



**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**  
**FACULTAD DE CIENCIAS**  
**CARRERA ESTADÍSTICA**

**ANÁLISIS Y PREDICCIÓN DE ACCIDENTES MENSUALES DE  
TRÁNSITO EN LA PROVINCIA DE CHIMBORAZO EN EL  
PERÍODO ENERO 2015 – DICIEMBRE 2021**

**Trabajo de Integración Curricular**

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

**INGENIERO/A EN ESTADÍSTICA**

**AUTORES: ERICK PAUL INCA MERINO**

**MÓNICA TATIANA TIERRA CHIMBOLEMA**

**DIRECTOR: Dr. JORGE WASHINGTON CONGACHA AUSHAY MSc.**

Riobamba – Ecuador

2022

**©2022, Erick Paúl Inca Merino & Mónica Tatiana Tierra Chimbolema**

Autorizamos la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Nosotros, ERICK PAUL INCA MERINO y MÓNICA TATIANA TIERRA CHIMBOLEMA 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, 25 de Noviembre del 2022



**Erick Paul Inca Merino**  
**060453885-0**



**Mónica Tatiana Tierra Chimbolema**  
**060531362-6**

**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, **ANÁLISIS Y PREDICCIÓN DE ACCIDENTES MENSUALES DE TRÁNSITO EN LA PROVINCIA DE CHIMBORAZO EN EL PERÍODO ENERO 2015 – DICIEMBRE 2021**, realizado por: **ERICK PAUL INCA MERINO y MÓNICA TATIANA TIERRA CHIMBOLEMA**, 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 autorizada su presentación.

	<b>FIRMA</b>	<b>FECHA</b>
Ing. Tania Paulina Morocho Barrionuevo MSc <b>PRESIDENTE DEL TRIBUNAL</b>		2022-11-25
Dr. Jorge Washington Congacha Aushay MSc. <b>DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR</b>		2022-11-25
Ing. Johanna Enith Aguilar Reyes MGs. <b>ASESORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR</b>		2022-11-25

## **DEDICATORIA**

Esta tesis se la dedico a Dios quien supo guiarme por un buen camino, darme fuerza para seguir adelante y no desmayar en los problemas que se me presentaban. A mi madre Ligia Merino, pues ella fue el principal cimiento para la construcción de mi vida profesional, sentó en mi la base de responsabilidad y deseos de superación.

Erick

Esta tesis va dedicada a mis padres Rosa Chimbolema y Gonzalo Tierra por ser mi fuente de motivación e inspiración, quienes me han apoyado moral y económicamente para poder llegar a esta instancia de mis estudios, a mis hermanos Elena, Patricia, Luis, Fabian, Sandra, Paola y Sarita ya que ellos con su experiencia me inculcaron buenos valores, cada día con sus palabras de aliento.

Mónica

## **AGRADECIMIENTO**

Primero damos gracias a Dios por permitirnos tener tan buena experiencia en la universidad, gracias a la ESPOCH por formarnos en profesionales en lo que tanto anhelamos. Queremos agradecer al Dr. Jorge Congacha, a la Ing. Johanna Aguilar por guiarnos en el desarrollo de este trabajo de investigación, a su vez a todos los docentes de la carrera de Estadística quienes hicieron parte de este camino que día a día nos fueron formando. A nuestros compañeros con los cuales compartimos experiencias inolvidables.

Erick & Mónica

## ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS.....	ix
ÍNDICE DE ILUSTRACIONES.....	xi
ÍNDICE DE ECUACIONES .....	xiii
ÍNDICE DE ANEXOS .....	xiv
RESUMEN.....	xv
SUMMARY .....	xvi
INTRODUCCIÓN .....	1

### CAPÍTULO I

1. PROBLEMA DE INVESTIGACIÓN.....	2
1.1. Planteamiento del problema .....	2
1.2. Limitaciones y delimitaciones.....	2
1.3. Problema general de investigación.....	2
1.4. Problemas específicos de investigación.....	2
1.5. Objetivos.....	3
1.5.1. <i>Objetivo general</i> .....	3
1.5.2. <i>Objetivos específicos</i> .....	3
1.6. Justificación.....	3
1.6.1. <i>Justificación teórica</i> .....	3
1.6.2. <i>Justificación metodológica</i> .....	3
1.6.3. <i>Justificación práctica</i> .....	4

### CAPÍTULO II

2. MARCO TEÓRICO.....	5
2.1. Antecedentes de investigación .....	5
2.2. Referencia teórica .....	6
2.2.1. <i>Bases teóricas</i> .....	6
2.2.1.1. <i>Fiscalía general del estado</i> .....	6
2.2.1.2. <i>Accidentes de tránsito</i> .....	6
2.2.1.3. <i>Derecho penal</i> .....	6
2.2.1.4. <i>Infracciones de tránsito</i> .....	7

2.2.2.	<i>Bases conceptuales</i> .....	7
2.2.2.1.	<i>Análisis exploratorio de datos</i> .....	7

### CAPÍTULO III

3.	<b>MARCO METODOLÓGICO</b> .....	22
3.1.	<b>Enfoque de investigación</b> .....	22
3.2.	<b>Nivel de investigación</b> .....	22
3.3.	<b>Diseño de la investigación</b> .....	22
3.3.1.	<i>Según la manipulación o no de la variable independiente</i> .....	22
3.3.2.	<i>Según las intervenciones en el trabajo de campo</i> .....	23
3.4.	<b>Tipo de estudio</b> .....	23
3.5.	<b>Población y planificación, selección y cálculo del tamaño de la muestra</b> .....	24
3.5.1.	<i>Población</i> .....	24
3.5.2.	<i>Tamaño de la muestra</i> .....	24
3.6.	<b>Métodos, técnicas e instrumentos de investigación</b> .....	24
3.6.1.	<i>Métodos de investigación</i> .....	24
3.6.2.	<i>Técnicas de investigación</i> .....	24
3.6.3.	<i>Instrumentos de investigación</i> .....	25

### CAPÍTULO IV

4.	<b>MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS</b> .....	31
4.1.	<b>Diagrama de Pareto</b> .....	31
4.2.	<b>Análisis exploratorio de datos</b> .....	32
4.2.1.	<i>Daños materiales (DM)</i> .....	32
4.2.2.	<i>Lesiones causadas por accidentes de tránsito (LCPATR)</i> .....	37
4.3.	<b>Análisis de serie de tiempo</b> .....	41
4.3.1.	<i>Daños materiales (DM)</i> .....	41
4.3.1.1.	<i>Identificación del modelo</i> .....	41
4.3.1.2.	<i>Modelos ARIMA</i> .....	44
4.3.1.3.	<i>Comprobación del modelo</i> .....	45
4.3.2.	<i>Lesiones causadas por accidente de tránsito (LCPATR)</i> .....	47
4.3.2.1.	<i>Identificación del modelo</i> .....	47
4.3.2.2.	<i>Modelos ARIMA</i> .....	50
4.3.2.3.	<i>Comprobación del modelo</i> .....	51



<b>4.4.</b>	<b>Análisis de Poisson</b> .....	53
<b>4.4.1.</b>	<b><i>Daños materiales (DM)</i></b> .....	53
<b>4.4.1.1.</b>	<i>Identificación de los modelos</i> .....	53
<b>4.4.1.2.</b>	<i>Modelo Poisson</i> .....	57
<b>4.4.1.3.</b>	<i>Comprobación del mejor modelo</i> .....	59
<b>4.4.2.</b>	<b><i>Lesiones causadas por accidentes de tránsito (LCPATR)</i></b> .....	60
<b>4.4.2.1.</b>	<i>Identificación del mejor modelo</i> .....	61
<b>4.4.2.2.</b>	<i>Modelo Poisson</i> .....	65
<b>4.4.2.3.</b>	<i>Comprobación del mejor modelo</i> .....	66
<b>4.5.</b>	<b>Comparación de modelos</b> .....	67
<b>4.5.1.</b>	<b><i>Daños materiales (DM)</i></b> .....	67
<b>4.5.1.1.</b>	<i>Predicción de los accidentes de tránsito de DM</i> .....	68
<b>4.5.2.</b>	<b><i>Lesiones causadas por accidentes de tránsito (LCPATR)</i></b> .....	69
<b>4.5.2.1.</b>	<i>Predicción de los accidentes de tránsito de LCPATR</i> .....	69
	<b>CONCLUSIONES</b> .....	72
	<b>RECOMENDACIONES</b> .....	73
	<b>BIBLIOGRAFÍA</b>	
	<b>ANEXOS</b>	

## ÍNDICE DE TABLAS

<b>Tabla 1-2:</b>	Proceso MA(q), AR(p), ARIMA(p,d,q) .....	14
<b>Tabla 2-2:</b>	Principales modelos lineales generalizados.....	18
<b>Tabla 1-3:</b>	Operacionalización de variables.....	26
<b>Tabla 2-3:</b>	Operacionalización de objetivos.....	29
<b>Tabla 3-3:</b>	Matriz de consistencia .....	30
<b>Tabla 1-4:</b>	Regla del 80-20.....	31
<b>Tabla 2-4:</b>	Atípicos de DM .....	33
<b>Tabla 3-4:</b>	Cuartiles de DM .....	34
<b>Tabla 4-4:</b>	Serie de tiempo de DM.....	35
<b>Tabla 5-4:</b>	Atípicos de LCPATR .....	38
<b>Tabla 6-4:</b>	Cuartiles del gráfico de cajas de LCPATR.....	39
<b>Tabla 7-4:</b>	Serie de tiempo de LCPATR.....	39
<b>Tabla 8-4:</b>	Test de integración para la parte ordinal de DM .....	41
<b>Tabla 9-4:</b>	Test de integración para la parte estacional de DM.....	41
<b>Tabla 10-4:</b>	Modelos de DM.....	44
<b>Tabla 11-4:</b>	Selección del mejor modelo para DM .....	44
<b>Tabla 12-4:</b>	Prueba de normalidad.....	46
<b>Tabla 13-4:</b>	Prueba de homocedasticidad .....	46
<b>Tabla 14-4:</b>	Prueba de independencia .....	46
<b>Tabla 15-4:</b>	Test de integración para la parte ordinal de LCPATR .....	47
<b>Tabla 16-4:</b>	Test de integración para la parte estacional de LCPATR.....	47
<b>Tabla 17-4:</b>	Modelos para LCPATR.....	50
<b>Tabla 18-4:</b>	Selección del mejor modelo para LCPATR .....	50
<b>Tabla 19-4:</b>	Prueba de normalidad.....	52
<b>Tabla 20-4:</b>	Prueba de homocedasticidad .....	52
<b>Tabla 21-4:</b>	Prueba de independencia .....	52
<b>Tabla 22-4:</b>	Modelo lineal de Poisson DM.....	53
<b>Tabla 23-4:</b>	Prueba analítica de homocedasticidad.....	54
<b>Tabla 24-4:</b>	Modelo cuadrático de Poisson DM .....	55
<b>Tabla 25-4:</b>	Diagnóstico del segundo modelo de DM .....	56
<b>Tabla 26-4:</b>	Modelo cúbico de Poisson DM .....	56
<b>Tabla 27-4:</b>	Mejor modelo de Poisson DM.....	58
<b>Tabla 28-4:</b>	Prueba de normalidad.....	59

<b>Tabla 29-4:</b>	Prueba de homocedasticidad .....	60
<b>Tabla 30-4:</b>	Prueba de independencia .....	60
<b>Tabla 31-4:</b>	Modelo lineal de Poisson LCPATR .....	61
<b>Tabla 32-4:</b>	Modelo cuadrático de Poisson LCPATR.....	62
<b>Tabla 33-4:</b>	Modelo cubico de Poisson LCPATR.....	64
<b>Tabla 34-4:</b>	Mejor modelo de Poisson LCPATR.....	65
<b>Tabla 35-4:</b>	Prueba de normalidad .....	66
<b>Tabla 36-4:</b>	Prueba de homocedasticidad .....	67
<b>Tabla 37-4:</b>	Prueba de independencia .....	67
<b>Tabla 38-4:</b>	Comparación del modelo DM .....	67
<b>Tabla 39-4:</b>	Predicciones para el 2022 con el modelo ARIMA (0,0,0) (2,1,2) [12] .....	69
<b>Tabla 40-4:</b>	Comparación del modelo LCPATR .....	69
<b>Tabla 41-4:</b>	Predicciones para el 2022 con el modelo ARIMA (1,0,1) (1,0,1) [12] .....	70

## ÍNDICE DE ILUSTRACIONES

<b>Ilustración 1-4:</b>	Diagrama de Pareto.....	32
<b>Ilustración 2-4:</b>	Serie de tiempo de DM.....	32
<b>Ilustración 3-4:</b>	Boxplot de DM.....	33
<b>Ilustración 4-4:</b>	Boxplot sin datos atípicos de DM.....	34
<b>Ilustración 5-4:</b>	Sin datos atípicos DM.....	35
<b>Ilustración 6-4:</b>	Descomposición de DM.....	36
<b>Ilustración 7-4:</b>	Gráfico mensual de la Serie de Tiempo de DM.....	36
<b>Ilustración 8-4:</b>	Serie de tiempo LCPATR.....	37
<b>Ilustración 9-4:</b>	Boxplot de LCPATR.....	37
<b>Ilustración 10-4:</b>	Boxplot sin datos atípicos LCPATR.....	38
<b>Ilustración 11-4:</b>	Sin datos atípicos LCPATR.....	39
<b>Ilustración 12-4:</b>	Descomposición de LCPATR.....	40
<b>Ilustración 13-4:</b>	Gráfico mensual de la serie de tiempo de LCPATR.....	40
<b>Ilustración 14-4:</b>	Parte regular de la autocorrelación simple.....	42
<b>Ilustración 15-4:</b>	Parte regular de la autocorrelación parcial.....	42
<b>Ilustración 16-4:</b>	Parte estacional de la autocorrelación simple.....	43
<b>Ilustración 17-4:</b>	Parte estacional de la autocorrelación parcial.....	43
<b>Ilustración 18-4:</b>	Diagnóstico del modelo ARIMA (0,0,0) (2,1,2) [12].....	45
<b>Ilustración 19-4:</b>	Parte regular de la autocorrelación simple.....	48
<b>Ilustración 20-4:</b>	Parte regular de la autocorrelación parcial.....	48
<b>Ilustración 21-4:</b>	Parte estacional de la autocorrelación simple.....	49
<b>Ilustración 22-4:</b>	Parte estacional de la autocorrelación parcial.....	49
<b>Ilustración 23-4:</b>	Diagnóstico del modelo ARIMA (1,0,1) (1,0,1) [12].....	51
<b>Ilustración 24-4:</b>	Diagrama de dispersión accidentes de DM.....	53
<b>Ilustración 25-4:</b>	Diagnóstico del primer modelo de DM.....	55
<b>Ilustración 26-4:</b>	Diagnóstico del tercer modelo de DM.....	57
<b>Ilustración 27-4:</b>	Ajuste del mejor modelo DM.....	58
<b>Ilustración 28-4:</b>	Diagnóstico del mejor modelo de DM.....	59
<b>Ilustración 29-4:</b>	Diagrama de dispersión LCPATR.....	61
<b>Ilustración 30-4:</b>	Diagnóstico del primer modelo de LCPATR.....	62
<b>Ilustración 31-4:</b>	Diagnóstico del segundo modelo de LCPATR.....	63
<b>Ilustración 32-4:</b>	Diagnóstico del tercer modelo de LCPATR.....	64
<b>Ilustración 33-4:</b>	Ajuste del mejor modelo de LCPATR.....	65

<b>Ilustración 34-4:</b>	Diagnóstico del mejor modelo de LCPATR.....	66
<b>Ilustración 35-4:</b>	Predicción del modelo ARIMA (0,0,0) (2,1,2) [12] <i>con constante</i> .....	68
<b>Ilustración 36-4:</b>	Predicción del modelo ARIMA (1,0,1) (1,0,1) [12] .....	70

## ÍNDICE DE ECUACIONES

<b>Ecuación 1-2:</b>	ARIMA (1,0,0) o AR (1) o de primer orden.....	10
<b>Ecuación 2-2:</b>	ARIMA(p,0,0) o simplemente por AR(p) .....	10
<b>Ecuación 3-2:</b>	ARIMA mediante el operador de cambio retroactivo B.....	10
<b>Ecuación 4-2:</b>	ARIMA (0,0,1), también denotado por MA (1) .....	11
<b>Ecuación 5-2:</b>	ARIMA (0,0, q) o también por Ma(q) .....	11
<b>Ecuación 6-2:</b>	ARIMA (p, 0, q) .....	11
<b>Ecuación 7-2:</b>	ARIMA (p, 0, q) 2 .....	12
<b>Ecuación 8-2:</b>	ARMA (p, q).....	12
<b>Ecuación 9-2:</b>	ARMA (p, q) estacionario .....	12
<b>Ecuación 10-2:</b>	ARMA (p, q) invertible .....	12
<b>Ecuación 11-2:</b>	ARIMA (p, d, q) .....	12
<b>Ecuación 12-2:</b>	SARIMA (p, d, q) [(P,D,Q)] _s.....	14
<b>Ecuación 13-2:</b>	Coficiente AIC .....	15
<b>Ecuación 14-2:</b>	Coficiente AICc .....	15
<b>Ecuación 15-2:</b>	Bayesiano BIC .....	15
<b>Ecuación 16-2:</b>	Normalidad .....	16
<b>Ecuación 17-2:</b>	Independencia .....	16
<b>Ecuación 18-2:</b>	Homocedasticidad.....	17
<b>Ecuación 19-2:</b>	MLG .....	19
<b>Ecuación 20-2:</b>	Distribución de Poisson .....	19
<b>Ecuación 21-2:</b>	Distribución de Poisson binomial .....	19
<b>Ecuación 22-2:</b>	Regresión de Poisson .....	20
<b>Ecuación 23-2:</b>	Probabilidad para n observaciones .....	20
<b>Ecuación 24-2:</b>	Aproximación a la distribución normal .....	20
<b>Ecuación 25-2:</b>	Contraste de la bondad del ajuste .....	21
<b>Ecuación 1-4:</b>	Modelo aditivo.....	54
<b>Ecuación 2-4:</b>	Modelo multiplicativo .....	54

## **ÍNDICE DE ANEXOS**

**ANEXO A:** AVAL DE LA FISCALÍA PROVINCIAL DE CHIMBORAZO

**ANEXO B:** CÓDIGO EN EL ENTORNO COMPUTACIONAL

## RESUMEN

El objetivo de esta investigación fue predecir el número de accidentes de tránsito mensuales que ingresan a la Fiscalía Provincial de Chimborazo (FPCH), mediante modelos lineales generalizados de ARIMA los cuales fueron comparados con modelos de Poisson para la toma de decisiones confiables y mejorar la atención de la FPCH. Para el desarrollo de la investigación se utilizó datos mensuales del periodo; Enero 2015 – Diciembre 2021, proporcionados por la unidad Gestión Procesal de la FPCH de las cuales se aplicó la regla del 80-20 donde se obtuvo como las variables más significativas: Daños Materiales (DM) ya que representa el 58% de los accidentes de tránsito y lesiones causadas por accidentes de tránsito (LCPATR) con el 28% de los accidentes de tránsito. Como resultados en el análisis exploratorio de datos univariante, se observó la presencia de datos atípicos, se realizó la descomposición de las series de tiempo constatando así la presencia de estacionalidad. Se concluye que entre los modelos ARIMA Y POISSON mediante el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC) como el modelo con mejor calidad predictiva ARIMA (0,0,0) (2,1,2) [12] para la variable DM y ARIMA (1,0,1) (1,0,1) [12] para la variable LCPATR con el cual se procedió a realizar las respectivas predicciones. Se recomienda a la FPCH utilizar más variables a fin de aplicar métodos multivariantes que se ajusten a situaciones de la realidad.

**Palabras clave:** <ACCIDENTES DE TRÁNSITO>, <MODELO LINEAL GENERALIZADO>, <REGRESIÓN DE POISSON>, <SERIES DE TIEMPO>, <AUTORREGRESIVO INTEGRADO DE MEDIA MÓVIL>.

0037-DBRA-UPT-2023





## SUMMARY

The objective of this research was to predict the number of monthly traffic accidents that enter the Provincial Prosecutor's Office of Chimborazo (FPCH), through generalized linear models of ARIMA which were compared with Poisson models for reliable decision making and to improve the service of the FPCH. For the development of the study, monthly data, provided by the Procedural Management unit of the FPCH, of the period January 2015 – December 2021 was used, of which the 80-20 rule was applied, where the most significant variables were obtained: Material Damage (DM) since it represents 58% of accidents and injuries caused by traffic accidents (LCPATR) with 28% of traffic accidents. As results in the exploratory analysis of univariate data, the presence of atypical data was observed, the decomposition of the time series was done, thus verifying the presence of seasonality. It is concluded that between the ARIMA and POISSON models using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) as the model with the best predictive quality ARIMA (0,0,0) (2,1,2) [12] for the DM variable and ARIMA (1,0,1) (1,0,1) [12] for the LCPATR variable with which the respective predictions were made. It is recommended that the FPCH use more variables in order to apply multivariate methods that adjust to real situations.

**Keywords:** <TRAFFIC ACCIDENTS>, <GENERALIZED LINEAR MODEL>, <POISSON REGRESSION>, <TIME SERIES>, <AUTOREGRESSIVE INTEGRATED MOVING AVERAGE>.



Edgar Mesías Jaramillo Moyano

0603497397

## **INTRODUCCIÓN**

Los accidentes de tránsito son analizados por la fiscalía general del Estado, (FGE) de la provincia de Chimborazo, los cuales son realizados por el Servicio de Investigaciones de Accidentes de Tránsito (SIAT), esto es importante para determinar las responsabilidades en contra de los causantes.

En el pensum de la carrera de Estadística de la ESPOCH, que viene dictando la asignatura de Series de Tiempo la misma que toma en cuenta el estudio de predicciones de observaciones tomadas en el tiempo. Para realizar predicciones aplicamos la metodología Box-Jenkins de los modelos ARIMA y sus derivados.

Es importante mencionar que existe el enfoque de los modelos ARIMA (procesos autorregresivos integrados con promedios móviles) y sus derivados, que requieren de una mayor conceptualización estadística. La validación de los supuestos es importante para verificar si nuestro modelo es óptimo, para ello consideramos los supuestos de normalidad, homocedasticidad e independencia, estos se comprueban en los errores generados por los modelos planteados.

Para la realización de esta investigación la FGE de la provincia de Chimborazo ha facilitado los datos de los accidentes de tránsito mensuales, en vista que existen diversos tipos de accidentes de tránsito aplicaremos la ley del 80-20 para determinar cuáles son los más representativos. Se realizó un análisis exploratorio para poder tener una idea del comportamiento de las series de tiempo en estudio y cumplir los objetivos que nos hemos planteado en esta investigación.

En la presente investigación se pretende analizar y pronosticar los accidentes de tránsito en la provincia de Chimborazo para el año 2022, para poder predecir lo haremos mediante el modelo ARIMA y derivados para después comparar los como modelos de POISSON con el fin de elegir el mejor modelo con calidad predictiva.

Este trabajo de titulación consta de cuatro capítulos. Capítulo I describe el trabajo de titulación, el problema y justificar la realización del proyecto. Además, indica los objetivos del proyecto, es decir lo que se quiere conseguir con la realización de este trabajo. Capítulo II contiene toda la teoría para desarrollar el trabajo de titulación. Capítulo III indica la metodología que se sigue para el análisis es decir cómo se realizó la investigación con mayor detalle. Capítulo IV se encarga del análisis, discusión e interpretación los resultados. Finalmente, este trabajo también presenta conclusiones, recomendaciones, bibliografía y anexos.

## CAPÍTULO I

### 1. PROBLEMA DE INVESTIGACIÓN

#### 1.1. Planteamiento del problema

Las personas que se encargan de los análisis estadísticos de la FGE de la provincia de Chimborazo desconocen el comportamiento de los accidentes de tránsito que llegan a esta dependencia. Actualmente los Fiscales toman decisiones según su experiencia a lo largo de los años, por este desconocimiento, este proyecto ayudará a tener un estimado del número de accidentes que ingresan a la Fiscalía del cantón Riobamba y de esta manera se podrán tomar mejores decisiones.

#### 1.2. Limitaciones y delimitaciones

Para realizar las predicciones de accidentes mensuales de tránsito se delimitó la zona y se limitó el período de tiempo para hacerlo más factible y se realizará en la provincia de Chimborazo en el período Enero 2015 – Diciembre 2021.

#### 1.3. Problema general de investigación

¿Cuántos accidentes de tránsito ingresarán mensualmente en el 2022 a la FGE de la Provincia de Chimborazo?

#### 1.4. Problemas específicos de investigación

- ❖ ¿Qué tipos de accidentes de tránsito son más frecuentes en la provincia de Chimborazo?
- ❖ ¿Qué métodos estadísticos se pueden realizar para predecir los accidentes de tránsito de la provincia de Chimborazo?
- ❖ ¿Qué modelo es mejor, Poisson o ARIMA para predecir accidentes de tránsito de la provincia de Chimborazo?
- ❖ ¿Cuántos modelos de Poisson y ARIMA podemos realizar para predecir los accidentes de tránsito de la provincia de Chimborazo?
- ❖ ¿Qué modelo es más adecuado: Poisson o ARIMA, para predecir accidentes de tránsito en la provincia de Chimborazo?

## **1.5. Objetivos**

### ***1.5.1. Objetivo general***

- ❖ Predecir el número de accidentes de tránsito mensuales en la provincia de Chimborazo mediante MLG: modelos ARIMA y compararlos con modelos de POISSON en el período enero 2015 – diciembre 2021.

### ***1.5.2. Objetivos específicos***

- ❖ Aplicar la regla del 80-20 para identificar los accidentes de tránsito más representativos.
- ❖ Realizar un análisis exploratorio de datos univariante de los accidentes de tránsito para obtener una matriz más manejable y poder aplicar métodos estadísticos.
- ❖ Construir modelos de ARIMA y POISSON con el propósito de realizar predicciones de los accidentes de tránsito.
- ❖ Verificar los supuestos de los modelos ARIMA y POISSON, para poder elegir el mejor modelo.
- ❖ Comparar los dos tipos de modelos a fin de seleccionar el mejor.

## **1.6. Justificación**

### ***1.6.1. Justificación teórica***

La Fiscalía General de Chimborazo no está realizando actualmente un análisis exploratorio y predictivo del número de accidentes de tránsito, en la actualidad, los accidentes de tránsito son una de las principales causas, por lo que los organismos (Fiscalía General de Justicia, Policía, Tribunales) se encargan de atender los accidentes de tránsito. Lesiones humanas y muerte.

Los modelos MLG: ARIMA y POISSON nos ayudaran a pronosticar los accidentes de tránsito para el año 2022, y con estos datos la FGE de la provincia de Chimborazo estará mejor organizada.

### ***1.6.2. Justificación metodológica***

Para la recolección de datos, nos comunicamos con la Fiscalía General de Chimborazo para que nos proporcionó la base de datos de los accidentes mensuales de la provincia de Chimborazo desde enero de 2015 hasta diciembre de 2021. Hay muchos tipos de accidentes de tránsito para

analizar según las instrucciones, así que se aplica la regla del 80-20 para determinar cuál es el más representativo.

### ***1.6.3. Justificación práctica***

Como resultado, se debe realizar un análisis exploratorio e inferencial de datos para ver cómo se comportan y predecir utilizando los modelos MLG; ARIMA y POISSON. A partir de los resultados del pronóstico de frecuencia de accidentes de tránsito, se puede tener una aproximación de los accidentes de tránsito mensuales, y poder tomar una decisión.

## CAPÍTULO II

### 2. MARCO TEÓRICO

#### 2.1. Antecedentes de investigación

Un trabajo de Camila Conde Ochoa y Ariana E. Rodríguez Álvarez titulado “PREDICCIÓN DE ACCIDENTES VIALES EN CUENCA Y MODELOS DE SERIES TEMPORALES ARIMA” identificó cuatro grandes fases para el análisis. Estadísticas analíticas, pronósticos de series temporales con modelos ARIMA e identificación de los parámetros más relevantes para implementar medidas de seguridad vial (Ochoa y Álvarez, 2021, p.3).

En 2020, Gómez Antonio y Espinoza César realizaron una encuesta titulada “Comportamiento y predicción de la mortalidad por tránsito: ECUADOR, 2000-2018”. En este estudio, la mortalidad vehicular, el riesgo de mortalidad y el riesgo de mortalidad tienden a disminuir en 2018, las previsiones para 2020 y 2030 tienen tendencias similares (García y Espinoza, 2020, p.2).

Trabajo “Análisis Estadístico de Series Temporales y Predicción de Información sobre Denuncias de Accidentes de Tránsito, Comisión Distrital PNP-Huaraz: 2007-2012”. Como resultado de la ejecución del modelo ARIMA de Rosario Libia y Torres, se concluyó que la mayoría de los accidentes de tránsito son predecibles y prevenibles (Rosario y Espinoza, 2013, p.6).

En el trabajo realizado por Yefry Quintero Avendaño llamado “CALIBRACIÓN DE UN MODELO ECONOMÉTRICO PARA EVALUAR LA FRECUENCIA EN ACCIDENTES DE TRÁNSITO CON PEATONES EN LA CIUDAD DE OCAÑA NORTE DE SANTANDER UTILIZANDO DISTRIBUCIÓN POISSON Y BINOMIAL NEGATIVO” analizado mediante Poisson y Binomial Negativo se puede concluir que el uso de los espacios peatonales no se respetan ni por parte de las personas, ni por parte de los flujos vehiculares que transitan en la ciudad (Avendaño et al., 2016, p.1).

Se han registrado accidentes de tránsito en la provincia de Chimborazo, lo cual fue investigado por la FGE de Chimborazo. Hay estadísticas de todos modos, pero hay un software que facilita el análisis y la predicción de accidentes de tránsito. Debido a esta escasez, las autoridades toman decisiones basadas en la experiencia adquirida durante muchos años de trabajo.

La estadística nos permite analizar datos y tomar decisiones. Una de las ramas principales es las series de tiempo, que permite hacer pronósticos a corto y largo plazo.

## **2.2. Referencia teórica**

### **2.2.1. Bases teóricas**

#### *2.2.1.1. Fiscalía general del estado*

La Constitución de la República establece que el Ministerio Público dirige las investigaciones y los procedimientos penales a solicitud de las partes y realiza las acciones públicas conforme a los principios de mínima oportunidad e intervención en los delitos penales, con especial consideración al interés y los derechos públicos. Poner al presunto autor ante un juez competente y sustentar finamente la acusación. Para llevar a cabo su misión, el Ministerio Público organiza y dirige investigaciones profesionales integrales, forenses y judiciales, incluidas las forenses civiles y policiales. Cumplir con otras obligaciones exigidas por la ley (Chiriboga, 2012, p.1).

#### *2.2.1.2. Accidentes de tránsito*

Un accidente automovilístico es un evento inesperado y es independiente de los factores humanos que cambian la ruta normal o planificada de viaje en la carretera. En particular, es un evento en el cual el daño a las personas o la propiedad es causado repentinamente por una influencia externa no deseada. Daño causado a una persona o material por una determinada ruta de movilización o transporte (la mayoría o en general) por factores externos e inesperados que contribuyen al peligro, la negligencia o el comportamiento irresponsable del conductor, tales como: falla mecánica adversa repentina, condiciones ambientales (terremotos o cambio climático repentino y repentino), cruce de venados en el tráfico o derrumbe de árboles debido a fuertes vientos en caminos o carreteras (Tipán, 2017, p.19).

#### *2.2.1.3. Derecho penal*

Según Gonzalo Rodríguez Murlo define el derecho penal como un medio de control social, así como la familia, la escuela, la religión, las normas sociales y morales. Lo que distingue al derecho penal como una forma de control social es que está altamente formalizado. Esto quiere decir que sus normas y acciones, prohibiciones y sanciones, los tipos de sanciones y su aplicación real se manifiestan por y sólo por ciertos cauces formales (Saggese et al., 2019, pp.27-29).

#### 2.2.1.4. *Infracciones de tránsito*

El Código Orgánico Integral Penal (COIP) define las infracciones de tránsito como actos u omisiones negligentes relacionados con el tránsito y la seguridad vial y se refiere a las infracciones de tránsito e infracciones de tránsito en el artículo 19 del propio reglamento (Denisse, 2016, pp.35-36).

#### 2.2.2. *Bases conceptuales*

##### 2.2.2.1. *Análisis exploratorio de datos*

##### ❖ *Regla de 80-20*

Es una herramienta utilizada para priorizar problemas o sus causas. La practicidad y utilidad general de los diagramas está respaldada por el llamado principio de Pareto, conocido como la "ley 80-20" o "Pocos vitales, muchos triviales, insignificantes", según este concepto, si el problema se debe a muchas causas, podemos decir que la causa 20% resuelve el problema 80% y la causa 80% solo resuelve el problema 20% (Juan y Gonzalez, 2004, p.79).

Pasos para la construcción de un diagrama de Pareto

1. Las áreas de problemas o mejoras que deben abordarse deben identificarse y retratarse para aclarar qué objetivos se persiguen. Use lo anterior para visualizar qué diagramas de Pareto pueden ayudarlo a identificar prioridades y comprender mejor su problema.
2. Con base en lo anterior, analice y decida el tipo de datos que necesita y los factores que pueden ser importantes para la organización en niveles. Luego se crea una hoja de verificación de adquisición de datos bien diseñada para identificar dichos factores.
3. Si desea extraer o recopilar información de informes anteriores, debe definir durante cuánto tiempo desea extraer los datos y quién es responsable de ello.
4. Al final de la recolección de datos, se crea una tabla para cuantificar la frecuencia de cada error, su porcentaje y otra información.
5. Se decide si el criterio utilizado para priorizar las diferentes categorías es la frecuencia, o si es necesario multiplicarlos por el costo o la intensidad correspondiente. Si es así, tienes que multiplicarlo. Luego se crea el gráfico.
6. Documente las referencias a los diagramas de Pareto, como título, duración, área de trabajo, etc.



7. Si se interpreta el diagrama de Pareto y hay categorías principales, se realiza un análisis de Pareto de segundo nivel para identificar los factores más influyentes.

#### ❖ **Análisis descriptivo univariante**

Consiste en un análisis descriptivo separándolo cada variable investigada que intenta describir la información de cada variable, en función de las medidas de tendencia central y varianza, además de la correspondiente gráfica (Verdoy et al., 2015, pp.10-27).

#### ❖ **Análisis de series de tiempo**

También conocida como series temporales, comparten una característica (serie univariante o escalar) y/o algunas características (serie multivariante o vectorial) de unidades que se pueden observar en diferentes momentos. Una secuencia de observaciones equidistantes (Mauricio, 2007, p.1).

La representación matemática de series temporales univariantes está dada por:

$y_1, y_2, \dots, y_N; (y_t)_{t=1}^N; (y_t: t = 1, \dots, N)$  donde  $y_t$  es la observación  $t$  –ésima, ( $1 \leq t \leq N$ ) de la serie y  $N$  es el número de observaciones que consta de la serie completa (el tamaño o la longitud de la serie). Las  $N$  observaciones  $y_1, y_2, \dots, y_N$  pueden recogerse en un vector columna  $y \equiv [y_1, y_2, \dots, y_N]'$  de orden  $N$  (Mauricio, 2007, pp.1-2).

#### ❖ **Metodología Box-Jenkins**

La metodología de Box & Jenkins se desarrolló en 1970 para facilitar el trabajo de los estadísticos al construir un modelo, explicar su estructura y predecir la evolución de esa serie en el futuro. Se puede pensar en una serie de tiempo como una serie de observaciones (datos) de variables recopiladas a intervalos regulares. Específicamente, la metodología Box-Jenkins es un de análisis estadístico diseñado para ajustarse a un conjunto de tipos de modelos especiales llamados ARIMA (autorregresive Integrated Moving Average) (Aldo, 2015, p.3).

Box & Jenkins han desarrollado un modelo estadístico de series temporales que tiene en cuenta las dependencias entre los datos. Cada observación en un momento determinado se modela sobre la base de valores anteriores. El análisis se basa en un modelo explícito. El modelo se conoce con el nombre genérico media móvil integrada autorregresiva (ARIMA), que se deriva de sus tres

componentes AR (autoregressive), I(integrated) y MA (moving averages) (Pelham Box & Jenkins , 1970, p. 1).

❖ **La metodología de Box & Jenkins se resume en cuatro fases:**

- 1ª Identificación:** Proponemos un modelo ARIMA. Es decir, para determinar los valores o parámetros (p, d, q).
- 2ª Estimación:** En esta etapa, debe estimar los parámetros de media móvil (q) y autorregresivos (p) contenidos en el modelo. Este cálculo se puede realizar utilizando un método simple de mínimos cuadrados o un método de estimación no lineal.
- 3ª Validación:** Puede haber más de un modelo ARIMA que haga lo mismo, por lo que el modelo debe representar adecuadamente el proceso que generó las observaciones, en dicho caso se debe aplicar el coeficiente de AIC o BIC.
- 4ª Predicción:** Tan pronto como el modelo correcto está disponible, se usa para hacer predicciones.

❖ **Procedimiento de análisis de datos**

1. **Recolección de datos:** Se recomienda tener más de 50 datos y trabajar en una serie entre 6 y 10 años.
2. **Representación gráfica:** Es muy útil tener una serie de gráficos para definir la estacionariedad y otras características como las tendencias y la varianza.
3. **Transformación previa de la serie:** Si la varianza de la serie no es constante, se requiere una transformación logarítmica la cual es útil para transformar distribuciones con sesgo positivo (con cola más larga hacia la derecha).
4. **Eliminación de la tendencia:** Puede ver si hay una tendencia observando el gráfico de la serie.
5. **Identificación del modelo:** Consiste en determinar o seleccionar el tipo de modelo más adecuado que reconozca los parámetros p, d y q.
6. **Estimación de los coeficientes del modelo:** Decidido el modelo, se procede a la estimación de sus parámetros dado que se trata de un procedimiento iterativo de cálculo, pueden sugerirse valores iniciales.
7. **Contraste de validez del modelo:** Se utilizan varios métodos para evaluar uno o más modelos seleccionados inicialmente. Pruebas de significación de parámetros, covarianzas entre estimadores, coeficientes de correlación, suma de errores al cuadrado, etc.

8. **Análisis detallado de los errores:** Se debe tener en cuenta las diferencias entre valores reales y estimados por el modelo para su valoración final.
9. **Selección del modelo:** En base a los resultados de pasos anteriores, se decide sobre el modelo adoptado.
10. **Predicción:** El modelo seleccionado se utilizará como fórmula inicial de predicción.

❖ **Modelos Autorregresivos AR(p)**

Un modelo autorregresivo AR describe una clase particular de proceso en que las observaciones en un momento dado son predecibles a partir de las observaciones previas del proceso más un término de error. El caso más simple es el ARIMA (1,0,0) o AR (1) o de primer orden, cuya expresión matemática es (Bonifaz, 2018, p.19):

$$AR(1) \equiv Y_t \equiv \Phi_1 Y_{t-1} + a_t \quad \text{Ecu. 1-2}$$

El proceso autorregresivo de orden p, representado por ARIMA(p,0,0) o simplemente por AR(p):

$$AR(p)Y_t \equiv \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + a_t \quad \text{Ecu. 2-2}$$

**Donde:**

$a_t$  = es ruido blanco.

Que puede ponerse, mediante el operador de cambio retroactivo B, en la forma:

$$(1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p) Y_t \equiv a_t \quad \text{Ecu. 3-2}$$

$$B^k(Y_t) \equiv Y_{t-k}$$

- ❖ Un proceso autorregresivo AR(p) es estacionario si las raíces del polinomio en B dado por:  $(1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p)$  caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación:  $Y^p - \Phi_1 Y^{p-1} - \Phi_2 Y^{p-2} - \dots - \Phi_{p-1} Y - \Phi_p = 0$  sean todas inferiores a uno en módulo.
- ❖ Un proceso autorregresivo siempre es invertible.

### ❖ Modelo de medias móviles Ma(q)

Un modelo de medias móviles MA describe una serie temporal estacionaria. En este modelo el valor actual puede predecirse a partir de la componente aleatoria de este momento y, en menor medida, de los impulsos aleatorios anteriores. El modelo ARIMA (0,0,1), también denotado por MA (1), viene dado por la expresión (Bonifaz, 2018, p.19):

$$Y_t = a_t - v_1 a_{t-1} \quad \text{Ecu. 4-2}$$

**Donde:**

$a_t$  = es ruido blanco.

El proceso de medias móviles de orden q, representado por ARIMA (0,0, q) o también por Ma(q), viene dado por la expresión:

$$Y_t = a_t - v_1 a_{t-1} - v_2 a_{t-2} - \dots - v_q a_{t-q} \quad \text{Ecu. 5-2}$$

- ❖ Un proceso de medias móviles es siempre estacionario.
- ❖ Un proceso de medias móviles Ma(q) es invertible si las raíces del polinomio en B definido por  $(1 - v_1 B - v_2 B^2 - \dots - v_q B^q)$  caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación  $y^q - \Phi_1 y^{q-1} - \Phi_2 y^{q-2} - \dots - \Phi_{q-1} y - \Phi_q = 0$  sean todas inferiores a uno en módulo.

### ❖ Modelos arma (p, q)

Una extensión natural de los modelos AR(p) y MA(q) es un tipo de modelos que incluyen tanto términos autorregresivos como de medias móviles y se definen como ARIMA (p, 0, q). Se representan por la ecuación (Bonifaz, 2018, p. 19):

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + a_t - v_1 a_{t-1} - v_2 a_{t-2} - \dots - v_q a_{t-q} \quad \text{Ecu. 6-2}$$

**Donde:**

$a_t$  = es ruido blanco.

Que puede ponerse de la forma:

$$Y_t - \Phi_1 Y_{t-1} + \Phi_1 Y_{t-2} + \dots + \Phi_p Y_{t-p} + a_t = a_t - v_1 a_{t-1} - v_2 a_{t-2} - \dots - v_q a_{t-q} \quad \text{Ecu. 7-2}$$

**Es decir:**

$$Y_t(1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p) = a_t(1 - v_1 B - v_2 B^2 - \dots - v_q B^q) \quad \text{Ecu. 8-2}$$

El proceso ARMA (p, q) es estacionario si lo es su componente autorregresiva, y es invertible si lo es su componente de medias móviles.

- ❖ Un modelo ARMA (p, q) es estacionario si las raíces del polinomio definido por  $(1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p)$  caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación:

$$Y^p - \Phi_1 Y^{p-1} - \Phi_2 Y^{p-2} - \dots - \Phi_{p-1} Y - \Phi_p \quad \text{Ecu. 9-2}$$

Sean todas inferiores a uno en módulo.

- ❖ Un modelo ARMA (p, q) es invertible si las raíces del polinomio en B definido mediante  $(1 - v_1 B - v_2 B^2 - \dots - v_q B^q)$  caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación:

$$x^q - \Phi_1 x^{q-1} - \Phi_2 x^{q-2} - \dots - \Phi_{q-1} x - \Phi_q = 0 \quad \text{Ecu. 10-2}$$

Sean todas inferiores a uno en módulo.

#### ❖ **Modelo ARIMA (p, d, q)**

La serie temporal se conoce como ARIMA (p, d, q), que es la media móvil integrado autorregresiva. Donde p es el número de términos autorregresivos, d es el número de veces que la serie necesita ser diferenciada y estacionaria, y q es el término promedio móvil reversible del caso más general ARIMA (p, d, q) (Bonifaz, 2018, p. 19).

$$\Phi_p(B^p) \Delta^d Y_t = \delta + \theta_q(B^q) a_t \quad \text{Ecu. 11-2}$$

**Donde:**

$\Phi_p(B^p)$  = es el polinomio autorregresivo de B de orden p.

$\theta_q(B^q)$  = es el polinomio de medias móviles de B de orden q.

$\Delta^d$  = representa el orden de integración (diferencias).

$a_t$  = es el error de predicción un periodo hacia adelante.

❖ **Identificación práctica del modelo**

Identificar un modelo significa utilizar los datos recogidos, así como cualquier información de cómo se genera la serie temporal objeto de estudio, para sugerir un conjunto reducido de posibles modelos, que tengan muchas posibilidades de ajustarse a los datos. Ante una serie temporal empírica, se deben encontrar los valores (p, d, q) más apropiados (Pelham Box y Jenkins , 1970, p.3).

- ❖ Si la serie temporal presenta una tendencia, lo primero que debe de hacerse es convertirla en estacionaria mediante una diferenciación de orden d. Una vez diferenciada la serie, una estrategia consiste en comparar los correlogramas de la función de autocorrelación (ACF) y la función de autocorrelación parcial (ACFP), proceso que suele ofrecer una orientación para la formulación del modelo orientativo.
- ❖ Los procesos autorregresivos presentan función de autocorrelación parcial (ACFP) con un número finito de valores distinto de cero. Un proceso AR(p) tiene los primeros p términos de la función de autocorrelación parcial distintos de cero y los demás son nulos.
- ❖ Los procesos de medias móviles presentan función de autocorrelación con un número finito de valores distintos de cero. Un proceso MA(q) tiene los primeros q términos de la función de autocorrelación distintos de cero y los demás son nulos.

❖ **Pasos para identificar un modelo de series temporales**

1. Decidir si  $y_t$  necesita ser transformada para eliminar la no estacionariedad en media p.
2. Determinación del grado d de diferenciación adecuado. En general, la falta de estacionariedad se manifiesta en que los coeficientes de la función de autocorrelación estimada tienden a decrecer muy lentamente.
3. Decidir los valores de (p, q), y si existe una componente estacional, decidir los órdenes de los operadores estacionales (p, q). Para esto se utiliza la autocorrelación y autocorrelación parcial, según el siguiente cuadro.

**Tabla 1-2:** Proceso MA(q), AR(p), ARIMA(p,d,q)

Proceso	Función de autocorrelación (ACF)	Función de autocorrelación parcial (ACFP)
MA(q)	Solo los q primeros coeficientes son significativos. El resto se anulan bruscamente (coef. 0 para retardo >q).	Decrecimiento rápido exponencial atenuado u ondas sinusoidales.
AR(p)	Decrecimiento rápido exponencial atenuado u ondas sinusoidales.	Solo los p primeros coeficientes son significativos. El resto se anulan bruscamente (coef. 0 para retardo >q).
ARIMA (p, d, q)	Comportamiento irregular en los retardos (1, ..., q) con q picos. Decrecimiento para retardos posteriores a q.	Decrece (aproximadamente con exponenciales atenuados y ondas sinusoidales).

Fuente: Pelham Box y Jenkins , 1970.

Realizado por: Inca E y Tierra M, 2022.

❖ **Modelos SARIMA (p, d, q) (P, D, Q)[S]**

Este modelo analiza las partes regulares y estacionales al mismo tiempo. La mayoría de las series de tiempo se analizan en un patrón anual, trimestral o mensual. Tipos de fenómenos analizando partes estacionales y normales o no estacionales (Pelham Box y Jenkins , 1970, pp.1-8).

Dichos modelos tienen las siguientes características:

- ❖ El modelo ARIMA (p, d, q) se encarga de modelar la parte regular o no estacional.
- ❖ El modelo ARIMA (P, D, Q) se encarga de modelar la parte estacional.

EL modelo SARIMA (p, d, q) (P, D, Q)<sub>S</sub> viene definido por:

$$\Phi_p(B^S)\Phi_p(B)\Delta^d\Delta_S^D Y_t = \theta_q(B)\theta_Q(B^S)a_t \quad \text{Ecu. 12-2}$$

**Donde:**

$\Phi_p(B^S)$  = es el polinomio autorregresivo de la parte estacional.

$\Phi_p(B)$  = es el polinomio autorregresivo de la parte no estacional.

$\theta_Q(B^S)$  = es el polinomio de medias móviles en la parte estacional.

$\theta_q(B)$  = es el polinomio de medias móviles de la parte no estacional.

$\Delta^d$  = representa el orden de integración en la parte no estacional.

$\Delta_S^D$  = es el orden de integración en la parte estacional.

$B^S$  = es el número de observaciones pertenecientes a un patrón estacional.

$a_t$  = es el error de predicción.

### ❖ Criterios de información

Existe varios criterios de información, los más utilizados son:

#### ❖ Criterio de información Akaike, AIC

Este criterio de información es útil para medir la calidad de los modelos de series temporales. Si tienes más de un modelo, los criterios AIC te ayudarán a elegir el modelo más adecuado (Montesinos, 2011, pp.9-10). Representación del coeficiente AIC es:

$$AIC = \frac{2k}{N} - \ln(\sigma_e^2) \quad \text{Ecu. 13-2}$$

#### Donde:

**k** = es el número de parámetros en el modelo estadístico.

$\sigma_e^2$  = es la varianza del error.

**N** = número de observaciones.

#### ❖ Criterio de Akaike corregido AICc

Este criterio se utiliza en orden cronológico para evaluar la calidad del modelo propuesto cuando el tamaño de la muestra es finito (M., 2021, p.10). La expresión viene dada por:

$$AICc = AIC + \frac{2k^2 + 2k}{N - k - 1} \quad \text{Ecu. 14-2}$$

#### ❖ Criterio de información Bayesiano BIC

También es un factor que mide la calidad del modelo (Montesinos, 2011, pp.10-12). Su expresión es:

$$BIC = N * \ln(\sigma_e^2) + k - \ln(N) \quad \text{Ecu. 15-2}$$

#### Donde:

**N** = es el número de observaciones.

$\sigma_e^2$  = es la varianza del error.

**k** = es el número de parámetros del modelo.



## ❖ Supuestos de verificación

### ❖ Normalidad

Para verificar si el modelo es válido, debe realizar una prueba de normalidad del error. Se suele utilizar el test de Jarque-Bera, Tierra (2020, p.33) esta ocupa los coeficientes de asimetría y kurtosis, está definida por:

$$H_0: \text{Existe normalidad}$$

$$H_1: \text{No existe normalidad}$$

$$JB = N \left[ \frac{A^2}{6} + \frac{(K-3)^2}{24} \right] \quad \text{Ecu. 16-2}$$

#### Donde:

$N$  = número de observaciones.

$A$  = coeficiente de asimetría.

$K$  = coeficiente de kurtosis.

### ❖ Independencia

Una de las pruebas utilizadas para demostrar la independencia es la prueba Ljung-Box utilizada para el modelo de ARIMA (Sánchez y Martínez, 2019, p.28).

$$H_0: \text{Existe independencia}$$

$$H_1: \text{No existe independencia}$$

$$LB = N(N + 2) \sum_{k=1}^m \left( \frac{\hat{P}_k^2}{N-k} \right) \sim X_{(m)}^2 \quad \text{Ecu. 17-2}$$

#### Donde:

$N$  = número de observaciones.

$\hat{P}_k^2$  = es la autocorrelación de la muestra en el retraso  $k$ .

$k$  = es el número de retrasos existentes.

### ❖ Homocedasticidad

Para probar homocedasticidad se utiliza la prueba de White el cual está definido por (Bonifaz, 2018, p.25):

$$\begin{aligned}
H_0: & \text{Existe homocedasticidad} \\
H_1: & \text{No existe homocedasticidad} \\
\sigma_t^2 = & h(Z_t)\alpha_0 + \alpha_1 Z_{1t} + \dots + \alpha_p Z_{pt} + \varepsilon_t
\end{aligned}
\tag{Ecu. 18-2}$$

❖ **Criterio de Decisión para las pruebas de hipótesis**

Si  $p \leq \alpha$  se rechaza la hipótesis nula,  $H_0$ .

Si  $p > \alpha$  no se rechaza la hipótesis nula,  $H_0$ .

**Donde:**

$\alpha$  es el nivel de significancia predefinido.

$p$  es p-valor de la prueba de hipótesis.

❖ **Modelo Lineal Generalizado (MLG)**

Modelo Lineal Generalizado (MLG) como el modelado estadístico son herramientas metodológicas que permiten codificar todas las situaciones de análisis dentro de un mismo esquema general. Obviamente, esto facilita el aprendizaje de nuevos modelos de análisis porque se trata simplemente de contemplarlos como casos particulares de un modelo más general ya conocido, el Modelo Lineal (ML). Veamos las relaciones entre ambos.

**Tabla 2-2:** Principales modelos lineales generalizados

Naturaleza de la variable respuesta	Componentes		Función de enlace	Módulo lineal	
	Sistemático	Aleatorio			
Numérica cuantitativa	Numérica	Normal	Identidad	Regresión lineal	ML
	Catagórica	Normal	Identidad	ANOVA o de diseño experimental	
	Mixta	Normal	Identidad	ANCOVA o de diseño experimental con variables concomitentes	
<b>Catagórica binaria</b>	Mixto	Binomial (1)	Logit	Regresión logística	MLG
No agrupada		Bernulli			
Agrupada (Frecuencia)	Catagórico	Binomial(n)	Logit	Análisis logit	
			Probit	Regresión probit	
<b>Catagórica politómica</b>	Mixto	Multinomial	Logit generalizado	Regresión logística multinomial	
No agrupada					
Agrupada (Frecuencia)	Catagórico	Multinomial	Logit generalizado	Análisis logit multinomial	
<b>Recuento</b>	Mixto	Poisson	Logarítmica	Regresión de Poisson	
<b>Frecuencia</b>	Catagórico	Poisson	Logarítmica	Análisis loglineal	

Fuente: González y Soler, 2011.

Realizado por: Inca E y Tierra M, 2022.

En la Tabla 2-2 puede observarse cómo el Modelo Lineal es el caso más elemental del Modelo Lineal Generalizado. Las coincidencias y las diferencias entre uno y otro hacen posible, en el caso del MLG, un tratamiento matemático y estadístico adecuado a los niveles de medida de las variables que contiene.

El MLG tiene las siguientes funciones:

$$\begin{aligned}
 y_i &= \sum_j \beta_j X_{ij} + \varepsilon_i \\
 \mu_i &= E(Y_i) \\
 \eta_i &= \sum_j \beta_j X_{ij} \\
 \eta_i &= g(\mu_i)
 \end{aligned}
 \tag{Ecu. 19-2}$$

**Donde:**

$y_i$  = vector de la variable respuesta.

$X_{ij}$ : = matriz de variables predictoras de covariables.

$\beta_j$  = vector de parámetros.

$\eta_i$  = vector del predictor lineal.

#### ❖ Distribución de probabilidad de Poisson

La distribución de Poisson es un buen modelo probabilístico para evaluar la probabilidad de que ocurran eventos en intervalos de tiempo, longitud, área o volumen (Martínez & Marí, 2022, pp. 3-4), su fórmula es:

$$f(y, \mu) = \frac{\mu^y e^{-\mu}}{y!}, \text{ para } y = 0, 1, 2, \dots ; \mu > 0
 \tag{Ecu. 20-2}$$

La distribución de POISSON se obtiene, dada  $y$  es una variable aleatoria con distribución binomial, con parámetros  $n$  y  $p$  (Arboccó, 2005, pp.39-40), dada por:

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y}, \text{ para } y = 0, 1, \dots, n \text{ haciendo que } n \rightarrow +\infty \text{ y } p \rightarrow 0
 \tag{Ecu. 21-2}$$

De esta manera  $np = \mu$  permanezca constante, el límite de la distribución es una Poisson con parámetro  $\mu$ .

La regresión de Poisson asume:

- ❖ Tiene una distribución de Poisson.
- ❖ La varianza de  $Y = y$  dado  $x$  es.

$$\lambda = E(y|x) = e^{x'\beta} \quad \text{Ecu. 22-2}$$

- ❖ Por lo tanto,  $\log(E(y|x)) = x'\beta$  (el modelo de regresión de Poisson se refiere a veces como el modelo log-lineal).
- ❖ La probabilidad viene dada por  $\Pr(y|x) = \frac{e^{yx'\beta} e^{-e^{x'\beta}}}{y!}$ .
- ❖ Para una muestra de  $N$  observaciones.

$$\Pr(y_1, \dots, y_N) = \prod_{i=1}^N \frac{e^{y_i x_i' \beta} e^{-e^{x_i' \beta}}}{y_i!} \quad \text{Ecu. 23-2}$$

#### ❖ Aproximación a la distribución normal

La distribución de Poisson tiende a una normal a medida que aumenta su media. Se puede observar McCullagh y Nelder (1991, p.105) que cuando  $\mu \rightarrow \infty$ .

$$\frac{(y-\mu)^{\frac{1}{2}}}{\mu^2} \sim N(0,1) \quad \text{Ecu. 24-2}$$

Esta distribución se puede visualizar graficando  $P(y = \mu)$  vs  $\mu$ .

#### ❖ Modelo de regresión de Poisson

El modelo de regresión de Poisson (MRP) se deriva de la función de conexión GLM y parametriza las relaciones entre medias,  $m$  y predictores. La idea básica de este modelo es que la información de la variable predictora ( $x$ ) está relacionada con la razón o sensibilidad de la respuesta al aumento o disminución de los conteos ( $Y$ ) (Arboccó, 2005, pp.41-42).

El PRM tiene la siguiente función:

$$\log \mu_i = \eta = \beta^T x_i \text{ donde } i = 1, 2, \dots, n, \quad n \text{ es el número total de observaciones}$$

Existen tres componentes de MRP:

- ❖ Componente aleatorio: La variabilidad de  $Y$  no explicada por  $\eta$  sigue una distribución de Poisson

$$\varepsilon \sim \text{Poisson}(\mu)$$

- ❖ Componente sistemática: El predictor lineal que expresa la combinación lineal de las variables explicativas y proporciona el valor predicho es:

$$\eta = \beta^T x_i$$

- ❖ Función de enlace: aquella que relaciona  $\eta$  con  $\mu$  es:

$$g(\mu_i) = \log(\mu_i)$$

### ❖ Contraste de la bondad del ajuste

La prueba de ajuste, que es el nivel de significación  $\alpha$  de la hipótesis nula  $H_0$  frente a la hipótesis alternativa  $H_1$  de que la probabilidad especificada es incorrecta, se basa en una regla de decisión.

(Williams, 2015, p.374):

$$\text{Rechazar } H_0 \text{ si } \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} > X_{K-1, \alpha}^2, \quad \text{donde } K \text{ es el número de parámetros}$$

**Donde:**

$X_{K-1, \alpha}^2$  = es el número tal que

$$P(X_{K-1}^2 > X_{K-1, \alpha}^2) = \alpha \quad \text{Ecu. 25-2}$$

La variable aleatoria  $X_{K-1, \alpha}^2$  sigue una distribución ji-cuadrado con  $(K - 1)$  grados de libertad.

## CAPÍTULO III

### 3. MARCO METODOLÓGICO

#### 3.1. Enfoque de investigación

Este estudio aplica un enfoque de investigación cuantitativa para analizar el número de accidentes de tránsito cada mes utilizando datos de FGE de Chimborazo, y para identificar los accidentes de tránsito más típicos utilizando la regla 80-20. Se realizó un análisis exploratorio de datos de accidentes de tránsito univariado para obtener una matriz manejable y permitir la aplicación de métodos estadísticos. Aplicar el método Box-Jenkins del modelo ARIMA. Finalmente, se usó el modelo POISSON para los siguientes propósitos: Crear un pronóstico de accidentes a partir de los datos de tránsito mensuales del año 2022 y comparar los de modelos MLG: ARIMA y Poisson para seleccionar el mejor modelo.

#### 3.2. Nivel de investigación

Esta investigación es predictiva y aplica la metodología Box & Jenkins, los criterios de medidas de ajuste de Akaike (AIC) y métricas de conformidad Bayesiano (BIC), para construir modelos ARIMA  $(p, d, q)(P, D, Q)_S$  y Poisson. También se prueba y compara los supuestos de normalidad, homocedasticidad, independencia y compararlos con el propósito de elegir el mejor modelo y predecir el número de accidentes de tránsito mensuales en la provincia de Chimborazo para el año 2022.

#### 3.3. Diseño de la investigación

##### 3.3.1. *Según la manipulación o no de la variable independiente*

Este proyecto de investigación para la manipulación de la variable independiente es no experimental ya que la información fue obtenida de fuentes secundarias, dichas variables no pueden ser controladas por factores externos como: velocidad, clima, responsabilidad del conductor, consumo de drogas, desconocimiento de las señales de tránsito, etc.

### **3.3.2. Según las intervenciones en el trabajo de campo**

Este proyecto de investigación es a largo plazo ya que obtenemos datos mensuales del número de accidentes de tránsito del periodo Enero 2015 – Diciembre 2021 y se realiza las predicciones para el año 2022.

Este estudio utilice el software R con una base teórica para series temporales, aplica el principio 80-20 y utiliza la librería qcc para crear un gráfico de pared que identifica los accidentes de tránsito más comunes.

Se realiza un análisis exploratorio de datos univariado mediante un gráfico de caja para comprobar la presencia de datos atípico. Desde enero de 2015, los datos se configuran en una serie de tiempo mediante la función ts. En este caso la frecuencia 12 porque el dato es mensual. Luego se aplica la prueba de Dickey-Fuller con una verificación preliminar para ver si la serie temporal es estacionaria, graficando la serie original, descomponiéndolas y utilizando un gráfico de autocorrelación con la función de trazado. La serie temporal no es estable. De hecho, la diferencia se hace con el comando diff.

Para el análisis de Series de Tiempo aplica la metodología Box & Jenkins siguiendo estos pasos: identificar el modelo, se puede utilizar primero la función "auto.arima()" para encontrar p y q del modelo adecuado, estimación de los coeficientes del modelo, contraste de validez del modelo se realiza la gráfica de los correlogramas de los residuos para comprobar que existe estacionariedad y se ajusta bien el modelo identificado, análisis detallado de los errores, selección del modelo para corroborar se pueda realizar un diagnóstico para saber si el modelo es adecuado con la función "tsdiag()", para predecir se utiliza la función forecast, en este caso se pronostica los 12 siguientes meses para el año 2022.

### **3.4. Tipo de estudio**

Se aplica como investigación de campo debido a que es el dato de número real de accidentes de tránsito que se registran todos los meses y es necesario investigar cómo se presenta sin manipular variables.



### **3.5. Población y planificación, selección y cálculo del tamaño de la muestra**

#### **3.5.1. Población**

La población a la que se realiza el estudio corresponde a los accidentes de tránsito mensuales que ingresa a la FGE de la provincia de Chimborazo.

#### **3.5.2. Tamaño de la muestra**

El tamaño de la muestra corresponde a los accidentes de tránsito de los 72 meses en el período enero 2015 a diciembre 2021.

### **3.6. Métodos, técnicas e instrumentos de investigación**

#### **3.6.1. Métodos de investigación**

##### **❖ Método Inductivo**

Según el tipo de inferencia es inductivo, se realizará predicciones de los accidentes mensuales de tránsito que ocurrirán en el año 2022 y de esta manera poder tomar una decisión.

##### **❖ Método Analítico**

Este método requiere analizar series de tiempo, cuyo objetivo es comprender sus características claves y probar la hipótesis ya planteada, con el propósito de predecir el número de accidentes de tránsito mensuales en la provincia de Chimborazo para el año 2022.

#### **3.6.2. Técnicas de investigación**

##### **❖ Técnica de recolección de datos**

A través de un oficio aprobado por el coordinador de la Carrera de Estadística de la Escuela Superior Politécnica de Chimborazo dirigido a la FGE de la provincia de Chimborazo se obtuvo los datos de accidentes de tránsito desde Enero 2015 hasta Diciembre 2021, las cuales fueron enviadas al correo electrónico (ANEXO A).

### **3.6.3. Instrumentos de investigación**

Mediante las bases de datos proporcionados por la FGE presentamos las variables que se utilizaron en el estudio.

**Tabla 1-3:** Operacionalización de variables

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
DM	Daños materiales	Cuantitativa	Razón	$[0, +\infty[$
DMPAT_+6	Daños materiales la persona que como consecuencia del accidente de tránsito cause solamente daños materiales cuyo costo de reparación exceda los seis salarios básicos unificados del trabajador en general.	Cuantitativa	Razón	$[0, +\infty[$
DMPCVLLS_TD	Daños materiales la persona que conduzca un vehículo en el lapso en que la licencia de conducir se encuentre suspendida temporal o definitivamente.	Cuantitativa	Razón	$[0, +\infty[$
DMPEEE_EPP	Daños materiales si la persona se encontrare en estado de embriaguez o bajo los efectos de sustancias estupefacientes, psicotrópicas o preparados que las contenga.	Cuantitativa	Razón	$[0, +\infty[$
LCA_DA+9	Lesiones causadas por accidente de tránsito si produce a la víctima un daño, incapacidad o enfermedad de nueve a treinta días.	Cuantitativa	Razón	$[0, +\infty[$
LCAT_+90	Lesiones causadas por accidente de tránsito si produce a la víctima una grave enfermedad o una disminución de sus facultades físicas o mentales o una incapacidad o enfermedad, que, no siendo permanente, supere los noventa días.	Cuantitativa	Razón	$[0, +\infty[$
LCAT_DI48	Lesiones causadas por accidente de tránsito si como resultado de las lesiones se produce en la víctima un daño, enfermedad o incapacidad de cuatro a ocho días.	Cuantitativa	Razón	$[0, +\infty[$

LCAT_EPP	Lesiones causadas por accidente de tránsito en los delitos de tránsito que tengan como resultado lesiones, si la persona conduce el vehículo en estado de embriaguez o bajo los efectos de sustancias estupefacientes, psicotrópicas o preparados que las contengan.	Cuantitativa	Razón	$[0, +\infty[$
LCAT_HIP	Lesiones causadas por accidente de tránsito si produce a la víctima enajenación mental, pérdida de un sentido o de la facultad del habla, inutilidad para el trabajo, incapacidad permanente, pérdida o inutilización de algún órgano o alguna grave enfermedad transmisible e incurable.	Cuantitativa	Razón	$[0, +\infty[$
LCPATR	Lesiones causadas por accidente de tránsito	Cuantitativa	Razón	$[0, +\infty[$
LCPT_31_90	Lesiones causadas por accidente de tránsito si produce a la víctima un daño, incapacidad o enfermedad de treinta y uno a noventa días.	Cuantitativa	Razón	$[0, +\infty[$
LPD_IE+1	Lesiones si produce a la víctima un daño, incapacidad o enfermedad de treinta y uno a noventa días.	Cuantitativa	Razón	$[0, +\infty[$
MC	Muerte culposa	Cuantitativa	Razón	$[0, +\infty[$
MCPCE	Muerte causada por conductor en estado de embriaguez o bajo los efectos de sustancias estupefacientes, psicotrópicas o preparados que las contengan	Cuantitativa	Razón	$[0, +\infty[$

MCRDAIPE_2	Muerte culposa cuando el resultado dañoso es producto de acciones innecesarias, peligrosas e ilegítimas tales como: 2. conocimiento de las malas condiciones mecánicas del vehículo.	Cuantitativa	Razón	$[0, +\infty[$
MCRDAIPE_4	Muerte culposa cuando el resultado dañoso es producto de acciones innecesarias, peligrosas e ilegítimas tales como: 4. haber conducido el vehículo más allá de las horas permitidas por la ley o malas condiciones físicas de la o el conductor	Cuantitativa	Razón	$[0, +\infty[$
MCRDPAI_1	Muerte culposa cuando el resultado dañoso es producto de acciones innecesarias, peligrosas e ilegítimas tales como: 1. exceso de velocidad	Cuantitativa	Razón	$[0, +\infty[$
MCRDPAI_5	Muerte culposa cuando el resultado dañoso es producto de acciones innecesarias, peligrosas e ilegítimas tales como: 5. inobservancia de leyes, reglamentos, regulaciones técnicas u órdenes legítimas de las autoridades o agentes de tránsito	Cuantitativa	Razón	$[0, +\infty[$

Realizado por: Inca E. y Tierra M, 2022.

**Tabla 2-3:** Operacionalización de objetivos

OBJETIVO GENERAL	OBJETIVOS ESPECIFICOS	CONCEPTO	INDICADOR	INSTRUMENTO
Predecir el número de accidentes de tránsito mensuales en la provincia de Chimborazo mediante MLG: modelos ARIMA y compararlos con modelos de POISSON en el periodo enero 2015- diciembre 2021.	Aplicar la regla del 80-20 para identificar los accidentes de tránsito más representativos.	Es una herramienta que se utiliza para priorizar los problemas o las causas que los genera.	Realizar la regla 80-20 mediante un diagrama de Pareto de los datos proporcionados.	Base de datos de los accidentes de tránsito de la Fiscalía Provincial de Chimborazo.
	Realizar un análisis exploratorio de datos univariante de los accidentes de tránsito para obtener una matriz manejable y poder aplicar métodos estadísticos.	Hace referencia a ver el comportamiento descriptivo del número de accidentes de tránsito y obtener una matriz manejable para aplicar métodos estadísticos,	Tabla de frecuencias, Diagrama de barras, Boxplot, matriz de correlación.	Librerías del entorno computacional R
	Verificar los supuestos de los modelos ARIMA y POISSON, para poder elegir el mejor modelo.	Verificación de hipótesis para la comprobación de supuestos.	Supuestos de normalidad e independencia, homocedasticidad.	Librerías del entorno computacional R
	Construir modelos ARIMA y POISSON con el propósito de realizar predicciones de los accidentes de tránsito.	Hace referencia a la predicción del número de accidentes mensuales de tránsito mediante modelos estadísticos ARIMA y POISSON.	Ecuaciones de los modelos ARIMA y POISSON.	Librerías específicas de entorno computacional R como: auto.arima
	Comparar los dos tipos de modelos a fin de seleccionar el mejor.	Identificar el modelo óptimo que cumplan los supuestos de verificación.	Coefficientes para comparar los modelos ARIMA y POISSON.	Librerías específicas del entorno computacional R Como: AIC(), BIC().

Realizado por: Inca E. y Tierra M, 2022.

**Tabla 3-3:** Matriz de consistencia

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES
<p><b>PROBLEMA GENERAL:</b> ¿Cuántos accidentes de tránsito ingresarán mensualmente a la FGE de la Provincia de Chimborazo?</p>	<p><b>OBJETIVO GENERAL:</b> Predecir el número de accidentes de tránsito mensuales en la provincia de Chimborazo mediante MLG: modelos ARIMA y compararlos con modelos de POISSON en el periodo enero 2015 – diciembre 2021.</p> <p><b>OBJETIVOS ESPECÍFICOS:</b></p> <ul style="list-style-type: none"> <li>• Aplicar la regla del 80-20 para identificar los accidentes de tránsito más representativos.</li> <li>• Realizar un análisis exploratorio de datos univariante de los accidentes de tránsito para obtener una matriz más manejable y poder aplicar métodos estadísticos.</li> <li>• Verificar los supuestos de los modelos ARIMA y POISSON, para poder elegir el mejor modelo.</li> <li>• Construir modelos de ARIMA y POISSON con el propósito de realizar predicciones de los accidentes de tránsito.</li> <li>• Comparar los dos tipos de modelos a fin de seleccionar el mejor.</li> </ul>	<p><b>HIPOTESIS GENERAL:</b> Mediante el análisis de modelos lineales generalizados, MLG se puede obtener resultados relevantes o significativos de los accidentes de tránsito.</p>	<p>VARIABLE INDEPENDIENTE mes. VARIABLE DEPENDIENTE Número de accidentes de tránsito.</p>

Realizado por: Inca E y Tierra M, 2022.

## CAPÍTULO IV

### 4. MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

Aplicación de la regla 80-20 mediante el diagrama de Pareto, para obtener los accidentes de tránsito más comunes.

#### 4.1. Diagrama de Pareto

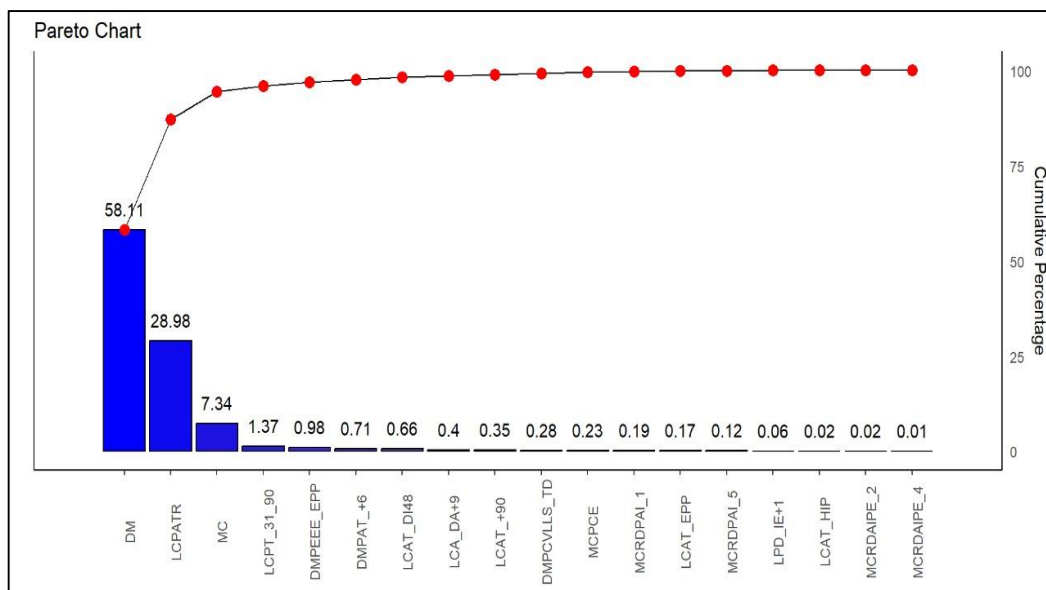
Para seleccionar las variables más comunes se aplica la regla del 80-20, donde se utiliza el diagrama de Pareto.

**Tabla 1-4:** Regla del 80-20

Accidentes	Frecuencia	Porcentaje	Porcentaje acumulado
DM	5256	58,11	58,11
DMPAT_+6	64	0,71	58,82
DMPCVLLS_TD	25	0,28	59,09
DMPEEE_EPP	89	0,98	60,08
LCA_DA+9	36	0,40	60,48
LCAT_+90	32	0,35	60,83
LCAT_DI48	60	0,66	61,49
LCAT_EPP	15	0,17	61,66
LCAT_HIP	2	0,02	61,68
LCPATR	2621	28,98	90,66
LCPT_31_90	124	1,37	92,03
LPD_IE+1	5	0,06	92,08
MC	664	7,34	99,43
MCPCE	21	0,23	99,66
MCRDAIPE_2	2	0,02	99,68
MCRDAIPE_4	1	0,01	99,69
MCRDDAI_1	17	0,19	99,88
MCRDDAI_5	11	0,12	100,00
Total	9045	100,00	

Realizado por: Inca E y Tierra M, 2022.





**Ilustración 1-4:** Diagrama de Pareto

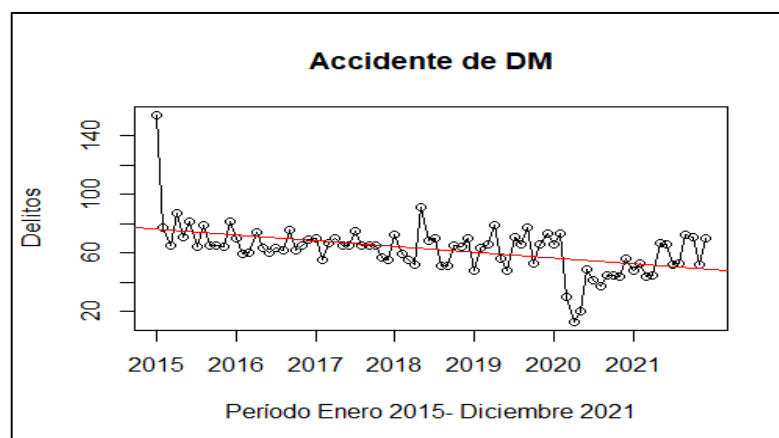
Realizado por: Inca E y Tierra M, 2022.

Aplicando la regla del 80 – 20 se obtuvo que la variable DM (Daños Materiales) que alcanza el 58.11% y la variable LCPATR (lesiones causadas por accidentes de tránsito) con el 28.98%, siendo estos los accidentes de tránsito más significativos que ingresan a la Fiscalía Provincial de Chimborazo.

#### 4.2. Análisis exploratorio de datos

##### 4.2.1. Daños materiales (DM)

##### ❖ Serie de Tiempo DM

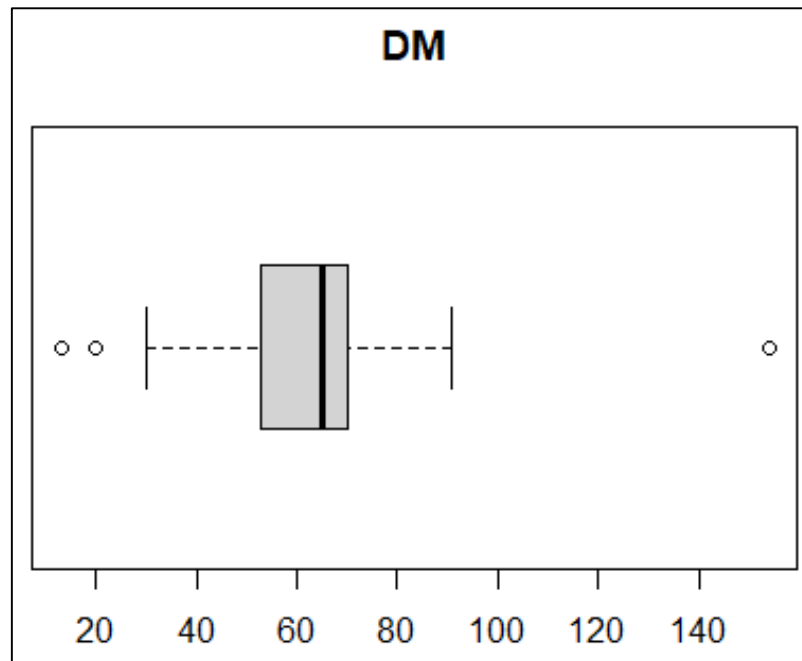


**Ilustración 2-4:** Serie de tiempo de DM

Realizado por: Inca E y Tierra M, 2022.

Se observa que en el año 2015 existe un incremento del número de accidentes de tránsito por DM y en el año 2020 en el mes de febrero descende el número de accidentes.

❖ **Diagrama de caja DM**



**Ilustración 3-4:** Boxplot de DM

Realizado por: Inca E y Tierra M, 2022.

Mediante la función `b$out` en donde `b` es el nombre asignado para el diagrama de caja, se observa que existe tres datos atípicos, los cuales se muestran en la tabla 2-4:

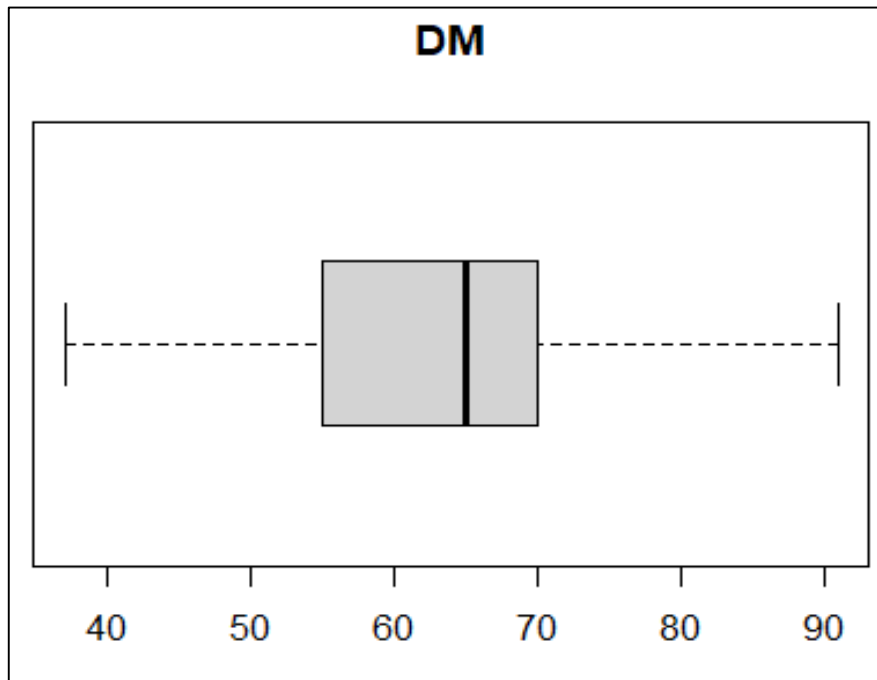
**Tabla 2-4:** Atípicos de DM

Fecha	Atípico (Frecuencia de DM)
Enero 2015	154
Abril 2020	13
Mayo 2020	20

Realizado por: Inca E y Tierra M, 2022.

Al encontrar los datos atípicos se procede a imputar por el promedio.

❖ **Diagrama de caja sin datos atípicos de DM**



**Ilustración 4-4:** Boxplot sin datos atípicos de DM

Realizado por: Inca E y Tierra M, 2022.

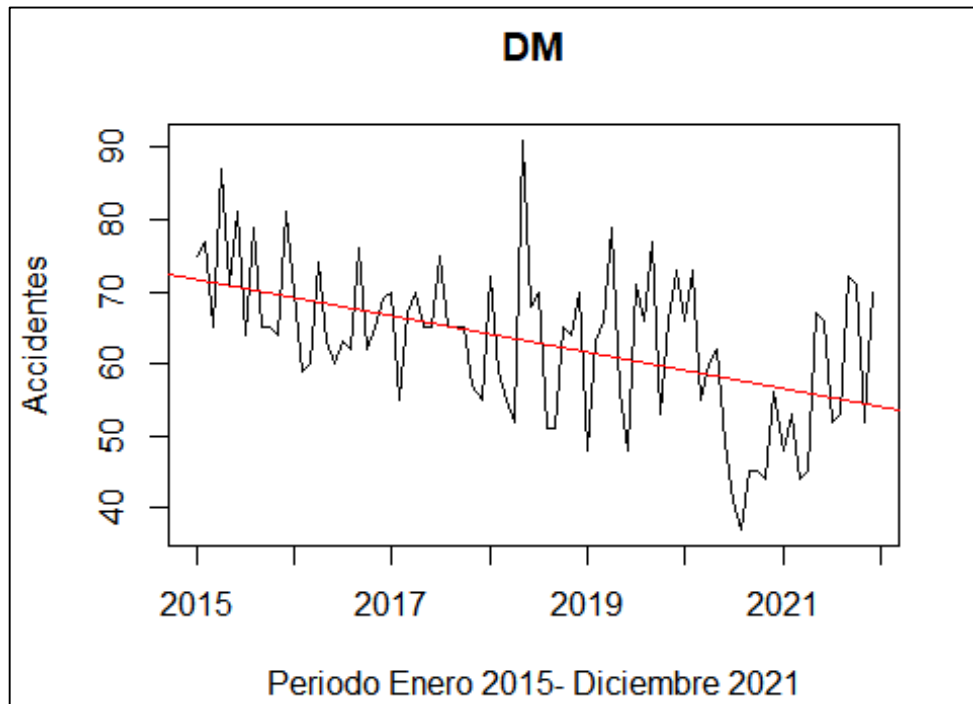
En el diagrama de cajas sin datos atípicos se observa que tiene asimetría negativa ya que está sesgada a la izquierda es decir existe una mayor variación en los datos, se concentran en la parte derecha de la distribución, mediante la función  $quantile()$  podemos observar los cuartiles de la ilustración, son los siguientes:

**Tabla 3-4:** Cuartiles de DM

Cuartil	Valor
Q1	55
Q2	65
Q3	70

Realizado por: Inca E y Tierra M, 2022.

❖ Sin datos atípicos DM



**Ilustración 5-4:** Sin datos atípicos DM

Realizado por: Inca E y Tierra M, 2022.

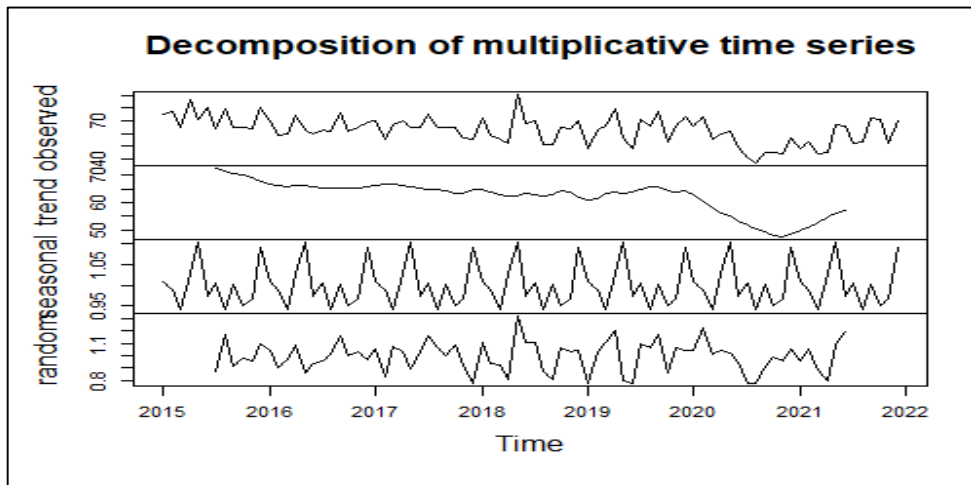
Se observa que existe tendencia negativa, es decir decrece a lo largo de la serie por lo cual las series es no estacionarias.

**Tabla 4-4:** Serie de tiempo de DM

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
2015	75	77	65	87	71	81	64	79	65	65	64	81
2016	70	59	60	74	63	60	63	62	76	62	65	69
2017	70	55	67	70	65	65	75	65	65	65	57	55
2018	72	59	55	52	91	68	70	51	51	65	64	70
2019	48	63	66	79	56	48	71	66	77	53	66	73
2020	66	73	55	60	62	49	41	37	45	45	44	56
2021	48	53	44	45	67	66	52	53	72	71	52	70

Realizado por: Inca E y Tierra M, 2022.

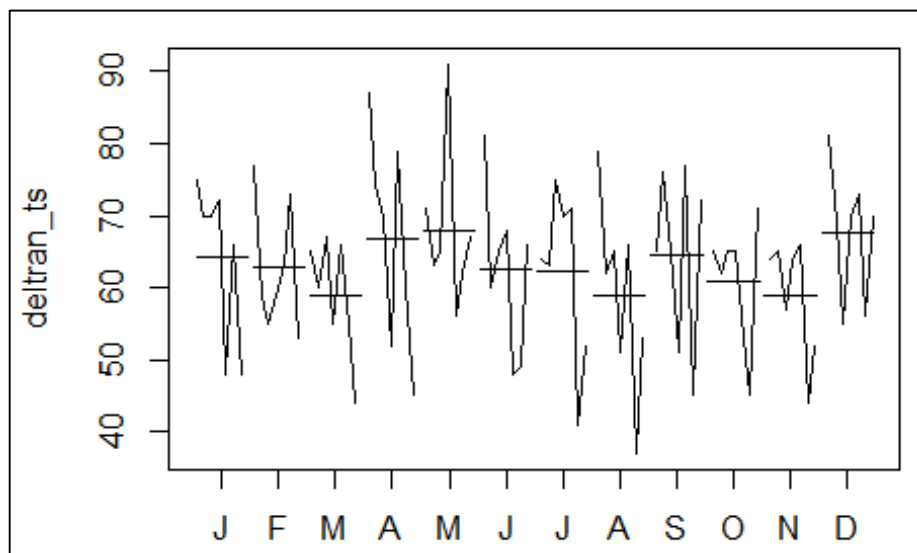
## ❖ Descomposición de la serie de tiempo DM



**Ilustración 6-4:** Descomposición de DM

Realizado por: Inca E y Tierra M, 2022.

Se puede apreciar que existe una tendencia decreciente ya que los datos van decreciendo a lo largo de la serie, existe estacionalidad dado que existe un patrón de un periodo a otro a lo largo del tiempo y existe aleatoriedad en la serie es decir no existe un patrón específico, por lo cual se aplica el modelo ARIMA con estacionalidad o sea es SARIMA.



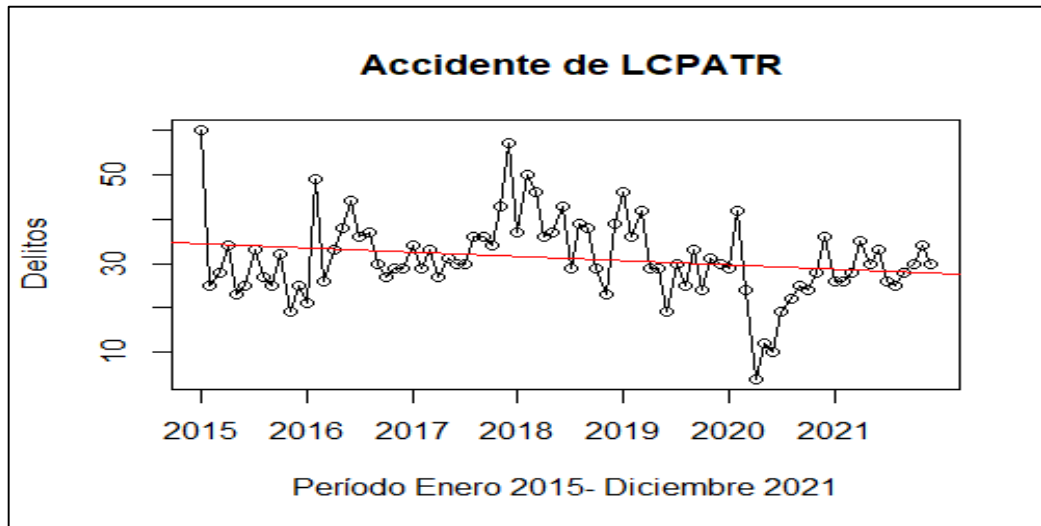
**Ilustración 7-4:** Gráfico mensual de la Serie de Tiempo de DM

Realizado por: Inca E y Tierra M, 2022.

Se aprecia que en el mes de mayo y diciembre existe mayor número de accidentes de tránsito por daños materiales, pero en el mes de Marzo, Agosto y Noviembre no existe muchos accidentes de tránsito por Daños Materiales.

#### 4.2.2. Lesiones causadas por accidentes de tránsito (LCPATR)

##### ❖ Serie de Tiempo LCPATR

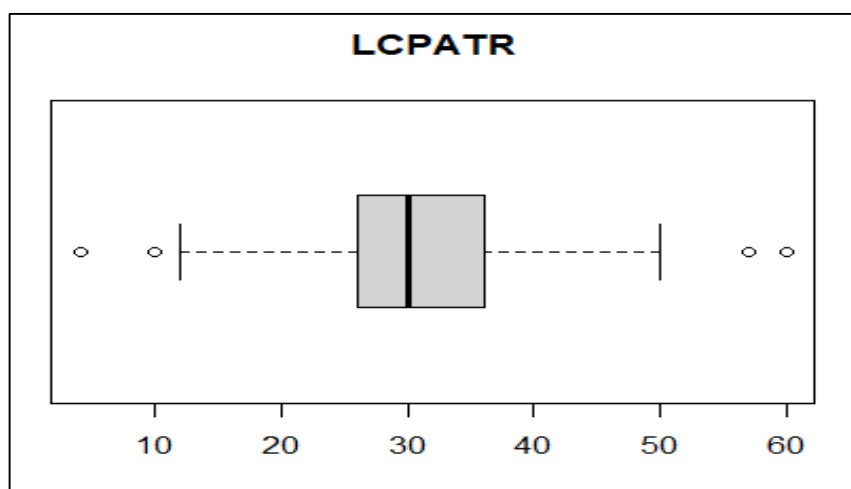


**Ilustración 8-4:** Serie de tiempo LCPATR

Realizado por: Inca E y Tierra M, 2022.

Existe datos atípicos lo cual será remplazada por el promedio de los meses a que corresponde estos datos, se observa que a principios del año 2020 tiene una fluctuación decreciente de los accidentes de tránsito, dado que tiene cambios bruscos lo que se puede decir que no sigue un patrón.

##### ❖ Diagrama de caja LCPATR



**Ilustración 9-4:** Boxplot de LCPATR

Realizado por: Inca E y Tierra M, 2022.

Mediante la función `c$out` en donde `c` es el nombre asignado para el diagrama de caja, se puede observar que existen cuatro datos atípicos los cuales se muestran en la tabla 5-4:

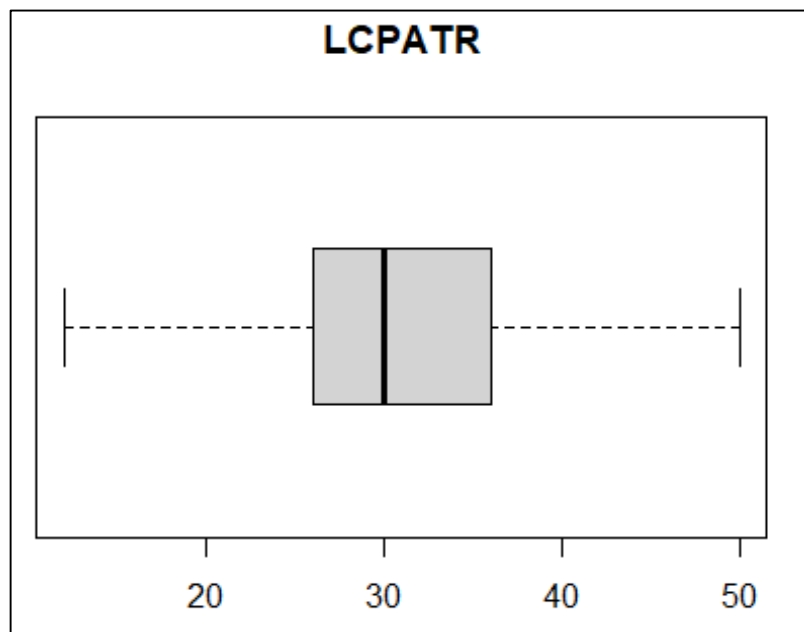
**Tabla 5-4:** Atípicos de LCPATR

Fecha	Atípico
Enero 2015	60
Diciembre 2017	57
Abril 2020	4
Junio 2020	10

Realizado por: Inca E y Tierra M, 2022.

Al encontrar los datos atípicos se procede a imputar por el promedio.

❖ **Diagrama de cajas sin datos atípicos LCPATR**



**Ilustración 10-4:** Boxplot sin datos atípicos LCPATR

Realizado por: Inca E y Tierra M, 2022.

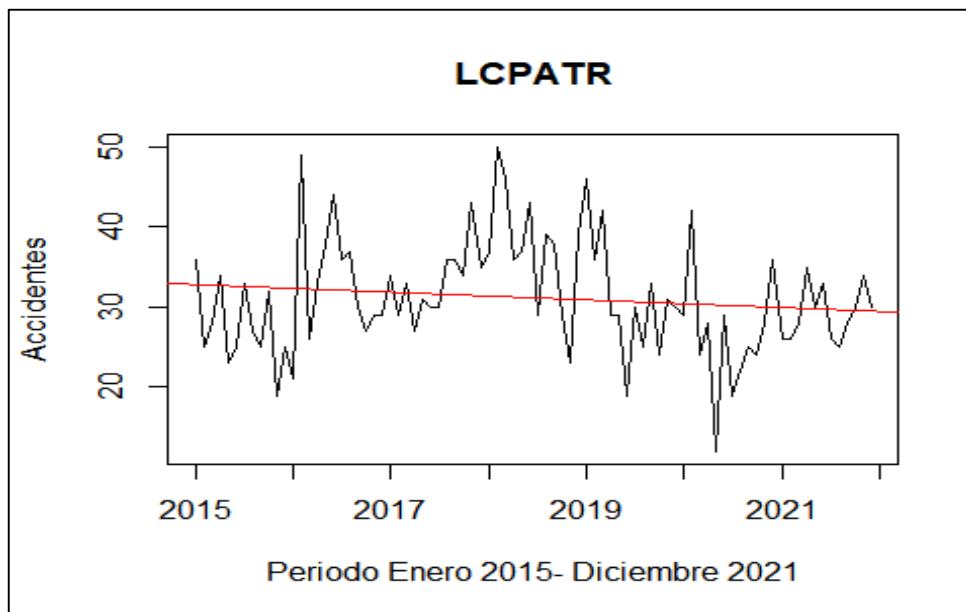
Se observa en el diagrama de cajas los datos no se encuentran agrupados en la mediana, los datos se distribuyen de manera simétrica ya que esta sesgada a la derecha es decir existe una mayor variación en los datos, se concentran en la parte izquierda de la distribución, además, mediante la función `quantile()` podemos obtener los cuartiles del gráfico, son los siguientes:

**Tabla 6-4:** Cuartiles del gráfico de cajas de LCPATR

Cuartil	Valor
Q1	26
Q2	30
Q3	36

Realizado por: Inca E y Tierra M, 2022.

❖ Sin datos atípicos LCPATR



**Ilustración 11-4:** Sin datos atípicos LCPATR

Realizado por: Inca E y Tierra M, 2022.

Se observa que existe tendencia negativa, es decir decrece a lo largo de la serie por lo cual las series es no estacionarias.

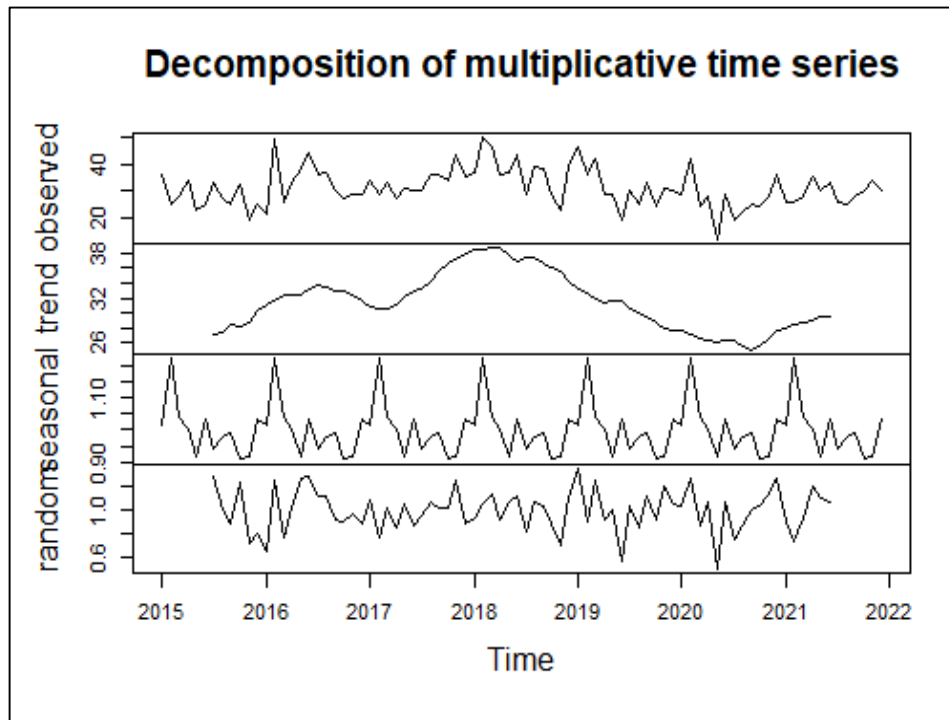
**Tabla 7-4:** Serie de tiempo de LCPATR

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
2015	36	25	28	34	23	25	33	27	25	32	19	25
2016	21	49	26	33	38	44	36	37	30	27	29	29
2017	34	29	33	27	31	30	30	36	36	34	43	35
2018	37	50	46	36	37	43	29	39	38	29	23	39
2019	46	36	42	29	29	19	30	25	33	24	31	30
2020	29	42	24	28	12	29	19	22	25	24	28	36
2021	26	26	28	35	30	33	26	25	28	30	34	30

Realizado por: Inca E y Tierra M, 2022.



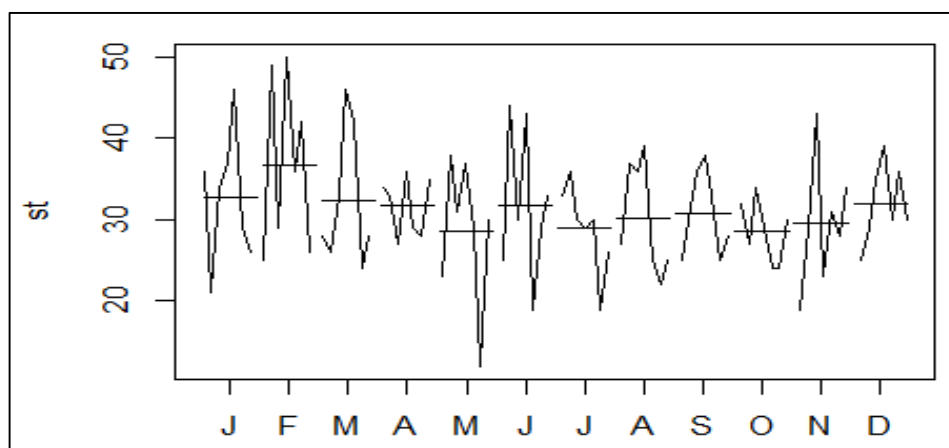
❖ **Descomposición de LCPATR**



**Ilustración 12-4:** Descomposición de LCPATR

Realizado por: Inca E y Tierra M, 2022.

En el componente tendencia se puede apreciar que existe una tendencia decreciente en el año 2018-2020 y en la parte estacional parece tener un patrón que se repite cada año por lo tanto se opta por el modelo ARIMA.



**Ilustración 13-4:** Gráfico mensual de la serie de tiempo de LCPATR

Realizado por: Inca E y Tierra M, 2022.

Mediante este gráfico se aprecia que en febrero existe un repunte de los accidentes por lo que nos indica que se deberá realizar un modelo ARIMA.

### 4.3. Análisis de serie de tiempo

#### 4.3.1. Daños materiales (DM)

##### 4.3.1.1. Identificación del modelo

##### ❖ ARIMA (p, d, q)

$H_0$ : La Serie de Tiempo es estacional

$H_1$ : La Serie de Tiempo no es estacional

**Tabla 8-4:** Test de integración para la parte ordinal de DM

Prueba	P-Value
kpss	0.01

Realizado por: Inca E y Tierra M, 2022.

En la parte regular se rechaza  $H_0$  teniendo un p-value del 0.01 menor a 0.05, por lo tanto, la serie DM no es estacional.

##### ❖ ARIMA (P, D, Q)

$H_0$ : La Serie de Tiempo es estacional

$H_1$ : La Serie de Tiempo no es estacional

**Tabla 9-4:** Test de integración para la parte estacional de DM

Prueba	P-Value
kpss	0,1125

Realizado por: Inca E y Tierra M, 2022.

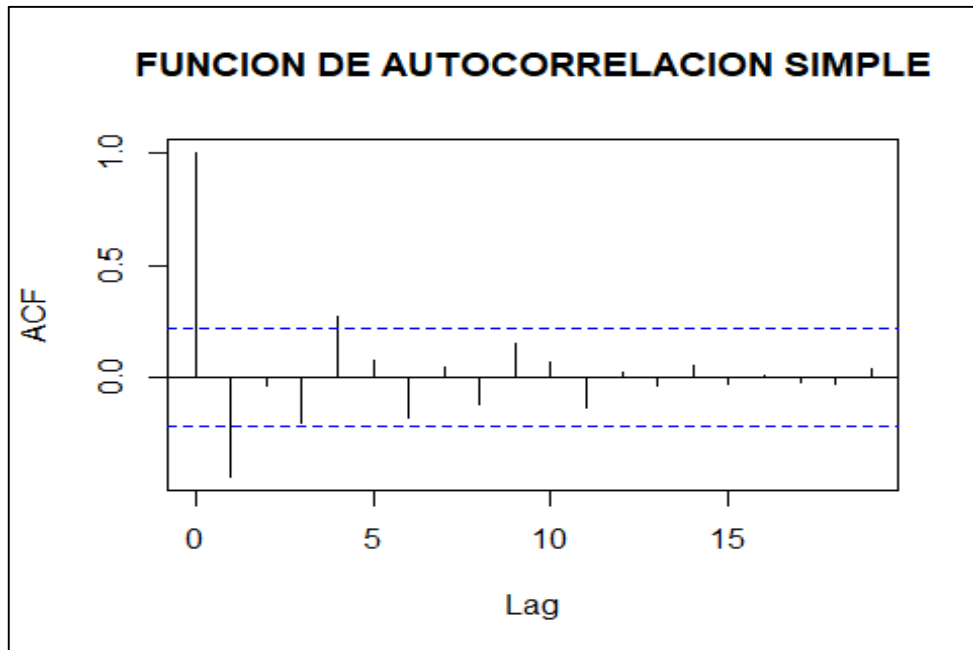
La serie DM es estacional dando el p-value del 0.11 mayor a 0.05, por lo tanto, no se rechaza  $H_0$ , hay que tener en cuenta el gráfico por mes que tiene un repunte estacional.

Por lo tanto, no es estacionaria en la parte regular y es estacional en la parte estacionaria, hay que tener en cuenta el gráfico por mes que tiene un repunte estacional.

❖ **Correlogramas**

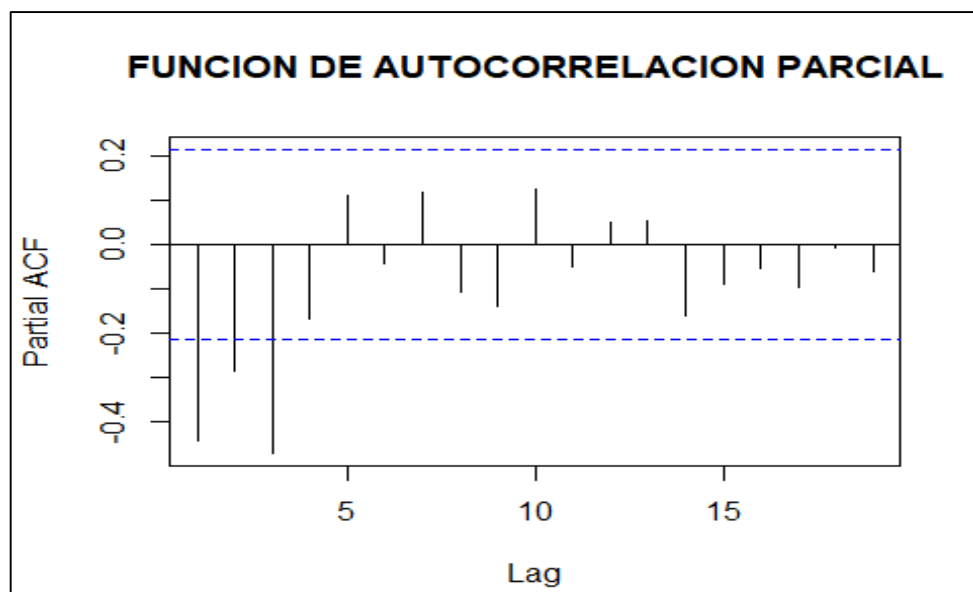
Se analiza los correlogramas de la parte no estacional, y la parte estacional a continuación:

❖ **Análisis para la parte no estacional ARIMA (p, d, q)**



**Ilustración 14-4:** Parte regular de la autocorrelación simple

Realizado por: Inca E y Tierra M, 2022.

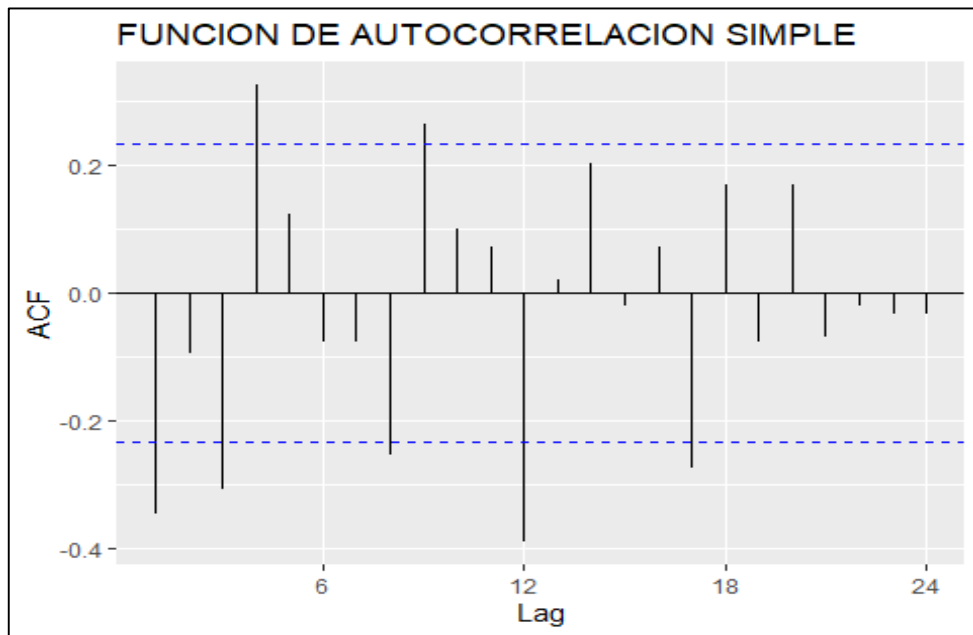


**Ilustración 15-4:** Parte regular de la autocorrelación parcial

Realizado por: Inca E y Tierra M, 2022.

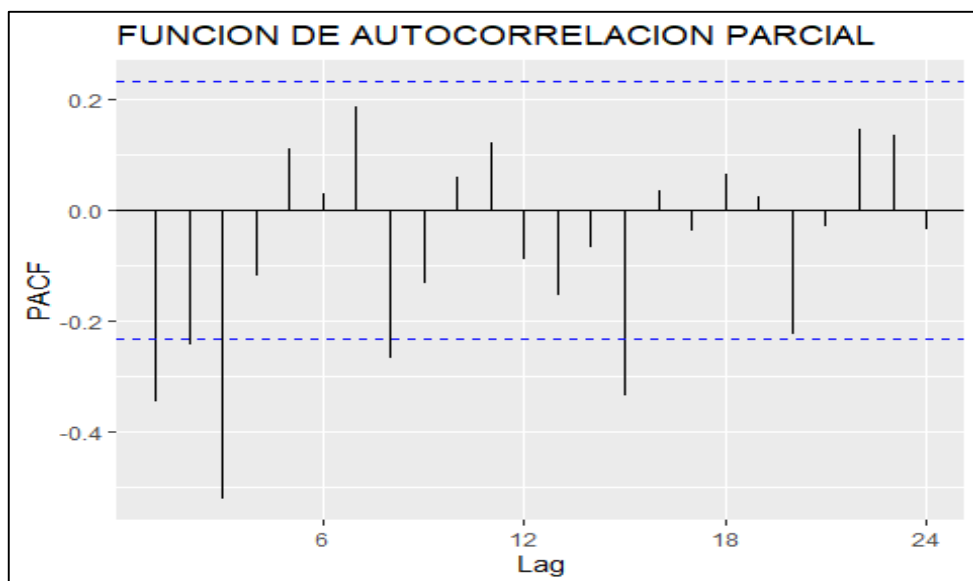
No se aprecia rezagos significativos en la parte estacional por lo tanto se aplicará  $ARIMA(p,d,q)$ . Observando los correlogramas de autocorrelación simple (ACF) el parámetro de medias móviles (MA) es  $q=3$ , la autocorrelación parcial (PACF) el parámetro autorregresivo (AR)  $p=3$  rezagos significativos.

❖ **Análisis para la parte estacional ARIMA (P, D, Q)**



**Ilustración 16-4:** Parte estacional de la autocorrelación simple

Realizado por: Inca E y Tierra M, 2022.



**Ilustración 17-4:** Parte estacional de la autocorrelación parcial

Realizado por: Inca E y Tierra M, 2022.

En la parte estacional se aprecia que la ilustración de autocorrelación simple el parámetro de medias móviles es Q=4, el parámetro autorregresivo P=4 y analizando el gráfico de autocorrelación parcial para medias móviles es Q=4, el parámetro autorregresivo P=1. Por lo tanto se aplicará un modelo ARIMA (p,d,q) (P,D,Q).

#### 4.3.1.2. Modelos ARIMA

Al obtener los ajustes del modelo se procede a cotejar.

**Tabla 10-4:** Modelos de DM

Modelos ARIMA	Parámetros	AIC	BIC
auto.arima()	(0,0,0) (2,1,2) [12]	561	575
2	(0,0,0) (1,1,2) [12]	575	584
3	(1,0,0) (1,1,1) [12]	569	578
4	(0,0,0) (0,1,0) [12]	596	599
5	(0,0,1) (0,1,2) [12]	572	581

Realizado por: Inca E y Tierra M, 2022.

Al cotejar los criterios AIC Y BIC se procede a seleccionar el mejor modelo.

**Tabla 11-4:** Selección del mejor modelo para DM

Modelo ARIMA	Coeficientes					AIC	BIC
	driff	sar1	sar2	sma1	sma2		
auto.arima()	-0.2076	0.2758	-0.0461	-1.4469	0.6399	561	575

Realizado por: Inca E y Tierra M, 2022.

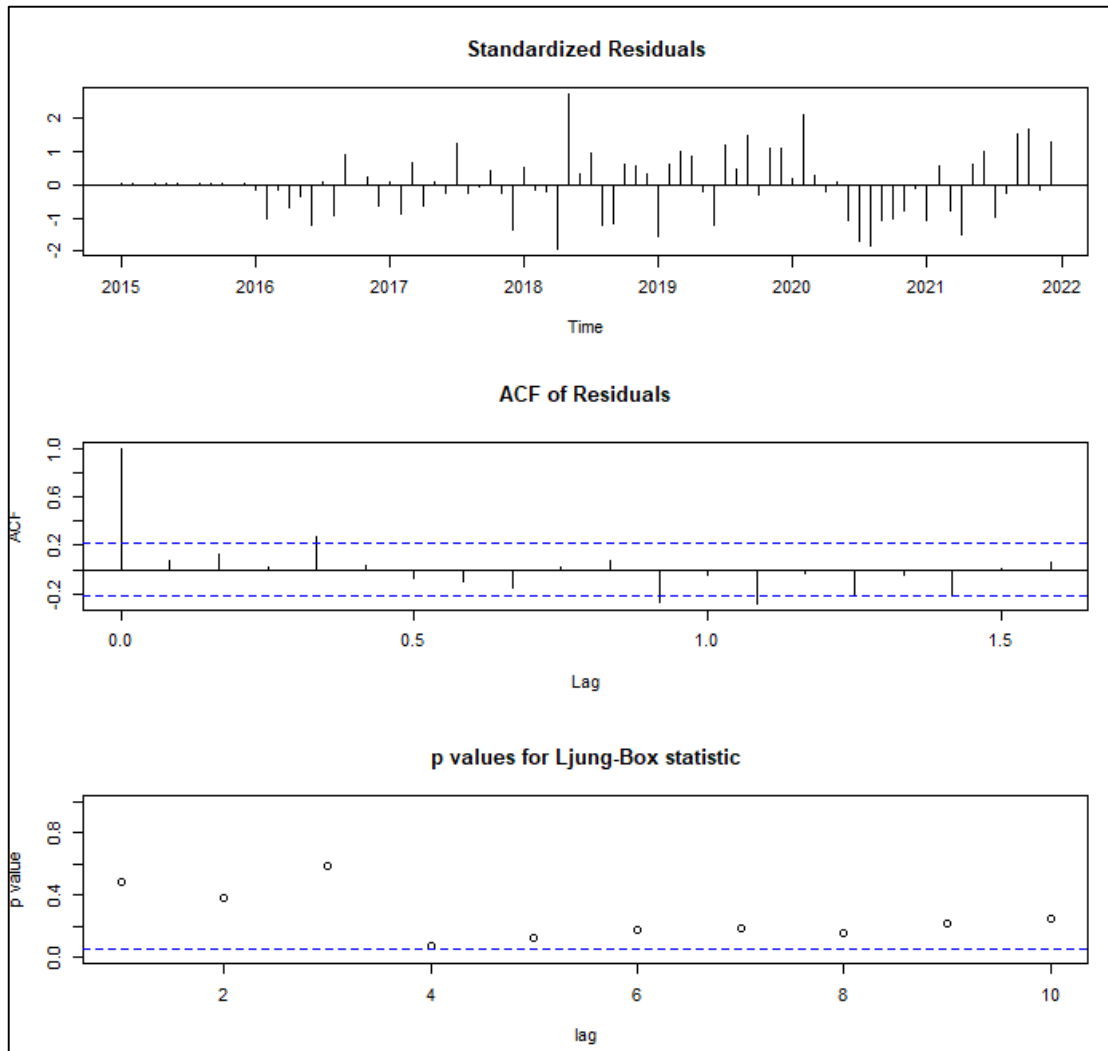
Mediante la información de la Tabla 11-4 se selecciona el modelo 1 que corresponde al modelo auto.arima (0,0,0) (2,1,2) [12], ya que presenta valores pequeños de los criterios de Akaike (AIC) y Bayesiano (BIC), es decir se adapta de mejor los datos.

Ecuación ARIMA (0,0,0) (2,1,2) [12].

$$Y_t = -0.2076 + 0.2758Y_{t-12} - 0.0461Y_{t-24} + 1.4465\varepsilon_{t-12} - 0.6399\varepsilon_{t-24} + \varepsilon_t$$

#### 4.3.1.3. Comprobación del modelo

##### ❖ Diagnóstico del modelo



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

**Realizado por:** Inca E y Tierra M, 2022.

En los residuos la serie tiene ruido blanco por lo que se puede ver que el modelo presenta estacionariedad, el p-value se ve que la mayoría de los puntos están encima de la línea de referencia, es decir es mayor a 0.05, por lo que eso nos quiere decir que la serie tiene ruido blanco y se corrobora con la prueba de Ljung-Box.

##### ❖ Prueba de Normalidad

$H_0$ : Errores normales

$H_1$ : Errores no normales

**Tabla 12-4:** Prueba de normalidad

Prueba	P-Value
Jarque Bera	0,57

Realizado por: Inca E y Tierra M, 2022.

No se rechaza  $H_0$  ya que el p-value encontrado por la prueba de normalidad de Jarque Bera es de 0,57 mayor a 0.05, se concluye que tiende a una distribución normal.

❖ **Prueba de Homocedasticidad**

$H_0$ : Existe homocedasticidad

$H_1$ : No existe homocedasticidad

**Tabla 13-4:** Prueba de homocedasticidad

Prueba	P-Value
White	0,34

Realizado por: Inca E y Tierra M, 2022.

Se deduce que mediante la prueba de White con un p-value de 0.34 mayor a 0.05, existe homocedasticidad, por lo que no se rechaza  $H_0$ .

❖ **Prueba de Independencia**

$H_0$ : Existe independencia

$H_1$ : No existe independencia

**Tabla 14-4:** Prueba de independencia

Prueba	P-Value
Ljung-Box	0,49

Realizado por: Inca E y Tierra M, 2022.

No se rechaza  $H_0$  ya que mediante la prueba de Ljung-Box con un p-value de 0.49 mayor a 0.05, se concluye que los errores son independientes.

### 4.3.2. Lesiones causadas por accidente de tránsito (LCPATR)

#### 4.3.2.1. Identificación del modelo

##### ❖ ARIMA (p, d, q)

$H_0$ : La Serie de Tiempo es estacional

$H_1$ : La Serie de Tiempo no es estacional

**Tabla 15-4:** Test de integración para la parte ordinal de LCPATR

Prueba	P-Value
Kpss	0,32

Realizado por: Inca E y Tierra M, 2022.

En la parte regular no se rechaza  $H_0$ , la serie LCPATR es estacional, por lo tanto, no se debe realizar diferenciación de los datos ya que el p-value es de 0.32 mayor a 0.05.

##### ❖ ARIMA (P, D, Q)

$H_0$ : La Serie de Tiempo es estacional

$H_1$ : La Serie de Tiempo no es estacional

**Tabla 16-4:** Test de integración para la parte estacional de LCPATR

Prueba	P-Value
Kpss	0,30

Realizado por: Inca E y Tierra M, 2022.

La serie LCPATR es estacional dando el p-value es de 0.30 y es mayor a 0.05.

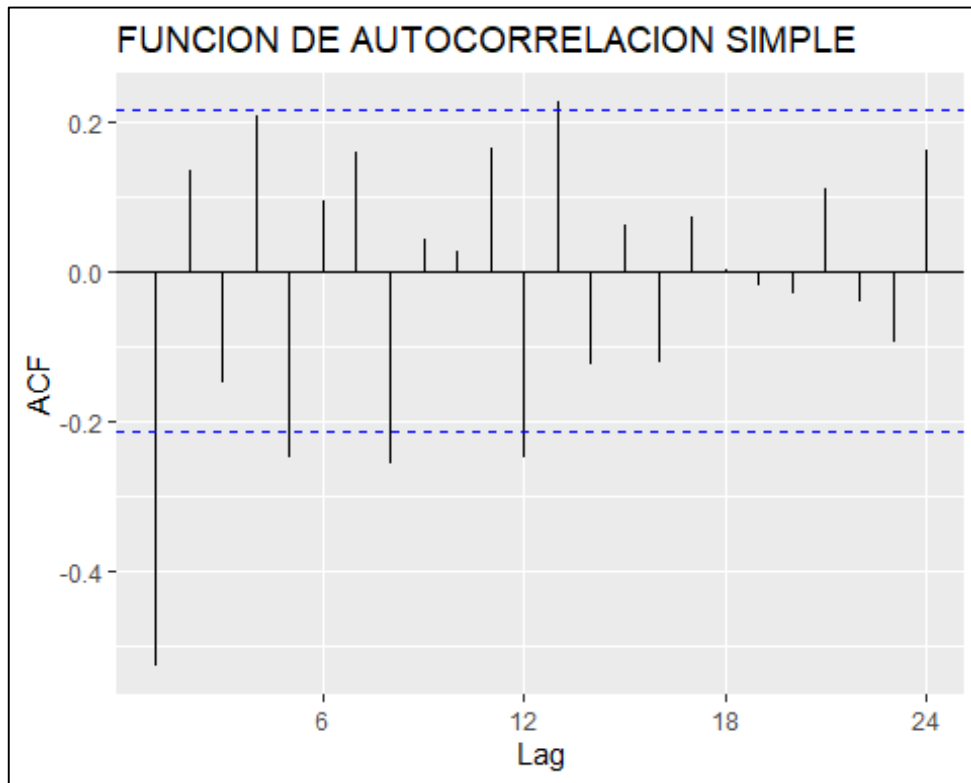
Por lo tanto, es estacional en la parte regular y es estacional en la parte estacionaria sin embargo se aprecia una clara estacionalidad hay que tener en cuenta ACF y PACF con una diferencia ordinal.

##### ❖ Correlogramas

Se analiza los correlogramas de la parte no estacional, a continuación, visualizaremos con los gráficos siguientes:

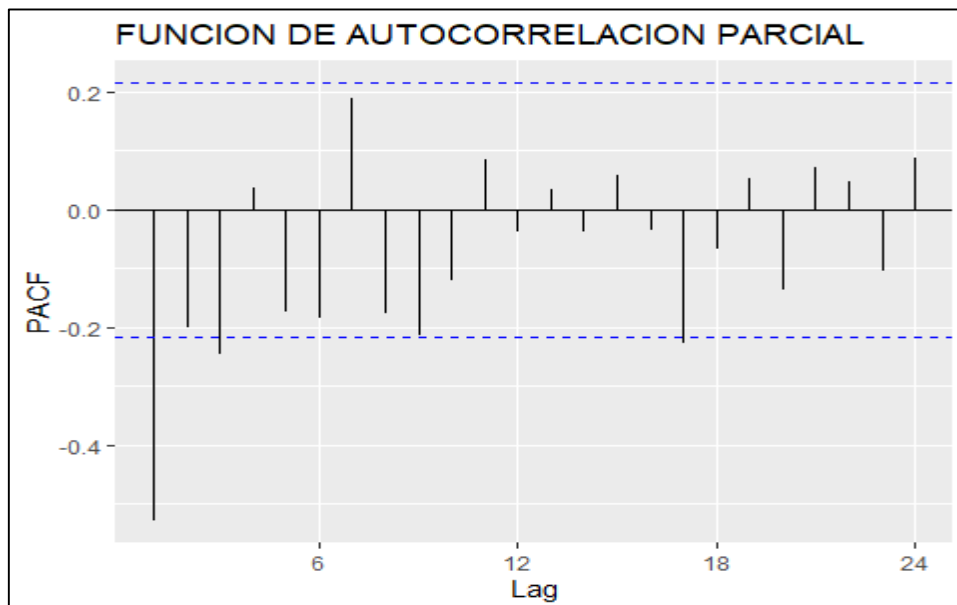


❖ **Análisis para la parte no estacional ARIMA (p, d, q)**



**Ilustración 19-4:** Parte regular de la autocorrelación simple

Realizado por: Inca E y Tierra M, 2022.

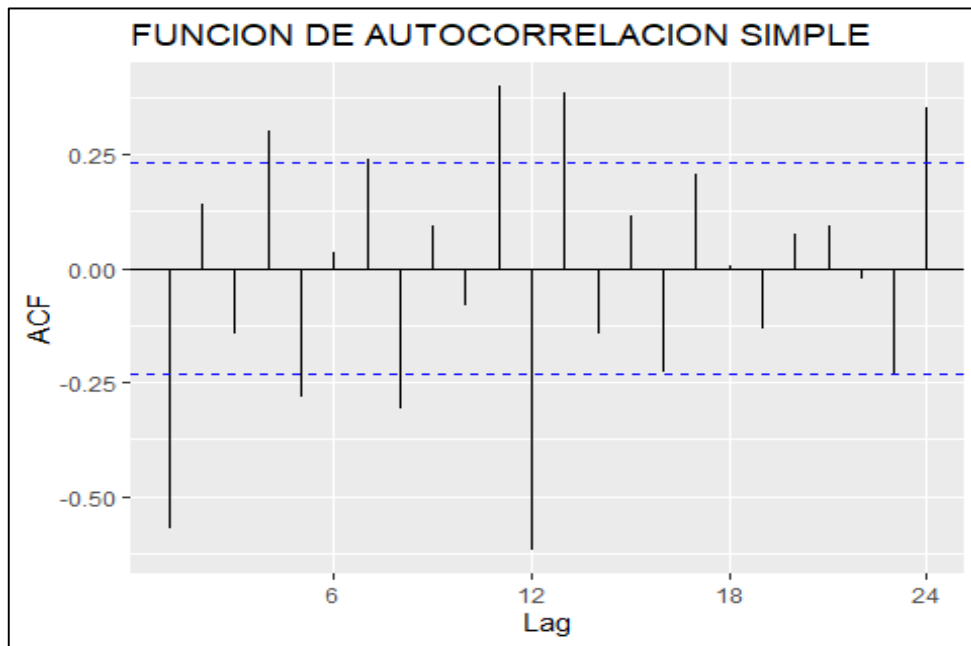


**Ilustración 20-4:** Parte regular de la autocorrelación parcial

Realizado por: Inca E y Tierra M, 2022.

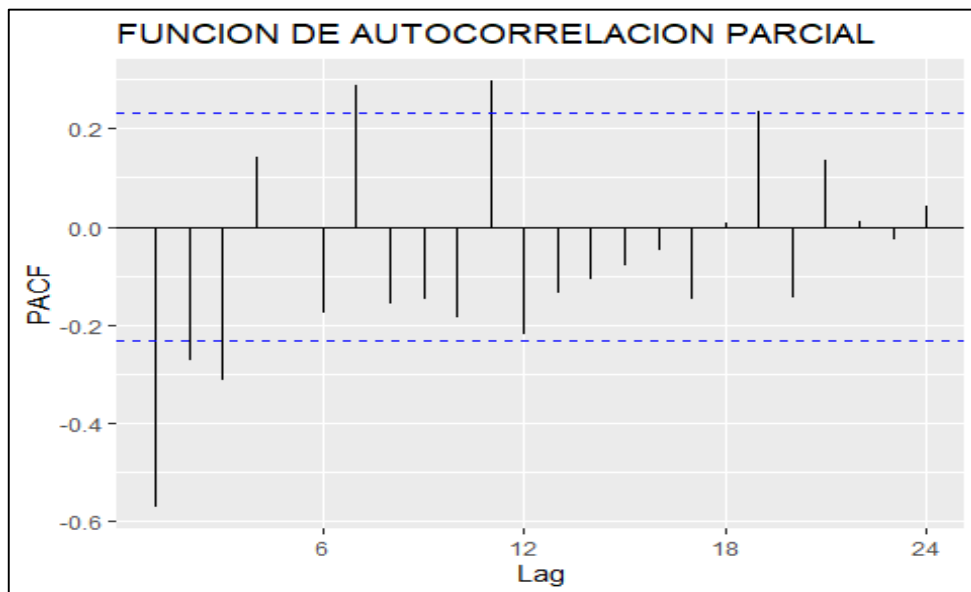
Se aplicará un modelo ARIMA(p,d,q), observando los correlogramas de autocorrelación simple (ACF) el parámetro de medias móviles (MA) existe q=4 rezagos significativos, no hay diferencia existe intercepto y la autocorrelación parcial (PACF) el parámetro autorregresivo (AR) existe p=2 rezagos significativos.

❖ **Análisis para la parte estacional ARIMA (P,D,Q)**



**Ilustración 21-4:** Parte estacional de la autocorrelación simple

Realizado por: Inca E y Tierra M, 2022.



**Ilustración 22-4:** Parte estacional de la autocorrelación parcial

Realizado por: Inca E y Tierra M, 2022.

En la parte estacional el gráfico de autocorrelación simple el parámetro de medias móviles es  $Q=7$ , no existe diferencia, el parámetro autorregresivo  $P=1$  y analizando el gráfico de autocorrelación parcial existe medias móviles  $Q=5$ , con una diferencia  $D=1$ , el parámetro autorregresivo  $P=1$ . Por lo tanto, se aplicará un modelo ARIMA.

#### 4.3.2.2. Modelos ARIMA

**Tabla 17-4:** Modelos para LCPATR

Modelos ARIMA	Parámetros	AIC	BIC
Auto.arima()	(2,0,1) (1,0,0) [12]	563	577
2	(1,0,0) (2,0,0) [12]	567	579
3	(2,0,1) (0,0,1) [12]	563	577
4	(1,0,2) (1,0,1) [12]	560	577
5	(1,0,1) (1,0,1) [12]	559	569

Realizado por: Inca E y Tierra M, 2022.

Al cotejar los criterios AIC Y BIC se procede a seleccionar el mejor modelo.

**Tabla 18-4:** Selección del mejor modelo para LCPATR

Modelo ARIMA	Coeficientes					AIC	BIC
	mean	ar1	ma1	sar1	sma1		
5	31.2580	0.8143	-0.5297	-0.9943	0.9502	559	569

Realizado por: Inca E y Tierra M, 2022.

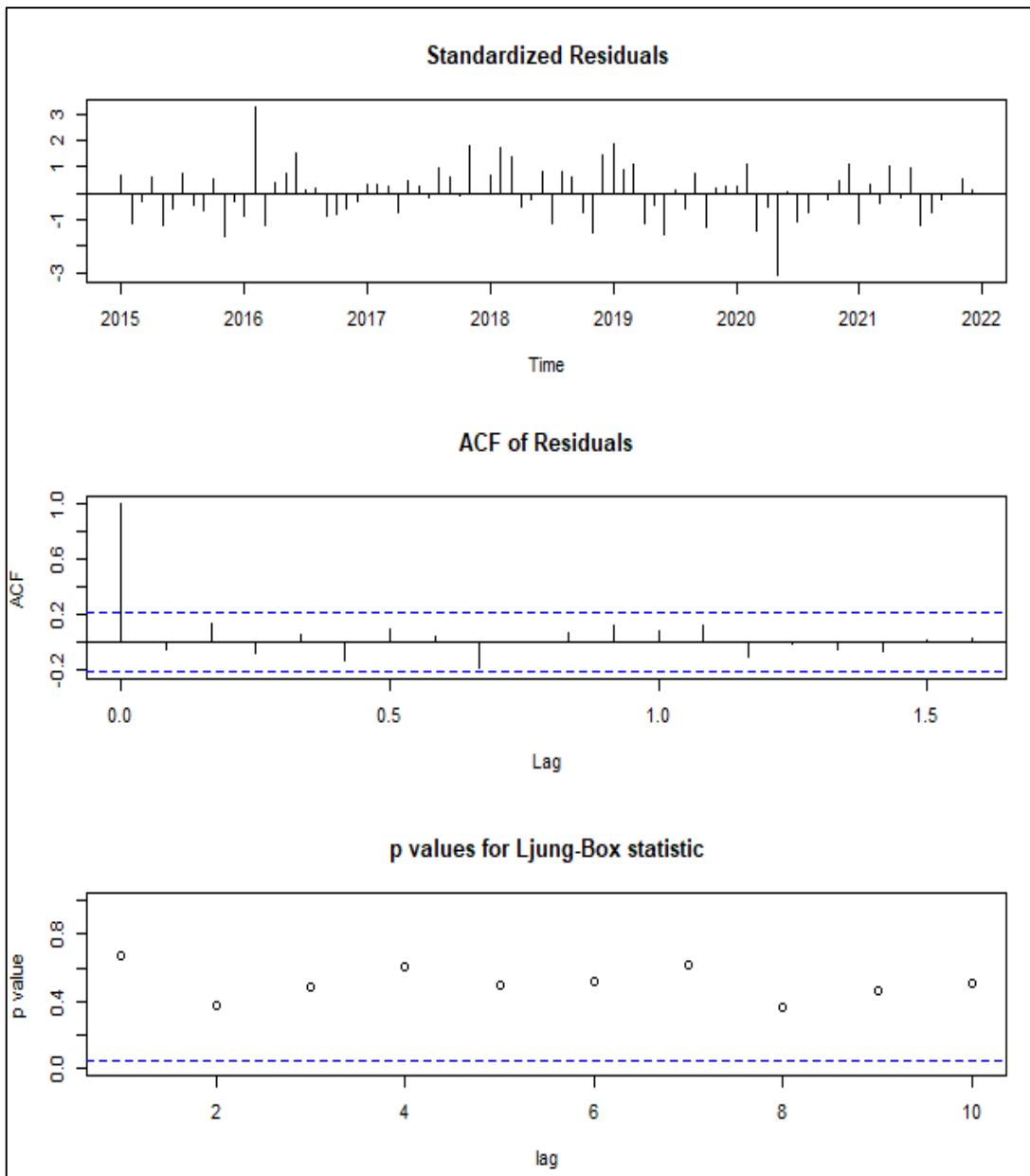
Mediante la información de la Tabla 18-4 se selecciona el modelo 5 que corresponde al modelo (1,0,1) (1,0,1) [12], ya que presenta valores pequeños de los criterios de Akaike (AIC) y Bayesiano (BIC), es decir se adapta mejor los datos.

Ecuación ARIMA (1,0,1) (1,0,1) [12].

$$Y_t = 31.2580 + 0.8143Y_{t-1} + 0.5297\varepsilon_{t-1} - 0.9943Y_{t-12} - 0.9502\varepsilon_{t-12} + \varepsilon_t$$

### 4.3.2.3. Comprobación del modelo

#### ❖ Diagnóstico del modelo



**Ilustración 23-4:** Diagnóstico del modelo ARIMA (1,0,1) (1,0,1) [12]

**Realizado por:** Inca E y Tierra M, 2022.

En los residuos la serie tiene ruido blanco por lo que se puede ver que el modelo presenta estacionariedad dado que tiene media cero y varianza constante, el ACF de los residuales se ve que la mayoría de los retardos se encuentran en el marco de referencia y el p-value se ve que la mayoría de los puntos están encima de la línea de referencia por lo que eso nos quiere decir que la serie tiene ruido blanco y se corrobora con la prueba de Ljung-Box.

### ❖ Prueba de Normalidad

$H_0$ : Errores normales

$H_1$ : Errores no normales

**Tabla 19-4:** Prueba de normalidad

Prueba	P-Value
Jarque Bera	0.13

Realizado por: Inca E y Tierra M, 2022.

No se rechaza  $H_0$  ya que el p-value encontrado por la prueba de normalidad de Jarque Bera es de 0,13 mayor a 0.05, se concluye que tiende a una distribución normal.

### ❖ Prueba de Homocedasticidad

$H_0$ : Existe homocedasticidad

$H_1$ : No existe homocedasticidad

**Tabla 20-4:** Prueba de homocedasticidad

Prueba	P-Value
White	0.95

Realizado por: Inca E y Tierra M, 2022.

Se deduce que mediante la prueba de White con p-value del 0.95 mayor a 0.05, existe homocedasticidad, por lo que no se rechaza  $H_0$ .

### ❖ Prueba de Independencia

$H_0$ : Existe independencia en errores

$H_1$ : No existe independencia en errores

**Tabla 21-4:** Prueba de independencia

Prueba	P-Value
Ljung-Box	0.67

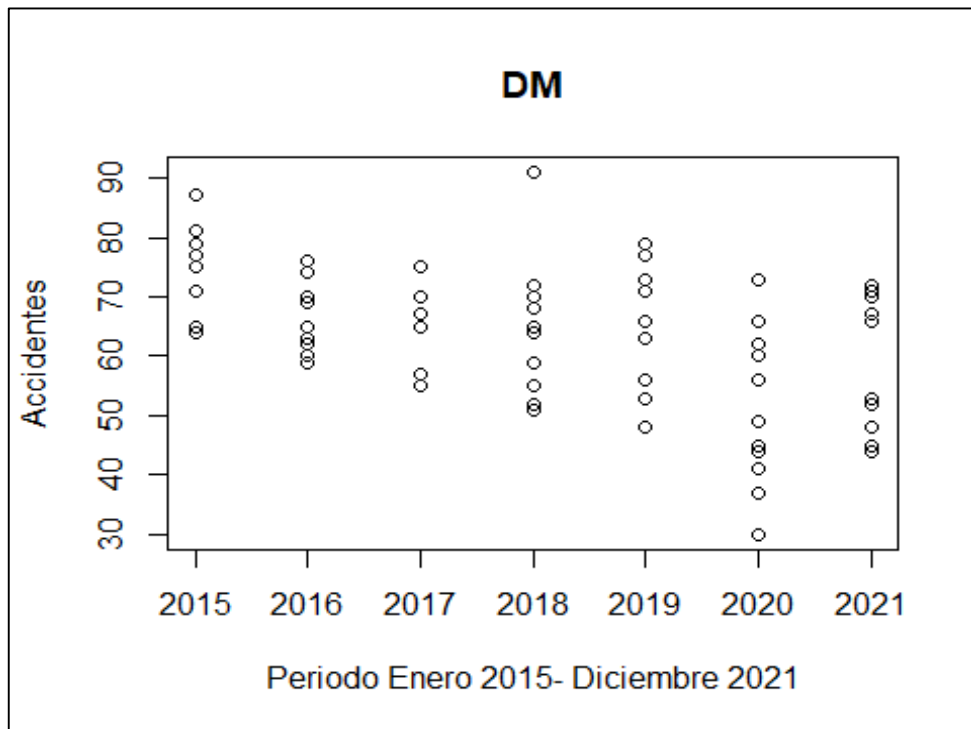
Realizado por: Inca E y Tierra M, 2022.

No se rechaza  $H_0$  ya que mediante la prueba de Ljung-Box se deduce que los errores son independientes por lo que tiene un p-value de 0,67 mayor a 0.05.

#### 4.4. Análisis de Poisson

##### 4.4.1. Daños materiales (DM)

Para realizar modelos de Poisson se tiene en cuenta el número de accidentes en un período de tiempo de enero 2015 a diciembre 2021.



**Ilustración 24-4:** Diagrama de dispersión accidentes de DM

Realizado por: Inca E y Tierra M, 2022.

Se aprecia los parámetros de los tres modelos teniendo la variable respuesta (consumado) y el predictor (mes), para dar respuesta al ajuste de los datos teniendo que es un modelo lineal generalizado (glm).

##### 4.4.1.1. Identificación de los modelos

Para obtener los coeficientes se utiliza la función glm() la cual se procede a construir los modelos de la regresión de Poisson.

**Tabla 22-4:** Modelo lineal de Poisson DM

Modelo	AIC	BIC	Intercept	mes
mod1	643	648	90.58522	-0.04284

Realizado por: Inca E y Tierra M, 2022.

Obteniendo la siguiente ecuación:

❖ **Modelo aditivo**

$$\log(mes) = 90.58522 - 0.04284 * mes \quad \text{Ecu 1-4}$$

❖ **Modelo multiplicativo**

$$mes = e^{90.58522} * e^{-0.04284mes} \quad \text{Ecu. 2-4}$$

❖ **Prueba analítica de Homocedasticidad**

$H_0$ : Existe homocedasticidad

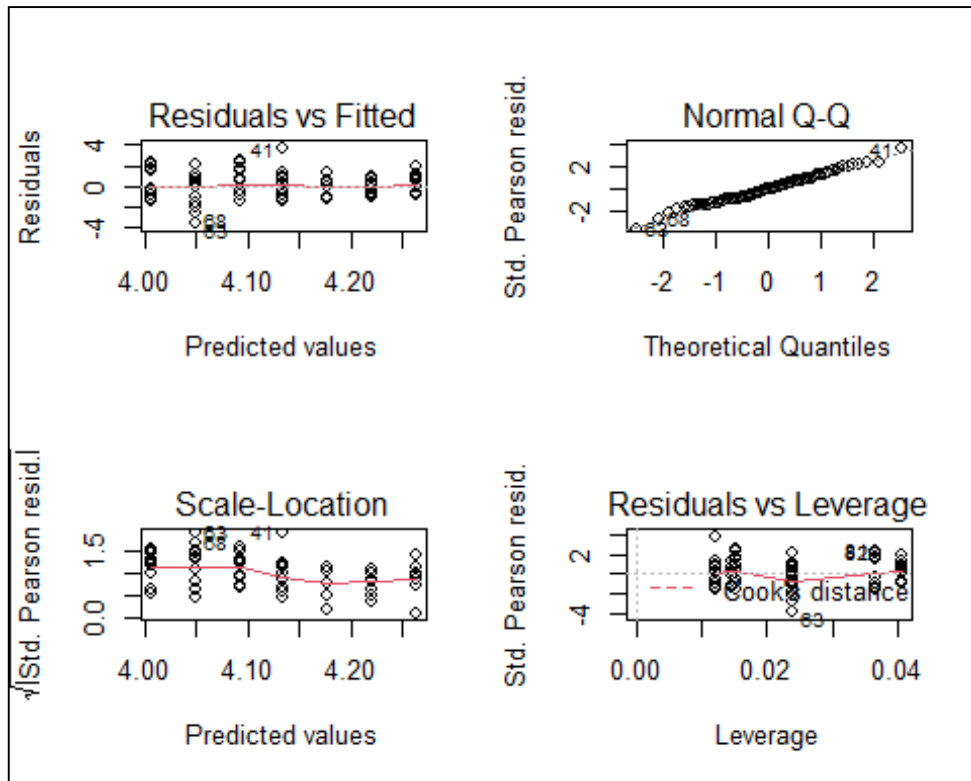
$H_1$ : No existe homocedasticidad

**Tabla 23-4:** Prueba analítica de homocedasticidad

Prueba	P-Value
White	0.90

Realizado por: Inca E y Tierra M, 2022.

Se deduce que mediante la prueba de White con p-value del 0.90 mayor a 0.05, se concluye que existe homocedasticidad, por lo que no se rechaza  $H_0$ .



**Ilustración 25-4:** Diagnóstico del primer modelo de DM

Realizado por: Inca E y Tierra M, 2022.

Podemos observar que los residuos no tienen una tendencia estacionaria, es decir no están corraladas, la normal no se ajusta a la línea de tendencia, el modelo de regresión no tiene una calidad predictiva buena, existe homocedasticidad ya que existe varianza de error constante a lo largo del tiempo.

Con la función `glm()` construimos, por lo tanto, un modelo nuevo añadiendo un término cuadrático.

**Tabla 24-4:** Modelo cuadrático de Poisson DM

Modelo	AIC	BIC	Intercept	mes	mes ^2
mod2	645	652	5.593e+03	-5.497e+00	1.351e-03

Realizado por: Inca E y Tierra M, 2022.

Obteniendo la siguiente ecuación:

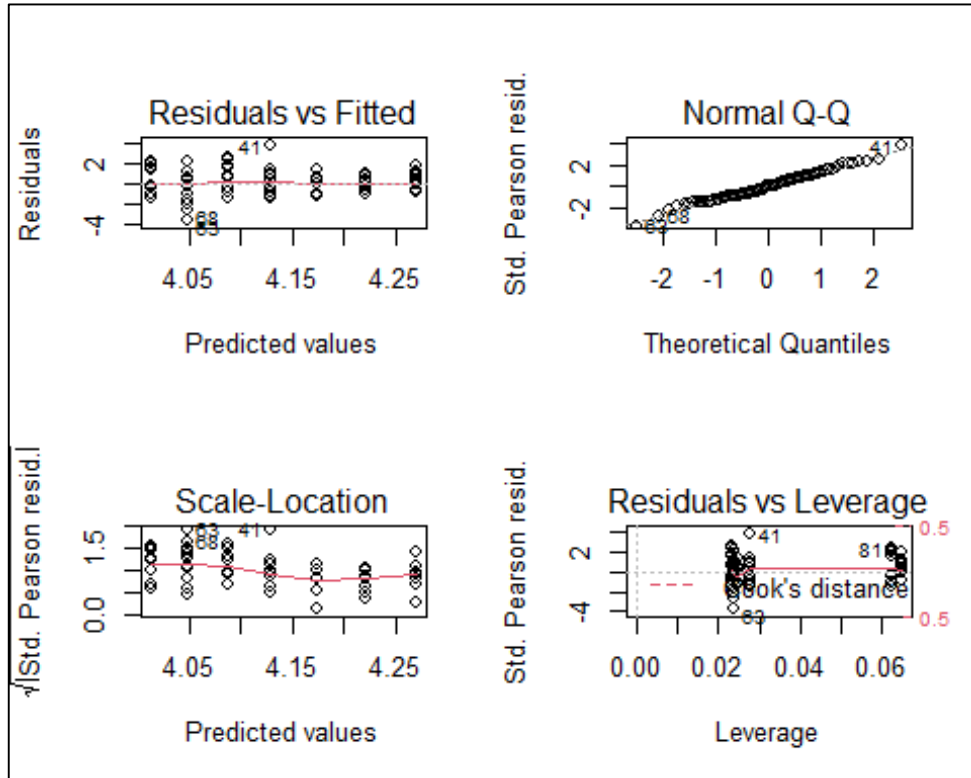
❖ **Modelo aditivo**

$$\log(mes) = (5.593e + 03) - (5.497e + 00)mes + (1.351e - 03)mes^2$$



❖ **Modelo multiplicativo**

$$mes = e^{5.593e+03} * e^{((-5.497e+00)mes+(1.351e-03)mes^2)}$$



**Tabla 25-4:** Diagnóstico del segundo modelo de DM

Realizado por: Inca E y Tierra M, 2022.

Podemos observar que los residuos no tienen una tendencia estacionaria, es decir no están corraladas, la normal se ajusta a la línea de tendencia, el modelo de regresión tiene una calidad predictiva buena, no existe homocedasticidad ya que no existe varianza de error constante a lo largo del tiempo.

Con la función glm() construimos, por lo tanto, un modelo nuevo añadiendo un término cubico.

**Tabla 26-4:** Modelo cúbico de Poisson DM

Modelo	AIC	BIC	Intercept	mes	mes ^2	mes ^3
mod3	647	656	-1.18e+06	1.765e+03	-8.761e-01	1.449e-04

Realizado por: Inca E y Tierra M, 2022.

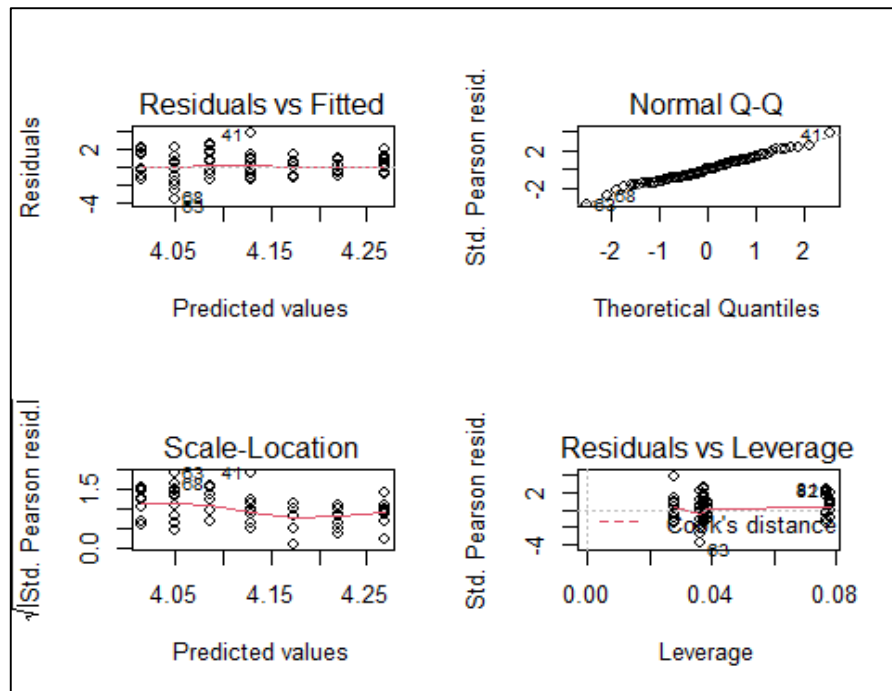
Obteniendo la siguiente ecuación:

❖ **Modelo aditivo**

$$\log(mes) = (-1.185e + 06) + (1.765e + 03)mes - (8.761e - 01)mes^2 + (1.449e - 04)mes^3$$

❖ **Modelo multiplicativo**

$$mes = e^{-1.185e+06} * e^{((1.765e+03mes)-(8.761e-01)mes^2+(1.449e-04)mes^3)}$$



**Ilustración 26-4:** Diagnóstico del tercer modelo de DM

Realizado por: Inca E y Tierra M, 2022.

Podemos observar que los residuos no tienen una tendencia estacionaria, es decir están corraladas, la normal se ajusta a la línea de tendencia, el modelo de regresión tiene una calidad predictiva buena, no existe homocedasticidad ya que no existe varianza de error constante a lo largo del tiempo.

4.4.1.2. *Modelo Poisson*

Para identificar el mejor modelo para la variable DM se procede a cotejar el modelo más bajo mediante los criterios AIC Y BIC.

**Tabla 27-4:** Mejor modelo de Poisson DM

Modelo	AIC	BIC
mod1	643	648
mod2	645	652
mod3	647	656

Realizado por: Inca E y Tierra M, 2022.

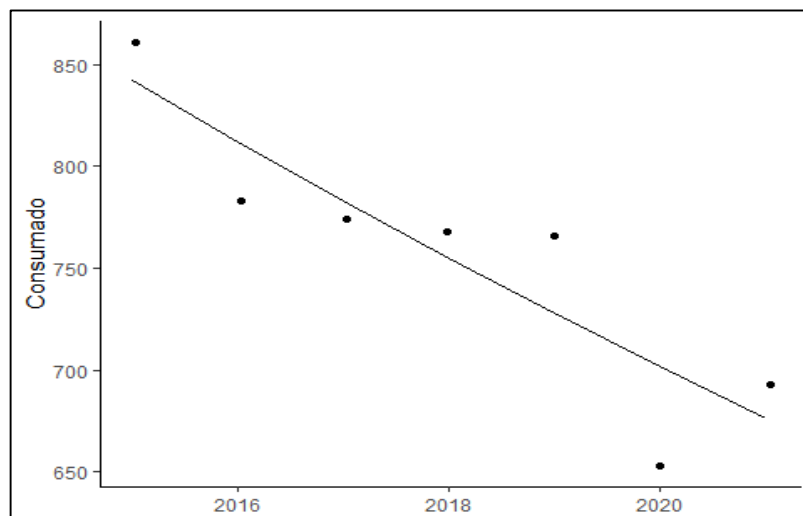
Teniendo el modelo lineal que es mod1 como el mejor modelo para realizar predicciones.

❖ **Modelo aditivo**

$$\log(mes) = 90.58522 - 0.04284 * mes$$

❖ **Modelo multiplicativo**

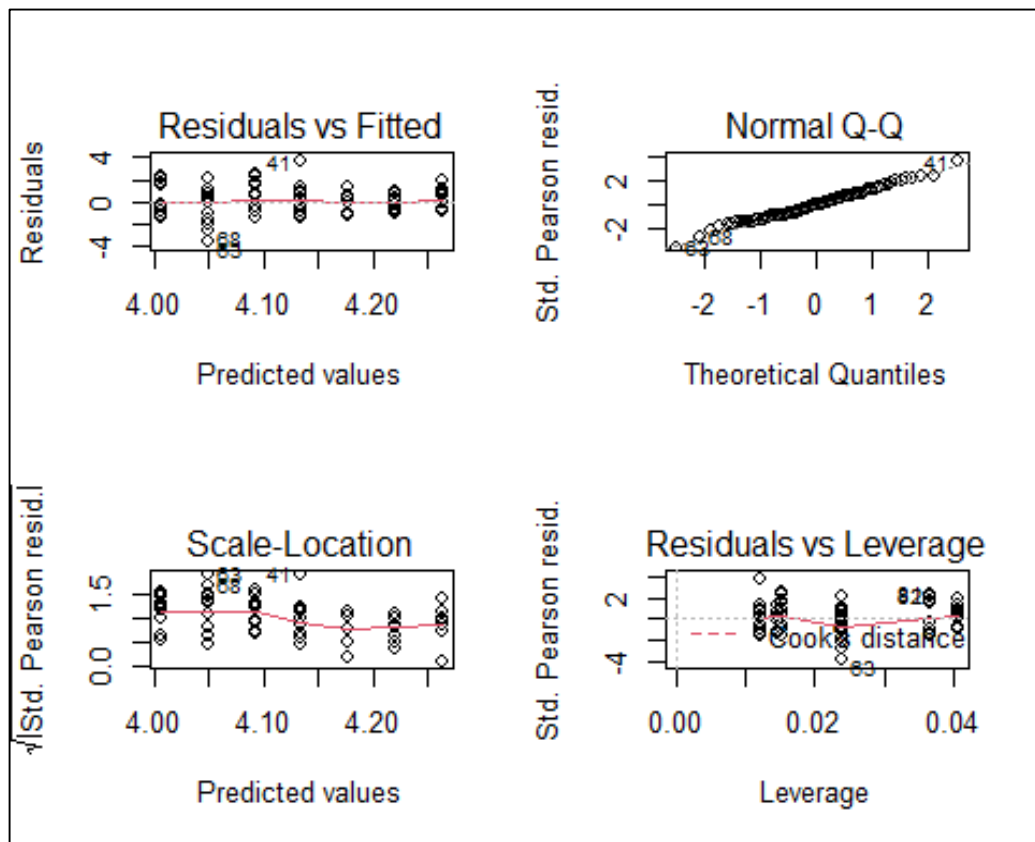
$$mes = e^{90.58522} * e^{-0.04284mes}$$



**Ilustración 27-4:** Ajuste del mejor modelo DM

Realizado por: Inca E y Tierra M, 2022.

#### 4.4.1.3. Comprobación del mejor modelo



**Ilustración 28-4:** Diagnóstico del mejor modelo de DM

Realizado por: Inca E y Tierra M, 2022.

Podemos observar que los residuos no tienen una tendencia estacionaria, es decir están corraladas, la normal se ajusta a la línea de tendencia, el modelo de regresión tiene una calidad predictiva buena, existe homocedasticidad ya que existe varianza de error constante a lo largo del tiempo.

#### ❖ Prueba de Normalidad

$H_0$ : Errores normales

$H_1$ : Errores no normales

**Tabla 28-4:** Prueba de normalidad

Prueba	P-Value
Jarque Bera	0.91

Realizado por: Inca E y Tierra M, 2022.

No se rechaza  $H_0$  ya que el p-value encontrado por la prueba de normalidad de Jarque Bera es de 0,91 mayor a 0.05, se concluye que tiende a una distribución normal.

❖ **Prueba de Homocedasticidad**

$H_0$ : Existe homocedasticidad

$H_1$ : No existe homocedasticidad

**Tabla 29-4:** Prueba de homocedasticidad

Prueba	P-Value
White	0.90

Realizado por: Inca E y Tierra M, 2022.

Se deduce que mediante la prueba de White con p-value del 0.90 mayor a 0.05, se concluye que existe homocedasticidad, por lo que no se rechaza  $H_0$ .

❖ **Prueba de Independencia**

$H_0$ : Existe independencia en errores

$H_1$ : No existe independencia en errores

**Tabla 30-4:** Prueba de independencia

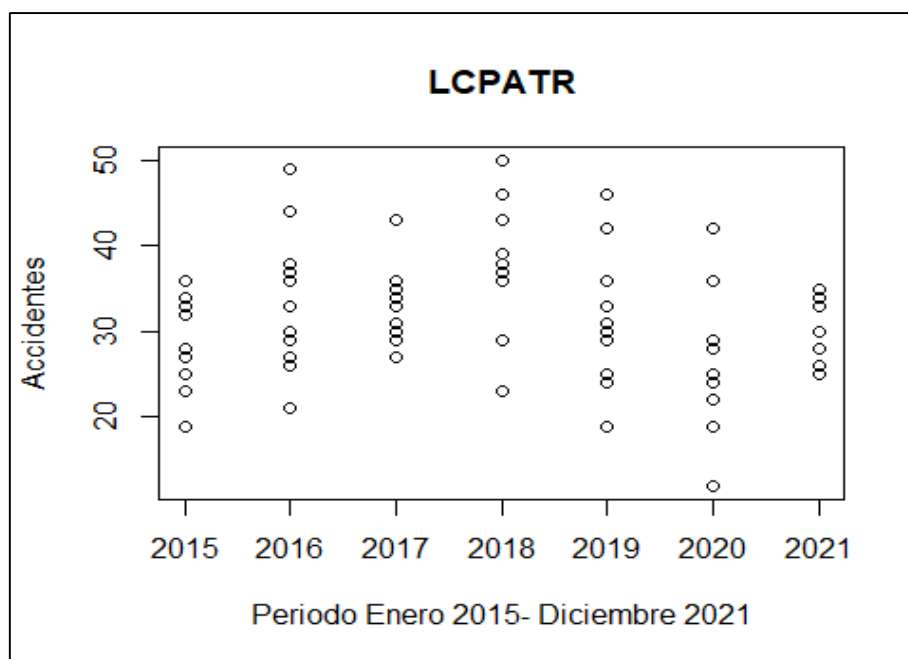
Prueba	P-Value
Ljung-Box	0.34

Realizado por: Inca E y Tierra M, 2022.

No se rechaza  $H_0$  ya que mediante la prueba de Ljung-Box se deduce que los errores son independientes por lo que tiene un p-value de 0,34 mayor a 0.05.

**4.4.2. Lesiones causadas por accidentes de tránsito (LCPATR)**

Para realizar los modelos de Poisson de estas variables se procede a realizar una gráfica para que tan dispersos están los datos.



**Ilustración 29-4:** Diagrama de dispersión LCPATR

Realizado por: Inca E y Tierra M, 2022.

Para dar respuesta al ajuste de estos datos pasamos a estimar los parámetros de tres modelos indicando que es un Modelo Lineal Generalizado(glm), teniendo la variable respuesta (consumado) y el predictor (mes).

#### 4.4.2.1. Identificación del mejor modelo

Para obtener los coeficientes se utiliza la función glm() la cual se procede a construir los modelos de la regresión de Poisson.

**Tabla 31-4:** Modelo lineal de Poisson LCPATR

Modelo	AIC	BIC	Intercept	Mes
mod1	578	583	28.30105	-0.01232

Realizado por: Inca E y Tierra M, 2022.

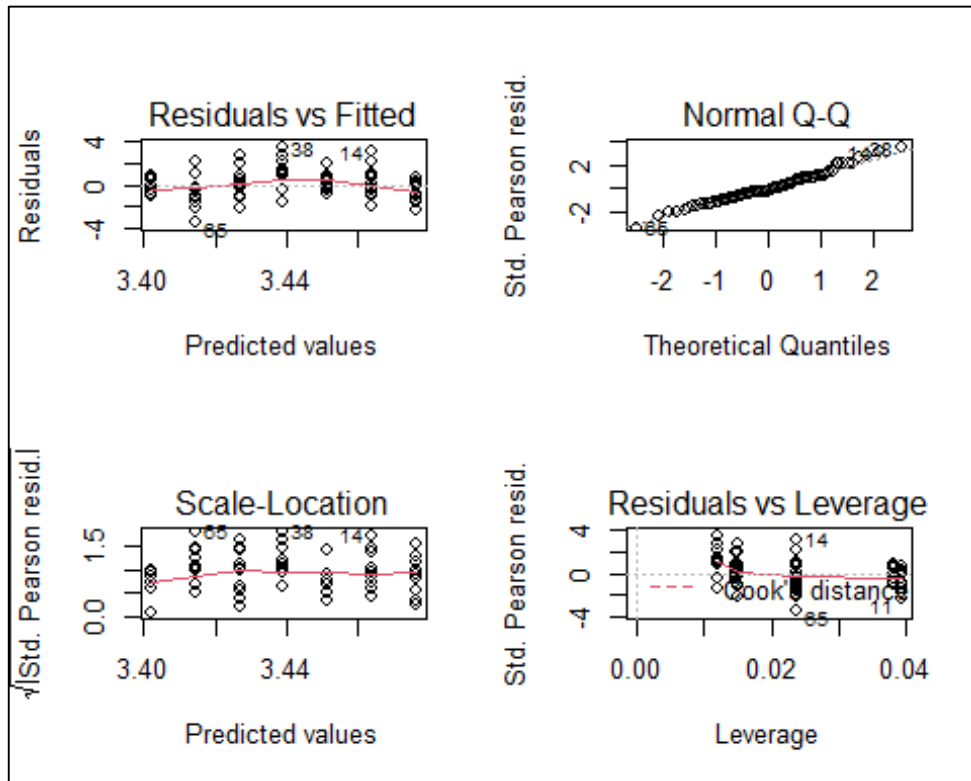
Obteniendo la siguiente ecuación:

#### ❖ Modelo aditivo en términos muestrales

$$\log(mes) = 28.30105 - 0.01232 * mes$$

❖ **Modelo multiplicativo**

$$mes = e^{28.30105} * e^{-0.01232mes}$$



**Ilustración 30-4:** Diagnóstico del primer modelo de LCPATR

Realizado por: Inca E y Tierra M, 2022.

Podemos observar que los residuos no tienen una tendencia estacionaria, es decir no están correladas, la normal no se ajusta a la línea de tendencia, el modelo de regresión es decir no tienen una calidad predictiva buena, no existe homocedasticidad ya que no existe varianza de error constante a lo largo del tiempo.

Con la función `glm()` construimos, por lo tanto, un modelo nuevo añadiendo un término cuadrático.

**Tabla 32-4:** Modelo cuadrático de Poisson LCPATR

Modelo	AIC	BIC	Intercept	mes	mes ^2
mod2	565	572	-9.081e+04	9.002e+01	-2.231e-02

Realizado por: Inca E y Tierra M, 2022.

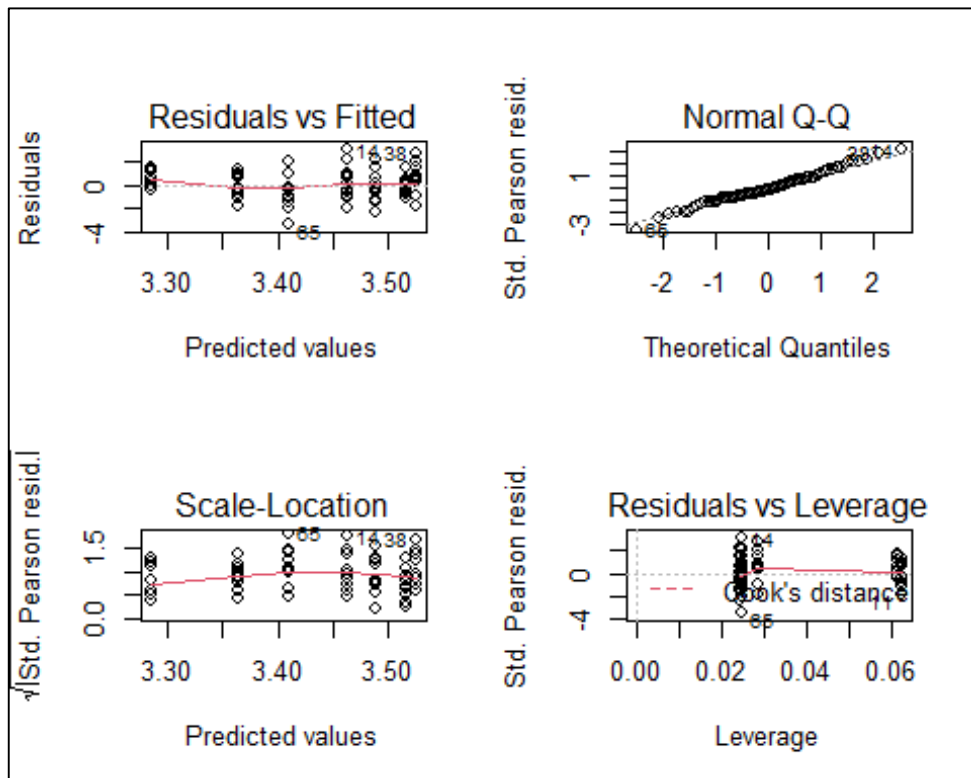
Obteniendo la siguiente ecuación:

❖ **Modelo aditivo**

$$\log(mes) = -9.081e + (04 + 9.002e + 01)mes - (2.231e - 02)mes^2$$

❖ **Modelo multiplicativo**

$$mes = e^{-9.081e+04} * e^{((9.002e+01)mes - (2.231e-02)mes^2)}$$



**Ilustración 31-4:** Diagnóstico del segundo modelo de LCPATR

**Realizado por:** Inca E y Tierra M, 2022.

Podemos observar que los residuos no tienen una tendencia no estacionaria ya que varían, es decir están correladas, la normal se ajusta a la línea de tendencia, el modelo de regresión tiene una calidad predictiva aproximadamente buena existe homocedasticidad ya que existe varianza de error constante a lo largo del tiempo.

Con la función glm() construimos, por lo tanto, un modelo nuevo añadiendo un término cúbico.



**Tabla 33-4:** Modelo cúbico de Poisson LCPATR

Modelo	AIC	BIC	Intercept	mes	mes ^2	mes ^3
mod3	560	570	-7.493e+07	1.113e+05	-5.516e+01	9.1107e-03

Realizado por: Inca E y Tierra M, 2022.

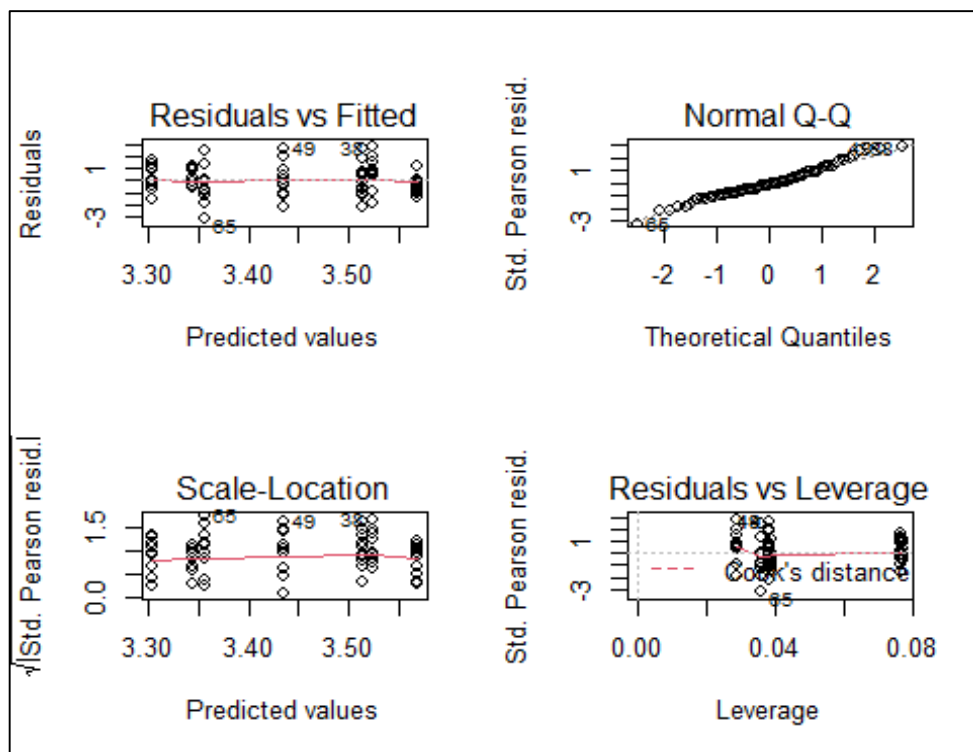
Obteniendo la siguiente ecuación:

❖ **Modelo aditivo**

$$\log(\text{mes}) = -7.493e + 07 + (1.113e + 05)\text{mes} - (5.516e + 01)\text{mes}^2 + (9.1107e - 03)\text{mes}^3$$

❖ **Modelo multiplicativo**

$$\text{mes} = e^{-7.493e+07} * e^{((1.113e+05)\text{mes} - (5.516e+01)\text{mes}^2 + (9.1107e-03)\text{mes}^3)}$$



**Ilustración 32-4:** Diagnóstico del tercer modelo de LCPATR

Realizado por: Inca E y Tierra M, 2022.

Podemos observar que los residuos no tienen una tendencia aproximadamente estacionaria, es decir están corraladas, la normal se ajusta a la línea de tendencia, el modelo de regresión tiene una calidad predictiva buena, existe homocedasticidad ya que existe varianza de error constante a lo largo del tiempo.

#### 4.4.2.2. Modelo Poisson

Para identificar el mejor modelo para la variable LCPATR se procede a cotejar el modelo más bajo mediante los criterios AIC Y BIC.

**Tabla 34-4:** Mejor modelo de Poisson LCPATR

Modelo	AIC	BIC
mod1	578	583
mod2	565	572
mod3	560	570

Realizado por: Inca E y Tierra M, 2022.

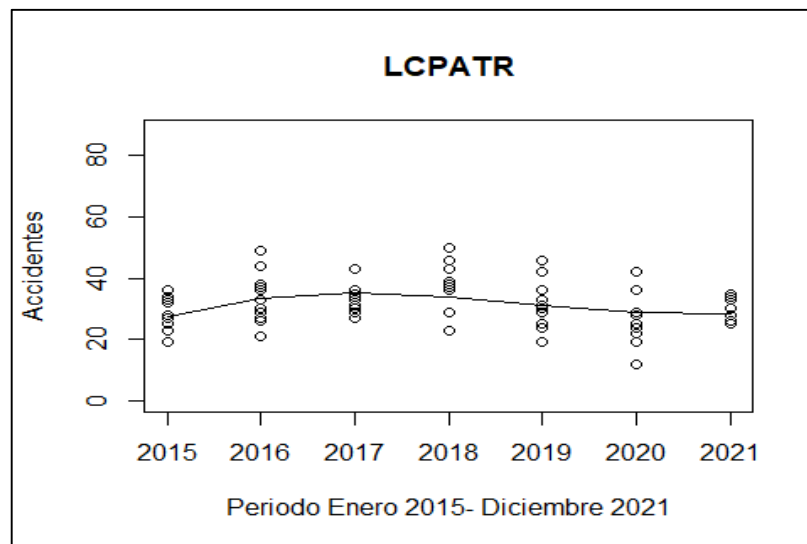
Teniendo el modelo cúbico que es el mod3 como el mejor modelo para realizar predicciones.

#### ❖ Modelo aditivo

$$\log(\text{mes}) = -7.493e + 07 + (1.113e + 05)\text{mes} - (5.516e + 01)\text{mes}^2 + (9.1107e - 03)\text{mes}^3$$

#### ❖ Modelo multiplicativo

$$\text{mes} = e^{-7.493e+07} * e^{((1.113e+05)\text{mes} - (5.516e+01)\text{mes}^2 + (9.1107e-03)\text{mes}^3)}$$

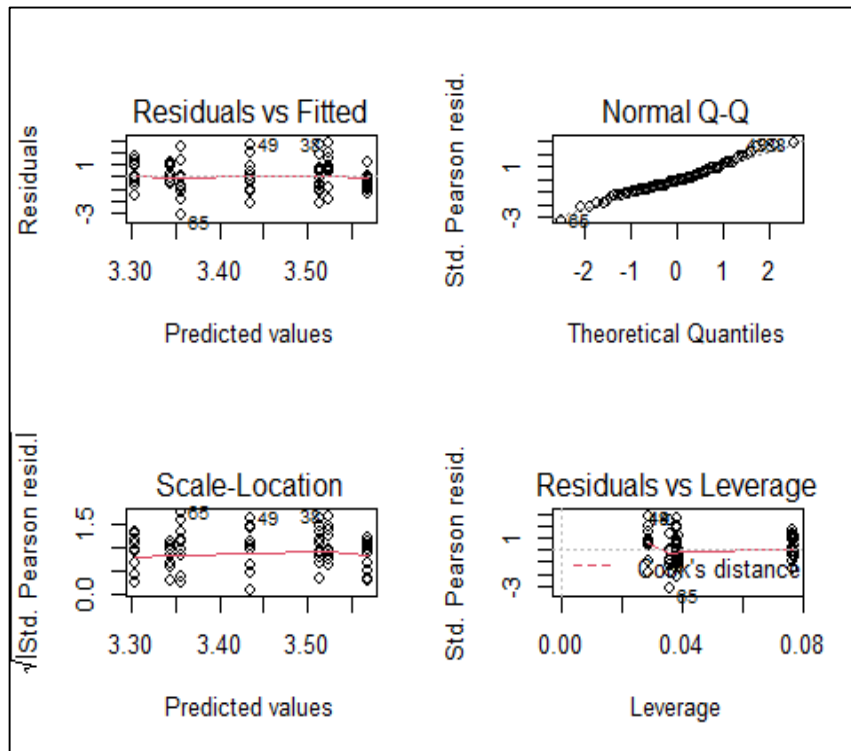


**Ilustración 33-4:** Ajuste del mejor modelo de LCPATR

Realizado por: Inca E y Tierra M, 2022.

En la representación de los valores ajustados por los tres modelos es fácil distinguir que el modelo mod3 se ajusta mejor al número de accidentes LCPATR observados que de los modelos 1 y 2.

#### 4.4.2.3. Comprobación del mejor modelo



**Ilustración 34-4:** Diagnóstico del mejor modelo de LCPATR

Realizado por: Inca E y Tierra M, 2022.

Podemos observar que los residuos no tienen una tendencia estacionaria, es decir están corraladas, la normal se ajusta a la línea de tendencia, el modelo de regresión tiene una calidad predictiva buena, existe homocedasticidad ya que existe varianza de error constante a lo largo del tiempo.

#### ❖ Prueba de Normalidad

$H_0$ : Errores normales

$H_1$ : Errores no normales

**Tabla 35-4:** Prueba de normalidad

Prueba	P-Value
Jarque Bera	0.76

Realizado por: Inca E y Tierra M, 2022.

No se rechaza  $H_0$  ya que el p-value encontrado por la prueba de normalidad de Jarque Bera es de 0,76 mayor a 0.05, se concluye que tiende a una distribución normal.

#### ❖ Prueba de Homocedasticidad

$H_0$ : Existe homocedasticidad

$H_1$ : No existe homocedasticidad

**Tabla 36-4:** Prueba de homocedasticidad

Prueba	P-Value
White	0.93

Realizado por: Inca E y Tierra M, 2022.

Se deduce que mediante la prueba de White con p-value del 0.93 mayor a 0.05, se concluye que existe homocedasticidad, por lo que no se rechaza  $H_0$ .

#### ❖ Prueba de Independencia

$H_0$ : Existe independencia en errores

$H_1$ : No existe independencia en errores

**Tabla 37-4:** Prueba de independencia

Prueba	P-Value
Ljung-Box	0.31

Realizado por: Inca E y Tierra M, 2022.

No se rechaza  $H_0$  ya que mediante la prueba de Ljung-Box se deduce que los errores son independientes por lo que tiene un p-value de 0,31 mayor a 0.05.

### 4.5. Comparación de modelos

#### 4.5.1. Daños materiales (DM)

Para la comparación de modelos se tiene en cuenta el criterio de información Akaike (AIC) y Bayesiano (BIC).

**Tabla 38-4:** Comparación del modelo DM

Modelo	AIC	BIC
ARIMA (0,0,0) (2,1,2) [12]	561	575
POISSON modelo lineal	643	648

Realizado por: Inca E y Tierra M, 2022.

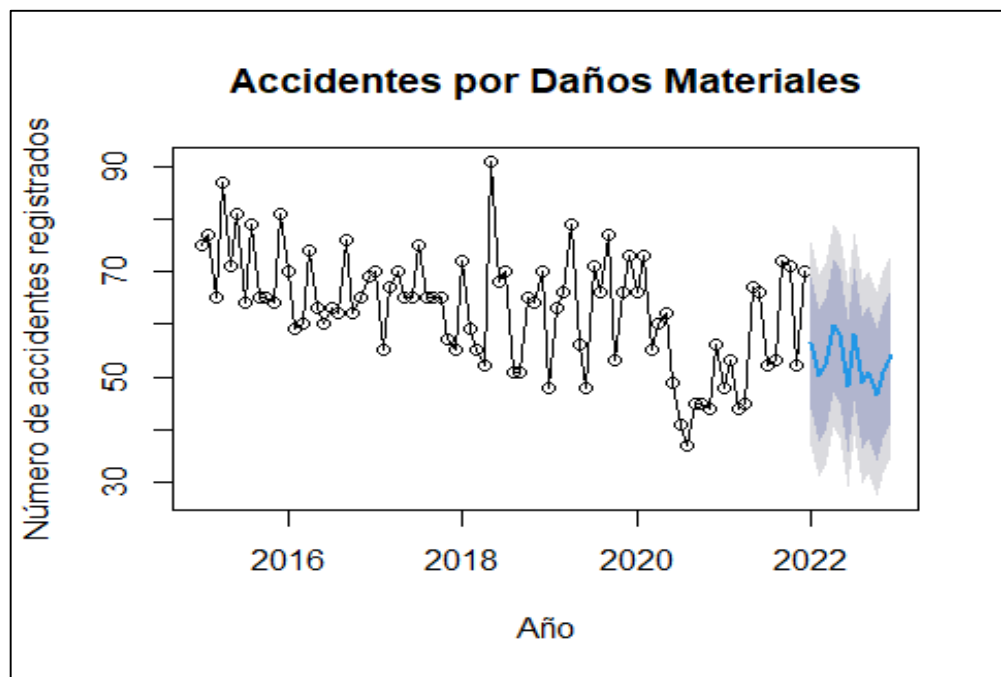
Comparando los resultados determinados en R los valores de AIC y BIC del modelo ARIMA son menores que el modelo de POISSON respectivamente, por lo tanto, el mejor modelo es ARIMA (0,0,0) (2,1,2) [12] con constante, a pronosticar para el año 2022 los accidentes de tránsito causadas por daños materiales.

#### 4.5.1.1. Predicción de los accidentes de tránsito de DM

El modelo ARIMA (0,0,0) (2,1,2) [12] con constante puede representar con la siguiente ecuación misma que nos ayuda a calcular las predicciones.

$$Y_t = -0.2076 + 0.2758Y_{t-12} - 0.0461Y_{t-24} + 1.4465\varepsilon_{t-12} - 0.6399\varepsilon_{t-24} + \varepsilon_t$$

#### ❖ Predicción



**Ilustración 35-4:** Predicción del modelo ARIMA (0,0,0) (2,1,2) [12] *con constante*  
Realizado por: Inca E y Tierra M, 2022.

La Ilustración de predicción muestra el mismo comportamiento de años pasados.

**Tabla 39-4:** Predicciones para el 2022 con el modelo ARIMA (0,0,0) (2,1,2) [12]

Mes	Predicción	Límite inferior 95%	Límite superior 95%
Enero 2022	56	37	76
Febrero 2022	50	31	70
Marzo 2022	52	33	72
Abril 2022	60	41	79
Mayo 2022	58	39	77
Junio 2022	48	29	68
Julio 2022	58	39	78
Agoto 2022	49	30	68
Septiembre 2022	51	32	70
Octubre 2022	47	27	66
Noviembre 2022	51	32	70
Diciembre 2022	54	35	73

Realizado por: Inca E y Tierra M, 2022.

En las predicciones del modelo muestran que en los meses de abril del 2022 existirá 60 accidentes por Daños Materiales, en cambio en el mes de octubre se tendrá menor número de delitos donde se presentará 47 accidentes por Daños Materiales.

#### 4.5.2. Lesiones causadas por accidentes de tránsito (LCPATR)

Para la comparación de modelos se tiene en cuenta el criterio de información Akaike (AIC) y Bayesiano (BIC).

**Tabla 40-4:** Comparación del modelo LCPATR

Modelo	AIC	BIC
ARIMA (1,0,1) (1,0,1) [12]	559	569
POISSON modelo cúbico	560	570

Realizado por: Inca E y Tierra M, 2022.

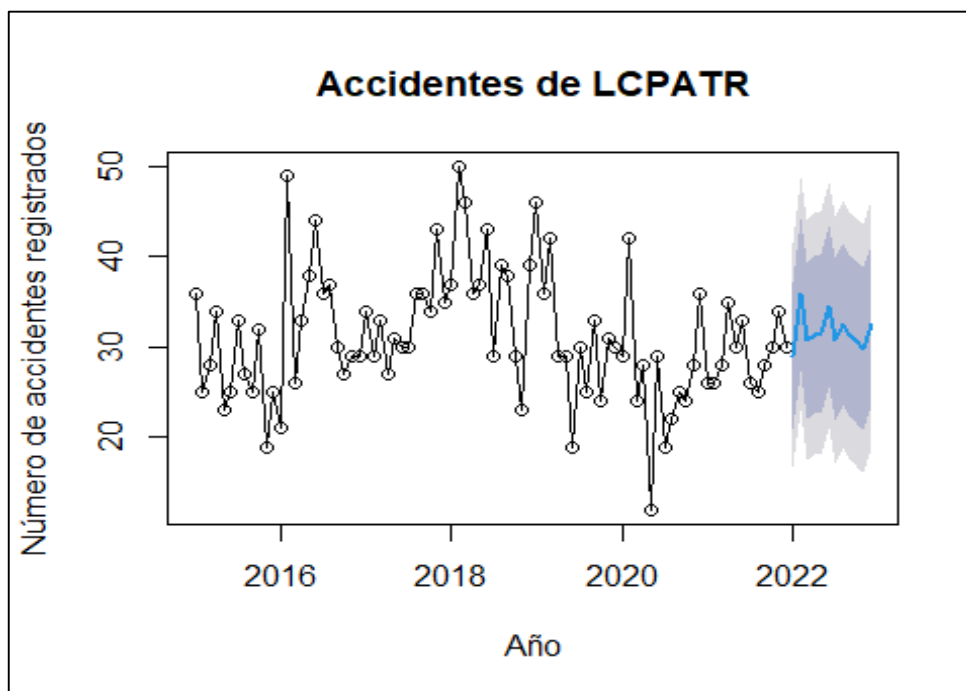
Comparando los resultados determinados en R los valores de AIC y BIC del modelo ARIMA son menores que el modelo de POISSON respectivamente, por lo tanto, el mejor modelo es ARIMA (1,0,1) (1,0,1) [12] con media, a pronosticar para el año 2022 los accidentes de tránsito causadas por daños materiales.

##### 4.5.2.1. Predicción de los accidentes de tránsito de LCPATR

El modelo ARIMA (1,0,1) (1,0,1) [12] con media, puede representarse con la siguiente ecuación misma que nos ayuda a calcular las predicciones.

$$Y_t = 31.2580 + 0.8143Y_{t-1} + 0.5297\varepsilon_{t-1} - 0.9943Y_{t-12} - 0.9502\varepsilon_{t-12} + \varepsilon_t$$

❖ **Predicción**



**Ilustración 36-4:** Predicción del modelo ARIMA (1,0,1) (1,0,1) [12]

Realizado por: Inca E y Tierra M, 2022.

En el gráfico muestra el mismo comportamiento de años pasados.

**Tabla 41-4:** Predicciones para el 2022 con el modelo ARIMA (1,0,1) (1,0,1) [12]

Mes	Predicción	Límite inferior 95%	Límite superior 95%
Enero 2022	29	17	41
Febrero 2022	36	23	49
Marzo 2022	31	17	44
Abril 2022	31	18	45
Mayo 2022	32	18	45
Junio 2022	35	21	48
Julio 2022	31	17	45
Agoto 2022	33	19	46
Septiembre 2022	34	18	45
Octubre 2022	30	17	44
Noviembre 2022	30	16	44
Diciembre 2022	33	19	46

Realizado por: Inca E y Tierra M, 2022.

En las predicciones del modelo muestran que en el mes de febrero del 2022 existirá 36 accidentes por lesiones causadas por accidentes de tránsito, en cambio en el mes de enero con menor número de delitos donde se presentará 29 accidentes por lesiones causadas por accidentes de tránsito.



## CONCLUSIONES

- ❖ Mediante el análisis de la regla del 80-20 o conocida como Ley de Pareto se encontró que la variable DM (daños materiales) representa el 58% de los accidentes de tránsito y la variable LCPATR (lesiones causadas por accidentes de tránsito) representa el 28% de los accidentes de tránsito de la Fiscalía General del Estado de la provincia de Chimborazo, siendo estas las más significativas.
- ❖ Con el análisis exploratorio de datos de las variables DM y LCPATR se detectó varios datos atípicos los mismos que fueron imputados, por el promedio de los meses que corresponden a esos datos.
- ❖ Para el modelo ARIMA y POISSON se analizó los supuestos de normalidad, homocedasticidad e independencia siendo ambos modelos válidos para predicción del año 2022.
- ❖ Mediante el análisis de series de tiempo para la variable DM se obtuvo el modelo ARIMA (0,0,0) (2,1,2) [12] con constante y para la variable LCPATR se obtuvo el modelo ARIMA (1,0,1) (1,0,1) [12] con media y mediante el análisis del modelo lineal generalizado Poisson se obtuvo que para la variable DM es el modelo lineal  $mes = e^{90.58522} * e^{-0.04284mes}$  y para la variable LCPATR es el modelo cúbico  $mes = e^{-7.493e+07} * e^{((1.113e+05)mes - (5.516e+01)mes^2 + (9.1107e-03)mes^3)}$ .
- ❖ Mediante la comparación de los criterios de información AIC y BIC tanto para la variable DM y LCPATR se seleccionó el mejor modelo ARIMA con el cual se realizó predicciones y esto aportará a la Fiscalía Provincial de Chimborazo tener una visión de la realidad y estar preparados para tomar buenas decisiones.

## **RECOMENDACIONES**

- ❖ Se recomienda a la Fiscalía Provincial de Chimborazo recolectar la información proporcionada, utilizando técnicas estadísticas.
- ❖ Se recomienda utilizar más variables a fin de aplicar métodos multivariantes que se ajusten a situaciones de la realidad

## BIBLIOGRAFÍA

**ALDO, FOURNIES.** *Modelos ARMA y Box and Jenkins.* 2015. DOI: 10.13140/RG.2.1.2907.0883.

**ARBOCCÓ FIGUEROA, Giulianna Teresa.** “La fecundidad y su relación con variables”. Perú: s.n., 2005. pp. 19-20.

**AVENDAÑO, Yefry, QUINTERO, Tamara, Andrea S. y GUERRERO, Thomas E.** “Calibración de un modelo econométrico para evaluar la frecuencia en accidentes de tránsito con peatones en la ciudad de Ocaña norte de Santander utilizando distribución POISSON y binomial negativo”. (2016), (Colombia) pp. 8-9.

**BONIFAZ BARAHONA, Jorge Iván.** “Pronósticos del consumo y demanda de potencia máxima de energía eléctrica en la ciudad de Riobamba para el periodo 2017-2020, mediante modelos arima”. (2018), (Riobamba) pp. 114-117.

**CHIRIBOGA, GALO.** “Estatuto orgánico por procesos de la Fiscal General Del Estado”. Ecuador:s.n, 2012. pp. 41-42.

**GORDILLO PROAÑO, Denisse Ambar.** Aplicación de penas alternativas en los accidentes de tránsito con muerte según el COIP. (Quito), (2016) pp. 125-126.

**FAMILYSEARCH.** Familysearch. [En línea]. 2020. [Consulta: 19 de abril 2022.]. Disponible en: [https://www.familysearch.org/es/wiki/Chimborazo,\\_Ecuador\\_-\\_Genealog%C3%ADa](https://www.familysearch.org/es/wiki/Chimborazo,_Ecuador_-_Genealog%C3%ADa).

**GARCÍA, Antonio Ramón y GÓMEZ ESPINOZA, Cesar.** “Comportamiento y pronóstico de la mortalidad en accidentes de tránsito- Ecuador, 2000 – 2018” Ecuador:s.n, 2020.

**IZAR, Juan y GONZALEZ, Jorge.** Diagrama de Pareto. (2004) pp. 8-9.

**MARTÍNEZ, Mónica y MARÍ, Manuel.** La distribución de POISSON. (2022), (Valencia) pp. 9-13.

**MAURICIO, José Alberto.** Análisis de series temporales. (2007), (Madrid) pp. 295-300.

**MCCULLAGH, P. Y NELDER, J.A.** “Generalized Linear Models”. Chicago: s.n., 1991. pp. 526-537.

**MONTESINOS, Abelardo.** “Estudio del AIC y BIC en la selección de modelos de vida con datos censurados”. Guanajuato: s.n., 2011. pp. 65-78.

**OCHOA, Camila Conde y ÁLVAREZ, Ariana E.** “Proyecciones de accidentes en el cantón Cuenca usando un modelo de series temporales ARIMA”. Cuenca: s.n., 2021. p. 175.

**PELHAM BOX, George Edward y JENKINS , Gwilym Meirion.** “Series temporales: modelo ARIMA”. Madrid: s.n., 1970. p. 57.

**ROSARIO CAMONES, Libia Justina y ESPINOZA TORRES, Irma Luisa.** “Análisis estadístico y pronósticos con series temporales de la información de las denuncias de accidente de tránsito, comisaria distrital PNP-Huaraz: 2007-2012”. Perú: s.n., 2013. p. 124.

**SAGGESE, Silvina, BACIGALUPO, y otros.** “Manual de Introducción al Derecho Penal. Primera”. Madrid: s.n., 2019. p. 312.

**SÁNCHEZ, Karina Victoria y MARTÍNEZ, Acurio David.** “Técnicas estadísticas para la modelación y predicción”. Riobamba, 2019. p. 145.

**TIERRA, Tania Fernanda.** “Análisis de series de tiempo y regresión logit del desempleo juvenil en el Ecuador, periodo 2013 – 2018”. Riobamba: s.n., 2020. p. 102.

**TIPÁN, Natalia Vanessa.** “Accidentes de Tránsito producidos por Imprudencia y Negligencia”. Quito: s.n., 2017. pp. 129-137.

**VERDOY, Pablo Juan, BELTRÁN, Modesto Joaquín y PERIS, María José.** *Problemas resueltos de estadística aplicada a las ciencias sociales*. 1ª ed. España: s.n., 2015, pp. 298-301.

**VILLALVA, Carlos Alfredo.** “Predicción del número de delitos de tránsito que ingresan a la Fiscalía Provincial de Chimborazo mediante ARIMA 2018-2021”. Riobamba: s.n., 2021. pp. 68-73.

**VIZCAÍNO, Silvio Toscano.** Derecho Ecuador. [En línea], 2005 [Consulta: 24 de noviembre. 2005]. Disponible en: <https://derechoecuador.com/queacutecute-es-un-accidente-de-traacutensito/#:~:text=Silvio%20Toscano%20Vizca%C3%ADno%20Definciones%20En,que%20ocasiona%20da%C3%B1os%20materiales%20%5B%E2%80%A6%5D>.

**WILLIAMS, Anderson Sweeney.** *Estadística para administración y economía*. 10<sup>ma</sup> ed. Cincinnati: s.n., 2015, pp. 1091-1096.



## ANEXOS

### ANEXO A: AVAL DE LA FISCALÍA PROVINCIAL DE CHIMBORAZO

	<h1>ESPOCH</h1> <p>ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO</p> <p>CARRERA DE ESTADÍSTICA</p>
Riobamba, 31 de enero de 2022	
Ing. Mayra Moreno <b>FISCALÍA PROVINCIAL DE CHIMBORAZO</b> Presente	
De mi consideración:	
<p>Con un atento y cordial saludo, me dirijo a usted con el fin de solicitar muy comedidamente, se otorgue, siempre y cuando sea posible, la base de datos sobre los accidentes mensuales de tránsito que ingresan a la Fiscalía Provincial de Chimborazo desde enero de 2015 hasta diciembre de 2021, esto con la finalidad de que los estudiantes de la carrera de Estadística de la ESPOCH: Mónica Tatiana Tierra Chimbolema (0605313626) y Erick Paul Inca Merino (0604538850), puedan concretar su intención de realizar el trabajo de titulación que han denominado "Análisis y predicción de accidentes mensuales de tránsito en la Provincia de Chimborazo".</p>	
Atentamente <b>"Saber para Ser"</b>	
PABLO JAVIER FLORES MUNOZ	Firmado digitalmente por PABLO JAVIER FLORES MUNOZ Fecha: 2022.01.31 11:10:20 -05'00'
Ing. Pablo Flores Muñoz. MSc. COORDINADOR DE LA CARRERA DE ESTADÍSTICA	
Recibido 31-01-2022 12:00. <i>[Signature]</i>	
Dirección: Panamericana Sur km 1 1/2, Teléfono: 593 (03) 2 998200 Código Postal: B00q0155 <a href="http://www.espoch.edu.ec">www.espoch.edu.ec</a>	

Oficio No.FPH-JGP-2022-000187-O

Riobamba, 01 de febrero de 2022

**Asunto:** ATENCIÓN A REQUERIMIENTO DE INFORMACIÓN

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

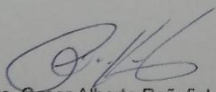
En atención a la petición suscrita por el Ing. Pablo Flores Muñoz, coordinador de la carrera de Estadística - de la Escuela Superior Politécnica de Chimborazo, de cuyo contenido, en lo pertinente, se requiere la entrega de los siguientes datos estadísticos:

- "(...) base de datos sobre los accidentes mensuales de tránsito que ingresan a la Fiscalía Provincial de Chimborazo, desde enero de 2015 hasta diciembre de 2021 (...)"

Al respecto, de la captura de información realizada en el Sistema Integrado de Actuaciones Fiscales SIAF 0.1 se procede a remitir a los correos monica.tierrach@esPOCH.edu.ec y erick.inca@esPOCH.edu.ec la información respectiva (mes a mes), de las causas ingresadas como noticias del delito por accidentes de tránsito, en todas las unidades fiscales especializadas y cantonales de la provincia de Chimborazo, en el periodo 2015 - 2021.

Particular que pongo en su conocimiento para los fines pertinentes.

Atentamente,

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



Referencia: FPH-GD-2022-000143-EXT

Fecha de elaboración	Elaborado por:	Revisado por:	Aprobado por:
2022-02-01 16:22:10	Peñafiel Andino Cesar Alberto	Peñafiel Andino Cesar Alberto	Peñafiel Andino Cesar Alberto

**FGE FISCALÍA DE CHIMBORAZO**  
**RECIBIDO HOY**

01 FEB 2022

SIENDO LAS 16:426 FS H/1  
SECRETARÍA PROVINCIAL

## ANEXO B: CÓDIGO EN EL ENTORNO COMPUTACIONAL

```
library(openxlsx)
library(tidyverse)
library(qcc)
library(knitr)
library(tseries)
library(forecast)
library(urca)
library(tseries)
deltran <- read.csv("deltran.csv",header=T, sep=";")
## serie de tiempo
deltran_ts <- ts(data = deltran, start = c(2015,1), frequency = 12)
plot(deltran_ts, main="DM",xlab="Periodo 2015-2021",ylab="Accidentes")
da <- boxplot.stats(x = deltran_ts)
boxplot(deltran_ts, horizontal = T)
quantile(deltran_ts)
# descomposicion
des <- decompose(deltran_ts, type = " multiplicative ")
plot(des)
## grafica por mes
monthplot(deltran_ts)
# diferenciacion
ur.kpss(deltran_ts)
kpss.test(deltran_ts)
ur.kpss(diff(deltran_ts,12))
# parte regular
ndiffs(deltran_ts)
# parte estacional
nsdiffs(deltran_ts)
ggAcf(diff(deltran_ts,1),main="FUNCION DE AUTOCORRELACION SIMPLE")
ggPacf(diff(deltran_ts,1),main="FUNCION DE AUTOCORRELACION PARCIAL")
ggAcf(diff(diff(deltran_ts,12)),main="FUNCION DE AUTOCORRELACION SIMPLE")
ggPacf(diff(diff(deltran_ts,12)),main="FUNCION DE AUTOCORRELACION PARCIAL")
#####auto arima
m1<-auto.arima(deltran_ts, D=1, stepwise=T,approximation=T, trace = T)
```



```

#### modelos propuestos
m2<-Arima(deltran_ts,order = c(0,0,0),seasonal = c(1,1,2))
m3<-Arima(deltran_ts,order = c(1,0,0),seasonal = c(1,1,1))
m4<-Arima(deltran_ts,order = c(0,0,0),seasonal = c(0,1,0))
m5<-Arima(deltran_ts,order = c(0,0,1),seasonal = c(0,1,2))
### criterios de informacion AIC y BIC
AIC(m1,m2,m3,m4,m5)
BIC(m1,m2,m3,m4,m5)
##representación grafica y proyeccion para 2022
plot(forecast(m1,h=12),xlab="Año",ylab="Número de accidentes registrados"
      ,main="Accidentes por Daños Materiales",type="o")
### datos de delitos proyectados para 2022
pro<- forecast(m1,h=12)
pro
#supuestos
tsdiag(m1)
resia<-as.vector(residuals(m1))
jarque.bera.test (resia)
Box.test(resia, type="Ljung-Box")
white.test(m1$residuals)
#####
# analisis de poisson
accitra <-read.csv("datat.csv",header=T, sep=";")
plot(accitra$Accidentes~accitra$mes)
## identificacion de modelos
mod0 <- glm(accitra$Accidentes~accitra$mes, family=poisson)
## diagnostico del modelo
plot(mod0)
## identificacion de modelos
mod1<-glm(accitra$Accidentes ~ accitra$mes+I(accitra$mes^2),family=poisson)
## diagnostico del modelo
plot(mod1)
mod2<-glm(accitra$Accidentes
accitra$mes+I(accitra$mes^2)+I(accitra$mes^3),family=poisson)
plot(mod2)
## criterios de informacion

```

~

```

BIC(mod0,mod1,mod2)
AIC(mod0,mod1,mod2)
plot(accitra$mes,accitra$Accidentes, ylim=c(0,280))
lines(mod2$fitted~accitra$mes)
## ANALISIS DE LA VARIABLE LCPATR
d<-read.csv("lcpatr_.csv",header=T, sep=";")
# serie de tiempo
d_ts <- ts(data = d, start = c(2015,1), frequency = 12)
plot(d_ts, main="DM",xlab="Periodo 2015-2021",ylab="Accidentes")
# box plot
da <- boxplot.stats(x = d_ts)
boxplot(d_ts, horizontal = T)
# descomposicion de la st
des <- decompose(d_ts, type = "multiplicative")
plot(des)
## grafica por mes
monthplot(d_ts)
### diferenciacion
ur.kpss(d_ts)
kpss.test(d_ts)
ur.kpss(diff(d_ts,12))
# parte regular
ndiffs(d_ts)
# parte estacional
nsdiffs(d_ts)
ggAcf(diff(d_ts,1),main="FUNCION DE AUTOCORRELACION SIMPLE")
ggPacf(diff(d_ts,1),main="FUNCION DE AUTOCORRELACION PARCIAL")
ggAcf(diff(diff(d_ts,12)),main="FUNCION DE AUTOCORRELACION SIMPLE")
ggPacf(diff(diff(d_ts,12)),main="FUNCION DE AUTOCORRELACION PARCIAL")
##### auto arima
md1<-auto.arima(d_ts,stepwise=T,approximation=T, trace = T)
#### modelos propuestos
#propuesta de otros modelos
md2<-Arima(d_ts,order = c(1,0,0),seasonal = c(2,0,0))
md3<-Arima(d_ts,order = c(2,0,1),seasonal = c(0,0,1))
md4<-Arima(d_ts,order = c(1,0,2),seasonal = c(1,0,1))

```

```

md5<-Arima(d_ts,order = c(1,0,1),seasonal = c(1,0,1))
### criterios de informacion AIC y BIC
AIC(md1,md2,md3,md4,md5)
BIC(md1,md2,md3,md4,md5)
##representación grafica y proyeccion para 2022
plot(forecast(md5,h=12),xlab="mes",ylab="Número de accidentes registrados"
     ,main="Accidentes de LCPATR",type="o")
### datos de delitos proyectados para 2022
pro<- forecast(md5,h=12)
pro
#supuestos
tsdiag(md5)
resia<-as.vector(residuals(md5))
jarque.bera.test (resia)
Box.test(resia, type="Ljung-Box")
white.test(md5$residuals)
#####
## ANALISIS POISSON
acc <-read.csv("lcpatr.csv",header=T, sep=";")
plot(acc$mes,acc$Accidentes)
## identificacion de modelos
mo1 <- glm(acc$Accidentes ~acc$mes, family=poisson)
## diagnostico del modelo
plot(mo1)
## identificacion de modelos
mo2<-glm(acc$Accidentes ~ acc$mes+I(acc$mes^2),family=poisson)
## diagnostico del modelo
plot(mo2)
## identificacion de modelos
mo3<-glm(acc$Accidentes ~ acc$Ames+I(acc$mes^2)+I(acc$mes^3),family=poisson)
## diagnostico del modelo
plot(mo3)
## criterios de informacion
AIC(mo1,mo2,mo3)
BIC(mo1,mo2,mo3)
## diagnostico del mejor modelo

```

```
plot(acc$mes,acc$Accidentes, ylim=c(0,88))  
lines(mo3$fitted~acc$mes)
```



**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:** 18 / 02 / 2022

<b>INFORMACIÓN DEL AUTOR/A (S)</b>
<b>Nombres – Apellidos: Mónica Tatiana Tierra Chimbolema Erick Paul Inca Merino</b>
<b>INFORMACIÓN INSTITUCIONAL</b>
<b>Facultad: Ciencias</b>
<b>Carrera: Estadística</b>
<b>Título a optar: Ingeniero/a en Estadística</b>
<b>f. Analista de Biblioteca responsable: Ing. Rafael Inty Salto Hidalgo</b>

0037-DBRA-UPT-2023