ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA ESTADÍSTICA

MODELIZACIÓN ESTADÍSTICA - INFORMÁTICA CON R DE DELITOS EN LA PROVINCIA DE CHIMBORAZO

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERO ESTADÍSTICO

AUTORES:

JHEFRI BLADIMIR LÓPEZ CASTILLO

MARIO GUSTAVO MOROCHO SALINAS

Riobamba – Ecuador

2022

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA ESTADÍSTICA

MODELIZACIÓN ESTADÍSTICA - INFORMÁTICA CON R DE DELITOS EN LA PROVINCIA DE CHIMBORAZO

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERO ESTADÍSTICO

AUTORES: JHEFRI BLADIMIR LÓPEZ CASTILLO

MARIO GUSTAVO MOROCHO SALINAS

DIRECTOR: DR. JORGE WASHINGTON CONGACHA AUSHAY, MSc.

Riobamba – Ecuador

2022

Nosotros, JHEFRI BLADIMIR LÓPEZ CASTILLO y MARIO GUSTAVO MOROCHO SALINAS, declaramos que el presente Trabajo de Integración Curricular es de nuestra autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autores asumimos la responsabilidad legal y académica de los contenidos de este Trabajo de Integración Curricular; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 23 de noviembre 2022

Jhefri Bladimir López Castillo
1400762017

Mario Gustavo Morocho Salinas
1400848170

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Integración Curricular, certifica que: El Trabajo de Integración Curricular; Tipo: Proyecto de Investigación, **MODELIZACIÓN ESTADÍSTICA – INFORMÁTICA CON R DE DELITOS EN LA PROVINCIA DE CHIMBORAZO**, realizado por los señores: **JHEFRI BLADIMIR LÓPEZ CASTILLO** y **MARIO GUSTAVO MOROCHO SALINAS**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Integración Curricular. El mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal autoriza su presentación.

	FIRMA	**FECHA**
Ing. Johanna Enith Aguilar Reyes, Mgs. **PRESIDENTE DEL TRIBUNAL**		2022-11-23
Dr. Jorge Washington Congacha Aushay, MSc. **DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR**		2022-11-23
Dra. Jaqueline Elizabeth Balseca Castro, Mgs. **ASESORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR**		2022-11-23

DEDICATORIA

El presente trabajo va dedicado principalmente a Dios por darme la oportunidad de finalizar esta etapa que es mi formación como profesional, a mi familia por la confianza y el apoyo incondicional que me han brindado en este proceso formativo, a mis amigos por estar presentes y compartir momentos y anécdotas inolvidables.

Mario

Dedico el presente trabajo primeramente a Dios, que fue el que me permitió culminar con éxito esta excelente etapa de vida, la cual pude entender y valorar cada una de las bendiciones con las cuales él me rodea. a mi hija, que es el motor de mi vida fue parte muy importante de lo que hoy puedo presentar como tesis, por cada palabra de apoyo, gracias por cada momento en familia.

Jhefri

AGRADECIMIENTOS

Queremos agradecer primeramente a la Escuela Superior Politécnica de Chimborazo por brindarnos la oportunidad de formarnos como profesionales en esta prestigiosa institución y con ello aportar con nuestros conocimientos a la sociedad y ser parte del desarrollo del país.

Un más sincero agradecimiento a los docentes que con su paciencia y dedicación han sabido inculcar en nosotros sus conocimientos.

Jhefri & Mario

ÍNDICE DE CONTENIDO

CAPÍTULO III

ÍNDICE DE TABLAS

ÍNDICE DE ILUSTRACIONES

ÍNDICE DE ECUACIONES

ÍNDICE DE ANEXOS

RESUMEN

El presente trabajo de investigación propuso implementar un modelo SARIMA con el fin realizar un pronóstico del número de delitos en la provincia de Chimborazo. Para la ejecución de la investigación se trabajó con la información mediante datos de delitos mensuales que concierne al periodo enero 2015-abril 2022 proporcionada por la Fiscalía Provincial de Chimborazo. Para el proceso de modelización se utilizó la metodología de Box-Jenkins mediante el software estadístico R. Mediante el análisis exploratorio univariante se detectó la presencia de datos anómalos que fueron reemplazados con el promedio de delitos correspondiente al mes del dato atípico. Una vez estacionarizada la serie, mediante un análisis de autocorrelación se estimó los parámetros AR y MA de la parte regular y estacional. A través de la función auto.arima de R y modelos planteados se obtuvo el modelo SARIMA(2,1,1)(1,0,0)[12]; siendo este el más adecuado. Determinado el mejor modelo se verificó que cumple con los supuestos de normalidad, independencia y homocedasticidad garantizando así que el modelo se ajusta de manera adecuada a los datos. Por último, se procedió a realizar el pronóstico de los delitos para los próximos 12 meses en el periodo mayo 2022-abril 2023 indicando que existirá un incremento en el mes de mayo con un total de 727 delitos siendo este el más elevado en comparación con los meses restantes que descienden oscilando entre 680 a 696 delitos mensuales, por lo que se recomienda a la Fiscalía Provincial de Chimborazo tomar acciones de control y fortalecer la seguridad durante estos meses.

Palabras clave: <SERIES DE TIEMPO>, <METODOLOGÍA BOX-JENKINS>, <PRONÓSTICO DE DELITOS>, <MODELOS SARIMA >, <ANÁLISIS ESTADÍSTICO>.

2432-DBRA-UTP-2022

SUMMARY

The present research work proposed to implement a SARIMA model in order to make a forecast of the number of crimes in the province of Chimborazo. For the execution of the investigation, the information through monthly crime data that concerns the period January 2015-April 2022 provided by the Provincial Prosecutor's Office of Chimborazo was used. For the modeling process, the Box-Jenkins methodology was utilized employing the R statistical software. Through univariate exploratory analysis, the presence of anomalous data was detected, which was replaced with the average number of crimes corresponding to the month of the atypical data. Once the series was stationary, with an autocorrelation analysis, the AR and MA parameters of the regular and seasonal part were estimated. Applying the auto.arima function of R and the proposed models, the SARIMA(2,1,1)(1,0,0)[12] model was obtained, being the most appropriate. Once the best model was determined, it was verified that it complies with the assumptions of normality, independence and homoscedasticity, thus guaranteeing that the model adequately adjusts to the data. Finally, the forecast of crimes for the next 12 months in the period May 2022-April 2023 was done, indicating that there will be an increase in the month of May with a total of 727 crimes, this being the highest compared to the remaining months that descend ranging from 680 to 696 monthly crimes, so it is recommended that the Provincial Prosecutor’s Office of Chimborazo take control actions and strengthen security during those months.

Keywords: <TIME SERIES>, <BOX-JENKINS METHODOLOGY>, <CRIME FORECAST>, <SARIMA MODELS>, <STATISTICAL ANALYSIS>.

Edgar Mesias Jaramillo Moyano
0603497397

INTRODUCCIÓN

En la actualidad la delincuencia ha tomado fuerza en varias provincias del Ecuador, en particular en la provincia de Chimborazo, que debido a la crisis económica post pandemia y fragilidad de seguridad por parte de la Policía Nacional, se ha convertido en una zona vulnerable para los actos delictivos.

El índice de delitos de los últimos años ha ido incrementando y su vez se está perdiendo el control de la seguridad y el bienestar de la sociedad, dado que se está produciendo delitos en plena luz del día, sin esta tener represalias y sin dar con los posibles culpables, dejando a la sociedad sin ese derecho de seguridad en cuanto a su salud, bienestar y pertenencia.

Los delitos son considerados acciones en donde se violentan las reglas o leyes constituidas para proteger y velar el bienestar y la integridad de una sociedad, estos delitos pueden ser de forma violenta como física, psicológica o moral. Todo acto que provoque o vaya en contra del bienestar de la ciudadanía es sancionado y se aplicará a toda persona que incumpla con las leyes sin importar raza, rango o función que desempeñe, sin privilegio alguno.

El derecho a la seguridad necesita emplearse en un ambiente de respeto a la ley, a los derechos de cada ciudadano y al bienestar social, para que esto puede emplearse se requiere de acciones de control concretas de forma integral y de manera permanente, tomando en cuenta la participación de los entes de seguridad de forma activa y responsable y destacando de manera especial la organización y la transparencia de las autoridades.

Por ende, este proyecto estadístico aportará como una referencia para diseñar o proyectar un plan de mejora en el ámbito de la seguridad social, ya sea esta por la creación de políticas de orden y seguridad, haciendo referencia en información real, en base a un análisis estadístico verídico y conforme las necesidades de los datos en relación con los actos delictivos de la Provincia de Chimborazo.

CAPÍTULO I

1. PROBLEMA DE INVESTIGACIÓN

1.1. Planteamiento del Problema

El derecho de la seguridad ciudadana se ha convertido en unos de los problemas principales por los actos delictivos que aborda la sociedad, por tal razón para la toma de decisiones de seguridad es de importancia tener una referencia previa, para ello es indispensable conocer la cantidad de actos delictivos que se generan y forman parte de fiscalización. De esta manera se logrará focalizar de manera estadística las características, las cifras e índices más relevantes sobre la percepción de delitos en la provincia de Chimborazo.

1.2. Limitaciones y delimitaciones

Limitación

La principal complicación dentro del proyecto es el tiempo de realización de la investigación, siendo una necesidad para el análisis estadístico completo y por ende en un pronóstico adecuado, tomando en cuenta esto de vital importancia ya que la disponibilidad del tiempo se puede alterar de acuerdo con lo establecido en la planificación y no cumplir con los objetivos planteados a tiempo.

Delimitaciones

El presente Proyecto de Investigación consta de las siguientes delimitaciones:

Delimitación Espacial: Provincia de Chimborazo
Delimitación Temporal: Periodo Enero 2015 - Abril 2022

1.3. Problema General de Investigación

¿Cuántos delitos ingresarán mensualmente a la fiscalía provincial de Chimborazo, FGE?

1.4. Problemas específicos de investigación

- ¿Cuáles son las características que presentan los datos sobre los delitos generales en la provincia de Chimborazo?
- ¿Existirá algún modelo óptimo que ayude a pronosticar los delitos en la provincia de Chimborazo?
- ¿Qué modelo se ajustará para la adecuada ejecución en la predicción de los delitos?
- ¿Cuál será el mejor criterio de información para la elección del mejor modelo?
- ¿Será el modelo planteado la mejor elección para realizar predicciones de los delitos?

1.5. Objetivos

1.5.1. Objetivo General

- Predecir el número de delitos en la provincia de Chimborazo mediante modelos SARIMA utilizando la metodología de Box-Jenkins.

1.5.2. Objetivos Específicos

- Realizar un análisis exploratorio de datos univariante de los delitos.
- Identificación del modelo adecuado.
- Estimación de parámetros del modelo adecuado.
- Comprobar o verificar con AIC o BIC los supuestos de los modelos ARIMA estacionales.
- Aplicar modelos SARIMA con el fin de realizar predicciones de los delitos.

1.6. Justificación

1.6.1. Justificación Teórica

La técnica de análisis de datos se ha convertido en una herramienta de utilidad para la Fiscalía General del Estado y la entidad policial, que pueda analizar la probabilidad y pronosticar el número de delitos en determinadas zonas de la provincia de Chimborazo.

El procesamiento de grandes volúmenes de datos tiene la capacidad de predecir donde y cuando se puede llevar acabo un delito por medio de modelos SARIMA o también llamado ARIMA Estacional cuyos modelos captan el comportamiento de la serie de observaciones tomadas en el tiempo.

Por lo tanto, dicha investigación tiene como finalidad comprobar la efectividad de los modelos estimados por medio de la metodología Box-Jenkins y permitir la elección del mejor modelo para predecir.

1.6.2. Justificación Metodológica

Para la realización de la investigación se procedió mediante métodos de análisis predictivo, descriptivos e inferencial, lo cual se pudieron realizar gracias a las técnicas de recolección de datos que tiene la Fiscalía General del Estado, por lo tanto, es necesario para obtener una predicción clara y concisa mediante la metodología Box-Jenkins con el modelo optimo que sirva como aporte a futuros estudios y pronosticar de mejor manera los resultados.

1.6.3. Justificación Práctica

La Fiscalía General del Estado del cantón Riobamba es una institución autónoma, que dedica tiempo a investigar y registrar los procesos delictivos. por ende, dicha entidad necesita de un modelo estadístico para predecir futuros números de delitos que se puedan ocasionar y obtener conocimiento sobre la tasa de delitos, que permitirá conocer el estado actual del riesgo que corre la sociedad. La institución en la actualidad no cuenta con un análisis exploratorio e inferencial univariante y multivariante de sus datos que ingresan mensualmente. por ello el presente trabajo pretende realizar un análisis exploratorio y predictivo de datos de todo tipo de delitos con el propósito de pronosticar los mismos para futuros años, dado un historial en periodos mensuales. Esto ayudará a tomar decisiones en cuándo y cómo actuar ante resultados que se encuentren a futuro.

CAPÍTULO II

2. MARCO TEÓRICO

2.1. Antecedentes de investigación

A nivel mundial el problema de la delincuencia se ha convertido en una constante que afecta a nuestra sociedad. Según datos aportados por la Organización Mundial de la Salud (OMS), es un problema de salud pública, 500.000 víctimas se ven afectadas en el mundo por este problema, los brotes de inseguridad no han cesado en poblaciones de: Europa, África, Asia central y América, los constantes secuestros, robos y atracos violan los derechos humanos. En consecuencia, los esquemas tradicionales de protección consignados en el derecho internacional se ponen cada vez más en tela de juicio, lo que revela la necesidad de desarrollar nuevas estrategias para prevenir la violencia desatada por la delincuencia (Gavilanes, 2018).

Según Rodrigo Toledo Astudillo en su tesis "MÉTODOS ECONOMÉTRICOS PARA EL PRONÓSTICO DE DELITOS EN EL GRAN SANTIAGO" pretende probar cuan efectivos pueden ser los modelos multiecuacionales del tipo vectores autorregresivos (VAR) y los modelos de sistemas de ecuaciones, para desarrollar pronósticos a corto plazo de delitos, aplicado a la comuna de Santiago en el período que comprende el primero de enero de 2001 y el 30 de junio de 2004. Se encontró que los modelos apropiados para la formulación de pronósticos difieren, dependiendo del sector que se esté tratando (Toledo, 2005).

Según Carlos Villalva Basantes en su tesis "PREDICCIÓN DEL NÚMERO DE DELITOS DE TRÁNSITO QUE INGRESARÁN A LA FISCALIA PROVINCIAL DE CHIMBORAZO MEDIANTE ARIMA 2018-2021" predice el número de delitos de tránsito que ingresarán a la Fiscalía Provincial de Chimborazo (FPCH), mediante modelos ARIMA estacionales (SARIMA) para que con este conocimiento se pueda tomar decisiones más acertadas, mejorar la atención y reparto de bienes en la FPCH (Villalba, 2021).

2.2. Referencias Teóricas

2.2.1. Análisis exploratorio de datos

"Antes de iniciar el análisis estadístico, es conveniente realizar una fase previa, encaminada a que el analista vaya tomando contacto con los datos que va a analizar y se familiarice con la naturaleza de los mismos. Este estudio previo se denomina análisis exploratorio de datos, y se realiza sin ninguna hipótesis a priori, utilizando técnicas muy sencillas, donde abundan las representaciones graficas. Es en esta fase donde se empezarán a revelar las relaciones más evidentes existentes entre las variables que posteriormente se estudiarán con el rigor correspondiente" (Rojo, 2006, p. 31).

El AED (Análisis exploratorio de datos) proporciona métodos sencillos para organizar y preparar los datos, detectar fallos en el diseño y recogida de datos, tratamiento y evaluación de datos ausentes, identificación de casos atípicos y comprobación de los supuestos subyacentes en la mayor parte de las técnicas multivariantes (Salvador & Gargallo, 2003).

2.2.1.1. Etapas del A.E.D. (análisis exploratorio de datos)

Para realizar un A.E.D. conviene seguir las siguientes etapas:

1. Preparar los datos para hacerlos accesibles a cualquier técnica estadística.
2. Realizar un examen gráfico de la naturaleza de las variables individuales a analizar y un análisis descriptivo numérico que permita cuantificar algunos aspectos gráficos de los datos.
3. Realizar un examen gráfico de las relaciones entre las variables analizadas y un análisis descriptivo numérico que cuantifique el grado de interrelación existente entre ellas.
4. Evaluar, si fuera necesario, algunos supuestos básicos subyacentes a muchas técnicas estadísticas como, por ejemplo, la normalidad, linealidad y homocedasticidad.
5. Identificar los posibles casos atípicos (outliers) y evaluar el impacto potencial que puedan ejercer en análisis estadísticos posteriores.
6. Evaluar, si fuera necesario, el impacto potencial que pueden tener los datos ausentes (NA) sobre la representatividad de los datos analizados (Salvador & Gargallo, 2003).

2.2.2. Entorno R

R es un programa de última generación para realizar análisis de datos, siendo también un lenguaje de programación, lo cual lo hace muy versátil.
Como lenguaje de programación, R es un dialecto de un lenguaje de programación denominado S, Dentro de los lenguajes de programación se puede clasificar como un lenguaje orientado a objetos de tipo interpretado. Lo que lo hace flexible, potente y posee un tiempo de aprendizaje corto. Actualmente se encuentran disponibles más de 15303 paquetes desarrollados en R, que cubren multitud de campos desde aplicaciones Bayesianas, financieras, traficación de mapas, wavelets, análisis de datos espaciales, etc.

2.2.2.1. Características de R

- R proporciona muchísimas herramientas estadísticas para el análisis de datos: modelos lineales y no lineales para regresión, test estadísticos, análisis de series temporales, algoritmos de clasificación y agrupamiento, graficas, etc.
- Como es un lenguaje de programación, permite que los usuarios lo extiendan definiendo sus propias funciones. Gran parte de las funciones de R están escritas en el mismo R,
- R también puede usarse como herramienta de cálculo numérico, donde puede ser tan eficaz como otras herramientas específicas tales como GNU Octave y su equivalente comercial, MATLAB (Jiménez, 2019, p. 4).
- Es libre.

2.2.3. Serie temporal

Una serie temporal (o simplemente una serie) es una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o sobre varias características (serie multivariante o vectorial) de una unidad observable en diferentes momentos (Mauricio, 2007, p. 1).

2.2.4. Componentes de una serie temporal

Una serie de tiempo por lo general consta de tres componentes, se basa en el comportamiento que pueden tomar los datos u observaciones de una serie para que se puedan tomar algunas decisiones. Estos componentes son:

- **Tendencia (T):** Representa la evolución o comportamiento que tiende a tomar la serie a largo plazo.
- **Estacional (E):** Refleja las oscilaciones que son producidas de repeticiones inferiores o iguales a un año, generalmente en observaciones mensuales.
- **Cíclica (C):** Enfoca el comportamiento de oscilaciones de forma periódica de amplitud superior a un año.
- **Aleatorio o irregular (I):** Comportamiento de la serie el cual se visualiza fluctuaciones generadas de eventos imprevisibles (García, 2016, p. 17).

Los componentes tendencia y estacional son deterministas mientras que el último componente es aleatorio. Los componentes de series temporales se denotar como:

$$X_t = T_t + E_t + C_t + I_t \quad 1\text{-}2$$

De modo que T_t es el componente tendencia, E_t es el componente estacional, el tercero es el componente cíclico y por último I_t como el componente aleatorio o irregular.

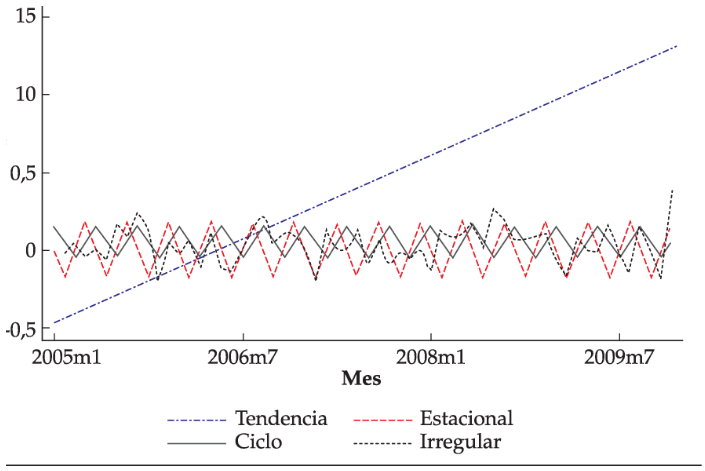

Ilustración 1-2: Componentes de una serie temporal

Fuente: Fundamentos de econometría intermedia: teoría y aplicaciones *(Rosales, et al., 2013, p. 176).*

Mediante la evolución de manera conjunta de los componentes antes mencionados y su interacción considerando un modelo establecido da como resultado a los valores de la serie de tiempo, es decir, $X_t = f(T_t, E_t, C_t, I_t)$. De manera general, se acepta la combinación de los componentes de la serie temporal de acuerdo con las tres siguientes formas:

Modelo aditivo: $X_t = T_t + E_t + C_t + I_t + \varepsilon_t$,

Modelo multiplicativo: $X_t = T_t \cdot E_t \cdot C_t \cdot I_t \cdot \varepsilon_t$,

Modelo mixto: $X_t = T_t \cdot E_t \cdot C_t \cdot I_t + \varepsilon_t$

2.2.5. *Clasificación descriptiva de las series temporales*

Las series temporales tienden a clasificarse de acuerdo con el comportamiento siguiente:

- **Estacionaria:** Es cuando la serie de tiempo definida como X_t, no presenta tendencia y la variabilidad permanece estable en el tiempo, estocásticamente los datos son independientes e idénticamente distribuidos con media cero y varianza constante.
- **No estacionaria:** Son series no estacionarias las que tienden a cambiar con el tiempo, es decir, la media o variabilidad determinan un comportamiento a crecer o decrecer a lo largo del tiempo, por lo tanto, la serie no es constante.

2.2.6. *Funciones de Autocorrelación*

Las funciones de autocorrelación simple y parcial son dos medidas características de la estructura de dependencia temporal que rigen el comportamiento del proceso estocástico generador de la serie temporal (Cáceres, et al., 2007, p. 107).

2.2.6.1. Función de Autocorrelación Simple (ACF)

Mide la relación lineal entre las observaciones de una serie de dato X_t, distanciados en un lapso de tiempo k. El lapso de tiempo k se le conoce como retardo o retraso. Este retardo denota el periodo de tiempo entre los valores de la serie, para el cual se mide el tipo y grado de correlación de la variable considerada (Melo & Santana, 2016, p. 25).

$$p_j = corr\left(X_j, X_{j-k}\right) = \frac{cov(X_j, X_{j-k})}{\sqrt{V(X_j)}\sqrt{V(X_{j-k})}} \qquad 2\text{-}2$$

La ACF simple está constituida de las siguientes propiedades

- $p_o = 1$
- $-1 \leq p_j \leq 1$
- $Simetria\, p_j = p_{-j}$

A través del siguiente gráfico se muestra cómo se representa la función de autocorrelación simple en función al análisis exploratorio de la serie temporal.

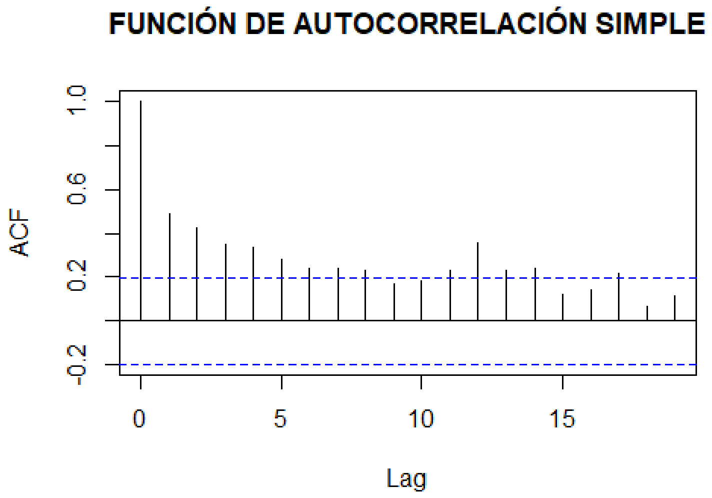

Ilustración 2-2: Función de autocorrelación simple

Realizado por: López, J.; Morocho, M. 2022.

2.2.6.2. Función Autocorrelación Parcial (PACF)

La función de autocorrelación parcial (fap o PACF) se define como la correlación lineal existente entre X_t y X_{t-k}, pero descontando del efecto que tienen los retardos intermedios sobre ellas (García, 2016, p. 67).

$$\pi_j = \frac{cov(X_j - \hat{X}_j, X_{j-k} - \hat{X}_{j-k})}{\sqrt{V(X_j - \hat{X}_j)}\sqrt{V(X_{j-k} - \hat{X}_{j-k})}} \quad 3\text{-}2$$

Para una mejor interpretación a través del siguiente gráfico se representa la función de autocorrelación parcial.

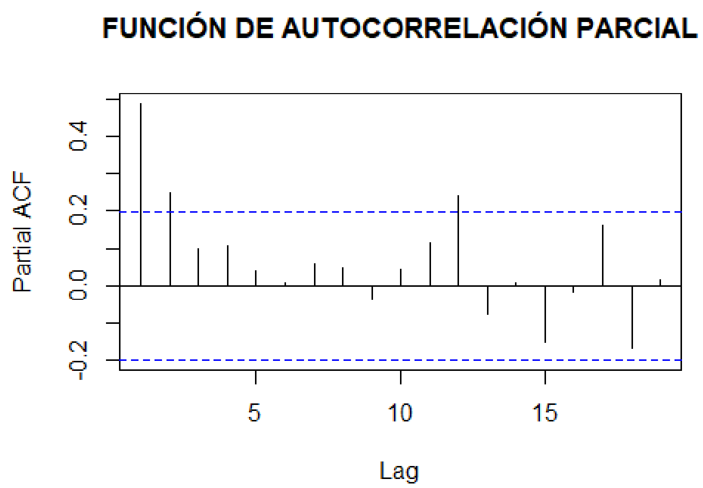

Ilustración 3-2: Función de autocorrelación parcial

Realizado por: López, J.; Morocho, M. 2022.

2.2.7. Procesos Lineales Estacionarios

2.2.7.1. Procesos autorregresivos AR(p)

Los modelos autorregresivos se fundamentan en el concepto en donde el valor actual de la serie definida X_t, puede ser explicada en función de p valores pasados $X_{t-1}, X_{t-2}, \dots, X_{t-p}$, en el cual p asigna el número de rezagos necesarios para poder predecir el valor actual (Villavicencio, 2010, p. 5).

El proceso autorregresivo AR(p) se denota mediante la fórmula:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_P X_{t-P} + \varepsilon_t \qquad \text{4-2}$$

Donde

$\phi_1, \dots, \phi_p$: son parámetros desconocidos,

ε_t: proceso ruido blanco.

2.2.7.2. Procesos de media móvil MA(q)

Una familia de procesos que tienen memoria muy corta, son la de media móvil, o procesos MA. Estos procesos son función de un número finito, y generalmente pequeño de las innovaciones pasadas (Capel, 2016, p. 18).

Los procesos MA(q) son considerados aquellos donde el valor actual depende de las q últimas observaciones o innovaciones (los ε_t). Se denota mediante su fórmula general dada la serie X_t es:

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \qquad \text{5-2}$$

Donde:

$\theta_1, \dots, \theta_q$: son parámetros desconocidos,

ε_t: proceso ruido blanco.

2.2.7.3. Procesos Autorregresivos de Medias Móviles ARMA (p,q)

El proceso autorregresivo de Medias Móviles o ARMA(p,q) es un proceso mixto en el que se combinan los modelos AR(p) y MA(q) vistos anteriormente (Bonilla, 2019, p. 8).

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_P X_{t-P} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \qquad \text{6-2}$$

Donde:
$\phi_1, \ldots, \phi_p$ y $\theta_1, \ldots, \theta_q$: son parámetros desconocidos,
ε_t: proceso ruido blanco.

Es complicado en la práctica identificar los procesos ARMA, por ende, es aconsejable buscar modelos MA o AR de bajo orden que asuman los rasgos más claros. A continuación, se explica algunas características para su correcta interpretación:

Tabla 1-2: Resumen de la ACF y PACF de procesos ARMA

	ACF	**PACF**
AR (p)	Muchos coeficientes no nulos	Primeros p no nulos, resto 0
MA (q)	Primeros q no nulos, resto 0	Muchos coeficientes no nulos
ARMA (p,q)	Muchos coeficientes no nulos	Muchos coeficientes no nulos

Fuente: Análisis de series temporales (Peña, 2005, p. 158).

2.2.7.4. Modelo Autorregresivo Integrado de Medias Móviles ARIMA(p,d,q)

El modelo autorregresivo integrado y de medias móviles ARIMA se define como una extensión de modelo visto anteriormente ARMA aplicada a series temporales no estacionales (el tipo de serie de tiempo que integra una o más raíces unitarias), es decir, se considera un proceso ARMA con la serie temporal diferenciada (López & Martínez, 2013, p. 55).

$$\Phi_P(L)\Delta^d X_t = \delta + \Theta_q(L)\varepsilon_t \qquad 7\text{-}2$$

Donde:
ε_t: proceso de ruido blanco,
$\Phi_p(B^s)$: rezago del componente autorregresivo AR(p),
$\Theta_q(L^s)$: rezago del componente de medias móviles MA(q),
Δ^d: componente diferencial regular.

2.2.7.5. Modelos Puramente Estacionales SARIMA(p,d,q)(P,D,Q)

El modelo SARIMA viene siendo una extensión del modelo ARIMA en donde la serie temporal integra un componente estacional, en donde parte del proceso multiplicativo de dos procesos ARMA de la serie temporal diferenciada en base en el análisis tanto de la parte regular y estacional, generalmente descrito por SARIMA(p,d,q)(P,D,Q)s (López & Martínez, 2013, p. 55).

$$\Phi_P(L^s)\phi_p(L)\Delta_s^D\Delta^d X_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t \quad 8\text{-}2$$

Donde

s: componente estacional,

ε_t: proceso de ruido blanco,

$\phi_p(L)$: rezago del componente autorregresivo AR(p),

$\theta_q(L)$: rezago del componente de medias móviles MA(q),

$\Phi_P(B^s)$: rezago del componente estacional autorregresivo AR(P),

$\Theta_Q(L^s)$: rezago del componente estacional de medias móviles MA(Q),

Δ^d: componente diferencial regular,

Δ_s^D: componente diferencial estacional.

2.2.8. Metodología Box - Jenkins

La metodología de Box-Jenkins hace referencia a una serie de procedimientos con el propósito de identificar, estimar, diagnosticar y pronosticar mediante modelos ARIMA con un formato de datos de series de tiempo, con el fin de realizar las predicciones apropiadas.

2.2.8.1. ¿En qué se basa la Metodología Box - Jenkins?

Este método predictivo se fundamenta en el análisis de las propiedades probabilísticas o estocásticas de las series temporales.

Dado que una variable X_t que determina la serie temporal se expresa como una función de los valores pasados.

2.2.8.2. Objetivo de la Metodología Box - Jenkins

"El objetivo de la metodología Box-Jenkins es identificar, estimar un modelo estadístico que pueda ser interpretado como generador de la información muestral, si este modelo va a ser usado para predicción, se debe suponer que sus características son constantes a través del tiempo, y particularmente, en periodos de tiempo futuro. Así, la simple razón para requerir información estacionaria es que cualquier modelo que sea inferido a partir de esta información pueda ser interpretado como estacionario o estable, proporcionando, por consiguiente, una base válida para predicción" (De la Oliva de Con, et al., 2016).

2.2.8.3. Etapas de la metodología Box - Jenkins

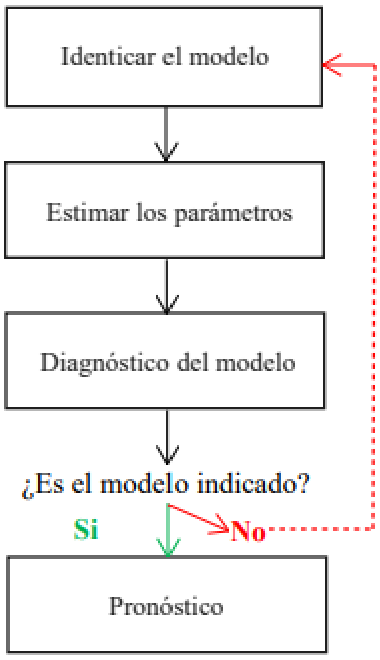

Ilustración 4-2: Etapas de la metodología Box - Jenkins

Realizado por: López, J.; Morocho, M. 2022.

La metodología de Box-Jenkins sigue un procedimiento de cuatro etapas:

1. **Identificar el modelo:** Tiene como objetivo identificar el modelo ARIMA adecuado que se ajusta a la serie temporal, el cual cumple los requisitos:
    - Convertir de la serie de tiempo a estacionaria.
    - Determinar los componentes (p,d,q).

2. **Estimar los parámetros:** Se estiman los parámetros AR y MA de orden p y q respectivamente del modelo seleccionado.

3. **Diagnóstico del modelo:** En esta etapa se corrobora que los residuos cumplan con un proceso de ruido blanco, es decir, que sigan una distribución en media cero y varianza constante, cumpliendo con los supuestos de normalidad, independencia y homocedasticidad.

4. **Pronóstico:** Determinado el mejor modelo y cumpliendo con todos los requisitos y supuestos se realiza el pronóstico con el modelo seleccionado (De la Fuente Fenández, 2013, p. 3).

2.2.9. *Supuestos de hipótesis*

2.2.9.1. Ruido blanco

Para que el modelo seleccionado de una serie temporal sea estacionario debe cumplir con el proceso de ruido blanco, esto quiere decir que estocásticamente los datos se distribuyen en media cero y varianza constante en el transcurso de la serie, este proceso se denota a través de la siguiente expresión:

$$H_0: Existe\ ruido\ blanco$$
$$H_1: No\ existe\ ruido\ blanco$$

$$\varepsilon_t \sim N(0, \sigma^2) \qquad cov\left(\varepsilon_{t_i}, \varepsilon_{t_j}\right) = 0 \qquad \forall t_i \neq t_j \tag{9-2}$$

Donde:

ε_t: es un proceso de ruido blanco.

2.2.9.2. Prueba de raíz unitaria KPSS

Este contraste se implementa para determinar si la serie temporal en estudio es estacionaria o no, por ende, si la serie integra una o más raíces unitarias se procede a realizar la diferencia respectiva a la serie de tiempo con el fin de establecer las hipótesis que se plantean a continuación:

$$H_0: La\ serie\ es\ no\ estacionarios: Tiene\ raíz\ unitaria$$
$$H_1: La\ serie\ es\ estacionarios: No\ tiene\ raíz\ unitaria$$

La prueba se denota matemáticamente por:

$$X_t = \xi_t + r_t + \varepsilon_t \tag{10-2}$$

Donde:

ξ_t: tendencia determinística,

r_t: caminata aleatoria,

ε_t: proceso de ruido blanco o error.

2.2.9.3. *Normalidad*

Uno de los requisitos para la selección de modelo es si cumple con el supuesto de normalidad de los errores, es decir, si los datos siguen o no una distribución normal, para ello habitualmente se utiliza la prueba de Jarque Bera que se fundamente de acuerdo con el coeficiente de simetría y curtosis, en donde se denota de la siguiente manera:

$$H_0: Los\ errores\ siguen\ una\ distribución\ normal$$

$$H_1: Los\ errores\ no\ siguen\ una\ distribución\ normal$$

$$JB = n[\frac{A^2}{6} + \frac{(K - 3)^2}{24}] \quad 11\text{-}2$$

Donde:

A: es el coeficiente de asimetría.

K: es el coeficiente de kurtosis.

2.2.9.4. *Independencia*

Un supuesto que se debe cumplir en el modelo a elegir es la independencia de los residuos, ya que la ausencia aleatoria en la serie no permite corregir fácilmente e influye de manera notable en el análisis estadístico, Se emplea para esta hipótesis el contraste de Ljung-Box en los modelos ARIMA, denotado matemáticamente a continuación:

$$H_0: Existe\ independencia\ en\ los\ residuos$$

$$H_1: No\ existe\ independencia\ en\ los\ residuos$$

$$LB = n(n + 2)\sum_{k=1}^{m}(\frac{\hat{P}_k^2}{n - k}) \sim X_{(m)}^2 \quad 12\text{-}2$$

Donde:

n: tamaño muestral,

$\hat{P}_k^2$: autocorrelación muestral en el retraso k,

k: número de retrasos existentes.

2.2.9.5. *Homocedasticidad*

Consiste es verificar si la variabilidad es constante en la serie temporal, por otra parte, si la variabilidad cambia de un periodo a otro de forma creciente o decreciente la serie vendría a ser heterocedástica, por lo que le modelo se descartaría por no cumplir homocedasticidad, para verificar este supuesto se utiliza la prueba de hipótesis de White el cual se define de la siguiente manera (Villalba, 2021, pp. 11-12):

$$H_0: Existe\ homocedasticidad$$
$$H_1: No\ existe\ homocedasticida$$

$$\sigma_t^2 = h(X_t')\alpha_0 + \alpha_1 X_{1t} + \cdots + \alpha_p X_{pt} + \varepsilon_t \quad 13\text{-}2$$

Regla de decisión

La toma de decisión de las hipótesis antes emocionado se basa a partir de estos criterios:
Si $p - valor \leq \alpha$ Se rechaza H_0.
Si $p - valor > \alpha$ No se rechaza H_0.
Donde:
α: es el nivel de significancia.

2.2.10. Criterios de información

Para determinar el mejor modelo se necesita comparar el valor critico de todos los modelos considerados candidatos. Para lo cual se evalúan dichos modelos para posterior seleccionar el modelo óptimo. En el cual se considera al mejor modelo el que contenga el menor valor del criterio, confirmando este el más adecuado conforme a los datos. Los criterios de información a implementar se detallan a continuación.

2.2.10.1. Criterio de información de Akaike (AIC)

El criterio de información Akaike mide la bondad de ajuste de determinado modelo definido matemáticamente de la siguiente forma:

$$AIC = -2\, x\, ln(L) + 2k \quad 14\text{-}2$$

Donde:
k: es el número de parámetros del modelo,
ln: es la función de verosimilitud,
L: valor máximo para la función de verosimilitud.

2.2.10.2. Criterio de información Bayesiano (BIC)

El criterio de información bayesiano es la alternativa al clásico criterio de evaluación de modelos AIC. Matemática se define cómo:

$$BIC = -2\,log(L) + k\,log(T) \qquad 15\text{-}2$$

Donde:
k: es el número de parámetros del modelo,
T: cantidad de datos disponibles.

Una de las diferencias fundamentales entre los criterios de evaluación AIC y BIC es que este último penaliza los modelos con mayor número de parámetros estimados obteniéndose modelos de menor orden y por ende más parsimoniosos (Osorio & Ángel, 2019, pp. 22-23).

2.2.11. Fiscalía General del Estado

La Fiscalía General del Estado es una entidad de administración autónoma e independiente, destinada a la inspección de procesos penales, es decir, delitos con objeto de fiscalización, manteniendo la justicia como prioridad en función judicial y sus principios de acuerdo a la oportunidad y la participación mínima penal, velando por la seguridad, bienestar y los derechos humanos de las partes involucradas ya sea esta en función de víctima, victimario, demandado o demandante, con el fin de llegar a un término donde la justicia actúa basado a los principios constitucionales y garantizando el respectivo proceso judicial.

2.2.12. Definiciones básicas

Victima: Persona física que haya sufrido daños, incluidos los daños corporales o de salud o los daños morales o patrimoniales, o la privación de libertad, que hayan sido causados puntualmente por un delito penal.

Victimario: Se considera aquella persona que inflige, realiza o causa un daño o perjuicio a otro individuo ya sea de forma física, psicológica o moral.

Delito: Es considerado todo aquello que ya sea por acto voluntario u omisión el legislador la considera como una conducta plenamente relevante merecedora de una pena.

El Código Orgánico Integral Penal, articula al delito como fracción penal, de tal modo que se considera una conducta atípica, antijurídica y culpable, cuyos actos son sancionados y determinados en la ley penal (Carrión, 2018).

Tipos de delito

En base al informe general presentado por el Instituto Nacional de Estadística y Censos (INEC), la Comisión Especial de Estadísticas de Seguridad, Justicia, Crimen y Transparencia del Ecuador enfocado en los delitos de los años 2020 y 2021 se presentan a continuación:

Tabla 2-2: Principales delitos del Ecuador, años 2020-2021

Descripción Actos Delictivos	CASOS		
	Años		TOTAL
	2020	2021	
Homicidios / Asesinatos	1,372	2,492	3,864
Femicidios	75	68	143
Robo a personas	20,126	25,33	45,456
Robo a domicilios	7,369	8,153	15,522
Robo a unidades económicas	4,078	4,782	8,860
Robo de motos	6,666	9,006	15,672
Robo de carros	4,596	6,866	11,462
Robo de bienes, accesorios y autopartes	6,214	7,932	14,146
Violaciones	7,886	5,613	13,499
Fallecidos por siniestros de tránsito	1,591	2,131	3,722

Fuente: Instituto Nacional de Estadística y Censos (INEC).

Realizado por: López, J.; Morocho, M. 2022.

De acuerdo con los resultados proporcionados por el INEC, el robo a personas es el delito que más se ha registrado en los años 2020 y 2021 con un total de 45.456 casos, seguido de robo de motos con 15.672, robo a domicilios con 15.522, robo de bienes, accesorios y autopartes con 14.146, violaciones con 13.499, robo de carros con 11.462, robo a unidades económicas con 8.860, homicidios o asesinatos con 3.864, fallecidos mediante un siniestro de tránsito con 3.722 y femicidios con un total 143 casos; el estudio se realizó en base a los casos de ingresados a fiscalización por la entidad de Fiscalía General y la Policía Nacional del Ecuador.

CAPÍTULO III

3. MARCO METODOLÓGICO

3.1. Enfoque de investigación

La investigación ejecutada tiene un enfoque cuantitativo en consideración de que se hará un análisis estadístico utilizando datos de series de tiempo, en donde se aplicará la metodología Box-Jenkins para modelar y mediante ello realizar un pronóstico sobre los delitos en la provincia de Chimborazo con el propósito de observar sus cambios a través del tiempo y realizar un diagnóstico general.

3.2. Nivel de Investigación

Según el nivel de profundización en el objeto de estudio es predictivo, dado que se desea emplear un análisis a la cantidad de delitos en general que son objetos de fiscalización en la provincia de Chimborazo, empleando la metodología Box-Jenkins con fin de identificar los componentes de serie, es decir, el comportamiento que tiene la serie a través del tiempo y en base a ello verificar si la serie es estacionaria o no, para lo cual se estimara mediante modelos Arima en el cual se verificará que cumpla todos los supuestos como normalidad, independencia y homocedasticidad y si la serie es no estacionaria aplicar las diferencias si es necesario para que los modelos cumplan los requisitos y mediante los criterios de información AIC y BIC elegir el mejor modelo y conforme a ello realizar pronósticos.

3.3. Diseño de investigación

3.3.1. Según la manipulación o no de la variable independiente

Según la manipulación de variables es no experimental debido a que la variable en estudio es de un registro real de los delitos en la provincia de Chimborazo en el cual no existe control alguno, dado que la información es otorgada de una fuente secundaria, es decir, el conjunto de datos fue proporcionado por la entidad en estudio.

3.3.2. Según las intervenciones en el trabajo de campo

Según el periodo temporal es longitudinal ya que los datos proporcionados detallan los registros mensuales que concierne a los años entre enero de 2015 hasta abril de 2022 del número de actos delictivos y mediante un análisis de series de tiempo utilizando la metodología de Box-Jenkins se desea hacer un pronóstico referente a los próximos 12 meses.

3.4. Tipo de estudio

Según el tipo de estudio se utilizará fuentes de información de campo debido a que los datos son proporcionados directamente de la Fiscalía de Chimborazo que son registrados por medio del Servicio de Atención Integral (SAI) a escala provincial. Aquí se receptan las denuncias de los delitos penales, para lo cual se va a realizar un análisis de series de tiempo empleando la metodología de Box-Jenkins, de tal manera que se pueda emplear un pronóstico y de acuerdo con ello realizar un diagnóstico general.

3.5. Población y Planificación, selección y cálculo del tamaño de la muestra

3.5.1. Localización de estudio

El presente proyecto de investigación se ejecutará en la provincia de Chimborazo, en donde se obtuvo la información mediante la entidad de la Fiscalía Provincial de Chimborazo

Ilustración 5-3: Provincia de Chimborazo

Realizado por: López, J.; Morocho, M. 2022.

3.5.2. Población de estudio

La población en estudio está conformada por los delitos registrados mensualmente por la Fiscalía Provincial de Chimborazo.

3.5.3. Tamaño de muestra

Se establece el tamaño de la muestra de 88 observaciones proporcionados por la Fiscalía Provincial de Chimborazo en base a registros de los delitos mensuales que concierne a los periodos de enero 2015 - abril 2022.

3.6. Métodos, técnicas e instrumentos de investigación

3.6.1. Métodos de investigación

3.6.1.1. Método de muestreo

Los datos son proporcionados de manera directa por la Fiscalía Provincial de Chimborazo por medio de registros.

3.6.1.2. Método inductivo

Según el tipo de inferencia es inductivo, dado que se desea conocer un aproximado de delitos en los meses futuros de acuerdo con el año en estudio, en donde se busca obtener resultados que contraste la realidad de esta problemática social.

3.6.1.3. Método analítico

Se busca emplear un análisis de series de tiempo para conocer sus principales características y en base a ello emplear los procedimientos necesarios según los requerimientos para el modelado estadístico – informático con R utilizando la metodología de Box-Jenkins, con el objetivo de establecer un pronóstico que refleje la realidad en cuanto al número de actos delictivos.

3.6.2. Técnicas de investigación

3.6.2.1. Técnica de recolección de datos

La obtención de los datos se sostuvo en atención al requerimiento de información estadística realizado por medio del Coordinador de la Carrera de Estadística de la Escuela Superior Politécnica de Chimborazo, por medio de un documento adjunto, se remite la información concerniente, relacionado con los datos estadísticos de recepción de noticias de delito, en todas las unidades de la Fiscalía Provincial de Chimborazo, durante el periodo 2015-Abril 2022 **(ANEXO A)**.

3.6.2.2. Modelo estadístico

Obtención y depurado de la información: La información que fue depurada y recolectada de acuerdo con lo solicitado en base a los delitos que concierne desde enero de 2015 hasta abril 2022, fue directamente proporcionada por la Fiscalía Provincial de Chimborazo, en el cual se nos facilitó con una base de datos de los números de actos delictivos y con ello se consolida la información necesaria para el estudio y análisis del proyecto de investigación a tratar.

Análisis exploratorio univariante: Con la obtención depurada de la información a través de una base de datos procederemos a identificar las características principales de la serie temporal, para lo cual emplearemos un análisis descriptivo y un análisis para identificar datos atípicos.

Aplicación de la metodología Box-Jenkins: Se ejecutará de manera ordenada con cada una de las etapas que presenta la metodología de Box-Jenkins con el propósito de determinar un modelo adecuado, el cual nos permitirá emplear un pronóstico y ejercer un diagnóstico.

Pronóstico: Determinado y validado el modelo adecuado procederemos a realizar el cálculo y su respectiva gráfica en función a los valores pronosticados con el fin de plantear conclusiones en relación de los valores que se tiene a futuro en la que se podrían presentar los delitos mensuales en la provincia de Chimborazo, y mediante ello podremos discutir y comparar con respecto al año pronosticado con los años que conciernen a la información que se nos otorgó por parte de FPCH y realizar conclusiones del mismo.

3.6.3. Instrumentos de investigación

3.6.3.1. Variables en estudio

Identificación de variables

Variable Dependiente:

- Delitos totales mensuales

Variable Independiente

- Tiempo: Mes

3.6.3.2. Matriz de consistencia

Tabla 3-3: Matriz de consistencia

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES
PROBLEMA GENERAL: ¿Cuántos delitos ingresarán mensualmente a la fiscalía provincial de Chimborazo, FGE? **PROBLEMAS ESPECÍFICOS**: a) ¿Cuáles son las características que presentan los datos sobre los delitos generales en la provincia de Chimborazo? b) ¿Existirá algún modelo óptimo que ayude a pronosticar los delitos en la provincia de Chimborazo? c) ¿Qué modelo se ajustará para la adecuada ejecución en la predicción de los delitos? d) ¿Cuál será el mejor criterio de información para la elección del mejor modelo? e) ¿Será el modelo planteado la mejor elección para realizar predicciones de los delitos?	**OBJETIVO GENERAL:** Predecir el número de delitos en la provincia de Chimborazo mediante modelos SARIMA utilizando la metodología de Box-Jenkins. **OBJETIVOS ESPECÍFICOS:** a) Realizar un análisis exploratorio de datos univariante de los delitos. b) Identificación del modelo adecuado. c) Estimación de parámetros del modelo adecuado. d) Comprobar o verificar con AIC o BIC los supuestos de los modelos ARIMA estacionales. e) Aplicar modelos SARIMA con el fin de realizar predicciones de los delitos.	**HIPOTESIS GENERAL:** Mediante modelización estadística informática con R podemos obtener resultados relevantes o significativos que nos acerquen a la realidad y/o predecir delitos en la provincia de Chimborazo.	**VARIABLE DEPENDIENTE:** • Delitos mensuales totales **VARIABLE INDEPENDIENTE:** • Tiempo: mes

Elaborado por: López, J.; Morocho, M. 2022.

3.6.3.3. Operacionalización de objetivos

Tabla 4-3: Operacionalización de objetivos

OBJETIVO GENERAL	OBJETIVOS ESPECIFICOS	CONCEPTO	INDICADOR	INSTRUMENTO
Predecir el número de delitos en la provincia de Chimborazo mediante modelos SARIMA utilizando la metodología de Box-Jenkins.	• Realizar un análisis exploratorio de datos univariante de los delitos.	El análisis exploratorio de datos (AED), emplea un estudio mediante el uso de gráficas con el fin de visualizar, explorar y analizar un colectivo de datos.	Análisis descriptivo aplicando histograma y diagrama de cajas.	Base de datos de delitos en Chimborazo
	• Identificación del modelo adecuado.	Identificar el modelo ARIMA adecuado que se ajusta a los datos, cumpliendo estacionariedad y validación de supuestos.	Modelos SARIMA estacionales del mejor modelo.	Base de datos de delitos en Chimborazo
	• Estimación de parámetros del modelo adecuado.	Se estiman los parámetros de orden p y q respectivamente y el componente diferencial d del modelo seleccionado.	Componentes (p,d,q) de modelos SARIMA estacionales.	Base de datos de delitos en Chimborazo
	• Comprobar o verificar con AIC o BIC los supuestos de los modelos ARIMA estacionales.	Criterios de información AIC Y BIC se emplea como medidas de bondad de ajuste aplicado a un modelo estadístico.	Mejor modelo respecto al valor más bajo del criterio de información.	Base de datos de delitos en Chimborazo
	• Aplicar modelos SARIMA con el fin de realizar predicciones de los delitos.	La metodología Box Jenkins estima modelos ARIMA y SARIMA para el pronóstico de la variable en estudio.	Modelos SARIMA estacionales y validación de supuestos al mejor modelo.	Base de datos de delitos en Chimborazo

Realizado por: López, J.; Morocho, M. 2022.

CAPÍTULO IV

4. MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

4.1. Análisis Exploratorio

En base a los datos obtenidos en relación con los delitos en la provincia de Chimborazo y por ser esta la variable como objeto de estudio, se realiza primeramente un análisis exploratorio, con el objetivo de identificar el comportamiento que tiene la serie a través del tiempo. Debido a que los datos presentan registros mensuales que conciemen a partir de enero de 2015 hasta abril de 2022 estructurado en la siguiente tabla:

Tabla 5-4: Tabla de delitos mensuales desde enero 2015 a abril 2022

Mes	2015	2016	2017	2018	2019	2020	2021	2022
Enero	714	736	738	813	713	728	527	643
Febrero	828	828	602	749	681	634	468	750
Marzo	890	765	828	809	702	398	596	753
Abril	842	721	683	748	728	114	564	412
Mayo	748	760	890	821	632	263	621	
Junio	799	746	742	829	629	410	643	
Julio	856	697	773	763	736	504	692	
Agosto	649	749	796	717	642	432	617	
Septiembre	756	630	691	635	658	518	970	
Octubre	730	618	803	675	669	549	652	
Noviembre	680	653	737	616	721	639	673	
Diciembre	831	732	794	609	681	645	647	
Total	9323	8635	9077	8784	8192	5834	7670	2558

Fuente: Fiscalía Provincial de Chimborazo.

Realizado por: López, J.; Morocho, M. 2022.

4.1.1. *Serie temporal*

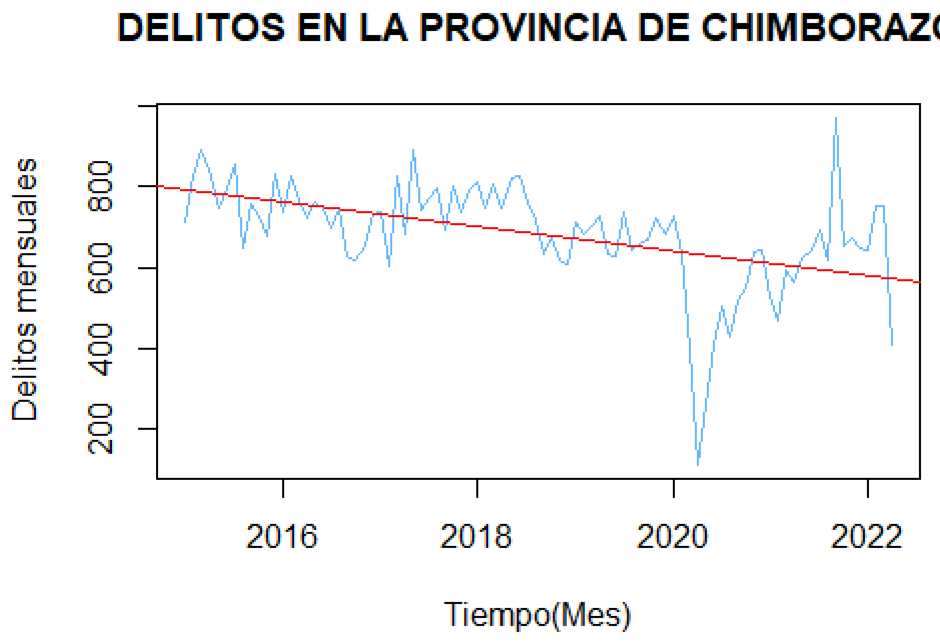

Ilustración 6-4: Serie temporal de los delitos mensuales en Chimborazo

Realizado por: López, J.; Morocho, M. 2022.

Mediante la visualización grafica de la serie temporal original podemos observar que los delitos registrados en la Provincia de Chimborazo presentan claramente una tendencia negativa, vale recalcar que en ciertos intervalos de tiempo se observa como tienden a crecer o decrecer los delitos mensuales, también es notorio que al parecer presenta estacionalidad, es decir, tiene una estructura repetitiva año tras año, esta se puede corroborar de una manera más efectiva mediante la descomposición de la serie, finalmente se evidencia que existe homocedasticidad, dado que su variabilidad se muestra constante a lo largo del tiempo, de acuerdo a lo mencionado, podemos identificar que el tipo de descomposición a emplear es aditiva.

4.1.2. Análisis descriptivo de la serie temporal

Tabla 6-4: Información estadística descriptiva

Análisis Descriptivo	
Mínimo	114.0
Primer Cuartil	633.5
Mediana	699.5
Media	682.6
Tercer Cuartil	753.8
Máximo	970.0
Desviación Estándar	133.3
Asimetría	-1.357375
Curtosis	3.435107

Realizado por: Morocho, M y López, J, 2022.

Conforme al análisis de tendencia central, de dispersión, y de forma podemos concluir que en promedio los delitos registrados en el lapso de tiempo de estudio de los datos son de 682.6, con un máximo de delitos de 970 que concierne al mes de septiembre de 2021, siendo el pico más alto de actos delictivos ocurridos, en cuanto a la variabilidad es de 133.3 que nos da a conocer la baja variabilidad de datos dispersos con respecto a la media, y la vez se refleja una asimetría negativa.

4.1.3. Representación gráfica para el estudio e identificación de la serie temporal

4.1.3.1. Histograma

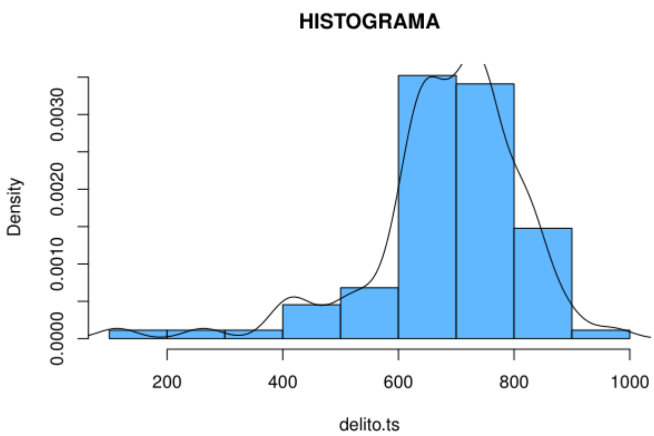

Ilustración 7-4: Histograma de los delitos mensuales

Realizado por: López, J.; Morocho, M. 2022.

Respecto al histograma se confirma la presencia de la asimetría negativa tal como se indicó en el análisis descriptivo en la tabla 6.4 en donde se refleja un valor de -1.36, es decir, la distribución en cuanto a la agrupación de los datos es asimétrica a la izquierda, por ende, se establece que se los delitos en Chimborazo correspondiente al periodo enero 2015 - abril 2022 representan la cantidad de delitos inferiores en relación con la mediana en lapso del intervalo de tiempo de estudio.

4.1.3.2. Datos atípicos

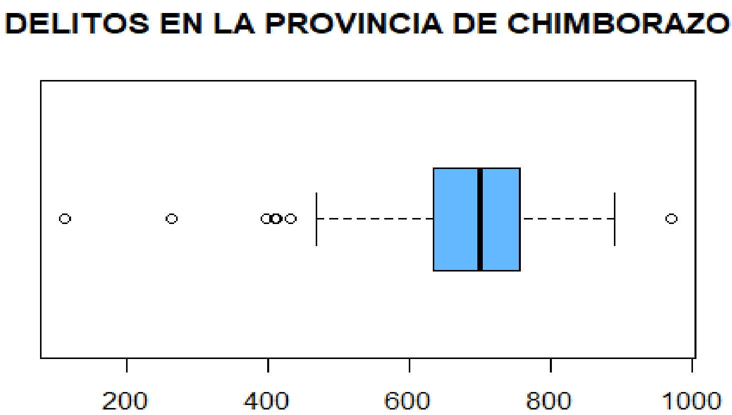

Ilustración 8-4: Boxplot de delitos mensuales de Chimborazo

Realizado por: López, J.; Morocho, M. 2022.

A través del diagrama de cajas se puede observar la presencia de siete datos anómalos o atípicos, los mismos que pueden ser causantes problemáticos al instante de estimar un modelo para la adecuada implementación del pronóstico.

Los datos atípicos identificados se detallan a continuación:

Tabla 7-4: Datos atípicos de delitos en Chimborazo

Fecha	Dato Atípico
Marzo 2020	398
Abril 2020	114
Mayo 2020	263
Junio 2020	410
Agosto 2020	432
Febrero 2021	468
Septiembre 2021	970
Abril 2022	412

Realizado por: López, J.; Morocho, M. 2022.

Con la identificación de los datos atípicos se establece la solución de los mismos, para lo cual se emplea la sustitución por la media o promedio total de los delitos establecidos en los meses respectivos en donde existe la presencia de estos valores atípicos detallados en la **Tabla 8-4**, con el fin de que la serie temporal sea la más adecuada para emplear la metodología de Box-Jenkins, debido a que la presencia de datos atípicos pueden traer problemas en la hora de establecer el modelado estadístico. Con el respectivo reemplazo de los datos atípicos con el promedio correspondiente y en función a ello se establece los siguientes resultados:

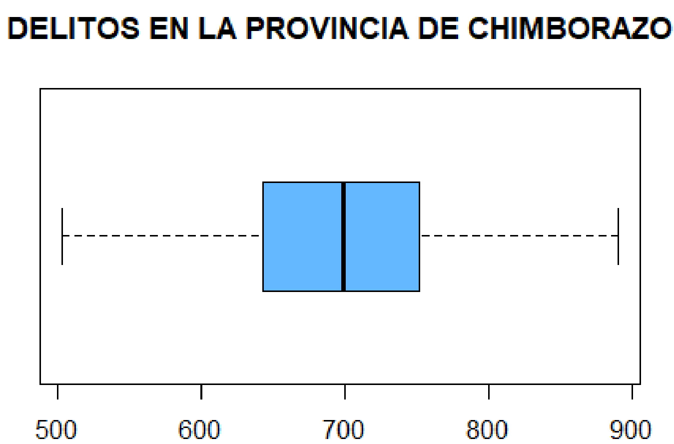

Ilustración 9-4: Boxplot sin datos atípicos

Realizado por: López, J.; Morocho, M. 2022.

Como se puede observar en el gráfico de Boxplot ya no existe la presencia de datos atípicos, además se visualiza una asimetría positiva, al parecer los datos son casi simétricos, dado que el bigote al lado izquierdo es menor en comparación con el bigote de lado derecho, respecto a estos resultados nos hace entender que existe poca variación en relación con la baja y alta cantidad de actos delictivos mensuales en la provincia de Chimborazo.

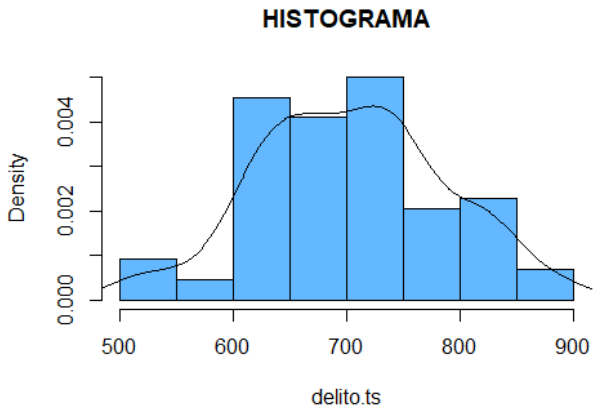

Ilustración 10-4: Histograma sin datos atípicos

Realizado por: López, J.; Morocho, M. 2022.

Mediante el gráfico de histograma sin la presencia de datos atípicos se puede observar la tendencia central alrededor de los 700 delitos con la distribución de los datos que siguen una ley normal ya que forma a una campana de Gauss, en cuanto a la agrupación de los datos se aprecia que existe poca variabilidad y se encuentran centrados respecto a la mediana.

4.1.4. Serie temporal sin presencia de datos atípicos

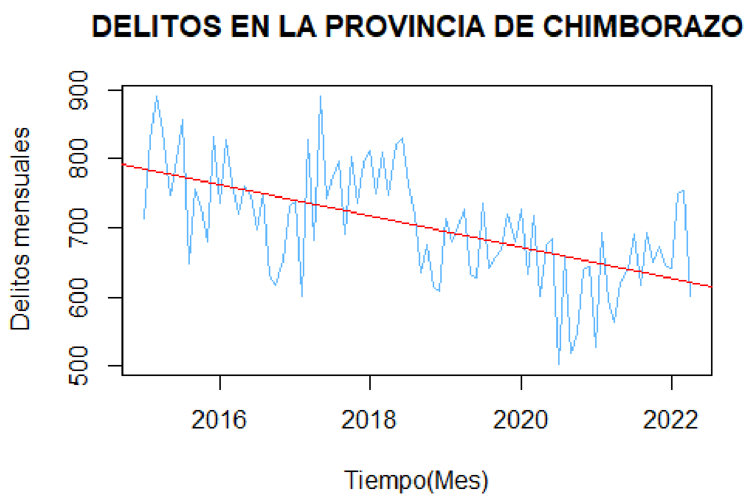

Ilustración 11-4: Serie temporal sin datos atípicos

Realizado por: López, J.; Morocho, M. 2022.

Una vez eliminados los datos anómalos se observar en la gráfica de la serie la presencia de tendencia esto implica que la serie tiende a decrecer con el tiempo, por lo que no existe estacionariedad. Por otra parte, la serie es homocedástica dado que la varianza presenta variabilidad y regularidad a lo largo de la serie, esto conlleva a realizar una descomposición aditiva de la serie para una adecuada identificación de cada uno de los componentes de la serie temporal.

4.1.4.1. Descomposición de la serie temporal

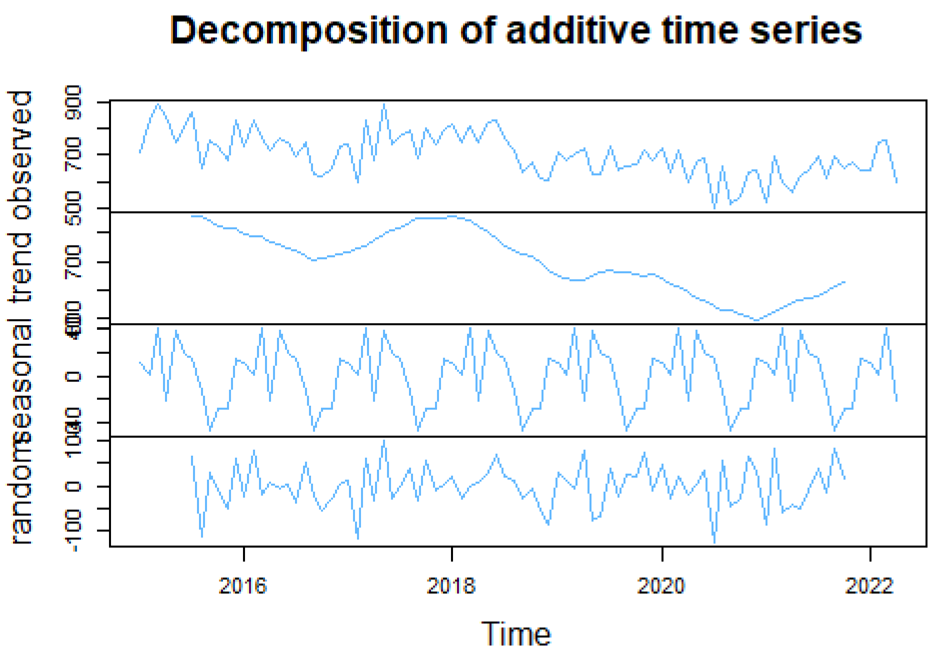

Ilustración 12-4: Descomposición aditiva de la serie temporal

Realizado por: López, J.; Morocho, M. 2022.

Podemos afirmar que la serie presenta una tendencia decreciente, esto se debe a que los actos delictivos van decreciendo progresivamente cada año. Se observa también que los datos no oscilan en relación con el valor constante, por lo que se debe aplicar diferencias para transfórmala a estacionaria y eliminar la tendencia que esta posee. A simple vista podemos observar que el componente estacional está presente en la serie temporal, es decir, la serie presenta un ciclo estacional debido a que tiene una estructura el cual se repite año tras año. Se aprecia que existe aleatoriedad en la serie, ya que los datos son irregulares o aleatorios en el tiempo, dado que hay periodos de crecimiento con otras de decrecimiento.

4.2. Aplicación de la metodología Box - Jenkins

4.2.1. Identificación del modelo

Consiste en determinar el modelo óptimo de una serie de candidatos, con respecto al orden p y q de los procesos autorregresivos y de medias móviles respectivamente, tomando en cuenta el componente regular o estacional. Técnicamente en base a las funciones de autocorrelación simple y parcial se toma una decisión. Al implementar en R podemos hacer el uso de las funciones “acf()” y “pacf()”. Para identificar el parámetro d se realiza un contraste para establecer si la serie es estacionaria o no, la prueba de Dickey-Fuller indica las diferencias a aplicar según las raíces unitarias que posee la serie, en donde se obtuvo el siguiente resultado:

$$Ho: La\ serie\ es\ no\ estacionaria$$
$$H_1: La\ serie\ es\ estacionaria$$

Tabla 9-4: Test de Dickey-Fuller

Test de Dickey Fuller	
p-valor	0.2393

Realizado por: López, J.; Morocho, M. 2022.

Se obtuvo un p-valor de 0.2393 mayor al nivel de significancia 0.05, por lo que existe suficiente evidencia para no rechazar la hipótesis nula, con esto se entiende que la serie posee raíz unitaria por ende es no estacionaria. Conforme a esto nos indica que se debe aplicar diferencias con el fin de eliminar la tendencia y transformar la serie a estacionaria.

4.2.1.1. Transformación de la serie: Aplicación de diferencias

Con respecto a las diferencias necesarias a implementar, en R mediante la función “ndiffs()” se responde esta cuestión. En la cual se obtuvo que se debe realizar una diferencia para corregir la serie a través del comando “diff()”. Para posterior realizar la prueba de estacionariedad con la prueba de Dickey-Fuller para determinar si la serie es estacionaria o no.

Prueba de estacionariedad aplicando una diferencia

Tabla 10-4: Test de Dickey-Fuller

Test de Dickey Fuller	
p-valor	0.01

Realizado por: López, J.; Morocho, M. 2022.

Dado que se obtuvo un p-valor de 0.01 siendo este menor al grado de significancia 0.05, se concluye que existe suficiente evidencia para rechazar la hipótesis nula, con esto se determina que aplicando una diferencia se eliminó la tendencia y la serie paso a ser estacionaria.

4.2.1.2. Funciones de autocorrelación del componente regular (p,q)

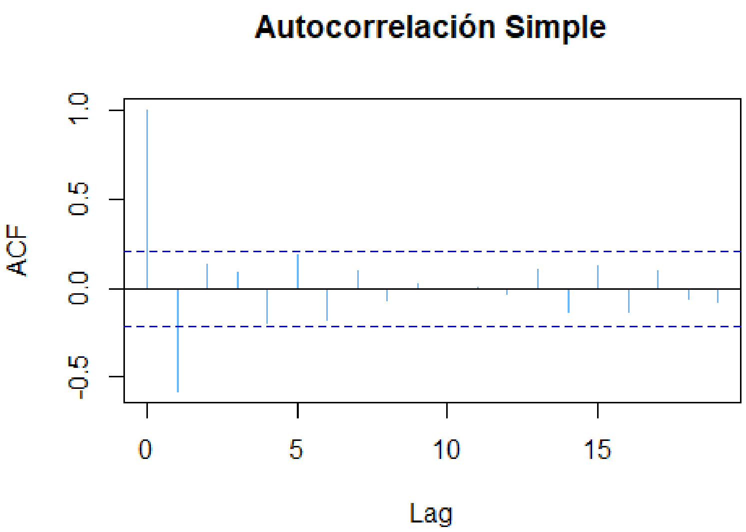

Ilustración 13-4: Correlograma ACF del componente regular

Realizado por: López, J.; Morocho, M. 2022.

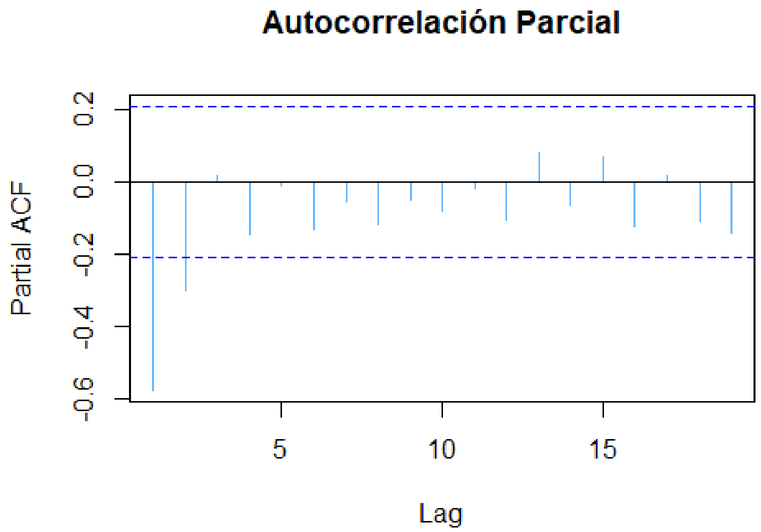

Ilustración 14-4: Correlograma PACF del componente regular

Realizado por: López, J.; Morocho, M. 2022.

A simple vista el gráfico del correlograma de ACF muestra la presencia de dos valores significativos de la función de autocorrelación simple del componente regular, por lo que se trataría de un MA de orden 2, mientras que el en gráfico del correlograma PACF se observa que sobresalen dos valores significativos correspondiente a la función de la autocorrelación parcial del componente regular, el mismo que representa un AR de orden 2, mediante este análisis se puede plantea varios modelos tentativos para seleccionar el más adecuado.

4.2.1.3. Funciones de autocorrelación del componente estacional (P,Q)

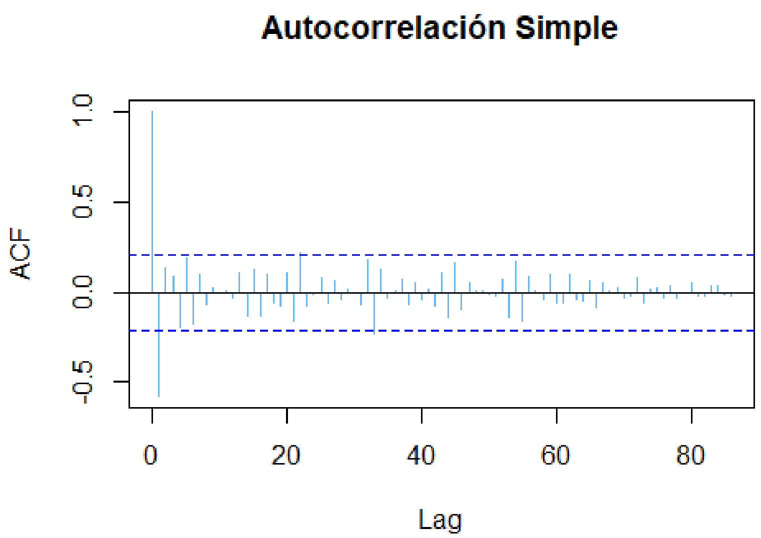

Ilustración 15-4: Correlograma ACF del componente estacional

Realizado por: López, J.; Morocho, M. 2022.

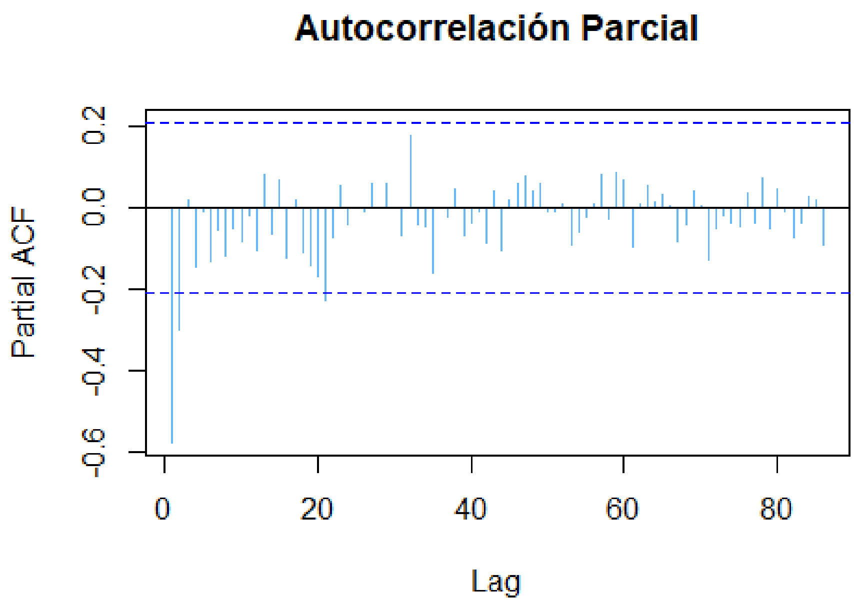

Ilustración 16-4: Correlograma PACF del componente estacional

Realizado por: López, J.; Morocho, M. 2022.

Mediante el análisis gráfico del correlograma de ACF y PACF correspondiente a la parte estacional, se aprecia que existe un valor que no es significativo por lo que tenemos un MA estacional de Q = 0, por otra parte, para el parámetro AR existe dos valores que sobresalen en donde solo uno es significativo por ende tenemos un P = 1.

4.2.2. Estimación de los parámetros

4.2.2.1. Modelo función auto.arima

A través del software estadístico de R podemos implementar una función "auto.arima()"en donde nos otorga de manera automática el mejor modelo que se ajusta a la serie, en base a los criterios de información AIC y BIC, el cual se podrá evaluar con otros modelos candidatos y así elegir el modelo adecuado.

Tabla 11-4: Modelo auto.arima

Modelo	Parámetros
auto.arima	(2,1,1)(1,0,0)[12]

Realizado por: López, J.; Morocho, M. 2022.

4.2.2.2. Ajuste del modelo

A partir de varios modelos conforme al análisis de los procesos autorregresivos y de medias móviles y en base a los correlogramas ACF y PACF tanto para el componente regular y estacional, se propondrá cuatro modelos candidatos con el fin de comparar con el modelo otorgado por la función auto.arima, los cuales serán evaluados para determinar el modelo que mejor se ajuste con el criterio AIC y BIC.

Tabla 12-4: Modelos SARIMA planteados

Modelo candidato	Parámetros
auto.arima	**(2,1,1)(1,0,0)[12]**
2	(2,1,1)(2,0,1)[12]
3	(2,1,1)(2,0,0)[12]
4	(1,1,2)(2,0,1)[12]

Realizado por: López, J.; Morocho, M. 2022.

4.2.2.3. Selección del mejor modelo

Mediante la evaluación de los modelos candidatos implementando el criterio de información de Akaike y Bayesiano se obtuvo los siguientes resultados:

Tabla 13-4: Evaluación AIC y BIC de los modelos

Modelos candidatos	Parámetros	AIC	BIC
auto.arima	**(2,1,1)(1,0,0)[12]**	**985.19**	**997.52**
2	(2,1,1)(2,0,1)[12]	986.9	1004.16
3	(2,1,1)(2,0,0)[12]	987.36	1002.16
4	(1,1,2)(2,0,1)[12]	991.18	1008.44

Realizado por: López, J.; Morocho, M. 2022.

Conforme a la información proporcionada al realizar la evaluación de los modelos candidatos a través de los criterios de Akaike y Bayesiano, el cual nos indica que el mejor modelo que se ajusta a nuestra serie es el que contenga el menor valor de los criterios, el cual viene siendo el modelo generado por la función auto.arima, descartando de manera directa a los modelos planteados.

Seleccionado modelo auto.arima como el mejor a continuación daremos a conocer los coeficientes respectivos del modelo:

Tabla 14-4: Coeficientes del modelo auto.arima

	ar1	ar2	ma1	sar1
Coeficiente	-0.7881	-0.3428	-0.0643	0.0613
Error estándar	0.3693	0.2449	0.4026	0.1217

Realizado por: López, J.; Morocho, M. 2022.

Obtenido los coeficientes del modelo ARIMA(2,1,1)(1,0,0)[12] la ecuación se define de la siguiente manera:

$$X_t = -0.7881X_{t-1} - 0.3428X_{t-2} + 0.0643\varepsilon_{t-1} + 0.0613X_{t-12} + \varepsilon_t \qquad 16\text{-}4$$

Donde:

X_t: es la serie en estudio (delitos),

ε_t: proceso de ruido blanco.

4.2.3. Diagnóstico y validación del modelo

4.2.3.1. Diagnóstico gráfico de los residuos

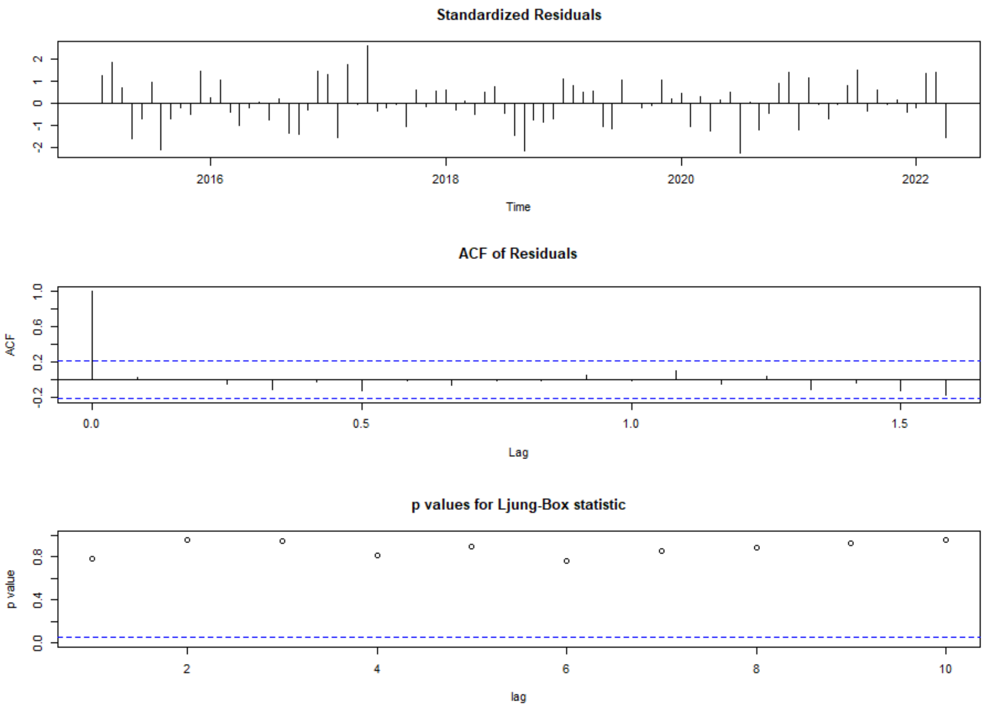

Ilustración 17-4: Diagnóstico de los residuos del modelo ARIMA(2,1,1)(1,0,0)[12]

Realizado por: López, J.; Morocho, M. 2022.

Al realizar un diagnóstico se puede apreciar que los residuos estandarizados en la primera gráfica son aleatorios, es decir, es un proceso de ruido blanco con media cero y varianza constante, mientras que en la segunda gráfica que representa la autocorrelación simple de los residuos se ve como los retardos están entre los límites de confianza, mientras que la última gráfica se verifica que exactamente los residuos son ruido blanco ya que en base al estadístico de Ljung-Box en el cual los puntos de referencia se encuentran por encima del valor de 0.05, por lo que nos indica que el modelo seleccionado se ajusta bien a los datos, por lo tanto el modelo encontrado nos ayuda a realizar las predicciones.

4.2.3.2. Verificación de supuestos

Prueba de Ruido blanco

$$H_0: Existe\ ruido\ blanco$$
$$H_1: No\ existe\ ruido\ blanco$$

Tabla 15-4: Prueba de Ljung-Box de ruido blanco

Prueba de Ljung-Box	
p-valor	0.7822

Realizado por: López, J.; Morocho, M. 2022.

Mediante la Prueba de Ljung-Box se determina que, como el p-valor de 0.7822 obtenido es mayor al nivel de significancia de 0.05, se acepta la hipótesis nula, es decir el modelo estimado se ajusta de manera correcta a los datos y sigue un proceso de ruido blanco, con esto se corrobora que los residuos del error siguen una distribución con media cero y varianza contante como se muestra en la siguiente gráfica.

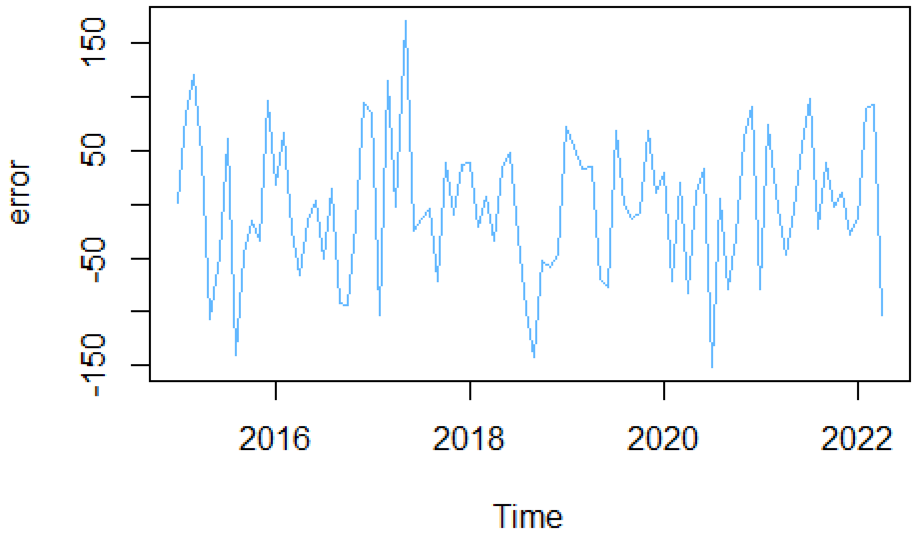

Ilustración 18-4: Residuos del error del modelo ARIMA(2,1,1)(1,0,0)[12]

Realizado por: López, J.; Morocho, M. 2022.

Como se observa en la gráfica de los residuos de la serie de tiempo con media y varianza constante siguiendo así un proceso de ruido blanco, por lo que el modelo indicado es aceptable.

Prueba de Normalidad

$$H_0: Los\ errores\ siguen\ una\ distribución\ normal$$
$$H_1: Los\ errores\ no\ siguen\ una\ distribución\ normal$$

Tabla 16-4: Pruebas de normalidad de Jarque y Wilk

Prueba de Jarque Bera		**Prueba de Shapiro Wilk**	
p-valor	0.8323	**p-valor**	0.9184

Realizado por: López, J.; Morocho, M. 2022.

Para verificar este supuesto de normalidad se implementó el uso de dos contrastes con el fin de garantizar el cumplimiento del mismo, en donde se obtuvo en las dos pruebas un p-valor mayor al nivel de significancia de 0.05, por lo que existe suficiente evidencia para aceptar la Ho, con esto se confirma que los errores o residuos del modelo estimado tienen una distribución normal.

Prueba de Independencia

$$Ho\text{: } Existe\ independencia\ en\ los\ residuos$$
$$H_1\text{: } No\ existe\ independencia\ en\ los\ residuos$$

Tabla 17-4: Prueba de independencia de Ljung-Box

Prueba de Ljung-Box	
p-valor	0.7822

Realizado por: López, J.; Morocho, M. 2022.

Con un nivel de significancia de 0.05, y al implementar la Prueba de Ljung-Box el cual no arroja un p-valor de 0.7822 siendo este mayor, se acepta la hipótesis nula y así se verifica la existencia de independencia en los residuos cumpliendo así este supuesto.

Prueba de Homocedasticidad

$$H_0\text{: } Existe\ homocedasticidad$$
$$H_1\text{: } No\ existe\ homocedasticidad$$

Tabla 18-4: Prueba de homocedasticidad de White

Prueba de White	
p-valor	0.4742

Realizado por: López, J.; Morocho, M. 2022.

Aplicando la Prueba de White para validar este supuesto, se determina que al obtener un p-valor de 0.4742 y este al ser mayor al nivel de significancia establecido de 0.05, se acepta la hipótesis nula, de acuerdo con esto se confirma que la variabilidad es constante, es decir, existe homocedasticidad dado que la variabilidad se conserva constante en el transcurso de la serie temporal.

4.2.4. Pronóstico del modelo

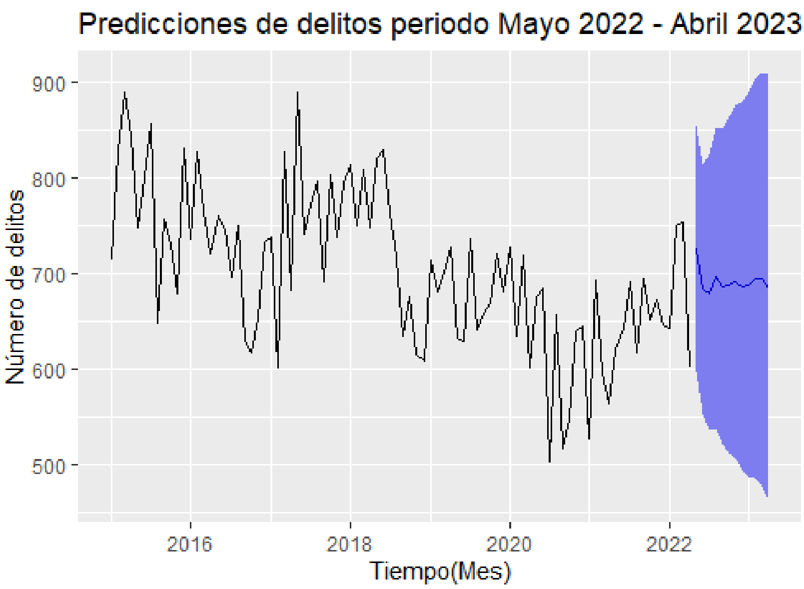

Ilustración 19-4: Pronóstico del modelo ARIMA(2,1,1)(1,0,0)[12]

Realizado por: López, J.; Morocho, M. 2022.

Con el pronóstico grafico se observa el incremento del número de delitos para el mes de mayo de 2022 con un total de 727 aproximadamente, para el transcurso de los próximos meses restantes se puede ver cómo va decreciendo levemente el número de actos delictivos con un número aproximado entre 680 a 696 delitos con respecto a los meses de junio de 2022 a abril de 2023, en el cual se detallan los valores pronosticados a continuación:

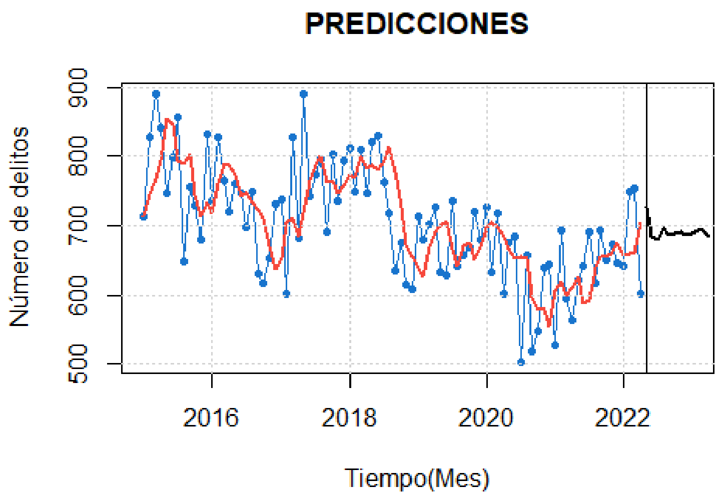

Ilustración 20-4: Ajuste del Modelo ARIMA(2,1,1)(1,0,0)[12]

Realizado por: López, J.; Morocho, M. 2022.

Mediante la ilustración se evidencia el ajuste adecuado del modelo ARIMA(2,1,1)(1,0,0)[12] con la serie de tiempo original, al realizar la predicción con una proyección de los próximos 12 meses.

Tabla 19-4: Pronóstico de delitos periodo Mayo 2022 - Abril 2023

Mes	Valor pronosticado	Límite inferior 95%	Límite superior 95%
Mayo 2022	727	595	858
Junio 2021	684	551	816
Julio 2022	680	533	827
Agosto 2022	696	535	857
Septiembre 2022	687	518	855
Octubre 2022	688	509	867
Noviembre 2022	691	503	879
Diciembre 2022	687	490	883
Enero 2023	688	483	893
Febrero 2023	694	481	907
Marzo 2023	694	473	914
Abril 2023	685	457	913

Realizado por: López, J.; Morocho, M. 2022.

Como se observa en la **Tabla 20-4** los valores pronosticados los mismos que se encuentran dentro de los intervalos de confianza al 95% para los 12 meses, lo que nos indica que efectivamente el modelo seleccionado es el más adecuado para predecir un estimado de delitos mensuales en la provincia de Chimborazo.

CONCLUSIONES

- Mediante el análisis exploratorio univariante de la serie temporal se pudo reconocer las principales características en cuanto al promedio estimado de delitos mensuales que oscila entre los 682.6, con un mínimo de 114 y un pico elevado de 970, al igual que la detección de datos atípicos, que luego se corrigió con el reemplazo a través del promedio de delitos que corresponde al mes en el cual se encuentra el dato anómalo, para no tener problemas en el momento de realizar el modelado estadístico.

- Para identificar el modelo adecuado se fundamentó en la transformación de la serie temporal a estacionaria aplicando una diferencia, se verifico mediante la prueba de estacionariedad de Dickey-Fuller donde se obtuvo un p-valor de 0.01 siendo este menor al grado de significancia de 0.05 lo cual nos da a conocer que se eliminó la tendencia y la serie paso a ser estacionaria, y en base al análisis de las funciones de autocorrelación simple y parcial tanto para parte regular y estacional.

- Implementando en R la función de auto.arima() se puedo estimar los parámetros con respecto al proceso autorregresivo y de medias móviles tomando en cuenta los correlogramas de la función de autocorrelación simple y parcial del componente regular y estacional con la diferencia integrada, en donde nos proporcionó un modelo ARIMA(2,1,1)(1,0,0)[12] siendo este el que mejor se ajusta a los datos.

- Tomando en cuenta el modelo indicado por la función auto.arima(), se realiza una evaluación comparativa con varios modelos planteados, con el fin de determinar el modelo adecuado en base al menor valor en este caso de 985.19 y 997.52 en los criterios de información AIC y BIC y con la validación de los supuestos de normalidad, independencia y homocedasticidad, en el que se destaca el modelo ARIMA(2,1,1)(1,0,0)[12] como el mejor.

- Con la selección del mejor modelo, que en este caso es un ARIMA(2,1,1)(1,0,0)[12] denotada por $X_t = -0.7881X_{t-1} - 0.3428X_{t-2} + 0.0643\varepsilon_{t-1} + 0.0613X_{t-12} + \varepsilon_t$, se procedió a realizar el pronóstico respectivo, que por medio de los valores predichos se corrobora lo antes mencionado, esto nos brinda un panorama real en cuanto a actos delictivos futuros, en el cual se estima que para el mes de mayo de 2022 un pico alto con un total de 727 delitos siendo este el más elevado con relación a los meses restantes pronosticados que oscilan entre los 680 a 696 delitos totales mensuales.

RECOMENDACIONES

- Se recomienda a la Fiscalía Provincial de Chimborazo tomar como referencia a los resultados obtenidos en esta investigación con respecto a la cantidad de delitos pronosticados, para la toma de decisiones pertinentes para el adecuado control de la seguridad en estos meses.

- Haciendo énfasis a la información otorgada por la entidad en estudio, se aconseja un registro adecuado de los datos, dado que para estudios próximos de dicha información se obtenga de forma inmediata beneficiando así al investigador.

- Considerar el software estadístico R como una herramienta para el uso de pronósticos ya sea en temas de delitos, accidentes de tránsito, etc. Con el fin de obtener resultados relevantes o significativos que nos acerquen a la realidad en la provincia de Chimborazo.

GLOSARIO

Serie temporal: Conjunto infinito de observaciones respecto a los datos de una variable, medidos a través de periodos secuenciales o puntos sucesivos en el tiempo (Mauricio, 2007, p. 1).

Tendencia: Comportamiento de una serie en el trascurso del tiempo, que tiende a decrecer o crecer a largo plazo (García, 2016, p. 17).

Homocedasticidad: La serie no cambia con el tiempo, es decir, la variabilidad es contante (Villalba, 2021, pp. 11-12).

Heteroscedasticidad: La variabilidad de la serie no se mantiene constante, aumenta o disminuye a través del tiempo (Villalba, 2021, pp. 11-12).

Estacionalidad: Comportamiento o patrón especifico de la serie de tiempo de un periodo a otro (García, 2016, p. 17).

Ruido blanco: Los datos de una serie de tiempo son independientes e idénticamente distribuidos con media cero y varianza constante (Villalba, 2021, pp. 11-12).

ARIMA: Modelo conformado por un proceso autorregresivo y media móvil, que al no ser estacionaria la serie de tiempo, integra un diferencial de corrección (López & Martínez, 2013, p. 55).

SARIMA: Extensión del modelo ARIMA con influencia de un componente estacional (López & Martínez, 2013, p. 55).

Correlograma: Representa de manera gráfica las funciones de autocorrelación simple y parcial (Cáceres, et al., 2007, p. 107).

Pronóstico: Predicción estimada de valores futuros de una serie temporal (De la Fuente Fenández, 2013, p. 3).

BIBLIOGRAFÍA

GAVILANES PILCO, Valeria. Análisis estadístico de las posibles zonas vulnerables de los delitos registrados en la ciudad de Riobamba, periodo 2015-2017 [En línea] (Trabajo de titulación). Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador. 2018. p. 8. [Consulta: 2022-04-15]. Disponible en: http://dspace.espoch.edu.ec/handle/123456789/9232

TOLEDO ASTUDILLO, Rodrigo. Métodos econométricos para el pronóstico de delitos en el gran Santiago [En línea] (Trabajo de titulación). Universidad de Chile, Santiago-Chile. 2005. p. 18. [Consulta: 2022-04-15]. Disponible en: https://repositorio.uchile.cl/handle/2250/108338

VILLALBA BASANTES, Carlos. Predicción del número de delitos de tránsito que ingresaran a la fiscalía provincial de Chimborazo mediante ARIMA 2018-2021 [En línea] (Trabajo de titulación). Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador. 2021. p. 8. [Consulta: 2022-04-15]. Disponible en: http://dspace.espoch.edu.ec/handle/123456789/14807

ROJO, J. Análisis descriptivo y exploratorio de datos [en línea]. Madrid-España: Instituto de Economía y Geografía, 2006. [Consulta: 25 abril 2021]. Disponible en: http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/SPSSAnDescripExplorat.pdf. P. 31

SALVADOR FIGUERAS, Manuel.; & GARGALLO VALERO, Pilar. Análisis Exploratorio de Datos [blog]. [Consulta: 24 mayo 2022]. Disponible en: http://www.5campus.com/leccion/aed

JIMÉNEZ, J. Introducción a R y RStudio [en línea]. Panamá-Panamá: Universidad Tecnológica de Panamá, 2019. [Consulta: 19 mayo 2022]. Disponible en: https://ridda2.utp.ac.pa/bitstream/handle/123456789/9428/manual-introduccion-R.pdf. P. 4

MAURICIO, J. Análisis de Series Temporales [en línea]. Madrid-España: Universidad Complutense de Madrid, 2007. [Consulta: 15 abril 2022]. Disponible en: https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf. P. 1

GARCÍA DÍAZ, J. Predicción en el dominio del tiempo: análisis de series temporales para ingenieros [en línea]. Valencia-España: Editorial de la Universidad Politécnica de Valencia, 2016. [Consulta: 17 abril 2022]. Disponible en: https://elibro.net/es/ereader/espoch/57439?page=17. P. 17

ROSALES ÁLVAREZ, Ramón; et al. Fundamentos de econometría intermedia: teoría y aplicaciones [en línea]. Universidad de los Andes, 2013. [Consulta: 26 abril 2021]. Disponible en: https://elibro.net/es/ereader/espoch/69432?page=176. P.176

CÁCERES HERÁNDEZ, José; et al. Introducción al Análisis Univariante de series temporales económicas [en línea]. Delta Publicaciones, 2009. [Consulta: 17 abril 2021]. Disponible en: https://elibro.net/es/ereader/espoch/170132?page=118. P.107

MELO MORÍN, Julia.; & SANTANA ESPARZA, Gil. "Minado de series de tiempo utilizando la metodología. Revista de Investigación y Desarrollo". Revista de Investigación y Desarrollo [en línea], 2016, (México) 2(5), pp. 21-31. [Consulta: 25 abril 2022]. ISSN 2444-4987. Disponible en: https://www.ecorfan.org/spain/researchjournals/Investigacion_y_Desarrollo/vol2num5/Revista_de_Investigaci%C3%B3n_y_Desarrollo_V2_N5_3.pdf. P.25

GARCÍA DÍAZ, J. Predicción en el dominio del tiempo: análisis de series temporales para ingenieros [en línea]. Valencia-España: Editorial de la Universidad Politécnica de Valencia, 2016. [Consulta: 17 abril 2022]. Disponible en: https://elibro.net/es/ereader/espoch/57439?page=17. P. 67

VILLAVICENCIO, John. Introducción a las series de tiempo. Puerto Rico, 2010, P.5.

CAPEL SANZ, Irene. Estudio de datos temporales con R [En línea] (Trabajo de titulación). Universidad de Almería, Almería-España. 2019. p. 18. [Consulta: 2022-04-25]. Disponible en: http://repositorio.ual.es/handle/10835/4671. P.18

BONILLA RUIZ, Ana. Estudio estadístico de un caso real. Series temporales [En línea] (Trabajo de titulación). Universidad de Almería, Almería, España. 2019. p. 8. [Consulta: 2022-04-26]. Disponible en: http://repositorio.ual.es/handle/10835/7878

PEÑA, D. Análisis de series temporales [en línea]. Madrid-España: Alianza Editorial, 2005. [Consulta: 17 abril 2021]. Disponible en: https://drive.google.com/file/d/12HxEsMNFhOnj1DMwiwGSRgugIR55LiLa/view?fbclid=IwAR2u1d519pJQ0esAurZkFvk6O2bTPIPqKU60aKZHgwNJrRAqlY9718iP6lU. P. 158

LÓPEZ SARMIENTO, Danilo Alfonso.; & MARTÍNEZ ALAYÓN, Carlos Andrés. "Modelado de pérdidas en una transmisión de video por medio de series de tiempo ARIMA y SARIMA". Ternura [en línea], 2013, (Bogotá) 17(37), pp. 53-63. [Consulta: 19 junio 2022]. ISSN 0123-921X. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-921X2013000300006. P.55

DE LA OLIVA DE CON, Fidel; JIMENO LIENS, Raydel.; & DIAZ DE VILLEGAS JORDAN, Lissette. "Aproximación a la metodología Box-Jenkins para la predicción de la tasa de cambio EUR/USD". Cofin [en línea], 2016, (Cuba) 10(1), pp. 57-75. [Consulta: 15 abril 2022]. ISSN 2073-6061. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2073-60612016000100005

DE LA FUENTE FENÁNDEZ, S. Series Temporales, Modelo ARIMA Metodología de Box - Jenkins [en línea]. Madrid-España. Universidad Autónoma de Madrid, 2013. [Consulta: 31 mayo 2022]. Disponible en: https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf. P.3

VILLALBA BASANTES, Carlos. Predicción del número de delitos de tránsito que ingresaran a la fiscalía provincial de Chimborazo mediante ARIMA 2018-2021 [En línea] (Trabajo de titulación). Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador. 2021. p. 8. [Consulta: 2022-04-15]. Disponible en: http://dspace.espoch.edu.ec/handle/123456789/14807, pp. 11-12

OSORIO HERNÁNDEZ, Jony.; & ÁNGEL OCHOA, Víctor. Aplicación del modelo Box-Jenkins en la estimación del flujo de llamadas en servicios de atención al cliente en el sector contact center [en línea]. Cartagena-Colombia: Fundación Universitaria Los Libertadores, 2019. [Consulta: 6 junio 2021]. Disponible en: https://repository.libertadores.edu.co/handle/11371/2798, pp. 22-23.

CARRIÓN, José. Delito y la Pena [blog]. [Consulta: 15 abril 2022]. Disponible en: https://derechoecuador.com/delito-y-la-pena/

ANEXOS

ANEXO A: AVAL DE LA FISCALÍA PROVINCIAL DE CHIMBORAZO

FGE
Fiscalía General del Estado
ECUADOR

43 - AV. MIGUEL ANGEL LEON NO. 2335 Y JOSE VELOZ

Oficio No.FPH-UGP-2022-000671-O

Riobamba, 21 de abril de 2022

Asunto: ATENCIÓN A REQUERIMIENTO DE INFORMACIÓN ESTADISTICA.

Sr.
Flores Muñoz Pablo Javier
Coordinador de la Carrera de Estadistica Espoch
CIUDADANO
riobamba

En atención al pedido de información realizado por el Ing. Pablo Flores Muñoz, Coordinador de la carrera de Estadistica de la Escuela Superior Politécnica de Chimborazo.

En documento adjunto, se remite la información concerniente, relacionada con los datos estadísticos mensuales de recepción de noticias del delito (NDDs), en todas las unidades de la Fiscalía Provincial de Chimborazo, durante el periodo 2015 – Abril 2022.

Particular que pongo en su conocimiento para los fines académicos pertinentes.

Atentamente,

Ab. Mgs. Cesar Alberto Peñafiel Andino
Analista Provincial de Gestión Procesal 2
Fiscalías Provinciales
FISCALÍA PROVINCIAL DE CHIMBORAZO

FISCALÍA DE CHIMBORAZO
FGE
Fiscalía General del Estado
ECUADOR
RIOBAMBA

Referencia: FPH-GD-2022-000619-EXT

Fecha de elaboración	Elaborado por:	Revisado por:	Aprobado por:
2022-04-21 11:30:56	Peñafiel Andino Cesar Alberto	Peñafiel Andino Cesar Alberto	Peñafiel Andino Cesar Alberto

ANEXO B: CODIFICACIÓN EN EL SOFTWARE ESTADÍSTICO R

```
# Paquetes y librerías a utilizar

library(tseries)
library(astsa)

library(forecast)

library(timeDate)

library(ggplot2)

# Importar base de datos

delito <- read.table("delitos.csv", header=F,sep=";", dec=",")

# Convertir objeto tipo serie temporal

delito.ts <- ts(data = delito,start = c(2015,1), frequency =12)

# Gráfica de la serie temporal

plot(delito.ts, main= "DELITOS EN LA PROVINCIA DE CHIMBORAZO", xlab =
"Años"
     ,ylab= "Delitos mensuales", col= "steelblue1")
tendencia<- lm(delito.ts~time(delito.ts))
abline(tendencia,col="red")

# Análisis descriptivo

summary(delito.ts)

var(delito.ts) # Varianza

sd(delito.ts) # Desviación estándar

skewness(delito.ts) # Simetría

kurtosis(delito.ts) # Curtosis
```

```r
# Análisis exploratorio

# Histograma

hist(delito.ts,freq = FALSE,main = "HISTOGRAMA",col= "steelblue1")
lines(density(delito.ts))

# Diagrama de cajas

boxplot(delito.ts, main= "DELITOS EN LA PROVINCIA DE CHIMBORAZO",
        horizontal = TRUE,col= "steelblue1")

# Identificación de datos atípicos

datip <- boxplot.stats(x = delito.ts)

# Reemplazo de datos atípicos

delito1 <- read.table("delito.csv", header=F,sep=";", dec=",")

delito.ts <- ts(data = delito1, start = c(2015,1), frequency =12)

# Diagrama de cajas sin datos atípicos

boxplot(delito.ts, main= "DELITOS EN LA PROVINCIA DE CHIMBORAZO",
        horizontal = TRUE,col= "steelblue1")

# Histograma sin datos atípicos

hist(delito.ts,freq = FALSE,main = "HISTOGRAMA",col= "steelblue1")
lines(density(delito.ts))

# Identificación de datos atípicos

datip1 <- boxplot.stats(x = delito.ts)

# Serie temporal sin presencia de datos atípicos
```

```r
plot(delito.ts, main= "DELITOS EN LA PROVINCIA DE CHIMBORAZO", xlab =
"Años",
     ylab= "Delitos mensuales", col= "steelblue1")
tendencia<- lm(delito.ts~time(delito.ts))
abline(tendencia,col="red")

# Descomposición de la serie

desc<-decompose(delito.ts, type = "additive")
plot(desc,  col= "steelblue1")

# IDENTIFICACIÓN DEL MODELO

# Test de estacionariedad

adf.test(delito.ts)

# Asignación de diferencias parte regular

ndiffs(delito.ts)

# Primera diferencia

delitodif <- diff(delito.ts)

# Asignación de diferencias parte estaocional

nsdiffs(delito.ts)

# Test de estacionariedad aplicando una diferencia

adf.test(delitodif)

# Correlograma parte regular

acf(ts(delitodif),main="Autocorrelación Simple",col="steelblue1")

pacf(ts(delitodif),main="Autocorrelación Parcial",col="steelblue1")
```

```r
# Correlograma parte estacional

acf(ts(delitodif),main="Autocorrelación Simple",col="steelblue1",lag.max = 90)

pacf(ts(delitodif),main="Autocorrelación Parcial",col="steelblue1",lag.max = 90)

# ESTIMACIÓN DE LOS PARÁMETROS

# Modelo de auto.arima

m1 <- auto.arima(delito.ts, stepwise = T, approximation = T,trace = T)

summary(m1)

# Modelos candidatos

auto_arima<- arima(delito.ts,order = c(2,1,1),seasonal=list(order=c(1,0,0),period=12))
m2<- arima(delito.ts,order = c(2,1,1),seasonal=list(order=c(2,0,1),period=12))
m3<- arima(delito.ts,order = c(2,1,1),seasonal=list(order=c(2,0,0),period=12))
m4<- arima(delito.ts,order = c(1,1,2),seasonal=list(order=c(2,0,1),period=12))

# Evaluación criterios AIC y BIC

AIC(auto_arima,m2,m3,m4)

BIC(auto_arima,m2,m3,m4)

# DIAGNÓSTICO Y VALIDACIÓN DEL MODELO

# Diagnóstico gráfico de residuos

tsdiag(auto_arima)
```

```r
# Verificación de supuestos

# Ruido Blanco

Box.test(residuals(auto_arima), type = "Ljung-Box")

error <- residuals(auto_arima)
plot(error, col="steelblue1")

# Normalidad

jarque.bera.test(auto_arima$residuals)

shapiro.test(auto_arima$residuals)

# Homocedasticidad

white.test(auto_arima$residuals)

# Independencia

Box.test(residuals(auto_arima),type = "Ljung-Box")

# PRONÓSTICO DEL MODELO

pronost <- forecast(auto_arima,h=12,level = 0.95)

autoplot(pronost, main = "Pronóstico de delitos periodo Mayo 2022 - Ab
ril 2023",
         xlab= "Año",ylab= "Número de delitos")



plot(delito.ts,xlab="Año",ylab="Número de delitos",col=4, xlim=c(2015,
2023),
     ylim=range(c(delito.ts,pronost$mean)),lwd=1,pch=20,type="o",main=
"PREDICCIONES")
lines(fitted(auto_arima),col=2,lwd=2)
lines(pronost$mean,col="black",lwd=2)
```

```
grid()
abline(v=c(2022+4/12))
```

espoch | Dirección de Bibliotecas y Recursos del Aprendizaje

UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y DOCUMENTAL

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 16 / 01 / 2023

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: Mario Gustavo Morocho Salinas Jhefri Bladimir López Castillo
INFORMACIÓN INSTITUCIONAL
Facultad: Ciencias
Carrera: Estadística
Título a optar: Ingeniero Estadístico
f. Analista de Biblioteca responsable: Ing. Leonardo Medina Ñuste MSc.

2432-DBRA-UTP-2022