

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN



**“ANÁLISIS Y PRONÓSTICO DE LA RECAUDACIÓN DEL
IMPUESTO VEHICULAR MUNICIPAL MEDIANTE LA
METODOLOGÍA BOX -JENKINS”**

Presentado por:

CINTHIA KATHERYN HUAMÁN INGA

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR EL
TÍTULO DE INGENIERO ESTADÍSTICO INFORMÁTICO**

LIMA-PERÚ

2020

**UNIVERSIDAD NACIONAL AGRARIA LA
MOLINA**

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

**“ANÁLISIS Y PRONÓSTICO DE LA RECAUDACIÓN DEL
IMPUESTO VEHICULAR MUNICIPAL MEDIANTE LA
METODOLOGÍA BOX -JENKINS”**

**PRESENTADO POR
CÍNTHIA KATHERYN HUAMÁN INGA**

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR EL
TÍTULO DE INGENIERO ESTADÍSTICO INFORMÁTICO**

SUSTENTADA Y APROBADA ANTE EL SIGUIENTE JURADO

.....
Dr. Rino Nicanor Sotomayor Ruiz
Presidente

.....
Dr. Jorge Chue Gallardo
Asesor

.....
Dr. Cesar Higinio Menacho Chiok
Miembro

.....
Mg. Jesús Walter Salinas Flores
Miembro

Lima – Perú

2020

DEDICATORIA

A mi madre Elizabeth Inga por su paciencia, comprensión y brindarme lo
mejor cada día y ser mi modelo a seguir.

A mis abuelos Leonor y Gaudencio por su cariño incondicional.

A mi tío Ángel Inga, a pesar de que ya no está en nosotros siempre lo llevo presente
en mi corazón.

AGRADECIMIENTOS

A Dios por su amor infinito quien me guía en todo momento y me da fuerzas para seguir esforzándome y sabiduría para tomar las mejores decisiones.

A mi casa de estudios la Universidad Nacional Agraria la Molina por la formación académica y profesional.

A mi asesor el Prof. Jorge Chue Gallardo por su apoyo constante y orientación en el desarrollo del presente trabajo.

A mis amistades que me brindaron su apoyo y motivación, gracias por sus consejos y amistad.

ÍNDICE DE CONTENIDO

I.	PRESENTACIÓN.....	1
II.	INTRODUCCIÓN	2
III.	OBJETIVOS.....	4
3.1	OBJETIVO GENERAL	4
3.2	OBJETIVOS ESPECÍFICOS.....	4
IV.	CUERPO DEL TRABAJO.....	4
4.1	FUNCIONES DESEMPEÑADAS	4
4.2	PUESTA EN PRÁCTICA DE LO APRENDIDO EN LA CARRERA.....	5
4.2.1.	DESCRIPCIÓN DE LA TÉCNICA ESTADÍSTICA UTILIZADA.....	5
4.2.2.	REVISIÓN DE LA LITERATURA.....	16
4.2.3.	PROPUESTA DE ALTERNATIVA DE SOLUCIÓN	18
4.3	CONTRIBUCIÓN EN LA SOLUCIÓN DE SITUACIONES PROBLEMÁTICAS	38
4.4	ANÁLISIS DE LA CONTRIBUCIÓN EN TÉRMINOS DE COMPETENCIA Y HABILIDADES.....	38
4.5	NIVEL DE BENEFICIO OBTENIDO POR EL CENTRO LABORAL	38
V.	CONCLUSIONES Y RECOMENDACIONES.....	39
5.1	CONCLUSIONES	39
5.2	RECOMENDACIONES.....	39
VI.	REFERENCIAS BIBLIOGRÁFICAS	41
VII.	ANEXOS	42

ÍNDICE DE TABLAS

Tabla 1: Criterios para determinar el modelo idóneo	13
Tabla 2: Porcentaje de representación de los principales ingresos de la entidad	19
Tabla 3: Principales estadísticas de la data.....	21
Tabla 4: Comparación del coeficiente de determinación de los modelos lineal y cuadrático.	22
Tabla 5: Prueba de normalidad de los datos	24
Tabla 6: Prueba de Kruskal-Wallis para analizar la tendencia	25
Tabla 7: Prueba de Homogeneidad de Varianzas a nivel anual.....	26
Tabla 8: Prueba de Homogeneidad de Varianzas a nivel mensual	27
Tabla 9 Test de Dickey – Fuller	31
Tabla 10: test de KPSS (Kwiatkowski-Phillips-Schmidt-Shin)	31
Tabla 11: Coeficientes de la Función de Correlación Simple y Parcial (ACF) y (PACF) ..	32
Tabla 12: Modelos ARIMA tentativos	33
Tabla 13: Criterios de Selección para cada modelo tentativo.....	34
Tabla 14: Prueba de Box-Price	35
Tabla 15: Prueba de Ljung Box	36
Tabla 16: Prueba de Jarque Bera Fuente: Elaboración Propia	36
Tabla 17: Prueba de Shapiro Wilk Fuente: Elaboración Propia	36
Tabla 18: Pronostico del impuesto vehicular para el periodo Nov. 2019 a Dic 2020	37
Tabla 19: Datos de registro de los ingresos del Impuesto Vehicular.....	42

ÍNDICE DE FIGURAS

Figura 1: Procedencia de una serie de tiempo observada	7
Figura 2: Metodología de Box y Jenkins	12
Figura 3: Serie de tiempo del impuesto vehicular entre enero del 2009 a octubre del 2019.	22
Figura 4: Gráficas de residuos para el modelo lineal.	23
Figura 5: Gráficas de residuos para el modelo cuadrático.....	23
Figura 6: Gráficas de cajas anual.....	24
Figura 7: Gráficas de cajas mensual.	25
Figura 8: Prueba de Homogeneidad de Varianzas a nivel anual	27
Figura 9: Prueba de Homogeneidad de Varianzas a nivel mensual.....	28
Figura 10: Gráfica de la serie temporal estabilizando la varianza.....	29
Figura 11: Gráfica de la serie temporal eliminando la tendencia	30
Figura 12: Gráfica de la serie temporal eliminando la estacionalidad.....	30
Figura 13: Función de Correlación Simple y Parcial (ACF) y (PACF)	33
Figura 14: Función de autocorrelación simple y parcial de los residuales	34
Figura 15: Gráficos de residuos del modelo elegido	35
Figura 16: Pronostico de la serie de tiempo impuesto vehicular	37

I. PRESENTACIÓN

Al término de la carrera universitaria, la autora del presente trabajo de suficiencia profesional prestó sus servicios profesionales en el área de consultoría de una empresa de investigación del rubro inmobiliario. En dicha entidad, las principales funciones fueron: realizar el proceso de validación y supervisión de la data correspondiente a las principales variables de la oferta inmobiliaria a nivel distrital de la Provincia de Lima, elaboración y desarrollo de los reportes trimestrales con información actualizada.

Posteriormente y hasta la actualidad, la labor profesional es en el área presupuestaria de la Gerencia de Finanzas de una entidad del sector público. Entre las principales funciones que desarrolla en dicha entidad son: realizar el seguimiento continuo de la ejecución presupuestaria de la entidad en función a la estructura organizativa y las diversas áreas, así como de las empresas municipales y organismos públicos descentralizados que se encuentran a su cargo; realizar proyecciones y estimaciones de los principales conceptos de ingresos como: Impuesto Predial, Impuesto Vehicular, Impuesto de Alcabala, Impuestos a los Juegos y Espectáculos y los arbitrios de Limpieza Pública, Serenazgo, Parques y Jardines; así mismo, elaborar y presentar informes gerenciales y reportes mensuales de los principales indicadores presupuestarios de ingresos y gastos a la alta dirección los cuales permiten medir el nivel de avance de las diversas áreas que comprende la entidad y tomar medidas para cumplir con los objetivos y planes en el marco del Plan Estratégico Institucional de la entidad.

Para poder desarrollar las labores dentro de la entidad, se hace uso de las principales técnicas estadísticas e informáticas aprendidas durante la carrera universitaria tales como implementación de técnicas de clasificación de los diversos centros de costos en relación a los servicios que presta la entidad; estimaciones utilizando métodos de regresión, mediante el uso de modelos lineales y series de tiempo para realizar pronósticos de los ingresos presupuestarios de la entidad.

II. INTRODUCCIÓN

En el ámbito del proceso presupuestario de las entidades públicas, se considera al presupuesto público como un instrumento de gestión del Estado para el logro de resultados a favor de la población mediante la prestación de servicios y cumplimiento de metas estableciendo límites de gastos durante el periodo fiscal y los ingresos que los financian de acuerdo a los recursos disponibles a fin de mantener el equilibrio fiscal; así mismo, se enmarcan de acuerdo las disposiciones que establece el ente rector que es el Ministerio de Economía y Finanzas a través de la Dirección General de Presupuesto Público y el Sistema Nacional de Presupuesto Público.

En la fase de programación del presupuesto público, la elaboración del presupuesto de ingresos y gastos se enmarca en una serie de actividades y metodologías que deben estar alineadas al logro de los objetivos estratégicos e institucionales; considerando el principio de equilibrio fiscal. La metodología a emplear para la estimación de los ingresos que las entidades de gobierno local esperan recaudar u obtener para el siguiente año fiscal se realiza sin considerar las transferencias establecidas por el Ministerio de Economía y Finanzas. En tal sentido, a las entidades les corresponde estimar los ingresos propios que obtendrán por los Impuestos Municipales, Recursos Directamente Recaudados, Recursos por Operaciones Oficiales de Crédito Interno y Externo.

Actualmente, uno de los principales problemas observados en relación a la programación de los ingresos presupuestarios es la falta de técnicas estadísticas que estimen con mayor precisión la recaudación de los principales impuestos y arbitrios municipales, los cuales se venían elaborando solamente a través de promedios anuales que finalmente, al realizar comparaciones con la ejecución real obtenida, presentan notorias diferencias.

Así mismo, dado que se dispone de información histórica anual y mensual del registro de los ingresos que percibe la entidad, de acuerdo al clasificador de ingresos establecido por el Ministerio de Economía y Finanzas; se consideró conveniente implementar el pronóstico a través de las series de tiempo y su proyección mediante un modelo matemático que se ajuste al comportamiento de los datos de ingresos de la entidad.

Para el análisis y pronóstico, en el presente trabajo de suficiencia profesional se considera la metodología propuesta por Box y Jenkins el cual desarrolla una secuencia de cuatro (04) fases que comprenden desde la identificación del mejor modelo, la estimación de los parámetros, diagnóstico del modelo propuesto y finalmente la predicción para tiempos futuros. Así mismo, se emplea el modelo paramétrico ARIMA dado que su desarrollo flexible para series temporales es aplicado tanto para estructuras estacionarias y no estacionarias.

Considerando las características de los registros de ingresos presupuestarios, se consideró factible el uso del análisis y pronóstico a través de las series de tiempo; del mismo modo, tomando en consideración que los conceptos de ingresos presupuestarios que registra la entidad presentan características diferentes, se decidió desarrollar el análisis de la recaudación del Impuesto Vehicular en forma específica debido a que es un impuesto que tiene una tendencia creciente y a la gran demanda por la compra de vehículos en la Provincia de Lima.

Por lo tanto, el principal objetivo del presente trabajo de suficiencia profesional es: determinar el mejor modelo que se ajuste a las características de la recaudación del Impuesto Vehicular de la Provincia de Lima mediante la técnica estadística de series de tiempo, con la finalidad de pronosticar el registro de ingresos de dicho concepto en tiempos futuros.

III. OBJETIVOS

3.1 OBJETIVO GENERAL

Determinar y estimar los pronósticos del registro de ingresos del Impuesto Vehicular mediante el uso de la metodología de Box - Jenkins, para un periodo de catorce (14) meses en el horizonte temporal noviembre 2019 – diciembre 2020.

3.2 OBJETIVOS ESPECÍFICOS

- Describir las principales características y componentes del registro de ingresos mensuales del Impuesto Vehicular, durante el periodo enero 2009 – octubre 2019.
- Establecer el mejor modelo para el pronóstico del registro de ingresos mensuales del Impuesto Vehicular con los datos del periodo enero 2009 – octubre 2019.

IV. CUERPO DEL TRABAJO

4.1 FUNCIONES DESEMPEÑADAS

Las funciones que actualmente son desempeñadas son las siguientes:

- a) Organizar y explorar las bases de datos correspondientes al área presupuestaria.
- b) Efectuar el procesamiento automático y análisis de datos correspondientes a las informaciones presupuestarias presentadas por las municipalidades distritales que integran la provincia de lima y de las empresas municipales y Organismos Públicos Descentralizados de la entidad.

- c) Elaborar cuadros y proyecciones estadísticos correspondientes al Presupuesto Público.
- d) Controlar la Programación, Ejecución y Evaluación Presupuestaria de la entidad.
- e) Preparar los reportes de la información presupuestaria, correspondiente al consolidado de los presupuestos y evaluaciones presentadas por las municipalidades distritales en ejecución a las directivas que emita la dirección general de presupuesto público-Ministerio de Economía y Finanzas, y los reportes de actualización de sus modificaciones y ejecuciones presupuestarias.
- f) Generar estadísticas presupuestales a nivel del gobierno local y Entidades de Tratamiento Empresarial.
- g) Crear y analizar los indicadores presupuestales e institucionales.
- h) Formular, analizar e interpretar cuadros estadísticos para su diagnóstico y/o estudio.
- i) Documentar los procesos estadísticos a su cargo, así como el análisis efectuado.

4.2 PUESTA EN PRÁCTICA DE LO APRENDIDO EN LA CARRERA

4.2.1. DESCRIPCIÓN DE LA TÉCNICA ESTADÍSTICA UTILIZADA

SERIES DE TIEMPO

Las **series de tiempo** son conjuntos de datos cuantitativos obtenidos del resultado de observar o medir algún fenómeno determinado a lo largo de un periodo de tiempo determinado. Dichos datos pueden presentar comportamientos diferentes a través del

tiempo tales como tendencia, ciclos, variaciones estacionales o no tener un comportamiento definido. Los modelos de series de tiempo son técnicas que tienen un enfoque netamente predictivo y se basan en el análisis del registro histórico de una o más variables determinadas para realizar pronósticos futuros.

En relación a la representación matemática de las series de tiempo univariantes, podemos definirla como un proceso estocástico el cual se presenta por una secuencia de N observaciones ordenadas de manera cronológica y equidistantes sobre una característica de una unidad observable en diferentes momentos.

$$\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\} = (\mathbf{y}_t)_{t=1}^N \quad ; \quad (\mathbf{y}_t: t = 1, 2, \dots, N)$$

Donde \mathbf{y}_t es la observación en el tiempo t ($1 \leq t \leq N$) de la serie y N es el número de observaciones que consta la serie completa. Las N observaciones pueden representarse en un vector de columna $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]'$ de orden $N \times 1$.

(Reyes Flores, 2016) señala que al igual que una muestra aleatoria, el análisis de las series de tiempo puede llevarse a cabo con herramientas estadísticas descriptivas e inferenciales. Así mismo, considera que el análisis descriptivo es fundamental en una serie de tiempo, ya que puede darnos una idea del comportamiento de la serie en el periodo de análisis; y considerando la metodología clásica el análisis inferencial de las series de tiempo es la descomposición para identificar los tres(03) componentes: *tendencia*, que presenta el comportamiento de la serie a lo largo del tiempo, *estacionalidad*, que representa los comportamientos que se repiten en determinados periodos del tiempo, y la *irregularidad*, que representa los movimientos aleatorios.

El análisis de las series de tiempo se refiere a las técnicas para el análisis de esta dependencia. Esto requiere el desarrollo de modelos estocásticos y dinámicos para los datos de series temporales y el uso de tales modelos en áreas importantes de aplicación. (Box, Jenkins, Reinsel, & Ljung, 2016)

El objetivo del análisis de las series de tiempo consiste en desarrollar un modelo estadístico que describa adecuadamente el comportamiento de los datos a lo largo del

tiempo. Posteriormente el modelo estadístico idóneo tendrá la capacidad de describir la evolución de la serie considerando la relación en el tiempo y sus componentes; prever la evolución en momentos futuros y presentar evidencia empírica respecto a las características a las que se refiere los componentes de la serie temporal.

PROCESOS ESTOCÁSTICOS

Un proceso estocástico es una secuencia de variables aleatoria, ordenadas y equidistantes cronológicamente, sobre una o varias características de una unidad observable en diferentes momentos. (Box, Jenkins, Reinsel, & Ljung, 2016)

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

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

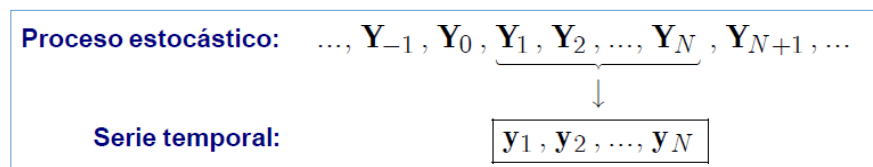


Figura 1: Procedencia de una serie de tiempo observada.

Fuente:(Mauricio, 2007)

Un proceso estocástico es considerado estacionario cuando las propiedades de cualquier secuencia finita de los componentes de dicha serie son semejantes a las de la secuencia para cualquier otro periodo de dicha serie.

Media: $E(y_t) = \mu$, constante para todo t

Varianza: $Var(y_t) = E(y_t - \mu)^2 = \sigma^2 = \gamma_0$, constante para todo t

Covariancia: $Cov(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k$, siendo $t \neq t + k$

COMPONENTES DE UNA SERIE DE TIEMPO

La metodología del análisis de las series de tiempo establece que para poder determinar los componentes de una serie de tiempo se debe realizar la descomposición en sus fuentes de variaciones. (Box, Jenkins, Reinsel, & Ljung, 2016)

- **TENDENCIA.** - Se considera que la serie no presenta estacionalidad, la estructura de la tendencia puede ser expresado de la siguiente manera: (Box, Jenkins, Reinsel, & Ljung, 2016)

$$Y_t = m_t + \varepsilon_t, \text{ para } t=1,2,\dots,n \text{ donde } E[\varepsilon_t]=0$$

- **ESTACIONALIDAD.** -Se considera que la serie se encuentra libre de tendencia, la estructura de la serie quedará definida por el componente de variación estacional expresado de la siguiente manera: (Box, Jenkins, Reinsel, & Ljung, 2016)

$$Y_t = s_t + \varepsilon_t, \text{ para } t=1,2,\dots,n$$

- **CICLO.** - El componente de ciclo considera las oscilaciones a lo largo del tiempo en relación a la tendencia de la serie. (Box, Jenkins, Reinsel, & Ljung, 2016)
- **IRREGULARIDAD.** – Se considera cuando a lo largo del tiempo la serie no presenta un comportamiento definido y no sigue un patrón específico. (Box, Jenkins, Reinsel, & Ljung, 2016)

MODELOS PARA UN PROCESO ESTOCÁSTICO

Para la descripción de los diversos modelos para el análisis de las series de tiempo univariados, se define una serie temporal como el conjunto de observaciones repetidas de la misma variable $\{y_1, y_2, \dots, y_N\}$ donde los sufijos representan el periodo en el tiempo en el que se observa la variable y_t .

En la actualidad existe una gran variedad de modelos para el análisis de las series de tiempo. En esta oportunidad, se delimita la descripción del uso de modelos para datos univariados. A continuación, se detalla los modelos fundamentales tales como: ruido blanco, paseo aleatorio, modelo autorregresivo (AR), modelo de media móvil (MA), modelo autorregresivo de media móvil (ARMA) y modelo autorregresivo integrado de media móvil (ARIMA).

- **RUIDO BLANCO – ARIMA (0,0,0)**

Un proceso de ruido blanco es una secuencia de variables aleatorias escalares e independientemente distribuidas con media 0 y varianza σ_A^2 , el cual suele representarse como $(A_t) \sim IID(0, \sigma_A^2)$, .Cuando cada A_t sigue una distribución Normal, (A_t) se denomina u proceso de ruido blanco Normal o Gaussiano, el cual suele representarse como $(A_t) \sim NIID(0, \sigma_A^2)$. (Mauricio, 2007)

$$E(A_t) = 0 \quad , \quad V(A_t) = \sigma_A^2$$

- **PASEO ALEATORIO – ARIMA (0,1,0)**

Un proceso estocástico univariante no estacionario (Y_t) es considerado como un paseo aleatorio cuando $y_t = \mu + y_{t-1} + A_t$ para todo $t = 0, \pm 1, \pm 2, . . .$, donde μ es un parámetro que muchas veces es cero y $(\varepsilon_t) \sim IID(0, \sigma_\varepsilon^2)$. (Mauricio, 2007)

- **MODELO AUTORREGRESIVO DE ORDEN p: AR(p) - ARIMA(p,0,0)**

Un proceso estocástico univariante estacionario (y_t) sigue un modelo AR(1), modelo autorregresivo de orden 1 cuando: $y_t = \mu + \phi_1 y_{t-1} + s_t$, para todo $t = 0, \pm 1, \pm 2, . . .$, donde μ y ϕ_1 son parámetros, $|\phi_1| < 1$ (condición de estacionalidad) y ε_t es ruido blanco. (Mauricio, 2007)

Generalizando a un modelo autorregresivo de orden p, AR(p) se define como:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + s_t$$

- **MODELO DE PROMEDIOS MOVILES DE ORDEN q: MA(q) - ARIMA(0,0,q)**

Un proceso estocástico univariante estacionario (y_t) sigue un modelo MA(1), modelo de media móvil de orden 1 cuando: $y_t = \mu + s_t - \theta_1 s_{t-1}$, para todo $t = 0, \pm 1, \pm 2, \dots$, donde μ y θ_1 son parámetros, $|\theta_1| < 1$ (condición de invertibilidad) y ε_t es ruido blanco. (Mauricio, 2007)

Generalizando a un modelo de media móvil de orden q, AR(q) se define como:

$$y_t = \mu - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_q y_{t-q} + s_t$$

- **MODELO AUTORREGRESIVO DE PROMEDIO MÓVIL DE ORDEN (p,q): ARMA(p,q) – ARIMA(p,0,q)**

Un proceso estocástico univariante estacionario (y_t) sigue un modelo ARMA(1,1), modelo autorregresivo de media móvil de orden (1,1) cuando: $y_t = \mu + \phi_1 y_{t-1} + s_t + \theta_1 s_{t-1}$, para todo $t = 0, \pm 1, \pm 2, \dots$, donde μ , θ_1 y ϕ_1 son parámetros, $|\phi_1| < 1$ (condición de estacionariedad) y $|\theta_1| < 1$ (condición de invertibilidad) y $(\varepsilon_t) \sim IID(0, \sigma_\varepsilon^2)$. (Mauricio, 2007)

Generalizando a un modelo autorregresivo de orden (p,q), ARMA(p,q) se define como:

$$y_t = a + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} - \theta_1 y_{t-1} - \theta_2 y_{t-2} - \dots - \theta_q y_{t-q} + s_t$$

- **MODELO AUTORREGRESIVO E INTEGRADO DE PROMEDIO MÓVIL: ARIMA(p,d,q)**

Un proceso estocástico **no** estacionario debe ser diferenciando para así poder trabajar el modelo; en caso se realice dicha diferenciación se considera el proceso de modelo autorregresivo e integrado de promedio móvil ARIMA(p,d,q), donde p denota el orden del rezago en Y_t , q el orden de rezago del promedio móvil asociado a A_t y d es el orden de diferenciación de la serie.

$$\Delta^d y_t = a + \phi_1 \Delta^d Y_{t-1} + \dots + \phi_p \Delta^d Y_{t-p} - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_q Y_{t-q} + s_t$$

METODOLOGÍA DE BOX Y JENKINS

La metodología de Box – Jenkins fue propuesta por los investigadores George Edward Pelham Box¹ y Gwilym Meirion Jenkins² en el año de 1970 el cual consta de una estructura de cuatro (04) fases que permitirán analizar y modelar un proceso estacionario y determinar el mejor modelo de series de tiempo.

El objetivo de la metodología de Box – Jenkins es identificar y estimar un modelo estadístico que pueda ser interpretado como un generador de la información de los datos para la predicción, en donde se supone que las características de la serie temporal serán constantes en el tiempo.

i. IDENTIFICACIÓN DEL MODELO

La fase de identificación consiste en detectar el tipo de proceso estocástico que implique determinar características teorías de un modelo en relación a las propiedades de la serie temporal en análisis. Así mismo, para identificar los componentes p, q y d se puede hacer uso de los correlogramas de las funciones de autocorrelación simple (ACF) ρ_k y funciones de autocorrelación parcial (PACF) ρ_{kk} de las series.

¹ Nació el año de 1919 en Inglaterra.

² Nació el año de 1933 en Inglaterra.

Es importante considerar que antes de utilizar las funciones de autocorrelación simple y parcial, la serie temporal en análisis debe ser estacionaria, de no cumplirse con dicha condición, se debe aplicar los procedimientos necesarios para lograr la estacionariedad utilizando la diferenciación d veces sea necesario.

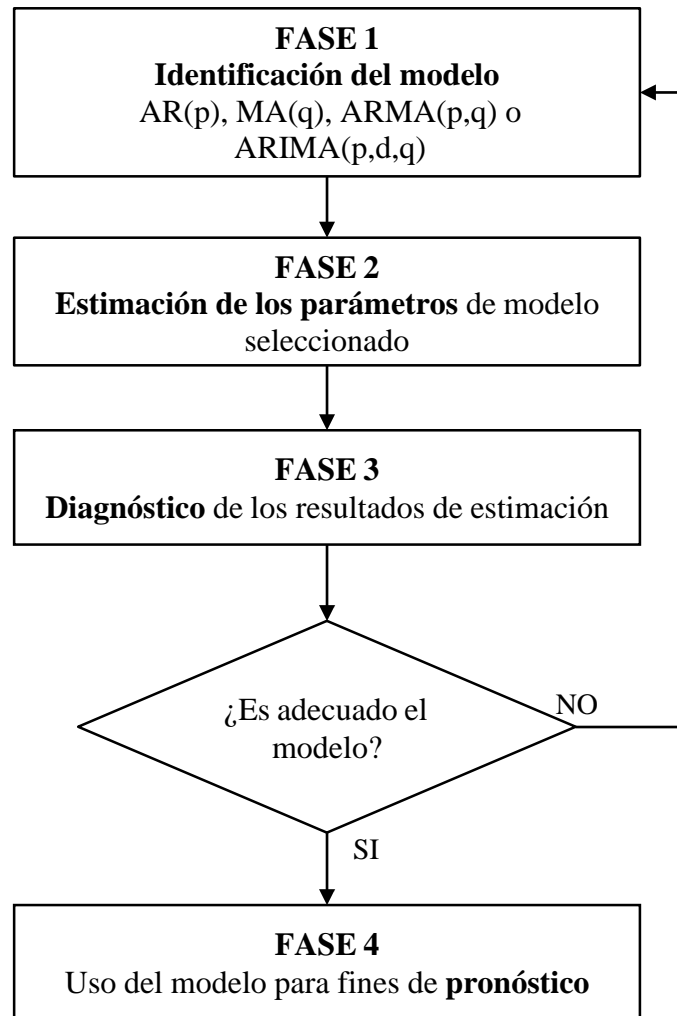


Figura 2: Metodología de Box y Jenkins.

Fuente: (Castillo P. & Varela L, 2008)

FUNCION DE AUTOCORRELACIÓN SIMPLE (ACF)

Se denomina función de autocorrelación a un proceso estocástico que describe las correlaciones en dos instantes diferentes de un proceso. (Castillo P. & Varela L, 2008)

$$\rho_k = \text{corr}(X_t, X_{t-k})$$

FUNCION DE AUTOCORRELACIÓN PARCIAL (PACF)

La función de autocorrelación parcial es un proceso estocástico que describe las correlaciones entre variables separadas por k retardos, eliminando el efecto de las $k-1$ variables intermedias. (Castillo P. & Varela L, 2008).

Mediante el análisis gráfico de las funciones de autocorrelación simple y autocorrelación parcial se puede determinar el nivel de orden adecuado para los modelos: de promedios móviles $MA(q)$, autorregresivo $AR(p)$ y autorregresivo en integrado de promedio móvil $ARMA(p,q)$. En la tabla 1 se detallan algunas características que ayudaran a determinar el mejor modelo para la serie de tiempo.

Tabla 1: Criterios para determinar el modelo idóneo

	FUNCION DE AUTOCORRELACIÓN SIMPLE (ACF)	FUNCIÓN DE AUTOCORRELACIÓN PARCIAL (PACF)
AR (p)	<ul style="list-style-type: none">- Disminución exponencial de forma regular o sinusoidal.- Puede presentar coeficientes positivos, negativos u una combinación de ambos.	<ul style="list-style-type: none">- Coeficientes significativos a lo largo de p retardos.
MA(q)	<ul style="list-style-type: none">- Coeficientes significativos a lo largo de q retardos.	<ul style="list-style-type: none">- Decrecimiento exponencial- Puede presentar coeficientes positivos, negativos u una combinación de ambos.
ARMA(p,q)	<ul style="list-style-type: none">- Decrecimiento exponencial	<ul style="list-style-type: none">- Decrecimiento exponencial

ii. ESTIMACIÓN DE PARÁMETROS

Posteriormente de haber determinado el modelo más idóneo que se ajuste a los datos de la serie temporal, se procede con la fase de estimación de parámetros del modelo con valores numéricos.

En esta fase se debe verificar la significancia estadística de los coeficientes estimados mediante el uso de la distribución normal; análisis de varianzas; en el caso

de que se cuente con más de un modelo tentativo se debe realizar comparaciones de los criterios de significancia.

iii. DIAGNÓSTICO DEL MODELO

En la fase de diagnóstico del modelo se utilizan técnicas que comprueben el ajuste del modelo a los datos de la serie temporal en análisis. El diagnóstico implica evaluar la confiabilidad de los resultados del modelo; es factible hacer uso de los correlogramas de los coeficientes de autocorrelación simple y autocorrelación parcial de los residuos del modelo identificado.

En dicho análisis, para determinar la confiabilidad del modelo, los coeficientes de autocorrelación simple y parcial deben ser iguales a cero; es decir, que cada residuo debe encontrarse dentro del intervalo de confianza de los correlogramas. En tal sentido se plantea que la hipótesis nula los residuos son ruido blanco; como resultado, no se rechaza dicha hipótesis si los estadísticos obtenidos son menores al valor crítico correspondiente a una X^2 , en caso la hipótesis sea rechazada se debe proceder a corregir el modelo o bien retornar a la fase 1 de la metodología.

BONDAD DE AJUSTE

En el caso de que, en la fase de identificación se plantee más de un modelo tentativo, se debe medir e identificar cuál de ellos presenta una mejor **bondad de ajuste**. Entre las principales pruebas para determinar el mejor modelo tenemos al análisis de R^2 *ajustado*, Criterio de Información de Akaike (AIC) o el Criterio de Schwartz (SC).

- **R^2 AJUSTADO**

El R^2 *ajustado* a pesar de presentar mejor resultados que el R^2 debido a que ayuda a evitar la sobreparametrización; solo tiene validez cuando se compara modelos que

consideran el mismo número de diferencias dado que la varianza total depende del número de diferencias que se haya considerado.

- **CRITERIO DE INFORMACIÓN DE AKAIKE (AIC) O EL CRITERIO DE SCHWARTZ (SC)**

Los criterios de información de Akaike y de Schwartz AIC y SC son técnicas estadísticas que utilizan la suma de cuadrados de los errores con el fin de minimizarla de acuerdo a las diversas combinaciones de p y q para poder determinar el número adecuado de rezagos p y q para el mejor modelo ARIMA.

Las fórmulas de cálculo de los criterios se presentan a continuación:

$$AIC = \ln\left(\frac{\hat{\varepsilon}\varepsilon}{n}\right) + \frac{2k}{n}$$

$$SC = \ln\left(\frac{\hat{\varepsilon}\varepsilon}{n}\right) + \frac{k \ln n}{n}$$

Donde n corresponde al número de observaciones de la serie temporal y k es el número agregado de parámetros incorporados a la estimación : $k = p + q + 1$.

Para determinar el mejor modelo se escoge a aquellos que presentan menores valores posibles de AIC y SC.

ANÁLISIS DE LOS RESIDUOS

- **CORRELOGRAMA DE LOS RESIDUOS Y EL ESTADÍSTICO LJUNG-BOX.**

El análisis del correlograma de los residuos del modelo debe determinar que dichos residuos son ruido blanco.

El estadístico de Ljung-Box (LB) es utilizado para realizar la prueba estadística en el cual no se rechaza la hipótesis nula si LB lo menor posible, entonces se afirma que los errores del modelo siguen un proceso de ruido blanco. (Reyes Flores, 2016)

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m$$

$$H_1: \text{Al menos un } \rho_k \neq 0 \quad \forall k = 1, 2, \dots, m$$

$$LB = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} + X^2_{1-\alpha, m-p-q \text{ g.l.}}$$

Para determinar si los errores del modelo siguen una distribución Normal se realiza la prueba utilizando el estadístico Jarque Bera, así mismo se puede realizar el histograma de frecuencias. (Reyes Flores, 2016)

iv. PREDICCIÓN

Una vez determinado la confiabilidad del modelo elegido se procede a la fase de pronóstico para futuros momentos aún no observados en la serie temporal. En el caso que se cuente con más de un modelo aceptado, se puede utilizar el análisis de los estadísticos t, F, Durbin-Watson y coeficiente de determinación múltiple considerando también los criterios que miel el margen de error cuadrático medio o el error absoluto medio.

4.2.2. REVISIÓN DE LA LITERATURA

(Jiménez G., Gázquez A., & Sánchez F., 2006) “La capacidad predictiva en los métodos Box-Jenkins y Holt-Winters: una aplicación al sector turismo”. El objetivo de dicha investigación es determinar la capacidad predictiva de dos métodos ampliamente utilizados tales como la metodología Box-Jenkins y el procedimiento de Holt-Winters. Para determinar la mejor capacidad predictiva, se utilizó el estudio de una serie con estacionalidad procedente del sector turismo en el cual la variable de

análisis es: Viajeros alojados en establecimientos hoteleros de la provincia de Almería durante el período 1993-2002. El análisis de dicha investigación concluye que los modelos ARIMA producen mejores resultados que los procedimientos clásicos de descomposición de series temporales; así mismo, señala que ambas metodologías: Box-Jenkins y Holt-Winters presentan una buena capacidad predictiva debido al componente estacional que presentan la variable.

(Marroquín Martínez & Chalita Tovar, 2011) “Application of Box-Jenkins methodology for forecasting prices in tomatoes” El objetivo de dicha investigación es identificar un modelo econométrico autorregresivo integrado de media móvil ARIMA que se ajuste al comportamiento de la serie de tiempo de precios nominales en vena al mayoreo del jitomate bola en México. Los datos utilizados para determinar el mejor modelo corresponden a los precios nominales por la venta al mayoreo de jitomate bola desde enero 1998 a noviembre 2008. El resultado de la investigación concluye que el mejor modelo ARIMA que se ajusta a los datos del precio de jitomate es el ARIMA (23,0,1); así mismo, permitió realizar pronósticos para 12 meses.

(De la Oliva de Con, Jimeno L., & Díaz de Villega, 2016) “An Approach to Box-Jenkins Methodology to Forecast EUR/USD”. Señala que el principal objetivo de dicha investigación es examinar la aplicación de los modelos autorregresivos para pronosticar las tasas de cambio EUR/USD. Para dicho análisis, los investigadores utilizaron datos de frecuencia mensual del tipo de cambio EUR/USD para el período enero 2004 – marzo 2014. El resultado de la investigación determina que los modelos autorregresivos y los modelos autorregresivos (MA) integrados de medias móviles (ARIMA) resultan favorables para predecir la trayectoria futura de las tasas de cambio de EUR/USD.

4.2.3. PROPUESTA DE ALTERNATIVA DE SOLUCIÓN

A) FASE DE COMPRENSIÓN DEL PROBLEMA

Todas las entidades del sector público del estado peruano se rigen a través del conjunto de principios, normas, políticas, procedimientos, técnicas e instrumentos mediante los cuales se organiza las actividades de la administración pública a través de los sistemas administrativos tales como: Gestión de Recursos humanos, Abastecimiento, Presupuesto Público, Tesorería, Endeudamiento Público, Contabilidad, Inversión Pública, Planeamiento Estratégico, Defensa Judicial del Estado, Control, Modernización de la Gestión Pública.

En el ámbito presupuestario, el sistema administrativo encargado es el Sistema Nacional de Presupuesto Público cuyo ente rector es el Ministerio de Economía y Finanzas a través de la Dirección General de Presupuesto Público. El Sistema Nacional de Presupuesto Público es el conjunto de principios, procesos, normas, procedimientos, técnicas e instrumentos que conducen el proceso presupuestario de todas las entidades del sector público en todas sus fases; está integrado por la Dirección General de Presupuesto Público, y por las Unidades Ejecutoras a través de las oficinas o dependencias en las cuales se conducen los procesos relacionados con el Sistema, a nivel de todas las entidades y organismos del Sector Público que administran fondos públicos La Ley vigente que tiene por objeto regular el Sistema de Presupuesto Público es el Decreto Legislativo N° 1440 publicado el 16 de setiembre del año 2018; dicho Decreto Legislativo en su artículo 22 señala que el proceso presupuestario comprende las fase de Programación Multianual, Formulación, Aprobación, Ejecución y Evaluación Presupuestaria. (Decreto Legislativo N°1440, 2018)

En la fase de Programación Multianual, las entidades públicas reciben la asignación presupuestal determinada por el Ministerio de Economía y Finanzas respecto a las fuentes de financiamiento: Recursos Ordinarios, Fondo de Compensación Municipal, Canon y Sobrecanon y regalías Mineras mientras que; para los recursos propios, las entidades tienen el deber de determinar y estimar los ingresos que recaudaran los

siguientes años fiscales, lo que corresponde a las fuentes de financiamiento: Impuestos Municipales y Recursos Directamente Recaudados.

En la actualidad, diversas entidades públicas presentan problemas respecto a la a la estimación de los ingresos presupuestarios de los siguientes años fiscales, por lo general, al no existir metodología definida, algunas entidades realizan sus estimaciones en base a promedios simples de la ejecución de ingresos de años anteriores el cual no siempre es una buena ya que no considera la tendencia y evolución de la recaudación de los recursos, y al finalizar los años fiscales se aprecia diferencias al comparar el presupuesto institucional aprobado y la ejecución presupuestaria real.

Entre los principales ingresos que recauda la entidad pública de la cual se evalúa en el presente trabajo de suficiencia profesional son: el impuesto de alcabala, impuesto al patrimonio vehicular, multas por infracciones al reglamento de tránsito, multas por infracciones al reglamento de transporte, arbitrios de limpieza pública, serenazgo y parques y jardines. En el siguiente cuadro se puede apreciar el porcentaje de participación que representan los principales ingresos recaudados en los periodos fiscales 2015 al 2018..

Tabla 2: Porcentaje de representación de los principales ingresos de la entidad

RUBRO/ PRINC. INGRESOS	INGRESOS (En millones de soles)				% de representación por rubro			
	2015	2016	2017	2018	2015	2016	2017	2018
RUBRO 08 IMPUESTOS MUNICIPALES								
ALCABALA	398	421	253	263	52%	52%	40%	41%
AL PATRIMONIO VEHICULAR	232	242	229	231	31%	30%	36%	36%
IMPUESTO A LOS JUEGOS DE MAQUINAS TRAGAMONEDAS	57	61	60	63	8%	7%	9%	10%
PREDIAL	54	55	57	60	7%	7%	9%	9%
INT.S POR DEP. DISTINTOS DE REC.S POR PRIVAT. Y CONC.	5	19	23	16	1%	2%	4%	2%
IMPUESTO A LOS CASINOS DE JUEGOS	4	4	4	4	1%	0%	1%	1%
FRACCIONAMIENTO TRIBUTARIO REGULAR	6	5	4	3	1%	1%	1%	1%
IMPUESTO A LOS ESPECTACULOS PUBLICOS NO DEPORTIVOS	2	2	2	2	0%	0%	0%	0%
OTROS	0	0	3	1	0%	0%	0%	0%
IMPUESTO A LAS APUESTAS	0	0	0	0	0%	0%	0%	0%
IMPUESTO A LOS JUEGOS	0	0	0	0	0%	0%	0%	0%
TOTAL RUBRO 09	759	809	633	644	100%	100%	100%	100%

RUBRO 09 RECURSOS DIRECTAMENTE RECAUDADOS	2015	2016	2017	2018	2015	2016	2017	2018
INFRACCIONES DE REGLAMENTO DE TRANSITO	95	92	94	114	26%	24%	24%	26%
INFRACCIONES DE REGLAMENTO DE TRANSPORTES	81	94	90	89	23%	25%	23%	21%
LIMPIEZA PUBLICA	49	52	55	62	14%	13%	14%	14%
OTROS DERECHOS E INGRESOS POR OTRAS CONCESIONES	18	21	22	24	5%	5%	6%	5%
SERENAZGO	26	21	20	22	7%	5%	5%	5%
INT. POR DEP. DISTINTOS DE REC.S POR PRIVAT. Y CONC.	9	9	9	19	2%	2%	2%	4%
OTRAS MULTAS	9	12	13	16	3%	3%	3%	4%
OTROS SERVICIOS CULTURALES Y RECREATIVOS	8	11	11	12	2%	3%	3%	3%
OTROS APORTES OBLIGATORIOS PARA INFRAESTRUCTURA	7	15	12	10	2%	4%	3%	2%
OTROS INGRESOS DIVERSOS	6	4	12	9	2%	1%	3%	2%
OTROS	51	52	51	55	14%	14%	13%	13%
TOTAL RUBRO 08	360	383	390	431	100%	100%	100%	100%

Fuente: Elaboración Propia

En el presente trabajo de suficiencia profesional se considera realizar el análisis y pronóstico del concepto de ingreso: **Impuesto Vehicular**, debido a que es uno de los ingresos con mayor recaudación anual tal como se puede apreciar en la tabla anterior.

Diversos estudios señalan que en las últimas décadas el parque automotor viene creciendo de manera sostenida, a nivel departamental, la mayor parte se concentra en la capital del país, lo cual representa una mayor variedad de oferta disponible para los diversos niveles socioeconómicos. Dicho crecimiento se ve reflejado también en el pago de impuesto vehicular que deben realizar por la adquisición de un vehículo. Como sabemos, el Impuesto Vehicular es un pago que se realiza anualmente el cual grava la propiedad de los vehículos, automóviles, camionetas, station wagons, buses y ómnibuses y es pagado durante un periodo de tres (03) años contados a partir del siguiente año en el que se realizó la primera inscripción en el Registro de Propiedad Vehicular.

B) FASE DE COMPRESIÓN DE LOS DATOS

El impuesto vehicular es recaudado por el Sistema de Administración Tributaria (SAT), quien se encarga de estimar y determinar los montos que los contribuyentes

deben de pagar por la adquisición de un vehículo. Posteriormente de realizar el cobro, el SAT transfiere los recursos a la Municipalidad Metropolitana de Lima.

Para efectos del presente análisis de series de tiempo, la data consta de ciento treinta (130) registros mensuales del ingreso del Impuesto Vehicular comprendidos entre enero del 2009 y octubre del 2019.

C) FASE DE PREPARACIÓN DE LOS DATOS

De acuerdo a la metodología de Box y Jenkins, antes de realizar el pronóstico de los datos de la serie de tiempo, primero se realiza el análisis del comportamiento de la serie temporal y se determina los componentes que presenta.

En la tabla 3 se observa un breve resumen de los datos, se aprecia que el promedio en el periodo establecido, desde enero del 2009 a octubre del 2019, del ingreso por concepto del impuesto vehicular asciende a S/ 16,288,979; sin embargo, los valores extremos registrados tanto mínimo y máximo ascienden a S/ 2,532,112 y S/ 91,153,836 respectivamente. Por lo tanto, se evidencia que los datos de la serie temporal no presentan normalidad y podrá ser corroborado mediante pruebas estadísticas.

Tabla 3: Principales estadísticas de la data.

summary(IV) En Soles (S/.)	
Min. :	2,532,112
1st Qu.:	6,619,614
Median :	10,353,255
Mean :	16,288,979
3rd Qu.:	15,813,920
Max. :	91,153,836

En la figura 3 se aprecia que la evolución temporal, la variable presenta una ligera tendencia creciente y comportamientos estacionales muy marcados lo cual conlleva a

considerar que no es un proceso estacionario y se requiere realizar transformaciones para estabilizar la varianza y la media.

SERIE TEMPORAL DE LOS INGRESOS DEL IMPUESTO VEHICULAR

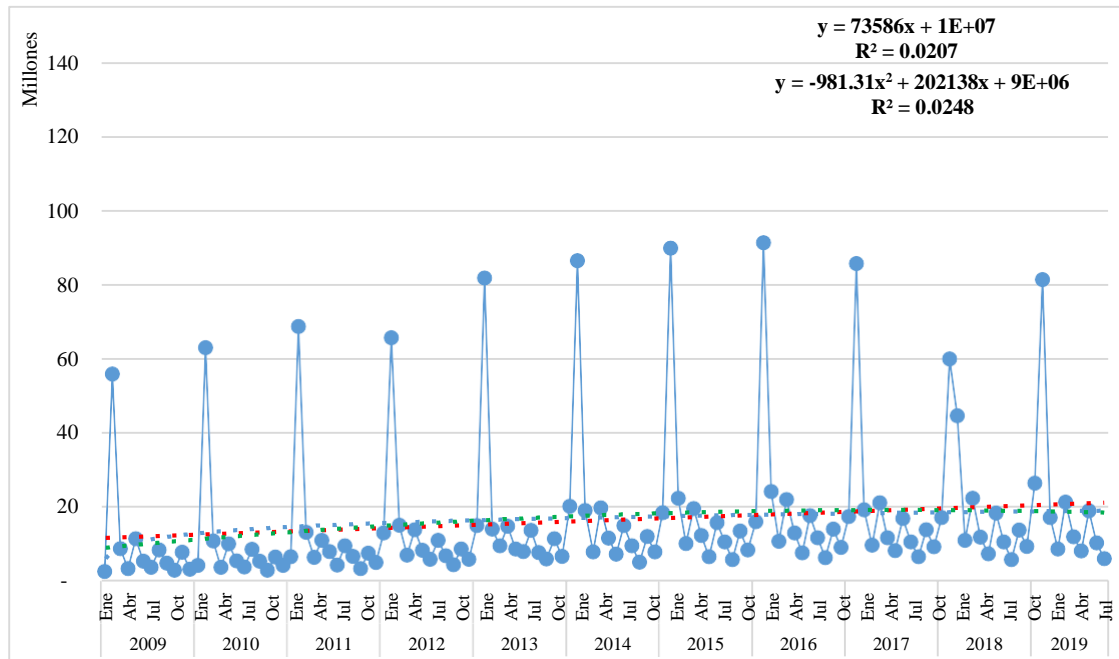


Figura 3: Serie de tiempo del impuesto vehicular entre enero del 2009 a octubre del 2019.

Fuente: Elaboración Propia

Para el análisis de la tendencia, se asume que la serie se encuentra libre del componente estacional. El porcentaje de variabilidad explicado por el coeficiente de determinación ajustado (R^2 ajustado) son bajos cercanos a cero. Se aprecia que los modelos: lineal o cuadrático no se ajustan a los datos de la serie temporal de ingresos del impuesto vehicular.

Tabla 4: Comparación del coeficiente de determinación de los modelos lineal y cuadrático.

<i>Modelo</i>	R^2 %	R^2 ajustado%
<i>Modelo Lineal</i>	2.07%	1.30%
<i>Modelo Cuadrático</i>	2.48%	0.94%

Al realizar el análisis de los residuos del modelo lineal y cuadrático, se aprecia que los residuos de dichos modelos no se encuentran normalmente distribuidos; el histograma de los residuos muestra la dispersión de estos los cuales son asimétricos; así mismo se aprecia que no tienen una varianza constante.

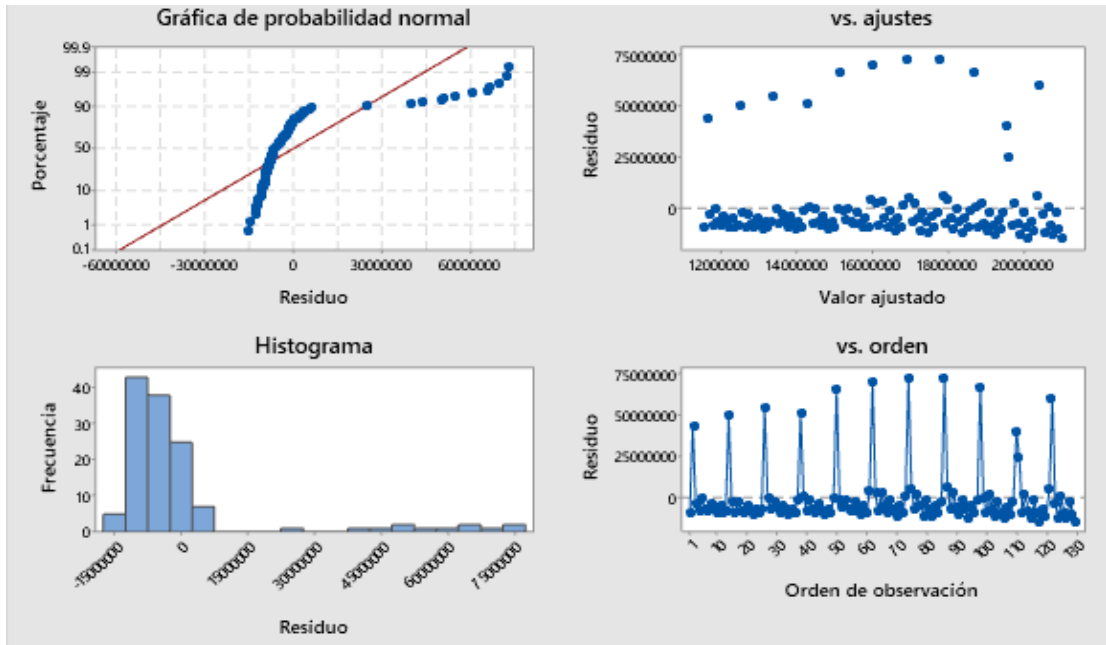


Figura 4: Gráficas de residuos para el modelo lineal.

.6

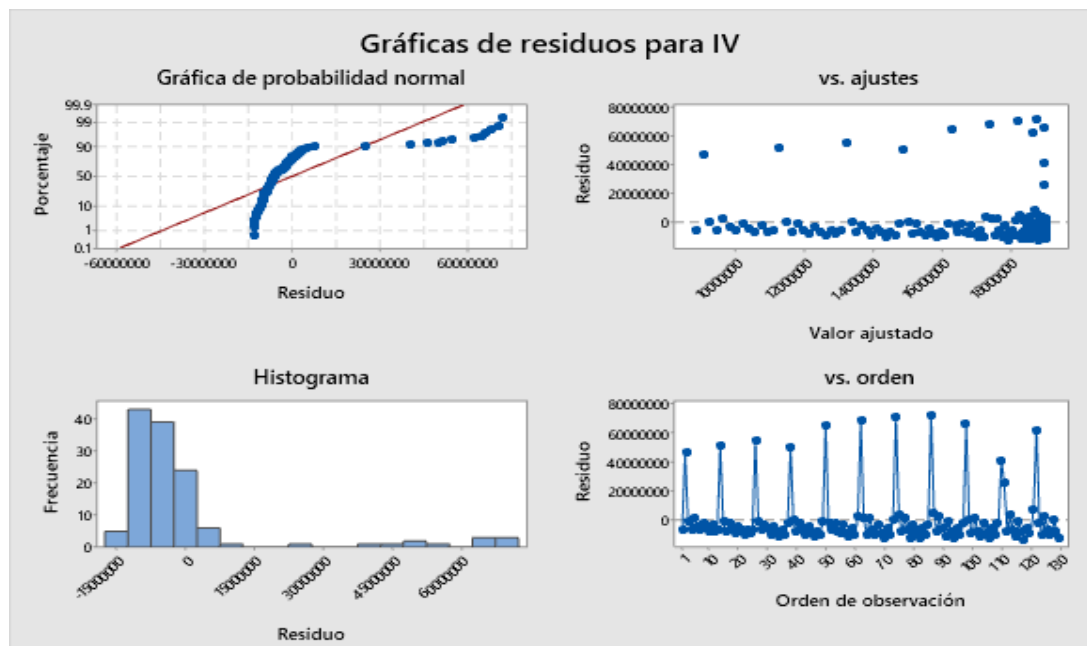


Figura 5: Gráficas de residuos para el modelo cuadrático.

PRUEBA DE NORMALIDAD DE LOS DATOS

H₀: Los datos se ajustan a una distribución normal

H₁: Los datos no se ajustan a una distribución normal

Tabla 5: Prueba de normalidad de los datos

Shapiro-Wilk normality test
data: datos[, 1]
W = 0.57985, p-value < 2.2e-16 : Rechaza Ho

Conclusión: A un nivel de significación de 0.05 se rechaza H₀. Por lo tanto, existe suficiente evidencia estadística para afirmar que la serie temporal de ingresos del impuesto vehicular desde el año 2009 hasta octubre del año 2019 no se ajustan a una distribución normal.

Analizando los diagramas de cajas por años tiene una ligera tendencia creciente; con respecto a su dispersión, se observa que en todos los años se presentan valores fuera de los rangos de los cuartiles, así mismo algunos años presentan mayor dispersión de los datos.

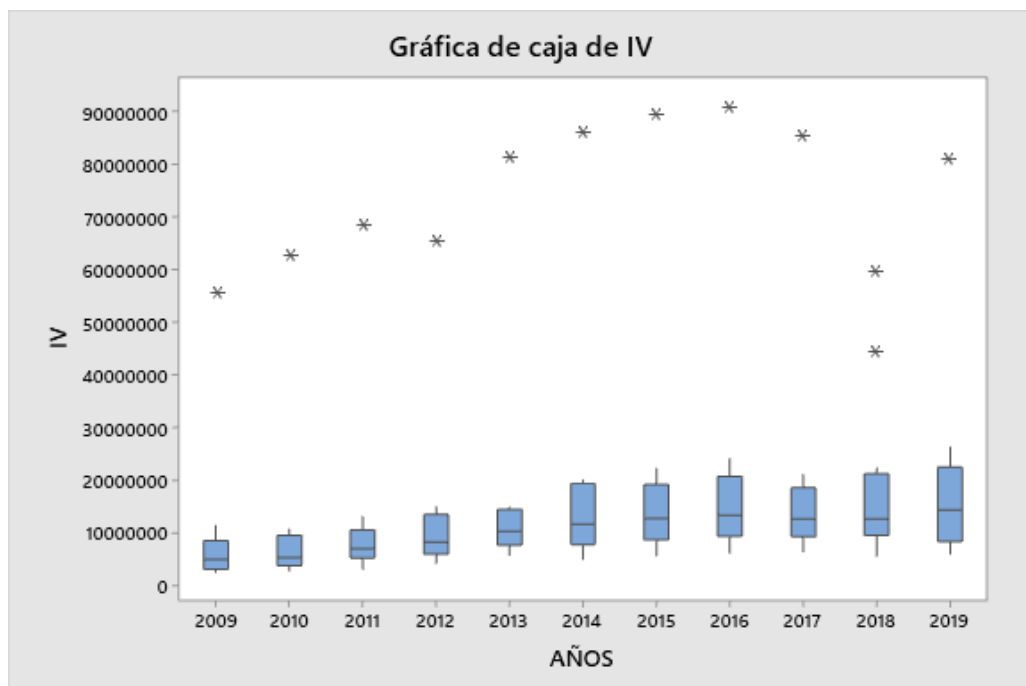


Figura 6: Gráficas de cajas anual

En relación a la gráfica de cajas por mes, se aprecia que no presenta tendencia; en relación a la dispersión, algunos meses presentan mayor dispersión tal como es el caso del mes de febrero, los datos se encuentran distribuidos asimétricamente en la mayoría de meses y se aprecia un valor atípico en el mes de marzo; asimismo, presenta patrones estacionales con montos de ingresos altos en el mes de febrero.

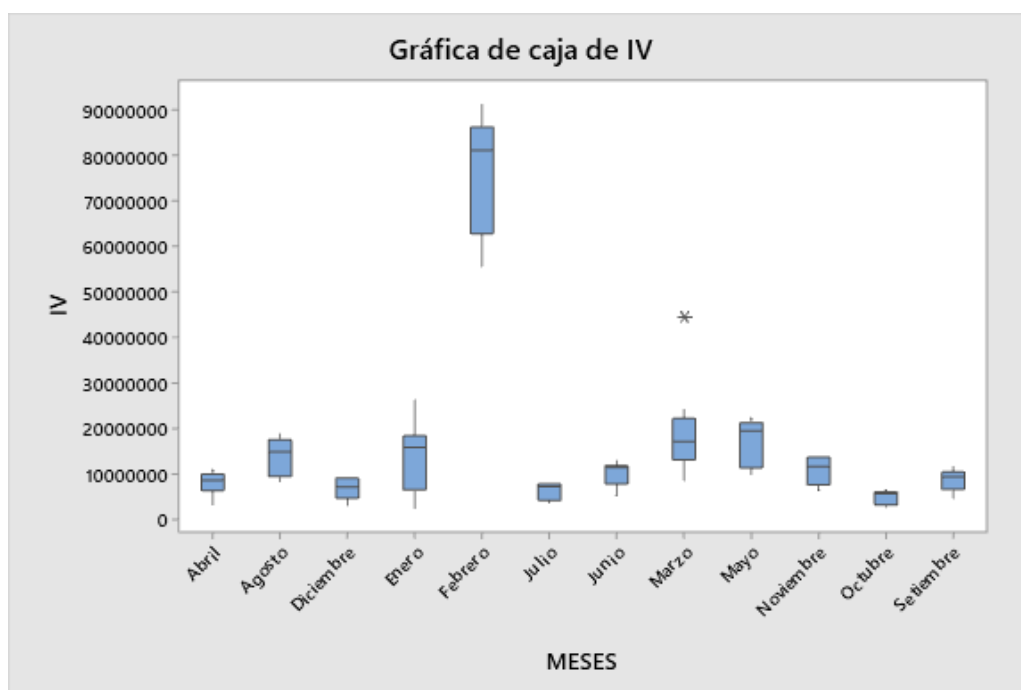


Figura 7: Gráficas de cajas mensual.

PRUEBA DE TENDENCIA

Tabla 6: Prueba de Kruskal-Wallis para analizar la tendencia

Estadísticas descriptivas					Prueba				
Año	N	Mediana	Clasificación de medias	Valor Z	Hipótesis nula	Hipótesis alterna	GL	Valor H	Valor p
2009	12	5047672	32.8	-3.15					
2010	12	5319260	34.2	-3.02					
2011	12	7030379	45.7	-1.91					
2012	12	8348720	57.6	-0.76					
2013	12	10362247	68.4	0.28					
2014	12	11722143	75.3	0.94					
2015	12	12799755	79.5	1.35					
2016	12	13385045	83.7	1.75					
2017	12	12677945	81.4	1.54					
2018	12	12742489	81.8	1.57					
2019	10	14446509	83.2	1.55					
General	130		65.5						

H₀:No existe tendencia en la serie

H₁:Existe tendencia en la serie

p-valor= 0.000 : Rechaza H₀

Conclusión: A un nivel de significación de 0.05 se rechaza H₀. Por lo tanto, existe suficiente evidencia estadística para afirmar que la serie temporal de ingresos del impuesto vehicular desde el año 2009 hasta octubre del año 2019 presenta tendencia.

PRUEBA DE HOMOGENEIDAD DE VARIANZAS

Tabla 7: Prueba de Homogeneidad de Varianzas a nivel anual

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar				Método		
Año	N	Desv.Est.	IC	Hipótesis nula	Todas las varianzas son iguales	
2009	12	14743017	(2326910; 122339045)	Hipótesis alterna	Por lo menos una varianza es diferente	
2010	12	16651946	(2560497; 141833109)	Nivel de significancia	$\alpha = 0.05$	
2011	12	17884974	(2779870; 150703729)	Pruebas		
2012	12	16679244	(2708592; 134518201)			
2013	12	20815253	(3288719; 172547431)	Método	Estadística de prueba	Valor p
2014	12	22006493	(3716769; 170650651)	Comparaciones múltiples	—	1.000
2015	12	22764313	(3822729; 177544588)	Levene	0.12	1.000
2016	12	22996707	(3858021; 179530630)			
2017	12	21436529	(3540818; 169971904)			
2018	12	16403636	(4553121; 77400341)			
2019	10	22162958	(3469822; 197646674)			

Nivel de confianza individual = 99.5455%

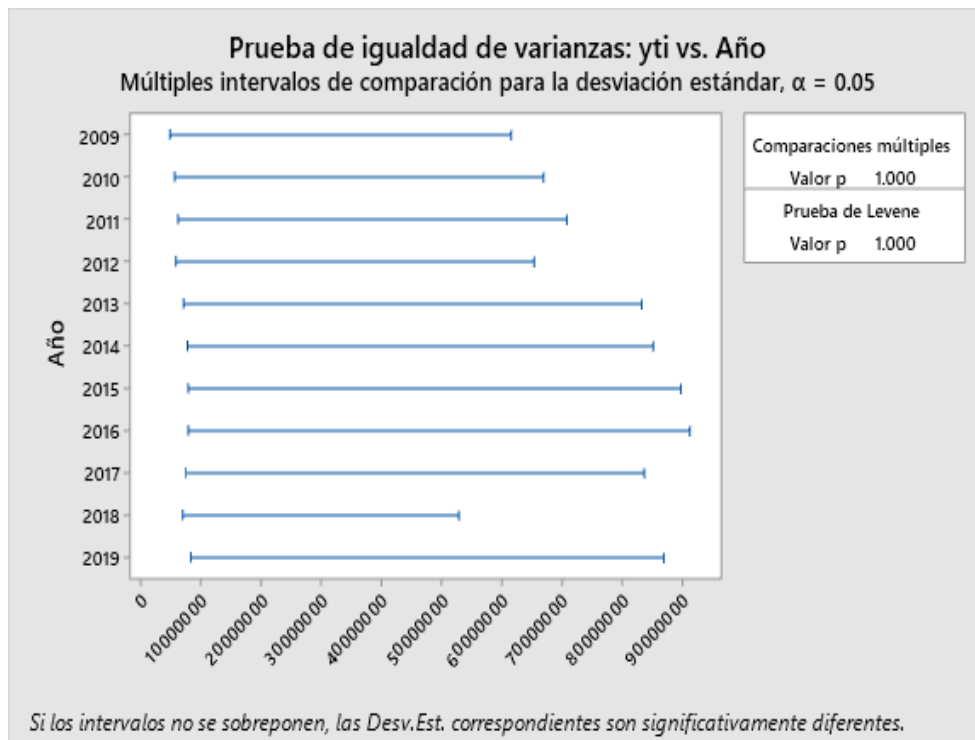


Figura 8: Prueba de Homogeneidad de Varianzas a nivel anual

Tabla 8: Prueba de Homogeneidad de Varianzas a nivel mensual

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

Mes	N	Desv.Est.	IC
Abr	11	2634208	(1235735; 7593181)
Ago	11	3971336	(2582317; 8258723)
Dic	10	2246589	(1248042; 5668132)
Ene	11	7179982	(3778912; 18447136)
Feb	11	12952127	(8710819; 26041872)
Jul	11	1747414	(918140; 4497092)
Jun	11	2775025	(1494148; 6969303)
Mar	11	9701254	(2381259; 53443894)
May	11	4823035	(3222546; 9760918)
Nov	10	2985914	(1875167; 6664024)
Oct	11	1388675	(774612; 3366404)
Set	11	2385187	(1445772; 5321010)

Nivel de confianza individual = 99.5833%

Método

Hipótesis nula Todas las varianzas son iguales
 Hipótesis alterna Por lo menos una varianza es diferente
 Nivel de significancia $\alpha = 0.05$

Pruebas

Método	Estadística de prueba	Valor p
Comparaciones múltiples	—	0.000
Levene	5.55	0.000

De acuerdo a los análisis realizados se determina que la serie temporal de ingresos de impuesto vehicular no es estacionaria. Dicha serie presenta los componentes de tendencia y una estacionalidad muy marcada, dado que el pago del impuesto vehicular se incrementa en los meses de febrero. Así mismo, la amplitud de las fluctuaciones no es constante. Antes de realizar el modelo estacionario se tiene que transformar la serie estacional.

ESTABILIZACIÓN DE LA VARIANZA

El uso de los logaritmos para poder estabilizar la varianza dado que funciona mejor cuando la variabilidad sea aproximadamente constante. A continuación, se presenta la serie temporal aplicando logaritmos.

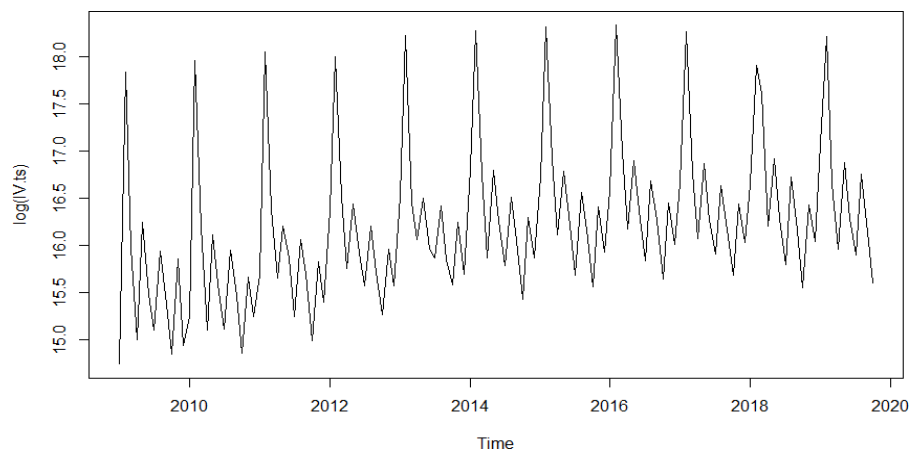


Figura 10: Gráfica de la serie temporal estabilizando la varianza

ELIMINACIÓN DE LA TENDENCIA

En cuanto a la eliminación de la tendencia, se realiza el proceso de diferenciación: $\nabla x_t = x_t - x_{t-1}$, debido a que la serie temporal presenta una ligera tendencia creciente.

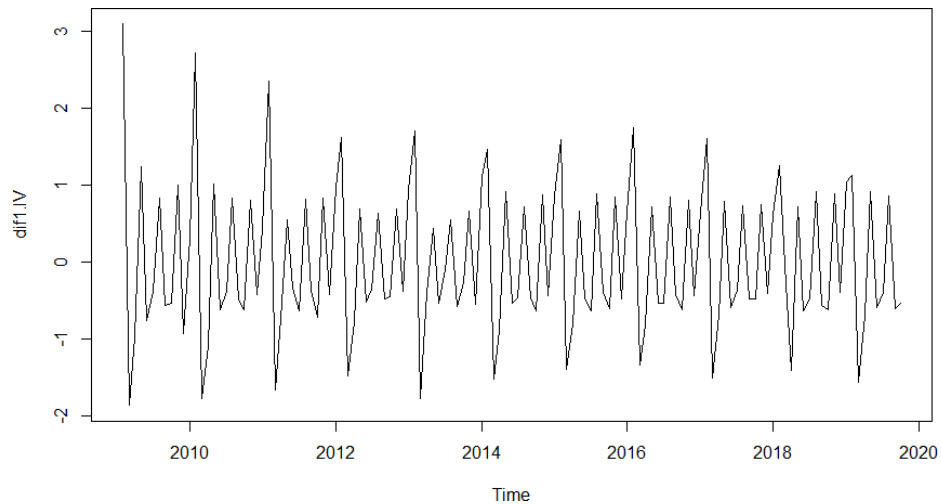


Figura 11: Gráfica de la serie temporal eliminando la tendencia

ELIMINACIÓN DE LA ESTACIONALIDAD

En relación al componente estacional muy marcado en la serie temporal de análisis, se realiza el proceso de diferenciación estacional de orden 12 debido a que la serie presenta ciclos estacionales a nivel mensual: $\nabla_{12}x_t = x_t - x_{t-12}$

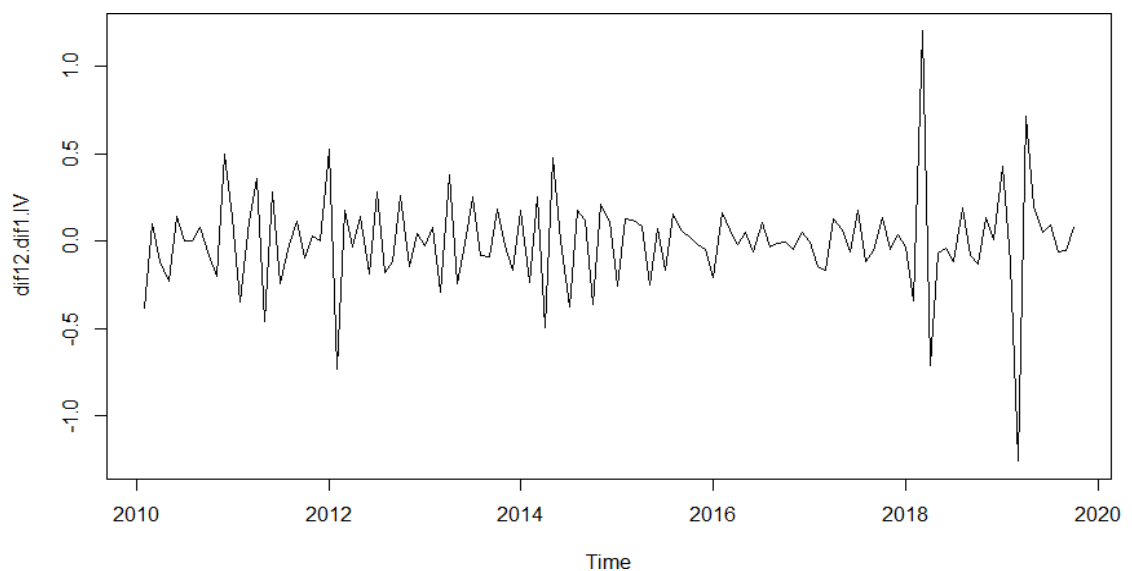


Figura 12: Gráfica de la serie temporal eliminando la estacionalidad

Habiendo eliminado ya los componentes de tendencia y estacionalidad, la serie parece ser estacionaria ya que es constante en media, asimismo, se presenta a

continuación las pruebas de estacionalidad como el Test de Dickey – Fuller y el test de KPSS (Kwiatkowski-Phillips-Schmidt-Shin).

H_0 : La serie es no estacionaria: tiene raíz unitaria

H_1 : La serie es estacionaria: no tiene raíz unitaria

Tabla 9 Test de Dickey – Fuller

Augmented Dickey-Fuller Test

data: Serie Estacional de Imp. Vehicular

Dickey-Fuller = -8.8731, Lag order = 4, **p-value = 0.01**

alternative hypothesis: stationary

Por lo tanto, se rechaza la hipótesis nula, existe suficiente evidencia estadística para afirmar que la serie temporal es estacionaria, es decir, no tiene raíz unitaria.

H_0 : La serie es estacionaria

H_1 : La serie es no estacionaria

Tabla 10: test de KPSS (Kwiatkowski-Phillips-Schmidt-Shin)

KPSS Test for Level Stationarity

data: Serie Estacional de Imp. Vehicular

KPSS Level = 0.042921, Truncation lag parameter = 4, **p-value = 0.1**

Por lo tanto, no se rechaza la hipótesis nula, no existe suficiente evidencia estadística para afirmar que la serie temporal no sea estacionaria.

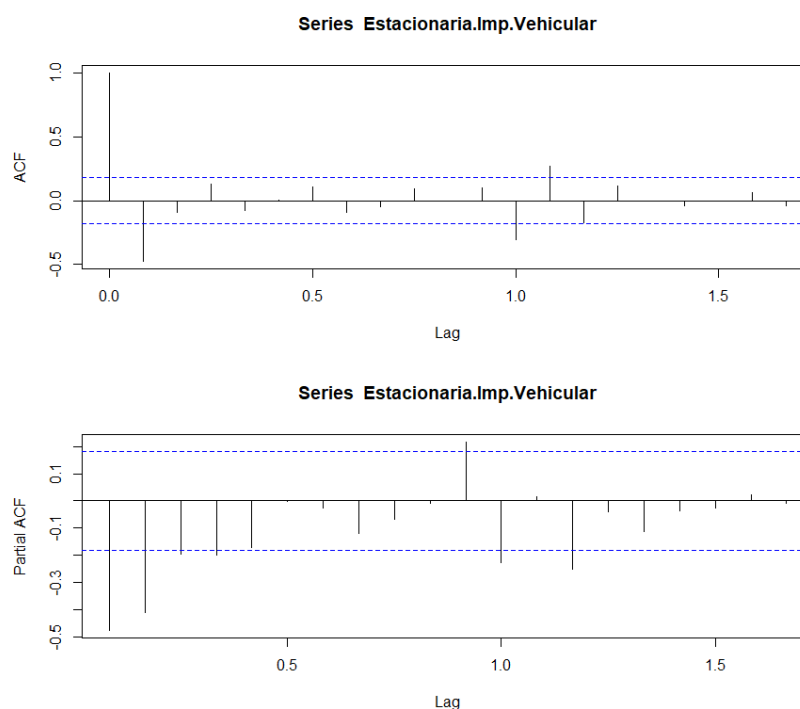
FUNCIONES DE AUTOCORRELACIÓN SIMPLE Y PARCIAL

Luego de haber realizado la transformación de la serie temporal de ingresos del impuesto vehicular para obtener una serie estacionaria, se realiza el análisis de las funciones de autocorrelación simple y parcial.

Tabla 11: Coeficientes de la Función de Correlación Simple y Parcial (ACF) y (PACF)

> acf(Estacionaria.Imp.Vehicular)\$acf	> pacf(Estacionaria.Imp.Vehicular)\$acf
[,1]	[,1]
[1,] 1.0000000000	[1,] -0.473637746
[2,] -0.4736377456	[2,] -0.408186799
[3,] -0.0922844321	[3,] -0.194704257
[4,] 0.1290213106	[4,] -0.197922375
[5,] -0.0775270068	[5,] -0.171746027
[6,] 0.0004149898	[6,] -0.002016193
[7,] 0.1066072057	[7,] -0.027496618
[8,] -0.0945391366	[8,] -0.119999777
[9,] -0.0505373311	[9,] -0.065867229
[10,] 0.0927915841	[10,] -0.009245258
[11,] -0.0065528583	[11,] 0.215119596
[12,] 0.0962505766	[12,] -0.226834430
[13,] -0.3092047108	[13,] 0.016197001
[14,] 0.2682372694	[14,] -0.250774705
[15,] -0.1767299415	[15,] -0.039429482
[16,] 0.1174277974	[16,] -0.113938900
[17,] -0.0040971499	[17,] -0.036140086
[18,] -0.0402904856	[18,] -0.024591968
[19,] -0.0057736647	[19,] 0.021235278
[20,] 0.0635593905	[20,] -0.007538732
[21,] -0.0371883906	

Figura 13: Función de Correlación Simple y Parcial (ACF) y (PACF)



En relación a las funciones de autocorrelación simple y parcial, se plantea los modelos tentativos que podrían ajustarse mejor a los datos de la serie temporal en análisis.

Tabla 12: Modelos ARIMA tentativos

Arima0	->	ARIMA (2,0,2) (1,1,2)
Arima1	->	ARIMA (0,1,2) (0,1,1)
Arima2	->	ARIMA (1,1,0) (2,1,0)
Arima3	->	ARIMA (1,1,2) (1,1,1)
Arima4	->	ARIMA (0,1,1) (0,1,1)
Arima5	->	ARIMA (1,1,0) (1,1,0)

D) FASE DE MODELADO

En relación a los modelos propuestos, se procedió a calcular los criterios AIC y BIC para cada modelo y así poder determinar qué modelo se ajusta mejor a los datos de la serie de tiempo.

Tabla 13: Criterios de Selección para cada modelo tentativo

MODELOS TENTATIVOS	DF	AIC	DF	BIC
Arima0 -> ARIMA (2,0,2) (1,1,2)	9	3970.736	9	3995.672
Arima1 -> ARIMA (0,1,2) (0,1,1)	4	3933.870	4	3944.919
Arima2 -> ARIMA (1,1,0) (2,1,0)	4	3985.733	4	3996.781
Arima3 -> ARIMA (1,1,2) (1,1,1)	6	3930.027	6	3946.600
Arima4 -> ARIMA (0,1,1) (0,1,1)	3	3937.445	3	3945.732
Arima5 -> ARIMA (1,1,0) (1,1,0)	3	3984.090	3	3992.377
Arima6 -> ARIMA (0,0,1) (1,1,2)	5	3967.959	5	3981.812
Arima7 -> ARIMA (2,0,2) (1,1,0)	6	3959.077	6	3975.701
Arima8 -> ARIMA (0,0,1) (0,1,2)	4	3969.094	4	3980.176

En el cuadro anterior se presentan los criterios AIC y BIC para cada modelo, se aprecia que el modelo que mejor se ajusta a los datos de la serie temporal es: **ARIMA (0,1,2)(0,1,1)**.

E) FASE DE EVALUACIÓN

Escogido el mejor modelo que se ajusta a la serie temporal de ingresos por concepto del impuesto vehicular, se procede a realizar la validación. A continuación se presenta las gráficas de autocorrelación simple y parcial de los residuos del modelo escogido ARIMA (0,1,2)(0,1,1).

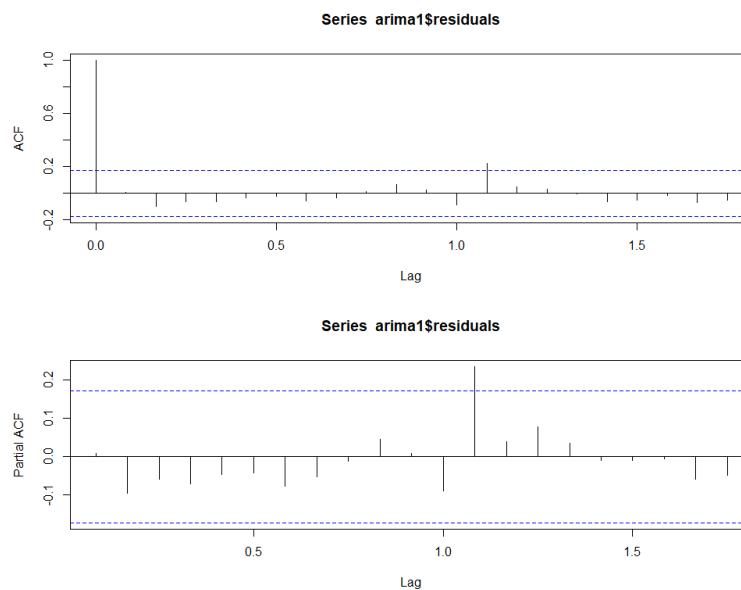


Figura 14: Función de autocorrelación simple y parcial de los residuales

Se aprecia en la Figura 15 que los correlogramas de los residuos no presentan residuos significativos, salvo en el gráfico de autocorrelación parcial donde se aprecia un rezago.

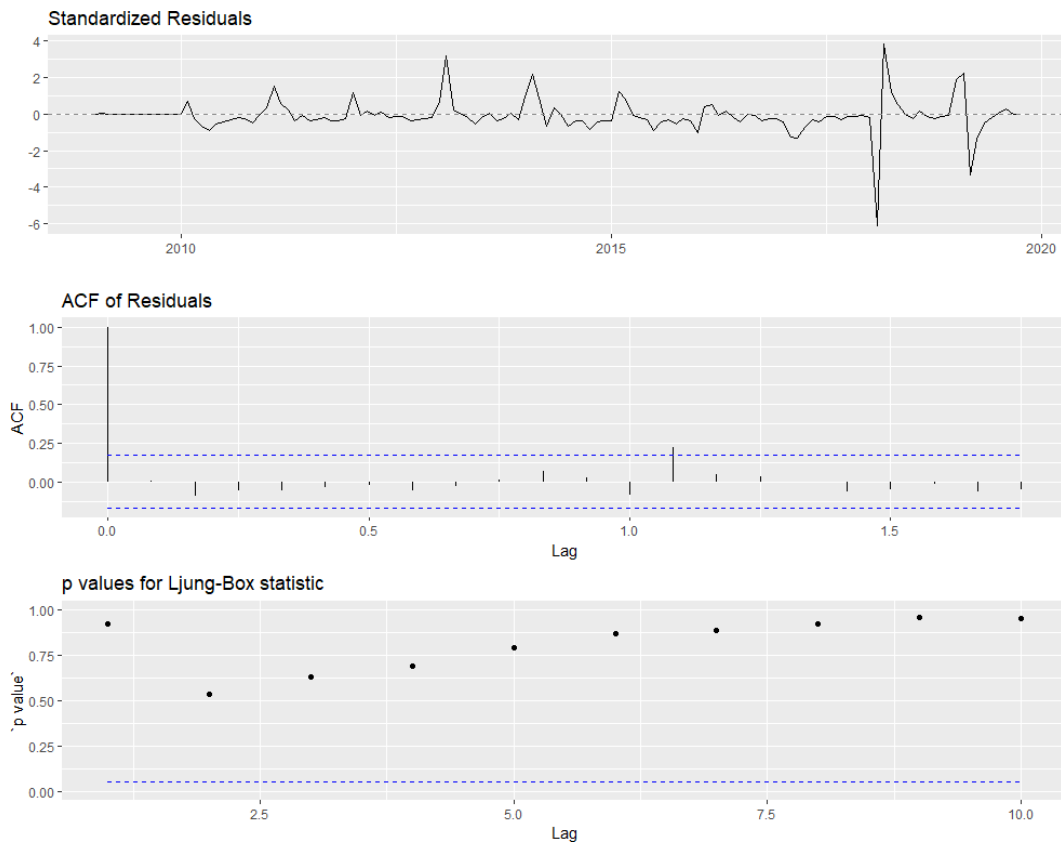


Figura 15: Gráficos de residuos del modelo elegido

Prueba de Box-Price

Tabla 14: Prueba de Box-Price

```
> bp <- Box.test(arima1$residuals) # Test de Box-Pierce
```

```
> bp
```

Box-Pierce test

data: arima1\$residuals

X-squared = 0.0091286, df = 1, p-value = 0.9239

Prueba de Ljung Box

Tabla 15: Prueba de Ljung Box

```
> lb <- Box.test(arima1$residuals, type="Ljung-Box") # Test de Ljung-Box
```

```
> lb
```

Box-Ljung test

data: arima1\$residuals

X-squared = 0.0093409, df = 1, p-value = 0.923

Prueba de Jarque Bera

Tabla 16: Prueba de Jarque Bera Fuente: Elaboración Propia

```
> jb <- jarque.bera.test(arima1$residuals) # Test de Jarque-Bera
```

```
> jb
```

Jarque Bera Test

data: arima1\$residuals

X-squared = 1394.7, df = 2, p-value < 2.2e-16

Prueba de Shapiro Wilk

Tabla 17: Prueba de Shapiro Wilk Fuente: Elaboración Propia

```
> sht<-shapiro.test(arima1$residuals) # Test de Shapiro-Wilk
```

```
> sht
```

Shapiro-Wilk normality test

data: arima1\$residuals

W = 0.69941, p-value = 5.672e-15

F) FASE DE IMPLEMENTACIÓN

Una vez elegido el modelo que mejor se ajusta a los datos de la serie de tiempo, se procede a realizar la implementación el cual consiste en realizar los pronósticos para el periodo que comprende los meses de noviembre 2019 a diciembre 2020.

Tabla 18: Pronostico del impuesto vehicular para el periodo Nov. 2019 a Dic 2020

> forecast1<-forecast(arima1, level = c(95), h = 14)

> autoplot(forecast1)

> forecast1

PERIODO	PRONÓSTICO	ICI(95%)	ICS(95%)
Nov 2019	13,893,260	4,833,581	22,952,939
Dec 2019	9,316,119	76,109	18,556,129
Jan 2020	22,270,186	13,016,829	31,523,543
Feb 2020	77,348,429	68,081,744	86,615,115
Mar 2020	24,837,955	15,557,961	34,117,949
Apr 2020	9,623,397	330,113	18,916,681
May 2020	21,616,690	12,310,135	30,923,244
Jun 2020	12,052,826	2,733,019	21,372,632
Jul 2020	8,006,542	-1,326,498	17,339,582
Aug 2020	18,437,342	9,091,088	27,783,596
Sep 2020	10,606,151	1,246,701	19,965,601
Oct 2020	6,196,728	-3,175,900	15,569,355
Nov 2020	14,000,397	3,181,195	24,819,598
Dec 2020	9,448,638	-1,381,528	20,278,805

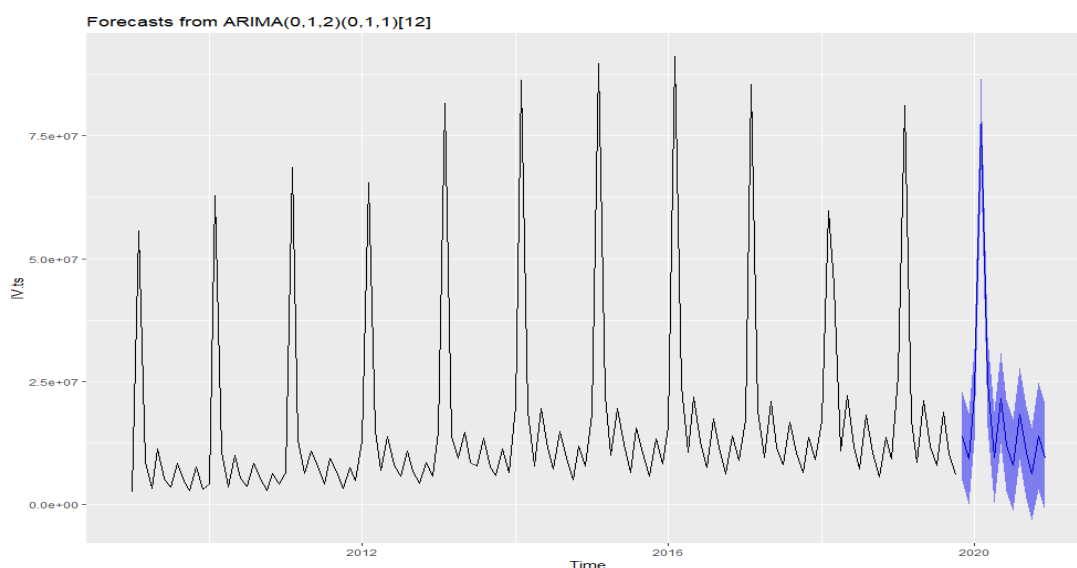


Figura 16: Pronostico de la serie de tiempo impuesto vehicular

4.3 CONTRIBUCIÓN EN LA SOLUCIÓN DE SITUACIONES PROBLEMÁTICAS

La aplicación de la técnica estadística para el pronóstico de los ingresos por concepto de Impuesto vehicular utilizando la metodología de Box y Jenkins permitió realizar una estimación más congruente con el comportamiento de la serie temporal permitiendo conocer sus componentes principales como tendencia y estacionalidad.

Así mismo, la estimación de pronósticos mejora la confiabilidad de los posibles ingresos que se registrarán durante el siguiente periodo fiscal en comparación con el uso de promedios simples para cada periodo estimado ya que como se pudo apreciar, durante los meses de febrero de todos los años considerados en el presente trabajo muestran picos altos de ingresos.

4.4 ANÁLISIS DE LA CONTRIBUCIÓN EN TÉRMINOS DE COMPETENCIA Y HABILIDADES

El uso de técnicas estadísticas como las series de tiempo desarrollado en el presente trabajo de especialización profesional permitió ampliar los conocimientos adquiridos en la etapa universitaria. La aplicación de la metodología de Box y Jenkins aplicado a una serie de tiempo real el cual presenta características propias de acuerdo a su naturaleza permitió realizar una investigación más profunda y revisión de la literatura de la técnica aplicada en diversas investigaciones aplicados a problemas de pronósticos y estimaciones los cuales presentan mejores resultados.

4.5 NIVEL DE BENEFICIO OBTENIDO POR EL CENTRO LABORAL

Entre los principales beneficios obtenidos por la entidad recaudadora del impuesto vehicular son:

- Estimación de la recaudación del impuesto vehicular más confiable en comparación a la aplicación de promedios simples.

- En relación a la fase de Programación Multianual del Proceso Presupuestario permitió determinar con mayor eficiencia el uso de los recursos por concepto del impuesto vehicular.

V. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

En relación al objetivo general y los objetivos específicos planteados en el presente trabajo de suficiencia profesional se concluye lo siguiente:

- Se logró determinar y estimar los pronósticos del registro de ingresos del Impuesto Vehicular mediante el uso de series de tiempo, para un periodo de catorce (14) meses en el horizonte temporal noviembre 2019 – diciembre 2020.
- Se identificó las principales características y componentes del registro de ingresos mensuales del Impuesto Vehicular, durante el periodo enero 2009 – octubre 2019.
- Se estableció el mejor modelo para el pronóstico del registro de ingresos mensuales del Impuesto Vehicular con los datos del periodo enero 2009 – octubre 2019.

5.2 RECOMENDACIONES

- Si bien, la Metodología de Box y Jenkins presenta una estructura que abarca desde el análisis de la serie de tiempo, la determinación del mejor modelo que se ajuste a los datos, la validación del modelo mediante pruebas estadísticas y finalmente el pronóstico para futuros escenarios; no se descarta la aplicación de otras metodologías tales como las Redes Neuronales las cuales han demostrado ser modelos que se ajustan a la estructura de los datos de series de tiempo. En tal sentido, es recomendable realizar comparaciones para medir la eficiencia entre Box y Jenkins y Redes Neuronales.

- Asimismo, habiéndose podido apreciar que es factible utilizar técnicas estadísticas para la estimación de uno de los principales ingresos que recauda la entidad, se recomienda para futuras investigación ampliar la metodología de series de tiempo para conceptos de ingresos que presenten características similares en relación a los componentes principales del Impuesto Vehicular, tales como el Impuesto de Acabala, los arbitrios de Limpieza Pública, Serenazgo y Parques y Jardines.

VI. REFERENCIAS BIBLIOGRÁFICAS

- Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2016). *Times Series Analysis: Forecasting and Control - 5th Edition*. Canada: Jhon Wiley & Sons.
- Castillo P., R., & Varela L, R. (2008). *Econometría Practicas: Fundamentos de Series de Tiempo*.
- De la Oliva de Con, F., Jimeno L., R., & Díaz de Villega, L. (2016). An approach to Box-Jenkins Methodology to Forecast EUR/USD Rate Exchange. *Cofin Habana*, 10(1), 57-75.
- Decreto Legislativo N°1440. (16 de diciembre de 2018). Decreto Legislativo del Sistema Nacional de Presupuesto Público. *Diario Oficial El Peruano*.
- Jiménez G., J. F., Gázquez A., J. C., & Sánchez F., R. (2006). La capacidad predictiva en los metodos Box-Jenkins y Holt-Winters: una aplicación al sector turismo. *Revista Europa de Dirección y Economía de la Empresa*, 15(3), 185-198.
- Marroquín Martínez, G., & Chalita Tovar, L. E. (2011). Application of Box-Jenkins methodology for forecasting prices in tomatoes. *Revista Mexicana de Ciencias Agrícolas*, 2(4), 573-577.
- Mauricio, J. (2007). *Introducción al Análisis de Series Temporales*. Madrid.
- Méndez Bustamante, A. (2013). *Series de tiempo: Una aplicación a registros hidrométricos en una cuenca de Oaxaca*. Mexico.
- Murdoch, B., Krause, P., & Guy, P. (2015). An Analysis Of Using Time-Series Current And Defferes Income Tax Expense To Forecast Income Taxes Paid. *The Journal of Applied Business Research*, 31(3), 1015-1022. Obtenido de <https://doi.org/10.19030/jabr.v31i3.9233>
- Reyes Flores, M. (2016). *Análisis de Series de Tiempo Univariadas y Multivariadas mediante modelos ARIMA y MVAR*. Universidad Autónoma Metropolitana - Iztapalapa, Ciudad de México.

VII. ANEXOS

Tabla 19: Datos de registro de los ingresos del Impuesto Vehicular

MES/ AÑO	2009	2010	2011	2012	2013	
Ene	2,532,112	4,184,990	6,489,216	12,894,860	14,912,107	
Feb	55,713,920	62,874,245	68,515,191	65,519,089	81,603,630	
Mar	8,664,211	10,750,371	13,101,827	14,936,108	13,894,414	
Abr	3,287,391	3,624,907	6,315,793	6,937,461	9,420,173	
May	11,355,490	9,960,658	10,935,569	13,773,543	14,664,756	
Jun	5,314,180	5,377,566	7,828,185	8,171,358	8,574,956	
Jul	3,625,422	3,678,607	4,195,331	5,799,811	7,830,916	
Ago	8,335,239	8,463,233	9,468,440	10,891,161	13,554,103	
Set	4,781,163	5,260,953	6,586,635	6,718,549	7,626,390	
Oct	2,822,625	2,854,778	3,242,581	4,288,501	5,840,440	
Nov	7,707,858	6,383,945	7,474,122	8,526,081	11,304,321	
Dic	3,072,618	4,171,460	4,892,435	5,820,525	6,542,492	
MES/ AÑO	2014	2015	2016	2017	2018	2019
Ene	20,035,147	18,358,870	15,878,422	17,271,875	17,020,780	26,277,665
Feb	86,293,448	89,705,822	91,153,836	85,536,992	59,846,787	81,200,059
Mar	18,922,654	22,228,249	24,094,258	19,133,571	44,477,059	17,082,296
Abr	7,806,001	9,990,606	10,624,562	9,582,803	10,906,465	8,547,568
May	19,619,563	19,501,624	21,850,622	21,031,363	22,268,106	21,238,827
Jun	11,475,651	12,233,399	12,848,747	11,627,399	11,782,738	11,810,721
Jul	7,193,066	6,469,175	7,529,407	8,094,276	7,289,750	8,022,198
Ago	14,841,084	15,620,413	17,542,756	16,774,053	18,230,591	18,825,513
Set	9,389,340	10,483,894	11,571,814	10,464,475	10,479,038	10,242,035
Oct	5,012,026	5,721,289	6,256,846	6,474,736	5,660,676	5,995,670
Nov	11,968,634	13,366,111	13,921,343	13,728,490	13,702,239	
Dic	7,771,076	8,263,905	9,032,710	9,197,657	9,277,457	