

UNIVERSIDAD NACIONAL PEDRO RUÍZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA



TESIS

“Modelos Arima univariante de series temporales para la producción y demanda de agua en el distrito de Lambayeque, periodo 2002 – 2017.”

PARA OPTAR TÍTULO PROFESIONAL DE:

LICENCIADO EN ESTADÍSTICA

AUTORES:

Bach. López Jiménez, Kevin Jherson

Bach. Villanueva Vásquez, Willian José

ASESOR:

Dr. Rodríguez López, Wilver Omero

LAMBAYEQUE, DICIEMBRE DEL 2019

UNIVERSIDAD NACIONAL PEDRO RUÍZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA



TESIS

“Modelos Arima univariante de series temporales para la producción y demanda de agua en el distrito de Lambayeque, periodo 2002 – 2017.”

PARA OPTAR TÍTULO PROFESIONAL DE:

LICENCIADO EN ESTADÍSTICA

AUTORES:

Bach. López Jiménez, Kevin Jherson

Bach. Villanueva Vásquez, Willian José

ASESOR:

Dr. Rodríguez López, Wilver Omero

LAMBAYEQUE, DICIEMBRE DEL 2019

UNIVERSIDAD NACIONAL PEDRO RUÍZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

TESIS

“MODELOS ARIMA UNIVARIANTE DE SERIES TEMPORALES PARA LA
PRODUCCIÓN Y DEMANDA DE AGUA EN EL DISTRITO DE
LAMBAYEQUE, PERIODO 2002 – 2017.”

Para optar el título de:

LICENCIADO EN ESTADÍSTICA

SUSTENTADO Y APROBADO ANTE LOS SIGUIENTES MIEMBROS DEL JURADO

Lic. Est. HUGO L. SAAVEDRA SAAVEDRA
PRESIDENTE



Dr. JORGE A. ACOSTA PISCOYA
SECRETARIO



M.Sc. FREDYS M. TRUJILLO CUSTODIO
VOCAL



**“MODELOS ARIMA UNIVARIANTE DE SERIES TEMPORALES PARA
LA PRODUCCIÓN Y DEMANDA DE AGUA EN EL DISTRITO DE
LAMBAYEQUE, PERIODO 2002 – 2017.”**

AUTORES:

Bach. KEVIN J. LÓPEZ JIMÉNEZ



Bach. WILLIAN J. VILLANUEVA VÁSQUEZ



DIRIGIDA POR:

Dr. WILVER O. RODRÍGUEZ LÓPEZ
ASESOR



DECLARACIÓN JURADA DE ORIGINALIDAD

Nosotros, investigadores principales: Kevin Jerson López Jiménez y Willian José Villanueva Vásquez, y Wilver Omero Rodríguez López asesor del trabajo de investigación “Modelos Arima Univariante de Series Temporales para la Producción y Demanda de agua en el Distrito de Lambayeque, Periodo 2002 – 2017”; declaramos bajo juramento que este trabajo no ha sido plagiado, ni contiene datos falsos. En caso se demostrará lo contrario, asumimos responsablemente la anulación de este informe y por ende el proceso administrativo a que hubiere lugar.

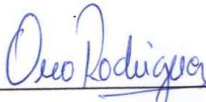
Lambayeque, diciembre 2019



Bach. López Jiménez, Kevin Jherson
AUTOR



Bach. Villanueva Vásquez, Willian José
AUTOR



Dr. Rodríguez López, Wilver Omero
ASESOR



UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DECANATO

Ciudad Universitaria - Lambayeque



ACTA DE SUSTENTACIÓN N° 071-2019-D/FACFyM

(Sustentación Autorizada por Resolución N° 1477-2019-D/FACFyM)

En la ciudad de Lambayeque, siendo las 4:00 pm del día 02 de Diciembre del 2019 se reunieron en la biblioteca del Laboratorio de Física de FacFyM los miembros del Jurado designados mediante Resolución N° 140-2019-D/FACFyM, los docentes:

Lic. Estad. Hugo Lorgio Saavedra Saavedra	Presidente
Dr. Jorge Antonio Acosta Piscoya	Secretario
M.Sc. Fredys Marino Trujillo Custodio	Vocal

Para recibir la tesis titulada:

Modelos Arima univariante de series temporales para la producción y demanda de Agua en el Distrito de Lambayeque periodo 2002-2017


desarrollada por los Bachilleres en Estadística, Villanueva Vásquez William José y López Jiménez Kevin Jherson.

Después de escuchar la exposición y las respuestas a las preguntas formuladas por los miembros del Jurado, se acordó Aprobar el trabajo por Unanimidad con el calificativo de Muy Bueno.

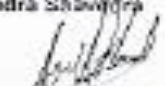
En consecuencia, los Bachilleres en referencia quedan aptos para recibir el Título Profesional de Licenciado en Estadística, de acuerdo a la Ley Universitaria, el Estatuto y Reglamento de la Universidad Nacional Pedro Ruiz Gallo de Lambayeque.

Observaciones:

Para constancia del hecho firman:


Lic. Estad. Hugo Lorgio Saavedra Saavedra
Presidente


Dr. Jorge Antonio Acosta Piscoya
Secretario


M.Sc. Fredys Marino Trujillo Custodio
Vocal

CERTIFICO: Que la presente reproducción
guarda conformidad con el original que se
me puso a la vista.
Lambayeque, **26 DIC 2019**



WALTER WENDACOCAY
NOTARIO - notario
Lambayeque



DEDICATORIA

El presente informe de tesis se la dedico con todo mi amor a mi madre Alicia Jimenez Araujo, por haberme forjado como la persona quien soy en la actualidad, y como mucho de mis logros que se los debo, incluyendo este. Sin olvidar a mi padre que en paz descansa Gilber Lopez Culqui que me guía y cuida en todo momento.

A mi abuela y familia por siempre darme sus consejos y buenos deseos de superación en cada etapa de mi vida profesional.

A mi amada novia Katherine Rucabado Llatas por ser mi compañera en toda etapa de mi carrera profesional, aunque hemos pasado momentos difíciles siempre ha estado brindándome su comprensión, cariño y amor.

A todos mis compañeros y docentes que siempre me apoyaron y permitieron entrar en su vida durante estos 5 años de mi carrera universitaria.

Kevin Jherson, Lopez Jimenez.

Esta tesis se la dedico con mucho cariño y amor a mi madre Elvia Vásquez y mi padre Gonzalo Villanueva, que son el pilar fundamental en todo lo que soy, por haberme dado la vida, por su apoyo incondicional, sus consejos, su comprensión y sobre todo que estaré eternamente agradecido con ustedes por promover el desarrollo y la unión familiar en nuestra familia.

A mis hermanos por apoyarme en aquellos momentos de necesidad por ayudar en la unión familiar que es lo más importante hoy en día. Ustedes por llenar mi vida de gratos momentos que hemos compartido durante todos estos años.

A la familia Ramirez Cabrera por apoyarme siempre durante este proceso de mi carrera profesional, les agradezco muchísimo. A mis amigos por todos los momentos que pasamos juntos. Por las tareas que juntos realizábamos en equipo, muchas gracias por la confianza que depositaron en mí.

Villanueva Vásquez, Willian José

AGRADECIMIENTO

Le agradecemos a Dios por habernos acompañado y guiado a lo largo de nuestra carrera, por ser nuestra fortaleza en los momentos de debilidad y por brindarnos una vida llena de aprendizaje, experiencias y sobre todo felicidad.

A nuestro asesor Dr. Wilver Omero Rodriguez López, por el apoyo brindado en el desarrollo de esta presente investigación. También por haber compartido sus conocimientos, su tiempo, su paciencia, comprensión, ser nuestro guía constante y sobre todo por su gran amistad.

A nuestro jurado Lic. Hugo Lorgio Saavedra Saavedra, Dr. Jorge Antonio Acosta Piscoya y M.Sc. Fredys Marino Trujillo Custodio, por brindarnos sus apreciaciones, sugerencias, observaciones y así haber podido culminar satisfactoriamente este proyecto. Además, por ayudarnos a tener una posición más crítica y razonable sobre nuestro problema de investigación.

A nuestros docentes de la Escuela profesional de Estadística, por enseñarnos a amar con pasión nuestra carrera profesional. También por brindarnos sus conocimientos impartidos durante toda su experiencia laboral, han sido nuestros guías para nuestra formación profesional.

A nuestra casa de estudios y Alma Mater Universidad Nacional “Pedro Ruiz Gallo”, por permitir estar durante los cinco años de estudio y habernos cedido un espacio en sus aulas para estudiar, practicar, aprender y sobre todo llegar a formar parte de la sociedad intelectual.

¡Muchas gracias!

ÍNDICE

CAPITULO I: INTRODUCCIÓN	12
1.1 Planteamiento del Problema	12
1.2 Formulación del Problema Científico	15
1.3 Hipótesis.....	15
1.4 Objetivos.....	15
1.4.1 Objetivo general	15
1.4.2 Objetivos específico.....	15
1.5 Justificación e Importancia	16
CAPITULO II: DISEÑO TEÓRICO.....	17
2.1 Antecedentes	17
2.2 Base Teórica.....	21
CAPITULO III: MÉTODOS Y MATERIALES	51
3.1 Diseño de Investigación	51
3.2 Tipo de Investigación	51
3.3 Materiales de Estudio	51
3.4 Técnicas e Instrumentos de Recolección de Datos	52
3.5 Análisis Estadístico de los Datos	52
IV. RESULTADOS Y DISCUSION.....	53
1. Aplicación de la Metodología Box-Jenkins para la producción de agua.....	53
1.1 Análisis de la estacionariedad de los datos.....	54
1.2 Identificación de los modelos.....	57
1.3 Estimación de los parámetros de los modelos tentativos	58
1.4 Validación del modelo.....	60
1.5 Pronóstico del modelo.....	63
2. Aplicación de la metodología Box- Jenkins para la demanda de agua.....	67
2.1 Análisis de la estacionariedad de los datos.....	68
2.2JN Identificación de los modelos.....	74
2.3 Estimación de los parámetros de los modelos tentativos	75
2.4 Validación del modelo.....	77
2.5 Pronóstico del modelo	80
CONCLUSIONES Y SUGERENCIAS.....	84
RECOMENDACIONES	86
BIBLIOGRAFÍA	87
APÉNDICE Y ANEXOS.....	91

Índice de Tabla

Tabla 1: Estadísticos descriptivos de la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	53
Tabla 2: Prueba de raíz unitaria de Dickey - Fuller para la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	57
Tabla 3: Prueba t para una muestra para la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	58
Tabla 4: Resumen de los modelos tentativos identificados de la serie de producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	59
Tabla 5: Prueba de normalidad para los residuales del modelo SARMA (1, 0, 2) (1, 0, 0).....	60
Tabla 6: Prueba de aleatoriedad para los residuales del modelo SARMA (1, 0, 2) (1, 0, 0).....	61
Tabla 7: Pronostico de la producción de agua en el distrito de Lambayeque, para el periodo enero 2018 a diciembre 2018.	63
Tabla 8: Pronostico de la producción de agua en el distrito de Lambayeque, para el periodo enero 2019 a diciembre 2019.	65
Tabla 9: Estadísticos descriptivos de la demanda de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	67
Tabla 10: Prueba de raíz unitaria de Dickey - Fuller para la demanda de agua del distrito de Lambayeque periodo enero 2002 a diciembre 2017.	69
Tabla 11: Prueba de raíz unitaria de Dickey - Fuller para la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.	73
Tabla 12: Prueba t para una muestra para la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.	75
Tabla 13: Resumen del desempeño de los modelos tentativos identificados de la serie de la demanda de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	76
Tabla 14: Prueba de normalidad para los residuales del modelo ARIMA (0, 1, 1)	77
Tabla 15: Prueba de aleatoriedad para los residuales del modelo ARIMA (0,1,1)	78
Tabla 16: Pronostico de la demanda de agua en el distrito de Lambayeque, para el periodo enero 2018 a diciembre 2018.	80
Tabla 17: Pronostico de la demanda de agua en el distrito de Lambayeque, para el periodo enero 2019 a diciembre 2019.	82
Tabla 18: Valores del Correlograma de la serie de la Producción de Agua en el distrito de Lambayeque. Periodo 2002-2017.	93
Tabla 19: Valores del Correlograma de la serie de la Demanda de Agua en el distrito de Lambayeque. Periodo 2002-2017.	94
Tabla 20: Valores del Correlograma de la serie a primera diferencia de la Demanda de Agua en el distrito de Lambayeque. Periodo 2002-2017.	95

Índice de Figura

Figura 1: Serie original de la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.....	54
Figura 2: Función de Autocorrelación Simple de la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.....	55
Figura 3: Función de Autocorrelación Parcial de la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.....	56
Figura 4: Función de Autocorrelación simple y parcial de los residuales del modelo SARMA (1, 0, 2) (1, 0, 0) de la producción de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017.....	62
Figura 5: Serie de la producción de agua en el distrito de Lambayeque, con el modelo SARMA (1, 0, 2) (1, 0, 0), periodo enero 2002 a diciembre 2017.....	64
Figura 6: Serie de la producción de agua en el distrito de Lambayeque, con el modelo SARMA (1, 0, 2) (1, 0, 0), periodo enero 2002 a diciembre 2018.....	66
Figura 7: Serie original de la demanda de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	68
Figura 8: Diferencia en 1 periodo de la demanda de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	70
Figura 9: Función de Autocorrelación Simple de la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.	71
Figura 10: Función de Autocorrelación Parcial de la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.	72
Figura 11: Función de Autocorrelación simple y parcial de los residuales del modelo ARIMA (0, 1, 1) de la demanda de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017.	79
Figura 12: Serie de la demanda de agua en el distrito de Lambayeque, con el modelo ARIMA (0, 1, 1) periodo enero 2002 a diciembre 2017.	81
Figura 13: Serie de la demanda de agua en el distrito de Lambayeque, con el modelo ARIMA (0, 1, 1) periodo enero 2002 a diciembre 2019.	83
Figura 14: Diagrama de cajas y bigotes de la producción de agua en el distrito de Lambayeque. Periodo 2002 a 2017.	105
Figura 15: Diagrama de cajas y bigotes de la demanda de agua en el distrito de Lambayeque. Periodo 2002 a 2017.	106
Figura 16: Distribución del número de habitantes del distrito de Lambayeque. Periodo 2000 a 2019.	107
Figura 17: Consumo per cápita en litros de agua por habitante del Distrito de Lambayeque. Periodo 2002 a 2019.	107

RESUMEN

El presente estudio tuvo como objetivo principal determinar el modelo univariante que permita predecir el comportamiento de la producción y demanda de agua en el distrito de Lambayeque, periodo 2002 al 2017. La investigación realizada es descriptiva predictiva de tipo longitudinal retrospectiva.

Los datos sobre producción y demanda en el periodo 2002 al 2017, fueron tomados del Instituto Nacional de Estadística e Informática (INEI).

El análisis de los datos de la serie, se realizó mediante la metodología de Box – Jenkins, para identificar el modelo que mejor se adecue a los datos observados. En función a ello se determinó:

El mejor modelo que explica el comportamiento de la producción de agua en el periodo indicado es el modelo SARMA (1,0,2) (1,0,0), con un Error Cuadrático Medio = 16 932.67, con Error Absoluto Medio = 13 254.61, el Error Porcentual Absoluto Medio = 3.98% y con coeficientes estimados AR (1) = 0.608, MA (1) = 0.299, MA (2) = -0.248, SAR (1) = 0.215. Mientras que el modelo ARIMA (0,1,1), con un Error Cuadrático Medio = 3 062.74, con Desviación Absoluta de la media = 2 303.09, el Porcentaje de Error Medio Cuadrado Absoluto = 1.44% y con coeficientes estimados MA (1) = 0.377 es el que explica mejor el comportamiento de la demanda de agua en el periodo analizado.

SUMMARY

The main objective of this study was to determine the univariate model that allows predicting the behavior of water production and demand in the district of Lambayeque, period 2002 to 2017. The research carried out is descriptive of retrospective longitudinal type.

The data on production and demand in the period 2002 to 2017, were taken from the National Institute of Statistics and Informatics (INEI).

The analysis of the series data was carried out using the Box - Jenkins methodology, to identify the model that best suits the observed data. Based on this, it was determined:

The best model that explains the behavior of water production in the indicated period is the SARMA model (1,0,2) (1,0,0), with a Mean Square Error = 16 932.67, with Average Absolute Error = 13 254.61, the Average Absolute Percentage Error = 3.98% and with estimated coefficients AR (1) = 0.608, MA (1) = 0.299, MA (2) = -0.248, SAR (1) = 0.215. While the ARIMA model (0,1,1), with a Mean Square Error = 3 062.74, with Absolute Deviation from the mean = 2 303.09, the Absolute Square Mean Error Percentage = 1.44% and with estimated MA coefficients (1) = 0.377 is the one that best explains the behavior of the water demand in the analyzed period.

CAPITULO I: INTRODUCCIÓN

1.1 Planteamiento del Problema

Contexto Internacional

El agua parte importante de los ecosistemas naturales y principal elemento para el sostenimiento de la vida en la tierra, representa el 70% de la superficie terrestre equivalente a un volumen total de 1 400 millones de kilómetros cúbicos, de los cuales el 97.5% es agua salada que no se puede beber y se encuentra en mares y océanos, 2.5% es apta para el consumo humano. Un recurso limitado cuya demanda, según las previsiones de la Organización para la Cooperación y el Desarrollo Económico (OCDE), se disparará globalmente hasta un 55% entre los años 2000 y 2050 (Boggiano, 2013-2017).

En un estudio hecho a 93 países en desarrollo, se mostró que sólo diez países utilizan el 40% de su agua dulce renovable de las cuales Asia aún sigue siendo la mitad del promedio mundial en agua dulce. En términos de escasez de agua a nivel mundial se determinó que EE. UU, Europa y África la escasez de este recurso es muy significativa (Organización de Naciones Unidas, 2013 y 2016).

Según Forbes (2019), tres países en Latinoamérica están en peligro de quedarse sin agua a nivel mundial en las próximas décadas. Todo indica que ante el estrés hídrico que registran los mantos acuíferos. Sao Paulo, Lima y Ciudad de México son las capitales que deben promover la seguridad hídrica del país, el riesgo es inminente donde se extrae generalmente más agua de la que se produce pues esto afecta el ciclo hídrico.

El 94% de la población de América del sur y centro, tienen acceso a fuentes de agua dulce (Organización de Naciones Unidas, 2013).

El 42% de la población activa dependen del agua, además es un componente sustancial de las economías internacionales, nacionales y locales (Organización de Naciones Unidas, 2016). La demanda del agua dulce en el mundo está en constante incremento y los recursos son cada vez más escasos (Bokava et al, 2017).

Contexto Nacional

El Perú es un país privilegiado por su gran volumen hídrico, que dispone alrededor de 2 046 287 Millones de Metros Cúbicos (MMC) de agua, ubicándose entre los 20 países más ricos del mundo (Ministerio de Agricultura y Riego, 2009). Tiene 159 unidades hidrográficas, para el desarrollo en las distintas formas de vida, ya sea agrícola, industrial, domestica. etc. (Instituto Nacional de Estadística e Informática, 2014). Más de 16 millones peruanos cuentan con acceso a sistemas de alcantarillo y recolección de aguas residuales (Superintendencia Nacional de Servicios de Saneamiento, 2015).

Sin embargo, al 2020 la población se incrementará en un 30%, esto ocasionará una sobreexplotación de las cuencas y acuíferos, dando como resultado un crecimiento de la demanda de agua (Ministerio de Agricultura y Riego, 2009). El estado peruano ha concesionado el diseño, financiamiento, construcción, operación y mantenimiento de las PTAR (Planta de tratamiento de aguas residuales) en Taboada y La Chira, será el más grande del país. Pero el gran problema que afecta el rendimiento de las PTAR son los vertimientos informales de efluentes industriales a los sistemas de alcantarillado. Además, introducen cargas orgánicas elevadas y compuestos tóxicos que afectan negativamente los procesos biológicos del tratamiento PTAR (Organización Naciones Unidas, 2017).

En el caso de Lima tiene una producción de agua de 1,81 MMC/día y el consumo promedio es de 250L/habitante al día. Las pérdidas de agua dulce en dicha región son aproximadamente del 40% (Reinhard, 2009).

Es importante remarcar que en todo el Perú el agua es un bien cada vez más limitado, pues en diferentes regiones hay presencia de un déficit hídrico muy alto, problemática que se presenta en lugares como Amazonas, Inkahuasi, etc. (Pimentel y Palacios, 2017).

Contexto Local

En Lambayeque más de 390 mil habitantes no tienen acceso a los servicios de saneamiento básico de agua potable o alcantarillado, donde el 35% son de zonas urbanas y el resto residen en distritos altos andinos, esto se debe al crecimiento de la población (Ramos, 2013). Aproximadamente hay 457 pueblos jóvenes o AA. HH sin acceso al agua potable, esto se debe al desorden del catastro urbano e implica que las ciudades continúen sin dirección o a un ritmo incongruente (Gordillo, 2015).

El estado peruano está invirtiendo en proyectos para mejorar y abastecer agua potable en la región, pues se desea cubrir la necesidad de agua potable durante las 24 horas al día, ya que hay zonas que no cuentan con acceso a este recurso (Camasca, 2017). Además, EPSEL no cuenta con un modelo de pronóstico que permita cuantificar las necesidades del agua potable.

Acorde a la problemática observada en la población Lambayecana, la presente investigación tiene como objetivo principal determinar los modelos univariantes óptimos que permitan predecir el comportamiento de la producción y demanda de agua en el distrito de Lambayeque en el periodo 2002 – 2017.

1.2 Formulación del Problema Científico

¿Cuáles son los modelos univariantes óptimos que permiten predecir el comportamiento de la producción y demanda de agua en el distrito de Lambayeque, mediante la metodología de Box Jenkins; con los datos del periodo 2002 al 2017?

1.3 Hipótesis

El modelo univariante integrado de Box Jenkins proporciona un mejor modelo de ajuste que el modelo no integrado de Box Jenkins en la producción y demanda de agua del distrito de Lambayeque, periodo 2002 – 2017.

1.4 Objetivos

1.4.1 Objetivo general

Determinar el modelo univariante que permitan de predecir el comportamiento de la producción y demanda de agua en el distrito de Lambayeque, mediante la metodología de Box Jenkins; periodo 2002-2017.

1.4.2 Objetivos específico

- Analizar o evaluar los componentes de series producción y demanda de agua en el distrito de Lambayeque, periodo 2002 – 2017.
- Construir los posibles modelos de predicción para la producción y demanda de agua en el distrito de Lambayeque, periodo 2002 – 2017.
- Predecir el comportamiento de las series de producción y demanda de agua en el distrito de Lambayeque, para el periodo enero 2018 a diciembre 2019.

1.5 Justificación e Importancia

El agua potable es indispensable para la vida y es escasa a nivel mundial, sobre todo en diferentes regiones del Perú donde inclusive nos enfrentamos a la falta de saneamiento de agua potable o alcantarillado. La oferta hídrica se ve afectada por las grandes pérdidas de agua en las redes urbanas, esto genera un problema para la demanda del agua ya que no satisface las necesidades de la población.

En la región Lambayeque el consumo de agua potable ha aumentado a cambio de la expansión de territorio o asentamientos humanos, los cambios meteorológicos, los factores demográficos, la falta de conciencia social frente al desperdicio del agua, etc.

Frente a estos acontecimientos sobre el problema del agua potable y las necesidades de los ciudadanos lambayecanos, nace esta investigación para poder facilitar un modelo óptimo de pronóstico de la producción y demanda de agua en el distrito de Lambayeque, en este caso el modelo ARIMA basado en la metodología de Box y Jenkins que nos permitirá cuantificar las necesidades de agua potable en la población de estudio.

Además, este proyecto también se justifica porque la Entidad Prestadora de Servicios de Saneamiento de Lambayeque (EPSEL) no cuenta con una metodología que explique la evolución histórica y proyección de la producción y demanda de agua dulce.

Los resultados de la investigación aportarán a la toma de decisiones de la empresa responsable del servicio, contribuyendo a que la producción sea más eficiente, la demanda más rentable y la distribución de agua potable sea más justa, mejorando la calidad de servicio, beneficiando a la empresa EPSEL y las personas interesadas en investigar la producción y demanda de agua en diferentes puntos del país.

CAPITULO II: DISEÑO TEÓRICO

2.1 Antecedentes

Contexto Internacional

Daza (2008). Realizó una investigación titulada “Demanda de Agua en Zonas Urbanas de Andalucía”, en la que utilizó la técnica de Redes Neuronales Artificiales (RNA) y la metodología de Box – Jenkins o modelos ARIMA, que implementadas en un modelo híbrido permiten conjugar sus ventajas y así explicar el comportamiento y predecir a corto plazo la demanda urbana de agua, consiguiendo unos resultados más satisfactorios que la utilización de los modelos anteriores por separado. El objetivo fue desarrollar una metodología de predicción del consumo de agua urbana durante el periodo de enero 1984 hasta diciembre de 2007, y concluye que el principal objetivo de este estudio se ha cumplido, consiguiendo resultados más satisfactorios para el modelo híbrido construido, siendo susceptible de utilización como instrumento de predicción del consumo de agua, con todos los beneficios que este avance implica para el desarrollo de políticas de aguas integradas sobre la base de la sostenibilidad de los usos presentes y futuros y la garantía de abastecimiento a la población.

Murat y Hamidreza (2015). Realizó una investigación titulada “Un Marco para la Gestión Sostenible del Agua Urbana a través de oferta y demanda de pronóstico: el caso de Estambul”, el objetivo fue analizar la demanda y suministro de agua, para evaluar la sostenibilidad de abastecimiento de agua hasta el 2018, los datos recogidos son de oferta y demanda de agua de Estambul en el periodo 2006 y 2014. Utilizando un modelo autorregresivo de medias móviles (ARIMA) se construye la previsión para el periodo entre 2015 y 2018. Los resultados de los pronósticos indican que los proyectos de tuberías serán fundamentales en el futuro próximo debido a los incrementos esperados en la demanda total de agua en Estambul.

Cuyo, J. (2016). Realizó una investigación titulada “Propuesta metodológica para el pronóstico de la demanda de agua potable a corto plazo (días, semanas, meses) en la planta de tratamiento de agua potable de Bellavista en el norte de la ciudad de Quito”, su objetivo fue proponer una metodología eficaz para realizar pronósticos de corto plazo (días, semanas, meses) de la demanda de agua potable con la información disponible en la zona norte de la ciudad de Quito que es atendida por la PTAP de Bellavista, se utilizaron variables exógenas climatológicas (precipitación y temperatura) y la variable caudal para el pronóstico de la demanda de agua potable con la metodología VAR y con las series de tiempo del caudal se procedió a aplicar las diferentes metodologías ARIMA, los resultados obtenidos con la metodología ARIMA fueron favorables al aplicarla en la predicción a corto plazo a nivel semanal y mensual, a nivel diario los modelos VAR tienen una mayor concordancia con los valores reales, existiendo diferencias mínimas en la comparación de los dos pronósticos.

Contexto Nacional

Vidal, G. (2015). Realizó una investigación titulada “Modelo SARIMA para el pronóstico de la masa mensual del río Jequetepeque”, el objetivo es obtener un modelo adecuado que describa el comportamiento de la masa mensual del río Jequetepeque en la región La Libertad, utilizando información de la gerencia Regional de Agricultura desde el periodo de enero de 1994 a diciembre del 2014. La metodología aplicada en el trabajo fue la propuesta por Box - Jenkins y la serie se dividió en: enero del 1994 a diciembre del 2013 para la estimación de modelo y de enero a diciembre del 2014 para la validación del pronóstico. El modelo de pronóstico estimado para la masa mensual del río Jequetepeque de enero del 1994 a diciembre del 2013 es un SARIMA (1,0,0,) (3,0,0).

Leonardo, J. (2017). Realizó una investigación titulada “Modelo Univariante para el Consumo Doméstico mensual de Agua Potable en el Distrito de Ilave – EMSA Puno, periodo 2002-2013”, el objetivo fue determinar un modelo univariante de ajuste que permita describir y predecir el consumo doméstico mensual de agua potable en el distrito Ilave – EMSA Puno, los datos de la muestra estuvo conformada por el Consumo mensual de Agua Potable (m³) del Distrito de Ilave en el periodo 2002 – 2013, recopilados por la Empresa Municipal de Saneamiento (EMSA Puno). Utilizando la metodología Box – Jenkins, se logró conseguir un modelo univariante que describa y ajuste a los datos. Los resultados fueron, que los modelos univariantes integrados proporcionan un mejor ajuste para la serie Consumo mensual de Agua Potable de Ilave en el periodo 2002 – 2013.

Contexto Local

Vásquez (2017). En su investigación titulada “Modelos de Box Jenkins y Redes Neuronales para pronosticar el precio del dólar del sistema bancario en moneda nacional año 2015”, el objetivo fue estimar un modelo de Box Jenkins y Redes Neuronales para pronosticar el precio del dólar del sistema bancario en moneda nacional año 2015, se estimó el modelo mediante la metodología de Box y Jenkins, de la compra y venta del dólar en moneda nacional de enero a diciembre para el año 2015, es un modelo ARIMA (2,2,3) para la compra y ARIMA (3,1,2) para la venta.

Siesquén y Riojas (2017). En su investigación titulada “Modelos de pronóstico de precios al consumidor región de Lambayeque y a nivel nacional. Enero 2010 a Diciembre 2016”, el objetivo fue construir los modelos de pronóstico del índice de precios al consumidor de la región Lambayeque y a nivel nacional. Enero 2010 a Diciembre 2016, estimó un modelo de pronóstico obtenido por el modelo determinado ARIMA (0, 1, 1) para el Índice de Precios al Consumidor región Lambayeque y los datos observados reales, donde tienden a estar cercanos del valor real a medida que se prolonga el tiempo con un menor Error Absoluto Promedio MAE = 0.170 por mes.

Damián y Sandoval (2018). En su investigación titulada “Modelos óptimos de pronóstico del índice mensual de pronóstico de electricidad. Perú en el periodo 2006-2015”, el objetivo fue determinar un modelo óptimo de pronóstico del índice mensual de producción de electricidad. Perú en el periodo 2006-2015, estimó un modelo de Índice Mensual de Producción de Electricidad en el Perú de acuerdo a la validez del pronóstico: SARIMA (1,1,0) (0,1,1), cuya ecuación fue: $Y_t = 0.5246Y_{t-1} + 0.4754Y_{t-2} + Y_{t-12} - 0.5246Y_{t-13} - 0.4754Y_{t-14} + e_t - 0.1e_{t-12}$, AR(1)= - 0.4754, SMA(12)= 0.7768.

2.2 Base Teórica.

2.2.1 Series temporales

Mason, Lind y Marchal (2001) define que una serie de tiempo es el conjunto de datos registrados durante un periodo que puede ser semanas, meses, trimestres o años (p. 647).

Spiegel y Stephens (2002) sostiene que una serie de tiempo es el conjunto de observaciones que el transcurso del tiempo es tomado en momentos específicos y pueden ser en intervalos iguales (p. 436).

En las matemáticas una serie de tiempo se define por un medio de valores y_1, y_2, \dots, y_N de una variable Y en los tiempos t_1, t_2, \dots, t_N . Además, Y es una función de t , esto se denomina por $Y = F(t)$ (p. 436).

Mauricio (2007) define a una serie temporal como una secuencia de N observaciones ordenadas y equidistantes cronológicamente sobre una característica o sobre varias características de una unidad observable en diferentes momentos (p.1 - 2).

Representaciones matemáticas frecuentes de series temporales univariantes:

$$y_1, y_2, \dots, y_N; (y_t)_{t=1}^N; (y_t: t = 1, \dots, N)$$

Donde y_t es la observación en t ($1 \leq t \leq N$) de la serie y N es el número de observaciones de que consta la serie completa (el tamaño o la longitud de la serie).

Las N observaciones y_1, y_2, \dots, y_N pueden recogerse en un vector columna:

$$y \equiv [y_1, y_2, \dots, y_N]'$$
 de orden $N \times 1$.

2.2.2 Procesos Estocásticos

Mauricio (2007) sostiene que un proceso estocástico es una secuencia de variables aleatorias, ordenadas y equidistantes cronológicamente, referidas a una (proceso univariante o escalar) o varias (proceso multivariante o vectorial) características de una unidad observable en diferentes momentos (p.09).

Wooldridge, J. (2010) considera que el proceso estocástico es una secuencia de variables aleatorias indexadas en el tiempo. Una base de datos de series de tiempo, se obtiene un resultado posible o realización. Se puede ver una sola realización, más no es posible retroceder en el tiempo y empezar de nuevo el proceso.

Además, el análisis de corte transversal en el que únicamente se puede reunir una sola muestra aleatoria. Por otro lado, el tamaño de muestra para una base de datos de series de tiempo es el número de periodos durante los cuales se observan las variables de interés (p. 341 a 342).

Representaciones matemáticas frecuentes de procesos estocásticos univariantes:

$$\dots, Y_{-1}, Y_0, Y_1, Y_2, \dots; (Y_t; t = 0, \pm 1, \pm 2, \dots); (Y_t)$$

Donde Y_t es una variable aleatoria escalar referida a la unidad observable considerada en el momento t .

Proceso estocástico: $\dots, Y_{-1}, Y_0, Y_1, Y_2, \dots, Y_N, Y_{N+1}, \dots$

Serie temporal: Y_1, Y_2, \dots, Y_N

En general, una serie temporal se refiere a un periodo muestral que tan sólo es una parte de la historia del proceso estocástico del que procede dicha serie.

2.2.3 Proceso de ruido blanco

González, M. (2009) sostiene que es una secuencia de variables aleatorias de media cero, varianza constante y covarianzas nulas (p. 18).

Se denotará habitualmente por a_t , $t = 0, \pm 1, \pm 2, \dots$

$$E(y_t) = 0, \forall t \quad V(y_t) = \sigma^2, \forall t \quad Cov(y_t y_s) = 0, \forall t \neq s$$

Un proceso ruido blanco $a_t \sim RB(0, \sigma^2)$ es estacionario si la varianza σ^2 es finita con función de autocovarianzas (FACV):

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

Y función de autocorrelación (FAC):

$$\gamma_k = 1, k = 0 \quad y \quad \gamma_k = 0, k > 0$$

Novales (1993) define que se llama ruido blanco $\{\varepsilon_t\}$ a una sucesión de variables aleatorias si: (p. 414-415).

$$E(y_t) = 0 \quad \text{para } t = 1, 2, \dots$$

$$V(y_t) = \sigma^2 < \infty \quad \text{para } t = 1, 2, \dots$$

$$E(e_t e_{t-j}) = Cov(e_t e_{t-j}) = 0 \quad \text{para } t = 1, 2, \dots j \neq 0$$

2.2.4 Paseo aleatorio

Novales (1993) define que el modelo paseo aleatorio es un proceso estocástico $\{y_t\}$ cuyas primeras diferencias forman un proceso de ruido blanco, es decir: (p. 414-415).

$$y_t = y_{t-1} + \varepsilon_t, \quad t = -\infty, \dots, -2, -1, 0, 1, 2, \dots$$

Donde ε_t es un ruido blanco.

González, M. (2009) describe que el modelo de paseo aleatorio es simplemente un modelo AR(1) con parámetros $\phi = 1$. Además, el valor de Y en el momento t es igual a su valor en el momento $t-1$ más una perturbación aleatoria (p.51).

2.2.5 Estacionariedad

Novales (1993) sostiene que un proceso estocástico y_t es estacionario en sentido estricto si para toda m-tupla (y_1, y_2, \dots, y_m) todo entero k es el vector de variables $(y_{t_1}, y_{t_2}, \dots, y_{t_m})$ tiene la misma distribución de probabilidad conjunta que el vector $(y_{t_1+k}, y_{t_2+k}, \dots, y_{t_m+k})$ (p. 415).

González, M. (2009) define que hay dos formas de estacionariedad. La estacionariedad de covarianza (conocido también como débil) y estacionariedad estricta. Además, la más importante es la estacionariedad en covarianza porque modela la media de un proceso. Por otro lado, la estacionariedad estricta es útil en tornos más difíciles, como los modelos no lineales (p. 13).

2.2.5.1 Estacionariedad de covarianza

Las series de tiempo es llamado estacionaria débil cuando sus propiedades estadísticas pertenecen constantes en el tiempo. Esto quiere decir que los diferentes intervalos de tiempo, las covarianzas muestrales y la media muestral de una serie durante los dos intervalos de tiempo será lo mismo.

Las condiciones se cumplen precisamente en una serie de tiempo y_t es llamada estacionaria en el sentido débil (p. 13).

$$E(Y_t) = 0 \quad \text{para } t = 1, 2, \dots$$

$$V(Y_t) = \sigma^2 < \infty \quad \text{para } t = 1, 2, \dots$$

$$E[(Y_t - \mu)(Y_{t-s} - \mu)] = Y_s \quad \text{para } t = 1, 2, \dots$$

2.2.5.2 Estacionariedad estricta

Es un proceso estocástico $\{y_t\}$ es estrictamente estacionario si la distribución conjunta $(y_1, y_{t+1}, \dots, y_{t+h})$ solo depende de h y no de t.

La distribución conjunta de un proceso estocástico es que no depende del tiempo y el único factor que afecta es la relación entre dos observaciones

es la brecha entre ellos (p.14).

2.2.6 Componentes

Los movimientos característicos en serie de tiempo o también llamados componentes de las series de tiempo que se clasifican en cuatro tipos principales.

2.2.6.1 Movimiento a largo plazo o seculares o tendencia

Mason, Lind y Marchal (2001) define que las tendencias a largo plazo en variables Y que se ajustan en diversos esquemas sociales por consiguiente se mueven hacia arriba, otras declinan y otras permanecen igual en un cierto periodo o intervalo de tiempo. Además, la tendencia a largo plazo sin alteraciones de una serie de tiempo (p. 647).

Spiegel y Stephens (2002) sostiene que el movimiento secular está indicado por una resta de tendencia, para otras series de tiempo puede ser la más adecuada una curva de tendencia (p. 437).

2.2.6.2 Movimientos cíclicos o variaciones cíclicas

Mason, Lind y Marchal (2001) define el ascenso y descenso de una serie de tiempo en periodos mayores de un año (p. 649).

Spiegel y Stephens (2002) sostiene que las oscilaciones o movimientos de una recta o curva de tendencia. Es decir, que se les denomina periódicos o no periódicos que pueden o no seguir patrones exactamente similares en intervalos iguales de tiempo. Además, se considera cíclico cuando se repiten después de intervalos mayores de un año (p. 437).

2.2.6.3 Movimientos estacionales o variaciones estacionales

Mason, Lind y Marchal (2001) define que es un esquema (patrones) de cambio en una serie de tiempo de uno año. Además, los patrones tienden a repetirse cada año (p.650).

Spiegel y Stephens (2002) sostiene que los patrones idénticos o casi idénticos que la serie de tiempo deben seguir los meses o trimestres que corresponden a los años sucesivos (p. 437).

2.2.6.4 Movimientos irregulares o aleatorios o erráticos

Mason, Lind y Marchal (2001) manifiesta que deben subdividir la variación irregular en variaciones episódicas y residuales. Las episódicas no son predecibles, pero pueden identificarse. Por otro lado, ninguna variación episódica o residual puede proyectarse a futuro (p.650).

Spiegel y Stephens (2002) sostiene que los movimientos esporádicos de las series de tiempo son eventos aleatorios a diferentes variables Y (p. 437).

2.2.7 Procesos estocásticos estacionarios

Damodar y Dawn (2010). Un proceso estocástico que ha recibido gran atención y ha sido objeto de escrutinio por parte de los analistas de series de tiempo es el proceso estocástico estacionario. (p. 740).

En términos generales, se dice que un proceso estocástico 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 (p. 740).

$$\text{Media} \quad E(y_t) = E(y_{t+k}) = \mu$$

$$\text{Varianza} \quad V(y_t) = E(y_{t+k}) = \mu$$

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

2.2.8 Función de autocovarianzas y de autocorrelación

2.2.8.1 Función de autocovarianzas (FACV)

Novales (1993) manifiesta que la función de autocovarianzas de un proceso estocástico estacionario $\{y_t\}$ es la función, a lo que nos referimos como FAC, es por ello que t y cada número entero k toma un valor, denominado $y_t(t)$, igual a la covarianza entre y_{t-k} (p.416).

$$Y_k(t) = cov(Y_t, Y_{t-k}) \quad \text{para } k = 0, 1, 2, \dots$$

González, M. (2009). La función de autocovarianzas de un proceso estocástico estacionario es una función de k (número de periodos de separación entre variables) que recoge el conjunto de las autocovarianzas del proceso (p. 15).

Características de la función de autocovarianzas:

- Incluye la varianza del proceso para

$$k = 0: \gamma_0 = E[Y_t - \mu][Y_t - \mu] = V(Y_t)$$

- Es una función simétrica:

$$\gamma_k = E[Y_t - \mu][Y_{t+k} - \mu]$$

$$\gamma_k = E[Y_t - \mu][Y_{t-k} - \mu] = \gamma_{-k}$$

La función de autocovarianzas de un proceso estocástico recoge toda la información sobre la estructura dinámica lineal del mismo. Pero depende de las unidades de medida de la variable, por lo que, en general, se suele utilizar la función de autocorrelación (p. 15).

2.2.8.2 Función de autocorrelación (FAC)

González, M. (2009). El coeficiente de autocorrelación de orden k de un proceso estocástico estacionario mide el grado de asociación lineal existente entre dos variables aleatorias del proceso separadas k periodos: (p.

16).

$$\rho_k = \frac{Cov(Y_t Y_{t+k})}{\sqrt{V(Y_t) V(Y_{t+k})}} = \frac{\gamma_k}{\sqrt{\gamma_0 \gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

La función de autocorrelacion se suele representar por medio de un gráfico de barras denominado correlograma (p. 16).

2.2.8.3 Función de autocorrelación simple (FAS)

Novales (1993). La función de autocorrelacion simple de un proceso estocástico y_t , a la que en lo sucesivo nos referimos por fas., es una función que para cada instante t y cada entero k toma un valor $\rho_k(t)$ igual a la correlación entre y_t e y_{t-k} : (p. 416).

$$\rho_k(t) = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var y_t} \sqrt{Var y_{t-k}}} = \frac{y_k(t)}{\sqrt{Var y_t} \sqrt{Var y_{t-k}}}$$

2.2.8.4 Función de autocorrelación parcial (FAP)

La función de autocorrelacion parcial de un proceso estocástico y_t , a la que en lo sucesivo nos referimos como fap, es una función que para cada instante t y cada entero toma un valor igual a la correlación entre y_t e y_{t-k} , ajustada por el efecto de los retardos intermedios ($y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$) (p. 416-417).

Las características de la función de autocorrelación de un proceso estocástico estacionario:

- El coeficiente de autocorrelación de orden 0 es por definición 1.

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$

- Es una función simétrica:

$$\rho_0 = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_{-k}}{\gamma_0} = \rho_{-k}$$

Por ello, en el correlograma se representa la función de autocorrelación solamente para los valores positivos del retardo k.

La función de autocorrelación de un proceso estocástico estacionario tiende a cero rápidamente cuando k tiende a ∞ .

El proceso estocástico estacionario de FAC, FAS y FAP son independientes del tiempo t, por lo que puede omitirse dicho argumento temporal (p.417).

2.2.9 Modelos lineales estacionarios

2.2.9.1 Procesos autorregresivos: AR (p)

González, M. (2009). Un proceso autorregresivo finito AR (p) es estacionario sí y solo sí el módulo de las raíces del polinomio autorregresivo $\phi_p(L)$ está fuera del círculo unidad (p. 27).

Este proceso expresa a Y_t en función de su pasado hasta el retardo $t - p$ y una innovación contemporánea:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t$$

$$a_t \sim RB(0, \sigma^2) \quad t = 1, 2, \dots$$

En términos del operador de retardos:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) Y_t = a_t \rightarrow \phi_p(L) Y_t = a_t$$

Donde $\phi_p(L)$ recibe el nombre de polinomio autorregresivo y $(\phi_1, \phi_2, \dots, \phi_p)$ es el vector de parámetros autorregresivos.

- **Modelo AR (1): Un proceso autorregresivo finito AR (1).**

$$Y_t = \emptyset Y_{t-1} + a_t \Rightarrow (1 - \emptyset L)Y_t = a_t$$

$$\text{Polinomio autorregresivo: } \emptyset_1(L) = 1 - \emptyset L$$

$$\text{Raíces: } 1 - \emptyset L = 0 \Rightarrow L = \frac{1}{\emptyset}$$

Condición de estacionariedad:

$$|L| = \left| \frac{1}{\emptyset} \right| > 1 \Rightarrow |\emptyset| < 1$$

- **Modelo AR (2): Un proceso autorregresivo finito AR (2).**

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + a_t \Rightarrow (1 - \emptyset_1 L - \emptyset_2 L^2)Y_t = a_t$$

$$\text{Polinomio autorregresivo: } \emptyset_2(L) = 1 - \emptyset_1 L - \emptyset_2 L^2$$

$$\text{Raíces: } 1 - \emptyset_1 L - \emptyset_2 L^2 = 0 \Rightarrow L_1, L_2 = \frac{\emptyset_1 \pm \sqrt{\emptyset_1^2 + 4 \times \emptyset_2}}{-2 \times \emptyset_2}$$

Condición de estacionariedad:

$$|L_1| = \left| \frac{\emptyset_1 + \sqrt{\emptyset_1^2 + 4 \times \emptyset_2}}{-2 \times \emptyset_2} \right| > 1 \quad y \quad |L_2| = \left| \frac{\emptyset_1 - \sqrt{\emptyset_1^2 + 4 \times \emptyset_2}}{-2 \times \emptyset_2} \right| > 1$$

2.2.9.2 Procesos de Medias Móviles: MA (p)

González, M. (2009). Un modelo MA (q) no es invertible para cualquier valor de parámetros de medias móviles, sino que estos tendrán que cumplir algunas restricciones, un proceso de medias móviles finito MA (q) es invertible sí y solo sí el módulo de las raíces del polinomio de medias móviles $\theta_q(L)$ esta fuera del círculo unidad (p. 37).

El modelo expresa el valor de Y_t en función de la innovación contemporánea y de su pasado hasta el retardo q:

$$Y_t = a_t - \theta_1 a_{t-1} + \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

$$a_t \sim RB(0, \sigma^2) \quad t = 1, 2, \dots$$

En términos del operador de retardos:

$$Y_t = (1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3 - \dots - \theta_q L^q) a_t \rightarrow Y_t = \theta_q(L) a_t$$

Donde $\theta_q(L)$ se denomina polinomio de medias móviles y $(\theta_1, \theta_2, \theta_3, \dots, \theta_q)$ es el vector de parámetros de medias móviles.

- **Modelo MA (1): Es un proceso de medias móviles finito MA (1)**

$$Y_t = a_t - \theta a_{t-1} \Rightarrow Y_t = (1 - \theta L) a_t$$

Polinomio medias móviles: $\theta_1(L) = 1 - \theta L$

$$\text{Raíces: } 1 - \theta L = 0 \Rightarrow L = \frac{1}{\theta}$$

$$\text{Condición de invertibilidad: } |L| = \left| \frac{1}{\theta} \right| > 1 \Rightarrow |\theta| < 1$$

- **Modelo MA (2): Es un proceso de medias móviles finito MA (2)**

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \Rightarrow Y_t = (1 - \theta_1 L - \theta_2 L^2) a_t$$

Polinomio de medias móviles: $\theta_2(L) = 1 - \theta_1 L - \theta_2 L^2$

$$\text{Raíces: } 1 - \theta_1 L - \theta_2 L^2 = 0 \Rightarrow L_1, L_2 = \frac{\theta_1 \pm \sqrt{\theta_1^2 + 4 \theta_2}}{-2 \theta_2}$$

Condición de invertibilidad:

$$|L_1| = \left| \frac{\theta_1 + \sqrt{\theta_1^2 + 4 \theta_2}}{-2 \theta_2} \right| > 1$$

$$|L_2| = \left| \frac{\theta_1 - \sqrt{\theta_1^2 + 4 \theta_2}}{-2 \theta_2} \right| > 1$$

2.2.10 Proceso Autorregresivo de Medias Móviles: ARMA (p,q)

González, M. (2009). Los procesos autorregresivos de medias móviles determinan Y_t en función de su pasado hasta el retardo p, de la innovación contemporánea y el pasado de la innovación hasta el retardo q:

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

En términos del operador de retardos como sigue:

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

Donde $\phi_p(L)$ es el polinomio autorregresivo y $\theta_q(L)$ es el polinomio medias móviles.

2.2.10.1 Condiciones de estacionariedad

Un proceso autorregresivo de medias móviles finito $ARMA(p, q)$ es estacionario sí y solo sí el módulo de las raíces del polinomio autorregresivo $\phi_p(L)$ esta fuera del círculo unidad. Las condiciones de estacionariedad del modelo $ARMA(p, q)$ vienen impuestas por la parte autorregresiva, dado que la parte de medias móviles finita siempre es estacionaria.

2.2.10.2 Condiciones de invertibilidad

Un proceso autorregresivo de medias móviles finito $ARMA(p, q)$ es invertible sí y solo sí el módulo de las raíces del polinomio medias móviles $\theta_q(L)$ esta fuera del círculo unidad. Las condiciones de invertibilidad del modelo $ARMA(p, q)$ vienen impuestas por la parte de medias móviles, dado que la parte autorregresiva finita siempre es invertible porque está directamente escrita en forma autorregresiva (p.38-39).

2.2.11 Modelos Autorregresivos Integrados de Promedio Móvil ARIMA (p, d, q)

Los procesos no estacionarios que pueden transformarse en estacionarios mediante sus diferencias de orden d se conocen como procesos estocásticos no estacionarios homogéneos de orden d , que se caracterizan porque, al ser diferenciadas una o más veces, se vuelven estacionarias.

González, M. (2009). Un proceso Y_t es integrado de orden d , si Y_t no es estacionario, pero su diferencia de orden d , sigue un proceso $ARMA(p - d, q)$ estacionario e invertible. El orden de integración del proceso es el número de diferencias que hay que tomar al proceso para conseguir la estacionariedad en media, o lo que es lo mismo, el número de raíces unitarias del proceso (p.46).

El modelo de una serie Y_t integrada de orden d se representa de la siguiente forma:

$$\phi_p(L)\Delta^d Y_t = \delta + \theta_q(L)a_t$$

Donde el polinomio autorregresivo estacionario $\phi_p(L)$ y el invertible de medias móviles $\theta_q(L)$ no tienen raíces comunes.

A este modelo se le denomina modelo Autorregresivo Integrado de Medias Móviles de orden (p, d, q) o $ARIMA(p, d, q)$, donde p es el orden del polinomio autorregresivo estacionario, d es el orden de la integración o las diferencias que hay que tomar para que la serie sea estacionaria y q es el orden del polinomio de medias móviles invertible.

2.2.12 Modelos lineales no estacionarios

2.2.12.1 No estacionariedad en varianza

González, M. (2009). Cuando la serie no es estacionaria en varianza, es decir, no se puede mantener el supuesto de que ha sido formada por un proceso con varianza constante en el tiempo, la solución es transformar la serie mediante algún método que estabilice la varianza.

En general, para estabilizar la varianza se utilizan las transformaciones Box-Cox:

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^{(\lambda)} - 1}{\lambda} & \lambda \neq 0 \\ \ln(Y_t) & \lambda = 0 \end{cases}$$

Donde λ es el parámetro de transformación.

Las transformaciones Box-Cox no solo estabilizan la varianza, sino que también mejoran la aproximación a la distribución normal del proceso Y_t (p. 43-44).

2.2.12.2 No estacionariedad en media

Una de las características significativas en las series económicas es la tendencia.

González, M. (2009). La tendencia es el movimiento a largo plazo de la serie una vez eliminados los ciclos y el término irregular. En economía esta tendencia se suele producir debido a evolución de las preferencias, la tecnología, de la demografía, etc (p. 44).

La no estacionariedad en media se puede modelar de diferentes maneras. Por un lado, es posible modelar tendencias, cambios sistemáticos de nivel, mediante modelos globales en los que se especifica la tendencia como una función del tiempo:

$$Y_t = T_t + u_t$$

Donde $T_t = f(t)$, es una función determinística del tiempo (lineal, cuadrática, exponencial) y u_t es un proceso estocástico estacionario con media cero.

2.2.13 Metodología de Box-Jenkins.

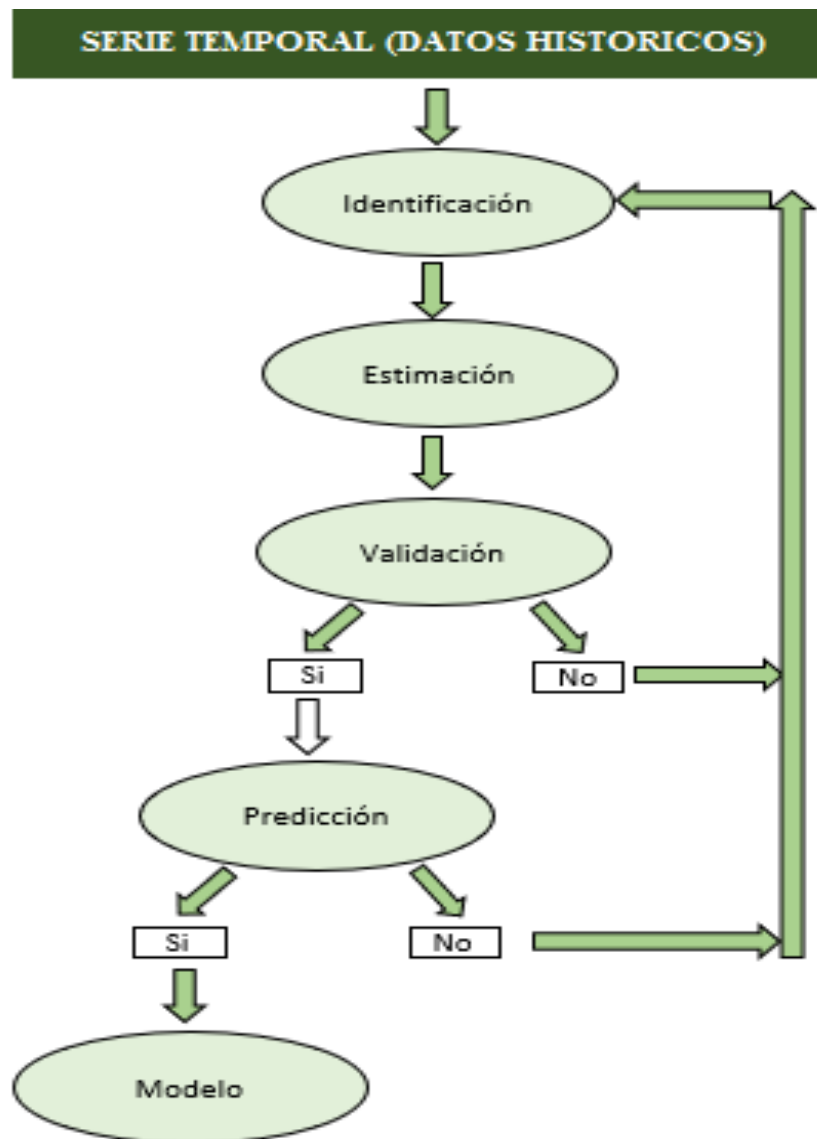
En el año 1970, la metodología propuesta por George Box y Gwilym Jenkins, dos ingenieros con formación estadística, sistematiza modelos modernos estadísticos para el análisis de series temporales univariantes que tienen en cuenta la dependencia existente entre los datos, cada observación es modelada en función de los valores anteriores, la variable tiempo, por tanto, juega un papel fundamental.

Estos modelos se conocen con el nombre genérico de ARIMA (AutoRegresive Integrated Moving Average), el cual deriva de sus tres componentes: Autorregresivo (AR), Integrado (I) de Medias Móviles (MA).

Modelar una serie temporal supone identificar un modelo ARIMA adecuado que se ajuste a la serie, debe contener los mínimos elementos necesarios para describir el fenómeno y ser útil para realizar previsiones.

González, M. (2009) sostiene que la aplicación de la metodología Box-Jenkins el punto de partida es el contrario: se conocen los valores de la serie temporal Y_1, Y_2, \dots, Y_T y se trata de determinar la estructura $ARIMA(p,d,q)$ que la ha podido generar (p.54).

La construcción de los modelos ARIMA se lleva a cabo de forma iterativa mediante un proceso en el que se pueden distinguir cuatro etapas:



Fuente: George B. y Gwilym J. Time series analysis: Forecasting.
Elaboración propia

González, M. (2009) sostiene que es el principio de parametrización escueta, también denominado parsimonia. Además, se trata de proponer un modelo capaz de representar la serie con el mínimo de parámetros posibles y únicamente acudir a una ampliación del mismo caso para describir el comportamiento de la serie (p.55).

2.2.13.1 Fase de Identificación

González, M. (2009). La identificación se basa en el objetivo de la etapa en modelizar un modelo que sea apropiado para una serie. La identificación del modelo tiene dos fases: El análisis de estacionariedad y la elección de órdenes (p y q) (p.51).

- El análisis de estacionariedad es la que determina las transformaciones para obtener una serie estacionaria, por lo cual, tiene dos apartados la estacionariedad en varianza y la estacionariedad en media.
- **Serie estacionariedad en varianza:** Es cuando existe una sola varianza para toda la serie, es decir, cuando la variabilidad de la serie en torno a su media se mantenga constante a lo largo del tiempo.
- **Serie no estacionariedad en varianza:** Se utiliza la transformación estabilizadora de varianza Box-Cox, que incluye las funciones de la raíz cuadrada, inversa, etc.
- **Serie estacionariedad en media:** Es cuando existe una única media para toda la serie temporal, esto quiere decir, que fluctúa alrededor de una media constante. Además, la función de autocorrelación teórica de un proceso estacionario en media que decae rápidamente.
- **Serie no estacionariedad en media:** Es cuando presenta tendencia o varios tramos con media diferente. Además, el proceso con alguna raíz unitaria presenta una función de autocorrelación muestral con un decaimiento muy lento.

Elección de los órdenes (p y q) es una serie estacionaria, que tiene por objetivo determinar el proceso estacionario AR (p), MA (q) o ARMA (p,q) (p.51-54).

➤ **La prueba de Dickey Fuller**

Dickey y Fuller (1979) diseñaron un procedimiento para probar formalmente la presencia de Raíces Unitarias. La prueba comienza por suponer que la serie X_t sigue un proceso autorregresivo de primer orden AR (1), de forma:

$$X_t = rX_{t-1} + e_t$$

González, M. (2009) sostiene que el estadístico t se le denomina estadístico de Dickey Fuller. El estadístico t no sigue una distribución conocida bajo la H_0 de no estacionariedad, por lo que Dickey Fuller calcula los percentiles del estadístico bajo, para proporcionar en función del tamaño muestral (T) y el nivel de significación (p. 55-56).

Wooldridge, J. (2010) sostiene que se conoce la prueba de Dickey Fuller como la distribución asintótica del estadístico t bajo H_0 . Cuando $|\rho| < 1$; $\{X_t\}$ es un proceso estable AR (1), lo que significa que es débilmente dependiente o que no se correlaciona asintóticamente (p.631).

En dicha prueba, en el caso ($r = 1$) se consideran tres formas diferentes ecuaciones de regresión para probar la presencia de una raíz unitaria y estas son:

Primera ecuación: X_t es un modelo puro de caminata aleatoria sin tendencia (estocástica).

$$\Delta X_t = rX_{t-1} + e_t$$

Segunda ecuación: X_t es un modelo de intercepto o término de deriva.

$$\Delta X_t = a_0 + rX_{t-1} + e_t$$

Tercera ecuación: Incluye ambas ecuaciones; X_t es un modelo de intercepto y una línea de tendencia.

$$\Delta X_t = a_0 + rX_{t-1} + a_1t + e_t$$

La diferencia entre las tres ecuaciones de regresión es por la presencia de los elementos determinísticos a_0 y a_1t . El parámetro de la ecuación de regresión es r ; la variable de tendencia es t . Las hipótesis son:

$H_0: r = 1$ (En este caso, existe una raíz unitaria, por lo tanto, la serie no es estacionaria o tiene tendencia estocástica)

$H_1: r < 1$ (En este caso, existe una serie estacionaria) (p.06-07).

2.2.13.2 Fase de Estimación

Luego de identificar los modelos seleccionados, se procede a la estimación de los parámetros.

González, M. (2009). Los parámetros se pueden estimar de forma consistente por mínimos cuadrados o máxima verosimilitud. Ambos métodos se basan en el cálculo de las innovaciones, a_t , a partir de los valores de la serie estacionaria (pág. 69).

Inclusión del término independiente:

La media de un proceso ARMA (p, q) estacionario está directamente relacionada con la constante δ . Si esta constante es cero, entonces la media del proceso es cero. Para decidir si se incluye un término independiente no nulo en el modelo, se contrastará si la media de la serie estacionaria es o no cero.

$$H_0: E(Z_t) = 0$$

$$H_\alpha: E(Z_t) \neq 0$$

Donde el estadístico de contraste es:

$$t = \frac{\bar{Z}}{\hat{\sigma}_Z} \sim t(T^* - 1)$$

Se rechazará la H_0 , a un nivel de significancia de α y, por lo tanto, se incluirá el parámetro δ en el modelo si:

$$t > t_{\alpha/2}(T^* - 1)$$

Esta etapa nos da señales de cual modelo es el más adecuado para poder hacer los respectivos pronósticos, ya que los coeficientes deberán tener ciertas condiciones para que el modelo sea aceptado.

2.2.13.3 Fase de Validación

Luego de elegir el modelo que más se adecue a la serie y sus coeficientes cumplan con ciertas condiciones se procede hacer la validación del modelo.

González, M. (2009). Propone que se debe tener en cuenta lo siguiente:

- Si las estimaciones de los coeficientes del modelo son significativas y cumplen con las condiciones de estacionariedad e invertibilidad que deben satisfacer los parámetros del modelo.
- Si los residuos del modelo tienen un comportamiento similar a las innovaciones, es decir, sin son ruido blanco (pag.73).

➤ Prueba de Ljung-Box

Villavicencio (2007) expresa, que la prueba Ljung-Box permite probar en forma conjunta si que todos los coeficientes de autocorrelación son simultáneamente iguales a cero, esto es que son independientes y está definida por:

H_0 : Las autocorrelaciones son independientes.

H_1 : Las autocorrelaciones no son independientes.

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{b_k^2}{n-k} \right) \sim X_{(m)}^2$$

Identificaremos a cada uno de la ecuación y son:

n = Tamaño de la muestra.

m = Longitud del rezago.

k = Número de rezagos que se analizan (k=12,24,36,48)

b_k = Coeficiente de correlación.

El p-valor se define como:

Es cierta: Si p-valor $< \alpha$ Se rechaza H_0 ; por lo contrario, Si p-valor $> \alpha$ Se acepta H_0

Además, si el Q calculado excede el valor Q crítico de la tabla ji cuadrada, al nivel de significancia seleccionado, no se acepta H_0 de que todos los coeficientes de autocorrelación son iguales a cero; por lo menos algunos de ellos deben ser diferentes de cero (p.05).

➤ Prueba de Kolmogorov-Smirnov de una muestra

Siegel y Castellan (1998) define que la prueba de Kolmogorov-Smirnov de una muestra es otra prueba de la bondad de ajuste, ambas son pruebas estadísticamente no paramétricas. La prueba de Kolmogorov-Smirnov determina si las puntuaciones en una muestra pueden razonablemente provenir de una población que tiene una distribución teórica.

Además, la prueba ji cuadrado es definitivamente menos potente o aproximadamente (asintóticamente) exacta que la prueba de Kolmogorov-

Smirnov es exacta, más aún para muestras muy pequeñas. La prueba ji cuadrada supone que las distribuciones son nominales, mientras que la prueba de Kolmogorov-Smirnov supone una distribución continua, ambas pruebas pueden aplicarse a datos ordinales (p.73-77).

H_0 : La distribución observada se ajusta a la distribución normal.

H_1 : La distribución observada no se ajusta a la distribución normal.

La fórmula se calcula en función a la máxima desviación de “D” según esta expresión es:

$$S_n(X_i) = \frac{F_i}{N}$$

$$D = \max |F_0(X_i) - S_n(X_i)| \quad i = 1, 2, \dots, N$$

F_0 = Función de distribución de frecuencia relativa acumulada esperada.

S_n = Función de distribución de frecuencia relativa acumulada observada.

➤ **Prueba de Kolmogorov-Smirnov de dos muestras**

Siegel y Castellan (1998) afirma que es una prueba de dos muestras independientes que se han extraído de la misma población. Esta prueba determina la distribución de la frecuencia acumulada para cada muestra de observaciones, utilizando los mismos intervalos para ambas distribuciones. La prueba se centra en las mayores de las desviaciones observadas.

Además, se comparó con la prueba t, la prueba de Kolmogorov-Smirnov tuvo una potencia-eficiencia alta (alrededor 95%) para muestras pequeñas. Al incrementarse el tamaño de la muestra, la potencia-eficiencia disminuye levemente. En pocas palabras la prueba de Kolmogorov-Smirnov es más poderosa en todos los casos de la ji cuadrada y la prueba de la mediana

(p.174-182).

La diferencia máxima absoluta D es:

$$S_m(X_i) = \frac{K}{m} ; \quad S_n(X_i) = \frac{K}{n}$$

$$D_{m,n} = \max |S_m(X_i) - S_n(X_i)|$$

K = Es el número de datos iguales o menores que X_i

S_m = Distribución acumulativa observada para una muestra.

S_n = Distribución acumulativa observada para otra muestra.

El p-valor se define como:

$$p - \text{valor} = P(D > D_{abs}/H_0)$$

Es cierta: Si p-valor $< \alpha$ Se rechaza H_0 ; por lo contrario, Si p-valor $> \alpha$ Se acepta H_0

➤ Prueba de Anderson-Darling

Villaseñor (2011) define que la prueba Anderson-Darling puede ser utilizada para probar base de datos transformados y estimando los parámetros (media y varianza) por máxima verosimilitud. Además, la prueba es estadísticamente no paramétrica (p.4-10).

H_0 : La distribución observada se ajusta a la distribución normal.

H_1 : La distribución observada no se ajusta a la distribución normal.

$$AD = -n - S$$

$$S = \sum_{i=1}^n \frac{2i-1}{n} [\ln F(X_i) + \ln(1 - F(Y_{n+1-i}))]$$

2.2.13.4 Fase de Predicción.

Una vez haber identificado el modelo ARIMA, estimado sus respectivos parámetros y haber pasado la etapa de validación, se utiliza el modelo para realizar pronósticos, con el menor error de predicción posible.

2.2.14 Producción y Demanda

2.2.14.1 Producción

Marx y Engels (1859) define que es un proceso de compra de la fuerza de trabajo humano por un determinado periodo (semana, mes, etc.). Es un intercambio de servicios productivos de trabajo propio, de tierra y de capitales. La producción no está determinada por la cantidad que se produce sino por la estructura en cómo se lleva adelante la producción.

Además, la producción es la acción de generar un proceso de diferentes productos que beneficien al consumidor. En la sociedad hay muchos campos concretos de los cuales son: La economía, la música, el cine, la televisión, etc. Son campos que se desarrollan por una fuerza de trabajo a cambio de un salario.

Marx (1864) define que es un medio histórico que desarrolla la fuerza productiva material y crea un mercado mundial, en beneficio a la sociedad de los productores.

Además, es la creación en procedimientos de bienes y mercancías, donde el aumento de la productividad de trabajo se denomina el proceso de industria. Lo más importante es la ganancia media global del capital social.

2.2.14.2 Demanda.

Marx (1864) Define que es la compra de medios de producción que pueden servir para fines mercantiles para entrar en el consumo productivo o individual.

La demanda es la acción de solicitar o pedir algo que realmente sea fundamental para su necesidad. En el caso de la Economía, el Derecho, la Comunicación, el Empleo, etc. Son campos concretos que se relacionan a la demanda.

En la economía la demanda es representada por los consumidores en adquirir un servicio. También está estrechamente unido a la oferta, esto quiere decir que, durante la evolución del precio, la demanda depende de la oferta o viceversa.

2.2.14.3 Agua potable en la Región de Lambayeque

El departamento de Lambayeque está entre las 10 regiones con más población en el país, esto quiere decir que el consumo del agua será cada vez más demandante por la población. La fuente principal del consumo de agua en Lambayeque es la primera etapa del Proyecto de Tinajones (El río Conchano y el río Chotano que alimentan a la cuenca del Pacífico con una masa anual de 250 millones de m³ de agua).

La captación del agua se realiza a través de la bocatoma Raca-Rumi ubicado en el río Chancay, cuenta con dos compuertas radiales que comunican al canal alimentador a través de 6 cámaras desarenadoras, 3 compuertas tipo vagón que regulan el volumen de embalse y un aliviadero. Las aguas derivadas del Río Chancay por la bocatoma Raca-Rumi ingresan al Reservorio Tinajones a través de un canal alimentador de 16 km. de

longitud con una capacidad máxima de 70 m³/s. La capacidad de almacenamiento de la laguna Boró I es de 500,000 m³/s y la laguna Boró II es de 1'500,000 m³/s.

La conducción hacia el Reservorio Tinajones, es cuando el agua cruda pasa a través de “las Cascadas”, y posteriormente se almacena en el Reservorio Tinajones que tiene una capacidad de embalse de 320 millones de m³, profundidad máxima de 37 metros, además el Río Chancay - Lambayeque tiene una capacidad máxima de 70 m³/s.

2.2.14.4 Proceso del tratamiento de agua.

La Entidad Prestadora de Servicios de Saneamiento de Lambayeque tiene dos Plantas de Tratamiento de agua y tiene por nombre (N1 y N2). Los procesos de tratamiento de agua son: Precloración, Coagulación, Floculación, Decantación, Filtración y Desinfección. Por lo tanto, la planta de tratamiento de agua realiza dos tramos de tuberías, por lo cual se divide en dos partes. La primera es el tramo Boró-Chéscope de 40 pulg. de diámetro y 5, 835 m.l. y el segundo Chéscope-Planta de 34 pulg. de diámetro y 8,520 m.l.

➤ Precloración:

El proceso de precloración su mismo nombre lo dice es antes de la cloración. En este proceso se coloca una inyección de solución clorada fijamente a la línea de ingreso de agua cruda. La función de la precloración es eliminar parcialmente la cantidad de microorganismos.

➤ Coagulación:

El proceso de la coagulación es cuando se realiza en la cámara de mezcla rápida, por lo que se debe esperar un periodo de retención de 30

segundos. Además, se produce la desestabilización de las partículas coloidales que trae el agua cruda. Este proceso se realiza por medio de la adición de sustancias químicas tal como sulfato de aluminio. Hay otras sustancias químicas que se aplica y son: Polímeros, Cal y Sulfato de cobre (dependiendo la calidad del agua cruda).

➤ **Floculación:**

El proceso de floculación es cuando en la cámara ocurre una mezcla lenta, por lo que promueve el crecimiento de los flocus o flósculos hasta un tamaño determinado y los pesos adecuados para su posterior sedimentación luego de un periodo de 13 minutos.

➤ **Decantación:**

El proceso decantación es de dos tipos: El convencional y el laminar. La decantación convencional produce primero la sedimentación simple o arrastre de los flocus de mayor tamaño y la decantación laminar es un sistema presentado de angotubos.

➤ **Filtración:**

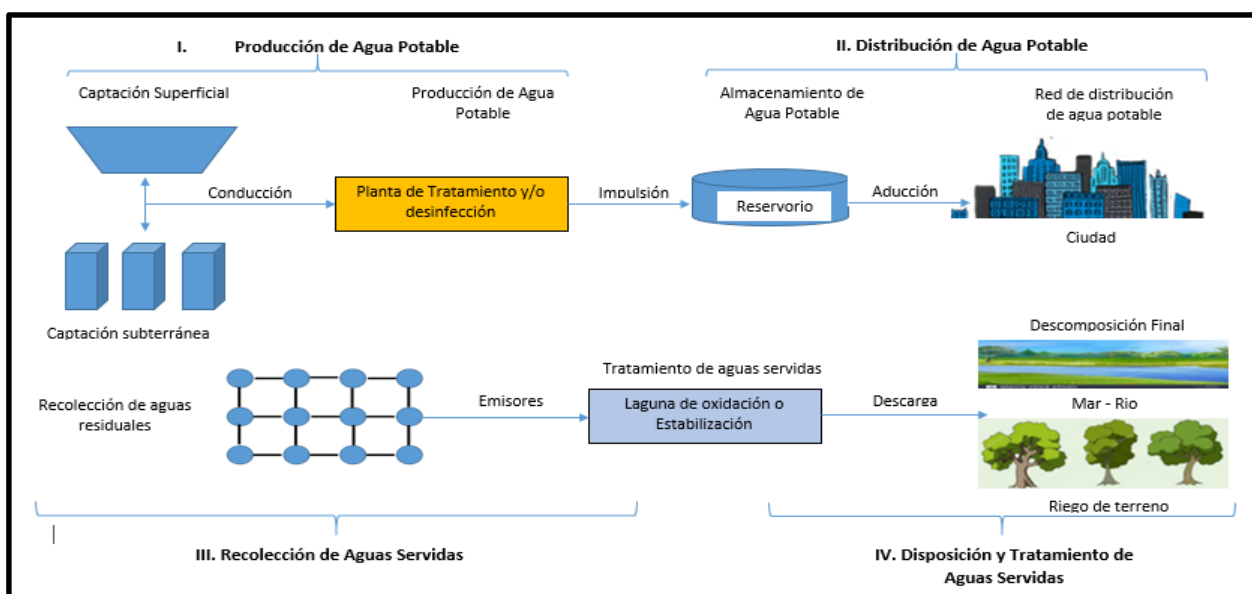
El proceso de filtración es a través de 08 unidades de filtración descendente. Por otro lado, cada unidad formada por dos secciones, donde el agua por gravedad pasa a través de lechos formado por antracita, arena y grava. Además, la filtración finaliza por medio de una tubería ubicada debajo de la galería de válvulas, en el sótano de la Planta.

➤ **Desinfección:**

El proceso de la desinfección es mediante solución clorada que se inyecta puntualmente en la tubería de recolección de agua filtrada, es allí donde se destruyen los gérmenes patógenos (causantes de enfermedades)

existentes en el agua. Además, mantiene un residual de cloro en los sistemas de almacenamiento y distribución, protegiendo así al agua potable de contaminaciones posteriores.

CICLO DE PRODUCCIÓN, DISTRIBUCIÓN, RECOLECCIÓN Y TRATAMIENTO



Elaboración propia / Fuente: EPSEL

3.3 Definición y Operacionalización de variables

- **EPSEL:** Entidad Prestadora de Servicios de Saneamiento de Lambayeque.
- **Producción:** Es la acción de generar un proceso de diferentes productos que beneficien al consumidor. En la sociedad hay muchos campos concretos de los cuales son: La economía, la música, el cine, la televisión, etc. Son campos que se desarrollan por una fuerza de trabajo a cambio de un salario (Marx, 1864).
- **Demanda:** Es la compra de medios de producción que pueden servir para fines mercantiles para entrar en el consumo productivo o individual (Marx, 1864).
- **Serie de Tiempo:** Es serie de tiempo es el conjunto de observaciones que el transcurso del tiempo que son tomadas en momentos específicos y que pueden ser en intervalos iguales (Spiegel y Stephens, 2002; p.436).
- **Procesos Estocásticos:** Es un proceso estocástico es una secuencia de variables aleatorias, ordenadas y equidistantes cronológicamente (Mauricio, 2007; p.9).
- **Estacionalidad:** Es un esquema (patrones) de cambio en una serie de tiempo de uno año. Estos patrones tienden a repetirse cada año (Mason, Lind y Marchal, 2001; p.650).
- **Estacionariedad:** Hay dos formas de estacionariedad. La estacionariedad de covarianza (conocido también como débil) y estacionariedad estricta (González, M. 2009; p.13).
- **MMC:** Millones de metros cúbicos. / **MC:** Metros cúbicos.

3.4 Operacionalización de Variables

VARIABLES	DEFINICIÓN	DIMENSION	INDICADOR	UNIDAD DE MEDIDA	ESCALA
Producción	Es un proceso de compra de la fuerza de trabajo humano por un determinado periodo (semana, mes, etc.). Es un intercambio de servicios productivos de trabajo propio, de tierra y de capitales.	Producción continua	Cantidad Disponible en Metros Cúbicos	MC	Razón
Demanda	Es la compra de medios de producción que pueden servir para fines mercantiles para entrar en el consumo productivo o individual.	Dimensión de Uso	Cantidad Consumida en Metros Cúbicos	MC	Razón

Elaboración propia.

CAPITULO III: MÉTODOS Y MATERIALES

3.1 Diseño de Investigación

El diseño es no experimental.

3.2 Tipo de Investigación

La presente investigación es Descriptiva Predictiva de tipo temporal longitudinal retrospectiva, dado que está destinado a encontrar explicación racional del comportamiento de las variables producción y demanda de agua en el distrito de Lambayeque, periodo enero 2002 hasta diciembre 2017.

3.3 Materiales de Estudio

➤ Población

La población objetivo en estudio está conformada por todos los datos históricos de la producción y demanda de agua en el distrito de Lambayeque. La empresa EPSEL tiene datos históricos sobre dichas variables, y que solo INEI tiene registrados desde un periodo limitado.

➤ Muestra

La muestra de estudio está constituida por 180 observaciones de la producción y demanda de agua en el distrito de Lambayeque, que corresponden al periodo enero 2002 hasta diciembre 2017. Por temas de investigación hemos delimitamos en tal periodo de tiempo.

INEI tiene los datos registrados desde un cierto periodo. Los datos descargados se obtuvieron en la página web del Instituto Nacional Estadística e Informática (INEI).
www.inei.gob.pe

3.4 Técnicas e Instrumentos de Recolección de Datos

Técnica de tipo documental. Los datos se obtuvieron mediante la observación indirecta, a través del “Avance Económico y Social Departamental, Diciembre 2018”, la cual contenía los datos de la serie histórica de producción y demanda de agua potable en la ciudad de Lambayeque, proporcionada por la página web del (INEI), para la cual se creará un almacenamiento de Base de datos utilizando el programa de paquetes estadísticos.

3.5 Análisis Estadístico de los Datos

El análisis se realizó haciendo uso de las técnicas de serie de tiempo (Metodología Box Jenkins), utilizando los softwares SPSS 22 (Licencia de prueba) y Eviews v.6.

Se inició con el análisis de los datos, observando si hay o no hay datos atípicos, para eso se utilizó el diagrama de caja y bigotes. Luego, se hizo el proceso estadístico de una serie temporal. Se inició con la identificación y para eso se realizó un análisis exploratorio del comportamiento de la variable en estudio para determinar su estacionariedad, también se utilizó la prueba de Dickey Fuller aumentado para evidenciar estacionariedad en la serie, si no cumple con estacionariedad se tuvo que volver a transformar la serie temporal. Por lo tanto, si cumple con estacionariedad se estima un modelo óptimo que explique el comportamiento de la producción y demanda de agua en el periodo de estudio. Además, se evaluó la adecuacidad del modelo estimado, analizando las propiedades de la serie de los residuales. Por otro lado, se estimó los parámetros utilizando los estadísticos de error RMSE, MAPE y MAE. La siguiente etapa es la validación del modelo (prueba de normalidad, prueba de aleatoriedad y ruido blanco) y finaliza con el pronóstico del modelo. Además, los datos estimados para el año 2018 se compararon con los datos registrados en la INEI. Nuestra investigación finalizó con el pronóstico de ambas variables para el año 2019.

IV. RESULTADOS Y DISCUSION

A través de la Metodología Box – Jenkins que consta de cuatro etapas importantes, se llega a construir los modelos ARIMA univariante que permiten realizar los pronósticos de las series de producción y demanda de agua en el distrito de Lambayeque, para el periodo enero 2002 a diciembre del 2017.

Se presentan las tablas y figuras para el análisis, discusión e interpretación de los datos.

1. Aplicación de la Metodología Box-Jenkins para la producción de agua

Tabla 1: Estadísticos descriptivos de la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

Estadístico	Producción
Media	335,005.19
Desviación estándar	20,044.97
Coeficiente de Variación	5.98%
Mínimo	275,813
Máximo	384,759

Fuente: Elaboración propia.

En la Tabla 1, se presentan los estadísticos descriptivos de la producción de agua del distrito de Lambayeque, donde se tiene que en promedio es 335,005.19 metros cúbicos, con una desviación típica de $\pm 20,044.97$, un coeficiente de variación de 5.98% es decir que los datos son homogéneos en torno a la media, el mínimo de producción de agua ha sido de 275,813 y un máximo de 384,759 metros cúbicos respectivamente.

1.1 Análisis de la estacionariedad de los datos

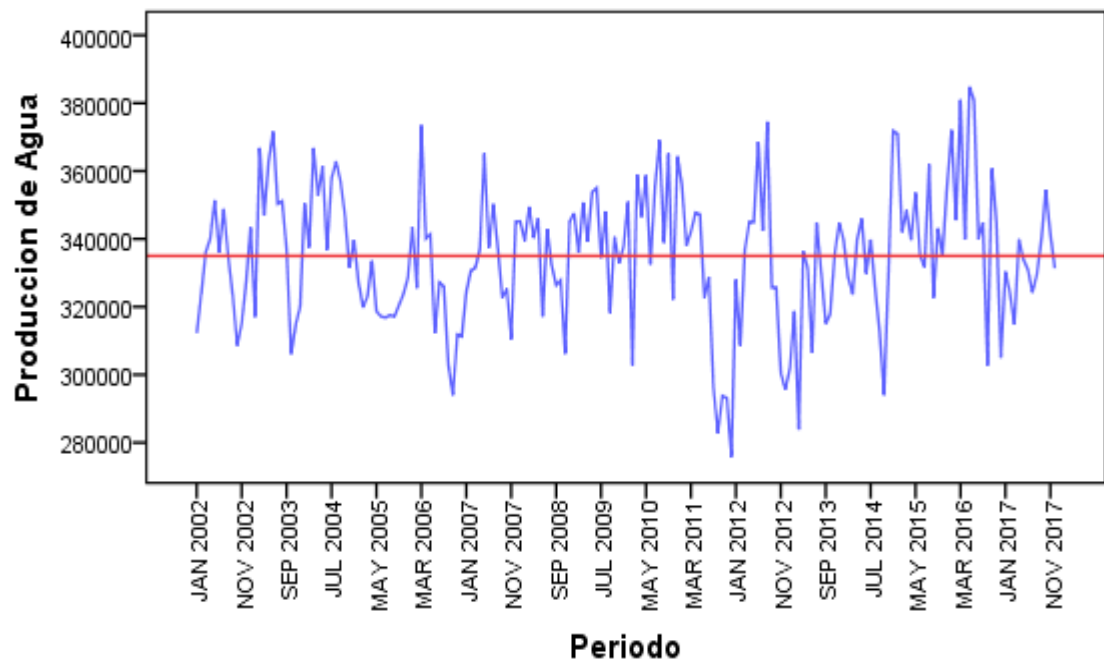


Figura 1: Serie original de la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

En el Figura 1, se observa que la serie presenta un comportamiento estacionario, es decir los datos oscilan alrededor de la media y tiene una varianza constante. Para confirmar dicha estacionariedad, se utiliza las funciones de autocorrelación y la prueba de raíces unitarias de Dickey–Fuller Aumentada.

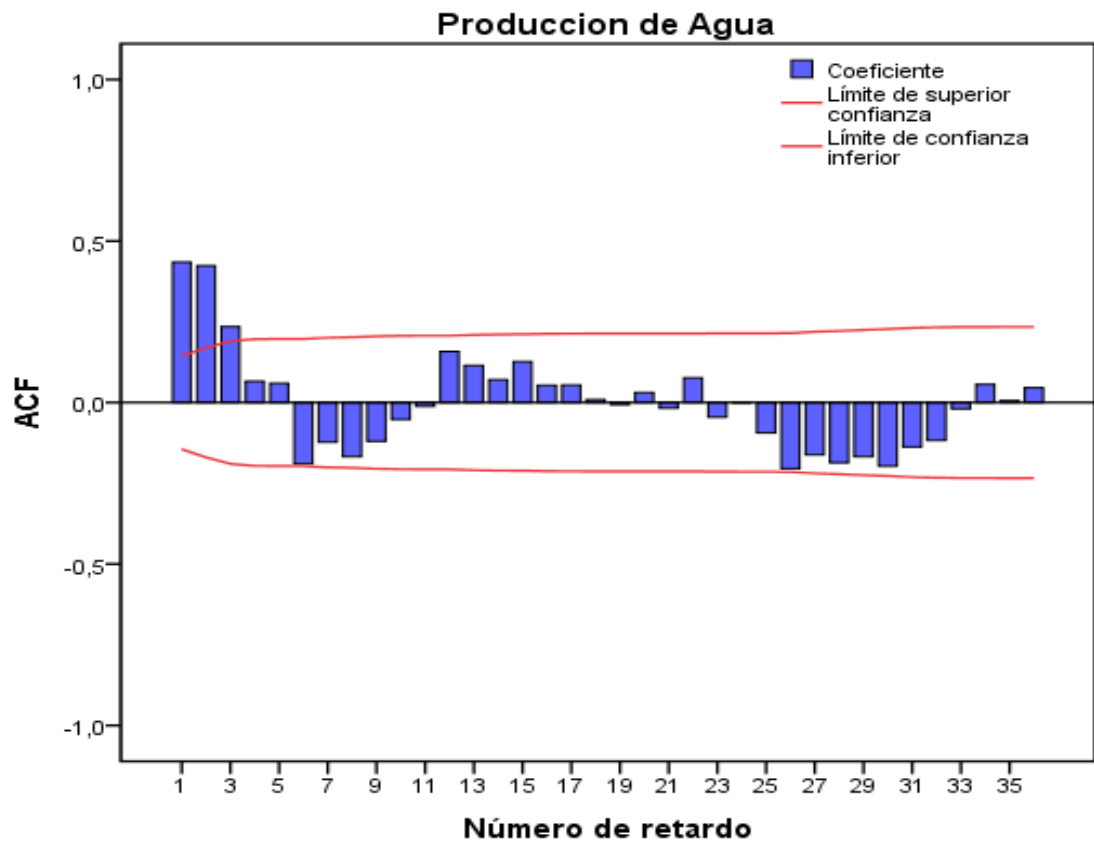


Figura 2: Función de Autocorrelación Simple de la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

En la Figura 2, observamos un decaimiento brusco a partir del tercer coeficiente de autocorrelación por lo que se puede afirmar que la serie de producción de agua del distrito de Lambayeque en el periodo enero 2002 a diciembre 2017, tiene un comportamiento estacionario. A si mismo se observa que la serie no presenta un comportamiento estacional, por lo que se presume que el modelo buscado tenga componentes MA en la parte regular.

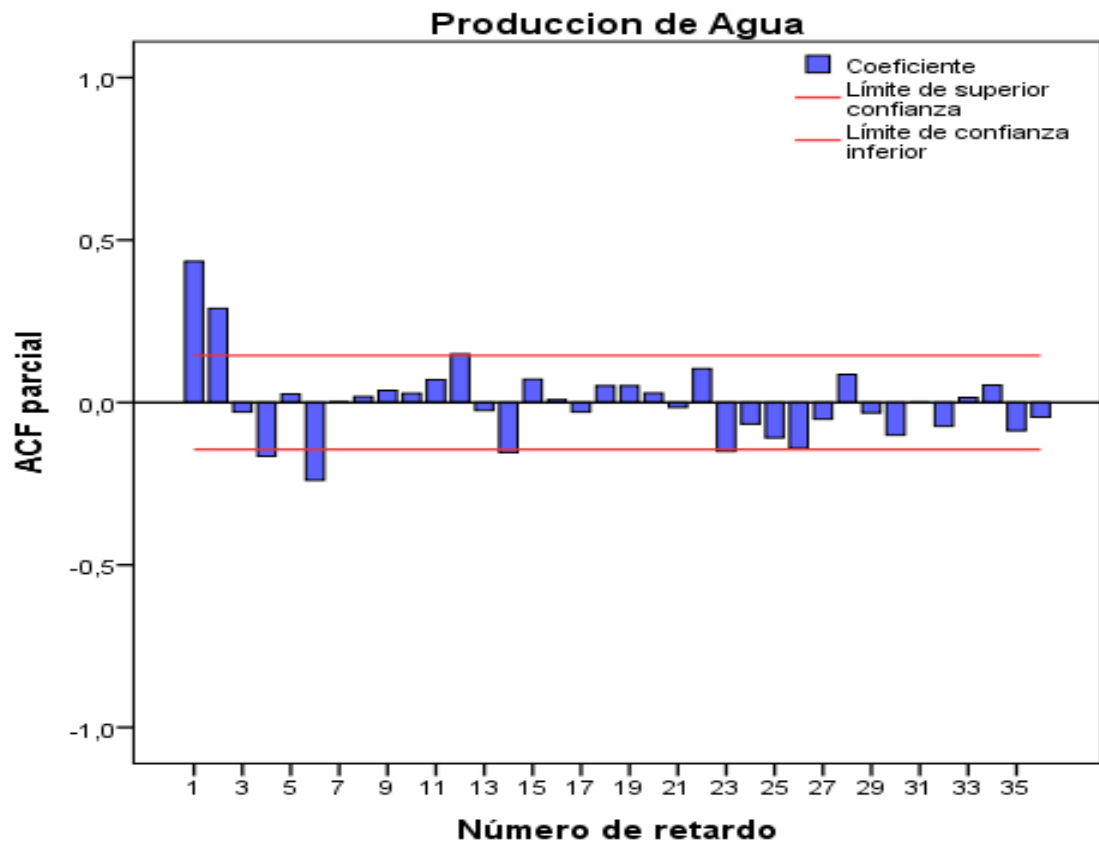


Figura 3: Función de Autocorrelación Parcial de la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

En la Figura 3, se observa un decaimiento brusco a partir del segundo coeficiente de autocorrelación parcial, lo que nos indica que la serie de la producción de agua del distrito de Lambayeque en el periodo enero 2002 a diciembre 2017 es estacionario, además existen coeficientes significativos que sobresalen los límites de confianza, lo cual se presume que el modelo buscado tenga componentes AR en la parte regular y estacional.

Tabla 2: Prueba de raíz unitaria de Dickey - Fuller para la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.427251	0.000
1% level	-3.464827	
5% level	-2.876595	
10% level	-2.574874	

Fuente: Elaboración propia – Eviews

En la Tabla 2, se observa la prueba de Dickey – Fuller con un p- value < 0.05, por lo que se concluye que la serie de la producción de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017, es estacionaria, con un nivel de significancia del 5%.

Cumpliendo de esta manera las condiciones de estacionariedad para la metodología ARIMA de Box – Jenkins. (VER ANEXO G)

1.2 Identificación de los modelos

Una vez confirmada la estacionariedad en media y varianza de la serie de producción de agua, procedemos a la etapa de identificación para determinar cuál es el modelo que mejor se ajusta, a partir de las funciones de autocorrelación que sirven como herramientas en esta etapa.

En base a las autocorrelaciones obtenidas en las figuras N°2 y 3, se identificaron diferentes modelos, a los cuales se les realizó pruebas siguiendo el proceso de la metodología Box – Jenkins, descartando los que no cumplían con los requisitos.

MODELOS TENTATIVOS IDENTIFICADOS:

- MODELO 1: ARMA (2, 0, 2)
- MODELO 2: SARMA (2, 0, 0) (1, 0, 0)
- MODELO 3: SARMA (1, 0, 2) (1, 0, 0)

Después de llevar a cabo el proceso de identificación de los modelos tentativos se procederá a la estimación de los parámetros, para la serie de producción de agua en el distrito de Lambayeque.

1.3 Estimación de los parámetros de los modelos tentativos

La siguiente etapa consiste en determinar el modelo más adecuado para la serie de producción de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017, donde se evaluará los modelos tentativos que se identificaron en la anterior etapa.

Para decidir si se incluye un término independiente (Constante), en la estimación de los parámetros del modelo, se contrastará si la media de la serie estacionaria es o no cero.

Tabla 3: Prueba t para una muestra para la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

	t	gl	Sig. (bilateral)
Producción de Agua	231.578	191	0.000

Fuente: Elaboración propia – SPSS

En la tabla 3, observamos la prueba t para una muestra con un p-valor < 0.05 , concluyendo a un nivel de significancia del 5%, que la media de la serie de producción de agua en el distrito de Lambayeque en el periodo enero 2002 a

diciembre 2017 es diferente cero, por lo tanto, se incluirá un término independiente (constante) al evaluar los parámetros de los modelos tentativos. (VER ANEXO H)

Tabla 4: Resumen de los modelos tentativos identificados de la serie de producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

MODELO	Parámetro	Estimación	Significancia	RMSE	MAPE	MAE
ARMA (2,0,2)	C	334,806.33	0.000	17,006.38	4.00%	13,367.95
	AR 1	1.616	0.000			
	AR 2	-0.844	0.000			
	MA 1	1.380	0.000			
	MA 2	-0.763	0.000			
SARMA (2,0,0)(1,0,0)	C	334,649.05	0.000	16,981.99	3.99%	13,304.18
	AR 1	0.302	0.000			
	AR 2	0.318	0.000			
	SAR 1	0.219	0.003			
SARMA (1,0,2)(1,0,0)	C	334,648.07	0.000	16,932.67	3.98%	13,254.61
	AR 1	0.608	0.000			
	MA 1	0.299	0.022			
	MA 2	-0.248	0.004			
	SAR 1	0.215	0.004			

Fuente: Elaboración propia – SPSS

En la tabla 4, se observa que los parámetros son significativos en los 3 modelos, entonces se procede a evaluar los estadísticos de error que son el RMSE (Error cuadrático medio), MAPE (Error porcentual absoluto medio) y el MAE (error absoluto medio), los cuales nos ayudaran a determinar qué modelo tiene menos errores de pronóstico, concluyendo que el mejor modelo es el SARMA (1, 0, 2) (1, 0, 0), ya que tiene el menor RMSE igual a 16 932.67, MAE igual a 13 254.61 y el MAPE que es igual a 3.98% el cual mide el error de pronóstico de manera porcentual

comparado con los valores reales de la serie.

1.4 Validación del modelo

Después de seleccionar un modelo y de estimar sus parámetros, se procede a la etapa de validación para ver si el modelo seleccionado para la serie de producción de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017, se ajusta a los datos en forma razonablemente buena, el detalle se encuentra en el análisis de los residuales, estos se deben aproximar a un comportamiento de ruido blanco.

Tabla 5: Prueba de normalidad para los residuales del modelo SARMA (1, 0, 2) (1, 0, 0)

Producción de Agua	Kolmogorov-Smirnova		
	Estadístico	gl	Sig.
	0.034	192	0,200*

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Fuente: Elaboración propia

En la Tabla 5, se observa la prueba de normalidad de Kolmogorov – Smirnov, donde se obtiene que el nivel de significación es mayor al 0.05, por lo tanto, se concluye que los residuales del modelo SARMA (1, 0,2) (1, 0,1) siguen una distribución normal, lo que indica que el modelo es satisfactorio. (VER ANEXO I)

Tabla 6: Prueba de aleatoriedad para los residuales
del modelo SARMA (1, 0, 2) (1, 0, 0)

Valor de prueba	-876
Casos < Valor de prueba	96
Casos >= Valor de prueba	96
Casos totales	192
Número de rachas	105
Z	1.158
Sig. asintótica (bilateral)	0.247

a. Mediana

Fuente: Elaboración propia

En la Tabla 6, se observa los resultados de la prueba de aleatoriedad o rachas, donde la significación de la prueba es de 0.247 mayor que 0.05, concluyendo al 5% de significancia que los residuos del modelo SARMA (1, 0, 2) (1, 0, 0) son aleatorios. (VER ANEXO J)

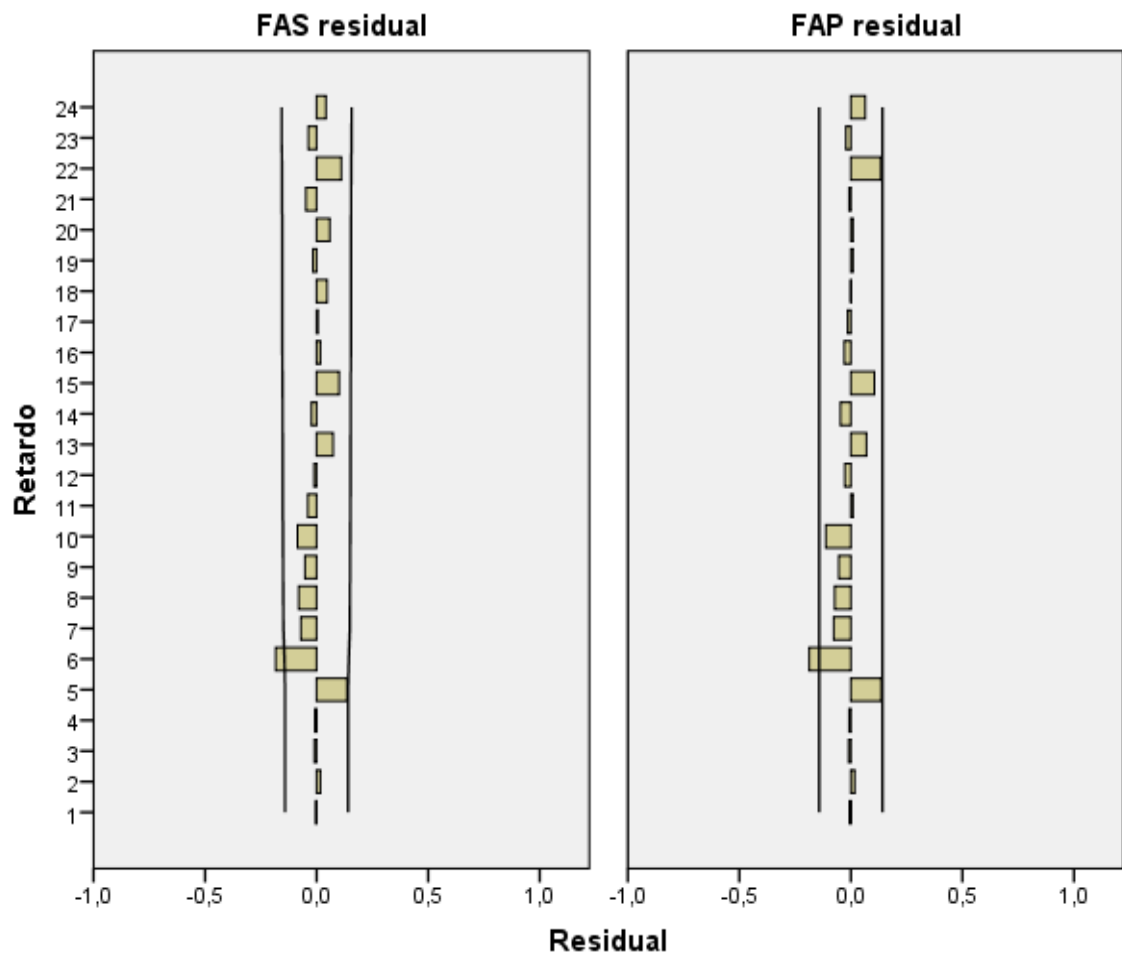


Figura 4: Función de Autocorrelación simple y parcial de los residuales del modelo SARMA (1, 0, 2) (1, 0, 0) de la producción de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

En la Figura N°04, se observa que los coeficientes de autocorrelación simple y parcial de los residuales de la serie se encuentran dentro de los límites de confianza, por lo que se concluye que los residuales del modelo SARMA (1, 0, 2) (1, 0, 0) tienden a un ruido blanco.

1.5 Pronóstico del modelo

➤ Pronóstico 2018:

En esta etapa usando el modelo que ha sido validado, se procede a realizar los pronósticos para el periodo enero 2018 a diciembre del 2018.

Tabla 7: Pronostico de la producción de agua en el distrito de Lambayeque, para el periodo enero 2018 a diciembre 2018.

Periodo	Pronostico SARMA (1, 0, 2) (1, 0, 0)	Observado	Límite inferior al 95%	Límite superior al 95%
Ene-18	335,693	355,246	302,360	369,027
Feb-18	332,519	365,602	297,626	367,412
Mar-18	330,483	393,498	292,683	368,282
Abr-18	335,828	360,819	297,007	374,648
May-18	334,526	372,016	295,334	373,717
Jun-18	333,883	355,246	294,556	373,211
Jul-18	332,427	352,035	293,049	371,805
Ago-18	333,584	359,136	294,187	372,981
Set-18	335,720	344,758	296,317	375,124
Oct-18	338,915	347,607	299,509	378,321
Nov-18	336,116	352,342	296,709	375,523
Dic-18	333,948	329,794	294,541	373,355

Fuente: Elaboración propia – SPSS

En la Tabla 7, se tiene los pronósticos a través del modelo SARMA (1, 0, 2) (1, 0, 0), para todos los meses de enero a diciembre del 2018. Además, se encuentra dentro de los límites de confianza predictiva.

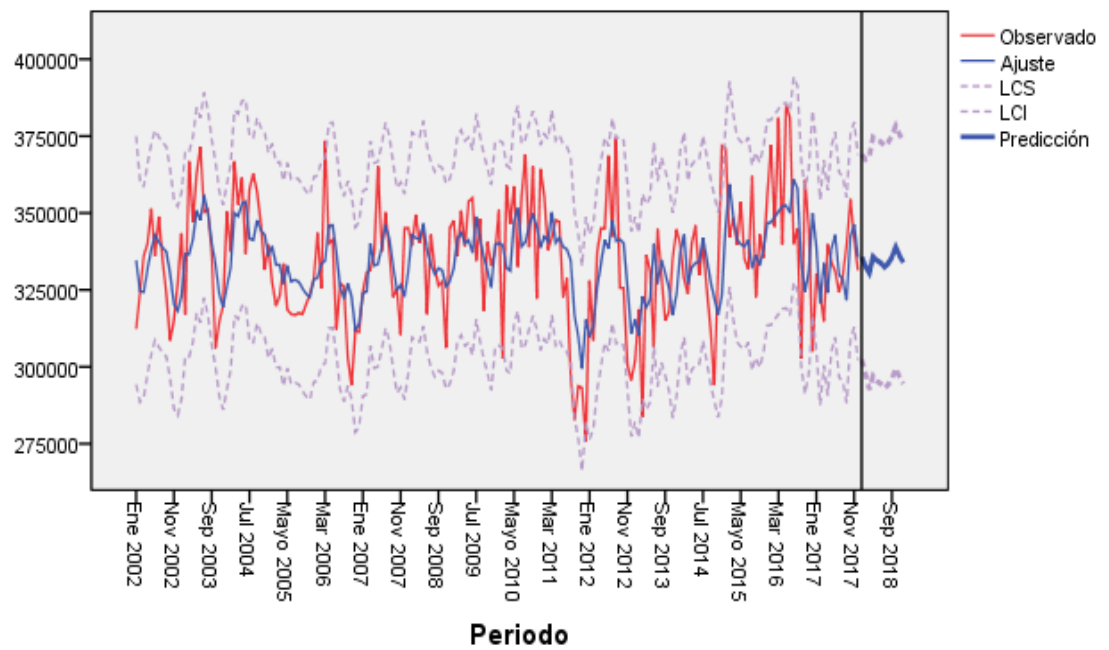


Figura 5: Serie de la producción de agua en el distrito de Lambayeque, con el modelo SARMA (1, 0, 2) (1, 0, 0), periodo enero 2002 a diciembre 2018.

En la Figura 5, se observa que el periodo predictivo calculado tiene congruencia en el procedimiento obtenido de los valores adecuados, lo que significa que la serie original de la producción de agua es de muy buena calidad predictiva. Además, el periodo predictivo tiene bajos porcentajes de error calculados.

➤ **Pronóstico 2019:**

En esta etapa usando el modelo que ha sido validado, se procede a realizar los pronósticos para el periodo enero 2019 a diciembre del 2019.

Tabla 8: Pronostico de la producción de agua en el distrito de Lambayeque, para el periodo enero 2019 a diciembre 2019.

Periodo	Pronostico SARMA (1, 0, 2) (1, 0, 0)	Límite inferior al 95%	Límite superior al 95%
Ene-19	334,873	294,804	374,942
Feb-19	334,192	294,058	374,325
Mar-19	333,755	293,497	374,012
Abr-19	334,901	294,598	375,205
May-19	334,622	294,301	374,942
Jun-19	334,484	294,157	374,811
Jul-19	334,172	293,843	374,500
Ago-19	334,420	294,090	374,750
Set-19	334,878	294,548	375,208
Oct-19	335,564	295,233	375,894
Nov-19	334,963	294,633	375,293
Dic-19	334,498	294,168	374,828

Fuente: Elaboración propia – SPSS

En la Tabla 8, se tiene los pronósticos a través del modelo SARMA (1, 0, 2) (1, 0, 0), para todos los meses de enero a diciembre del 2019. Además, se encuentra dentro de los límites de confianza predictiva.

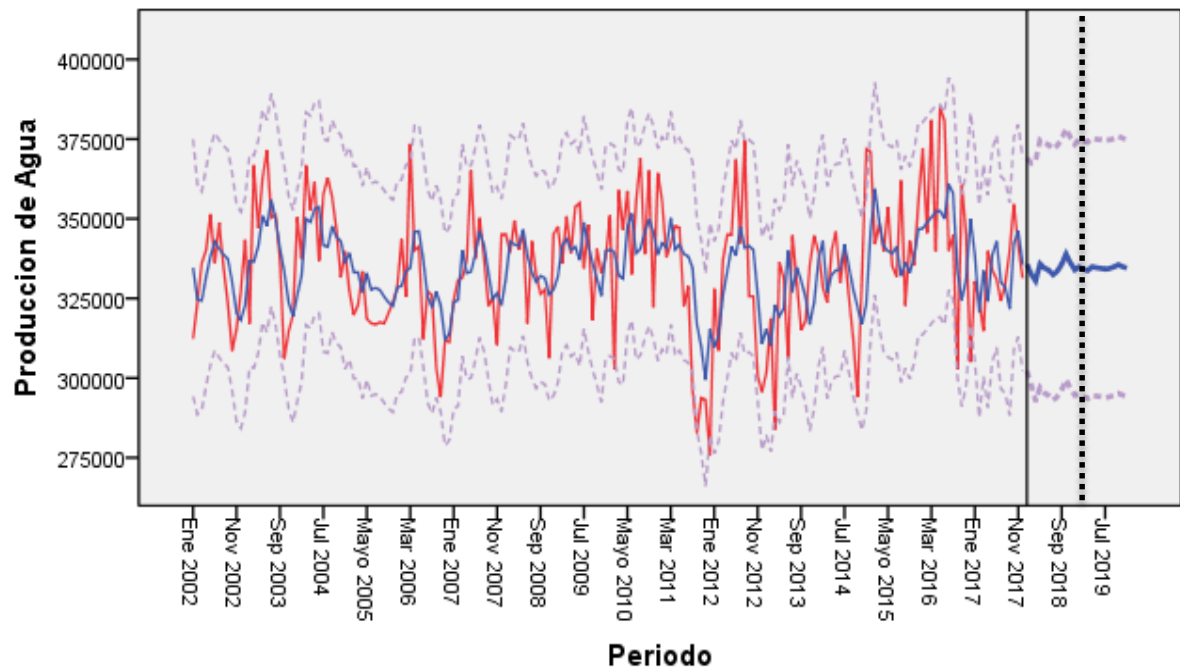


Figura 6: Serie de la producción de agua en el distrito de Lambayeque, con el modelo SARMA (1, 0, 2) (1, 0, 0), periodo enero 2002 a diciembre 2019.

En la Figura 6, se observa que el periodo predictivo calculado tiene un comportamiento ligeramente constante en el tiempo de los meses pronosticado, concluyendo que el procedimiento obtenido de los valores adecuados, lo que significa que la serie original de la producción de agua es de muy buena calidad predictiva. Además, el periodo predictivo tiene bajos porcentajes de error calculados.

2. Aplicación de la metodología Box- Jenkins para la demanda de agua

Tabla 9: Estadísticos descriptivos de la demanda de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

Estadístico	Demanda
Media	156,633.58
Desviación estándar	13,866.68
Coeficiente de Variación	8.85%
Mínimo	128,517
Máximo	181,660

Fuente: Elaboración propia.

En la Tabla 9, se presentan los estadísticos descriptivos de la demanda de agua del distrito de Lambayeque, donde se tiene que en promedio es 156,633.58 metros cúbicos, con una desviación típica de $\pm 13,866.68$, un coeficiente de variación de 8.85% es decir que los datos son homogéneos en torno a la media, el mínimo de producción de agua ha sido de 128,517 y un máximo de 181,660 metros cúbicos respectivamente.

2.1 Análisis de la estacionariedad de los datos

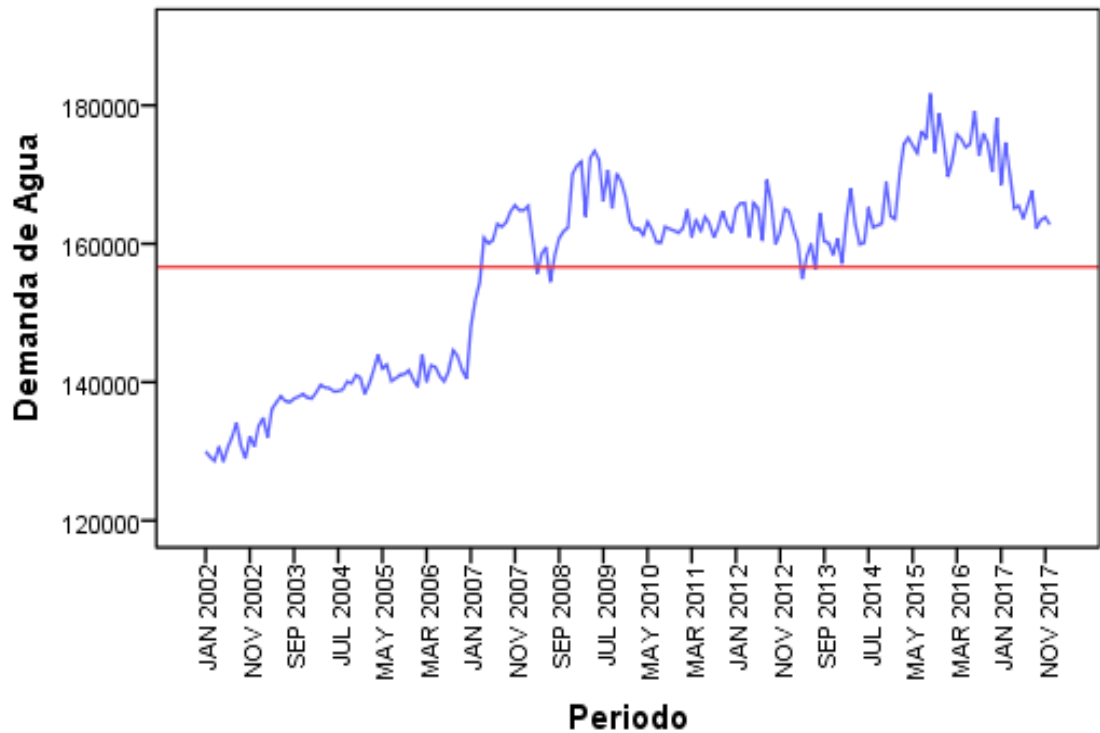


Figura 7: Serie original de la demanda de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

En la Figura 7, se observa que el comportamiento de la demanda de agua tiene una tendencia creciente, esto nos indica que la serie no es estacionaria en varianza y media, ya que no se mantienen constantes a través del tiempo, y será verificada por la prueba de Dickey Fuller aumentada.

Tabla 10: Prueba de raíz unitaria de Dickey - Fuller para la demanda de agua del distrito de Lambayeque periodo enero 2002 a diciembre 2017.

Augmented Dickey-Fuller test statistic	t-Statistic	Prob.*
	-1.976456	0.2971
1% level	-3.464827	
5% level	-2.876595	
10% level	-2.574874	

Fuente: Elaboración propia - Eviews

En la Tabla 10, se observa la prueba de Dickey – Fuller con un p- valor > 0.05, por lo que se concluye que la serie de la demanda de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017, no es estacionaria, con un nivel de significancia del 5% . En consecuencia, la serie tiene que ser diferenciada y volverla estacionaria, para poder aplicar la metodología Box – Jenkins. (VER ANEXO K)

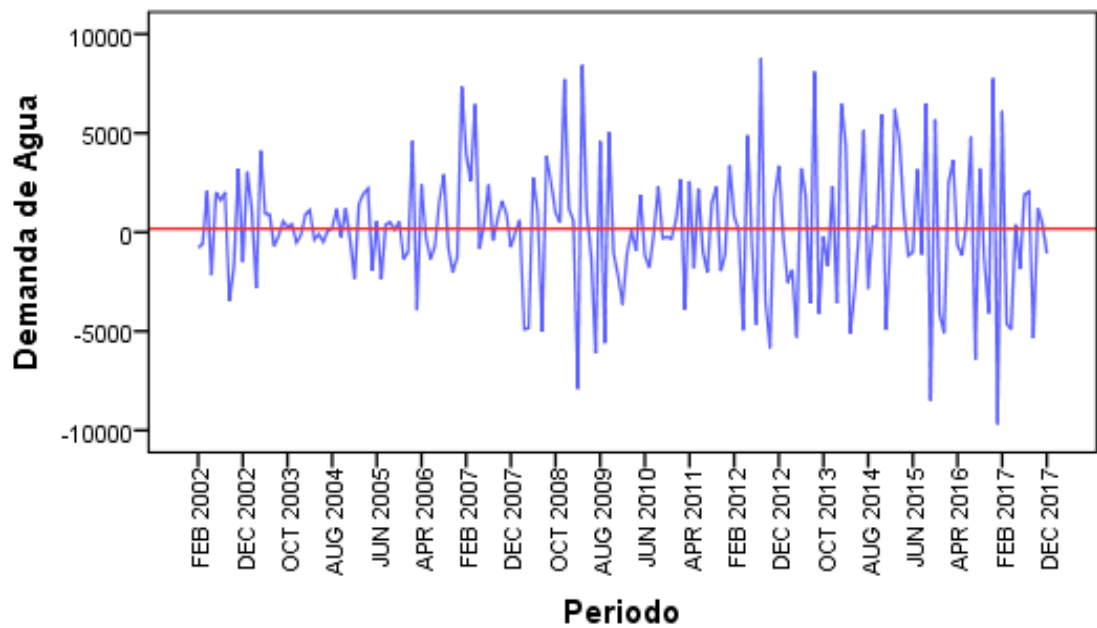


Figura 8: Diferencia en 1 periodo de la demanda de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

En el Figura 8, se presenta la serie diferencia en un periodo de la demanda de agua del distrito de Lambayeque, en la cual observamos que los datos oscilan alrededor de la media y tiene una varianza constante, perdiendo su tendencia y volviéndose estacionaria.

Para cumplir con los supuestos de estacionariedad que exige la metodología de Box –Jenkins, se utiliza las funciones de autocorrelación y la prueba de raíces unitaria de Dickey –Fuller Aumentada.

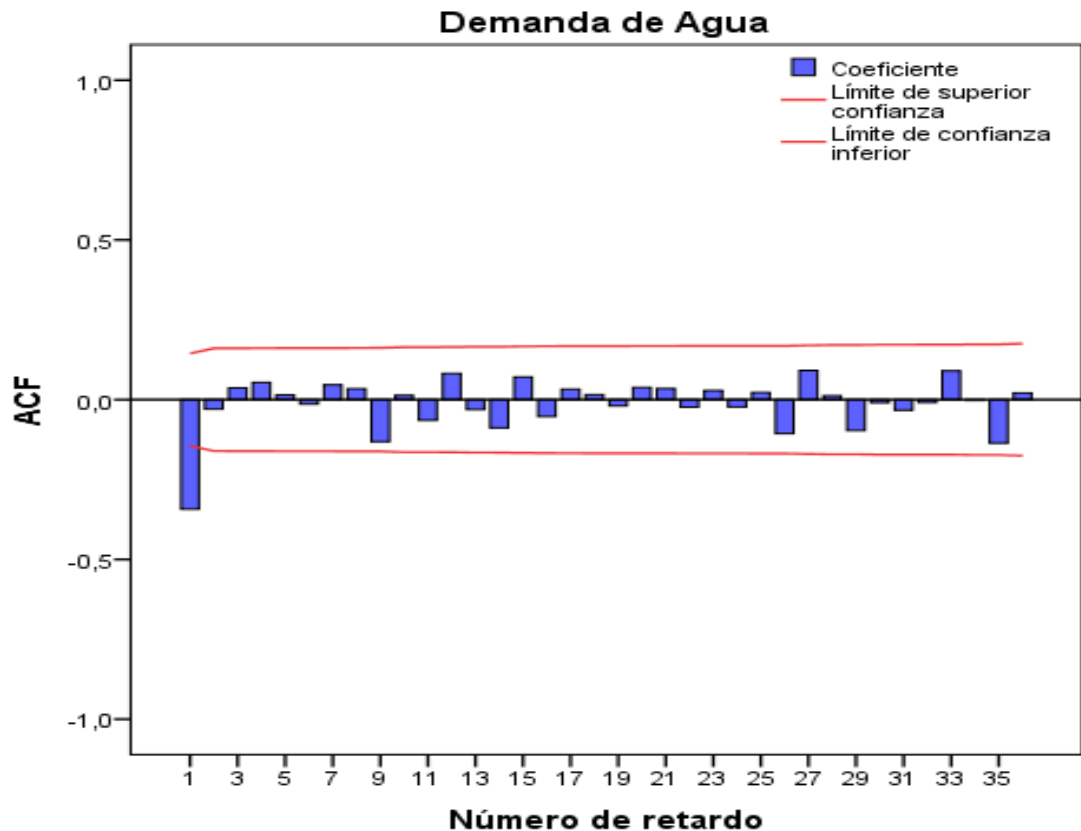


Figura 9: Función de Autocorrelación Simple de la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.

En la Figura 9, observamos un decaimiento brusco a partir del primer coeficiente de autocorrelación por lo que se puede afirmar que la serie de la demanda de agua en el distrito de Lambayeque en el periodo enero 2002 a diciembre 2017 diferenciada en un periodo, tiene un comportamiento estacionario. Así mismo se observa que la serie no presenta un comportamiento estacional, por lo que se presume que el modelo buscado tenga componentes MA en la parte regular.

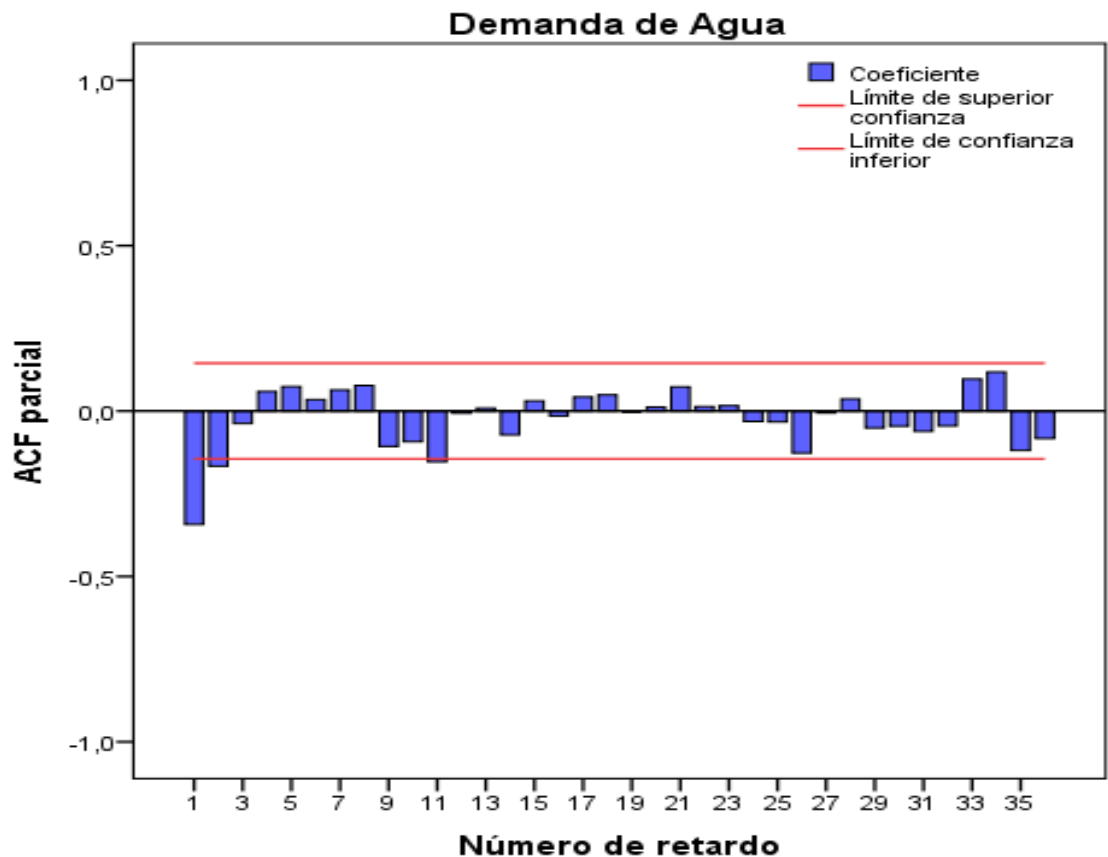


Figura 10: Función de Autocorrelación Parcial de la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.

En la Figura 10, se observa un decaimiento brusco a partir del segundo coeficiente de autocorrelación parcial, lo que nos indica que la serie de la demanda de agua en el distrito de Lambayeque en el periodo enero 2002 a diciembre 2017 diferencia en un periodo es estacionario. A si mismo se observa que no presenta evidencia de comportamiento estacional, lo cual se presume que el modelo buscado tenga componentes AR en la parte regular.

Tabla 11: Prueba de raíz unitaria de Dickey - Fuller para la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.

Augmented Dickey-Fuller test	t-Statistic	Prob.*
statistic	-19.56337	0.0000
1% level	-2.57719	
5% level	-1.942508	
10% level	-1.615589	

Fuente: Elaboración propia – Eviews.

En la Tabla 11, se observa la prueba de Dickey – Fuller con un p- value < 0.05, por lo que se concluye que la serie de la demanda de agua en el distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017, es estacionaria, con un nivel de significancia del 5%. (VER ANEXO L)

Cumpliendo de esta manera las condiciones de estacionariedad para la metodología ARIMA de Box – Jenkins.

2.2 Identificación de los modelos

Una vez confirmada la estacionariedad de la serie de la demanda de agua, procedemos a la etapa de identificación para determinar cuál es el modelo que mejor se ajusta, a partir de las funciones de autocorrelación que sirven como herramientas en esta etapa.

En base a las autocorrelaciones obtenidas en las figuras N° 7 y 8, se identificaron diferentes modelos, a los cuales se les realizó pruebas siguiendo el proceso de la metodología Box – Jenkins, descartando los que no cumplían con los requisitos.

MODELOS TENTATIVOS IDENTIFICADOS:

- MODELO 1: ARIMA (1,1,0)
- MODELO 2: ARIMA (0,1,1)
- MODELO 3: ARIMA (2,1,0)

Después de llevar a cabo el proceso de identificación de los modelos tentativos se procederá a la estimación, para la serie de la demanda de agua en el distrito de Lambayeque.

2.3 Estimación de los parámetros de los modelos tentativos

La siguiente etapa consiste en determinar el modelo más adecuado para la serie de la demanda de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

Para decidir si se incluye un término independiente (Constante), en la estimación de los parámetros del modelo, se contrastará si la media de la serie estacionaria es o no cero.

Tabla 12: Prueba t para una muestra para la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.

Demanda de Agua	t	gl	Sig. (bilateral)
	0.72	190.00	0.47

Fuente: Elaboración propia – SPSS

En la tabla 12, observamos la prueba t para una muestra con un p-valor > 0.05, concluyendo que la media de la serie de la demanda de agua en el distrito de Lambayeque a primera diferencia, en el periodo enero 2002 a diciembre 2017 no es diferente a cero, por lo tanto, no se incluirá un término independiente (constante) al evaluar los parámetros de los modelos tentativos. (VER ANEXO LL)

Tabla 13: Resumen del desempeño de los modelos tentativos identificados de la serie de la demanda de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

MODELO	Parámetro	Estimación	Significancia	RMSE	MAPE	MAE
ARIMA (1,1,0)	AR 1	0.315	0.00	3,094.66	1.47%	2,336.95
ARIMA (0,1,1)	MA 1	0.377	0.000	3,062.74	1.44%	2,303.09
ARIMA (2,1,0)	AR 1	-0.391	0.000	3,062.97	1.44%	2,305.21
	AR 2	-0.158	0.029			

Fuente: Elaboración propia - SPSS

En la Tabla 13, se presentan la estimación de los modelos tentativos propuestos en la etapa de la identificación. Se observa que los parámetros son significativos en los 3 modelos, entonces se procede a evaluar los estadísticos de error que son el RMSE (Error cuadrático medio), MAPE (Error porcentual absoluto medio) y el MAE (error absoluto medio), los cuales nos ayudaran a determinar qué modelo tiene menos errores de pronóstico. El mejor modelo es el ARIMA (0, 1,1) ya que tiene menos errores con un RMSE igual a 3 062.74, MAE igual a 2 303.09 y el MAPE que es igual a 1.44% el cual mide el error de pronóstico de manera porcentual comparado con los valores reales de la serie.

2.4 Validación del modelo

Después de seleccionar un modelo ARIMA y de estimar sus parámetros, se procede a la etapa de validación para ver si el modelo seleccionado para la serie de la demanda de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017, se ajusta a los datos en forma razonablemente buena, el detalle se encuentra en el análisis de los residuales, estos se deben aproximar a un comportamiento de ruido blanco.

Tabla 14: Prueba de normalidad para los residuales del modelo ARIMA (0, 1, 1)

Demanda de Agua	Kolmogorov-Smirnova		
	Estadístico	gl	Sig.
	0.05	191.00	0,200*

*, Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Fuente: Elaboración propia - SPSS

En la Tabla 14, se observa la prueba de normalidad de Kolmogorov – Smirnov, donde se obtiene que el nivel de significación es mayor al 0.05, por lo tanto se concluye que los residuales del modelo ARIMA (0, 1, 1) siguen una distribución normal, lo que indica que el modelo es satisfactorio. (VER ANEXO M)

Tabla 15: Prueba de aleatoriedad para los residuales del modelo ARIMA (0,1,1)

Valor de prueba ^a	234
Casos < Valor de prueba	95
Casos >= Valor de prueba	96
Casos totales	191
Número de rachas	99
Z	0.36
Sig. asintótica (bilateral)	0.72

a. Mediana

Fuente: Elaboración propia

En la Tabla 15, se observa la prueba de aleatoriedad o rachas, la hipótesis nula a contrastar es que los residuos del modelo son aleatorios. Como la significación de la prueba es de $0.72 > 0.05$, se concluye que los residuales del modelo ARIMA (0, 1, 1) son aleatorios. (VER ANEXO N)

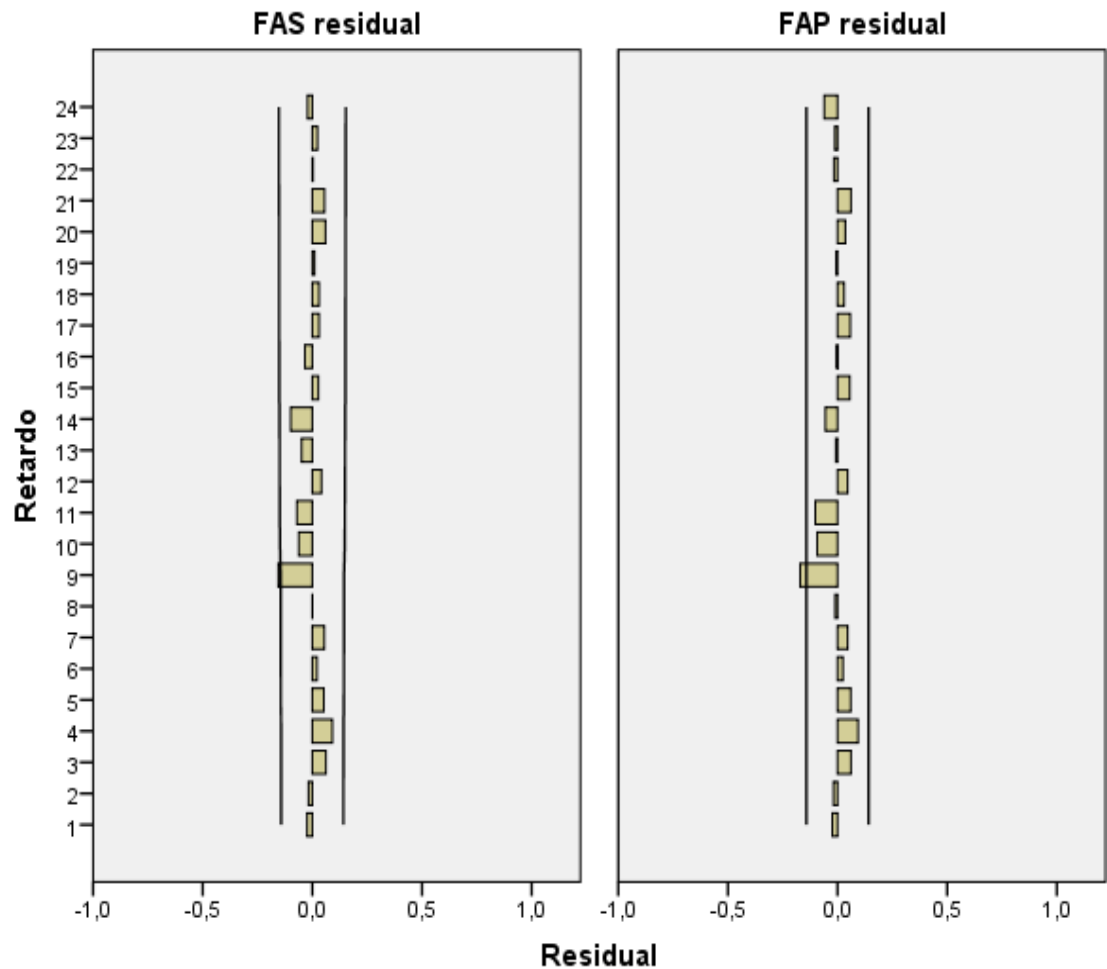


Figura 11: Función de Autocorrelación simple y parcial de los residuales del modelo ARIMA (0, 1, 1) de la demanda de agua en el distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

En la Figura 11, se observa que los coeficientes de autocorrelación simple y parcial de los residuales están dentro de los límites de confianza, por lo que se concluye que los residuales del modelo ARIMA (0, 1, 1), tienden a un ruido blanco.

2.5 Pronóstico del modelo

➤ Pronóstico 2018

En esta etapa usando el modelo que ha sido validado, se procede a realizar los pronósticos para el periodo enero 2018 a diciembre del 2018.

Tabla 16: Pronostico de la demanda de agua en el distrito de Lambayeque, para el periodo enero 2018 a diciembre 2018.

Periodo	Pronostico ARIMA (0, 1, 1)	Observado	Límite inferior al 95%	Límite superior al 95%
Ene-18	163,143	155,482	157,102	169,185
Feb-18	163,143	162,465	156,025	170,262
Mar-18	163,143	167,531	155,090	171,197
Abr-18	163,143	161,198	154,253	172,034
May-18	163,143	168,480	153,488	172,799
Jun-18	163,143	169,498	152,780	173,507
Jul-18	163,143	162,644	152,117	174,170
Ago-18	163,143	165,039	151,492	174,795
Set-18	163,143	162,791	150,898	175,389
Oct-18	163,143	175,529	150,332	175,955
Nov-18	163,143	170,789	149,790	176,497
Dic-18	163,143	174,174	149,269	177,018

Fuente: Elaboración propia

En la Tabla 16, se tiene los pronósticos a través del modelo ARIMA (0, 1, 1) para todos los meses de enero a diciembre del 2018.

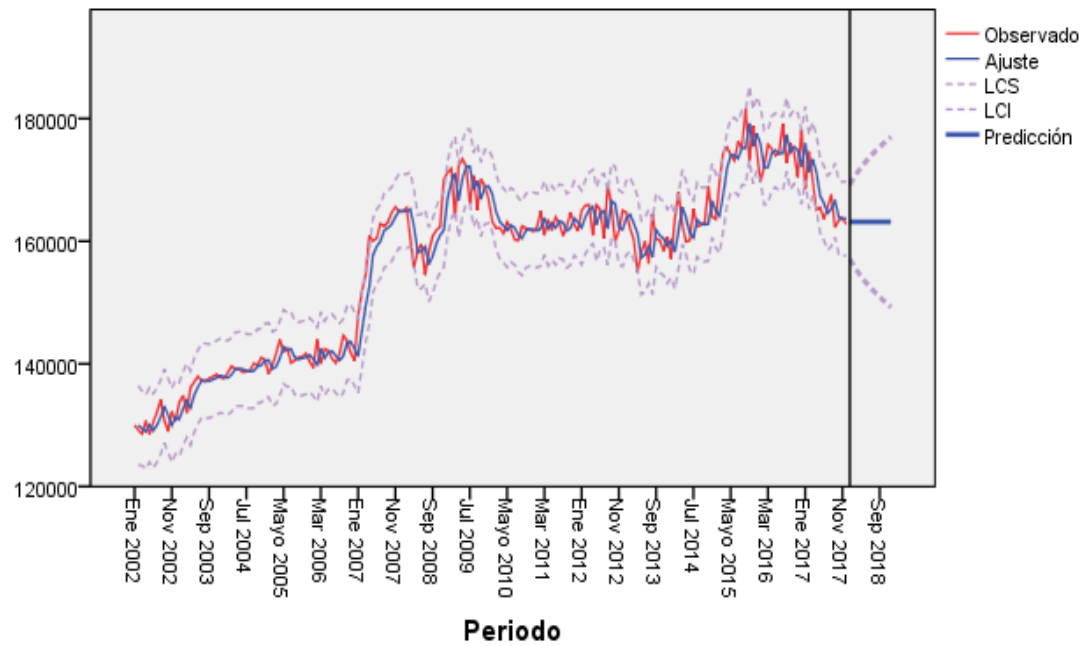


Figura 12: Serie de la demanda de agua en el distrito de Lambayeque, con el modelo ARIMA (0, 1, 1) periodo enero 2002 a diciembre 2018.

En la Figura 12, se observa que el periodo predictivo calculado sigue constante en el tiempo, ya que depende del modelo ARIMA (0,1,1) lo que significa que la serie original de la demanda de agua es de muy buena calidad predictiva para el año 2018.

➤ **Pronóstico 2019**

En esta etapa usando el modelo que ha sido validado, se procede a realizar los pronósticos para el periodo enero 2019 a diciembre del 2019.

Tabla 17: Pronostico de la demanda de agua en el distrito de Lambayeque, para el periodo enero 2019 a diciembre 2019.

Periodo	Pronostico ARIMA (0, 1, 1)	Límite inferior al 95%	Límite superior al 95%
Ene-19	163,143	148,767	177,520
Feb-19	163,143	148,282	178,005
Mar-19	163,143	147,812	178,474
Abr-19	163,143	147,357	178,930
May-19	163,143	146,914	179,373
Jun-19	163,143	146,483	179,804
Jul-19	163,143	146,062	180,224
Ago-19	163,143	145,652	180,635
Set-19	163,143	145,251	181,035
Oct-19	163,143	144,859	181,427
Nov-19	163,143	144,476	181,811
Dic-19	163,143	144,099	182,187

Fuente: Elaboración propia

En la Tabla 17, se tiene los pronósticos a través del modelo ARIMA (0, 1, 1) para todos los meses de enero a diciembre del 2019.

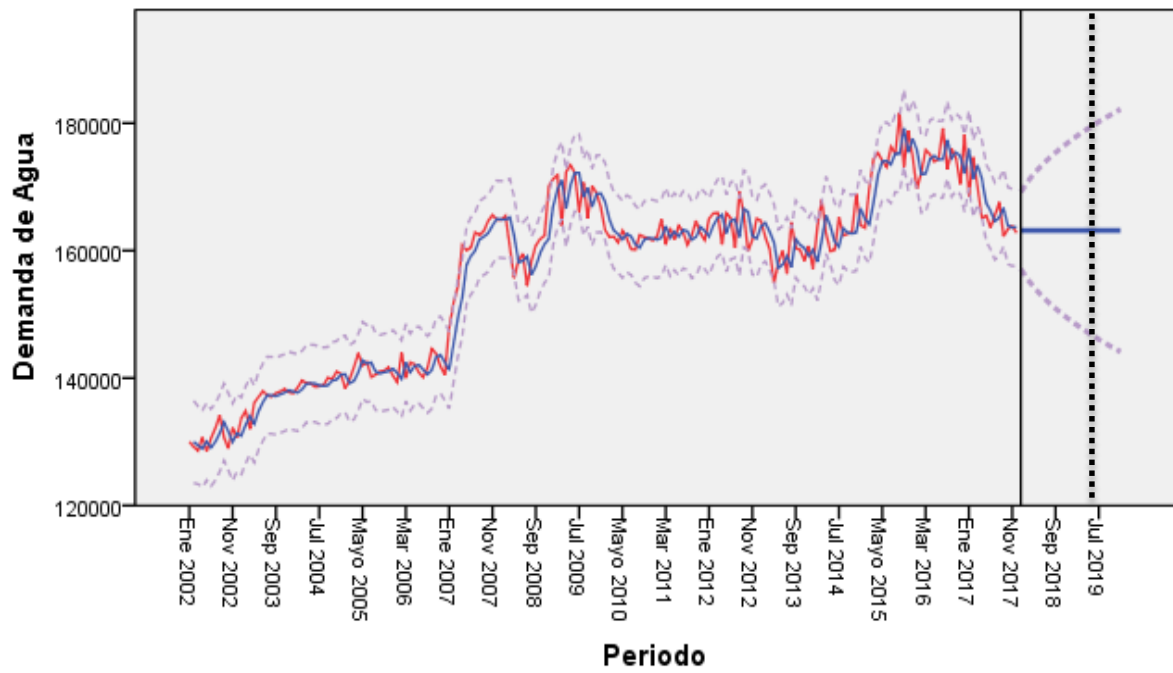


Figura 13: Serie de la demanda de agua en el distrito de Lambayeque, con el modelo ARIMA (0, 1, 1) periodo enero 2002 a diciembre 2019.

En la Figura 13, se observa que el periodo predictivo calculado sigue constante en el tiempo, ya que depende del modelo ARIMA (0,1,1) lo que significa que la serie original de la demanda de agua es de muy buena calidad predictiva.

CONCLUSIONES Y SUGERENCIAS

PRODUCCIÓN DE AGUA

1. En la serie de los datos originales de la variable producción de agua se observa a simple vista estacionariedad (con media y varianza constante en el tiempo). Además, no se observan componentes de tendencia.
2. Para confirmar que la serie presenta estacionariedad, se aplicó la prueba de Dickey Fuller que resultó ser estacionaria con una significancia del 5%, de acuerdo a dicha prueba.
3. En el proceso de identificación se encontraron modelos tentativos y satisfactorios de pronóstico para la serie original y los cuales son los siguientes: ARMA (2, 0, 2); SARMA (2, 0, 0) (1, 0, 0); SARMA (1, 0, 2) (1, 0, 0).
4. Se evaluó los estadísticos de error y son: (RMSE, MAPE, MAE). Nosotros elegimos el menor estadístico de error de pronóstico, siendo el mejor modelo SARMA (1, 0, 2) (1, 0, 0) con un RMSE = 16,932.67 un MAPE = 3.98% y un MAE = 13,254.61; por lo que se procedió a pronosticar para el año 2018 y 2019.
5. El modelo no integrado de la metodología Box-Jenkins proporciona un mejor modelo de ajuste que el modelo integrado.

DEMANDA DE AGUA

1. La serie de los datos originales de la variable demanda de agua se observa que no tiene estacionariedad (sin media y ni varianza constante en el tiempo). Además, presentó componente de tendencia.
2. La serie no presenta estacionariedad, por lo que se optó diferenciarla en un periodo, para luego aplicar la prueba de Dickey Fuller con primera diferencia. La serie resultó ser estacionaria con un nivel de significancia del 5%, de acuerdo a dicha prueba.
3. En el proceso de identificación se encontraron modelos tentativos y satisfactorios de pronóstico para la serie original y son las siguientes: ARIMA (1,1,0); ARIMA (0,1,1); ARIMA (2,1,0).
4. Se evaluó los estadísticos de error y son: (RMSE, MAPE, MAE). Eligiéndose el menor estadístico de error de pronóstico, siendo el mejor modelo ARIMA (0,1,1) con un RMSE = 3,062.74 un MAPE = 1.44% y un MAE = 2,303.09; por lo que se procedió a pronosticar para el año 2018 y 2019.
5. El modelo integrado de la metodología Box-Jenkins proporciona un mejor modelo de ajuste que el modelo no integrado.

RECOMENDACIONES

1. Dar conocimiento al jefe del Instituto Nacional de Estadística e Informática (INEI) y la Entidad Prestadora de Servicios de Saneamiento de Lambayeque (EPSEL), para que utilicen el modelo óptimo de pronóstico SARMA(1,0,2)(1,0,0) para la producción y el modelo óptimo ARIMA (0,1,1) para la demanda de agua del distrito de Lambayeque para realizar futuros indicadores.
2. Utilizar otras metodologías de pronóstico para la producción y demanda de agua del distrito de Lambayeque; como el procedimiento de inteligencia artificial (redes neuronales), utilizando software más potente como el SAS, Rstudio y Phyton.
3. Actualizar, controlar, analizar, evaluar y renovar periódicamente la base de datos de los modelos óptimos de pronóstico para identificar los cambios que se puedan producir en cierto periodo.
4. Dado que se desarrolló un agregado en la investigación sobre la evolución poblacional y el consumo per cápita en litros de agua del distrito de Lambayeque, recomendamos que futuras investigaciones integren la data histórica de la población para dar a conocer el déficit del agua potable.

BIBLIOGRAFÍA

- Alegría, M. (2013). Ampliación y mejoramiento del sistema de agua potable de la ciudad de Bagua Grande. (Tesis Pregrado). Universidad Nacional de Ingeniería, Lima. Recuperado de: http://cybertesis.uni.edu.pe/bitstream/uni/1175/1/alegria_mj.pdf
- Bokava et al (2017). Desarrollo de los Recursos Hídricos en el mundo, “Aguas residuales, el recurso desaprovechado”. Recuperado de: <http://unesdoc.unesco.org/images/0024/002476/247647s.pdf>
- Bogginano, M. (2013-2017). Los 10 países con más reservas de agua del mundo. Carta Financiera. Recuperado de: <http://www.cartafinanciera.com/tendencia-actual/los-10-paises-con-mas-reservas-de-agua-del-mundo>
- Box, G., Jenkins, G., Reinsel, G., Ljung, G. (2015). Time series analysis: Forecasting and control.- Fifth edition. San Francisco. USA: Holden day.
- Cáceres, V. (2013). La provisión del servicio de agua y saneamiento en la provincia de Buenos Aires: Fragmentación y Desigualdad. (Revista Científica). Universidad Nacional de General Sarmiento, Buenos Aires.
- Camasca, L. (2017). Lambayeque sufre por falta de agua pese a inversión de S/ 1,500 millones en obras. Diario Correo. Recuperado de: <https://diariocorreo.pe/edicion/lambayeque/lambayeque-sufre-por-falta-de-agua-pese-a-inversion-de-s-1-500-millones-en-obras-752803>
- Cuyo, J. (2016). “Propuesta metodológica para el pronóstico de la demanda de agua potable a corto plazo (días, semanas, meses) en la planta de tratamiento de agua potable de bellavista en el norte de la ciudad de quito”. Ecuador. Recuperado de: <http://www.dspace.uce.edu.ec/bitstream/25000/8729/1/T-UCE-0005-013-2016.pdf>
- Damián, Ll. y Sandoval, S. (2018). Modelo óptimo de pronóstico del índice mensual de producción de electricidad, Perú en el periodo 2006 a 2015 (Tesis de Pregrado). Universidad Nacional “Pedro Ruiz Gallo”, Lambayeque. Recuperado de: <http://repositorio.unprg.edu.pe/handle/UNPRG/2192>
- Daza, F. (2008). "Demanda de agua en zonas urbanas en Andalucía". Córdoba. Recuperado de: <https://helvia.uco.es/bitstream/handle/10396/279/9788478019281.pdfsequence=1&isAllowed=y>
- Dickey, D. A. y Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, J. Amer. Stat. Assoc.
- Entidad Prestadora de Servicios de Saneamiento de Lambayeque (Enero 2017). Obtenido de EPSEL Sitio Web: <http://www.epsel.com.pe/Portal/servicio-de-alcantarilladoFAO>

- Forbes México (08 de junio del 2019). “CDMX, entre las 3 capitales en el mundo en riesgo de quedarse sin agua”. Recuperado de: <https://www.forbes.com.mx/cdmx-entre-las-3-capitales-en-el-mundo-en-riesgo-de-quedarse-sin-agua/>
- Gonzalo, R. (2008). Tópicos en Minería de Datos. Series de Tiempo (p.16-18). Chile. Recuperado de: https://www.ucursos.cl/ingenieria/2010/1/CC52A/1/material_docente/bajar_arid_material=2_96003
- González, M. (2009). Análisis de series temporales: Modelos ARIMA (p.19-48) España: Vasco. Recuperado de: <https://addi.ehu.es/bitstream/10810/12492/1/04-09gon.pdf>
- Gordillo, H. (2015). Lambayeque: 90 mil familias aún o tienen agua potable. Diario Correo. Recuperado de: <https://diariocorreo.pe/edicion/lambayeque/lambayeque-90-mil-familias-aun-no-tienen-agua-potable-612760/>
- Gujarati, D. y Porter, D. (2009). Econometría quinta edición. México: The McGraw - Hill. Recuperado de: https://scalleruizunp.files.wordpress.com/2015/04/econometria_damodarr_n-gujarati.pdf
- Hernández, S., Hernández, C. y Baptista, L. (2014). Metodología de la Investigación. Sexta edición. Mexico: McGraw-Hill Education.
- Instituto Nacional de Estadística e Informática (10 de diciembre del 2014). Anuario de Estadística ambientales. Recuperado de: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1197/libro.pdf / https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1531/lambaye.htm
- Instituto Nacional de Estadística e Informática (2015), “Magnitud y crecimiento poblacional” y “Estimaciones y proyecciones de población”. Periodo 2002-2015. Recuperado de: <https://www.inei.gob.pe/estadisticas/indice-tematico/population/>
- Koohafkan et al (1 de octubre del 2011). El estado de los recursos de tierras y aguas del mundo para alimentación y la agricultura. Recuperado de: <http://www.fao.org/docrep/015/i1688s/i1688s00.pdf>
- Koncagül, Connor y Tran (21 de marzo del 2014). Desarrollo de los Recursos Hídricos en el mundo, “Agua y Energía”. Recuperado de: <http://unesdoc.unesco.org/images/0022/002257/225741E.pdf>
- Leonardo, J. (2017). “Modelo Univariante para el Consumo Doméstico mensual de Agua Potable en el Distrito de Ilave – EMSA Puno, periodo 2002-2013”. Universidad Nacional del Altiplano. Perú – Puno.

- Limache, S. (2015). Relación entre las variables de influencia y variables de resultado del consumidor de agua potable en la ciudad de Tacna - 2014. (Tesis Doctoral). Universidad privada de Tacna, Tacna.
- Marx, K. y Engels, F. (1859). "El proceso de Producción del Capital". Biblioteca de Autores Socialistas. Tomo I. Alemania: Siglo XXI Editores. Recuperado de: <https://webs.ucm.es/info/bas/es/marx-eng/capital1/>
- Marx, K. (1864). "El proceso global de la producción capitalista". Biblioteca de Autores Socialistas. Tercer edición. Alemania: Siglo XXI Editores. Recuperado de: <https://webs.ucm.es/info/bas/es/marx-eng/capital3/MRXC3601.htm>
- Mason, D., Lind, A. y Marchal, G. (2001). Estadística para Administración y Economía. The University of Toledo. Décima edición. Colombia: Alfaomega. Series de tiempo y predicción (p. 647 – 652).
- Mauricio, J. (2007). Introducción al Análisis de series Temporales. Universidad Complutense de Madrid. España. Recuperado de: <https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf>
- Ministerio de la salud (2016). "Población estimada por edades simples y grupos de edad, según provincia y distrito". Periodo 2004-2016. Recuperado de: <http://www.minsa.gob.pe/estadisticas/estadisticas/Poblacion/PoblacionMarcos.asp?14>
- Ministerio de Agricultura y Riego (24 de abril del 2009). Política y estrategia Nacional de Recursos Hídricos del Perú. Recuperado de: http://www.ana.gob.pe/media/290336/politicas_estrategias_rh.pdf
- Murat, K., Hamidreza, S. (2015). "A Framework for Sustainable Urban Water Management through Demand and Supply Forecasting: The Case of Istanbul". Turquía. Recuperado de: <http://www.mdpi.com/2071-1050/7/8/11050>
- Novales, A. (1993). Econometría segunda Edición. Madrid, España: Isabel Capella. Recuperado de: <https://econometria101.files.wordpress.com/2013/02/econometria-2ed-a-novales.pdf>
- Organización de Naciones Unidas (18 de enero del 2013). La desertificación, "Esa invisible línea de frente". Recuperado de: https://www.unccd.int/sites/default/files/documents/12112014_Invisible%20frontline_SP_0.pdf
- Pimentel y Palacios (22 de marzo del 2017). El agua es un bien escaso que el Perú no sabe administrar. RPP Noticias. Recuperado de: <https://rpp.pe/peru/actualidad/la-falta-de-agua-potable-afecta-a-8-millones-de-peruanos-noticia-998969>

- Spiegel, R. y Stephens J. (2002). Estadística. University of Nebraska at Omaha. Tercera edición. Mexico: McGraw-Hill Interamericana. Análisis de series de tiempo, CAP 18 (p. 436 – 442).
- Siegel, S., & Castellan, J. (1998) Estadística no Paramétrica: Aplicada a las ciencias de la conducta (Cuarta ed.). México: Trillas.
- Siesquén, A. y Riojas, Y. (2018). Modelos de pronóstico del índice de precios al consumidor región Lambayeque y a nivel nacional. Enero 2010 a diciembre 2016. (Tesis de Pregrado). Universidad Nacional “Pedro Ruiz Gallo”, Lambayeque. Recuperado de:
<http://repositorio.unprg.edu.pe/handle/UNPRG/2977>
- Ramos, N. (2013). 390 mil habitantes de Lambayeque no cuentan con agua ni desagüe. La Republica. Recuperado de <http://larepublica.pe/archivo/733520-390-mil-habitantes-de-lambayeque-no-cuentan-con-agua-ni-desague>
- Reinhard, S. (2009). Análisis de la situación del agua (cantidad y residual) en Lima Metropolitana. Recuperado de:
http://www.lima-water.de/documents/rseifert_studie.pdf
- Vásquez, M. (2017). Modelo de Box Jenkins y Redes Neuronales para pronosticar el precio el dólar del sistema bancario en moneda nacional al año 2015. (Tesis de Pregrado). Universidad Nacional “Pedro Ruiz Gallo”, Lambayeque.
- Vidaurre, Y. (2012). “Aplicación de las Redes Neuronales Artificiales para el Pronóstico de la Demanda de Agua Potable en la Empresa EPSEL S.A. de la ciudad de Lambayeque”. Universidad Católica Santo Toribio de Mogrovejo. Perú – Chiclayo. Recuperado de:
http://tesis.usat.edu.pe/bitstream/usat/525/1/TL_Vidaurre_Siaden_Yasmin.pdf
- Villavicencio, J. (2007). Introducción a series de tiempo. Obteniendo de Instituto de Estadística de Puerto Rico. Recuperado de
http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D
- Villaseñor, A. (2011). Prueba de bondad de ajuste para distribuciones con parámetro de forma. Obteniendo por el Colegio de Graduados de México.
- Wooldridge, J. (2010). Introducción a la econometría un enfoque moderno. Universidad del Estado Michigan, 4ª. Edición: Cengage Learning Editores S.A.
- Vidal, G. (2015). Realizó una investigación titulada “Modelo SARIMA para el pronóstico de la masa mensual del río Jequetepeque”, (Tesis de Pregrado). Universidad Nacional de Trujillo, Trujillo.

APÉNDICE Y ANEXOS

ANEXO A

PRODUCCIÓN DE AGUA EN MC DEL DISTRITO DE LAMBAYEQUE – PERIODO 2002-2017 INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA (INEI). LAMBAYEQUE – PERÚ

MESES	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Enero	312,290	343,367	350,478	327,268	343,542	324,418	345,149	347,530	350,958	355,549	328,037	301,703	339,500	370,944	372,075	330,392
Febrero	323,794	317,038	337,476	319,835	325,490	330,714	339,445	336,010	302,703	338,010	308,584	318,700	289,100	306,785	345,520	324,048
Marzo	336,083	366,797	366,796	323,068	373,506	331,458	349,445	350,749	358,982	342,134	336,964	283,827	323,680	348,443	380,985	314,823
Abril	340,111	346,991	352,720	333,512	340,111	336,929	340,304	339,212	346,494	347,692	345,012	336,377	339,831	339,690	339,831	339,895
Mayo	351,396	362,747	361,526	318,628	341,386	365,169	346,191	353,827	358,708	347,203	344,945	331,301	345,940	353,584	384,759	333,925
Junio	336,082	371,570	336,781	317,159	312,206	337,479	317,168	354,974	332,520	322,691	368,666	306,408	329,840	335,118	380,912	330,991
Julio	348,834	350,342	357,752	316,817	327,156	350,241	342,958	334,435	355,295	328,837	342,374	344,754	339,780	331,716	339,970	324,240
Agosto	334,836	351,115	362,818	317,551	326,040	337,312	332,193	348,131	369,057	296,304	374,522	330,207	325,442	361,990	344,820	329,655
Septiembre	323,565	292,147	356,871	317,142	303,081	322,679	326,383	318,213	338,895	282,743	281,674	315,082	312,365	322,678	302,684	339,625
Octubre	308,602	306,130	346,718	320,409	294,028	325,265	327,807	340,603	365,106	293,742	280,926	317,924	294,018	343,101	360,735	354,521
Noviembre	315,230	314,558	331,546	323,788	311,831	310,408	306,209	332,846	322,177	293,169	300,619	335,586	331,201	335,384	276,965	341,483
Diciembre	327,218	319,882	339,736	328,542	311,290	345,098	345,187	337,764	364,322	275,813	295,643	344,680	371,840	355,191	305,010	331,381

ANEXO B

DEMANDA DE AGUA EN MC DEL DISTRITO DE LAMBAYEQUE – PERIODO 2002-2017.

INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA (INEI). LAMBAYEQUE - PERÚ

MESES	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Enero	129,988	133,708	137,612	138,323	139,346	162,102	164,843	178,265	163,204	162,215	165,024	164,667	157,116	163,572	169,677	168,530
Febrero	129,216	134,844	138,510	139,761	143,961	161,175	165,452	171,842	162,122	164,890	165,842	162,121	167,414	183,948	172,146	174,640
Marzo	128,611	132,015	139,618	141,705	140,023	164,252	154,764	163,930	162,204	160,970	165,928	154,044	167,949	189,133	175,781	170,021
Abril	130,692	136,127	139,238	143,911	142,433	160,867	155,712	172,357	161,273	163,535	160,982	154,910	153,598	186,389	175,116	160,556
Mayo	128,517	137,060	139,137	141,972	142,185	160,039	158,469	173,420	163,156	161,731	165,874	158,131	159,906	174,141	173,947	158,536
Junio	130,531	137,960	138,646	142,525	140,843	160,481	159,506	172,195	161,994	163,908	165,116	159,957	160,113	173,132	174,395	163,637
Julio	132,143	137,270	138,715	140,152	140,113	162,858	154,495	166,096	160,226	162,910	160,436	156,363	165,242	176,292	179,193	165,542
Agosto	134,157	137,075	138,908	140,543	141,670	162,440	158,335	177,317	160,129	160,902	169,210	164,456	162,383	175,167	172,743	167,586
Septiembre	130,706	137,633	140,074	141,053	144,586	163,092	160,828	165,078	162,450	162,399	165,800	160,350	162,636	181,660	175,964	162,234
Octubre	128,989	137,891	139,825	141,173	143,803	164,671	161,830	178,700	162,142	164,688	159,962	160,125	162,932	173,162	174,541	163,423
Noviembre	132,196	138,296	141,015	141,704	141,799	165,582	162,325	169,044	161,907	162,734	161,661	158,401	170,664	178,831	170,450	163,847
Diciembre	130,677	137,768	140,659	140,355	140,520	164,834	170,054	166,848	161,577	161,662	165,007	168,907	163,963	174,746	178,217	162,773

ANEXO C

NÚMERO DE HABITANTES DEL DISTRITO DE LAMBAYEQUE – PERIODO 2000-2019.

INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA (INEI) / MINSA. LAMBAYEQUE - PERÚ

Año	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Total	57,364	58,644	59,923	56,863	58,298	59,160	64,420	66,694	67,997	68,464	69,100	71,927	73,252	74,576	75,905	77,234	77,855	79,772	81,736	83,749

Nota: Los últimos años son estimados por los autores.

ANEXO D

Tabla 18: Valores del Correlograma de la serie de la Producción de Agua en el distrito de Lambayeque. Periodo 2002-2017.

	AC	PAC	Q-Stat	Prob
1	0.435	0.435	36.837	0.000
2	0.424	0.290	72.036	0.000
3	0.235	-0.029	82.903	0.000
4	0.066	-0.165	83.760	0.000
5	0.059	0.026	84.461	0.000
6	-0.190	-0.239	91.675	0.000
7	-0.123	0.001	94.727	0.000
8	-0.168	0.018	100.420	0.000
9	-0.120	0.037	103.370	0.000
10	-0.052	0.028	103.930	0.000
11	-0.012	0.070	103.960	0.000
12	0.158	0.149	109.140	0.000
13	0.115	-0.024	111.890	0.000
14	0.071	-0.154	112.940	0.000
15	0.127	0.071	116.310	0.000
16	0.053	0.008	116.910	0.000
17	0.054	-0.029	117.520	0.000
18	0.009	0.052	117.540	0.000
19	-0.007	0.051	117.550	0.000
20	0.031	0.029	117.760	0.000
21	-0.017	-0.015	117.820	0.000
22	0.076	0.104	119.090	0.000
23	-0.046	-0.150	119.550	0.000
24	0.000	-0.066	119.550	0.000
25	-0.095	-0.109	121.540	0.000
26	-0.205	-0.141	130.970	0.000
27	-0.162	-0.051	136.900	0.000
28	-0.187	0.086	144.810	0.000
29	-0.168	-0.032	151.230	0.000
30	-0.197	-0.101	160.140	0.000
31	-0.137	0.000	164.510	0.000
32	-0.117	-0.073	167.690	0.000
33	-0.020	0.015	167.780	0.000
34	0.056	0.053	168.520	0.000
35	0.006	-0.087	168.530	0.000
36	0.046	-0.045	169.020	0.000

ANEXO E

Tabla 19: Valores del Correlograma de la serie de la Demanda de Agua en el distrito de Lambayeque. Periodo 2002-2017.

	AC	PAC	Q-Stat	Prob
1	0.962	0.962	180.410	0.000
2	0.942	0.224	354.330	0.000
3	0.923	0.065	522.260	0.000
4	0.904	0.006	684.080	0.000
5	0.879	-0.086	837.900	0.000
6	0.855	-0.042	984.150	0.000
7	0.833	0.005	1123.700	0.000
8	0.809	-0.021	1256.100	0.000
9	0.781	-0.076	1380.200	0.000
10	0.758	0.024	1497.700	0.000
11	0.735	0.015	1608.900	0.000
12	0.716	0.053	1715.000	0.000
13	0.693	-0.033	1814.900	0.000
14	0.673	0.014	1909.700	0.000
15	0.655	0.025	2000.100	0.000
16	0.636	-0.014	2085.800	0.000
17	0.621	0.049	2168.000	0.000
18	0.604	-0.028	2246.100	0.000
19	0.586	-0.040	2319.900	0.000
20	0.568	-0.013	2389.800	0.000
21	0.549	-0.039	2455.300	0.000
22	0.527	-0.060	2516.200	0.000
23	0.507	0.000	2572.900	0.000
24	0.486	-0.031	2625.300	0.000
25	0.465	-0.015	2673.500	0.000
26	0.442	-0.021	2717.400	0.000
27	0.427	0.084	2758.500	0.000
28	0.405	-0.055	2795.800	0.000
29	0.383	-0.041	2829.200	0.000
30	0.365	0.042	2859.900	0.000
31	0.349	0.014	2888.100	0.000
32	0.334	0.031	2914.000	0.000
33	0.320	0.026	2938.000	0.000
34	0.301	-0.091	2959.400	0.000
35	0.283	-0.053	2978.500	0.000
36	0.274	0.117	2996.400	0.000

ANEXO F

Tabla 20: Valores del Correlograma de la serie a primera diferencia de la Demanda de Agua en el distrito de Lambayeque. Periodo 2002-2017.

	AC	PAC	Q-Stat	Prob
1	-0.343	-0.343	22.777	0.000
2	-0.030	-0.166	22.947	0.000
3	0.037	-0.037	23.213	0.000
4	0.054	0.059	23.778	0.000
5	0.015	0.073	23.823	0.000
6	-0.014	0.035	23.860	0.001
7	0.046	0.064	24.286	0.001
8	0.034	0.077	24.515	0.002
9	-0.132	-0.106	28.024	0.001
10	0.013	-0.093	28.060	0.002
11	-0.064	-0.153	28.906	0.002
12	0.082	-0.006	30.278	0.003
13	-0.031	0.009	30.480	0.004
14	-0.089	-0.072	32.114	0.004
15	0.071	0.031	33.162	0.004
16	-0.053	-0.015	33.756	0.006
17	0.033	0.042	33.984	0.008
18	0.016	0.048	34.037	0.012
19	-0.019	-0.002	34.116	0.018
20	0.038	0.012	34.432	0.023
21	0.034	0.073	34.688	0.031
22	-0.024	0.014	34.810	0.041
23	0.029	0.016	34.990	0.052
24	-0.023	-0.032	35.104	0.067
25	0.022	-0.033	35.216	0.084
26	-0.106	-0.127	37.729	0.064
27	0.091	-0.006	39.594	0.056
28	0.012	0.037	39.629	0.071
29	-0.096	-0.051	41.738	0.059
30	-0.010	-0.045	41.759	0.075
31	-0.033	-0.061	42.015	0.090
32	-0.009	-0.045	42.034	0.110
33	0.090	0.097	43.918	0.097
34	-0.002	0.118	43.919	0.119
35	-0.136	-0.119	48.285	0.067
36	0.021	-0.083	48.390	0.081

ANEXO G
PRODUCCIÓN DE AGUA

Prueba de Dickey – Fuller

1. Planteamiento de Hipótesis:

H_0 : La serie no es estacionaria, tiene raíz unitaria

H_1 : La serie es estacionaria

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

$$X_t = rX_{t-1} + e_t$$

4. Región Crítica:

El valor crítico es $t^* = F(t) = P(T \leq t) \rightarrow T = 180$ y $\alpha = 0.05 \rightarrow t^* = -2.885$, verificar Apéndice A.

5. Cálculo:

Tabla 2: Prueba de raíz unitaria de Dickey - Fuller para la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

Augmented Dickey-Fuller test	t-Statistic	Prob.*
statistic	-5.427251	0.000
1% level	-3.464827	
5% level	-2.876595	
10% level	-2.574874	

Fuente: Elaboración propia – Eviews

6. Decisión:

Si $|t^*| > |\text{valor crítico DF}| \longrightarrow$ Se rechaza la H_0 .

Si $|t^*| \leq |\text{valor crítico DF}| \longrightarrow$ No se rechaza la H_0 .

El valor $|-2.885|$ es mayor al valor crítico al 5% de $|-2.876595|$, por lo que la H_0 Se rechazó.

7. Conclusión:

Existe evidencia suficiente para afirmar que la serie es estacionaria, es decir no tiene raíz unitaria.

ANEXO H

Prueba t para los coeficientes del modelo SARMA (1,0,1) (1,0,0)

1. Planteamiento de Hipótesis:

H_0 : Los coeficientes son iguales a cero

H_1 : Los coeficientes son distintos a cero

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

Parte regular:

$$t_c = \frac{\theta_{t-1} - \theta_{t-1}}{ET} \sim t_{\frac{1-\alpha}{2n-2}}$$

Donde:

n: Número de retardos.

θ_{t-1} : Coeficientes de media móvil estimado de orden 1.

ET: Error de estimación del coeficiente θ_{t-1} .

Parte estacional:

$$t_c = \frac{\phi_{t-1} - \phi_{t-1}}{ET} \sim t_{\frac{1-\alpha}{2n-2}}$$

Donde:

n: Número de retardos.

ϕ_{t-1} : Coeficientes de media móvil estimado de orden 1.

ET: Error de estimación del coeficiente ϕ_{t-1} .

4. Región Crítica:

La prueba de t se rechaza si $|t_c| > |t_T| \rightarrow t_{\frac{1-0.05}{2(36)-2}} = 0.0135 = t_T$

5. Cálculo:

TABLA 7 Prueba t para una muestra para la producción de agua del distrito de Lambayeque, periodo enero 2002 a diciembre 2017.

	T	Gl	Sig. (bilateral)
Producción de Agua	231.578	191	0.000

Fuente: Elaboración propia – SPSS

6. Decisión:

Dado que $|t_c| < |t_T|$, $0.000 < 0.0135$; En ambos casos, tanto para el coeficiente de la parte regular, así como también para la parte estacional. Por lo tanto, no se rechaza la H_0 .

7. Conclusión:

Los coeficientes del modelo SARMA (1,0,2) (1,0,0) son significativamente iguales a

cero.

ANEXO I

Prueba de Normalidad de Kolmogorov Smirnov para el modelo SARMA (1,0,2)

(1,0,0)

1. Planteamiento de Hipótesis:

H_0 : Los residuos del modelo siguen aproximadamente una distribución normal.

H_1 : Los residuos del modelo no siguen aproximadamente una distribución normal.

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

$$D = \max |F_0(X) - S_n(X)|$$

4. Región Crítica:

Como $N > 35$, se usa la aproximación a muestras grandes. Por lo tanto, el valor crítico

$$\text{es } U = \frac{1.36}{\sqrt{N}} \Rightarrow U = \frac{1.36}{\sqrt{180}} = 0.101$$

Rechazar H_0 si $D > U$

5. Cálculo:

Tabla 5: Prueba de normalidad para los residuales del modelo SARMA (1, 0, 2) (1, 0, 0)

Producción de Agua	Kolmogorov-Smirnova		
	Estadístico	gl	Sig.
	0.034	192	0,200*

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Fuente: Elaboración propia

6. Decisión:

No se rechaza la H_0 dado que $D = 0.034 < U = 0.101$

7. Conclusión:

Los residuales del modelo SARMA (1,0,2) (1,0,0) siguen una distribución normal.

ANEXO J

Prueba de Aleatoriedad para el modelo para el modelo SARMA (1,0,2) (1,0,0)

1. Planteamiento de Hipótesis:

H_0 : Los residuos del modelo son aleatorios.

H_1 : Los residuos del modelo no son aleatorios.

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

$$|Z| = \frac{R - \frac{n}{2} - 1}{\sqrt{\frac{n^2 - 2n}{4(n-1)}}} \sim N(0,1)$$

4. Región Crítica:

Rechazar H_0 si $|Z| > Z_{\frac{\alpha}{2}} \rightarrow Z_{\frac{0.05}{2}} = 1.96$

5. Cálculo:

Tabla 6: Prueba de aleatoriedad para los residuales del modelo SARMA (1, 0, 2) (1, 0, 0)

Valor de prueba	-876
Casos < Valor de prueba	96
Casos >= Valor de prueba	96
Casos totales	192
Número de rachas	105
Z	1.158
Sig. asintótica (bilateral)	0.247

a. Mediana

Fuente: Elaboración propia

6. Decisión:

No se rechaza la H_0 dado que $|Z| = 1.158 < Z = 1.96$

7. Conclusión:

Se concluye que estadísticamente los residuos del modelo SARMA (1,0,2) (1,0,0) son aleatorios.

ANEXO K

DEMANDA DE AGUA

Prueba de Dickey – Fuller

1. Planteamiento de Hipótesis:

H_0 : La serie no es estacionaria, tiene raíz unitaria

H_1 : La serie es estacionaria

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

$$X_t = rX_{t-1} + e_t$$

4. Región Crítica:

El valor crítico es $t^* = F(t) = P(T \leq t) \rightarrow T = 180$ y $\alpha = 0.05 \rightarrow t^* = -1.95$, verificar Apéndice A.

5. Cálculo:

Tabla 10: Prueba de raíz unitaria de Dickey - Fuller para la demanda de agua del distrito de Lambayeque periodo enero 2002 a diciembre 2017.

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.976456	0.2971
1% level	-3.464827	
5% level	-2.876595	
10% level	-2.574874	

Fuente: Elaboración propia – Eviews

6. Decisión:

Si $|t^*| > |\text{valor crítico DF}| \longrightarrow$ Se rechaza la H_0 .

Si $|t^*| \leq |\text{valor crítico DF}| \longrightarrow$ No se rechaza la H_0 .

El valor $|-1.95|$ es mayor al valor crítico al 5% de $|-2.876595|$, por lo que la H_0 Se rechazó.

7. Conclusión:

Existe evidencia suficiente para afirmar que la serie es no estacionaria, es decir tiende a tener una raíz unitaria. Por otro lado, según p-valor es mayor al 5%.

ANEXO L

Prueba de Dickey – Fuller

1. Planteamiento de Hipótesis:

H_0 : La serie no es estacionaria, tiene raíz unitaria

H_1 : La serie es estacionaria

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

$$X_t = rX_{t-1} + e_t$$

4. Región Crítica:

El valor crítico es $t^* = -F(t) = P(T \leq t) \rightarrow T = 180$ y $\alpha = 0.05 \rightarrow t^* = -1.95$, verificar Apéndice A.

5. Cálculo:

Tabla 11: Prueba de raíz unitaria de Dickey - Fuller para la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.

Augmented Dickey-Fuller test	t-Statistic	Prob.*
statistic	-19.56337	0.0000
1% level	-2.57719	
5% level	-1.94250	
10% level	-1.615589	

Fuente: Elaboración propia – Eviews.

6. Decisión:

Si $|t^*| > |\text{valor crítico DF}| \longrightarrow$ Se rechaza la H_0 .

Si $|t^*| \leq |\text{valor crítico DF}| \longrightarrow$ No se rechaza la H_0 .

El valor $|-1.95|$ es mayor al valor crítico al 5% $|-1.942508|$, por lo que la H_0 Se rechazó.

7. Conclusión:

Existe evidencia suficiente para afirmar que la serie es estacionaria, es decir tiende a no tener raíz unitaria.

ANEXO LL

Prueba t para los coeficientes del modelo el modelo ARIMA (0,1,1)

1. Planteamiento de Hipótesis:

H_0 : Los coeficientes son iguales a cero

H_1 : Los coeficientes son distintos a cero

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

Parte regular:

$$t_c = \frac{\theta_{t-1} - \theta_{t-1}}{ET} \sim t_{\frac{1-\alpha}{2n-2}}$$

Donde:

n: Número de retardos.

θ_{t-1} : Coeficientes de media móvil estimado de orden 1.

ET: Error de estimación del coeficiente θ_{t-1} .

Parte estacional:

$$t_c = \frac{\phi_{t-1} - \phi_{t-1}}{ET} \sim t_{\frac{1-\alpha}{2n-2}}$$

Donde:

n: Número de retardos.

ϕ_{t-1} : Coeficientes de media móvil estimado de orden 1.

ET: Error de estimación del coeficiente ϕ_{t-1} .

4. Región Crítica:

La prueba de t se rechaza si $|t_c| > |t_T| \rightarrow t_{\frac{1-0.05}{2(36)-2}} = 0.0135 = t_T$

5. Cálculo:

Tabla 12: Prueba t para una muestra para la demanda de agua del distrito de Lambayeque a primera diferencia, periodo enero 2002 a diciembre 2017.

	t	gl	Sig. (bilateral)
Demanda de Agua	0.72	190.00	0.47

Fuente: Elaboración propia – SPSS

6. Decisión:

Dado que $|t_c| > |t_T|$, para el coeficiente de la parte regular. Por lo tanto, se rechaza la H_0 .

7. Conclusión:

Los coeficientes del modelo ARIMA (0,1,1) son significativamente diferentes a cero.

ANEXO M

Prueba de Normalidad de Kolmogorov Smirnov para el modelo ARIMA (0,1,1)

1. Planteamiento de Hipótesis:

H_0 : Los residuos del modelo siguen aproximadamente una distribución normal.

H_1 : Los residuos del modelo no siguen aproximadamente una distribución normal.

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

$$D = \max |F_0(X) - S_n(X)|$$

4. Región Crítica:

Como $N > 35$, se usa la aproximación a muestras grandes. Por lo tanto, el valor crítico

$$\text{es } U = \frac{1.36}{\sqrt{N}} \rightarrow U = \frac{1.36}{\sqrt{180}} = 0.101$$

Rechazar H_0 si $D > U$

5. Cálculo:

Tabla 14: Prueba de normalidad para los residuales del modelo ARIMA (0, 1, 1)

Demanda de Agua	Kolmogorov-Smirnova		
	Estadístico	Gl	Sig.
	0.05	191.00	0,200*

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Fuente: Elaboración propia – SPSS

6. Decisión:

No se rechaza la H_0 dado que $D = 0.05 < U = 0.101$

7. Conclusión:

Los errores del modelo ARIMA (0,1,1) siguen una distribución normal.

ANEXO N

Prueba de Aleatoriedad para el modelo para el modelo ARIMA (0,1,1)

1. Planteamiento de Hipótesis:

H_0 : Los residuos del modelo son aleatorios.

H_1 : Los residuos del modelo no son aleatorios.

2. Nivel de Significancia:

$$\alpha = 0.05$$

3. Estadístico de Prueba:

$$|Z| = \frac{R - \frac{n}{2} - 1}{\sqrt{\frac{n^2 - 2n}{4(n-1)}}} \sim N(0,1)$$

4. Región Crítica:

Rechazar H_0 si $|Z| > Z_{\frac{\alpha}{2}} \Rightarrow Z_{\frac{0.05}{2}} = 1.96$

5. Cálculo:

Tabla 15: Prueba de aleatoriedad para los residuales del modelo ARIMA (0,1,1)

Valor de prueba ^a	234
Casos < Valor de prueba	95
Casos >= Valor de prueba	96
Casos totales	191
Número de rachas	99
Z	0.36
Sig. asintótica (bilateral)	0.72

a. Mediana

Fuente: Elaboración propia

6. Decisión:

No se rechaza la H_0 dado que $|Z| = 0.36 < Z = 1.96$

7. Conclusión:

Se concluye que estadísticamente los residuos del modelo ARIMA (0,1,1) son aleatorios.

DIAGRAMA DE CAJAS Y BIGOTES / BOX-PLOT

➤ Producción de Agua

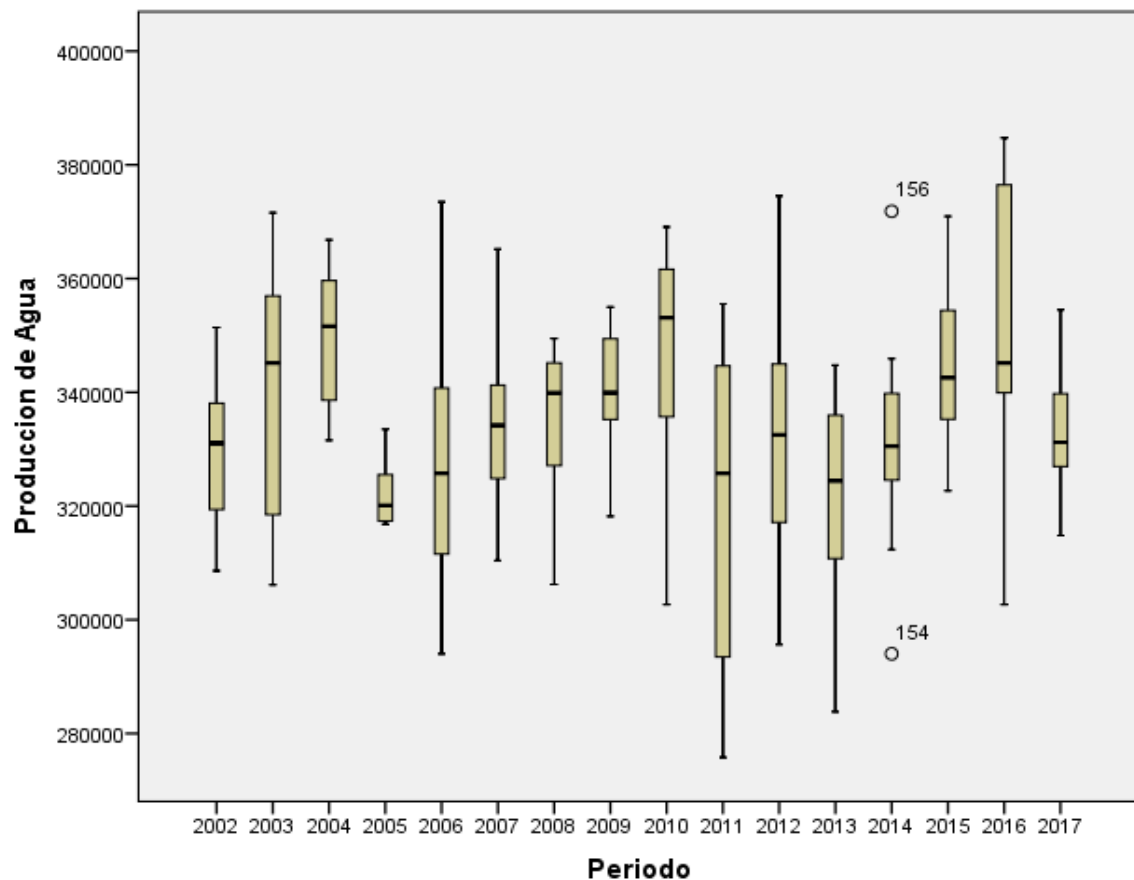


Figura 14: Diagrama de cajas y bigotes de la producción de agua en el distrito de Lambayeque. Periodo 2002 a 2017.

En la Figura 14, observamos que muestra dos datos atípicos en el año 2014. Además, en la gran mayoría de años vemos gran expansión en los cuartiles (Q1 y Q3) esto quiere decir que los datos muestran mayor dispersión de la producción de agua, las medianas por cada año indican asimetría tanto negativa como positiva en la producción de agua, esto también hace referencia al sesgo por cada año. Por otro lado, las distribuciones de los años expresan diferencia de variabilidad según la forma de las cajas.

➤ **Demanda de Agua**

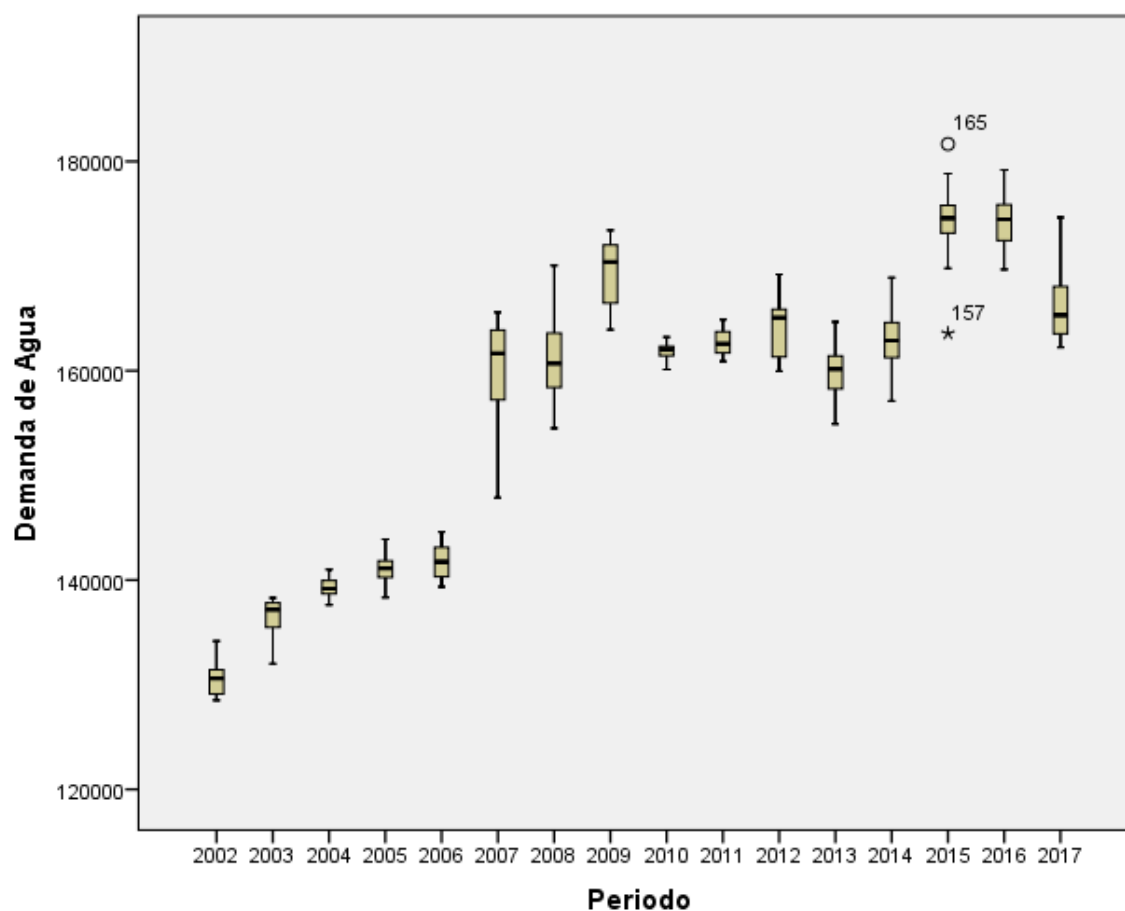


Figura 15: Diagrama de cajas y bigotes de la demanda de agua en el distrito de Lambayeque. Periodo 2002 a 2017.

En la Figura 15, observamos que muestra dos datos atípicos en el año 2015. Además, en la gran mayoría de años vemos cajas achatadas en los cuartiles (Q1 y Q3) esto quiere decir que los datos muestran mayor concentración de la demanda de agua, las medianas por cada año indican asimetría tanto negativa como positiva en la demanda de agua, esto también hace referencia al sesgo por cada año. Por otro lado, las distribuciones de los años expresan diferencia de variabilidad según la forma de las cajas.

EVOLUCIÓN DE LA POBLACIÓN DEL DISTRITO DE LAMBAYEQUE.

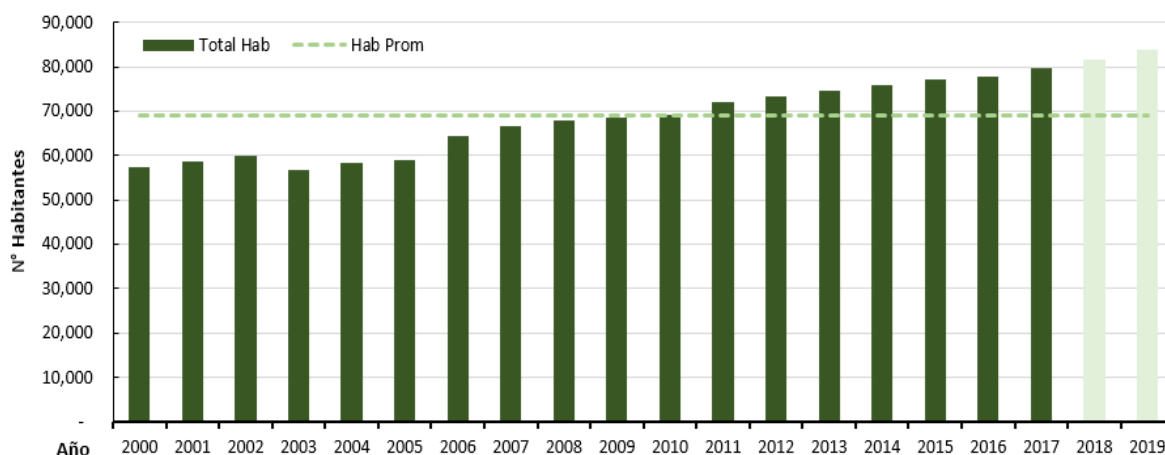


Figura 16: Distribución del número de habitantes del distrito de Lambayeque. Periodo 2000 a 2019.

Fuente: INEI / MINSA (Datos) – Elaboración Propia

En la Figura 16, observamos una tendencia en la evolución de la población del distrito Lambayeque. Además, se obtuvo una tasa promedio anual de 2.04% desde el 2000 al 2019. En el año 2000, se registró 57 mil habitantes y hasta el 2017 se observa un crecimiento ligeramente alto de 39% registrando 79 mil habitantes. Se estima un incremento de 5% desde el 2017 al 2019, registrando un valor de 83.7 mil habitantes.

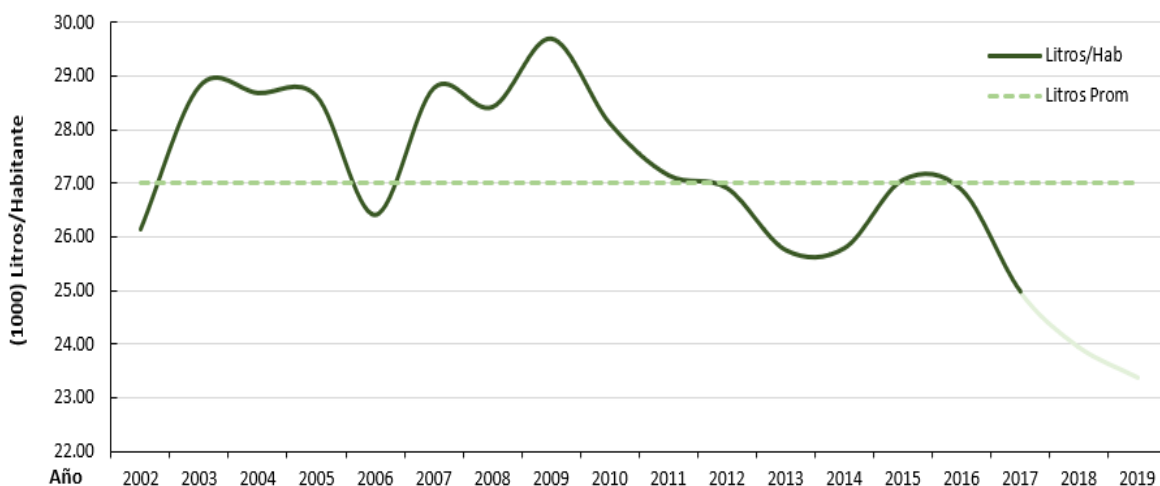


Figura 17: Consumo per cápita en litros de agua por habitante del Distrito de Lambayeque. Periodo 2002 a 2019.

En la Figura 17, observamos fluctuaciones sobre el consumo promedio per cápita en litros de agua por habitante del distrito Lambayeque. Además, se obtuvo una tasa promedio anual de 2.56% menos desde el 2002 al 2019. En el año 2000, se registró 26.1 mil litros por habitante y hasta el 2017 se observa un decaimiento ligeramente bajo de 4.43% menos registrando 24.9 mil litros de agua por habitante. Se estima una baja en el consumo de 6% menos desde el 2017 al 2019, registrando un valor de 23.3 mil litros de agua por habitante. Por lo cual, hay un déficit en el consumo de agua.

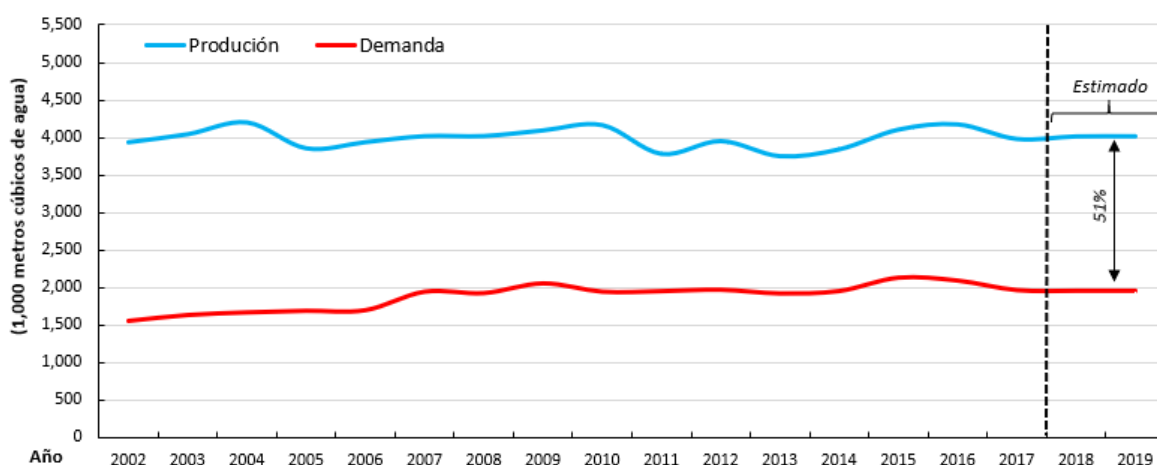


Figura 188: Producción y demanda de agua en (metros cúbicos) del Distrito de Lambayeque. Periodo 2002 a 2019.

En la Figura 18, observamos la comparación de ambas variables de la producción y demanda de agua del distrito de Lambayeque, periodo 2002 hasta el 2019. La producción a lo largo del tiempo ha tenido un comportamiento ligeramente constante, manteniéndose muy por debajo y por encima de los 4 millones de metros cúbicos. La demanda tiene un comportamiento de tendencia a lo largo del tiempo, superando los 2 millones de metro cúbicos de agua al año. Según lo estimado por los investigadores, la variación es del (-51%) en comparación a la producción.

Apéndice A

Tablas de valores críticos del estadístico de Dickey-Fuller

Modelo son constante

T	Nivel de significancia α								
	0.01	0.03	0.05	0.10	0.90	0.95	0.98	0.99	
25	- 2.66	- 2.26	- 1.95	- 1.60	0.92	1.33	1.70	2.16	
50	- 2.62	- 2.25	- 1.95	- 1.61	0.91	1.31	1.66	2.08	
100	- 2.60	- 2.24	- 1.95	- 1.61	0.90	1.29	1.64	2.03	
250	- 2.58	- 2.24	- 1.95	- 1.62	0.89	1.29	1.63	2.01	
500	- 2.58	- 2.23	- 1.95	- 1.62	0.89	1.28	1.62	2.00	
∞	- 2.58	- 2.23	- 1.95	- 1.62	0.89	1.28	1.62	2.00	

Fuente: González, M. (2009).

T	Nivel de significancia α								
	0.01	0.03	0.05	0.10	0.90	0.95	0.98	0.99	
25	- 3.75	- 3.33	- 3.00	- 2.63	0.37	0.00	0.34	0.72	
50	- 3.58	- 3.22	- 2.93	- 2.60	0.40	0.03	0.29	0.66	
100	- 3.51	- 3.17	- 2.89	- 2.58	0.42	0.05	0.26	0.63	
250	- 3.46	- 3.14	- 2.88	- 2.57	0.42	0.06	0.24	0.62	
500	- 3.44	- 3.13	- 2.87	- 2.57	0.43	0.07	0.24	0.61	
∞	- 3.43	- 3.12	- 2.86	- 2.57	0.44	0.07	0.23	0.60	

Fuente: González, M. (2009).