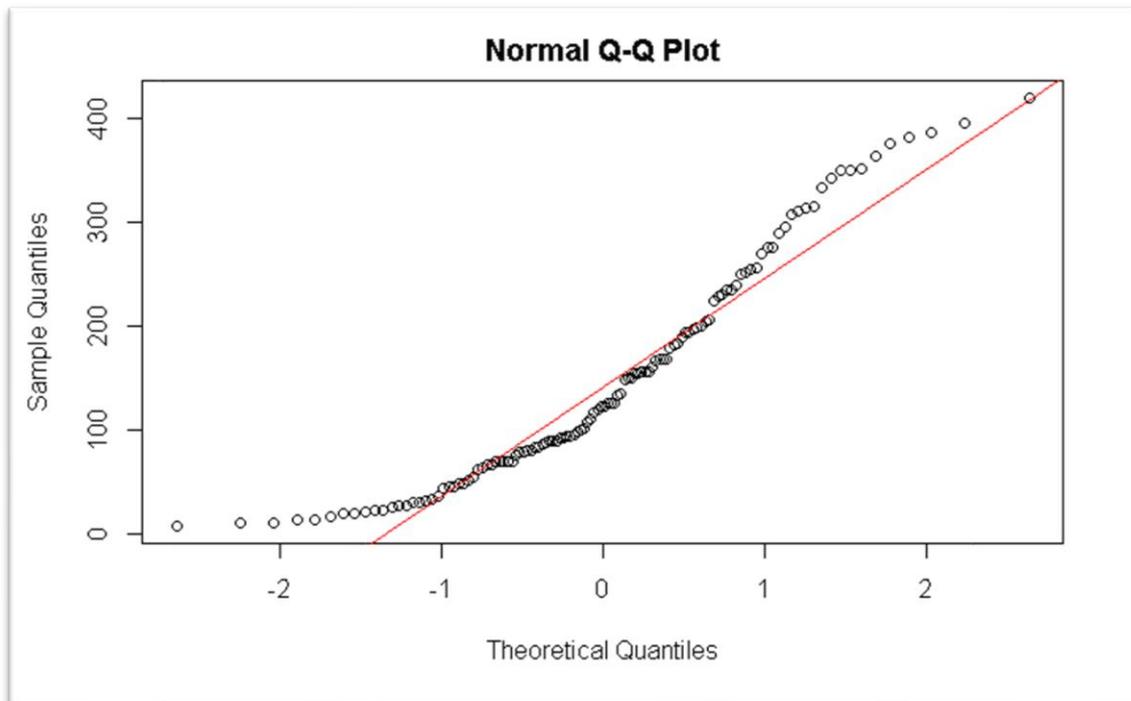


Figura 17 Grafico de normalidad $Q - Q$ Plot de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la figura 17, se observa de la serie de horas de trabajo de las maquinarias agrícola de la Municipalidad Distrital de Caracoto que no se ajusta a una distribución normal.

TEST DE PEARSON DE LA SERIE DE DATOS

H_0 : los datos de las horas de trabajo de las maquinarias provienen de una población normal

H_1 : los datos de las horas de trabajo de las maquinarias no provienen de una población normal

```
##  
## Pearson chi-square normality test  
##  
## data: shora  
## P = 42.4, p-value = 1.381e-05
```

Figura 18 Resultados del test de Pearson de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

De la salida anterior se tiene que el valor-P para la prueba fue de 0.00001 y que es menor al nivel de significancia 0.05, lo cual indica que hay evidencias para rechazar la hipótesis nula de normalidad.

Por lo cual para una buena predicción no se puede aplicar la metodología de BOX – JENKINS; ya que es recomendable que la serie se ajuste a una distribución normal.

APLICAMOS AUTO-ARIMA A LA SERIE HORA

El comando Auto - arima no proporciona un modelo, que nos ayudara a tomar en cuenta cuantos AR(p), d y MA(q).

```
## Series: shora  
## ARIMA(0,1,1) (0,1,0) [12]  
##  
## Coefficients:  
##          ma1  
##          -0.9487  
## s.e.      0.0348  
##  
## sigma^2 estimated as 483.9: log likelihood=-483.21  
## AIC=970.42   AICc=970.54   BIC=975.77
```

Figura 19 Resultado del comando Auto-arima de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.



Interpretación:

El comando *Auto – arima*, nos muestra como el mejor modelo es un SARIMA y se muestra como: *ARIMA (0,1,1)(0,1,0)*[12], como coeficiente: $MA(1) = -0.9487$; y una diferenciación en la parte regular $d = 1$ y una diferenciación en la parte estacionaria $D = 12$.

4.2. TRANSFORMACIÓN DE LA SERIE DE DATOS

(Parra, Estadística y Machine Learning con R, 25 de Enero de 2019) En caso de que la serie Y_t presente un comportamiento estacional no estacionario, habrá que tomar diferencias entre aquellas observaciones separadas por el periodo que presenta el comportamiento estacional, aplicando para ello el operador diferencia estacional

La detección del comportamiento estacional de la serie y su carácter estacionario es importante ya que, tal y como Box y Jenkins plantearon, es posible incorporar a un modelo *ARIMA (p, d, q)* las correlaciones existentes entre pares de observaciones separadas por periodos estacionales suponiendo que el término de error de un modelo ARIMA para la parte estacional está correlacionado en serie.

4.2.1. Aplicando la Primera Diferencia de la serie de datos

Para que la serie de horas de trabajo de las maquinarias de la municipalidad de Caracoto, sea estacionaria en varianza se debe transformar con la primera diferencia regular a la serie y se tiene como resultado lo siguiente:

Tabla 3: Datos con primera diferencia de la serie horas de trabajo de la municipalidad distrital de Caracoto 2010 al 2019.

Años /Meses	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2010	14	25	94	81	198	155	168	20	84	181	190	70
2011	8	21	96	89	197	94	156	44	98	224	200	66
2012	10	23	78	90	179	90	156	46	119	290	206	66
2013	10	23	76	108	183	100	168	46	125	308	240	70
2014	20	31	88	123	194	110	168	48	256	343	270	70
2015	14	27	70	150	204	116	194	54	230	350	315	80
2016	16	27	86	154	234	122	234	64	250	364	350	70
2017	48	31	92	126	296	134	228	78	255	351	396	62
2018	32	33	126	148	313	150	252	84	275	381	376	92
2019	52	37	156	161	333	167	276	102	310	420	386	135
promedios mensuales	22.4	27.8	96.2	123	233.1	124	200	58.6	200.2	321.2	292.9	78.1

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

GRAFICO DE CAJAS DE LA PRIMERA DIFERENCIA

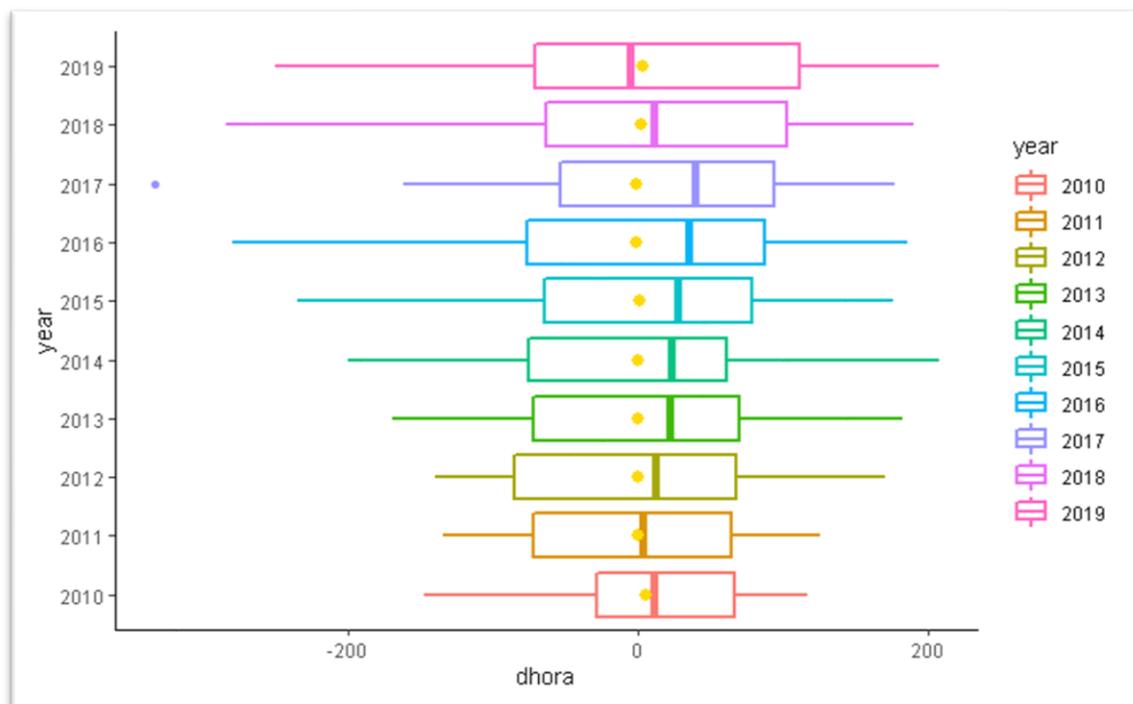


Figura 20 Grafico de caja con primera diferencia de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 al 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En el grafico se muestra que después de realizar la primera diferencia en la serie de horas de trabajo de las maquinarias de la municipalidad de Caracoto

vemos que la varianza mejora en la parte estacionaria pero aun presenta variación en el año 2016 y 2018 también ya no hay mucha variación en la mediana.

GRAFICO DE HISTOGRAMA DE LA PRIMERA DIFERENCIA

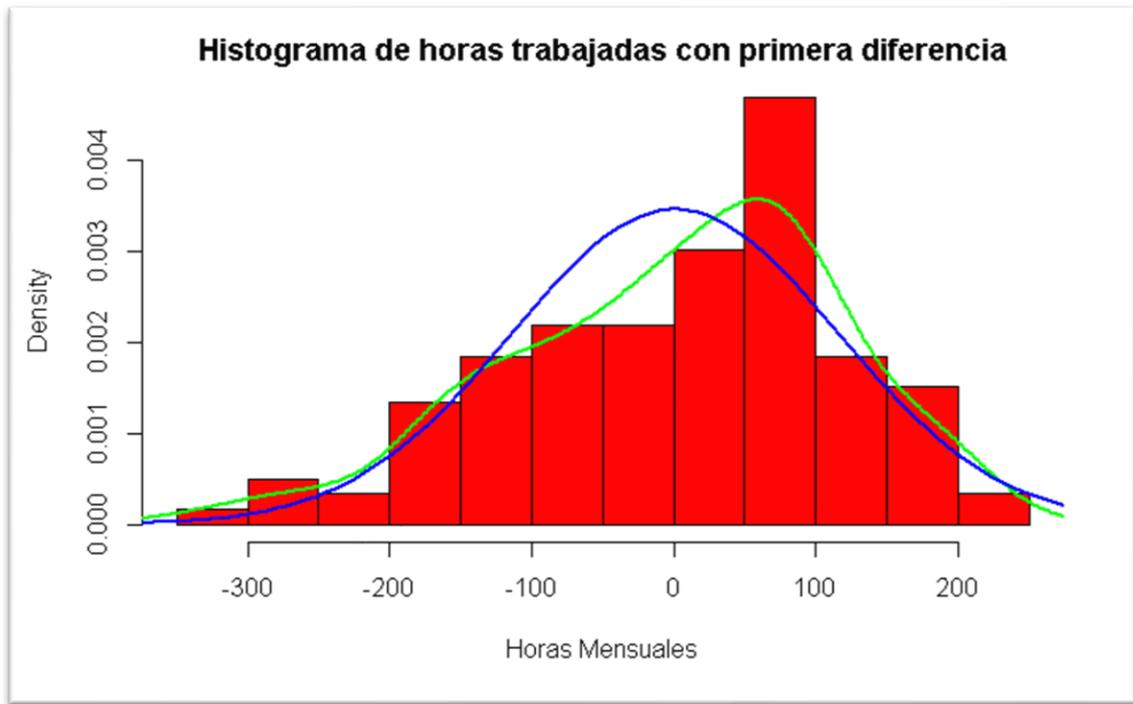


Figura 21 Histograma y distribución normal con primera diferencia de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 al 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la figura 20, se observa el histograma de la serie de horas de trabajo de las maquinarias de la municipalidad de Caracoto con primera diferencia que todavía no se distribuye según una distribución normal.

GRAFICO DE Q-Q PLOT DE LA PRIMERA DIFERENCIA

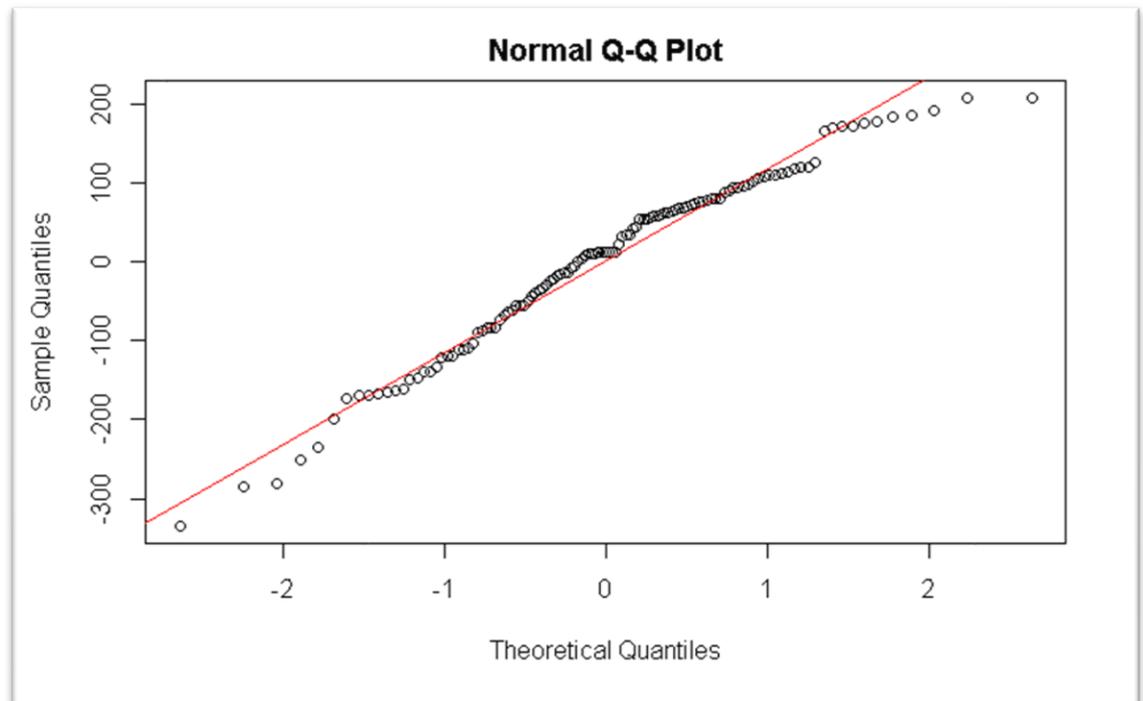


Figura 22 Grafico de normalidad $Q - Q$ Plot de la serie con primera diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la figura 21, se observa de la serie con primera diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto que al parecer se ajusta a una distribución normal para lo cual aplicamos el test de normalidad.

Test de Pearson primera diferencia

H_0 : los datos de las horas de trabajo de las maquinarias provienen de una población normal

H_1 : los datos de las horas de trabajo de las maquinarias no provienen de una población normal

```
##  
## Pearson chi-square normality test  
##  
## data: dhora  
## P = 17.353, p-value = 0.09786
```

Figura 23 Resultados del test de Pearson de la serie con primera diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la figura se tiene que el valor-P para la prueba fue de 0.09786 y que es mayor al nivel de significancia 0.05, lo cual indica que hay evidencias para rechazar la hipótesis nula de normalidad. Como se ve en el estadístico de PEARSON concluimos que la serie si se ajusta a una distribución normal; ya podemos aplicar la metodología de BOX – JENKINS.

4.3. APLICACIÓN DE LA METODOLOGIA BOX JENKINS PARA TIEMPO DE TRABAJO DE LAS MAQUINARIAS AGRICOLAS

Teniendo los datos transformados procesamos para el análisis de las horas de trabajo de las maquinarias, se aplicó la metodología BOX – JENKINS que consta de cuatro fases que acompañaran en todo el proceso para encontrar el modelo adecuado para la presente investigación son:

4.3.1. Identificación del Modelo

EL primer paso para la identificación del modelo fue verificar si la serie era estacionaria o no estacionaria, para lo cual se realizó el ploteo de la serie y/o revisión grafica de las horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.

El análisis de la serie original (ver figura N° 11), nos mostró las características de interés para su correspondiente análisis.

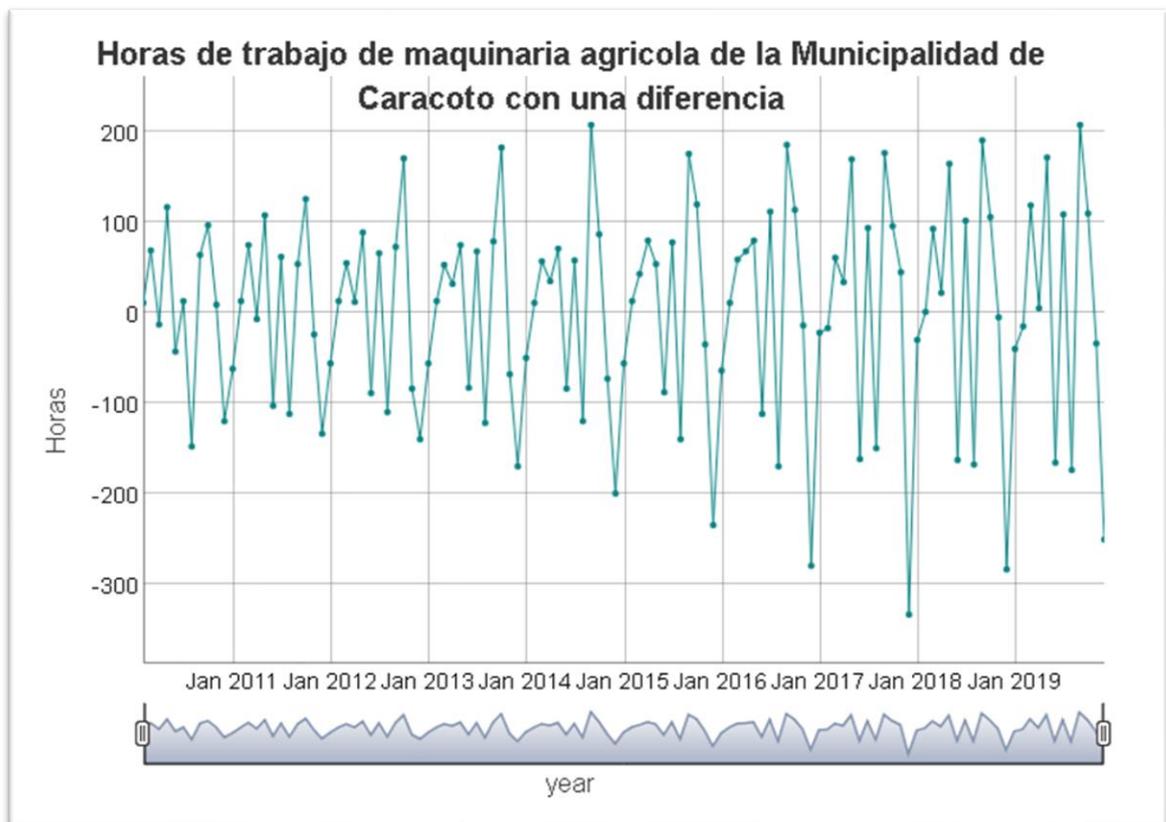


Figura 24 Grafico de la serie con una diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En el grafico 23 de la serie con una diferencia se observa que la serie no tiene tendencia en media y en varianza; ahora se tiene que probar que la serie no tenga raíz unitaria (serie estacionaria).

TEST DE DICKEY – FULLER

d. Formulación de la hipótesis Estadística

$H_0: \rho = 1$ La serie tiene raíz unitaria (serie no estacionario)

$H_1: \rho < 1$ La serie no tiene raíz unitaria (serie estacionario)

e. Regla de Decisión

- Aceptamos H_0 si el valor p – valor es igual a 1 que el nivel de significancia elegido (0.05)
- Rechazamos H_0 si el valor p – valor es menor a 1 que el nivel de significancia elegido (0.05)

f. Estadístico de prueba

Para ver la estacionariedad de la serie de tiempo se utilizó la prueba de Dickey - Fuller

g. Decisión

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: dhora  
## Dickey-Fuller = -9.4505, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary
```

Figura 25 Resultados del test de Dickey – Fuller de la serie con segunda diferencia de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto 2010 – 2019.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la figura 28 nos muestra que a un nivel de significancia del 5% se tiene que el valor-P 0.01 el cual es menor al nivel de significancia 0.05. Por lo tanto, indica que se rechaza la hipótesis nula H_0 y se acepta la hipótesis alterna H_1 quiere decir, que los datos de la serie con una diferencia de horas de trabajo de las maquinarias de la municipalidad de Caracoto no presentan raíz unitaria por lo cual la serie es estacionaria.

4.3.2. Estimación del modelo

Se estima parámetros del modelo identificado mediante el software R.

Aplicamos *Auto-Arima* A La Serie Con una Diferencias Y Graficamos La Autocorrelación Simple Y Autocorrelación Parcial.



CUADRO DE MODELOS PROPUESTOS POR AUTO-ARIMA

##	ARIMA (2,0,2) (1,1,1) [12] with drift	: Inf
##	ARIMA (0,0,0) (0,1,0) [12] with drift	: 1032.429
##	ARIMA (1,0,0) (1,1,0) [12] with drift	: 1011.343
##	ARIMA (0,0,1) (0,1,1) [12] with drift	: Inf
##	ARIMA (0,0,0) (0,1,0) [12]	: 1030.377
##	ARIMA (1,0,0) (0,1,0) [12] with drift	: 1009.453
##	ARIMA (1,0,0) (0,1,1) [12] with drift	: 1011.373
##	ARIMA (1,0,0) (1,1,1) [12] with drift	: 1013.236
##	ARIMA (2,0,0) (0,1,0) [12] with drift	: 1000.839
##	ARIMA (2,0,0) (1,1,0) [12] with drift	: 1002.837
##	ARIMA (2,0,0) (0,1,1) [12] with drift	: 1002.886
##	ARIMA (2,0,0) (1,1,1) [12] with drift	: 1004.207
##	ARIMA (3,0,0) (0,1,0) [12] with drift	: 997.6634
##	ARIMA (3,0,0) (1,1,0) [12] with drift	: 999.4645
##	ARIMA (3,0,0) (0,1,1) [12] with drift	: 999.5566
##	ARIMA (3,0,0) (1,1,1) [12] with drift	: 1000.894
##	ARIMA (4,0,0) (0,1,0) [12] with drift	: 998.2301
##	ARIMA (3,0,1) (0,1,0) [12] with drift	: Inf
##	ARIMA (2,0,1) (0,1,0) [12] with drift	: Inf
##	ARIMA (4,0,1) (0,1,0) [12] with drift	: Inf
##	ARIMA (3,0,0) (0,1,0) [12]	: 995.566
##	ARIMA (3,0,0) (1,1,0) [12]	: 997.317
##	ARIMA (3,0,0) (0,1,1) [12]	: 997.4107
##	ARIMA (3,0,0) (1,1,1) [12]	: 998.686
##	ARIMA (2,0,0) (0,1,0) [12]	: 998.7515
##	ARIMA (4,0,0) (0,1,0) [12]	: 996.1103
##	ARIMA (3,0,1) (0,1,0) [12]	: 976.5267
##	ARIMA (3,0,1) (1,1,0) [12]	: 978.3886
##	ARIMA (3,0,1) (0,1,1) [12]	: 978.4742
##	ARIMA (3,0,1) (1,1,1) [12]	: 979.7282
##	ARIMA (2,0,1) (0,1,0) [12]	: 974.5072
##	ARIMA (2,0,1) (1,1,0) [12]	: 976.2285
##	ARIMA (2,0,1) (0,1,1) [12]	: 976.3348
##	ARIMA (2,0,1) (1,1,1) [12]	: 977.4523
##	ARIMA (1,0,1) (0,1,0) [12]	: 972.4447

```
## ARIMA(1,0,1)(1,1,0)[12] : 974.2743
## ARIMA(1,0,1)(0,1,1)[12] : 974.3418
## ARIMA(1,0,1)(1,1,1)[12] : Inf
## ARIMA(0,0,1)(0,1,0)[12] : 970.5358
## ARIMA(0,0,1)(1,1,0)[12] : 972.3239
## ARIMA(0,0,1)(0,1,1)[12] : 972.3964
## ARIMA(0,0,1)(1,1,1)[12] : 973.369
## ARIMA(0,0,2)(0,1,0)[12] : 972.4322
## ARIMA(1,0,0)(0,1,0)[12] : 1007.377
## ARIMA(1,0,2)(0,1,0)[12] : Inf
## ARIMA(0,0,1)(0,1,0)[12] with drift : Inf
##
## Best model: ARIMA(0,0,1)(0,1,0)[12]
```

Figura 26 Modelos propuestos mediante Auto-Arima a la serie con una diferencia de las Horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

AUTO-ARIMA DE LA SERIE HORACON UNA DIFERENCIA

```
## Series: dhora
## ARIMA(0,0,1)(0,1,0)[12]
##
## Coefficients:
##          ma1
##          -0.9487
## s.e.      0.0348
##
## sigma^2 estimated as 483.9: log likelihood=-483.21
## AIC=970.42  AICc=970.54  BIC=975.77
```

Figura 27 Coeficientes del mejor modelo propuesto con Auto – Arima de la serie dhora con un MA(1), una diferencia en la parte regular y una en la parte estacionaria de las Horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la gráfica 30 nos muestra que el mejor modelo generado por *AUTO-ARIMA* es un *SARIMA(0,0,1)(0,1,0)[12]*; de la serie aplicado con una diferencia en la parte regular y una diferencia en la parte estacionara.

CORRELACIÓN SIMPLE Y CORRELACIÓN PARCIAL

Función de autocorrelación simple y parcial de horas de trabajo de las maquinarias agrícolas de la municipalidad distrital de Caracoto (con límites de significancia de 5% para las autocorrelaciones)

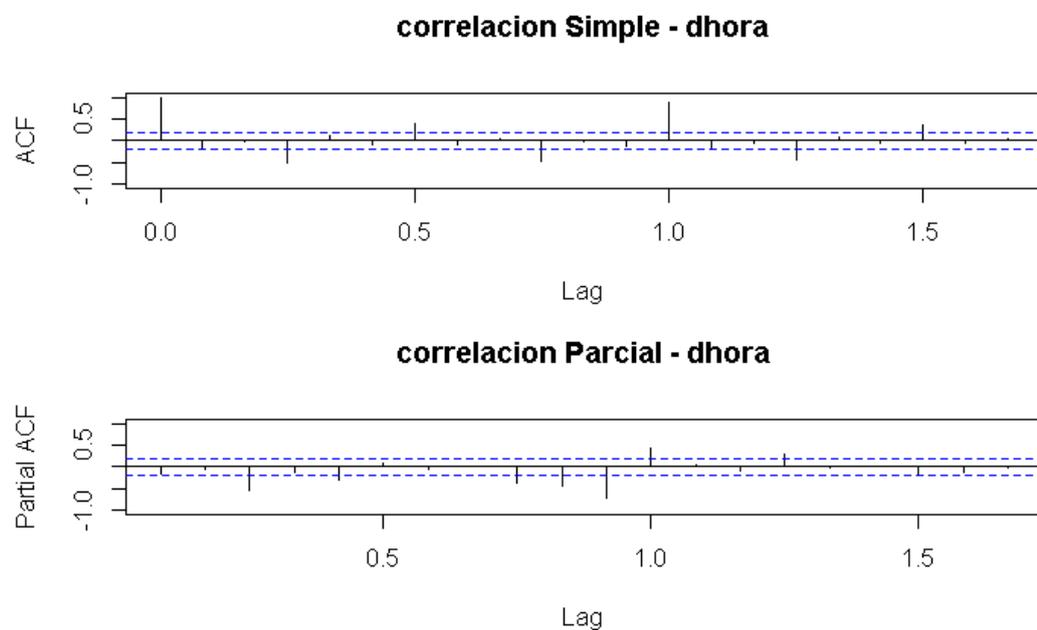


Figura 28 *Correlación simple y parcial del modelo recomendado por AUTO – ARIMA de la serie de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.*

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En el grafico N° 31 se muestra la correlación simple y parcial del mejor modelo calculado por *AUTO – ARIMA*, en el gráfico de autocorrelación simple se observa que el modelo *SARIMA* se debe considerar un *MA(1)* en la parte

regular, con coeficiente -0.9487 ; y se tiene que aplicar una diferenciación en la parte estacional observamos en grafico de correlación parcial de la serie histórica, donde advierte el comportamiento de los coeficientes.

**PARA CALIBRAMOS LOS MODELOS ENCONTRADOS POR
AUTO-ARIMA PARTICIONAMOS LA SERIE EN DOS GRUPOS:**

Según (Rodriguez, 2017) se particiona la serie en dos grupos para medir la bondad de ajuste de modelo calibrado(aprendizaje) frente al grupo de testing.

- Primer grupo: De los años 2010 al 2018 de la serie con una diferencia (Aprendizaje)
- Segundo grupo: Del año 2019 de la serie con una diferencia (Testing)

Tabla 4 *Primer Grupo ARENDIZAJE serie de horas de trabajo de la municipalidad distrital de Caracoto del 2010 al 2018 y Grupo de Testing del 2019.*

AÑOS / MESES	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010		11	69	-13	117	-43	13	-148	64	97	9	-120
2011	-62	13	75	-7	108	-103	62	-112	54	126	-24	-134
2012	-56	13	55	12	89	-89	66	-110	73	171	-84	-140
2013	-56	13	53	32	75	-83	68	-122	79	183	-68	-170
2014	-50	11	57	35	71	-84	58	-120	208	87	-73	-200
2015	-56	13	43	80	54	-88	78	-140	176	120	-35	-235
2016	-64	11	59	68	80	-112	112	-170	186	114	-14	-280
2017	-22	-17	61	34	170	-162	94	-150	177	96	45	-334
2018	-30	1	93	22	165	-163	102	-168	191	106	-5	-284
2019	-40	-15	119	5	172	166	109	174	208	110	-34	251

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.



CALIBRACION DEL MODELO CON LOS DATOS DE APRENDIZAJE Y TESTIN

Según (Rodríguez, 2017) para encontrar el mejor modelo, calibraremos los modelos propuesto por *Auto – arima* y los mediremos con el Error Cuadrático Medio (ECM); y elegimos al modelo que tenga el menor *ECM*.

Leemos los modelos que fueron propuesto por *Auto – arima* de la serie con primera diferencia y los calibramos con el Grupo aprendizaje

CUADRO DE ERRORES CUADRATICOS MEDIOS

Tabla 5 Cuadro de los Errores Cuadráticos Medios de los modelos calibrados de la función programada.

N° Mod.	=	ECM Mod
1	=	17.55704987
2	=	20.98224647
3	=	17.55704987
4	=	19.50833263
5	=	20.65176795
6	=	22.35573145
7	=	15.91098333
8	=	15.64991914
9	=	15.71532466
10	=	15.2407394
11	=	18.8529456
12	=	18.21634967
13	=	18.3636504
14	=	18.1818378
15	=	17.79292965
16	=	18.8529456
17	=	18.21634967
18	=	18.3636504
19	=	18.1818378
20	=	15.91098333
21	=	17.79292965
22	=	17.9184085
23	=	17.34505405
24	=	17.49766129
25	=	17.06868072
26	=	17.15084915
27	=	16.27661042

28	=	16.49840542
29	=	16.48084184
30	=	17.36848959
31	=	16.85038897
32	=	16.97454978
33	=	17.45647452
34	=	16.8942744
35	=	17.03792339
36	=	17.16730446
37	=	17.34158085
38	=	19.50833263

Fuente: Elaboración propia en base a datos de la Municipalidad Distrital de Caracoto.

Interpretación:

En la tabla 5, se muestra los Errores Cuadráticos Medios de los modelos calibrados, para elegir el mejor modelo considerando el menor error cuadrático medio; el cual es el modelo número 10 con un error cuadrático medio de 15.240 del *SARIMA*(2,0,0) (1,1,1) (12).

IDENTIFICACIÓN DEL MODELO A UTILIZAR

Según (Rodríguez, 2017) en la fase de identificación de modelo se procede a identificar los valores del parte autorregresivo (p) y de la parte de medias móviles (q) y parte estacional de un modelo *SARIMA* (p, d, q) (P, D, Q)¹²

Conforme se observa en el gráfico de la parte regular se obtiene (p, d, q). La parte *AR*(1) proviene del decrecimiento rápido inicial de la función de autocorrelación parcial, Asimismo la parte *AR*(2) proviene que la función de autocorrelación parcial presenta el segundo retardo significativo en la mayoría de los periodos.

Calculamos los siguientes modelos sugeridos, al cual les medicaremos los errores en pronósticos de series (ER, ECM, TFA y PTFA) respectivamente para cada uno:

- SARIMA (0, 1, 1) (0, 1, 0) (12)

Modelo aplicado Auto-arima la serie con una diferencia

Cuya ecuación es:

$$y_t = \mu + \Phi_1 y_{t-12} + a_t + \theta_1 y_{t-1} + \theta_1 a_{t-12}$$

$$(1 - L^1)(1 - L^{12})y_t = (1 - \Phi^{12})(1 + \theta L^1)(1 + \theta L^{12})a_t$$

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

Modelo calibramos los ECM de la serie con una diferencia

Cuya ecuación es:

$$y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \Phi_1 y_{t-12} + a_t + \theta_1 a_{t-12}$$

$$(1 - L^1)(1 - L^{12})y_t = (1 - \varphi L^1 - \varphi L^2)(1 - \Phi^{12})(1 + \theta L^{12})a_t$$

- ARIMA (2, 1, 0) (0, 0, 0) (12)

Modelo que planteamos como investigador cuando se observa las gráficas de ACF y PACF de la serie con una diferencia

Cuya ecuación es:

$$y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + a_t$$

$$(1 - L^1)y_t = (1 - \varphi L^1 - \varphi L^2)a_t$$

CALCULAMOS LOS ERRORES PARA ENCONTRAR EL MEJOR MODELO SEGÚN EL GRAFICO DE RADAR.

Tabla 6 Cuadro de Medición de errores en predicción de series de los modelos propuestos para la serie de horas de trabajo de maquinarias de la municipalidad distrital de Caracoto.

##	Error Relativo	PFA	PTFA	Error Cuadratico Medio
## 1	51.453036	0.5000000	0.2014242	221.04198
## 11	4.051082	0.1666667	0.1671467	17.53775
## Errores Auto-arima	4.069767	0.5000000	0.1918748	17.53775
## Errores Calibrada	4.051082	0.5000000	0.1671467	18.91927
## Errores Planteado	51.453036	0.1666667	0.2014242	221.04198

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

COMPARAMOS LOS MEDELOS CON EL GRAFICO DE RADAR

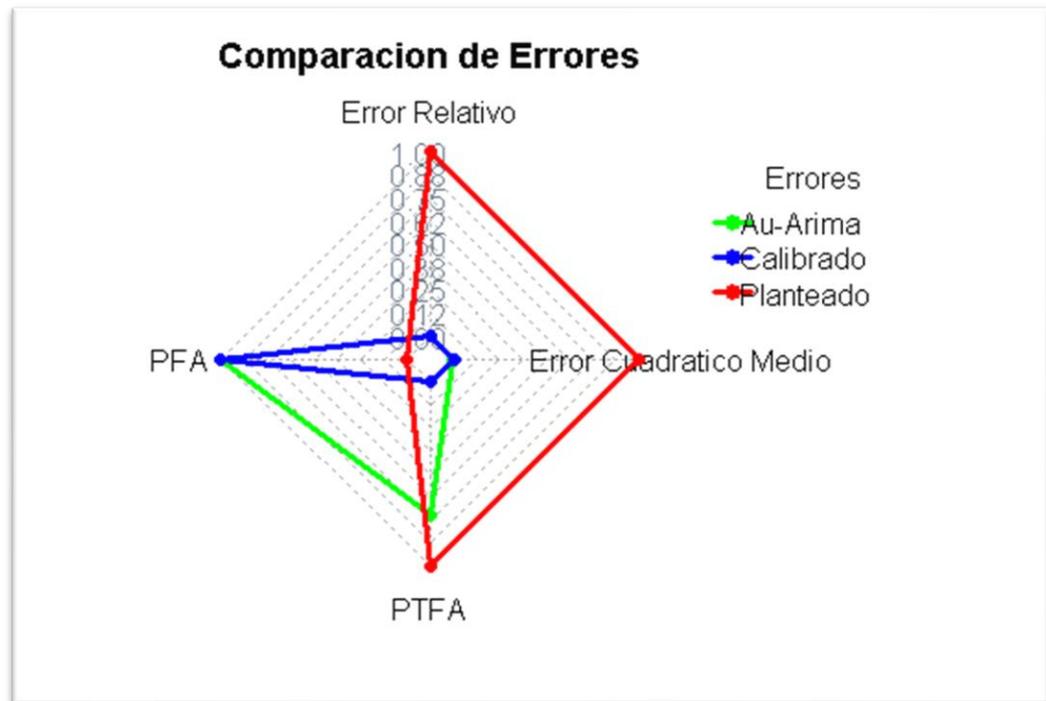


Figura 29 Grafico de radar comparación de Medición de errores (ER, ECM, TFA y PTFA) de los modelos propuestos

Fuente: Elaboración propia en base a los datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la figura 33 nos muestra el grafico de radar se observa que el modelo de *Auto – arima* y el modelo Calibrado tienen mediciones similares, a excepción de PTFA lo cual para elegir el menor modelo elegimos el menor PTFA que pertenece al modelo Calibrado de $SARIMA(2, 1, 0)(1, 1, 1) (12)$.

4.3.3. Validación del Modelo

En esta fase tenemos el modelo estimado para la serie (Transformada en esta investigación) y siendo estacionaria en media y varianza para eso se observó, en que los residuos del modelo estimado se aproximan al comportamiento del ruido blanco, que es lo primordial para validar el modelo tentativo. Los residuales deben ser aleatorios tal como se propone en la prueba de hipótesis planteada para esta fase.

Para la ecuación de este modelo para las horas de trabajo de las maquinarias agrícolas, se evaluó varias alternativas de esta manera se eligió al menor error cuadrático medio (ECM) y menor porcentaje de total de fallas hacia abajo (PTFA) que tuvo del gráfico de radar; por lo pronto se confirmó que el mejor modelo estimado apropiado fue $SARIMA(2, 1, 0)(1, 1, 1)(12)$.

Al llevar a cabo el modelo de pronóstico ARIMA siempre es necesario analizar el comportamiento de los residuos, en este caso nos interesa analizar si estos residuos se comportan como ruido blanco. Para ello se recomienda llevar a cabo la prueba de autocorrelación de forma conjunta de Ljung-Box (Gujarati, D., & Porter, D., 2010). En lo que a esta prueba se refiere, existen distintas posturas sobre el valor del rezago utilizado (Court Monteverde, E., y Rengifo, E. W. 2011).

Prueba de Ljung – Box Análisis de los residuales

Los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco.

1. Planteamiento de la hipótesis para el modelo

H_0 : Los residuales son ruido blanco (es decir, las correlaciones en la serie de los valores residuales son 0)

H_1 : Los residuales no son ruido blanco

2. Nivel de significancia:

$$\alpha < 0.05 = 5\%$$

3. Regla de Decisión

- Aceptamos H_0 si el p – valor > 0.05 quiere decir, que si el valor p es mayor que el nivel de significancia elegido (0.05)
- Rechazamos H_0 si el p – valor < 0.05 quiere decir, que si el valor p es menor que el nivel de significancia elegido (0.05)

4. Estadístico de prueba

Para probar la hipótesis de que todos los coeficientes de autocorrelación son simultáneamente iguales a cero se utilizó el estadístico Q Ljung – Box Pierce. Será más favorable al ruido blanco cuanto mayor sea la probabilidad p valor.

5. Decisión

Tabla 7 *Tabla de prueba Box – Ljung para los 24 retardos.*

n.corre	P.va.Re	inter
1	0.356636151	1
2	0.23257823	1
3	0.008506413	0
4	0.017833795	0
5	0.017442219	0
6	0.013258465	0
7	0.023862161	0
8	0.040225633	0
9	0.057464429	1
10	0.027814755	0
11	0.032696915	0
12	0.045150132	0
13	0.065440286	1
14	0.037425394	0
15	0.045461832	0
16	0.032668928	0
17	0.044267794	0
18	0.049211888	0
19	0.056097294	1
20	0.068055087	1
21	0.073157783	1
22	0.0938903	1
23	0.088064961	1
24	0.09386429	1

Fuente: Elaboración propia en base a los datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la tabla 7 se muestra que a un nivel de significancia del 95%, se ha calculado el estadístico p-value para los 24 retardos, donde se evaluó y se le interpreto según la hipótesis. Por lo que se acepta la hipótesis nula H_0 y se rechaza la hipótesis alterna H_1 ; si uno de los coeficientes es igual a 1 que los residuales son ruido blanco.

Los errores de pronóstico están distribuidos normalmente, con la media y la varianza constante, también concluimos que el ruido blanco. El modelo $SARIMA(2, 1, 0)(1, 1, 1)(12)$ es el adecuado para la serie de tiempo para las horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.

GRAFICO DE LA EVALUACION DEL MODELO

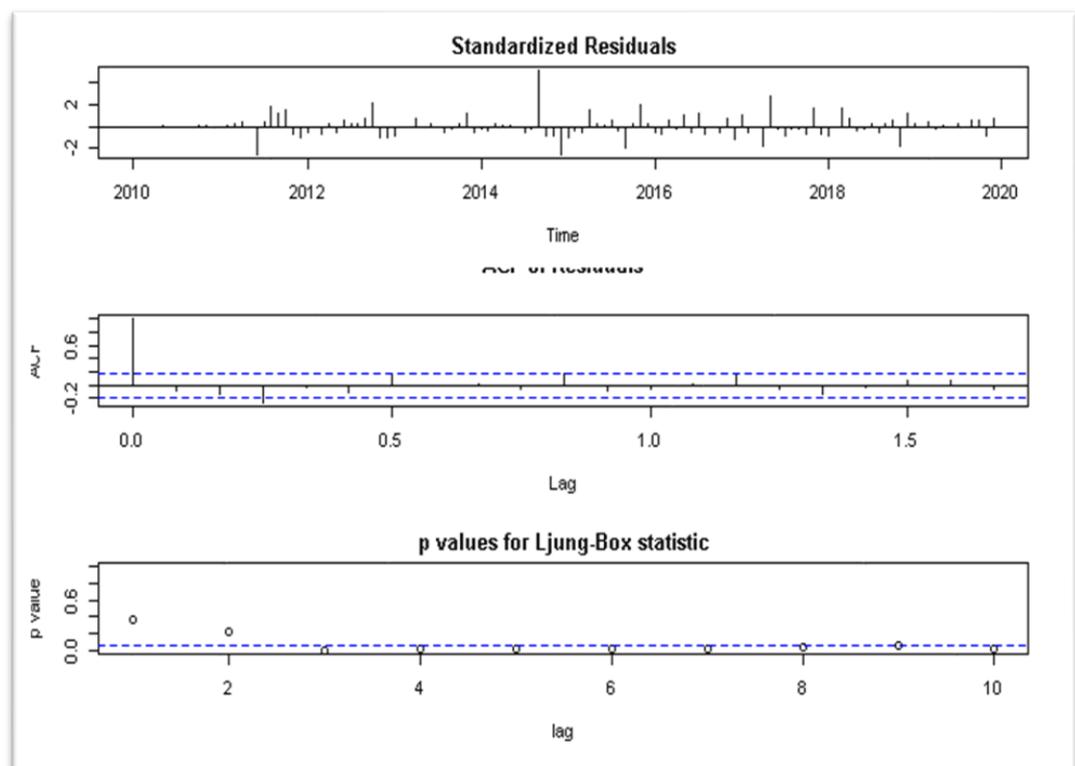


Figura 30 Autocorrelación y Box - Ljung para el modelo $ARIMA(2, 1, 0)(1, 1, 1)(12)$ de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto.

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

ECUACIÓN DE PRONOSTICO

Ecuación de pronóstico estimado para el número de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto, para futuras predicciones, mientras no exista factores externos que alteren la serie.

Modelo:

$$SARIMA(2.1.0)(1.1.1)_{[12]}$$

Ecuación del modelo:

$$y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \Phi_1 y_{t-12} + a_t + \theta_1 a_{t-12}$$

$$\theta_{13} = \hat{\theta}_1 * \hat{\theta}_{12}$$

Desfase del periodo:

“L” es el retardo o lag que se aplica a serie en la parte regular y parte estacional.

$$LY_t = Y_{t-1}$$

$$L^8 Y_t = Y_{t-8}$$

Aplicamos la diferencia en media regular y una diferencia en media Estacional, de periodo 12:

$$(1 - L^1)(1 - L^{12})y_t = (1 - 0.6111L^1 - 0.3410L^2)(1 - 0.6738L^{12})(1 + 0.5878L^{12})a_t$$

4.3.4. Pronostico con el Modelo Estimado

Una vez conseguido el modelo adecuado, calculamos los valores del pronóstico, para las horas de trabajo de las maquinarias agrícolas de la municipalidad distrital de Caracoto.

Tabla 8 Pronostico de la serie de las horas de trabajo de las maquinarias agrícolas de la municipalidad distrital de Caracoto 2020.

Mes Predic	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2020	83.89125	52.11985	115.6627	35.30109	132.4814
Feb-2020	62.68307	28.59414	96.772	10.54856	114.8176
Mar-2020	192.43076	155.80792	229.0536	136.42095	248.4406
Apr 2020	192.30438	150.8711	233.7377	128.93765	255.6711
May-2020	366.68141	322.49129	410.8715	299.09845	434.2644
Jun-2020	199.71209	152.79126	246.6329	127.95286	271.4713
Jul-2020	309.46975	259.60091	359.3386	233.20195	385.7376
Aug 2020	133.8052	81.41147	186.1989	53.6759	213.9345
Sep 2020	344.64246	289.80425	399.4807	260.77465	428.5103
Oct-2020	454.67224	397.43517	511.9093	367.1357	542.2088

Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En el cuadro se observa las predicciones para el año 2020 con límites de confianza al 80% y 95% respectivamente de la serie de datos de las horas de trabajo de las maquinarias de la municipalidad de Caracoto.

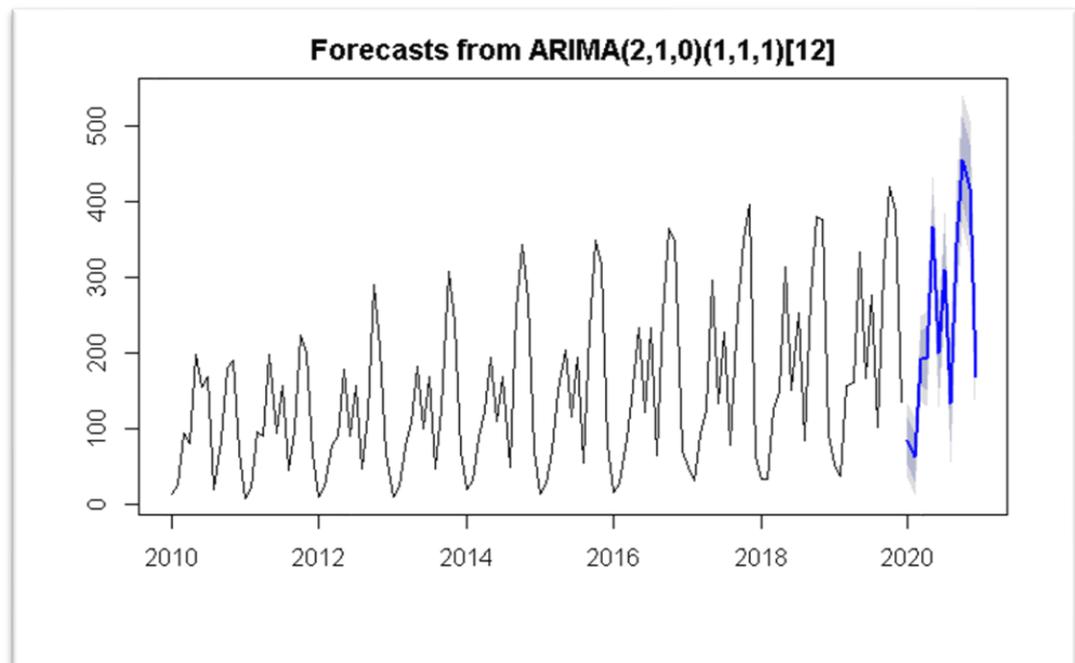


Figura 31 Grafico del Pronostico de la serie de las horas de trabajo de las maquinarias agrícolas de la municipalidad de Caracoto 2020.



Fuente: Elaboración propia en base a datos de la Municipalidad distrital de Caracoto.

Interpretación:

En la figura 32 se observa la predicción para 12 meses futuros para a serie de números de horas de trabajo de las maquinarias de la municipalidad distrital de Caracoto. El modelo es:

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

En el presente trabajo de investigación frente a los antecedentes mencionados se concuerda que en la metodología Box-Jenkins los modelos SARIMA son los que mejor se ajustan en la predicción de una serie de tiempo, por el comportamiento del factor estacional.

V. CONCLUSIONES

Se llegó a la conclusión que el modelo es $SARIMA(2, 1, 0)(1, 1, 1)(12)$ que tuvo un mejor ajuste, respecto a la serie histórica de las horas de trabajo mensual de las maquinarias agrícolas de la municipalidad distrital de Caracoto, que en su forma de operadores polinómicas es:

$$(1 - L^1)(1 - L^{12})y_t = (1 - 0.6111L^1 - 0.3410L^2)(1 - 0.6738L^{12})(1 + 0.5878L^{12})a_t$$

PRIMERO: Se realizó la exploración, análisis, transformación y diseño del modelo, utilizando la metodología de Box–Jenkins para encontrar el mejor modelo, se logró obtener la estacionariedad de la serie gráficamente y analíticamente, transformando la variable original. Los parámetros encontrados para el modelo cumplen las condiciones de estacionariedad y estos fueron significativos de acuerdo a las pruebas estadísticas de Pearson con sus respectivos niveles de significancia.

SEGUNDO: Se verificó la validez del método evaluando la bondad de ajuste del error cuadrado medio y porcentaje total de fallas hacia arriba, que es la que presenta menor valor con referencia a los demás; y se obtuvo los pronósticos para los 12 próximos meses.

TERCERO: Se realizó el pronóstico de los 12 meses para el año 2020 en horas mensuales para el trabajo de las maquinarias agrícolas.



VI. RECOMENDACIONES

PRIMERO: Se recomienda realizar trabajos de investigación del presente estudio cada cierto periodo de tiempo, ya que el análisis es recomendable para periodos cortos de tiempo y que nos permita tener una mejor visión de la población.

SEGUNDO: Se recomienda incluir otras variables como uso de combustible, cantidad de operarios que se requiere; también en otros distritos con el objetivo de conseguir modelos óptimos y tener pronósticos más útiles con el fin de planificar en la institución.

TERCERO: Para la elección del modelo se recomienda pronosticar para datos ya existentes de una serie, con el fin de comparar su similitud de las dos series.

CUARTO: En el proceso de estimación se recomienda usar las herramientas necesarias para comprobar la estacionariedad e invertibilidad del proceso, de preferencia la estacionariedad.

QUINTO: Las predicciones deben considerarse de mucha importancia, por que dan una visión de lo que puede suceder a futuro con las variables analizadas y lograr una planificación, para tomar acciones para mejor desarrollo de sociedad.

SEXTO: El análisis de serie de tiempo es de mucha importancia cuando se desea planificar el futuro, aunque en nuestro entorno no se toma en cuenta estas acciones y se lleva una vida al azar.



VII. Referencias bibliográficas

- Anderson, D., Sweeney, D., & Wilians, T. (2008). *Estadística para Administración y Economía*. Mexico.
- Bedoya Luza, S. L. (2018). *Modelamiento univariado del número de defunciones infantiles producidas por infecciones respiratorias agudas, a través de la metodología box-jenkins, puno 2008-2016*. Puno.
- Bernejo Mancera, A. (2011). *Metodos estadísticos de series temporales no lineales, con aplicacion a la predicción de energia eolica (Tesis de Doctoral)*. Departamento de Estadística universidad Carlos III de Madrid, Madrid.
- Betancourt, D. F. (7 de marzo de 2016). Obtenido de Medición del error en pronosticos de demanda: www.ingenioempresa.com/medicion-error-pronostico
- Bruce L, B., Richard T, O., & Anne B, K. (2007). *Pronóstico, series de tiempo y regresión. Cuarta Edición*. Mexico D. F.: Cengage Learning Editores, S.A. .
- Coelho, F. (08 de enero de 2018). *Significados.com*. Recuperado el 10 de septiembre de 2020, de <https://www.significados.com/municipio/>
- Fernandez, S. d. (2011). *Series temporales: Modelo Arima*. Madrid: Universidad Autonoma de Madrid.
- Frank, R. (1977). *Costos Adiministrativos de Maquinarias Agrícolas*. Buenos Aires: 1 ed. Edición Universitaria de Buenos Aires.
- Garcia Bellido, R., González Such, J., & Jornet Meliá, M. J. (2010). *SPSS: Pruebas no Parametricas*. Valencia: InnovaMIDE.
- Gilardi, J. (1977). *Reparacion de Motores de Tractores Agrícolas*. Lima Perú: Editorial IICA.
- Hernández Sampieri, R., Fernández Collado, C., & baptista Lucio, P. (1999). *Metodologia de la Investigación*. Colombia: Editorial Nomos S. A.



- Hernández, S. (2015). *Analisis de Series de Tiempo*. Mexico.
- Huillca Vargas, L. A. (2017). *Modelos de series de tiempo aplicado a los ingresos diarios de un Taller de Bicicletas en la Ciudad del Cusco 2016 (Tesis Pregrado)*. Universidad Nacional de San Antonio Abad Cusco Escuela Profesional de Matematica, Cusco.
- Kishurim, Tecnice, Tecnimat, Griduc, & Gidsaw. (2013). *El modelamiento matematico en la formación del ingeniero*. Bogota: L. F. M. Granados, Ed.
- Ljung, G., & Box, G. (1979). *On a Measure of Lack of Fit in Time Series Models*. Wesley.
- mclaranarvaez. (14 de febrero de 2017). *Slideshare*. Obtenido de <https://es.slideshare.net/mclaranarvaez/hans-gugelotescuola-de-ulm>
- Melo Mayta, E. (2016). *Modelo de Predicción Mensual de Mortalidad General Intrahospitalaria en el Hospital Regional Manuel Nuñez Butron - Puno, 2008 - 2016 - I (Tesis pregrado)*. Universidad Nacional del Altiplano Ingeniería Estadística e Informática, Puno.
- Mendez Lima, S. J. (05 de marzo de 2018). *Slideshare*. Obtenido de <https://es.slideshare.net/jaredmndz/gui-bonsiepe-metodologia>
- Modesto Arias, J. E. (2017). *Comportamiento de la Demanda de Depositos a Plazo Fijo en Moneda Nacional en las Cajas Municipales Periodo 2004 - 2016 (monografía)*. Universidad Nacional Agraria la Molina Facultad de Economía y Planificación, Lima.
- Morales Zapata, O. S. (2015). *Diseño de un sistema dearado para la preparación del suelo en cultivos organicos derigido a pequeños agricultores (Tesis de grado)*. Universidad Militar Nueva Granada Facultad en Mecatronica, Bogota.
- Munari, B. (1985). *DISEÑO Y COMUNICAICON VISUAL*. 8ª. Ed. Barcelona: G. Gili.



- Murillo, N. (1987). *Tractores y Maquinarias Agrícolas*. Costa Rica: 2 ed. Editorial EUNED.
- Murray Ralph, S., & Larry J, S. (2009). *Estadística Teoría y Problemas Resueltos. cuarta edición*. (Cuarta Edición ed.). (M. d. Roa, Trad.) Colombia: Schaum.
- Parra, F. (25 de enero de 2019). *Estadística y Machine Learning con R*. Obtenido de <https://bookdown.org/content/2274/series-temporales.html>
- Peña, D. (2005). *Análisis de series Temporales*. Alianza Editorial.
- Plackett, R. L. (1983). *International Statistical Review*. International Statistical Institute.
- Rios, G. (2008). *Series de Tiempo*. Universidad de Chile Facultad de Ciencias Físicas y Matemáticas., Chile.
- Rivera, D. P. (2010). *Análisis de series temporales*.
- Robert H., S., & David S., S. (September 2010). *Time Series Analysis And Its Applications*. Springer Texts in Statistics.
- Rodriguez, O. (24 de agosto de 2017). *Promidat Iberoamericano. programa iberoamericano de formacion de mineria de datos*. Obtenido de <https://www.promidat.com/promidat>
- Rugel Barreto, E. R. (2011). *Modelos de Pronostico de Estadísticas Vitales en la Provincia de Huaraz a Nivel de Centros Poblados 2011 (Tesis Pregrado)*. Universidad Nacional Santiago Antúnez de Mayolo Facultad de ciencias escuela profesional de Estadística e Infomática, Huaraz.
- Sarli, C. (2017). *Maestriadicom*. Obtenido de <https://maestriadicom.org/articulos/el-problema-de-disenar-en-la-complejidad-las-metodologias-y-sus-aportes-teoricos/>
- Tesis de Investigadores. (01 de octubre de 2011). *Tesis de Investigacion*. Obtenido de <https://tesisdeinvestig.blogspot.com/2011/10/caracteristicas-del-enfoque.html>



- Valadez Silvas, R. (15 de marzo de 2019). *Youtube*. Obtenido de <https://www.youtube.com/watch?v=2EA-wSWQopU>
- Villavicencio, J. (2010). *Introducción a Series de Tiempo*. Puerto Rico.
- Villazón Bustillos, D., Rubio Arias, H. O., & Ortega Gutierrez, J. A. (2015). Análisis en series de tiempo para el pronóstico de sequia en la región noroest del estado de Chihuahua. *Ecosistemas y Recursos Agropecuarios*.
- Wilfredo Toledo, D. (2013). *Una Introducción muy General a los modelos ARIMA*. Taller Junta de Planificación, Puesto Rico.

ANEXOS

ANEXO N° 01 Comandos en R Exploración de serie de datos.

```
#convertir Los datos en serie  
shora<-ts(hora[ ,3],start=c(2010,1), freq=12)  
#SERIE MENSUAL DESDE 2010 A 2019  
shora
```

Resumen de la serie de datos.

```
summary(hora$horas)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##      8.0   69.0   122.5   148.1  210.5   420.0
```

```
dygraph(shora, main =  
         "Horas de trabajo de maquinaria agricola de la Municipalidad de Caracoto  
         desde enero 2010 a diciembre 2019",xlab = "Año",ylab = "Horas")%>%  
dyOptions(fillGraph = TRUE, fillAlpha = 0.4)%>%  
dyOptions(drawPoints = TRUE,pointSize = 2)%>%  
dyRangeSelector()
```

```
# Descomponer la serie de tiempo shora. Se almacena en el objeto fit  
fit <- decompose(shora, type='additive')  
summary(fit)
```

```
##      Length Class  Mode  
## x      120    ts     numeric  
## seasonal 120    ts     numeric  
## trend    120    ts     numeric  
## random   120    ts     numeric  
## figure    12  -none- numeric  
## type       1  -none- character
```

ANEXO N° 02 Comando R Descomposición de la serie de datos

```
# Para graficar de la serie
autoplot(fit)+
  labs(title = "Descomposicion de la serie de tiempo",
       x = "Tiempo",
       y = "Horas",
       colour = "Gears")+
  theme_bw()
```

Grafica de la serie de datos con tendencia

```
# Grafica de la serie con tendencia
autoplot(shora, series="Serie tiempo") +
  autolayer(trendcycle(fit), series="Tendencia") +
  labs(title = "Serie de tiempo",
       x = "Tiempo",
       y = "Horas"
  ) +
  theme_bw()
```

ANEXO N° 03 Comando R Grafica de estacionalidad de la serie de datos

```
# Grafico de estacionalidad
ggseasonplot(shora,lwd=2.5)
```

Grafica de cajas de la serie de datos

```
#GRAFICA DE CAJAS
ggplot(hora, aes(x = year, y = horas))+
  theme_classic()+
  geom_boxplot(aes(col = year), lwd = 1)+
  coord_flip()
```

Histograma de la serie de datos

```
#HISTOGRAMA la serie (horas de trabajo mensual)
hist(shora, prob=T, col = "red",main="Histograma de horas trabajadas",
     xlab="Horas Mensuales")
lines(density(shora),lwd=2, col= "green")
me<-mean(shora)
me
```



```
y <- dnorm(x,me,st)  
lines(x,y,lwd=2,col="blue")
```

Grafica de test de normalidad de la serie de datos

```
#=====  
# Grafico de test de Normalidad  
qqnorm(shora)  
qqline(shora,col="red")
```

Test de Pearson de la serie de datos

```
#test de Pearson  
# Pearson chi-square normality test  
nortest::pearson.test(shora)
```

ANEXO N° 04 Comando R Análisis de la serie de datos

Comando Auto-arima de la serie shora

```
auto.arima(shora)
```

Primera diferencia de la serie shora

```
dhora<-diff(shora)  
#defhora <- data.frame(dhora)  
#write.table(defhora,file = "difserie.csv", sep =";", row.names = F)  
dhora
```

Grafica de cajas de la serie con primera diferencia

```
#GRAFICA DE CAJAS  
ggplot(hora, aes(x = year, y = dhora))+  
  theme_classic()+  
  coord_flip()+  
  geom_boxplot(aes(col = year), lwd = 1)+  
  stat_summary(fun.y=mean, col="gold")
```

Test de la primera diferencia



```
##  
## Pearson chi-square normality test  
##  
## data: dhora  
## P = 17.353, p-value = 0.09786
```

Auto-arima de la serie dhora con los posibles modelos

```
auto.arima(dhora, trace=TRUE)
```

Grafica de Auto Correlación Simple y Auto Correlación Parcial

```
# el mejor modelo generado por AUTO - ARIMA es un ARIMA(0,0,1)(0,1,0)[12]; de la serie  
con una diferencia.  
  
par(mfrow=c(2,1), mar=c(4,4,4,1+.1))  
  
#dividir el cuadro para las graficas  
acf(dhora, main="correlacion Simple - dhora",  
    ylim=c(-1,1),col="black")  
  
pacf(dhora, main="correlacion Parcial - dhora",  
     ylim=c(-1,1),col="black")
```

Partición de los datos de la serie Aprendizaje

```
ahora <- dhora[1:107]  
#Serie de arpende  
ahora <- ts(ahora, start =c(2010,2), freq = 12)  
ahora
```

Partición de los datos de la serie Testing

```
#serie de testin  
thora <- dhora[108:119]  
thora
```

ANEXO N° 05 Comando R Calibración de la serie de datos

Calibración el modelo con los datos de aprendizaje y testing

```
#ERROR RELATIVO (ER)
ER <- function(Pron, Real){
  return(sum(abs(Pron-Real))/abs(sum(Real)))
}
#ERROR CUADRATICO MEDIO (MSE)
ECM <- function(Pred, Real){
  N <- length(Real)
  SS <- sum((Real-Pred)^2)
  return(1/N*SS)
}
# PORCENTAJE DE FALLAS HACIA ARRIBA EN CONTEO (PFA)
PFA <- function(Pron, Real){
  Total <- 0
  N <- length(Pron)
  for(i in 1:N){
    if (Pron[i]>=Real[i])
      Total <- Total+1
  }
  return(Total/N)
}
# PORCENTAJE DE FALLAS HACIA ARRIBA EN ABSOLUTO (PTFA)
PTFA <- function(Pron, Real){
  Total <- 0
  SReal <- 0
  N <- length(Pron)
  for (i in 1:N) {
    if (Pron[i] >= Real[i]) {
      Total<-Total+(Pron[i]-Real[i])
      SReal<-SReal+abs(Real[i])
    }
  }
  if(Total==0)
    SReal = 1
  return(Total/SReal)
}
```

Calibrando los modelos con la Función Calibrar

```
# Funcion para calibrar los modelos
calibrar<-function(aprender,testing){
  error.c<-Inf
  p<-1
  while(p<=38) {
    mod.hora <- arima(x = aprender, order = c(tmod[p,3],tmod[p,4],tmod[p,5]),s
easonal=list(order=c(tmod[p,6],tmod[p,7],tmod[p,8]),period=12))
    pred.hora <- predict(mod.hora, n.ahead=12)
# CUADRO DE LOS ERRORES CUADRATICOS MEDIOS DE LOS MODELOS CALIBRADOS
    ecm.a <- sqrt(ECM(pred.hora$pred,testing))
    if(ecm.a<error.c){
      error.c<-ecm.a
      mejor<-mod.hora
    }
    errores<-print(paste(p,ecm.a,sep = " = "))
    p<-p + 1
  }
  return(mejor)
}
modelo<-calibrar(ahora,thora)
```



Calculamos los errores de los siguientes modelos propuestos

```
# SERIE ahora ARIMA(0,1,1)(0,1,0)[12]  
mod.ahora <- arima(x = ahora, order = c(0,1,1),seasonal=list(order=c(0,1,0),period=12))  
mod.ahora
```

```
pred.ahora <- predict(mod.ahora, n.ahead=12)  
pred.ahora
```

```
# SERIE chora AUTO-ARIMA(2,1,0)(1,1,1)[12]  
mod.chora <- arima(x = ahora, order = c(2,1,0),seasonal=list(order=c(1,1,1),period=12))  
mod.chora
```

```
pred.chora <- predict(mod.chora, n.ahead=12)  
pred.chora
```

```
# SERIE ihora AUTO-ARIMA(2,1,0)(0,0,0)[12]  
mod.ihora <- arima(x = ahora, order = c(2,1,0),seasonal=list(order=c(0,0,0),period=12))  
mod.ihora
```

```
pred.ihora <- predict(mod.ihora, n.ahead=12)  
pred.ihora
```

Comparar los modelos con el grafico de radar

```
# Cuadro de Medicion de Errores en pronosticos de series
errores<-rbind(verror.a,verror.c,verror.i)
rownames(errores)<-c("Errores Auto-arima","Errores Calibrada","Errores Planteado")
colnames(errores)<-c("Error Relativo","PFA","PTFA","Error Cuadratico Medio")
errores<-as.data.frame(errores)
maximos<-apply(errores,2, max)
minimos<-apply(errores,2, min)
errores<-rbind(minimos,errores)
errores<-rbind(maximos,errores)
errores
```

##	Error Relativo	PFA	PTFA	Error Cuadratico Medio
## 1	51.453036	0.5000000	0.2014242	221.04198
## 11	4.051082	0.1666667	0.1671467	17.53775
## Errores Auto-arima	4.069767	0.5000000	0.1918748	17.53775
## Errores Calibrada	4.051082	0.5000000	0.1671467	18.91927
## Errores Planteado	51.453036	0.1666667	0.2014242	221.04198

```
# CREANDO GRAFICO DE REDAR
radarchart(errores,maxmin = TRUE,axistype = 4,axislacol = "slategray4",
           centerzero = FALSE,seg = 8,cglcol = "gray67",
           pcol = c("green","blue","red"),
           plty = 1,
           plwd = 3,
           title = "Comparacion de Errores")
legenda<-legend(1.5,1,legend = c("Au-Arima","Calibrado","Planteado"),
               seg.len = -1.4,
               title = "Errores",
               pch = 21,
               bty = "n",lwd = 3,y.intersp = 1,horiz = FALSE,
               col = c("green","blue","red"))
```

ANEXO N° 06 Comando R Aplicamos el test de Ljung box Ruido blanco para los 24 coeficiente

```
#Test de Ljungbox
mejor.modelo <- arima(x = shora, order = c(2,1,0),seasonal=list(order=c(1,1,1),period=12))

retardos <- 24
P.va.Re <- c()
inter <- c()
pva<-0.05
n.corre<-rep(1:retardos,1)

for (i in 1:length(n.corre)) {
  lb<-Box.test(residuals(mejor.modelo),lag=i,type = "Ljung-Box")
  P.va.Re[i]<- lb$p.value
  if (pva<P.va.Re[i]) {
    inter[i]<-1
  }else{
    inter[i]<-0
  }
}
df.r<-data.frame(n.corre,P.va.Re,inter)
df.r
```

Test de diagrama del modelo

```
tsdiag(mejor.modelo)
```

Predicción de la serie de datos

```
# PREDICCION
# Analisis con los 10 year y prediccion con el modelo ARIMA(3,2,0)(1,1,1)[12]
prediccion<-forecast::forecast(mejor.modelo, h=12)
prediccion
```

```
plot(prediccion)
```

```
pred <- predict(mejor.modelo,n.ahead = 12)

preds<-pred$pred
LimInf<-preds-pred$se

LimSup<-preds+pred$se

per_1<-seq(as.Date("2010-1-1"),as.Date("2019-12-1"),"month")
per_2<-seq(as.Date("2020-1-1"),as.Date("2020-12-1"),"month")

todas.series<-cbind(shora=xts(shora,order.by=per_1),LimInf=xts(LimInf,order.by=per_2),Pronostico=xts(preds,order.by=per_2),LimSup=xts(LimSup,order.by=per_2))

dygraph(todas.series,main = "Prediccion de horas mensuales para el año 2020",ylab = "Cantidad de horas")>%
  dySeries(c("LimInf","Pronostico","LimSup"),label = "Horas")>%
  dyRangeSelector(height = 20,strokeColor = "")>%
  dyOptions(axisLineColor = "navy",gridLineColor = 'lightblue')
```