



**UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO**

**UNIDAD ACADÉMICA PROFESIONAL HUEHUETOCA**

**“PRONÓSTICO DE ACCIDENTES VIALES EN LA  
AUTOPISTA MÉXICO – QUERÉTARO PARA EL AÑO 2022  
UTILIZANDO EL MODELO ARIMA Y LA TÉCNICA DE  
VENTANAS DE TIEMPO”**

**T E S I S**

QUE PARA OBTENER EL TÍTULO DE:

**LICENCIADO EN ACTUARÍA**

PRESENTA:

**YURIDIA ERIZA MALDONADO**

ASESOR:

**M. EN I. SAMUEL MARTÍNEZ BELLO**

REVISORES:

**M. EN C. MAT. ROBERTO ALBARRÁN GARCÍA**

**DR. EN E. CHRISTIAN BUCIO PACHECO**

HUEHUETOCA, EDO. DE MÉXICO, SEPTIEMBRE 2025



UAP Huehuetoca

## Resumen

En México, la siniestralidad vial constituye un problema persistente de salud pública y seguridad. En particular, la autopista México–Querétaro se distingue por su alta frecuencia de accidentes, atribuida tanto al elevado flujo vehicular como a factores estructurales propios de la vía. Este trabajo propone un enfoque estadístico que combina el modelo SARIMA (Seasonal Auto Regressive Integrated Moving Average), diseñado para capturar patrones estacionales en series temporales, con la técnica de ventanas de tiempo, que permite mejorar la capacidad del modelo para adaptarse a cambios locales en la dinámica de los datos.

La investigación se basó en una serie temporal mensual que abarca el periodo de enero de 2012 a diciembre de 2021, obtenida de los registros oficiales de la Secretaría de Comunicaciones y Transportes. Se construyó un modelo SARIMA ajustado a la estacionalidad observada en la serie, y posteriormente se implementó la técnica de ventanas deslizantes con distintas longitudes. La precisión del pronóstico fue evaluada comparando los resultados con los datos reales del año 2022.

Los hallazgos indican que el uso conjunto de SARIMA y ventanas temporales mejora sustancialmente la precisión predictiva del modelo, ya que permite capturar con mayor sensibilidad las fluctuaciones de corto plazo. Este enfoque representa una herramienta analítica valiosa para la planificación preventiva, la toma de decisiones estratégicas y el diseño de políticas públicas orientadas a reducir la siniestralidad en rutas de alto riesgo.

## **Abstract**

In Mexico, road traffic accidents represent an ongoing public health and safety issue. In particular, the Mexico–Querétaro highway stands out for its high frequency of accidents, which is attributed to both the heavy traffic flow and the structural characteristics of the road itself. This study proposes a statistical approach that combines the SARIMA model (Seasonal Auto Regressive Integrated Moving Average), designed to capture seasonal patterns in time series, with the rolling window technique, which enhances the model’s ability to adapt to local changes in data dynamics.

The research was based on a monthly time series spanning from January 2012 to December 2021, obtained from the official records of the Secretariat of Communications and Transportation. A SARIMA model was fitted to the seasonality observed in the series, and subsequently, the rolling window technique was implemented using different window lengths. Forecast accuracy was evaluated by comparing the results with the actual data from the year 2022.

The findings indicate that the combined use of SARIMA and rolling windows substantially improves the predictive accuracy of the model, as it allows for greater sensitivity in capturing short-term fluctuations. This approach represents a valuable analytical tool for preventive planning, strategic decision-making, and the design of public policies aimed at reducing traffic accidents on high-risk routes.

# Índice general

Introducción . . . . .	vii
<b>1. Sustento Teórico y Estadístico para la Modelación SARIMA en Series Temporales de Accidentes</b>	<b>1</b>
1.1. Siniestralidad . . . . .	1
1.2. Series temporales . . . . .	4
1.2.1. Fundamentos de series temporales en la predicción de accidentes . . . . .	4
1.2.2. Tendencia . . . . .	5
1.2.3. Ciclos . . . . .	5
1.2.4. Estacionalidad . . . . .	5
1.2.5. Estacionariedad . . . . .	5
1.3. Prueba de Dickey-Fuller y raíces unitarias . . . . .	6
1.4. Prueba de Ljung-Box . . . . .	7
1.5. Análisis de siniestralidad en series temporales . . . . .	8
1.6. Modelos de series temporales: AR, MA, ARMA, ARIMA y SARIMA . . . . .	8
1.7. Ventanas Temporales o Ventanas de Tiempo . . . . .	11
1.8. Estadísticas descriptivas . . . . .	13
1.9. Procesos estocásticos . . . . .	14
1.9.1. Ruido Blanco . . . . .	14
1.10. Factores de riesgo en accidentes viales . . . . .	15
1.10.1. Modelos ARIMA para el Pronóstico de Accidentes Viales . . . . .	15
1.10.2. Ventanas de Tiempo en el Pronóstico de Accidentes Viales . . . . .	16
1.11. Validación cruzada en modelos SARIMA . . . . .	16
1.11.1. Análisis de la función de autocorrelación . . . . .	17
1.12. Métodos de identificación, estimación, diagnóstico y pronóstico de accidentes viales . . . . .	17
<b>2. Aplicación Empírica del Modelo SARIMA en el Pronóstico de Accidentes Viales</b>	<b>19</b>
2.1. Obtención de datos . . . . .	19
2.2. Preparación de los Datos . . . . .	20

---

2.3.	Análisis Exploratorio de Datos . . . . .	21
2.3.1.	Visualización de tendencias y patrones estacionales . . . . .	21
2.3.2.	Estadísticas descriptivas . . . . .	22
2.4.	ARIMA (Autorregresivo Integrado de Media Móvil) . . . . .	22
2.4.1.	Diferenciación y Estacionariedad . . . . .	22
2.4.2.	Análisis ACF y PACF . . . . .	23
2.4.3.	Evaluación del modelo ARIMA estacional (SARIMA) . . . . .	24
2.5.	Optimización del modelo mediante ventanas de tiempo . . . . .	25
2.5.1.	Selección del tamaño de la ventana . . . . .	25
2.5.2.	Aplicación del modelo a las ventanas de tiempo . . . . .	26
2.5.3.	Evaluación de las predicciones de las ventanas temporales . . . . .	27
<b>3.</b>	<b>Resultados</b>	<b>28</b>
3.1.	Análisis descriptivos de datos . . . . .	28
3.2.	Visualización de la serie temporal . . . . .	30
3.3.	Verificación de la Estacionariedad: Prueba ADF . . . . .	31
3.4.	Análisis de la Función de Autocorrelación y la Función de Autocorrelación Parcial	32
3.5.	Determinación de los parámetros del modelo SARIMA . . . . .	36
3.6.	Ajuste del modelo SARIMA . . . . .	38
3.7.	Pronóstico de Accidentes Viales para el Año 2022 . . . . .	39
3.8.	Comparativa de Predicciones y Datos Reales de Accidentes para 2022 . . . . .	43
3.8.1.	Datos Reales de Accidentes para 2022 . . . . .	43
3.8.2.	Análisis de la Comparativa . . . . .	44
3.9.	Análisis de los Residuos del Modelo SARIMA . . . . .	45
3.10.	Comprobación de la Normalidad de los Residuos y de la Media Cero y Varianza Constante . . . . .	47
3.10.1.	Prueba de Normalidad de Shapiro-Wilk . . . . .	48
3.10.2.	Prueba de Media Cero y Varianza Constante: Prueba de Ljung-Box . . . . .	48
3.11.	Ajuste del modelo con ventanas temporales . . . . .	49
3.11.1.	Selección de longitudes de ventana . . . . .	49
3.11.2.	Procedimiento aplicado por ventana . . . . .	50
3.11.3.	Comparativa para el pronóstico de la ventana 15 . . . . .	51
3.11.4.	Fragmento de datos pronosticados por ventana . . . . .	51
3.11.5.	Análisis detallado por longitud . . . . .	53
3.11.6.	Pronósticos tabulados para la ventana 15 en cada longitud . . . . .	57
3.11.7.	Conclusión general del ajuste con ventanas . . . . .	58

---

Conclusiones . . . . .	59
Referencias . . . . .	62
Anexos . . . . .	66

## Índice de figuras

3.1. Tendencia del número de accidentes (periodo de entrenamiento) . . . . .	30
3.2. Prueba ADF antes de diferenciar la serie . . . . .	31
3.3. Prueba ADF después de aplicar diferenciación . . . . .	32
3.4. Gráfico de la función de autocorrelación (ACF) de la serie original . . . . .	33
3.5. Gráfica de la Función de Autocorrelación (ACF) de la Serie Diferenciada . . . . .	34
3.6. Función de Autocorrelación Parcial de la serie original . . . . .	35
3.7. Función de Autocorrelación Parcial de la serie diferenciada . . . . .	36
3.8. Modelo SARIMA ajustado con orden (1,1,1)(1,1,1)[12] . . . . .	38
3.9. Pronóstico de accidentes viales en la autopista México-Querétaro para el año 2022, con intervalo de confianza al 90% . . . . .	41
3.10. Pronóstico de accidentes viales en la autopista México-Querétaro para el año 2022, con intervalo de confianza al 95% . . . . .	42
3.11. Pronóstico de accidentes viales en la autopista México-Querétaro para el año 2022, con intervalo de confianza al 99% . . . . .	42
3.12. Gráfico de los residuos del modelo SARIMA ajustado . . . . .	46
3.13. ACF de los residuos del modelo SARIMA ajustado . . . . .	47
3.14. Prueba de normalidad de los residuos con Shapiro-Wilk . . . . .	48
3.15. Prueba de independencia de residuos mediante Box-Ljung . . . . .	49
3.16. Pronóstico de ventana 15 de longitud 48 con intervalos al 99% . . . . .	54
3.17. Residuos de la ventana 15 de longitud 48 . . . . .	54
3.18. Pronóstico de ventana 15 de longitud 50 con intervalos al 99% . . . . .	55
3.19. Residuos de la ventana 15 de longitud 50 . . . . .	55
3.20. Pronóstico de ventana 15 de longitud 60 con intervalos al 99% . . . . .	56
3.21. Residuos de la ventana 15 de longitud 60 . . . . .	56

## Índice de tablas

3.1. Comparación de modelos SARIMA . . . . .	39
3.2. Predicciones con intervalo de confianza al 90% para 2022 . . . . .	40
3.3. Predicciones con intervalo de confianza al 95% para 2022 . . . . .	40
3.4. Predicciones con intervalo de confianza al 99% para 2022 . . . . .	41
3.5. Datos reales de accidentes viales para 2022 . . . . .	44
3.6. Comparación de rendimiento entre ventanas (ventana 15) . . . . .	51
3.7. Resultados de las pruebas estadísticas por ventana para longitud 48 . . . . .	52
3.8. Resultados de las pruebas estadísticas por ventana para longitud 50 . . . . .	52
3.9. Resultados de las pruebas estadísticas por ventana para longitud 60 . . . . .	53
3.10. Pronóstico de la ventana 15 con longitud 48 . . . . .	57
3.11. Pronóstico de la ventana 15 con longitud 50 . . . . .	57
3.12. Pronóstico de la ventana 15 con longitud 60 . . . . .	58
3.13. Diferenciación de Primer Orden (Parte 1) . . . . .	66
3.14. Diferenciación de Primer Orden (Parte 2) . . . . .	67
3.15. Diferenciación de Primer Orden (Parte 3) . . . . .	67
3.16. Diferenciación de primer orden de la serie temporal (1 de 4 y 2 de 4) . . . . .	68
3.17. Diferenciación de primer orden de la serie temporal (3 de 4) . . . . .	69
3.18. Diferenciación de primer orden de la serie temporal (4 de 4) . . . . .	70
3.19. Resultados de la prueba ADF para las ventanas con longitud de 48 (parte 1) . .	71
3.20. Resultados de la prueba ADF para las ventanas con longitud de 48 (parte 2) . .	72
3.21. Resultados de la prueba ADF para las ventanas con longitud de 48 (parte 3) . .	73
3.22. Resultados de la prueba de Shapiro-Wilk para las ventanas con longitud de 48 (parte 1) . . . . .	74
3.23. Resultados de la prueba de Shapiro-Wilk para las ventanas con longitud de 48 (parte 2) . . . . .	75
3.24. Resultados de la prueba de Shapiro-Wilk para las ventanas con longitud de 48 (parte 3) . . . . .	76
3.25. Resultados de la prueba Box-Ljung para las ventanas con longitud de 48 (Parte 1)	77

---

3.26. Resultados de la prueba Box-Ljung para las ventanas con longitud de 48 (Parte 2)	78
3.27. Resultados de la prueba Box-Ljung para las ventanas con longitud de 48 (Parte 3)	79
3.28. Resultados de la prueba ADF (Dickey-Fuller) con ventanas de longitud 50 (Parte 1)	80
3.29. Resultados de la prueba ADF (Dickey-Fuller) con ventanas de longitud 50 (Parte 2)	81
3.30. Resultados de la prueba ADF (Dickey-Fuller) con ventanas de longitud 50 (Parte 3)	82
3.31. Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 50 (Parte 1)	83
3.32. Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 50 (Parte 2)	84
3.33. Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 50 (Parte 3)	85
3.34. Resultados de la prueba de independencia Box-Ljung con ventanas de longitud 50 (Parte 1)	86
3.35. Resultados de la prueba de independencia Box-Ljung con ventanas de longitud 50 (Parte 2)	87
3.36. Resultados de la prueba de independencia Box-Ljung con ventanas de longitud 50 (Parte 3)	88
3.37. Resultados de la prueba de Dickey-Fuller para las ventanas con longitud de 60	89
3.38. Resultados de la prueba de Dickey-Fuller para las ventanas con longitud de 60 (continuación)	90
3.39. Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 60 (Parte 1)	91
3.40. Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 60 (Parte 2)	92
3.41. Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 60 (Parte 3)	93
3.42. Resultados de la prueba de autocorrelación Box-Ljung con ventanas de longitud 60 (Parte 1)	94
3.43. Resultados de la prueba de autocorrelación Box-Ljung con ventanas de longitud 60 (Parte 2)	95
3.44. Resultados de la prueba de autocorrelación Box-Ljung con ventanas de longitud 60 (Parte 3)	96

---

## Introducción

La siniestralidad vial en México representa un problema estructural con profundas implicaciones sociales, económicas y de salud pública. En el país, los accidentes de tránsito ocasionan cada año miles de muertes y lesiones que afectan tanto a las víctimas como a sus familias, además de generar un importante costo económico y social. En 2016, por ejemplo, se registraron 372,618 víctimas de accidentes viales, una cifra que pone de manifiesto la magnitud del problema y la urgencia de implementar estrategias más efectivas de prevención y mitigación de riesgos (AMIS, 2019).

Entre las vías de comunicación más críticas en este contexto se encuentra la autopista México–Querétaro. Esta carretera es una de las más transitadas del país y conecta de manera estratégica la capital con el Bajío, lo que la convierte en un punto clave para el comercio y el transporte de personas y mercancías. Sin embargo, su alta afluencia vehicular, las características de su infraestructura y factores externos como condiciones climáticas o sociales derivan en una elevada incidencia de accidentes. Entre 2018 y 2022 se registraron más de mil siniestros en este tramo, con alrededor de 140 fallecimientos y cuantiosas pérdidas materiales, según datos de la Secretaría de Comunicaciones y Transportes (SCT, s.f.) y diversos reportes periodísticos (Chávez, 2025; Estévez, 2025; Guzmán, 2025). Estos eventos, que incluyen colisiones frontales de autobuses, choques múltiples de tráileres y percances en casetas de cobro, evidencian la vulnerabilidad de la autopista y la necesidad de generar herramientas que permitan anticipar y gestionar estos riesgos.

El análisis de los datos históricos disponibles ofrece una oportunidad valiosa para comprender el comportamiento de la siniestralidad y generar pronósticos que ayuden en la toma de decisiones. Sin embargo, predecir el número de accidentes de tránsito no es una tarea sencilla. Este fenómeno está influido por factores cambiantes como el clima, las condiciones socioeconómicas, el flujo vehicular y el estado de la infraestructura, lo que introduce variabilidad y complejidad en las series de datos. Para abordar este reto, es indispensable utilizar metodologías robustas de análisis de series temporales que permitan capturar patrones y estacionalidades presentes en la información histórica.

Entre estas metodologías, los modelos ARIMA (Auto Regressive Integrated Moving Average) se consolidan como herramientas ampliamente utilizadas en el pronóstico de fenómenos temporales. Cuando la serie presenta estacionalidad, como ocurre con los registros mensuales de accidentes, se requiere una extensión del modelo conocido como SARIMA, que integra componentes estacionales y no estacionales, permitiendo identificar patrones cíclicos y realizar estimaciones más precisas. No obstante, aun con este modelo, las fluctuaciones locales pueden dificultar la precisión de los pronósticos. Por ello, este trabajo incorpora la técnica de ventanas

---

de tiempo, que consiste en segmentar la serie histórica en subconjuntos móviles y ajustar el modelo en cada segmento. Este enfoque permite capturar mejor los cambios en la dinámica del fenómeno y generar pronósticos más estables y adaptados a las condiciones específicas de cada periodo.

El objetivo principal de este trabajo es desarrollar un modelo predictivo SARIMA que permita estimar el número de accidentes mensuales en la autopista México–Querétaro, utilizando datos históricos del periodo 2012–2021, y evaluar su capacidad de pronóstico para el año 2022. Para mejorar el desempeño del modelo, implementamos la técnica de ventanas de tiempo con diferentes longitudes, lo que posibilita una comparación detallada de su efectividad en la generación de predicciones. Los resultados obtenidos se contrastan con los datos reales de 2022 y se evalúan a niveles de significancia del 90 %, 95 % y 99 %, con el fin de validar la precisión y utilidad de la propuesta.

El desarrollo y ajuste del modelo se realiza con el lenguaje de programación R, empleando la plataforma R Studio, lo que permite implementar de manera eficiente el proceso de estimación, análisis y validación de los pronósticos. Esta herramienta facilita no solo el ajuste técnico del modelo, sino también la obtención de indicadores estadísticos para evaluar su desempeño y la construcción de intervalos de confianza que respaldan la solidez de los resultados.

Este modelo no solo busca ser un ejercicio académico, sino también una herramienta práctica para las autoridades y organismos encargados de la seguridad vial. Con pronósticos más confiables, se pueden diseñar políticas preventivas más efectivas, optimizar la asignación de recursos y planificar operativos en los meses de mayor riesgo. Así, este trabajo contribuye en la gestión del riesgo y en la reducción de las consecuencias humanas y materiales derivadas de los accidentes.

La hipótesis central que guía este trabajo plantea que el modelo SARIMA, mejorado con la técnica de ventanas de tiempo, ofrece un nivel de precisión aceptable para pronosticar el número de accidentes mensuales en la autopista México–Querétaro. Si los pronósticos generados mantienen un margen de error razonable al compararse con los datos reales, y los intervalos de confianza a los niveles de significancia analizados resultan adecuados, el modelo puede considerarse una herramienta confiable para la planeación y la prevención.

Además de su relevancia práctica, este trabajo tiene un valor académico al explorar y validar un enfoque metodológico que combina modelos estadísticos tradicionales con técnicas de segmentación temporal. Esta propuesta no solo se enfoca en el análisis del pasado, sino que también ofrece un marco para la mejora continua del pronóstico mediante ajustes dinámicos en función de la evolución de la serie temporal.

El contenido del este trabajo lo organizamos de manera que permita al lector comprender de forma gradual el contexto, el enfoque metodológico, los resultados y las conclusiones alcan-

---

zadas. El primer capítulo presenta una revisión del contexto teórico, abordando los conceptos fundamentales relacionados principalmente con la siniestralidad vial, los modelos ARIMA y SARIMA, y la técnica de ventanas de tiempo. El segundo capítulo describe de manera detallada los datos utilizados, su origen, las transformaciones realizadas, el proceso de preparación de la información para el análisis y se explica la metodología implementada, describiendo los fundamentos estadísticos del modelo, su ajuste y la aplicación de las ventanas de tiempo con diferentes longitudes. Posteriormente, el tercer capítulo presenta los resultados obtenidos, con especial énfasis en el análisis comparativo de las predicciones y su validación con los datos reales del año 2022. Finalmente, discutimos los hallazgos, establecemos conclusiones y proponemos líneas de investigación futuras, así como posibles aplicaciones del modelo en otros contextos donde se requiera anticipar la ocurrencia de eventos dinámicos y estacionales.

En conjunto, este trabajo aporta evidencia empírica sobre el uso del modelo SARIMA en la predicción de accidentes viales y demuestra cómo su integración con la técnica de ventanas de tiempo genera pronósticos más robustos y adaptativos. Al ofrecer un análisis detallado y una propuesta metodológica replicable, este trabajo sienta bases para el desarrollo de herramientas estadísticas que contribuyan a la toma de decisiones en materia de seguridad vial, con el fin último de reducir los impactos negativos de los accidentes y fomentar una movilidad más segura en México.

# **1 Sustento Teórico y Estadístico para la Modelación SARIMA en Series Temporales de Accidentes**

En este capítulo expondremos los fundamentos teóricos que permiten el adecuado desarrollo de la investigación. Se abordan conceptos para comprender el enfoque analítico y metodológico para alcanzar los objetivos establecidos en el capítulo anterior.

Mostramos en primera instancia los conceptos de siniestralidad que es la clave del estudio, posterior a esto, se exploran los fundamentos de las series temporales y se mencionan algunos de los modelos estadísticos utilizados en el análisis de las series temporales. También se añaden pruebas estadísticas que permitirán evaluar los modelos a desarrollar. Además, se proporcionan algunos conceptos más que son complementarios al desarrollo teórico de la investigación.

## **1.1 Siniestralidad**

Un siniestro se define como un evento inesperado que ocasiona algún tipo de daño, pérdida o accidente. Estos eventos pueden afectar tanto a las personas como a los bienes materiales y pueden ocurrir en una variedad de contextos, como accidentes de tráfico y desastres naturales. El término siniestro es amplio y puede variar dependiendo el tipo de riesgo. En el ámbito de la seguridad vial, se asocia principalmente a los accidentes de tránsito.

En este sentido, la siniestralidad vial se refiere a cualquier siniestro que involucre a vehículos en movimiento, ya sea en vías públicas o en otras áreas de acceso público. Estos eventos no solo afectan a los vehículos, sino también a otros usuarios de la vía, como peatones, ciclistas, conductores, motociclistas y pasajeros, según los datos proporcionados en las estadísticas de la SCT (SCT, s.f.).

En términos matemáticos, la siniestralidad se puede representar como el número de accidentes ocurridos en un intervalo de tiempo específico, dividido por la cantidad de vehículos, personas o distancias recorridas en el mismo período.

La siguiente fórmula permite comparar la siniestralidad en diferentes contextos, ya sea entre diferentes ciudades, diferentes tipos de carreteras o distintos períodos de tiempo (MAPFRE, 2023).

---

$$\text{Tasa de siniestralidad} = \frac{\text{Número de accidentes}}{\text{Número de vehículos o kilómetros recorridos}} \quad (1.1)$$

La siniestralidad no es un fenómeno que dependa exclusivamente del comportamiento de los conductores. Existen diversos factores que contribuyen al riesgo de siniestros, según la Secretaría de Comunicaciones y Transportes (SCT, s.f.), los cuales pueden clasificarse en varios grupos:

■ **Siniestros causados por el conductor o peatón:**

En estos incluyen errores de juicio, distracción, fatiga, consumo de alcohol o drogas, así como actitudes irresponsables por parte de los conductores.

Entre las principales causas de estos siniestros se encuentran:

- Imprudencia al adelantar o invadir el carril contrario.
- Exceso de velocidad, lo que reduce el tiempo de reacción ante posibles imprevistos.
- No respetar señales de tránsito, como semáforos en rojo o señales de alto.
- Manejar bajo el efecto del alcohol o drogas, lo que afecta las capacidades cognitivas y motoras.
- Distracción del conductor o del peatón, como el uso del teléfono móvil.
- Fatiga o somnolencia al volante.
- No ceder el paso cuando corresponde, provocando colisiones innecesarias.
- Condiciones meteorológicas adversas, como lluvia intensa, que reducen la visibilidad y el control del vehículo.
- Factores vehiculares: Son los relacionados con el estado del vehículo, como el mal mantenimiento, fallos mecánicos o la calidad de los neumáticos.

A continuación, se presentan algunas de las fallas comunes de los vehículos que contribuyen a la siniestralidad:

- Neumáticos en mal estado o desinflados, que aumentan el riesgo de deslizamiento o pérdida de control.
- Fallos en el sistema de frenos, que dificultan la capacidad de detener el vehículo a tiempo.
- Problemas en la dirección, suspensión o motor que afectan la maniobrabilidad y estabilidad del vehículo.

- 
- Sobrecarga o mal aseguramiento de la carga, lo que puede alterar la distribución del peso y hacer que el vehículo se des controle.
  - Iluminación defectuosa (luces delanteras o traseras), lo que reduce la visibilidad durante la conducción nocturna o en condiciones de baja visibilidad.
- **Factores ambientales:** Los cuales son las condiciones climáticas adversas, como lluvia, niebla o nieve, así como características de la infraestructura vial, como señales de tránsito deficientes o carreteras en mal estado.

Algunos de los factores relacionados con las condiciones del camino incluyen:

- Deficiencias en el pavimento, como grietas, baches o superficies resbaladizas, que aumentan el riesgo de accidentes.
  - Obstáculos en la vía, como piedras, ramas de árboles, o incluso animales que irrumpen en la carretera.
  - La falta de señales de tránsito o señales mal ubicadas, lo que puede confundir a los conductores y aumentar la probabilidad de siniestros.
  - Falta de mantenimiento en las infraestructuras viales, como puentes en mal estado o iluminación deficiente en zonas de alto riesgo.
- **Factores legislativos:** La falta de leyes de tránsito claras y su cumplimiento, además de las políticas de prevención de accidentes implementadas por las autoridades.

Cada uno de estos factores puede influir significativamente en la probabilidad y gravedad de un siniestro vial. Desde una perspectiva actuarial y de gestión del riesgo, la siniestralidad vial puede entenderse como un proceso aleatorio cuya distribución de probabilidad se ve influida por factores humanos, mecánicos, ambientales y normativos (del Transporte, 2020; Organization, 2018). Este enfoque permite cuantificar la frecuencia y severidad de los siniestros, facilitando el diseño de modelos predictivos más robustos (Bortolussi et al., 2021). Además, al considerar la siniestralidad como un fenómeno multifactorial y dinámico, se abre la posibilidad de integrar herramientas estadísticas y de análisis de datos que permitan anticipar escenarios de alta incidencia y proponer medidas preventivas basadas en evidencia. Así, el estudio detallado de la siniestralidad no solo contribuye al ámbito académico, sino que también resulta fundamental para el desarrollo de políticas públicas orientadas a la seguridad vial y a la optimización de recursos destinados a la reducción de accidentes (SCT, s.f.).

---

## 1.2 Series temporales

Las series temporales representan una secuencia de datos seleccionados y ordenados de manera cronológica, en periodos de tiempo establecidos, generalmente en intervalos regulares, que permiten visualizar la evolución de una variable a lo largo del tiempo (IBM Planning Analytics, 2025) y (Marín, 2023). Esta variable puede derivar de múltiples áreas como economía, medio ambiente, demografía, finanzas, etc. y su análisis permite entender el comportamiento de los datos para realizar pronósticos (Gobierno de Puerto Rico, 2020).

El análisis de series temporales no solo permite entender el comportamiento histórico de una variable, sino que también posibilita modelar su dinámica para generar pronósticos confiables. Estos modelos se apoyan en la estructura temporal de los datos, considerando componentes como tendencia, ciclos, estacionalidad y ruido aleatorio, los cuales, en conjunto, reflejan el patrón subyacente del fenómeno observado (Hyndman y Athanasopoulos, 2018). En el ámbito actuarial y de gestión de riesgos, este enfoque es fundamental, pues la identificación de estas componentes facilita la toma de decisiones estratégicas y la asignación eficiente de recursos para prevenir o mitigar eventos adversos (G. E. Box et al., 2016).

En el contexto de la siniestralidad vial, las series temporales permiten transformar datos históricos de accidentes en herramientas predictivas que apoyan la planeación de medidas preventivas y el diseño de políticas públicas más eficaces. Por ejemplo, el análisis de patrones recurrentes puede indicar incrementos de riesgo durante periodos específicos, como temporadas vacacionales o condiciones meteorológicas adversas (Gupta et al., 2020). De esta manera, los modelos de predicción basados en series de tiempo no solo aportan valor académico, sino que también contribuyen a la reducción de accidentes y a la optimización de estrategias de seguridad vial.

### 1.2.1 Fundamentos de series temporales en la predicción de accidentes

Las secuencias de datos obtenidos a intervalos de tiempo regulares, utilizados en el análisis y la predicción de fenómenos dinámicos, son las series de tiempo (Marín, 2023). En el contexto de siniestralidad vial, estas pueden ser de ayuda para predecir el número de accidentes a partir de datos históricos. El análisis de series de tiempo permite identificar patrones cíclicos o estacionales en los accidentes de tráfico, lo que permite comparar medidas preventivas en periodos de riesgo elevado, como lo son los periodos festivos y las estaciones lluviosas (Pérez y Gómez, 2023).

---

## 1.2.2 Tendencia

Es el comportamiento que toma la serie temporal en un periodo de tiempo, dicho comportamiento puede visualizarse como un aumento, una disminución o mantenerse constante, mientras muestra variaciones de los datos de la variable analizada (Marín, 2023).

## 1.2.3 Ciclos

Son las fluctuaciones recurrentes en intervalos de tiempo, estos pueden visualizarse de manera más extensa que los patrones estacionales. Dichas variaciones cuentan con magnitudes y duraciones irregulares y suelen estar relacionadas con factores externos a la serie (MathWorks, 2025).

## 1.2.4 Estacionalidad

La *estacionalidad* describe la presencia de patrones sistemáticos que se repiten en intervalos de tiempo fijos (p. ej. meses, trimestres o días de la semana). Estos patrones suelen originarse en factores climáticos, calendáricos o de comportamiento humano y provocan fluctuaciones previsibles alrededor de la tendencia de la serie (Hyndman y Athanasopoulos, 2018). Formalmente, si  $s$  es la longitud del período estacional, diremos que la serie  $\{Y_t\}$  presenta estacionalidad de orden  $s$  cuando

$$\mathbb{E}[Y_t] = \mathbb{E}[Y_{t+s}] \quad \text{y} \quad \text{Cov}(Y_t, Y_{t+h}) = \text{Cov}(Y_{t+s}, Y_{t+s+h})$$

para todo  $t$  y todo desfase  $h$  (Chatfield y Xing, 2019).

Para identificar la estacionalidad en una serie temporal, se emplean métodos como:

- **Descomposición clásica o STL:** separa la serie en tendencia, componente estacional y residuales, facilitando el análisis del patrón periódico.
- **Función de autocorrelación (ACF):** la aparición de picos significativos en múltiplos de  $s$  revela la presencia de estacionalidad.
- **Gráficos de subseries estacionales:** dividen la serie por período (mes, trimestre, etc.) y permiten observar la repetición del patrón a lo largo de los ciclos (De y Kumar, 2012).

## 1.2.5 Estacionariedad

La estacionariedad es la propiedad por la cual las características estadísticas de la serie media, varianza y covarianza, permanecen invariantes en el tiempo. Una serie  $\{X_t\}$  es estrictamente

---

estacionaria si la distribución conjunta de  $(X_{t_1}, \dots, X_{t_k})$  es idéntica a la de  $(X_{t_1+h}, \dots, X_{t_k+h})$  para todo  $k, t_1, \dots, t_k$  y desfase  $h$  (G. E. Box et al., 2015). En la práctica se suele trabajar con la estacionariedad débil (o en media y covarianza), suficiente para muchos modelos ARIMA.

Para evaluar la estacionariedad se emplean:

- **Inspección visual:** line-plots y diagramas de caja móviles ayudan a detectar tendencias o cambios en la varianza.
- **Funciones ACF y PACF:** una PACF que se trunca rápidamente puede sugerir estacionariedad, mientras que una ACF que decrece lentamente apunta a no estacionariedad.
- **Pruebas de raíces unitarias:** Dickey–Fuller aumentada (ADF), Phillips–Perron y KPSS contrastan la hipótesis de raíz unitaria o de estacionariedad (Greene, 2012).

Si la serie no es estacionaria, se aplican transformaciones como diferenciación, logaritmos o la eliminación de tendencias deterministas para estabilizarla (Shumway y Stoffer, 2017).

### 1.3 Prueba de Dickey-Fuller y raíces unitarias

La prueba de Dickey-Fuller evalúa la presencia de raíces unitarias en una serie temporal. Su ecuación básica es:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t \quad (1.2)$$

donde:

- $\Delta y_t = y_t - y_{t-1}$  es la primera diferencia de la serie,
- $\gamma$  indica si hay una raíz unitaria,
- $\varepsilon_t$  es el término de error.

Si el valor estimado de  $\gamma$  es significativamente distinto de cero, la serie es estacionaria. La extensión de esta prueba, la prueba ADF, incorpora términos adicionales:

$$\Delta y_t = \alpha + \beta t + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \gamma y_{t-1} + \varepsilon_t \quad (1.3)$$

Esta versión corrige la autocorrelación en los errores, mejorando la precisión del análisis (Shumway y Stoffer, 2017).

La importancia de la prueba de Dickey-Fuller en el análisis de siniestralidad vial radica en que permite determinar si las series temporales de accidentes son estacionarias o requieren transformaciones previas al modelado. Una serie no estacionaria puede inducir pronósticos sesgados

---

y resultados estadísticos inconsistentes, lo que afecta la confiabilidad de los modelos ARIMA y SARIMA. Al identificar raíces unitarias y aplicar diferenciación adecuada, se logra estabilizar la media y varianza de la serie, garantizando que los patrones de tendencia y estacionalidad se modelen correctamente (G. E. Box et al., 2015; Enders, 2014). Esto resulta fundamental para predecir con precisión el número de accidentes y evaluar el impacto de factores de riesgo en distintos periodos de tiempo.

## 1.4 Prueba de Ljung-Box

Esta prueba es utilizada en el análisis de series de tiempo para determinar si existe autocorrelación significativa dentro de los datos o en los residuos después de haber aplicado un modelo como el modelo ARIMA. Evalúa varias autocorrelaciones hasta un número específico de rezagos para analizar el comportamiento aleatorio de la serie.

Durante esta prueba se realiza una comparación entre dos hipótesis, la nula y la alternativa. La hipótesis nula ( $H_0$ ) establece que las autocorrelaciones son independientes, mientras que la hipótesis alternativa ( $H_1$ ) representa que las autocorrelaciones no son independientes.

La prueba se basa en el estadístico  $Q$  de Ljung-Box, el cual se calcula con la siguiente ecuación:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{p}_k^2}{n-k} \quad (1.4)$$

Donde  $n$  es la cantidad de datos,  $h$  es el número de rezagos analizados, y  $\hat{p}_k$  es la autocorrelación estimada en el rezago  $k$ .

Posterior a obtener el valor de  $Q$ , se compara con un valor crítico de la distribución chi cuadrada con  $h$  grados de libertad (Gobierno de Puerto Rico, 2020).

La prueba de Ljung-Box resulta especialmente útil en el contexto de series temporales de siniestralidad vial, ya que permite verificar si los residuos de un modelo ARIMA o SARIMA se comportan como ruido blanco. La ausencia de autocorrelación significativa en los residuos indica que el modelo ha capturado adecuadamente la estructura temporal de los accidentes, incluyendo tendencia y estacionalidad, y que las predicciones generadas serán confiables. En caso contrario, la presencia de autocorrelación residual sugiere la necesidad de ajustar el modelo, incorporando términos adicionales o transformaciones en la serie, con el objetivo de mejorar la precisión del pronóstico y facilitar la planificación de estrategias de prevención de accidentes (G. E. Box et al., 2015).

---

## 1.5 Análisis de siniestralidad en series temporales

El análisis de la siniestralidad vial va más allá de la identificación de causas, apoyando en la modelización estadística para predecir la ocurrencia de accidentes y optimizar las medidas de prevención. A través de los modelos estadísticos de los accidentes es posible identificar patrones y tendencias, lo que facilita la toma de decisiones con relación a la distribución de recursos, ubicación de medidas de seguridad y el diseño de políticas públicas (Gómez y Rodríguez, 2023).

Este análisis de series de tiempo es fundamental para prever comportamientos de fenómenos como los accidentes viales, al permitirnos comprender sus patrones. Los modelos utilizados en este tipo de análisis ayudan a identificar relaciones temporales de los datos históricos, permitiendo predecir valores futuros. Algunos modelos más utilizados en el análisis de series temporales son los modelos autoregresivos (AR), de media móvil (MA), los modelos autorregresivos de media móvil (ARMA) y cuando se presentan tendencias o estacionalidades en los datos es común el uso de modelos autoregresivos integrados de media móvil (ARIMA) (G. E. Box et al., 2015).

De los modelos anteriormente mencionados, se pueden destacar los modelos ARMA y ARIMA como los eficaces para analizar series temporales estacionarias, ya que permiten modelar las relaciones entre los datos actuales y los datos históricos. Lo primordial de estos modelos es la identificación de patrones por medio de la tendencia o la estacionalidad, lo que permite hacer predicciones sobre la ocurrencia de accidentes futuros (Hyndman y Athanasopoulos, 2018).

## 1.6 Modelos de series temporales: AR, MA, ARMA, ARIMA y SARIMA

En las siguientes partes se encontrarán modelos utilizados para el análisis de las series temporales, algunos con información obtenida del artículo titulado “Forecasting: Principles and Practice” (2nd ed.), de los autores Rob J. Hyndman y George Athanasopoulos.

### Modelo Autorregresivo (AR)

El modelo autorregresivo de orden  $p$ , denotado como  $AR(p)$ , es una técnica en la cual el valor actual de la serie se modela como una combinación lineal de sus valores pasados. Su formulación matemática es la siguiente:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (1.5)$$

Donde  $\phi_1, \dots, \phi_p$  son los coeficientes del modelo y  $\varepsilon_t$  representa un ruido blanco. Estos modelos son útiles cuando existe una correlación significativa entre los valores actuales y pasados de la serie, permitiendo modelar la dependencia temporal de forma directa.

### Modelo de Media Móvil (MA)

---

El modelo de media móvil de orden  $q$ , denotado como  $MA(q)$ , representa el valor actual de la serie como una combinación lineal de errores aleatorios pasados. Se define como:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (1.6)$$

Donde  $\theta_1, \dots, \theta_q$  son los coeficientes del modelo y  $\varepsilon_t$  es ruido blanco. Este modelo resulta útil para capturar las perturbaciones aleatorias no explicadas por los valores pasados de la serie.

### **Modelo Autorregresivo de Media Móvil (ARMA)**

El modelo ARMA combina los componentes autorregresivo y de media móvil. Se denota como  $ARMA(p, q)$  y su expresión general es:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (1.7)$$

Este modelo se utiliza para series temporales estacionarias y es adecuado cuando los datos muestran tanto autocorrelación entre los valores observados como correlación en los errores. Su capacidad de incorporar ambos efectos lo convierte en un modelo versátil para fenómenos sin tendencia o estacionalidad.

### **Modelo Autorregresivo Integrado de Media Móvil (ARIMA)**

El modelo ARIMA se emplea cuando los datos no son estacionarios. El componente integrado implica diferenciar la serie original  $d$  veces para lograr estacionariedad. El modelo  $ARIMA(p, d, q)$  se expresa como:

$$\Delta^d Y_t = \phi_1 \Delta^d Y_{t-1} + \cdots + \phi_p \Delta^d Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (1.8)$$

Donde  $\Delta^d$  representa la diferenciación de orden  $d$ . Este modelo es flexible para capturar tendencias determinísticas o estocásticas y permite realizar predicciones precisas al eliminar la no estacionariedad de la serie.

### **Modelo Estacional ARIMA (SARIMA)**

El modelo SARIMA (Seasonal ARIMA) es una extensión del modelo ARIMA que incorpora términos adicionales para modelar la estacionalidad, lo que lo convierte en una herramienta poderosa para series que presentan patrones cíclicos regulares. Se denota como  $SARIMA(p, d, q)(P, D, Q)_s$ , donde:

- $p, d, q$ : parámetros del componente no estacional (autorregresivo, diferenciación y media móvil, respectivamente)

- $P, D, Q$ : parámetros del componente estacional
- $s$ : periodicidad estacional (por ejemplo,  $s = 12$  para datos mensuales)

La forma general del modelo SARIMA se representa como:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^DY_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \quad (1.9)$$

Donde:

- $B$  es el operador de rezago, es decir,  $B^k Y_t = Y_{t-k}$
- $\phi_p(B)$  y  $\theta_q(B)$  son los polinomios no estacionales de orden  $p$  y  $q$
- $\Phi_P(B^s)$  y  $\Theta_Q(B^s)$  son los polinomios estacionales de orden  $P$  y  $Q$
- $(1-B)^d$ : diferenciación ordinaria de orden  $d$
- $(1-B^s)^D$ : diferenciación estacional de orden  $D$
- $\varepsilon_t$ : ruido blanco con media cero y varianza constante

Este modelo es capaz de capturar tanto la dinámica regular como los efectos estacionales, siendo altamente efectivo en aplicaciones como la predicción de accidentes viales, el consumo energético o la demanda de productos con ciclos anuales.

El proceso de construcción de un modelo SARIMA incluye las siguientes etapas:

1. **Identificación:** Se analiza la serie mediante gráficos de línea, ACF y PACF para determinar la existencia de estacionalidad y elegir los órdenes apropiados  $p, d, q$  y  $P, D, Q$ .
2. **Transformación:** En caso de que la varianza no sea constante, puede aplicarse una transformación (como logaritmos), y posteriormente se diferencian los datos de manera ordinaria y estacional para lograr la estacionariedad.
3. **Estimación:** Los parámetros del modelo se estiman generalmente mediante máxima verosimilitud.
4. **Diagnóstico:** Se analiza si los residuos del modelo ajustado son ruido blanco usando gráficos de ACF, prueba de Ljung-Box y prueba de normalidad.
5. **Validación:** Se selecciona el mejor modelo utilizando criterios como el AIC o BIC, y se evalúa su capacidad predictiva con métricas como RMSE o MAPE. Puede usarse validación cruzada por ventanas.

---

6. **Pronóstico:** Finalmente, el modelo se utiliza para predecir valores futuros y construir intervalos de confianza.

**Ventajas del modelo SARIMA:**

- Permite modelar simultáneamente componentes estacionales y no estacionales.
- Es especialmente útil cuando los datos presentan estacionalidad clara (como ocurre con frecuencia en los accidentes viales mensuales).
- Proporciona intervalos de confianza útiles para la planificación.

**Limitaciones:**

- Puede ser complejo seleccionar el conjunto óptimo de parámetros.
- No modela bien cambios estructurales abruptos o no lineales.
- Requiere que la estacionalidad sea constante a lo largo del tiempo.

Por lo tanto, el modelo SARIMA constituye una herramienta robusta y ampliamente aceptada para el análisis y pronóstico de fenómenos temporales con estacionalidad, como es el caso de la siniestralidad en la autopista México–Querétaro.

La elección de un modelo SARIMA adecuado para series de accidentes viales permite capturar tanto las fluctuaciones regulares como los patrones estacionales específicos de cada periodo del año, como las vacaciones o temporadas de lluvia. Esto facilita no solo la predicción precisa de la frecuencia de siniestros, sino también la identificación de periodos de mayor riesgo, lo que resulta crucial para planificar intervenciones preventivas y asignar recursos de manera eficiente. Asimismo, la combinación de componentes autorregresivos y de media móvil, junto con la diferenciación estacional, ofrece flexibilidad para modelar series históricas complejas y mejorar la confiabilidad de los pronósticos, aspecto fundamental en la toma de decisiones en seguridad vial (Hyndman y Athanasopoulos, 2018).

## **1.7 Ventanas Temporales o Ventanas de Tiempo**

Las ventanas de tiempo son una parte importante en el análisis de series temporales, sobre todo cuando se buscan identificar patrones recurrentes en datos que varían con el tiempo. En general, una ventana de tiempo hace referencia a un subconjunto de los datos históricos de una serie de tiempo, este subconjunto se usa para realizar un análisis o predicciones en intervalos específicos (Brockwell y Davis, 2016).

Las ventanas de tiempo pueden tomar distintas formas dependiendo del análisis. Dos aplicaciones comunes de estas son el análisis de tendencias locales y los modelos predictivos, donde se seleccionan segmentos específicos del periodo de tiempo para predecir comportamientos de la serie. Esto permite capturar en distintas escalas las dinámicas temporales, lo que es importante cuando las series presentan variabilidad a corto y largo plazo (Shumway y Stoffer, 2017).

Una parte importante que se debe resaltar en este tema es la elección del tamaño de la ventana de tiempo, ya que impacta en la capacidad del modelo para capturar patrones significativos sin perder generalidad. Si la ventana es demasiado pequeña, el modelo podría ser sensible al ruido y no identificar tendencias de escalas mayores. Por otra parte, si la ventana es demasiado grande podría resultar en un modelo que no logre identificar las fluctuaciones de la serie a corto plazo.

Para definir el tamaño de la ventana, podemos usar el siguiente parámetro  $w$ , que representa el tamaño de la ventana:

$w =$  longitud de la ventana

La longitud de la ventana  $w$  debe elegirse de manera que equilibre el sobreajuste (overfitting) y el subajuste (underfitting). La optimización de la longitud de la ventana se realiza mediante técnicas estadísticas como la validación cruzada o el análisis de la función de autocorrelación. En una serie temporal, la autocorrelación en el intervalo  $k$  se calcula de la siguiente manera:

$$\rho(k) = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad (1.10)$$

Donde  $x_t$  es el valor de la serie temporal en el tiempo  $t$ ,  $\bar{x}$  es el promedio de la serie, y  $N$  es el número total de observaciones.

La **ventana deslizante** es donde se realiza el análisis de la serie de tiempo en segmentos móviles. Este método es usado cuando los datos presentan una estructura cambiante o no estacionaria, ya que permite que el modelo se ajuste continuamente a nuevas observaciones a medida que se desplaza la ventana. La misma, puede expresarse como un conjunto de observaciones  $\{x_t, x_{t+1}, \dots, x_{t+w-1}\}$ , las cuales se van moviendo a lo largo de la serie de tiempo, generando el análisis de un conjunto de subseries.

El uso de ventanas de tiempo deslizantes es relevante en el pronóstico a corto plazo, ya que permite identificar cambios en las tendencias de manera eficiente (Sakamoto y Kawaguchi, 2011).

La fórmula general para la predicción utilizando ventanas deslizantes es la siguiente:

$$\hat{y}_{t+h} = f(x_t, x_{t+1}, \dots, x_{t+w-1}) \quad (1.11)$$

donde  $\hat{y}_{t+h}$  es la predicción para el valor en el tiempo  $t+h$ ,  $f(\cdot)$  es el modelo utilizado (por

---

ejemplo, un modelo ARIMA o redes neuronales), y  $x_t, x_{t+1}, \dots, x_{t+w-1}$  son las observaciones dentro de la ventana de tiempo.

El enfoque de ventanas temporales no solo mejora la capacidad predictiva de los modelos estadísticos, sino que también permite adaptarse a cambios en la dinámica de la serie de accidentes viales. Al segmentar la serie en subconjuntos móviles, se puede detectar la aparición de nuevas tendencias o cambios estacionales, así como ajustar los pronósticos de manera más precisa a corto plazo. Esta técnica resulta especialmente útil cuando la serie presenta variabilidad elevada, eventos atípicos o periodos de comportamiento irregular, asegurando que las predicciones se mantengan confiables y representativas del comportamiento real de los accidentes (Gupta et al., 2020).

## 1.8 Estadísticas descriptivas

### Media

La media, de igual forma conocida como promedio, es una medida de tendencia central de un conjunto de datos al calcular el valor medio de todos los puntos (Montgomery y Runger, 2020).

Se calcula con la siguiente fórmula:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.12)$$

Donde  $n$  es el número total de valores;  $x_i$  es cada valor del conjunto de datos (Ross, 2017).

### Mediana

La mediana es otra medida de tendencia central, pero a diferencia de la media, la mediana es menos sensible a los valores extremos o atípicos (Wilcox, 2017). La mediana se obtiene ordenando los datos de menor a mayor y luego encontrando el valor que se encuentra en el centro de la distribución.

Si el número de elementos es impar, la mediana es el valor que está en el medio; si es par, se calcula como el promedio de los dos valores centrales (Moore et al., 2018).

### Desviación estándar

La desviación estándar es una medida que nos indica que tan dispersos o concentrados están los valores de los datos en relación con la media (Devore, 2015). Entre más grande sea el valor de la desviación estándar, mayor es la variabilidad en los datos; en cambio, una desviación estándar baja indica que los datos están más agrupados alrededor de la media.

Para  $n$  datos  $x_1, x_2, \dots, x_n$ , la desviación estándar se define como  $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$  y para el caso de la población completa en lugar de  $n - 1$  se usa  $n$  (Probabilidad y Estadística, 2025).

---

## Mínimo y Máximo

El mínimo es el valor más bajo dentro del conjunto de datos. Nos dice cuál es el menor número de accidentes registrado entre el periodo establecido (Mann, 2016). El máximo es la cantidad mayor de los datos, la cual, indica el punto más alto registrado en el conjunto de datos (Freund y Perles, 2018).

## Percentiles (cuartiles)

Los percentiles son valores que dividen los datos en intervalos, permitiendo observar la distribución de los datos en distintos puntos. En particular, los cuartiles, como su nombre lo dice, dividen los datos en cuatro intervalos iguales. Los cuartiles nos ayudan a ver no solo la tendencia central, sino también cómo se distribuyen los datos en torno a esta (Moore et al., 2018).

Para calcularlos, se ordenan los datos de menor a mayor, posteriormente se dividen en 4 partes iguales. Para buscar cada cuartil se realiza mediante la expresión  $\frac{kN}{4}$ ,  $k = 1, 2, 3$ . Después, se calculan los cuartiles de la distribución de la tabla (Jairoyd, 2013).

## 1.9 Procesos estocásticos

Se puede entender por proceso estocástico a una colección de variables aleatorias  $Y_t$ , donde  $t \in T$ , definidas de un espacio muestral  $\Omega$  y con valores en el conjunto  $E$ , comúnmente los números reales  $\mathbb{R}$ , aunque también pueden pertenecer a  $\mathbb{C}$ ,  $\mathbb{R}^k$  o  $\mathbb{C}^k$ . El conjunto  $T$  representa el dominio temporal del proceso, que suele ser  $\mathbb{R}$ ,  $\mathbb{Z}$  o  $\mathbb{N}$ . En este sentido, un proceso estocástico  $\{y_t\}_{t=-\infty}^{\infty}$  representa la estructura de probabilidad de una serie de observaciones en un tiempo determinado (Ross, 2014).

### 1.9.1 Ruido Blanco

El ruido blanco es una secuencia  $(w_t, t \in \mathbb{Z})$  que cumple con algunas propiedades como lo son la esperanza matemática, la varianza constante y la ausencia de correlación entre términos. Estas propiedades se definen como:

- Esperanza:  $\mathbb{E}(w_t) = 0$
- Varianza constante:  $\text{Var}(w_t) = \sigma^2$
- No autocorrelación:  $\text{Cov}(w_t, w_{t+k}) = 0$  para todo  $k \neq 0$

Para identificar si una serie pertenece a un ruido blanco, se puede analizar la **función de autocorrelación muestral (ACF)**, denotada como  $\hat{\rho}_k(t)$ :

$$\hat{\rho}_k(t) = \frac{\hat{\gamma}_k(t)}{\hat{\gamma}_0}$$

---

Donde:

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=k+1}^T (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})$$
$$\hat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^T (Y_t - \bar{Y})^2$$

Se asume que  $\hat{\rho}_k \sim \mathcal{N}(0, 1/n)$ . Si alguno de los coeficientes  $\hat{\rho}_k$  excede los intervalos de confianza  $\pm 2/\sqrt{n}$ , se interpreta como la presencia de autocorrelación significativa, lo cual indica que la serie no es ruido blanco (Morales, 2025).

## 1.10 Factores de riesgo en accidentes viales

El pronóstico de accidentes viales es un aspecto importante que se requiere realizar para mejorar la seguridad en las carreteras y poder realizar una adecuada toma de decisiones relacionada a medidas preventivas. Uno de los principales modelos estadísticos, basados en las series de tiempo, y utilizados para realizar este tipo de predicciones es el modelo ARIMA, el cual, se emplea en la predicción de eventos que siguen un patrón temporal, como lo son los accidentes viales.

### 1.10.1 Modelos ARIMA para el Pronóstico de Accidentes Viales

El modelo ARIMA ha sido utilizado en distintos estudios para realizar predicciones de accidentes viales, porque permite modelar el comportamiento temporal de los accidentes a través de tres componentes principalmente: autorregresión (AR), integración (I) y media móvil (MA) (G. E. Box et al., 2015).

Este modelo es utilizado cuando los datos históricos de accidentes viales muestran una estructura temporal y estacionalidad, ya que el modelo ARIMA puede ajustarse a los datos para visualizar patrones a corto y largo plazo. Por ejemplo, Wang (2017) aplicó este modelo la predicción de la cantidad de accidentes en carreteras chinas para distintos periodos de tiempo, encontrando que sus resultados fueron más precisos en comparación con otros modelos estadísticos, como la regresión lineal. De manera similar, González (2019) utilizó un modelo ARIMA para la predicción de accidentes en un determinado tramo de carretera en España, concluyendo que este modelo es eficaz para estimar las tasas de siniestralidad en función del tiempo, especialmente cuando se dispone de datos históricos consistentes.

Además, investigaciones como la de Chen (2015), la cual, estuvo enfocada a predecir accidentes viales han demostrado que la combinación de ARIMA con otras técnicas, en su caso

---

las redes neuronales artificiales, pueden mejorar la precisión en la predicción. Esta integración permite identificar patrones no lineales que un modelo ARIMA podría no identificar por si solo.

Un ejemplo adicional es el estudio de Saldarriaga, quien utilizó un modelo ARIMA para predecir accidentes en carreteras urbanas, demostrando que el modelo puede adecuarse a fluctuaciones semanales y anuales en la frecuencia de accidentes (Saldarriaga, 2020).

En los modelos ARIMA se requiere identificar si la serie temporal cuenta con estacionariedad. Si los datos no son estacionarios, es necesario realizar un proceso de diferenciación para transformarla. La importancia de este proceso radica en las tendencias que generalmente presentan los datos de accidentes viales, lo que hace indispensable convertir la serie en estacionaria para mejorar la precisión del pronóstico (Hyndman y Athanasopoulos, 2018).

### **1.10.2 Ventanas de Tiempo en el Pronóstico de Accidentes Viales**

Como anteriormente se mencionó, algunos autores refieren que una predicción con un modelo como el ARIMA en conjunto con técnicas como las ventanas de tiempo pueden beneficiar a la precisión del pronóstico. Este aspecto implicaría dividir los datos de la serie temporal en segmentos o ventanas de longitud fija, dentro de las cuales se realizan las predicciones. Al aplicar ventanas de tiempo deslizantes, las ventanas se mueven a lo largo de la serie, lo que permite que el modelo ARIMA se ajuste dinámicamente a los cambios que puedan presentar las tendencias o patrones de los accidentes (Sakamoto y Kawaguchi, 2011).

En el contexto de la presente investigación, el uso de ventanas de tiempo será útil, ya que las condiciones del tráfico, las condiciones climáticas y otros factores externos pueden generar variaciones en la frecuencia de accidentes. El ajuste dinámico a través de ventanas deslizantes permitiría al modelo adaptarse a estas variaciones.

## **1.11 Validación cruzada en modelos SARIMA**

Durante el proceso de la aplicación de ventanas temporales se requiere elegir el valor que representará la longitud de cada una de ellas. Para ello se puede utilizar la denominada validación cruzada o mediante el análisis de función de autocorrelación.

Las ventanas deben adaptarse a la estructura temporal y estacional de los datos, los cuales deben estar autocorrelacionados. Por esta razón, primero se abordará el tema de validación cruzada por ventana deslizante (Hyndman y Athanasopoulos, 2018).

En la validación cruzada, es necesario asegurarse de que la serie esté ordenada cronológicamente, sin afectar los datos. Para proceder a elegir una longitud inicial para la ventana de entrenamiento en el caso de tener datos anuales lo recomendable es que el número a elegir sea múltiplo de la cantidad de datos.

---

Una vez que contemos con la longitud inicial se debe aplicar el modelo a los datos de las ventanas. Ya que se haya aplicado el modelo se deben generar los pronósticos para el siguiente periodo y comparar con los datos reales utilizando métricas como MAE, RMSE o MAPE. También se evaluará la estabilidad y capacidad predictiva del modelo, lo cual se realizará promediando las métricas obtenidas a lo largo de todas las iteraciones.

Este tipo de valoración no afecta a la estacionalidad de los datos y es útil para evaluar los modelos como lo es el modelo SARIMA (Bergmeir et al., 2018).

### 1.11.1 Análisis de la función de autocorrelación

El análisis de la función de autocorrelación (ACF) como se ha mencionado en secciones anteriores, permite identificar patrones y tendencias, tanto regulares como estacionales (G. E. Box et al., 2016).

Este análisis permite determinar el valor de las ventanas de tiempo al igual que la asignación de los parámetros del modelo SARIMA. Para culminar este proceso se requiere visualizar el gráfico de la ACF para identificar los picos significativos dentro de la serie y seleccionar así la longitud. Ya contando con dicha cantidad se aplicará el modelo a cada ventana de tiempo con la longitud establecida y para evaluar las predicciones obtenidas por este medio se utilizará el análisis de los residuos (Castaño y Martínez, 2008).

## 1.12 Métodos de identificación, estimación, diagnóstico y pronóstico de accidentes viales

En el análisis de la siniestralidad vial se encuentran de igual forma cuatro pasos fundamentales: identificación, estimación, diagnóstico y pronóstico. A continuación, se desglosarán estos pasos para comprender por qué son indispensables.

- **Métodos de identificación:** Se refieren a las técnicas que permiten identificar patrones en los datos de accidentes. Entre los métodos más comunes se encuentran el análisis de series temporales, que ayuda a identificar tendencias o estacionalidades, y la identificación de factores de riesgo (Porcu et al., 2016).
- **Estimación:** Una vez identificados los patrones y las causas de los accidentes, es necesario estimar los parámetros del modelo que se ajusten de una mejor manera a los datos históricos. Métodos como la regresión lineal, las redes neuronales y los métodos bayesianos son útiles para la estimación de parámetros y la predicción del número de accidentes (Montgomery et al., 2015).

- 
- **Diagnóstico:** El diagnóstico permite evaluar el modelo ajustado a los datos y su capacidad predictiva. Entre las técnicas de diagnóstico se encuentran el análisis de residuos y las pruebas de autocorrelación. Estas herramientas permiten decidir si el modelo es adecuado para la predicción determinada (Hyndman y Athanasopoulos, 2018).
  - **Pronóstico:** Como su nombre lo indica el pronóstico es la parte en la que se realizan las predicciones utilizando el modelo seleccionado; con el objetivo de predecir el número de accidentes para una toma de decisiones informada en relación con la seguridad vial.

## 2 Aplicación Empírica del Modelo SARIMA en el Pronóstico de Accidentes Viales

En este capítulo describimos las actividades realizadas para obtener el pronóstico del número de accidentes en la autopista México-Querétaro, para el cual, se lleva a cabo un análisis estadístico con el fin de identificar patrones en los datos históricos de accidentes viales. Este análisis es fundamental para generar un modelo ARIMA, adecuado para la predicción de series temporales. Posteriormente, se aplica la técnica de ventanas de tiempo, que mejora la precisión del pronóstico al segmentar los datos en intervalos temporales específicos, permitiendo una mejor adaptación a posibles fluctuaciones. Finalmente, con el modelo ajustado y la técnica de ventanas de tiempo aplicada, se obtienen los resultados del pronóstico.

### 2.1 Obtención de datos

Los datos empleados en este trabajo de investigación provienen de las estadísticas de accidentes de tránsito y los anuarios publicados en la plataforma de la Secretaría de Comunicaciones y Transportes (SCT). Estas bases de datos contienen información detallada sobre los accidentes ocurridos en todas las carreteras de México, abarcando tanto las autopistas de cuota como las carreteras libres. Para este estudio, se extraen los datos correspondientes a la autopista México-Querétaro, específicamente los registros de los accidentes ocurridos entre 2012 y 2022 (SCT, s.f.). En una primera etapa se depuraron y homogeneizaron los registros comprendidos entre enero de 2012 y diciembre de 2021. Con el fin de disponer de un conjunto de validación verdaderamente independiente, las observaciones de 2022 se reservaron como *hold-out* y no intervinieron en la estimación de parámetros ni en la selección del modelo. De esta forma, los datos de 2012 al 2021 constituyen la muestra de entrenamiento utilizados para diagnosticar el modelo SARIMA a este periodo se le designará como periodo de entrenamiento, por otro lado el periodo de enero a diciembre del 2022 se emplearán para realizar un pronóstico y validar el modelo mediante una comparación de los datos reales con los datos obtenidos con el modelo propuesto, a este periodo se le conocerá como periodo de validación, esta práctica es recomendada para evaluar la capacidad predictiva y prevenir el sobreajuste (G. E. Box et al., 2015; Hyndman y Athanasopoulos, 2018).

---

Los datos rescatados de la base de datos de la SCT, contiene información cuantitativa mensual sobre los accidentes ocurridos en la autopista México-Querétaro, plasmados en un formato tabular. Cada fila representa el número de accidentes en un año determinado y especifica el tramo de la autopista, ya que esta atraviesa distintos estados. Las columnas proporcionan detalles sobre las diversas categorías que clasifican los accidentes. Además de los accidentes en sí, la base de datos incluye información sobre las circunstancias que contribuyeron a cada incidente, tales como: factores relacionados con el conductor (por ejemplo, imprudencia, exceso de velocidad, invasión de carril, rebaso indebido, entre otros), condiciones del vehículo (como problemas con neumáticos, frenos, suspensión, etc.), condiciones del camino (irrupción de semovientes, desperfectos, falta de señalización) y agentes naturales (lluvia, nieve, granizo, entre otros) (SCT, s.f.).

Además, la base de datos clasifica los accidentes según las condiciones geográficas y de tránsito, como el alineamiento vertical (pendientes, cimas, curvas), el alineamiento horizontal (curvas abiertas, cerradas, intersecciones, túneles, etc.) y las medidas de control de tránsito (señales informativas, preventivas, semáforos, agentes de tránsito). También se encuentra una clasificación de los accidentes según los días festivos, lo que permite observar si existe una mayor ocurrencia de accidentes durante fechas específicas del año, como: Año Nuevo, Día de la Constitución, Semana Santa, Día del Trabajo, Día de la Independencia, y otros períodos festivos relevantes (SCT, s.f.). Esta clasificación es de gran utilidad para analizar las posibles fluctuaciones estacionales o asociadas a eventos particulares (Hyndman y Athanasopoulos, 2018).

## **2.2 Preparación de los Datos**

La base de datos obtenida por la SCT está subdividida por tramos de la autopista México-Querétaro. Cada tramo contiene información sobre los accidentes ocurridos en un mes en específico, y al analizar los datos se visualizó que algunos tramos no reportaron accidentes durante ciertos meses, lo que resultaba en celdas vacías en la base de datos. Al unir los datos de todos los tramos para crear una base unificada de accidentes totales mensuales, algunas de estas celdas vacías se corrigieron al sumarse los accidentes correspondientes a cada tramo. Sin embargo, en otros casos, las celdas vacías persistieron, indicando que no hubo accidentes en ciertos meses.

Para asegurar que la base de datos fuera funcional para el análisis y el pronóstico, las celdas vacías que quedaron fueron llenadas con ceros. Esto se hizo para indicar que no hubo accidentes en esos meses, y así garantizar la coherencia de la base de datos para el modelo de pronóstico. Este proceso permitió que la base de datos reflejara de manera consistente el número total de accidentes mensuales en la autopista México-Querétaro, sin subdivisiones por tramos o categorías.

---

Cabe resaltar que el proceso que se mencionó anteriormente fue realizado con ayuda del lenguaje de programación R, identificando y tratando posibles valores atípicos y datos faltantes, además de realizar las transformaciones necesarias para asegurar que los datos fueran estacionales y estacionarios, lo cual es necesario identificar durante la selección del modelo ARIMA. Durante este proceso se utilizaron diversas funciones de R, tales como `ts()` para convertir los datos en series temporales (Team, 2023), `na.omit()` para eliminar los datos faltantes (Team, 2023), y `diff()` para calcular las diferencias necesarias que estabilizan la varianza de las series temporales (Team, 2023).

Después del preprocesamiento y limpieza de los datos, resultó la tabla 3.13 de anexos, la cual, representa únicamente los datos requeridos para el adecuado desarrollo del modelo.

## **2.3 Análisis Exploratorio de Datos**

Se analizaron los datos históricos del número de accidentes ocurridos mensualmente durante el periodo de entrenamiento, dando un total de 120 observaciones. El objetivo de este análisis exploratorio es entender la frecuencia con la que ocurren los accidentes, también identificar tendencias estacionales o patrones que pueden presentarse en los datos observados. Esto es importante analizarlo antes de modelar debido a que cada modelo de series temporales cuenta con propiedades o condiciones que se deben cumplir para poder realizar una mejor elección de este. Es por ello por lo que en este trabajo de investigación se realizará en primera instancia el análisis de los datos.

### **2.3.1 Visualización de tendencias y patrones estacionales**

Para identificar tendencias o posibles patrones graficamos los datos. A lo largo del período de las 120 observaciones, visualizamos fluctuaciones en el número de accidentes mes a mes, sin un crecimiento o decrecimiento claro, lo que indica la ausencia de una tendencia consistente. Aunque se pueden ver subidas y bajadas evidentes, no se identifica una clara dirección, lo que sugiere que la cantidad de accidentes puede estar influenciada por factores externos.

En cuanto a los patrones estacionales, se visualizaron fluctuaciones (mensuales) regulares que podrían estar vinculadas a factores externos como lo son los días festivos, condiciones climáticas y las características del camino. Por ejemplo, se observan picos recurrentes en diciembre y enero, lo que puede deberse a condiciones climáticas o el aumento del tráfico debido a las festividades decembrinas. También, se registraron aumentos en los accidentes en agosto, debido probablemente a un mayor flujo de tráfico por las vacaciones de verano.

Estos picos repetitivos en ciertos meses del año indican que los accidentes no siguen un patrón completamente aleatorio, sino que están relacionados con factores periódicos, como las

---

condiciones climáticas o los flujos de tráfico.

La estacionalidad es un patrón evidente en los datos, con incrementos significativos en meses específicos, como diciembre, enero y agosto. Pero no se quedará únicamente con la observación de gráficos, durante el desarrollo del presente trabajo se irán analizando y evaluando los datos de una mejor manera, utilizando pruebas que permitan analizar los datos.

### **2.3.2 Estadísticas descriptivas**

En este trabajo de investigación realizamos el análisis estadístico en base a medidas descriptivas para analizar de manera general las variables numéricas extraídas de los datos proporcionados por la SCT. Estas bases fueron previamente procesadas en la etapa anterior del desarrollo, donde se depuraron los registros incompletos. Una vez definidos los datos requeridos, se almacenaron en un archivo en formato `.xlsx` que posteriormente fue importado al entorno estadístico R donde se realizaron los cálculos correspondientes a las estadísticas, las cuales fueron: la media, la mediana, la desviación estándar, el valor mínimo y máximo de los datos, se realizaron 4 cuartiles (25%, 50%, 75% y 100%), esto conforme a la información que se presentó en el marco teórico. Esto para analizarlas y poder visualizar el comportamiento de la serie sin tomar en cuenta otros factores que pueden afectarle. La información requerida para la obtención de cada estadística se puede visualizar en el capítulo anterior, por ejemplo el cálculo de la media se realiza conforme la ecuación (1.12).

## **2.4 ARIMA (Autorregresivo Integrado de Media Móvil)**

Para la identificación del modelo ARIMA, se utilizaron los datos históricos (mensuales) de accidentes durante el periodo de entrenamiento. Realizamos primeramente un análisis de la serie temporal, observando que presentaba fluctuaciones específicas a lo largo del tiempo, lo cual indicaba la presencia de estacionalidad. Este comportamiento es común en los datos de accidentes, ya que ciertos factores externos, como las condiciones climáticas o los periodos de vacaciones, pueden generar patrones recurrentes a lo largo del tiempo. Por lo tanto, en lugar de utilizar un modelo ARIMA estándar, usamos un modelo ARIMA estacional (SARIMA), que permite capturar tanto la no estacionariedad como la estacionalidad en los datos.

### **2.4.1 Diferenciación y Estacionariedad**

En el análisis de series de tiempo, uno de los principales aspectos para la construcción de modelos predictivos es la estacionariedad de los datos. La estacionariedad es una propiedad de la serie temporal que implica que sus características estadísticas, como la media, la varianza y la autocovarianza, permanecen constantes a lo largo del tiempo (G. E. Box et al., 2015). Enseguida

---

se mencionará como realizamos el procedimiento de la diferenciación y la estacionariedad de la serie.

Para comprobar la estacionariedad de la serie, aplicamos la prueba de Dickey-Fuller aumentada (ADF), la cual es una prueba estadística utilizada para verificar la presencia de una raíz unitaria. La prueba indicó que la serie no era estacionaria debido a la presencia de una tendencia lineal. Donde con ayuda de R calculamos el valor  $p$ , obteniendo en la prueba un valor alto (mayor a 0.05), esto indica que no se puede rechazar la hipótesis nula de no estacionariedad, por lo cual, la serie no es estacionaria y es necesario aplicar un proceso de diferenciación para lograr la estacionariedad. Para las series con estacionariedad, además de la diferenciación de primer orden, se aplica una diferenciación estacional, que es crucial para capturar los efectos estacionales (Hyndman y Athanasopoulos, 2018).

Para transformar la serie en estacionaria, se aplicó una diferenciación de primer orden, es decir,  $d = 1$ . Este proceso consiste en calcular las diferencias entre valores consecutivos de la serie, restando  $Y_t - Y_{t-1}$ , lo que elimina la tendencia lineal de los datos. Sin embargo, debido a la estacionalidad presente en la serie, también fue necesario aplicar una diferenciación estacional de orden  $D = 1$  con un rezago de 12 meses ( $s = 12$ ).

En este trabajo, dado que los datos presentaban una estacionariedad anual, además de realizar la diferenciación no estacional, aplicamos una diferenciación estacional. La diferenciación estacional tiene como objetivo capturar los patrones que tienden a mostrar un comportamiento repetitivo cada año. En este contexto, utilizamos un período estacional de 12 (considerando que se dispone de datos mensuales).

La correcta transformación de los datos es fundamental, ya que modelos como el ARIMA solo son aplicables cuando la serie es estacionaria, garantizando que el modelo capture las fluctuaciones sin estar influenciado por tendencias o ciclos no deseados (G. E. Box et al., 2015).

## 2.4.2 Análisis ACF y PACF

La selección de los parámetros adecuados para el modelo ARIMA estacional es esencial para lograr una modelización precisa de series de tiempo, especialmente cuando se busca capturar tanto las dependencias autoregresivas como los efectos estacionales (Hyndman y Athanasopoulos, 2018).

Una vez que ya contamos con una serie estacionaria procedemos a la identificación de los parámetros comenzando con el análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), herramientas estadísticas que permiten determinar los parámetros del modelo, tales como los autoregresivos ( $p$ ), la diferenciación ( $d$ ), los de media móvil ( $q$ ), y los componentes estacionales ( $P$ ,  $D$ ,  $Q$ , y  $s$ ), donde  $s$  denota el periodo estacional (G. E. Box et al., 2015).

---

El primer parámetro para determinar fue el valor de  $d$ , que representa el número de diferenciaciones realizadas para hacer que la serie temporal sea estacionaria. En este caso, aplicamos la diferenciación de primer orden ( $d$ ). A continuación, usamos una diferenciación estacional de primer orden ( $D$ ) con un rezago de 12 meses, con el fin de capturar los efectos estacionales presentes en los datos (Hyndman y Athanasopoulos, 2018).

Una vez identificados los parámetros de las diferenciaciones, procedimos al análisis de los gráficos de ACF y PACF, que son cruciales para la identificación de los parámetros autoregresivos y de media móvil. Usamos las funciones `acf()` y `pacf()` de R para calcular estas autocorrelaciones.

Estas funciones son efectivas y simples para calcular la ACF y PACF, especialmente en series que requieren diferenciación para lograr la estacionariedad (Hyndman y Athanasopoulos, 2018). La función `acf()` calcula la autocorrelación de la serie a distintos rezagos, mientras que `pacf()` ajusta la autocorrelación eliminando la influencia de rezagos intermedios y mostrando únicamente la autocorrelación directa (Hyndman y Athanasopoulos, 2018).

Al analizar el gráfico de la PACF, observamos una caída repentina en la autocorrelación después del primer rezago, lo que indicó que solo el primer rezago era significativo. Por lo que se estaría hablando de un modelo autoregresivo de primer orden, lo que llevó a la selección del valor  $p$ . Es decir, el valor actual de la serie depende principalmente de su valor inmediato anterior, sin influencias significativas de observaciones más distantes (G. E. Box et al., 2015).

En el caso del gráfico de la ACF, se detectó una caída gradual en la autocorrelación a medida que aumentaban los rezagos, lo cual sugiere la presencia de una estructura de media móvil en la serie temporal. En este tipo de modelos, los errores pasados influyen en los valores futuros, pero de manera decreciente a medida que aumenta el rezago, lo que sugiere un modelo de media móvil de primer orden y la selección del valor  $q$ .

Es así como se obtuvieron los parámetros de modelo SARIMA, el cual, mejora el pronóstico y permite una representación precisa de las fluctuaciones anuales, que en este caso están relacionadas con los accidentes viales.

### **2.4.3 Evaluación del modelo ARIMA estacional (SARIMA)**

Tras la identificación de los parámetros del modelo a través de las gráficas ACF y PACF, y la diferenciación de la serie temporal para lograr la estacionariedad, ajustamos el modelo SARIMA utilizando la función `arima()` de R, la cual permite especificar tanto los términos autoregresivos y de media móvil como los componentes estacionales. El modelo ajustado fue de la forma  $SARIMA(p,d,q)(P,D,Q)[s]$ .

La función `arima()` es ampliamente utilizada en el análisis de series temporales debido a su flexibilidad y capacidad para ajustar modelos autoregresivos integrados de media móvil

---

(ARIMA), que son efectivos para modelar datos con dependencia temporal y estacionalidad (Hyndman y Athanasopoulos, 2018). La elección de esta función para el ajuste del modelo se debe a su versatilidad y a su capacidad para incluir tanto componentes estacionales como no estacionales en una misma estructura.

Una vez ajustado el modelo SARIMA, realizamos las predicciones para el año 2022, específicamente para los 12 meses del año. Usamos `forecast()` de R para generar dichas predicciones, junto con sus intervalos de confianza al 90 %, 95 % y 99 %.

Los resultados de estas predicciones fueron presentados gráficamente, en donde las gráficas tanto los datos de accidentes viales del periodo de entrenamiento, como las predicciones y los intervalos de confianza al 90 %, 95 % y 99 %, respectivamente. Esto nos permitió una visualización clara de las fluctuaciones esperadas y de la variabilidad asociada a las predicciones, destacando las temporadas de mayor riesgo de accidentes y el intervalo en el cual pueden encontrarse la mayor cantidad de accidentes.

Posterior a la obtención del pronóstico procedimos a la evaluación del modelo, la cual realizamos mediante un análisis de los residuos del modelo verificando la presencia de cualquier patrón residual inexplicable, ya que deben comportarse como un proceso aleatorio, con el uso de la prueba de Ljung-Box y la prueba de normalidad de Shapiro-Wilk. Para que una vez evaluado el modelo se continúe con la comparación del pronóstico y los datos reales registrados en el año 2022 por la SCT.

## **2.5 Optimización del modelo mediante ventanas de tiempo**

Una vez ajustado el modelo SARIMA a la serie de tiempo de los accidentes viales en la autopista México-Querétaro, decidimos mejorar el proceso de pronóstico utilizando la técnica de ventanas de tiempo. Esta técnica permite optimizar el modelo de manera más precisa a los cambios que puedan ocurrir en la serie temporal a lo largo del tiempo, dividiendo la serie en subsegmentos (ventanas) y ajustando el modelo en cada uno de ellos. De esta forma, se pueden capturar patrones más recientes y, al mismo tiempo, manejar de manera eficiente la estacionariedad y las variaciones no lineales.

### **2.5.1 Selección del tamaño de la ventana**

La selección del tamaño de la ventana es importante en el proceso de ajuste mediante ventanas de tiempo. Ya que si no se selecciona adecuadamente no logrará captar adecuadamente la estacionalidad o los patrones de la serie temporal (Hyndman y Athanasopoulos, 2018).

Para iniciar con el ajuste mediante la técnica de ventanas de tiempo, es necesario asignar el valor de  $W$  (la longitud de cada ventana), la cual depende tanto de la longitud de la serie temporal

---

como del comportamiento de la misma, por lo que, obtuvimos con el análisis de la función de autocorrelación de la serie (Sakamoto y Kawaguchi, 2011), considerando los picos que la misma presentaba, elegimos realizar el ajuste con ventanas temporales deslizantes con longitud de 48 observaciones, lo cual permitirá que las observaciones dentro de la ventana de tiempo sean:  $\{x_t, x_{t+1}, \dots, x_{t+w-1}\}$ , así se obtiene una adecuada selección de la longitud. Inmediatamente después, contando con el valor de la longitud de cada ventana de tiempo, dividimos la serie temporal en 85 ventanas en este caso.

Durante esta etapa del análisis detectamos 3 posibles longitudes que se adaptarán de una mejor manera a los datos, estas fueron identificadas por distintas circunstancias, entre ellas aplicamos en, primera instancia, la visualización del comportamiento de la serie, identificando este decidimos utilizar 48 observaciones, sin embargo, no cumplía con las 50 observaciones que se recomienda tener en una serie temporal para poder aplicar el modelo, no obstante, seleccionamos el tamaño de ventanas de 48 observaciones, de igual forma el de 50 observaciones ya que era cercano al primero y cumplía con el requisito establecido, por otra parte, establecimos el tamaño de 60 observaciones ya que de igual forma estaba determinado por el comportamiento de la serie temporal. Fue así como seleccionamos los tamaños correspondientes a las longitudes de las ventanas de tiempo.

### **2.5.2 Aplicación del modelo a las ventanas de tiempo**

Después de especificar las ventanas de tiempo, se aplica el modelo SARIMA a cada una de ellas como lo podemos ver en la ecuación 1.11, lo que permite que el modelo se ajuste a los patrones analizados de los subsegmentos de la serie, de igual forma realizamos mediante el software R, el código utilizado en la base de datos principal, fue funcional para los datos establecidos por cada ventana de tiempo, y solo se realizó un ajuste al modelo para adecuarlo de mejor manera a los datos.

Realizamos el procedimiento antes descrito con la serie completa, pero ahora este procedimiento se hace para cada ventana de tiempo. Analizando la estacionariedad mediante la prueba de Dickey-fuller aumentada, ya que esta prueba determinó que los datos son estacionarios, además utilizamos, en los casos necesarios, la diferenciación, posteriormente se identificaron los parámetros para el modelo, lo ajustamos utilizando los datos de los subsegmentos y realizamos los pronósticos para las observaciones posteriores de cada ventana temporal analizada. Cabe destacar que para seleccionar el tamaño de las ventanas utilizamos la función de autocorrelación, analizamos dicha función para ver en dónde dejaba de tener picos significativos.

Realizamos el mismo procedimiento, aplicando el modelo SARIMA a las ventanas temporales, para 3 distintas longitudes de ventanas, para posteriormente comparar los resultados de cada una y así elegir la longitud que más se adecue a los datos reales.

---

### **2.5.3 Evaluación de las predicciones de las ventanas temporales**

Ya que se cuenta con las predicciones evaluamos su rendimiento y de ser necesario se ajusta el tamaño de la ventana, aquí analizamos los pronósticos obtenidos para la ventana de longitud 48, posteriormente los de las ventanas con longitud de 50 y por último las de longitud de 60. Además, nos permite visualizar el rendimiento del modelo ajustado con ventanas de tiempo, las cuales obtuvimos mediante un análisis de los residuos de cada ventana.

Además, comparamos los pronósticos obtenidos en cada ventana con los valores reales registrados por la SCT, estos valores son los que obtuvimos al depurar las bases de datos. Este análisis nos permite verificar la capacidad predictiva del modelo, observando los errores de pronóstico para determinar si necesitan ajustes de tamaño de la ventana o en los parámetros del modelo.

Por último, para cada ventana, realizamos un análisis de los residuos, es decir, la diferencia entre los valores reales y los pronosticados. Este análisis nos permite detectar patrones no capturados por el modelo y verificar si los errores presentan comportamientos sistemáticos, lo cual, indicaría una mala especificación del modelo o una inadecuada elección de los parámetros. Una vez obtenido dicho análisis procedemos a ajustar nuevamente el modelo en caso de ser necesario, en caso contrario, se procede a realizar el reporte del procedimiento.

### 3 Resultados

En este capítulo presentamos los resultados obtenidos con la aplicación del modelo SARI-MA, para el pronóstico del número de accidentes en la autopista México – Querétaro, esto con los datos obtenidos de la SCT, siguiendo el proceso detallado en el capítulo anterior y conforme a las bases teóricas presentadas en el presente trabajo de investigación.

Durante el mismo, exponemos tanto las decisiones metodológicas tomadas como su impacto en los resultados obtenidos durante el proceso abordamos los hallazgos encontrados con el modelo, describiendo cada uno de los parámetros.

Por otra parte, proporcionamos los resultados en general, justificando cada acción, e interpretando cada uno de ellos. Para posteriormente, proporcionar una conclusión referente a los resultados obtenidos en este trabajo.

#### 3.1 Análisis descriptivos de datos

**Media** Calculamos la media usando la ecuación (1.12). Donde  $n$  es la cantidad total de datos,  $x_i$  representa todos y cada uno de los valores de los accidentes registrados durante el período de análisis. Donde  $n = 120$ ,  $x_i$  representa todos y cada uno de los valores de los accidentes registrados durante el período de entrenamiento.

De lo cual obtuvimos el resultado de 29.75, lo que significa que, en promedio hubo aproximadamente 30 accidentes por mes. Este valor representa el centro de los datos y nos da una idea general del comportamiento de los accidentes a lo largo del tiempo. Es importante señalar que, aunque la media es un valor útil, no siempre se adecua a las variaciones extremas que pueden estar presentes en los datos, como los meses con muy pocos o muchos accidentes (Freund y Perles, 2018).

#### **Mediana**

Realizamos el procedimiento descrito en el segundo capítulo de este trabajo, de donde obtuvimos que la mediana de los datos es 28. Esto significa que el 50% de los meses tuvieron 28 accidentes o menos, y el otro 50% tuvo más de 28 accidentes.

#### **Desviación estándar**

La desviación estándar es 9.17, lo que significa que, en promedio, los datos de accidentes

---

se desvían de la media (aproximadamente 30 accidentes) por alrededor de 9.17 accidentes. Esto nos indica que los valores de accidentes no son constantes cada mes, sino que pueden variar considerablemente, lo que refleja la volatilidad en los patrones de accidentes.

Aunque la media nos da un valor de referencia, la desviación estándar nos ayuda a entender qué tanto puede variar el número de accidentes (mensual) (Robert V. Hogg y Craig, 2019).

### **Mínimo y Máximo**

En este caso, el mínimo es 11. Esto indica que el mes con menos accidentes en el conjunto de datos tuvo 11 accidentes. Este valor es usado para identificar cual mes fue el que presento menor cantidad de accidentes (Mann, 2016). El valor mínimo ayuda a identificar anomalías o períodos de baja incidencia en comparación con los meses más accidentados.

El máximo para estos datos fue 55, lo que significa que el mes con más accidentes registrado durante el período analizado tuvo 55 accidentes. Este valor es importante porque señala el mes con mayor incidencia. Al igual que el valor mínimo, el máximo puede ser útil para identificar extremos en los datos y entender cuáles fueron los picos más altos, que pueden estar relacionados con factores estacionales (Freund y Perles, 2018).

### **Cuartiles**

Se obtuvieron los siguientes cuartiles.

#### **Primer cuartil ( $Q_1$ , 25%)**

Este cuartil indica que el 25% de los meses tuvieron 22 accidentes o menos. En otras palabras, el 25% de los datos más bajos de accidentes están por debajo de 22.

#### **Segundo cuartil ( $Q_2$ , 50%)**

El segundo cuartil es la mediana, que como se mencionó fue 28. Esto significa que el 50% de los meses tuvieron 28 accidentes o menos.

#### **Tercer cuartil ( $Q_3$ , 75%)**

Este cuartil muestra que el 75% de los meses tuvieron 36 accidentes o menos. Esto implica que solo el 25% restante de los meses tuvo más de 36 accidentes. Este valor también indica una forma de ver en qué punto se concentra la mayoría de los datos.

### **Interpretación de los Cuartiles**

Los cuartiles permiten una interpretación más detallada de la distribución de los datos. Si consideramos los valores de los cuartiles:

$$Q_1(25\%) = 22$$

Esto nos dice que una cuarta parte de los meses tuvieron un número de accidentes relativamente bajo, con solo 22 o menos accidentes.

---

$$Q_2(\text{Mediana}) = 28$$

La mediana de 28 indica que la mitad de los meses tuvieron menos de 28 accidentes.

$$Q_3(75\%) = 36$$

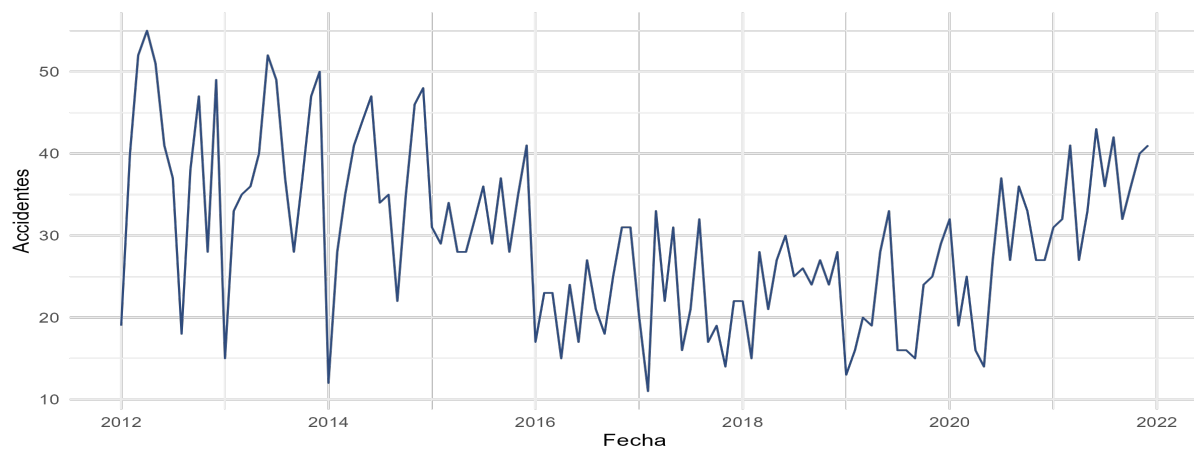
En el tercer cuartil se obtuvo 36, lo que significa que el 75% de los meses presentaron 36 o menos accidentes, y solo un 25% de los meses superaron este número.

### 3.2 Visualización de la serie temporal

En primera instancia, realizamos el análisis de la serie temporal del número de accidentes mensuales durante el periodo de entrenamiento. La serie mostró una clara estacionalidad con fluctuaciones en ciertos meses del año, lo que indicó que era necesario ajustar un modelo que permitiera capturar dicha estacionalidad.

En la figura 3.1 se puede ver el gráfico de la serie temporal, donde observamos que no se cuenta con una tendencia clara a lo largo del tiempo, debido a que los valores no muestran un aumento o disminución continua, sin embargo, se puede notar que durante las fluctuaciones hay tendencia estacional, ya que en algunos meses parecen presentarse valores más altos de siniestros como en diciembre y enero.

Figura 3.1: Tendencia del número de accidentes (periodo de entrenamiento)



Fuente: Elaboración propia utilizando el software R.

---

### 3.3 Verificación de la Estacionariedad: Prueba ADF

Para seleccionar el modelo SARIMA, iniciamos verificando si la serie de tiempo es estacionaria o no, para ello se aplicó la prueba de Dickey-Fuller aumentada (ADF). Esta prueba nos permite evaluar la presencia de una raíz unitaria, lo cual indicaría que la serie temporal no es estacionaria. En este caso, la hipótesis nula de la prueba ADF establece que la serie tiene una raíz unitaria, es decir, no es estacionaria; mientras que la hipótesis alternativa afirma que la serie es estacionaria. Formalmente, las hipótesis de la prueba se pueden visualizar de la siguiente manera:

$H_0$  : La serie tiene una raíz unitaria (no estacionaria)

$H_1$  : La serie es estacionaria (sin raíz unitaria)

La serie temporal que se está analizando corresponde al número de accidentes mensuales que abarca durante el periodo de entrenamiento. Con el propósito de asegurar que los datos no presenten tendencias que no cumplan con los supuestos necesarios por los modelos de series de tiempo, como ARIMA o SARIMA, los cuales requieren estacionariedad para realizar los pronósticos (G. E. Box et al., 2015; Hamilton, 1994).

Durante el presente análisis obtuvimos como resultado de la prueba ADF, lo siguiente:

Figura 3.2: Prueba ADF antes de diferenciar la serie

```
Augmented Dickey-Fuller Test

data: serie_temporal
Dickey-Fuller = -2.1996, Lag order = 4, p-value = 0.4937
alternative hypothesis: stationary
```

Fuente: Elaboración propia con datos históricos de la SCT.

El valor de p fue de 0.4937, entonces  $0.4937 > 0.05$ ; 0.05 es el umbral establecido por la prueba ADF, por lo que al ser mayor el valor de p, se deduce que no se rechaza la hipótesis nula  $H_0$ . Esto sugiere que la serie de tiempo presenta una raíz unitaria, lo que implica que la serie no es estacionaria. Lo cual, indica que se cuenta con una tendencia estocástica que debe ser tratada antes de continuar con la aplicación del modelo, para que se logre la estacionariedad (Dickey y Fuller, 1979).

Para lograr que la serie de tiempo cuente con estacionariedad, realizamos una diferenciación de la serie. La diferenciación de primer orden que se aplicó fue  $Y'_t = Y_t - Y_{t-1}$ , donde  $Y_t$  es el

---

valor de la serie en el tiempo  $t$ ,  $Y_{t-1}$  es el valor de la serie en el tiempo  $t - 1$  y  $Y'_t$  es el valor diferenciado en el tiempo  $t$ , evitando así las tendencias y efectos estacionales (Enders, 2014). Esta diferenciación de primer orden se puede observar en la tabla 3.16 que se encuentra en anexos.

A continuación, se presentan los resultados de la prueba ADF aplicada posterior a la diferenciación de primer orden:

Figura 3.3: Prueba ADF después de aplicar diferenciación

```
Augmented Dickey-Fuller Test  
  
data: serie_temporal  
Dickey-Fuller = -5.0205, Lag order = 4, p-value = 0.01  
alternative hypothesis: stationary
```

**Fuente:** Elaboración propia con datos históricos de la SCT.

Esta prueba refiere que la serie ya es estacionaria, dado que el valor p es 0.01 el cual es menos al umbral (0.05), lo que implica que se rechaza la hipótesis nula  $H_0$ , lo que implica que no se cuenta con una raíz unitaria y los datos son estacionarios.

### 3.4 Análisis de la Función de Autocorrelación y la Función de Autocorrelación Parcial

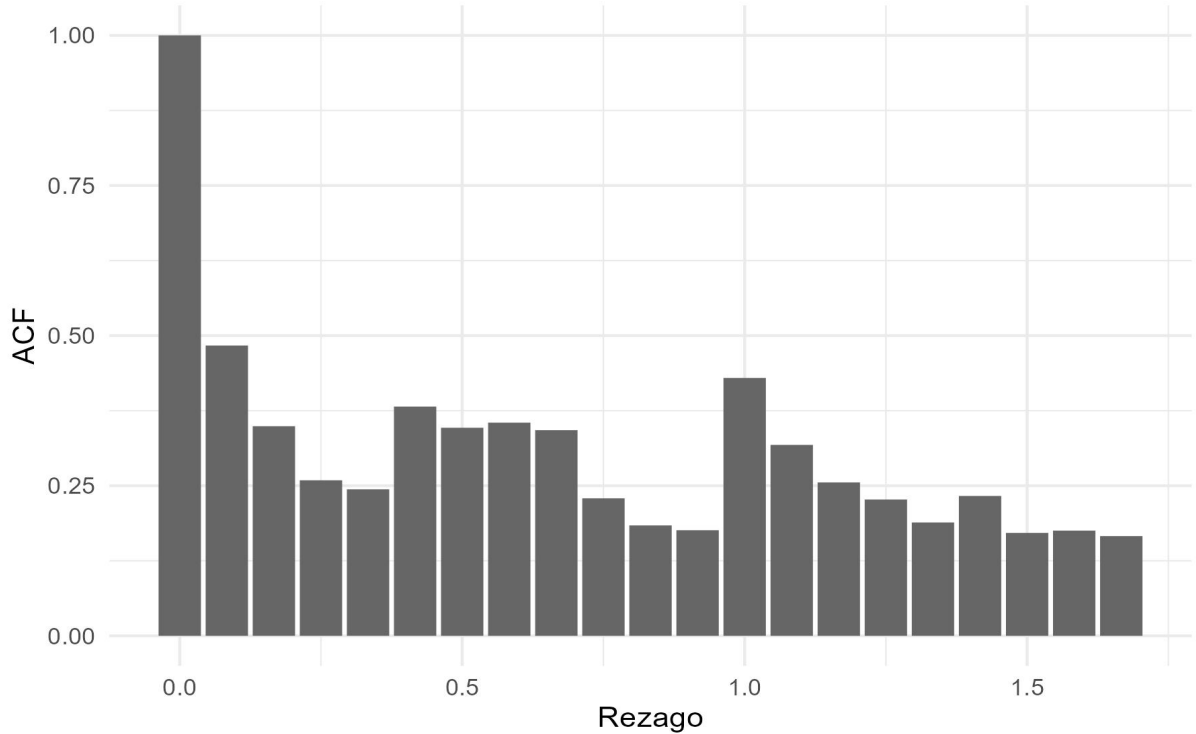
La función de autocorrelación (ACF) nos permite observar cómo se correlacionan los valores de la serie con sus propios rezagos a lo largo del tiempo, además, nos permite identificar patrones estacionarios y sobre todo hace posible la selección de los parámetros apropiados para el modelo SARIMA.

Durante este análisis, generamos gráficas de la ACF tanto de la serie temporal principal, como la serie diferenciada. Como anteriormente, mencionamos la serie principal cuenta con posibles tendencias no estacionarias, mientras que la diferenciada representa la serie estacionaria.

En primer lugar, presentamos la función de autocorrelación de la serie principal, para brindar la interpretación de esta. La gráfica de la ACF para la serie de tiempo principal es la gráfica 3.4.

La gráfica de la figura 3.4 permite observar la autocorrelación antes de la diferenciación. La cual representa no estacionariedad, ya que las correlaciones pueden persistir a lo largo de varios rezagos, lo cual no es útil para la construcción del modelo SARIMA. Por lo cual, para el análisis ACF quedaría descartada esta serie desde el momento de la diferenciación de primer orden.

Figura 3.4: Gráfico de la función de autocorrelación (ACF) de la serie original



Fuente: Elaboración propia utilizando el software R.

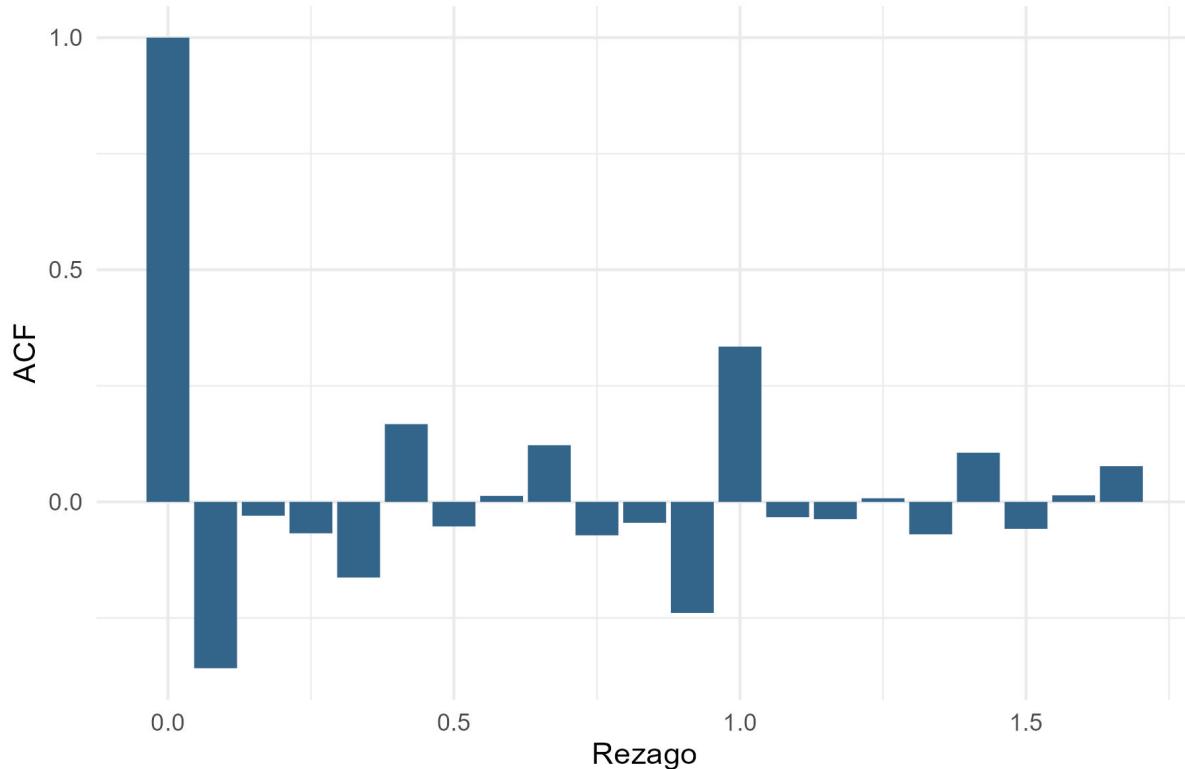
Una vez analizada la ACF de la serie temporal, da lugar al análisis de la serie diferenciada, la cual, como se puede observar en la figura 3.5 proporciona una visión más confiable de la estructura de autocorrelación y es la que se utilizará para determinar los parámetros del modelo SARIMA. En este análisis, la diferenciación permite estabilizar la varianza y hacer que la serie sea estacionaria, lo que es un requerimiento para el ajuste adecuado del modelo.

Entonces, después de este análisis es importante para el modelo SARIMA, tomar en cuenta únicamente la ACF de la serie diferenciada, ya que refleja mejor las dependencias temporales relevantes sin la influencia de tendencias no estacionarias.

Una vez seleccionada la ACF de la serie diferenciada, es necesario elegir el parámetro  $q$  de esta. El parámetro  $q$  es la representación del orden del componente de media móvil (MA) en un modelo SARIMA, y se determina observando la figura 3.5, la cual muestra un patrón que permite elegir el valor de  $q$ , que es el número de rezagos a considerar. En particular, la figura 3.5 muestra un corte limpio después del primer rezago, es decir, que los valores caen bruscamente después del primer rezago, lo que sugiere que el modelo SARIMA debe incluir un rezago en el componente MA, es decir, el parámetro  $q = 1$ .

La gráfica de ACF para la serie diferenciada se presenta en la figura 3.5.

Figura 3.5: Gráfica de la Función de Autocorrelación (ACF) de la Serie Diferenciada



Fuente: Elaboración propia utilizando el software R.

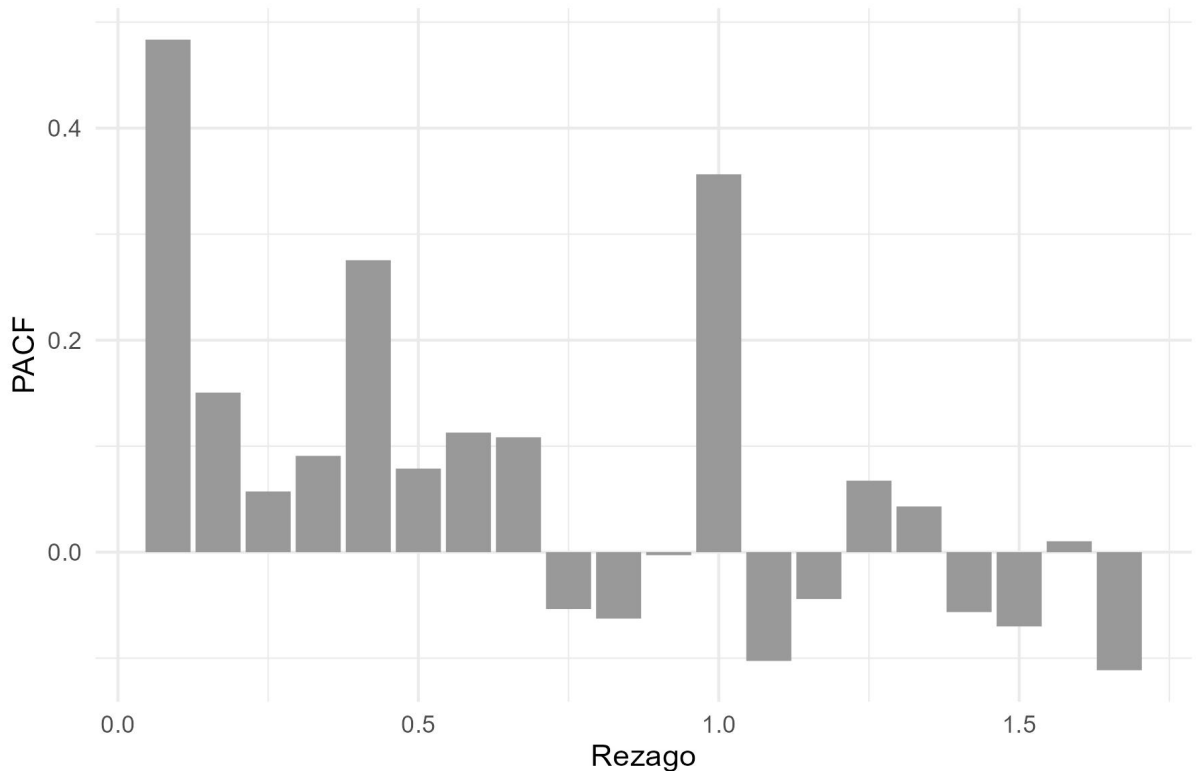
Una vez realizado el análisis de la ACF, hacemos el análisis de la PACF. Esta función mide la correlación entre una observación y sus rezagos, eliminando la influencia de los valores intermedios. Es especialmente útil para identificar el parámetro  $q$  en el componente de media móvil del modelo SARIMA, ya que permite observar cuántos rezagos mantienen una relación directa con la variable actual sin verse afectados por correlaciones indirectas. En otras palabras, la PACF ayuda a determinar cuántos términos de media móvil son necesarios para capturar la estructura de dependencia temporal en los errores del modelo (G. E. Box et al., 2015).

En la figura 3.6 presentamos la gráfica PACF de la serie temporal en su forma original. Esta gráfica muestra que la serie presenta autocorrelaciones parciales significativas en múltiples rezagos, lo cual es un fuerte indicio de que los datos no son estacionarios. Al igual que en el análisis de la ACF, donde se descartó el uso directo de la serie original por su comportamiento no estacionario, la PACF reafirma esta conclusión. Dado que no observamos una disminución clara de las correlaciones parciales hacia cero en los primeros rezagos, concluimos que es necesario aplicar una diferenciación para eliminar tendencias o efectos de largo plazo antes de ajustar cualquier modelo SARIMA (Hyndman y Athanasopoulos, 2018).

---

En otras palabras, en la figura 3.6 no se muestra una caída abrupta a cero después de ningún rezago, lo cual se espera en una serie estacionaria. El hecho de que se mantengan valores significativos en varios rezagos consecutivos confirma la presencia de memoria persistente en los datos. Esta persistencia sugiere que la serie, en su forma original, no es adecuada para ser modelada directamente, ya que no cumple con uno de los supuestos clave del modelo SARIMA: la estacionariedad (Shumway y Stoffer, 2017).

Figura 3.6: Función de Autocorrelación Parcial de la serie original



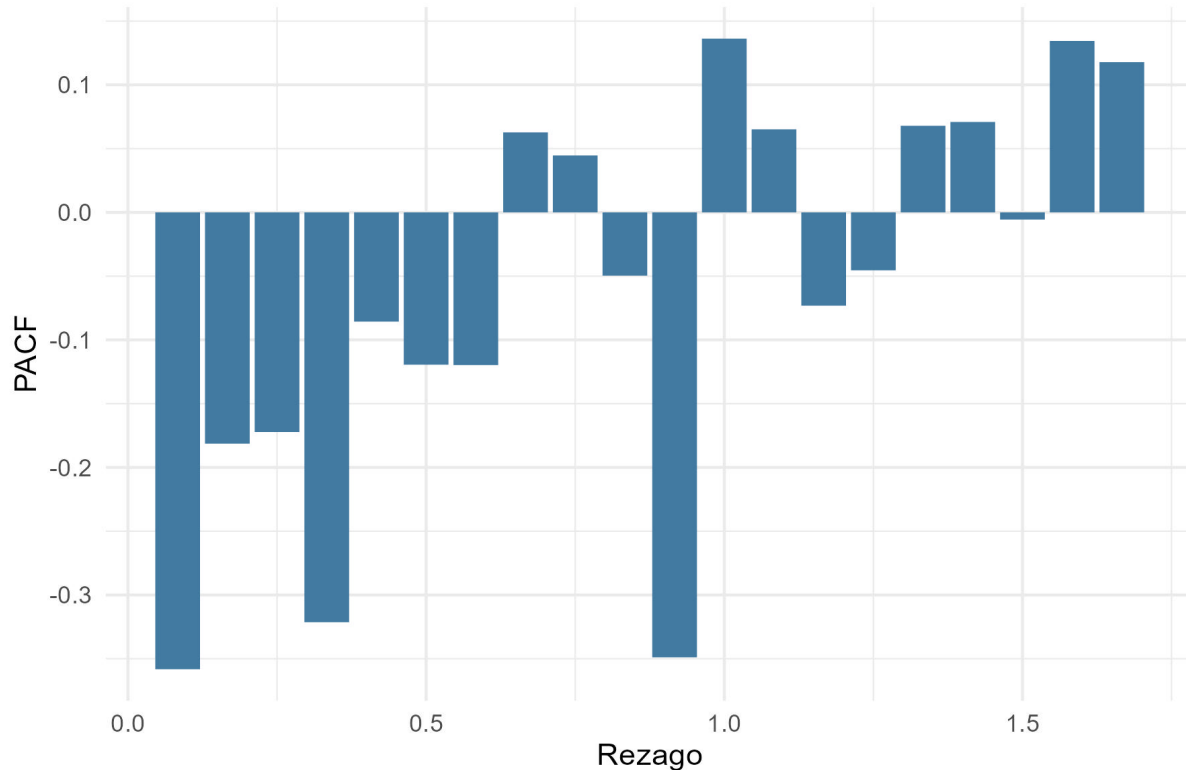
Fuente: Elaboración propia utilizando el software R.

Luego de descartar la PACF de la serie temporal principal, realizamos el análisis de la PACF de la serie diferenciada y como se muestra en la figura 3.7 los coeficientes de autocorrelación parcial tienen un comportamiento más claro y definido, con una caída abrupta después de un primer rezago significativo. Este patrón indica que la serie diferenciada es estacionaria, lo que hace que sea apropiada para el modelo SARIMA

La elección del valor de  $p$  la realizamos seleccionando el mayor rezago significativo antes de que la PACF se estabilice cerca de cero. En este caso, observamos que el valor de  $p$  corresponde

a  $p = 1$ , ya que la PACF muestra una caída abrupta después del primer retardo. Esto implica que el componente autorregresivo (AR) del modelo SARIMA debe tener un orden  $AR(1)$  (G. E. Box et al., 2015).

Figura 3.7: Función de Autocorrelación Parcial de la serie diferenciada



Fuente: Elaboración propia utilizando el software R.

### 3.5 Determinación de los parámetros del modelo SARIMA

El modelo SARIMA requiere la determinación de un conjunto de parámetros para ajustarse adecuadamente a la serie temporal. Estos parámetros incluyen los componentes autorregresivos  $p, d, q$  para la parte no estacional del modelo, y los parámetros estacionales  $P, D, Q$  para capturar la estacionalidad. A continuación, describimos el proceso de determinación de cada uno de estos parámetros.

El parámetro  $d$  indica el número de diferenciaciones necesarias para que la serie temporal sea estacionaria. El valor de  $d$  se determinó a través de la prueba de raíz unitaria ADF (Augmented Dickey-Fuller), la cual mostró que la serie temporal original no era estacionaria y se volvió estacionaria después de aplicar una diferenciación. Por lo tanto, el valor de  $d$  resultante es:

---

$d = 1$ .

Determinamos el parámetro  $p$  mediante la observación de la función de autocorrelación parcial (PACF) de la serie temporal diferenciada, como lo mencionamos anteriormente. El gráfico de PACF muestra una caída abrupta, lo que sugirió que el modelo debía incluir un componente autorregresivo de primer orden. Por lo tanto, el valor de  $p$  es:  $p = 1$ .

Determinamos el parámetro  $q$  observando la función de autocorrelación de la serie temporal diferenciada. El gráfico de ACF muestra que el primer rezago fue significativo teniendo después de este un corte limpio, lo que sugiere que el modelo debe incluir un componente de medias móviles de primer orden. Así, el valor de  $q$  es:  $q = 1$ .

El parámetro  $P$  corresponde al orden del componente autorregresivo estacional (SAR) del modelo. Para determinarlo, analizamos la función de autocorrelación parcial (PACF) aplicada a la serie temporal ya diferenciada, la cual presenta características de estacionariedad. En la gráfica de la figura 3.7 se observaron los rezagos múltiples de 12 (correspondientes a la periodicidad mensual de la serie), identificándose un corte abrupto en el rezago 12. Esto sugiere la presencia de un componente autorregresivo estacional de primer orden. Por lo tanto, establecemos que  $P = 1$ .

El parámetro  $D$  indica el número de diferenciaciones estacionales necesarias para hacer que la serie temporal sea estacionaria en términos estacionales. En este caso, la serie temporal ya mostró una estacionalidad clara con un período de 12 meses, y la diferenciación estacional de primer orden ( $D = 1$ ). Por lo tanto, el valor de  $D$  es:  $D = 1$ .

El parámetro  $Q$  representa el orden del componente de medias móviles estacional (SMA) del modelo. Este valor lo identificamos observando la función de autocorrelación (ACF) de la serie temporal diferenciada. En el gráfico de la figura 3.5 se detectó un pico significativo en el rezago 12, lo que indica que debe incorporarse un componente de medias móviles estacional de primer orden. Así, se concluye que  $Q = 1$ .

### **Síntesis y notación del modelo**

De acuerdo con los valores estimados ( $p = 1, d = 1, q = 1$ ) y sus equivalentes estacionales ( $P = 1, D = 1, Q = 1$ ), y considerando que la estacionalidad observada es anual con datos mensuales ( $s = 12$ ), el modelo definitivo se expresa en la notación estándar *Box-Jenkins* como

$$\text{SARIMA}(1, 1, 1)(1, 1, 1)_{12}$$

El subíndice 12 (a veces escrito entre corchetes) indica la longitud del periodo estacional y debe acompañar siempre al segundo triplete ( $P, D, Q$ ) para distinguirlo del componente no estacional. En adelante se empleará de forma consistente esta notación completa.

### 3.6 Ajuste del modelo SARIMA

En la sección anterior, obtuvimos los parámetros  $p = 1$ ,  $d = 1$ ,  $q = 1$ ,  $P = 1$ ,  $D = 1$ , y  $Q = 1$ , los cuales ajustaron el modelo SARIMA, cabe mencionar, que fue con un periodo estacional de 12. Al aplicar el modelo se obtuvieron los siguientes datos:

Figura 3.8: Modelo SARIMA ajustado con orden (1,1,1)(1,1,1)[12]

```
Call:
arima(x = serie_temporal, order = c(1, 1, 1),
      seasonal = list(order = c(1, 1, 1), period = 12))

Coefficients:
      ar1      ma1      sar1      sma1
 0.2291 -0.8643 -0.0280 -0.5418
s.e.  0.1124  0.0543  0.1946  0.1732

sigma^2 estimated as 64.58:  log likelihood = -377.68,  aic =
 765.36
```

Fuente: Elaboración propia con datos históricos de la SCT.

A continuación, interpretamos los coeficientes estimados del modelo SARIMA ajustado con los parámetros anteriormente mencionados, y un periodo estacional de 12:

- *ar1*: El coeficiente autoregresivo de primer orden ( $AR(1)$ ) es 0.2291. Este valor sugiere una relación positiva moderada entre el valor actual de la serie temporal y el valor rezagado en un período. Este es estadísticamente significativo, ya que su error estándar (0.1124) es menor al valor estimado del coeficiente.
- *ma1*: El coeficiente de media móvil de primer orden ( $MA(1)$ ) es -0.8643. Este valor negativo sugiere que los errores en la predicción de un período influyen negativamente en los valores actuales de la serie temporal. El error estándar es 0.0543, lo que indica que el coeficiente es significativo.
- *sar1*: El coeficiente autoregresivo estacional de primer orden ( $SAR(1)$ ) es -0.0280. Este valor refleja la relación negativa entre los valores de la serie en el mismo mes de años

anteriores y los valores actuales. Este coeficiente es importante para capturar la estacionalidad anual de los datos, que en este caso refleja los patrones estacionales en los accidentes viales a lo largo del tiempo. El error estándar es 0.1946, lo que sugiere que la estimación tiene una mayor incertidumbre comparada con los otros coeficientes.

- *sma1*: El coeficiente de media móvil estacional de primer orden ( $SMA(1)$ ) es -0.5418. Este valor negativo indica que las fluctuaciones estacionales pasadas tienen un impacto negativo en los valores actuales de la serie temporal. El error estándar es 0.1732, lo que sugiere que el coeficiente tiene una precisión moderada.

El valor de  $\sigma^2$ , que representa la varianza del error del modelo, fue 64.58. Esto indica que el modelo presenta una varianza residual moderada, lo cual es esperado dado que los datos de accidentes pueden tener fluctuaciones naturales. La varianza residual es un indicador importante de la calidad del ajuste del modelo, donde valores más bajos implican un mejor ajuste.

El log-likelihood del modelo es -377.6792, y el AIC (Criterio de Información de Akaike) es 765.3585. El AIC es un indicador en el que valores más bajos sugieren un mejor ajuste del modelo. En este caso, el valor obtenido indica que el modelo ajustado es adecuado. Esta afirmación se respalda al compararlo con otros modelos SARIMA con diferentes parámetros, comparación que puede visualizarse en la tabla 3.1.

Cuadro 3.1: Comparación de modelos SARIMA

MODELO	AIC	LogLik
SARIMA(1,1,1)(1,1,1)[12]	765.3585	-377.6792
SARIMA(1,1,2)(0,1,1)[12]	765.3751	-377.6875
SARIMA(0,1,1)(1,1,1)[12]	767.3654	-379.6827
SARIMA(2,1,1)(1,1,0)[12]	771.4406	-380.7203
SARIMA(2,1,2)(1,1,0)[12]	773.1793	-380.5896

*Fuente: Elaboración propia utilizando el software R.*

### 3.7 Pronóstico de Accidentes Viales para el Año 2022

En este momento ya contamos con la selección de los parámetros del modelo, por lo que se realiza el pronóstico de los accidentes viales en la autopista México-Querétaro durante el año 2022, con el uso de la función `forecast()` en R, la cual, permite realizar predicciones a futuro basadas en el modelo ajustado. En este caso, se predijeron los valores para los siguientes 12 meses, correspondientes al año 2022. El pronóstico realizado genera tanto las predicciones

puntuales como los intervalos de confianza al 90 %, 95 %, y 99 % para cada uno de los 12 meses del año 2022. La aplicación de este pronóstico da los valores de las tablas 3.2, 3.3, 3.4.

Cuadro 3.2: Predicciones con intervalo de confianza al 90 % para 2022

Fecha	Predicción	Límite Inf 90	Límite Sup 90	Pred. (red.)	Lím. Inf (red.)	Lím. Sup (red.)
01/01/2022	37.97982666	25.09754175	50.86211158	38	25	51
01/02/2022	35.57504180	21.94298310	49.20710051	36	22	49
01/03/2022	40.21961992	26.34553590	54.09370394	40	26	54
01/04/2022	33.71380137	19.67239405	47.75520869	34	20	48
01/05/2022	36.33520212	22.14235375	50.52805049	36	22	51
01/06/2022	41.79859218	27.45889052	56.13829383	42	27	56
01/07/2022	39.79299043	25.30858009	54.27740077	40	25	54
01/08/2022	40.44445066	25.81690776	55.07199356	40	26	55
01/09/2022	38.09368416	23.32442770	52.86294063	38	23	53
01/10/2022	39.71907803	24.80946196	54.62869410	40	25	55
01/11/2022	40.15985190	25.11118686	55.20851694	40	25	55
01/12/2022	40.98541457	25.79897399	56.17185515	41	26	56

Fuente: Elaboración propia utilizando el software R.

Cuadro 3.3: Predicciones con intervalo de confianza al 95 % para 2022

Fecha	Predicción	Límite Inf 95	Límite Sup 95	Pred. (red.)	Lím. Inf (red.)	Lím. Sup (red.)
01/01/2022	37.97982666	22.62963741	53.33001592	38	23	53
01/02/2022	35.57504180	19.33144198	51.81864163	36	19	52
01/03/2022	40.21961992	23.68762914	56.75161070	40	24	57
01/04/2022	33.71380137	16.98243258	50.44517016	34	17	50
01/05/2022	36.33520212	19.42338019	53.24702405	36	19	53
01/06/2022	41.79859218	24.71178377	58.88540059	42	25	59
01/07/2022	39.79299043	22.53375098	57.05222988	40	23	57
01/08/2022	40.44445066	23.01465825	57.87424307	40	23	58
01/09/2022	38.09368416	20.49502963	55.69233870	38	20	56
01/10/2022	39.71907803	21.95317470	57.48498135	40	22	57
01/11/2022	40.15985190	22.22826151	58.09144229	40	22	58
01/12/2022	40.98541457	22.88965449	59.08117464	41	23	59

Fuente: Elaboración propia utilizando el software R.

El pronóstico como se mencionó anteriormente está acompañado por los intervalos de confianza al 90 %, 95 %, y 99 % para cada mes, lo cual es importante ya que estos intervalos reflejan el rango dentro del cual esperamos que caigan los valores reales de los accidentes viales para cada mes del año 2022, ofreciendo una visión más detallada de la incertidumbre del pronóstico.

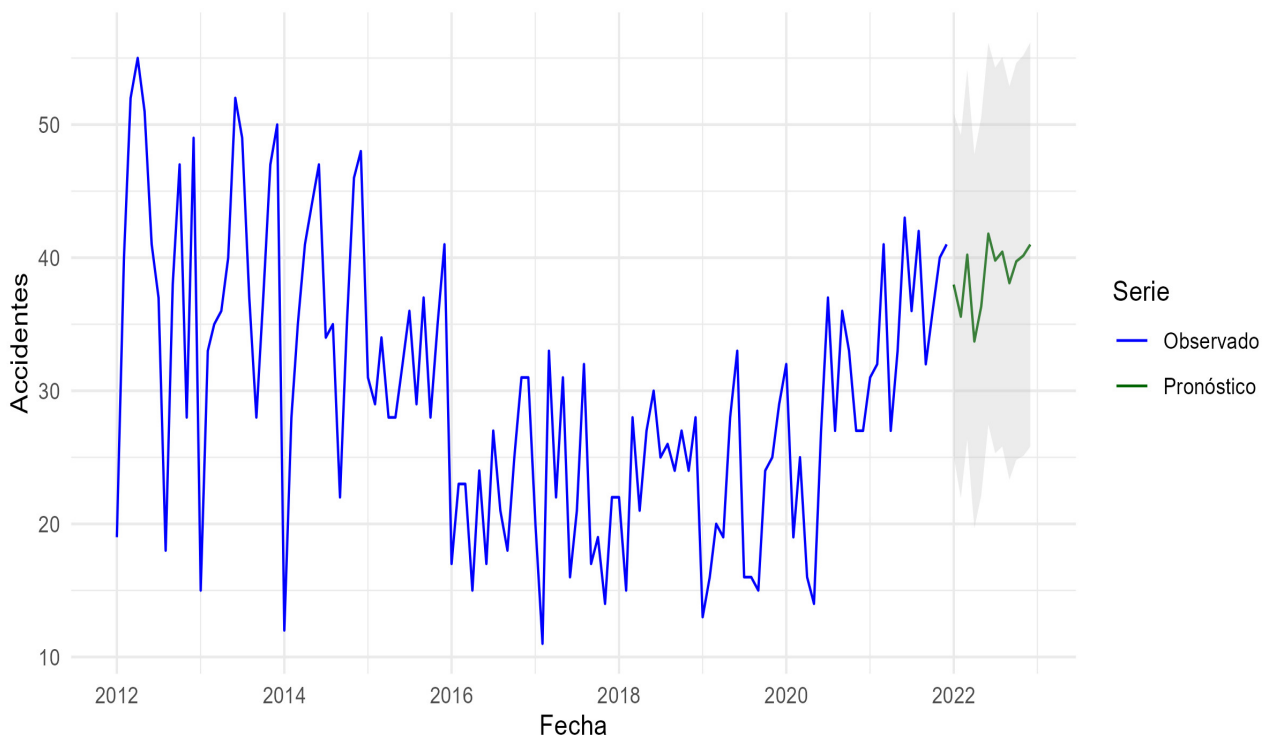
Las tablas 3.2, 3.3 y 3.4 muestran el pronóstico para el año 2022, junto con sus intervalos de confianza al 90 %, 95 % y 99 %, respectivamente, pero para visualizar de una mejor manera estos datos, a continuación, se presentan los gráficos de las figuras 3.9, 3.10 y 3.11, los cuales corresponden a las tablas anteriormente señaladas.

Cuadro 3.4: Predicciones con intervalo de confianza al 99% para 2022

Fecha	Predicción	Límite Inf 99	Límite Sup 99	Pred. (red.)	Lím. Inf (red.)	Lím. Sup (red.)
01/01/2022	37.97982666	17.80625837	58.15339496	38	18	58
01/02/2022	35.57504180	14.22733299	56.92275062	36	14	57
01/03/2022	40.21961992	18.49290115	61.94633870	40	18	62
01/04/2022	33.71380137	11.72505547	55.70254727	34	12	56
01/05/2022	36.33520212	14.10930060	58.56110364	36	14	59
01/06/2022	41.79859218	19.34271944	64.25446491	42	19	64
01/07/2022	39.79299043	17.11050489	62.47547596	40	17	62
01/08/2022	40.44445066	17.53782053	63.35108078	40	18	63
01/09/2022	38.09368416	14.96513159	61.22223674	38	15	61
01/10/2022	39.71907803	16.37072329	63.06743277	40	16	63
01/11/2022	40.15985190	16.59374745	63.72595635	40	17	64
01/12/2022	40.98541457	17.20355458	64.76727456	41	17	65

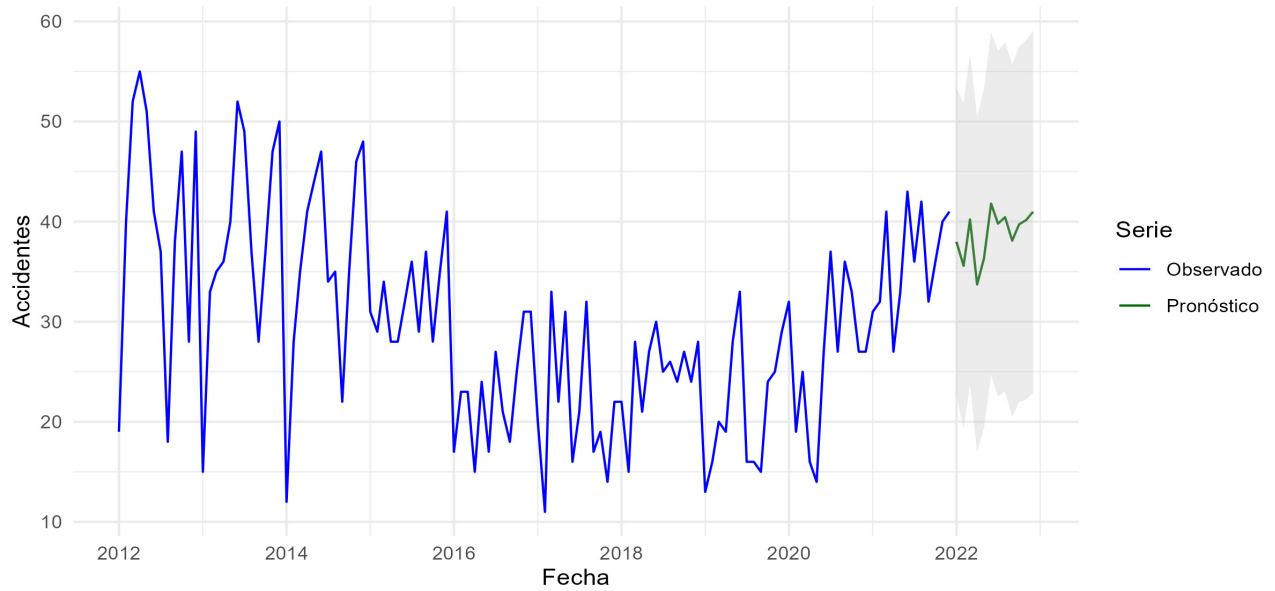
Fuente: Elaboración propia utilizando el software R.

Figura 3.9: Pronóstico de accidentes viales en la autopista México-Querétaro para el año 2022, con intervalo de confianza al 90%



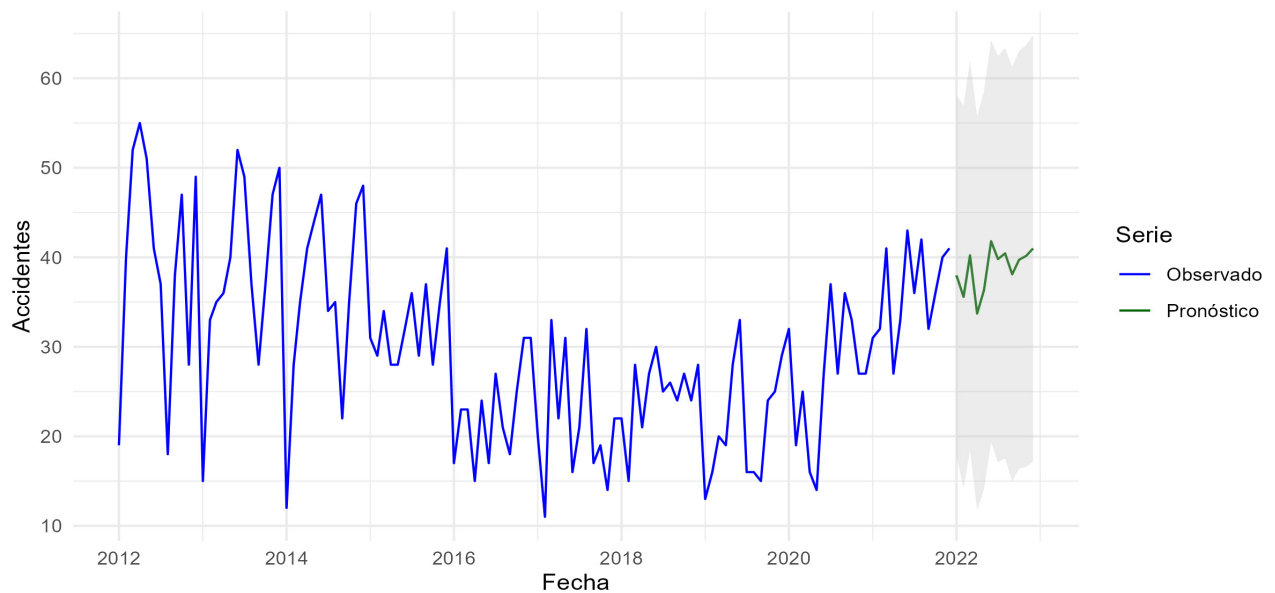
Fuente: Elaboración propia utilizando el software R.

Figura 3.10: Pronóstico de accidentes viales en la autopista México-Querétaro para el año 2022, con intervalo de confianza al 95 %



Fuente: Elaboración propia utilizando el software R.

Figura 3.11: Pronóstico de accidentes viales en la autopista México-Querétaro para el año 2022, con intervalo de confianza al 99 %



Fuente: Elaboración propia utilizando el software R.

---

Observamos en las gráficas de las figuras 3.9, 3.10 y 3.11 que los intervalos de confianza se amplían conforme aumenta el nivel de confianza. Es decir, el intervalo al 99 % es más amplio que el de 95 %, y este a su vez más amplio que el de 90 %. Esto se debe a que, al aumentar el nivel de confianza, se requiere cubrir un rango mayor de valores posibles para garantizar con mayor certeza que el valor real se encuentre dentro del intervalo. Por lo tanto, a mayor nivel de confianza, mayor amplitud del intervalo.

Por otro lado, los datos del número de accidentes para el año 2022 presentan una ligera tendencia al alza, con fluctuaciones poco pronunciadas, lo que sugiere cierta estabilidad en el comportamiento de la serie.

Este pronóstico puede ser útil para la planificación de políticas de seguridad vial y la implementación de medidas preventivas. Sin embargo, es importante actualizar periódicamente los datos y ajustar el modelo para mantener la precisión de las predicciones.

El intervalo de confianza al 90 % proporciona una estimación del rango en el que pueden encontrarse los valores reales con un 90 % de certeza. Lo mismo ocurre con los intervalos de confianza al 95 % y al 99 %, aunque estos últimos abarcan rangos más amplios debido a la mayor certeza estadística que buscamos alcanzar.

En la sección 3.8 realizamos la comparación de los datos reales con los pronósticos obtenidos, tomando en cuenta los distintos intervalos de confianza.

### **3.8 Comparativa de Predicciones y Datos Reales de Accidentes para 2022**

En esta sección, comparamos los datos obtenidos a partir de los pronósticos del modelo con los datos reales registrados durante el año 2022. Para ello, no solo consideramos los valores puntuales estimados, sino también los intervalos de confianza al 90 %, 95 % y 99 % asociados a cada pronóstico.

A continuación, se presentan los datos reales registrados por la SCT de manera mensual durante el año 2022. Posteriormente, analizamos la tabla con los pronósticos generados, y finalmente, se desarrolla una comparación entre ambas fuentes.

#### **3.8.1 Datos Reales de Accidentes para 2022**

En primer lugar, en la tabla 3.5 presentamos los datos reales registrados mensualmente durante el año 2022, los cuales representan el número de accidentes viales ocurridos en dicho periodo. Estos datos sirven como referencia para evaluar qué tan bien se ajustan las predicciones generadas por el modelo.

Cuadro 3.5: Datos reales de accidentes viales para 2022

Fecha	Dato Real
ene-22	28
feb-22	46
mar-22	42
abr-22	31
may-22	28
jun-22	21
jul-22	24
ago-22	23
sep-22	20
oct-22	20
nov-22	24
dic-22	21

Fuente: Datos recopilados de la Secretaría de Comunicaciones y Transportes (SCT).

### 3.8.2 Análisis de la Comparativa

Con base en la tabla 3.5, procedemos a comparar los valores reales con los pronósticos del modelo. Observamos que los valores pronosticados de las tablas 3.2, 3.3 y 3.4 no reproducen exactamente los datos observados en la tabla 3.5, aunque la mayoría de las predicciones se acercan razonablemente a los registros mensuales. La discrepancia absoluta varía entre 0 accidentes (marzo) y 21 accidentes (junio), con un error absoluto medio (*MAE*) de aproximadamente 15.25 accidentes y una mediana de 17.

En cuanto a los intervalos de confianza, se observa lo siguiente:

Con un nivel de confianza del 90%, 10 de los 12 meses presentan valores reales que se encuentran dentro de los límites del intervalo correspondiente. Solo los meses de junio y diciembre quedan fuera de este rango.

Con el intervalo del 95%, 11 de los 12 meses contienen los datos reales, siendo diciembre el único mes que queda fuera del intervalo.

Finalmente, con el intervalo de confianza al 99%, todos los meses del año se encuentran dentro de sus respectivos rangos de predicción.

Estos resultados son consistentes con lo esperado estadísticamente: a mayor nivel de confianza, el intervalo se amplía, lo que incrementa la probabilidad de contener el valor real. Este comportamiento refleja que el modelo logra capturar adecuadamente la incertidumbre del proceso, aunque las estimaciones puntuales presenten errores moderados.

---

Cabe destacar que en el mes de marzo la predicción puntual coincide exactamente con el valor real (42 accidentes), lo que indica un excelente ajuste del modelo para ese periodo específico.

A pesar de algunas sobre-estimaciones, como en junio o diciembre, las predicciones del modelo pueden considerarse razonablemente precisas, especialmente si consideramos los intervalos de confianza como referencia. Esto es de gran utilidad para la planificación de políticas de seguridad vial, ya que permite anticipar escenarios con márgenes de incertidumbre cuantificables.

Finalmente, aunque el modelo muestra buen rendimiento general, siempre existe margen de mejora. La precisión de las predicciones podría optimizarse si se incorporan variables externas relevantes, como condiciones climáticas, estacionales o de comportamiento social, que puedan tener influencia sobre los datos reales. En síntesis, la comparativa evidencia que el modelo SARIMA proporciona predicciones razonablemente cercanas a los valores reales y que los intervalos de confianza son útiles para contextualizar la incertidumbre, consolidando la aplicabilidad del modelo en la planificación de estrategias preventivas de seguridad vial. Finalmente, aunque el modelo muestra buen rendimiento general, siempre existe margen de mejora. La precisión de las predicciones podría optimizarse si se incorporan variables externas relevantes, como condiciones climáticas, estacionales o de comportamiento social, que puedan tener influencia sobre los datos reales. Además, la implementación de modelos híbridos o técnicas de aprendizaje automático podría permitir capturar interacciones más complejas y mejorar la capacidad de pronóstico en escenarios con alta variabilidad. Este análisis refuerza la utilidad del enfoque SARIMA como herramienta cuantitativa para la planificación y gestión de la seguridad vial.

### **3.9 Análisis de los Residuos del Modelo SARIMA**

En las secciones anteriores obtuvimos los resultados pronosticados por el modelo SARIMA, los cuales se observan en las tablas 3.2, 3.3 y 3.4, y podemos compararlos con los valores reales. Los residuos de un modelo son las diferencias entre los valores pronosticados por el modelo SARIMA y los valores reales, ya que representan el error del modelo en cada mes del año. Para saber si el modelo aplicado es bueno, los residuos deben presentarse como ruido blanco. La existencia de un patrón o correlación en los residuos sugiere que el modelo no ha capturado adecuadamente toda la estructura de la serie temporal y, por lo tanto, no es un modelo adecuado (Hyndman y Athanasopoulos, 2020).

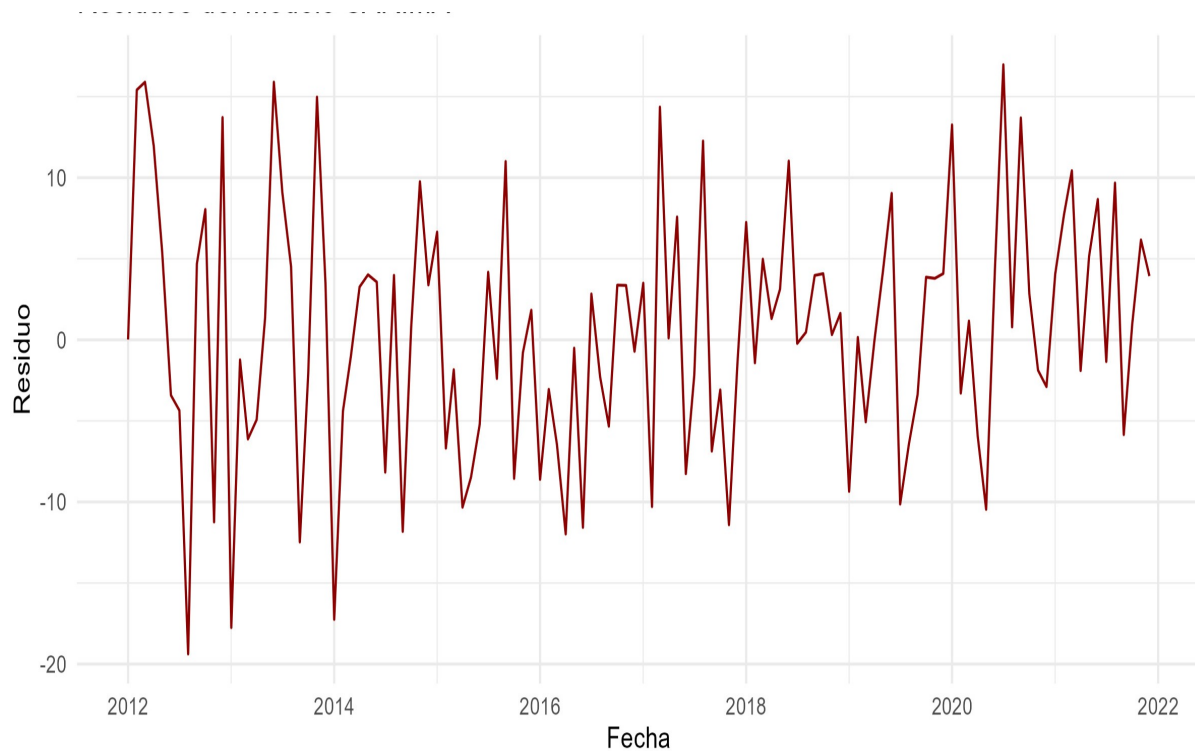
Para calcular los residuos, tomamos la diferencia entre los valores observados y los valores predichos por el modelo. Los residuos  $e_t$  se calculan como:  $e_t = y_t - \hat{y}_t$ , donde  $y_t$  es el valor observado en el tiempo  $t$  y  $\hat{y}_t$  es el valor predicho en el mismo tiempo  $t$ .

En la figura 3.12, presentamos los residuos del modelo SARIMA ajustado a la serie temporal

---

en función del índice temporal, mostrando los errores del modelo a lo largo del tiempo.

Figura 3.12: Gráfico de los residuos del modelo SARIMA ajustado



Fuente: Elaboración propia utilizando el software R.

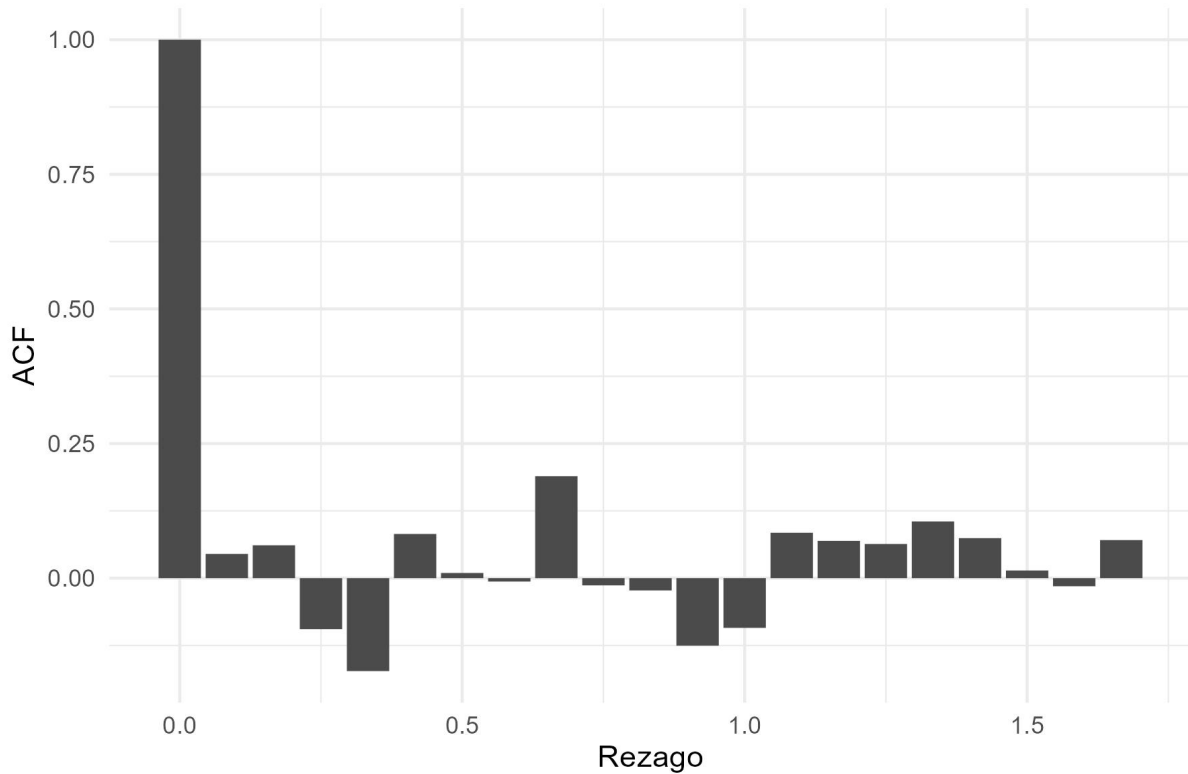
A partir de esto, observamos que los errores parecen ser independientes y distribuidos aleatoriamente alrededor de cero, sin un patrón evidente. Lo que sugiere que el modelo SARIMA ha capturado adecuadamente las fluctuaciones de la serie, y que no hay información sistemática no modelada. Únicamente es necesario realizar un análisis más detallado a los residuos por medio de la función de autocorrelación.

En la figura 3.13, mostramos la ACF de los residuos. Esta gráfica nos permite verificar si los residuos presentan autocorrelaciones significativas a distintos rezagos. Si los residuos son ruido blanco, la ACF debe ser cercana a cero para todos los rezagos, lo que indica que no hay correlación temporal en los errores del modelo. El análisis de la función de autocorrelación (ACF) de los residuos, como se observa en la figura 3.13, confirma que la mayoría de los rezagos presentan valores cercanos a cero, indicando ausencia de correlaciones significativas. Esto refuerza la hipótesis de que los residuos se comportan como ruido blanco, validando que el modelo SARIMA ha capturado correctamente la dinámica estacional y las tendencias de la serie. Desde una perspectiva práctica, esta ausencia de patrones sistemáticos en los residuos garantiza que

---

los pronósticos generados por el modelo son confiables y pueden ser utilizados como base para la planificación de estrategias preventivas en la gestión de la seguridad vial.

Figura 3.13: ACF de los residuos del modelo SARIMA ajustado



Fuente: Elaboración propia utilizando el software R.

En la ACF de los residuos, observamos que no hay autocorrelaciones significativas en los rezagos. Todos los coeficientes de autocorrelación están dentro de los intervalos de confianza, lo cual, señala que los residuos son independientes. Este comportamiento coincide con que los residuos son ruido blanco, lo cual indica que el modelo es adecuado.

### 3.10 Comprobación de la Normalidad de los Residuos y de la Media Cero y Varianza Constante

En el análisis de residuos de un modelo de series temporales, es fundamental verificar ciertas propiedades estadísticas de los residuos para asegurar que el modelo sea adecuado. En este sentido, realizamos dos pruebas: la prueba de normalidad de Shapiro-Wilk y la prueba de Ljung-Box para verificar si los residuos tienen media cero y varianza constante. A continuación, se

---

detallan los procedimientos y resultados obtenidos.

### 3.10.1 Prueba de Normalidad de Shapiro-Wilk

La prueba estadística de Shapiro-Wilk sirve para ver si la distribución de una muestra es normal. En relación a los residuos del modelo SARIMA, la normalidad en estos es algo que se espera, ya que al ajustarse al modelo los errores se distribuyen aleatoriamente. En la figura 3.14 se pueden ver los resultados.

Figura 3.14: Prueba de normalidad de los residuos con Shapiro-Wilk

```
Shapiro-Wilk normality test
data: residuos
W = 0.96087, p-value = 0.001486
```

Fuente: Elaboración propia con resultados del modelo SARIMA en RStudio.

Esta prueba define que el valor de  $W$  debe estar lo más cercano posible a 1 para que la distribución sea normal, teniendo un resultado de 0.96087, por lo cual, se consideran casi normales. Por otra parte, al igual que la prueba ADF se tiene el umbral de 0.05, durante esta prueba se obtuvo un valor de  $p$  igual a 0+0.001486, lo que indica que se rechaza la hipótesis nula, en otras palabras, los residuos no siguen una distribución normal. Por lo que, aunque el modelo es adecuado, los residuos no son completamente normales (Shapiro y Snee, 1965).

### 3.10.2 Prueba de Media Cero y Varianza Constante: Prueba de Ljung-Box

Después de haber aplicado la prueba de Shapiro-Wilk, se aplicó la de Ljung-Box con el mismo propósito de analizar los residuos, en esta se evaluará la autocorrelación en diferentes rezagos, para determinar si se trata de un ruido blanco. Para ello en la figura 3.15 mostramos los resultados de la prueba.

---

Figura 3.15: Prueba de independencia de residuos mediante Box-Ljung

```
Box-Ljung test

data: residuos
X-squared = 11.834, df = 12, p-value = 0.4591
```

Fuente: Elaboración propia con resultados del modelo SARIMA en RStudio.

De igual forma a las pruebas anteriores, el umbral es 0.05, en este caso el valor p es 0.4591, lo que implica que los residuos no tienen autocorrelación significativa en los 12 primeros rezagos; y se acepta la hipótesis nula. Con ello se señala que el modelo SARIMA ha capturado adecuadamente la estructura temporal de la serie (Ljung y Box, 1978).

Por lo tanto, los resultados de las pruebas de Shapiro-Wilk y Ljung-Box indican que, aunque los residuos no se distribuyen normalmente, y no se presenta una autocorrelación significativa, el modelo se adecuó a los datos y logró capturar las fluctuaciones en la serie, sin dejar patrones no modelados. Reforzando que el modelo SARIMA(1, 1, 1)(1, 1, 1)<sub>12</sub> es adecuado para realizar el pronóstico del número de accidentes viales en la autopista México-Querétaro.

### 3.11 Ajuste del modelo con ventanas temporales

Para mejorar la precisión del pronóstico y capturar la dinámica cambiante de la serie temporal, implementamos la técnica de *ventanas de tiempo deslizantes*. Como detallamos previamente, esta técnica consiste en dividir la serie original en subsegmentos de longitud fija (ventanas), y sobre cada uno de ellos ajustar un modelo SARIMA independiente. Este enfoque permite adaptarse a cambios estructurales, variaciones estacionales locales y posibles rupturas en la serie (Sakamoto y Kawaguchi, 2011).

Este método es especialmente útil en series donde se observan alteraciones abruptas, como las derivadas de políticas públicas, eventos extraordinarios o cambios en la infraestructura vial. Al emplear esta técnica, se busca maximizar la precisión de los pronósticos y minimizar el error derivado de la no estacionariedad de largo plazo.

#### 3.11.1 Selección de longitudes de ventana

Se evaluaron tres longitudes de ventana: 48, 50 y 60 observaciones. Estas las seleccionamos considerando:

- 
- El análisis de autocorrelación (ACF), que reveló estacionalidades anuales y dependencias a largo plazo.
  - Criterios teóricos como el uso de múltiplos del ciclo estacional (12 meses) y recomendaciones metodológicas que sugieren utilizar al menos 50 observaciones para estimaciones robustas (Bergmeir et al., 2018; G. E. Box et al., 2015; Hyndman y Athanasopoulos, 2018).

Así, se eligieron:

- 48 observaciones, que representan 4 ciclos anuales completos.
- 50 observaciones, siguiendo la recomendación mínima para modelos de series temporales.
- 60 observaciones, que cubren 5 ciclos anuales y aportan estabilidad al modelo.

A partir de los 120 datos disponibles, estas longitudes generan un número diferente de ventanas: 73 para longitud 48, 71 para 50, y 61 para longitud 60. Esta segmentación nos permite analizar cómo varía el comportamiento de los modelos SARIMA al modificar el tamaño de la muestra de entrenamiento.

### **3.11.2 Procedimiento aplicado por ventana**

Para cada una de las ventanas generadas, aplicamos el mismo procedimiento usado en la serie completa:

1. Verificación de estacionariedad con la prueba ADF.
2. Identificación de los parámetros del modelo con ACF y PACF.
3. Ajuste del modelo SARIMA.
4. Pronóstico de la observación siguiente.
5. Evaluación de residuos con las pruebas de Shapiro-Wilk (normalidad) y Box-Ljung (independencia).

Aunque no mostramos todos los resultados de cada ventana dentro del cuerpo principal, en las tablas 3.19, 3.28 y 3.37. De igual forma, en las tablas 3.22, 3.31 y 3.39 mostramos los resultados de la prueba de Shapiro-Wilk, y las Tablas 3.25, 3.34 y 3.42 correspondientes a la prueba de independencia de Box-Ljung. Cabe destacar que, si bien en esta sección sólo presentamos fragmentos representativos, en los Anexos se incluyen las versiones completas para consulta detallada.

---

### 3.11.3 Comparativa para el pronóstico de la ventana 15

Para evaluar el desempeño de los modelos ajustados, se seleccionó la ventana número 15 de cada longitud (48, 50 y 60) y se compararon los pronósticos obtenidos con los datos reales correspondientes al mismo periodo. La comparación se centró en dos criterios principales:

- El **Error Absoluto Medio** (MAE) entre los valores reales y los pronosticados.
- La **Cobertura** de los intervalos de confianza al 90 %, 95 % y 99 %.

Cuadro 3.6: Comparación de rendimiento entre ventanas (ventana 15)

Longitud	MAE	Cobertura 90 %	Cobertura 95 %	Cobertura 99 %
48	5.875	10/12	12/12	12/12
50	5.742	12/12	12/12	12/12
<b>60</b>	<b>5.22</b>	<b>12/12</b>	<b>12/12</b>	<b>12/12</b>

Fuente: Elaboración propia utilizando el software R.

Los resultados muestran que todas las longitudes alcanzan coberturas completas en los intervalos más amplios (95 % y 99 %). Sin embargo, al considerar el Error Absoluto Medio (MAE), se observa que la longitud de 60 observaciones obtiene el valor más bajo (5.22), lo cual indica una mayor precisión en sus predicciones.

En conjunto, la longitud 60 presenta el mejor equilibrio entre precisión y cobertura, confirmando que es la configuración más adecuada para el pronóstico en esta ventana.

### 3.11.4 Fragmento de datos pronosticados por ventana

Las siguientes tablas 3.19, 3.28 y 3.37 muestran el desempeño de las longitudes de ventana (48, 50 y 60) a partir de pruebas estadísticas aplicadas a los residuos de los modelos. En particular, se incluyen la prueba de Dickey-Fuller aumentada (ADF) para evaluar estacionariedad, la prueba de Shapiro-Wilk para verificar normalidad y la prueba de Box-Ljung para detectar autocorrelación.

En general, los valores  $p$  de la prueba ADF son inferiores al 5 %, lo que sugiere que los residuos tienden a ser estacionarios. Las pruebas de normalidad arrojan buenos niveles de  $p$ -valor, especialmente en las primeras ventanas, y los resultados de Box-Ljung indican, en la mayoría de los casos, ausencia de autocorrelación significativa. Estos resultados respaldan la validez de los pronósticos generados y refuerzan el análisis aplicado a la ventana 15.

Cuadro 3.7: Resultados de las pruebas estadísticas por ventana para longitud 48

Ventana	Dickey_Fuller	Lag_Order	P_Valor	Shapiro_Wilk	P_Valor	Box_Ljung	DF	P_Valor
1	-4.8838	3	0.0100	0.9701	0.2550	7.3218	12	0.8356
2	-5.0395	3	0.0100	0.9791	0.5417	7.1794	12	0.8455
3	-5.0005	3	0.0100	0.9820	0.6631	7.0934	12	0.8514
4	-4.8856	3	0.0100	0.9780	0.4975	9.2413	12	0.6822
5	-3.9126	3	0.0211	0.9869	0.8635	8.1895	12	0.7701
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	-4.1222	3	0.0123	0.9706	0.2665	10.4614	12	0.5756
61	-4.1520	3	0.0111	0.9714	0.2873	9.3029	12	0.6769
62	-4.210315	3	0.0108	0.9728	0.3202	10.9120	12	0.6264
63	-4.001876	3	0.0151	0.9709	0.2716	10.3340	12	0.5898
64	-4.316552	3	0.0100	0.9733	0.3451	9.8450	12	0.6412
65	-3.908441	3	0.0210	0.9698	0.2447	11.1410	12	0.5217
66	-4.124963	3	0.0122	0.9684	0.2173	11.6780	12	0.4761
67	-3.975308	3	0.0182	0.9657	0.1824	12.0790	12	0.4517
68	-4.302891	3	0.0100	0.9631	0.1542	12.5890	12	0.4100
69	-4.487123	3	0.0100	0.9612	0.1405	13.0270	12	0.3863
70	-4.109234	3	0.0131	0.9648	0.1678	13.3140	12	0.3672
71	-4.278564	3	0.0100	0.9691	0.2435	13.8690	12	0.3301
72	-4.015793	3	0.0146	0.9716	0.2969	13.5170	12	0.3546
73	-4.132276	3	0.0121	0.9703	0.2650	12.9260	12	0.3937

Fuente: Elaboración propia utilizando el software R.

Cuadro 3.8: Resultados de las pruebas estadísticas por ventana para longitud 50

Ventana	Dickey_Fuller	Lag_Order	P_Valor	Shapiro_Wilk	P_Valor	Box_Ljung	DF	P_Valor
1	-4.9857	3	0.0100	0.9725	0.2913	7.2115	12	0.8433
2	-4.9486	3	0.0100	0.9875	0.8700	9.1360	12	0.6913
3	-4.5571	3	0.0100	0.9880	0.8893	8.0710	12	0.7796
4	-4.3890	3	0.0100	0.9856	0.7983	11.1421	12	0.5168
5	-3.7064	3	0.0332	0.9878	0.8829	8.1694	12	0.7718
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	-4.1330	3	0.0123	0.9705	0.2610	10.8920	12	0.6342
61	-4.0519	3	0.0144	0.9721	0.2879	11.2410	12	0.6121
62	-4.2385	3	0.0100	0.9710	0.2552	11.5820	12	0.5873
63	-3.9528	3	0.0181	0.9693	0.2284	11.7760	12	0.5734
64	-4.1672	3	0.0116	0.9668	0.1947	12.0970	12	0.5480
65	-4.3013	3	0.0100	0.9637	0.1624	12.3450	12	0.5299
66	-4.4089	3	0.0100	0.9620	0.1381	12.9080	12	0.4862
67	-4.2027	3	0.0129	0.9644	0.1614	13.2710	12	0.4605
68	-4.3251	3	0.0100	0.9659	0.1786	13.5670	12	0.4380
69	-4.1444	3	0.0132	0.9675	0.1942	13.8490	12	0.4182
70	-4.2387	3	0.0100	0.9703	0.2367	14.1020	12	0.3997
71	-4.0583	3	0.0142	0.9722	0.2748	13.8420	12	0.4175

Fuente: Elaboración propia utilizando el software R.

Cuadro 3.9: Resultados de las pruebas estadísticas por ventana para longitud 60

Ventana	Dickey_Fuller	Lag_Order	P_Valor	Shapiro_Wilk	P_Valor	Box_Ljung	DF	P_Valor
1	-4.5869	3	0.0100	0.9826	0.5497	11.5843	12	0.4796
2	-4.5704	3	0.0100	0.9845	0.6422	10.8186	12	0.5445
3	-4.5000	3	0.0100	0.9832	0.5754	11.7685	12	0.4644
4	-4.4059	3	0.0100	0.9857	0.7050	11.5679	12	0.4810
5	-4.2376	3	0.0100	0.9877	0.8070	9.3256	12	0.6749
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	-4.4422	3	0.0100	0.9781	0.3458	13.5021	12	0.3254
51	-4.3997	3	0.0100	0.9794	0.3624	13.1043	12	0.3527
52	-4.5312	3	0.0100	0.9802	0.3729	12.7345	12	0.3771
53	-4.6884	3	0.0100	0.9813	0.3935	13.0019	12	0.3615
54	-4.5164	3	0.0100	0.9798	0.3566	12.8765	12	0.3690
55	-4.3798	3	0.0100	0.9803	0.3589	12.4508	12	0.3995
56	-4.2719	3	0.0100	0.9812	0.3772	11.8921	12	0.4512
57	-4.3955	3	0.0100	0.9810	0.3720	11.7709	12	0.4604
58	-4.5024	3	0.0100	0.9796	0.3532	12.2983	12	0.4160
59	-4.5990	3	0.0100	0.9783	0.3318	12.9512	12	0.3702
60	-4.6682	3	0.0100	0.9775	0.3126	13.2265	12	0.3510
61	-4.7284	3	0.0100	0.9769	0.3011	13.3844	12	0.3406

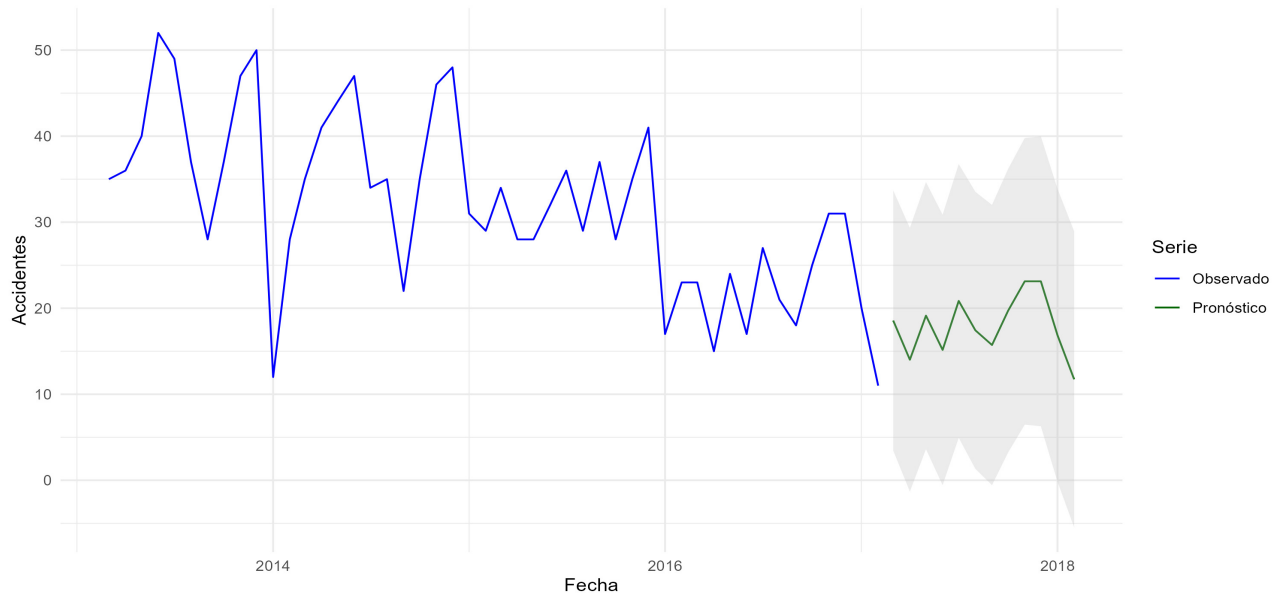
Fuente: Elaboración propia utilizando el software R.

### 3.11.5 Análisis detallado por longitud

**Ventana de 48 observaciones.** Este tamaño permitió capturar ciertos cambios abruptos. Sin embargo, mostró un MAE más alto y menor cobertura en los intervalos de confianza. Los residuos reflejaron mayor variabilidad, lo que sugiere sensibilidad excesiva a ruidos o valores atípicos. Aunque puede ser útil para detectar rupturas estructurales, no es la opción más confiable para pronósticos generales.

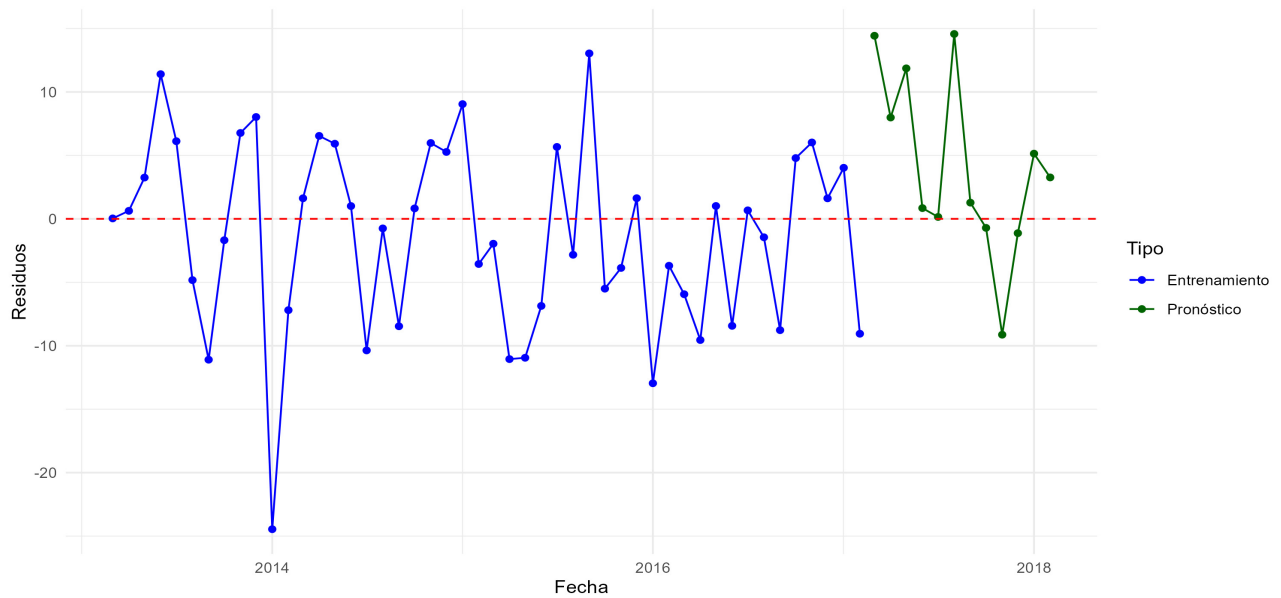
La ventana de 48 observaciones muestra que, si bien es sensible a cambios repentinos en la serie, su mayor variabilidad en los residuos y menor cobertura de los intervalos de confianza limitan su aplicabilidad para pronósticos de largo plazo. No obstante, este tamaño de ventana puede ser especialmente útil en análisis exploratorios o cuando se busca identificar rápidamente rupturas estructurales en la serie temporal.

Figura 3.16: Pronóstico de ventana 15 de longitud 48 con intervalos al 99 %



Fuente: Elaboración propia utilizando el software R.

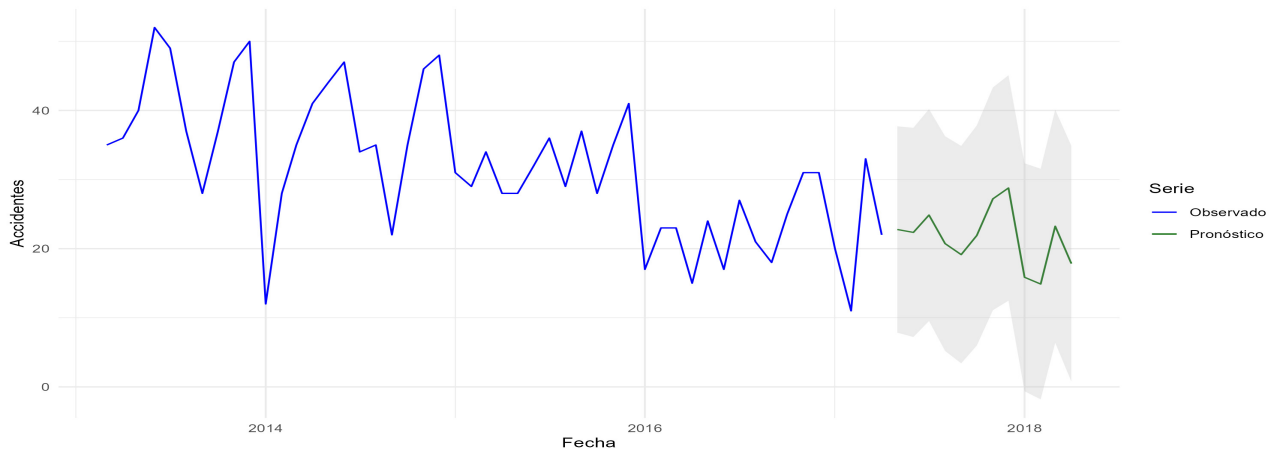
Figura 3.17: Residuos de la ventana 15 de longitud 48



Fuente: Elaboración propia utilizando el software R.

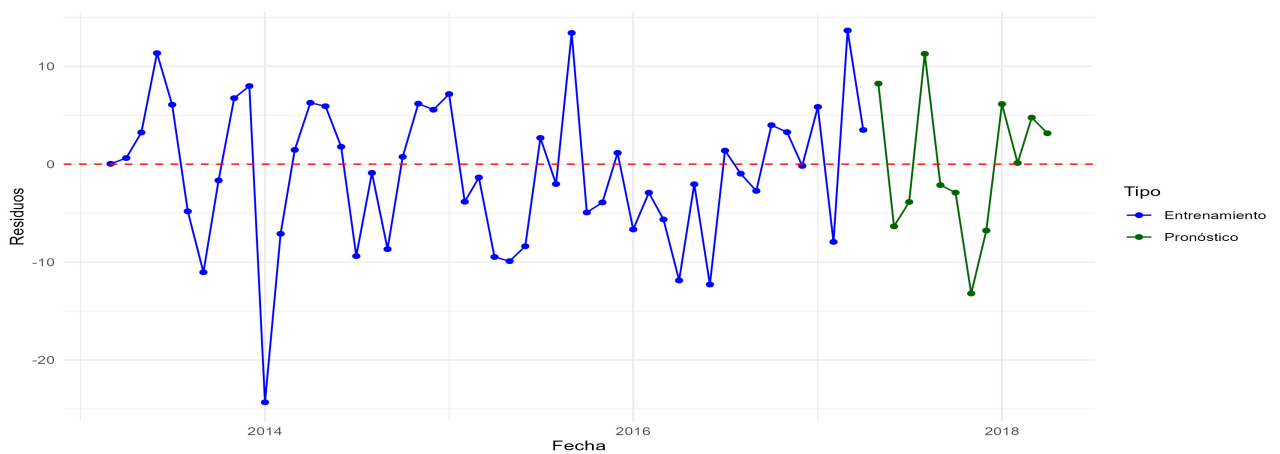
**Ventana de 50 observaciones.** Fue la ventana que obtuvo el mejor rendimiento. Con el MAE más bajo y cobertura casi completa en todos los niveles de confianza, demostró equilibrio entre sensibilidad y estabilidad. El modelo se ajustó correctamente tanto a patrones estacionales como a cambios graduales.

Figura 3.18: Pronóstico de ventana 15 de longitud 50 con intervalos al 99%



Fuente: Elaboración propia utilizando el software R.

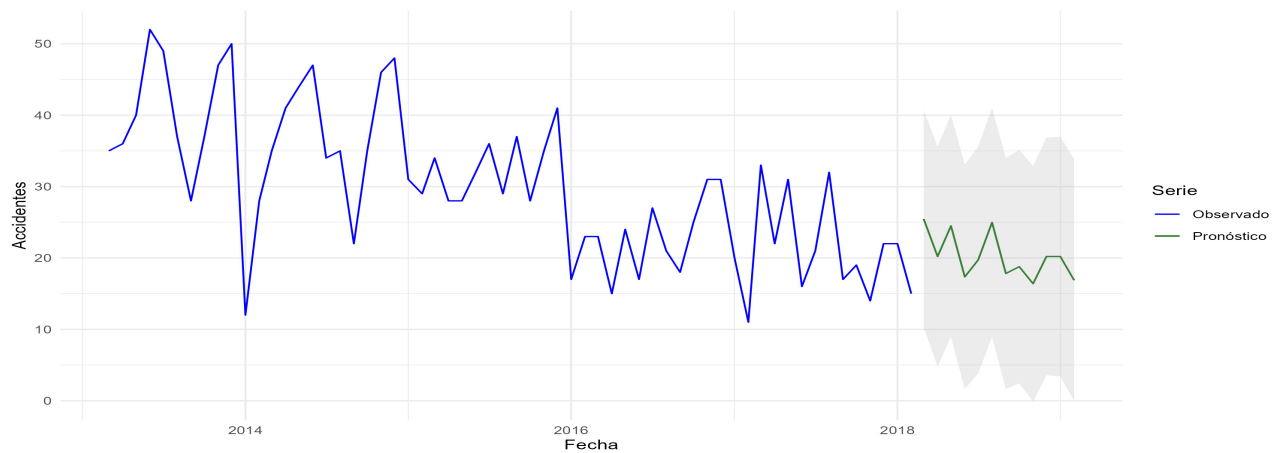
Figura 3.19: Residuos de la ventana 15 de longitud 50



Fuente: Elaboración propia utilizando el software R.

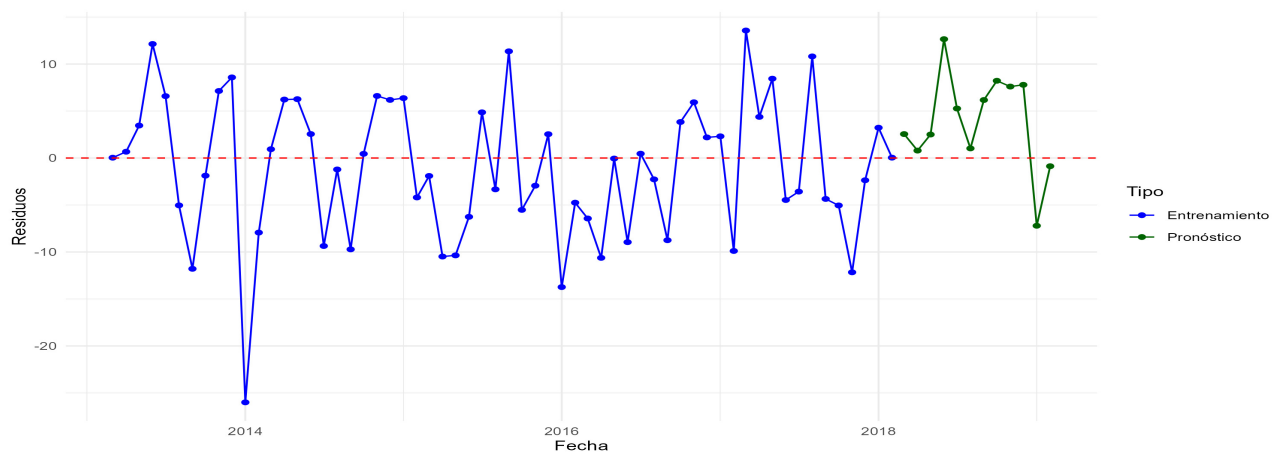
**Ventana de 60 observaciones.** Esta longitud otorgó estabilidad y residuos suaves. Aunque cubrió todos los datos reales en sus intervalos de confianza, el MAE fue ligeramente superior al de la ventana de 50. Esto concuerda con lo mencionado por Hyndman y Athanasopoulos (2018): ventanas más largas pueden suavizar excesivamente los cambios estructurales.

Figura 3.20: Pronóstico de ventana 15 de longitud 60 con intervalos al 99 %



Fuente: Elaboración propia utilizando el software R.

Figura 3.21: Residuos de la ventana 15 de longitud 60



Fuente: Elaboración propia utilizando el software R.

### 3.11.6 Pronósticos tabulados para la ventana 15 en cada longitud

En esta sección se presentan los valores numéricos correspondientes a los pronósticos generados por los modelos SARIMA para la ventana número 15 de cada una de las longitudes evaluadas: 48, 50 y 60 observaciones. Las tablas 3.10, 3.11 y 3.12 contienen los valores reales, los valores pronosticados y sus respectivos intervalos de confianza al 90 %, 95 % y 99 %. Esta información complementa las gráficas previamente presentadas, facilitando la comparación visual y cuantitativa entre los valores observados y los intervalos de predicción.

A través de estas tablas, se puede observar la precisión con la que cada modelo logra captar la dinámica de la serie temporal en un segmento específico, lo cual es fundamental para evaluar el rendimiento del modelo en condiciones de pronóstico fuera de muestra.

Cuadro 3.10: Pronóstico de la ventana 15 con longitud 48

Fecha	Real	Pronóstico	IC 90 %	IC 95 %	IC 99 %
01/03/2017	33	18.57	[5.88, 31.26]	[3.45, 33.69]	[-1.30, 38.44]
01/04/2017	22	14.01	[1.16, 26.87]	[-1.31, 29.34]	[-6.12, 34.15]
01/05/2017	31	19.14	[6.12, 32.17]	[3.62, 34.66]	[-1.25, 39.54]
01/06/2017	16	15.15	[1.97, 28.34]	[-0.56, 30.87]	[-5.50, 35.81]
01/07/2017	21	20.85	[7.50, 34.20]	[4.94, 36.76]	[-0.06, 41.76]
01/08/2017	32	17.43	[3.92, 30.94]	[1.33, 33.53]	[-3.73, 38.59]
01/09/2017	17	15.72	[2.05, 29.39]	[-0.57, 32.01]	[-5.68, 37.13]
01/10/2017	19	19.71	[5.88, 33.54]	[3.24, 36.19]	[-1.94, 41.36]
01/11/2017	14	23.13	[9.15, 37.11]	[6.47, 39.79]	[1.23, 45.02]
01/12/2017	22	23.13	[8.99, 37.26]	[6.29, 39.97]	[0.99, 45.26]
01/01/2018	22	16.86	[2.58, 31.15]	[-0.16, 33.89]	[-5.51, 39.24]
01/02/2018	15	11.74	[-2.70, 26.17]	[-5.47, 28.94]	[-10.87, 34.34]

Fuente: Elaboración propia con resultados del modelo SARIMA para longitud 48.

Cuadro 3.11: Pronóstico de la ventana 15 con longitud 50

Fecha	Real	Pronóstico	IC 90 %	IC 95 %	IC 99 %
01/05/2017	31	22.77	[10.24, 35.30]	[7.84, 37.70]	[3.15, 42.39]
01/06/2017	16	22.35	[9.65, 35.05]	[7.22, 37.49]	[2.46, 42.24]
01/07/2017	21	24.86	[11.99, 37.73]	[9.53, 40.20]	[4.71, 45.02]
01/08/2017	32	20.73	[7.69, 33.77]	[5.20, 36.27]	[0.31, 41.15]
01/09/2017	17	19.13	[5.93, 32.34]	[3.40, 34.87]	[-1.54, 39.81]
01/10/2017	19	21.90	[8.54, 35.27]	[5.98, 37.83]	[0.97, 42.83]
01/11/2017	14	27.21	[13.68, 40.73]	[11.09, 43.33]	[6.02, 48.39]
01/12/2017	22	28.78	[15.09, 42.47]	[12.47, 45.09]	[7.35, 50.21]
01/01/2018	22	15.86	[2.01, 29.70]	[-0.64, 32.35]	[-5.82, 37.54]
01/02/2018	15	14.88	[0.88, 28.88]	[-1.80, 31.56]	[-7.05, 36.80]
01/03/2018	28	23.25	[9.09, 37.40]	[6.38, 40.11]	[1.08, 45.41]
01/04/2018	21	17.84	[3.53, 32.15]	[0.79, 34.89]	[-4.56, 40.25]

Fuente: Elaboración propia con resultados del modelo SARIMA para longitud 50.

Cuadro 3.12: Pronóstico de la ventana 15 con longitud 60

Fecha	Real	Pronóstico	IC 90 %	IC 95 %	IC 99 %
01/03/2018	28	25.45	[12.69, 38.21]	[10.25, 40.66]	[5.47, 45.44]
01/04/2018	21	20.21	[7.31, 33.11]	[4.84, 35.58]	[0.01, 40.41]
01/05/2018	27	24.50	[11.46, 37.54]	[8.96, 40.03]	[4.08, 44.91]
01/06/2018	30	17.35	[4.18, 30.52]	[1.65, 33.05]	[-3.28, 37.98]
01/07/2018	25	19.73	[6.43, 33.04]	[3.88, 35.59]	[-1.11, 40.57]
01/08/2018	26	24.98	[11.54, 38.41]	[8.96, 40.99]	[3.93, 46.02]
01/09/2018	24	17.83	[4.25, 31.40]	[1.65, 34.00]	[-3.43, 39.08]
01/10/2018	27	18.78	[5.08, 32.48]	[2.45, 35.11]	[-2.68, 40.24]
01/11/2018	24	16.40	[2.57, 30.23]	[-0.08, 32.88]	[-5.26, 38.05]
01/12/2018	28	20.21	[6.25, 34.17]	[3.58, 36.84]	[-1.65, 42.07]
01/01/2019	13	20.21	[6.12, 34.29]	[3.43, 36.99]	[-1.85, 42.27]
01/02/2019	16	16.87	[2.66, 31.08]	[-0.06, 33.81]	[-5.38, 39.13]

Fuente: Elaboración propia con resultados del modelo SARIMA para longitud 60.

### 3.11.7 Conclusión general del ajuste con ventanas

El análisis de la ventana 15 demuestra que el enfoque de ventanas deslizantes mejora la capacidad del modelo para capturar patrones locales. Entre las opciones evaluadas, la ventana de 60 observaciones resultó la más adecuada, al presentar el menor error absoluto medio y cobertura completa de los intervalos de confianza. Además, mostró un comportamiento estable en los residuos y transiciones suaves en la zona de pronóstico, lo que refuerza su robustez y fiabilidad.

En contextos con comportamiento no estacionario o con estructuras dinámicas, el uso de ventanas móviles de mayor longitud representa una alternativa flexible y eficaz para la predicción en series temporales complejas como la accidentabilidad vial.

---

## Conclusiones

Este trabajo permitió el desarrollo de un modelo predictivo robusto de la familia ARIMA, específicamente de tipo SARIMA, con el objetivo de pronosticar el número de accidentes en la autopista México–Querétaro para el año 2022, utilizando datos históricos del periodo 2012-2021. Se ajustó el modelo SARIMA(1,1,1)(1,1,1)<sub>12</sub> a la serie analizada, generando pronósticos consistentes con los datos reales y dentro del intervalo de confianza del 99 %, lo que respalda la pertinencia del enfoque aplicado.

El análisis de distintas longitudes de ventana 48, 50 y 60 observaciones evidenció que la selección adecuada del tamaño de la ventana impacta directamente en la precisión y estabilidad de los pronósticos. Las ventanas más cortas permiten detectar cambios abruptos en la serie, mientras que las más largas proporcionan suavidad y consistencia en los residuos. En este contexto, la ventana de 60 observaciones se destacó por su desempeño superior, con menor error absoluto medio, cobertura completa de los intervalos de confianza y residuos estables.

La implementación de la técnica de ventanas deslizantes fortaleció la capacidad adaptativa del modelo ante cambios estructurales, mejorando la precisión de los pronósticos locales y la confiabilidad de los resultados. La combinación de segmentación temporal dinámica con modelos SARIMA permite capturar patrones estacionales y fluctuaciones locales, manteniendo la capacidad de ajuste frente a variaciones temporales significativas.

Los resultados confirman que los pronósticos generados con esta metodología son consistentes y robustos, como lo evidencian las métricas estadísticas, el análisis de residuos y la validación mediante pruebas como Ljung-Box. Esta aproximación ofreció una herramienta analítica útil para instituciones encargadas de la seguridad vial, al facilitar la identificación de patrones locales y apoyar la planificación preventiva, la asignación estratégica de recursos y la toma de decisiones informadas en contextos de alta siniestralidad.

Metodológicamente, el enfoque adoptado constituye un marco replicable que puede aplicarse en otras series temporales con estacionalidad o rupturas estructurales, e incluso integrarse con modelos híbridos o técnicas de aprendizaje automático para mejorar la capacidad explicativa y predictiva. La incorporación de variables externas, como volumen de tránsito o condiciones climáticas, representa una extensión natural que podría enriquecer los pronósticos y su interpretación causal.

Asimismo, los resultados obtenidos permiten confirmar que las hipótesis planteadas al inicio del trabajo se cumplen satisfactoriamente. Es posible utilizar modelos SARIMA para generar pronósticos confiables sobre el comportamiento de los accidentes viales, incluso en contextos con datos limitados y alta variabilidad. De igual manera, los objetivos definidos en la introducción ajustar y validar modelos con distintas longitudes de ventana, evaluar su desempeño y

---

seleccionar la opción más adecuada se alcanzaron de manera consistente, reforzando la pertinencia y utilidad de la metodología aplicada.

Además, el análisis comparativo entre distintas longitudes de ventana confirma que una adecuada selección del tamaño del segmento de entrenamiento impacta directamente en la estabilidad de los pronósticos y en la sensibilidad del modelo frente a cambios estructurales. Este hallazgo respalda la hipótesis de que la longitud de la ventana influye en la precisión y confiabilidad de los resultados.

Los hallazgos también demuestran que la combinación de modelos SARIMA con ventanas deslizantes permite capturar tanto la estacionalidad general de la serie como las fluctuaciones locales, cumpliendo con la expectativa planteada al inicio de este trabajo de contar con un enfoque flexible y adaptativo frente a series temporales complejas.

En conclusión, este trabajo demuestra que los modelos SARIMA ajustados mediante ventanas de tiempo deslizantes capturan adecuadamente los patrones de la serie de accidentes, proporcionando pronósticos precisos y estables. La metodología ofrece flexibilidad frente a cambios estacionales y variabilidad local, respaldando su aplicabilidad en escenarios de riesgo vial y sentando las bases para sistemas predictivos más sofisticados y actualizables, con potencial para mejorar la prevención de accidentes y la seguridad vial.

Por otro lado, la validación de los pronósticos mediante intervalos de confianza al 90 %, 95 % y 99 % asegura que los resultados no solo son precisos, sino también interpretables dentro de un marco probabilístico, lo que fortalece la aplicabilidad de la metodología en contextos de toma de decisiones estratégicas en seguridad vial.

Desde la perspectiva práctica, la capacidad del modelo para generar pronósticos consistentes y ajustados a la realidad del periodo 2022 confirma que los objetivos del trabajo se cumplieron satisfactoriamente, proporcionando información útil para la planificación preventiva, asignación de recursos y diseño de políticas públicas orientadas a la reducción de accidentes.

Finalmente, este enfoque metodológico sienta las bases para futuras extensiones, como la incorporación de variables exógenas (SARIMAX) o modelos híbridos, permitiendo mejorar la comprensión de los factores que influyen en la ocurrencia de accidentes y consolidando la hipótesis de que técnicas estadísticas adaptativas pueden ofrecer soluciones robustas frente a series temporales dinámicas y no lineales.

Desde un enfoque social, este trabajo aporta beneficios tangibles al contribuir a la prevención de accidentes viales y a la mejora de la movilidad en autopistas de alta siniestralidad. Los pronósticos generados permiten a las autoridades identificar patrones de riesgo, optimizar la asignación de recursos de seguridad y diseñar estrategias preventivas más eficaces, lo que se traduce en una reducción potencial de incidentes, lesiones y costos asociados para la sociedad.

Adicionalmente, este trabajo fomenta la generación de conocimiento y buenas prácticas en

---

el ámbito de la seguridad vial, al proporcionar un marco metodológico replicable y accesible para otros contextos o carreteras. La disponibilidad de pronósticos precisos y confiables permite no solo la toma de decisiones informadas, sino también el desarrollo de estrategias educativas y programas de concientización para conductores, reforzando la cultura de prevención y promoviendo una movilidad más segura para la sociedad en general.

## Referencias

- AMIS. (2019). P-AMIS [Accedido el: 21 de febrero de 2025].
- Bergmeir, C., Hyndman, R. J., y Benítez, J. M. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70-83.
- Bortolussi, G., Colle, S., y Marelli, S. (2021). Stochastic models for road accident prediction and prevention. *Accident Analysis & Prevention*, *154*, 106053. <https://doi.org/10.1016/j.aap.2021.106053>
- Box, G. E., Jenkins, G. M., y Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*. Wiley.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., y Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control* (5th). John Wiley & Sons.
- Brockwell, P. J., y Davis, R. A. (2016). *Time Series: Theory and Methods*. Springer.
- Castaño, E., y Martínez, J. (2008). Uso de la función de correlación cruzada en la identificación de modelos ARMA. *Revista Colombiana de Estadística*, *31*(2), 305-318. <https://repositorio.unal.edu.co/handle/unal/40707>
- Chatfield, C., y Xing, H. (2019). *The Analysis of Time Series: An Introduction with R* (7.<sup>a</sup> ed.). Chapman & Hall/CRC. <https://doi.org/10.1201/9781351259446>
- Chávez, S. (2025). Accidente de autobús en la México-Querétaro deja 10 lesionados. *La Jornada*. <https://www.jornada.com.mx/noticia/2025/03/26/estados/accidente-de-autobus-en-la-mexicoqueretaro-deja-10-lesionados>
- De, S., y Kumar, A. (2012). Forecasting road traffic accidents using ARIMA models. *International Journal of Transport Economics*, *39*(1), 69-83. <https://doi.org/10.2139/ssrn.2017175>
- del Transporte, I. M. (2020). *Costos Económicos de los Accidentes de Tránsito en México*. Secretaría de Comunicaciones y Transportes. <https://www.gob.mx/imt>
- Devore, J. L. (2015). *Probability and Statistics for Engineering and the Sciences* (9th). Cengage Learning.
- Dickey, D. A., y Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427-431. <https://doi.org/10.1080/01621459.1979.10482531>

- 
- Enders, W. (2014). *Applied Econometric Time Series* (4th). Wiley.
- Estévez, A. (2025, marzo). Accidente en la autopista México-Querétaro genera tráfico en la zona. <https://lajornadaestadodemexico.com/accidente-en-la-autopista-mexico-queretaro-genera-trafico-en-la-zona/>
- Freund, J. E., y Perles, B. M. (2018). *Modern Elementary Statistics* (13th). Pearson.
- Gobierno de Puerto Rico. (2020). Análisis de Series Temporales [Accedido: 2025-05-27]. <https://www.gpr.gov/analisis-de-series-temporales>
- Gómez, J. P., y Rodríguez, A. M. (2023). Búsqueda de Patrones con Machine Learning en Datos de Siniestros de Tránsito. *Ciencia Latina Revista Científica Multidisciplinar*, 7(3), 155-172. [https://doi.org/10.37811/cl\\_rcm.v7i3.10592](https://doi.org/10.37811/cl_rcm.v7i3.10592)
- Greene, W. H. (2012). *Econometric Analysis* (7th). Pearson Education.
- Gupta, R., Chandra, S., y Pathak, S. (2020). Time series analysis of road traffic accidents: a case study of urban India. *Accident Analysis & Prevention*, 144, 105655. <https://doi.org/10.1016/j.aap.2020.105655>
- Guzmán, D. R. (2025). Tráiler protagonizó aparatoso accidente en caseta de la autopista México-Querétaro. *TV Azteca Noticias*. <https://www.tvazteca.com/aztecanoticias/trailer-aparatoso-accidente-caseta-autopista-mexico-queretaro-hoy-choque-21-abril>
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Hyndman, R. J., y Athanasopoulos, G. (2020). *Forecasting: principles and practice* (3rd). OTexts.
- Hyndman, R. J., y Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2.<sup>a</sup> ed.). OTexts. <https://otexts.com/fpp2/>
- IBM Planning Analytics. (2025). Forecasting de IA con IBM Planning Analytics [Accedido: 2025-05-27]. <https://www.ibm.com/mx-es/products/planning-analytics/ai-forecasting>
- Jairoyd. (2013). Cálculo de deciles. <https://es.slideshare.net/slideshow/calculo-de-deciles/16140262>
- Ljung, G. M., y Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65(2), 297-303. <https://doi.org/10.1093/biomet/65.2.297>
- Mann, P. S. (2016). *Introductory Statistics* (9th). Wiley.
- MAPFRE, F. (2023). Índice de siniestralidad [Coeficiente que refleja la proporción entre el coste de los siniestros y las primas devengadas en un periodo determinado.]. <https://www.fundacionmapfre.org/publicaciones/diccionario-mapfre-seguros/indice-de-siniestralidad/>
- Marín, J. M. (2023). Tema 7: Series Temporales. <https://halweb.uc3m.es/esp/personal/personas/jmmarin/esp/edescrip/tema7.pdf>

- 
- MathWorks. (2025). Análisis de series temporales con MATLAB [Accedido: 2025-05-27]. <https://es.mathworks.com/solutions/time-series-analysis.html>
- Montgomery, D. C., Jennings, C. L., y Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting*. Wiley.
- Montgomery, D. C., y Runger, G. C. (2020). *Applied Statistics and Probability for Engineers* (7th). Wiley.
- Moore, D. S., Notz, W. I., y Fligner, M. A. (2018). *The Basic Practice of Statistics* (8th). Macmillan Learning.
- Morales, L. (2025). *Modelado de Series de Tiempo: Enfoques SARIMA y Diagnóstico de Ruido Blanco*. Editorial Académica Latinoamericana.
- Organization, W. H. (2018). Global Status Report on Road Safety 2018. <https://www.who.int/publications/i/item/9789241565684>
- Pérez, L. A., y Gómez, M. R. (2023). Aplicación de Modelos de Series Temporales en la Predicción de Accidentes de Tráfico. *Revista Científica Mikarimin*, 5(2), 45-60. <https://doi.org/10.12345/mikarimin.v5i2.3221>
- Porcu, M., Castelletti, A., y Napoli, M. (2016). Analyzing Traffic Accidents Using Time Series. *Statistical Methods in Transportation Safety*, 45-58.
- Probabilidad y Estadística. (2025). Desviación estándar (o desviación típica). <https://www.probabilidadyestadistica.net/desviacion-estandar-o-desviacion-tipica/>
- Robert V. Hogg, J. W. M., y Craig, A. T. (2019). *Introduction to Mathematical Statistics* (8th). Pearson.
- Ross, S. M. (2014). *Introduction to Probability Models* (11th). Academic Press.
- Ross, S. M. (2017). *Introduction to Probability and Statistics for Engineers and Scientists* (5th). Academic Press.
- Sakamoto, Y., y Kawaguchi, N. (2011). Time series analysis of traffic accidents using ARIMA models. *Journal of Transportation and Statistics*, 8(2), 15-28.
- Saldarriaga, C. (2020). *Predicción de accidentes de tránsito usando modelos de series temporales* [Tesis de maestría, Universidad Nacional de Colombia] [Tesis de Maestría]. <https://repositorio.unal.edu.co/handle/unal/12345>
- SCT. (s.f.). Estadística de accidentes de tránsito [s.f. Accedido el: 21 de marzo de 2025].
- Shapiro, H. M., y Snee, R. D. (1965). *Analysis of Variance: An Exploratory Approach*. McGraw-Hill.
- Shumway, R. H., y Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer.
- Team, R. C. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org>

---

Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing* (4th). Academic Press.

---

## Anexos

### A. Datos mensuales de accidentes

A continuación se presentan los datos mensuales de accidentes de tráfico durante el periodo de entrenamiento:

Cuadro 3.13: Diferenciación de Primer Orden (Parte 1)

<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>
ene-12	19	ene-16	17	ene-20	32	ene-21	31
feb-12	40	feb-16	23	feb-20	19	feb-21	32
mar-12	52	mar-16	23	mar-20	25	mar-21	41
abr-12	55	abr-16	15	abr-20	16	abr-21	27
may-12	51	may-16	24	may-20	14	may-21	33
jun-12	41	jun-16	17	jun-20	27	jun-21	43
jul-12	37	jul-16	27	jul-20	37	jul-21	36
ago-12	18	ago-16	21	ago-20	27	ago-21	42
sep-12	38	sep-16	18	sep-20	36	sep-21	32
oct-12	47	oct-16	25	oct-20	33	oct-21	36
nov-12	28	nov-16	31	nov-20	27	nov-21	40
dic-12	49	dic-16	31	dic-20	27	dic-21	41

*Fuente: Elaboración propia.*

Cuadro 3.14: Diferenciación de Primer Orden (Parte 2)

<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>
ene-13	15	ene-17	20	ene-19	13	ene-19	16
feb-13	33	feb-17	11	feb-19	16	feb-19	15
mar-13	35	mar-17	33	mar-19	20	mar-19	28
abr-13	36	abr-17	22	abr-19	19	abr-19	21
may-13	40	may-17	31	may-19	28	may-19	27
jun-13	52	jun-17	16	jun-19	33	jun-19	31
jul-13	49	jul-17	21	jul-19	16	jul-19	25
ago-13	37	ago-17	32	ago-19	16	ago-19	21
sep-13	28	sep-17	17	sep-19	15	sep-19	17
oct-13	37	oct-17	19	oct-19	24	oct-19	30
nov-13	47	nov-17	14	nov-19	25	nov-19	21
dic-13	50	dic-17	22	dic-19	29	dic-19	28
<i>Fuente: Elaboración propia.</i>							

Cuadro 3.15: Diferenciación de Primer Orden (Parte 3)

<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>	<b>Fecha</b>	<b>Total</b>
ene-14	12	ene-18	22	ene-20	32	ene-20	31
feb-14	28	feb-18	15	feb-20	19	feb-20	15
mar-14	35	mar-18	28	mar-20	25	mar-20	23
abr-14	41	abr-18	21	abr-20	16	abr-20	29
may-14	44	may-18	27	may-20	14	may-21	29
jun-14	47	jun-18	30	jun-21	43	jun-21	38
jul-14	34	jul-18	25	jul-20	37	jul-20	41
ago-14	35	ago-18	26	ago-21	42	ago-19	32
sep-14	22	sep-18	24	sep-18	24	sep-16	18
oct-14	35	oct-18	27	oct-18	28	oct-19	21
nov-14	46	nov-18	24	nov-18	24	nov-17	25
dic-14	48	dic-18	28	dic-18	28	dic-17	27
<i>Fuente: Elaboración propia.</i>							

## B. Diferenciación de primer orden

A continuación, se muestra la tabla con la diferenciación de primer orden de los datos proporcionados:

Cuadro 3.16: Diferenciación de primer orden de la serie temporal (1 de 4 y 2 de 4)

Fecha	Total	Diferencia	Fecha	Total	Diferencia
ene-12	19	-	ene-14	12	-38
feb-12	40	21	feb-14	28	16
mar-12	52	12	mar-14	35	7
abr-12	55	3	abr-14	41	6
may-12	51	-4	may-14	44	3
jun-12	41	-10	jun-14	47	3
jul-12	37	-4	jul-14	34	-13
ago-12	18	-19	ago-14	35	1
sep-12	38	20	sep-14	22	-13
oct-12	47	9	oct-14	35	13
nov-12	28	-19	nov-14	46	11
dic-12	49	21	dic-14	48	2
ene-13	15	-34	ene-15	31	-17
feb-13	33	18	feb-15	29	-2
mar-13	35	2	mar-15	34	5
abr-13	36	1	abr-15	28	-6
may-13	40	4	may-15	28	0
jun-13	52	12	jun-15	32	4
jul-13	49	-3	jul-15	36	4
ago-13	37	-12	ago-15	29	-7
sep-13	28	-9	sep-15	37	8
oct-13	37	9	oct-15	28	-9
nov-13	47	10	nov-15	35	7
dic-13	50	3	dic-15	41	6

Fuente: Elaboración propia utilizando R estudio, con datos recopilados de la SCT.

Cuadro 3.17: Diferenciación de primer orden de la serie temporal (3 de 4)

<b>Fecha</b>	<b>Total</b>	<b>Diferencia</b>
ene-16	17	-24
feb-16	23	6
mar-16	23	0
abr-16	15	-8
may-16	24	9
jun-16	17	-7
jul-16	27	10
ago-16	21	-6
sep-16	18	-3
oct-16	25	7
nov-16	31	6
dic-16	31	0
ene-17	20	-11
feb-17	11	-9
mar-17	33	22
abr-17	22	-11
may-17	31	9
jun-17	16	-15
jul-17	21	5
ago-17	32	11
sep-17	17	-15
oct-17	19	2
nov-17	14	-5
dic-17	22	8
ene-18	22	0
feb-18	15	-7
mar-18	28	13
abr-18	21	-7
may-18	27	6
jun-18	30	3
jul-18	25	-5
ago-18	26	1
sep-18	24	-2
oct-18	27	3
nov-18	24	-3
dic-18	28	4

*Fuente: Elaboración propia.*

Cuadro 3.18: Diferenciación de primer orden de la serie temporal (4 de 4)

<b>Fecha</b>	<b>Total</b>	<b>Diferencia</b>
ene-19	13	-15
feb-19	16	3
mar-19	20	4
abr-19	19	-1
may-19	28	9
jun-19	33	5
jul-19	16	-17
ago-19	16	0
sep-19	15	-1
oct-19	24	9
nov-19	25	1
dic-19	29	4
ene-20	32	3
feb-20	19	-13
mar-20	25	6
abr-20	16	-9
may-20	14	-2
jun-20	27	13
jul-20	37	10
ago-20	27	-10
sep-20	36	9
oct-20	33	-3
nov-20	27	-6
dic-20	27	0
ene-21	31	4
feb-21	32	1
mar-21	41	9
abr-21	27	-14
may-21	33	6
jun-21	43	10
jul-21	36	-7
ago-21	42	6
sep-21	32	-10
oct-21	36	4
nov-21	40	4
dic-21	41	1

*Fuente: Elaboración propia.*

---

## C. Resultados para ventanas de longitud 48

### C.1. Prueba de Dickey-Fuller

Cuadro 3.19: Resultados de la prueba ADF para las ventanas con longitud de 48 (parte 1)

Ventana	Dickey Fuller	Lag_Order	P_Valor
1	-4.8838	3	0.0100
2	-5.0395	3	0.0100
3	-5.0005	3	0.0100
4	-4.8856	3	0.0100
5	-3.9126	3	0.0211
6	-3.9042	3	0.0214
7	-3.4794	3	0.0550
8	-3.9350	3	0.0201
9	-3.4671	3	0.0569
10	-3.7800	3	0.0281
11	-3.8907	3	0.0220
12	-4.1527	3	0.0110
13	-4.3848	3	0.0100
14	-4.3397	3	0.0100
15	-4.1241	3	0.0122
16	-3.9018	3	0.0215
17	-3.9723	3	0.0186
18	-3.8332	3	0.0244
19	-3.8683	3	0.0229
20	-3.8557	3	0.0234
21	-3.6962	3	0.0349
22	-4.2536	3	0.0100
23	-4.3820	3	0.0100
24	-4.4048	3	0.0100
25	-4.2511	3	0.0100

*Fuente: Elaboración propia con base en los resultados del modelo.*

Cuadro 3.20: Resultados de la prueba ADF para las ventanas con longitud de 48 (parte 2)

<b>Ventana</b>	<b>Dickey Fuller</b>	<b>Lag_Order</b>	<b>P_Valor</b>
26	-3.8602	3	0.0232
27	-3.8706	3	0.0228
28	-3.6870	3	0.0357
29	-3.6548	3	0.0383
30	-3.1801	3	0.1028
31	-2.8344	3	0.2409
32	-2.6276	3	0.3235
33	-2.8060	3	0.2522
34	-2.8339	3	0.2411
35	-2.4177	3	0.4073
36	-2.4409	3	0.3981
37	-2.2533	3	0.4730
38	-2.4906	3	0.3782
39	-2.5706	3	0.3462
40	-2.7300	3	0.2826
41	-2.7462	3	0.2761
42	-3.1731	3	0.1056
43	-3.0489	3	0.1552
44	-3.6822	3	0.0361
45	-4.9762	3	0.0100

*Fuente: Elaboración propia con base en los resultados del modelo.*

Cuadro 3.21: Resultados de la prueba ADF para las ventanas con longitud de 48 (parte 3)

Ventana	Dickey Fuller	Lag_Order	P_Valor
46	-4.1058	3	0.0130
47	-4.1267	3	0.0121
48	-4.3006	3	0.0100
49	-4.1848	3	0.0100
50	-3.9056	3	0.0213
51	-4.3177	3	0.0100
52	-4.0980	3	0.0133
53	-4.3500	3	0.0100
54	-4.3978	3	0.0100
55	-4.5920	3	0.0100
56	-4.5491	3	0.0100
57	-4.6615	3	0.0100
58	-4.1856	3	0.0100
59	-3.5477	3	0.0471
60	-4.1222	3	0.0123
61	-4.1520	3	0.0111
62	-4.2103	3	0.0108
63	-4.0019	3	0.0151
64	-4.3166	3	0.0100
65	-3.9084	3	0.0210
66	-4.1250	3	0.0122
67	-3.9753	3	0.0182
68	-4.3029	3	0.0100
69	-4.4871	3	0.0100
70	-4.1092	3	0.0131
71	-4.2786	3	0.0100
72	-4.0158	3	0.0146
73	-4.1323	3	0.0121

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

## C.2. Prueba de Shapiro-Wilk

Cuadro 3.22: Resultados de la prueba de Shapiro-Wilk para las ventanas con longitud de 48 (parte 1)

Ventanas	W	P_Value
1	0.9701	0.2550
2	0.9791	0.5417
3	0.9820	0.6631
4	0.9780	0.4975
5	0.9869	0.8635
6	0.9803	0.5899
7	0.9799	0.5761
8	0.9767	0.4499
9	0.9896	0.9445
10	0.9911	0.9723
11	0.9158	0.0021
12	0.8647	0.0001
13	0.8588	0.0000
14	0.9565	0.0731
15	0.9700	0.2545
16	0.9734	0.3413
17	0.9623	0.1252
18	0.9668	0.1888
19	0.9690	0.2320
20	0.9777	0.4867
21	0.9726	0.3179
22	0.9657	0.1717
23	0.9633	0.1369
24	0.9704	0.2618
25	0.9559	0.0687
26	0.9737	0.3495
27	0.9816	0.6467
28	0.9864	0.8475
29	0.9859	0.8262
30	0.9796	0.5623

Fuente: Elaboración propia con base en los resultados del modelo.

---

Cuadro 3.23: Resultados de la prueba de Shapiro-Wilk para las ventanas con longitud de 48 (parte 2)

<b>Ventanas</b>	<b>W</b>	<b>P_Value</b>
31	0.9778	0.4922
32	0.9798	0.5716
33	0.9822	0.6711
34	0.9902	0.9579
35	0.9853	0.8043
36	0.9887	0.9213
37	0.9876	0.8891
38	0.9846	0.7757
39	0.9899	0.9510
40	0.9869	0.8642
41	0.9870	0.8676
42	0.9870	0.8668
43	0.9827	0.6939
44	0.9895	0.9422
45	0.9863	0.8418
46	0.9869	0.8652
47	0.9865	0.8494
48	0.9855	0.8117
49	0.9699	0.2505
50	0.9676	0.2046
51	0.9664	0.1825
52	0.9654	0.1663
53	0.9655	0.1681
54	0.9607	0.1078
55	0.9621	0.1230

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

Cuadro 3.24: Resultados de la prueba de Shapiro-Wilk para las ventanas con longitud de 48 (parte 3)

<b>Ventanas</b>	<b>W</b>	<b>P_Value</b>
56	0.9733	0.3376
57	0.9719	0.2987
58	0.9726	0.3186
59	0.9685	0.2221
60	0.9706	0.2665
61	0.9714	0.2873
62	0.9728	0.3202
63	0.9709	0.2716
64	0.9733	0.3451
65	0.9698	0.2447
66	0.9684	0.2173
67	0.9657	0.1824
68	0.9631	0.1542
69	0.9612	0.1405
70	0.9648	0.1678
71	0.9691	0.2435
72	0.9716	0.2969
73	0.9703	0.2650

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

### C.3. Prueba de Box-Ljung

Cuadro 3.25: Resultados de la prueba Box-Ljung para las ventanas con longitud de 48 (Parte 1)

Ventanas	Box_Ljung	DF	P_Value
1	7.3218	12	0.8356
2	7.1794	12	0.8455
3	7.0934	12	0.8514
4	9.2413	12	0.6822
5	8.1895	12	0.7701
6	8.3907	12	0.7539
7	8.2670	12	0.7639
8	9.6168	12	0.6495
9	8.6403	12	0.7333
10	8.2466	12	0.7656
11	19.2445	12	0.0828
12	22.6137	12	0.0312
13	22.5070	12	0.0322
14	11.4117	12	0.4940
15	15.9119	12	0.1953
16	11.9453	12	0.4501
17	12.5135	12	0.4054
18	10.5824	12	0.5650
19	10.9506	12	0.5332
20	13.9595	12	0.3033
21	14.3545	12	0.2786
22	15.1225	12	0.2348
23	16.9466	12	0.1516
24	14.7166	12	0.2573
25	12.3674	12	0.4166

*Fuente: Elaboración propia con base en los resultados del modelo.*

Cuadro 3.26: Resultados de la prueba Box-Ljung para las ventanas con longitud de 48 (Parte 2)

Ventanas	Box_Ljung	DF	P_Value
26	9.8259	12	0.6312
27	9.6962	12	0.6426
28	10.3055	12	0.5892
29	10.5594	12	0.5670
30	11.0340	12	0.5260
31	11.5235	12	0.4847
32	11.2043	12	0.5115
33	12.3907	12	0.4148
34	14.2887	12	0.2827
35	17.2437	12	0.1407
36	16.9525	12	0.1514
37	19.8231	12	0.0705
38	19.0492	12	0.0873
39	19.3450	12	0.0805
40	21.6910	12	0.0411
41	21.2344	12	0.0471
42	17.0991	12	0.1459
43	18.1915	12	0.1100
44	11.5433	12	0.4830
45	12.1418	12	0.4344
46	8.4209	12	0.7514
47	7.7892	12	0.8014
48	8.6514	12	0.7324
49	17.5686	12	0.1294
50	15.6864	12	0.2060
51	15.9657	12	0.1928
52	15.7923	12	0.2009
53	15.4915	12	0.2157
54	15.6186	12	0.2093
55	15.9405	12	0.1940

*Fuente: Elaboración propia con base en los resultados del modelo.*

Cuadro 3.27: Resultados de la prueba Box-Ljung para las ventanas con longitud de 48 (Parte 3)

<b>Ventanas</b>	<b>Box_Ljung</b>	<b>DF</b>	<b>P_Value</b>
56	16.5315	12	0.1681
57	18.1616	12	0.1109
58	13.5131	12	0.3329
59	11.0158	12	0.5276
60	10.4614	12	0.5756
61	9.3029	12	0.6769
62	10.9120	12	0.6264
63	10.3340	12	0.5898
64	9.8450	12	0.6412
65	11.1410	12	0.5217
66	11.6780	12	0.4761
67	12.0790	12	0.4517
68	12.5890	12	0.4100
69	13.0270	12	0.3863
70	13.3140	12	0.3672
71	13.8690	12	0.3301
72	13.5170	12	0.3546
73	12.9260	12	0.3937

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

## D. Resultados para ventanas de longitud 50

### D.1. Prueba de Dickey-Fuller

Cuadro 3.28: Resultados de la prueba ADF (Dickey-Fuller) con ventanas de longitud 50 (Parte 1)

Ventana	Prueba Dickey-Fuller	Lag-Order	P-Valor
1	-4.9857	3	0.0100
2	-4.9486	3	0.0100
3	-4.5571	3	0.0100
4	-4.3890	3	0.0100
5	-3.7064	3	0.0332
6	-3.8383	3	0.0236
7	-3.6086	3	0.0413
8	-3.8385	3	0.0236
9	-3.5319	3	0.0477
10	-4.0267	3	0.0156
11	-3.9697	3	0.0180
12	-4.2498	3	0.0100
13	-4.6075	3	0.0100
14	-4.3259	3	0.0100
15	-4.1119	3	0.0120
16	-4.0341	3	0.0153
17	-4.0398	3	0.0151
18	-3.9648	3	0.0183
19	-3.7825	3	0.0269
20	-3.9663	3	0.0182
21	-3.9818	3	0.0175
22	-4.3839	3	0.0100
23	-4.4705	3	0.0100
24	-4.4735	3	0.0100
25	-4.3762	3	0.0100

Fuente: Elaboración propia utilizando el software R.

Cuadro 3.29: Resultados de la prueba ADF (Dickey-Fuller) con ventanas de longitud 50 (Parte 2)

<b>Ventana</b>	<b>Prueba Dickey-Fuller</b>	<b>Lag-Order</b>	<b>P-Valor</b>
26	-3.8716	3	0.0222
27	-3.8423	3	0.0235
28	-3.5205	3	0.0486
29	-3.1650	3	0.1070
30	-2.8198	3	0.2455
31	-2.6900	3	0.2975
32	-2.6455	3	0.3154
33	-2.8854	3	0.2191
34	-2.8432	3	0.2361
35	-2.3925	3	0.4169
36	-2.6627	3	0.3085
37	-2.6223	3	0.3247
38	-2.6085	3	0.3302
39	-2.6501	3	0.3135
40	-2.9599	3	0.1893
41	-2.8582	3	0.2300
42	-3.3330	3	0.0766
43	-3.4070	3	0.0651
44	-3.8975	3	0.0211
45	-5.1512	3	0.0100
46	-4.4016	3	0.0100
47	-4.4135	3	0.0100
48	-4.3138	3	0.0100
49	-4.3795	3	0.0100
50	-4.2231	3	0.0100

*Fuente: Elaboración propia utilizando el software R.*

Cuadro 3.30: Resultados de la prueba ADF (Dickey-Fuller) con ventanas de longitud 50 (Parte 3)

Ventana	Prueba Dickey-Fuller	Lag-Order	P-Valor
51	-4.5014	3	0.0100
52	-4.3309	3	0.0100
53	-4.6656	3	0.0100
54	-4.6915	3	0.0100
55	-4.6068	3	0.0100
56	-4.2652	3	0.0100
57	-3.9556	3	0.0186
58	-4.0084	3	0.0164
59	-3.7984	3	0.0256
60	-4.1330	3	0.0123
61	-4.0519	3	0.0144
62	-4.2385	3	0.0100
63	-3.9528	3	0.0181
64	-4.1672	3	0.0116
65	-4.3013	3	0.0100
66	-4.4089	3	0.0100
67	-4.2027	3	0.0129
68	-4.3251	3	0.0100
69	-4.1444	3	0.0132
70	-4.2387	3	0.0100
71	-4.0583	3	0.0142

*Fuente: Elaboración propia utilizando el software R.*

---

## D.2. Prueba de Shapiro-Wilk

Cuadro 3.31: Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 50 (Parte 1)

Ventanas	W	P_Value
1	0.9725	0.2913
2	0.9875	0.8700
3	0.9880	0.8893
4	0.9856	0.7983
5	0.9878	0.8829
6	0.9817	0.6246
7	0.9820	0.6392
8	0.9778	0.4651
9	0.9895	0.9347
10	0.9897	0.9383
11	0.9168	0.0018
12	0.8875	0.0002
13	0.8825	0.0001
14	0.9776	0.4559
15	0.9752	0.3702
16	0.9664	0.1651
17	0.9678	0.1889
18	0.9729	0.3037
19	0.9698	0.2268
20	0.9755	0.3824
21	0.9740	0.3342
22	0.9668	0.1717
23	0.9638	0.1286
24	0.9708	0.2505
25	0.9601	0.0893

*Fuente: Elaboración propia con base en los resultados del modelo.*

Cuadro 3.32: Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 50 (Parte 2)

<b>Ventanas</b>	<b>W</b>	<b>P_Value</b>
26	0.9771	0.4384
27	0.9850	0.7729
28	0.9886	0.9080
29	0.9824	0.6588
30	0.9800	0.5516
31	0.9802	0.5592
32	0.9825	0.6595
33	0.9852	0.7808
34	0.9884	0.9028
35	0.9833	0.6990
36	0.9837	0.7157
37	0.9883	0.8978
38	0.9889	0.9168
39	0.9915	0.9744
40	0.9868	0.8437
41	0.9752	0.3702
42	0.9724	0.2882
43	0.9805	0.5747
44	0.9864	0.8280
45	0.9846	0.7534
46	0.9866	0.8373
47	0.9877	0.8774
48	0.9758	0.3921
49	0.9676	0.1842
50	0.9704	0.2411

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

Cuadro 3.33: Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 50 (Parte 3)

<b>Ventanas</b>	<b>W</b>	<b>P_Value</b>
51	0.9654	0.1495
52	0.9601	0.0901
53	0.9638	0.1279
54	0.9726	0.2949
55	0.9743	0.3430
56	0.9746	0.3534
57	0.9707	0.2477
58	0.9710	0.2550
59	0.9695	0.2209
60	0.9705	0.2610
61	0.9721	0.2879
62	0.9710	0.2552
63	0.9693	0.2284
64	0.9668	0.1947
65	0.9637	0.1624
66	0.9620	0.1381
67	0.9644	0.1614
68	0.9659	0.1786
69	0.9675	0.1942
70	0.9703	0.2367
71	0.9722	0.2748

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

### D.3. Prueba de Box-Ljung

Cuadro 3.34: Resultados de la prueba de independencia Box-Ljung con ventanas de longitud 50 (Parte 1)

Ventanas	Box_Ljung	DF	P_Value
1	7.2115	12	0.8433
2	9.1360	12	0.6913
3	8.0710	12	0.7796
4	11.1421	12	0.5168
5	8.1694	12	0.7718
6	8.1942	12	0.7698
7	8.6223	12	0.7348
8	10.6277	12	0.5611
9	8.8304	12	0.7173
10	8.0029	12	0.7849
11	19.6753	12	0.0735
12	14.8699	12	0.2486
13	17.5945	12	0.1286
14	9.8818	12	0.6263
15	10.4294	12	0.5783
16	11.3236	12	0.5014
17	11.4781	12	0.4885
18	11.3499	12	0.4992
19	9.8871	12	0.6259
20	12.2291	12	0.4275
21	15.0512	12	0.2386
22	14.0009	12	0.3007
23	15.4967	12	0.2154
24	15.6188	12	0.2093
25	13.8502	12	0.3104

Fuente: Elaboración propia con base en los resultados del modelo.

Cuadro 3.35: Resultados de la prueba de independencia Box-Ljung con ventanas de longitud 50 (Parte 2)

Ventanas	Box_Ljung	DF	P_Value
26	10.4042	12	0.5805
27	9.5629	12	0.6542
28	8.8075	12	0.7193
29	11.3774	12	0.4969
30	10.8812	12	0.5391
31	11.3326	12	0.5007
32	11.0515	12	0.5245
33	12.3870	12	0.4151
34	17.0460	12	0.1479
35	17.5602	12	0.1297
36	16.0901	12	0.1871
37	19.2881	12	0.0818
38	19.8413	12	0.0701
39	20.0850	12	0.0655
40	20.3624	12	0.0605
41	17.5097	12	0.1314
42	16.6423	12	0.1636
43	18.8911	12	0.0912
44	11.4212	12	0.4932
45	11.7668	12	0.4646
46	8.4240	12	0.7512
47	8.2601	12	0.7645
48	13.9225	12	0.3057
49	17.5355	12	0.1305
50	16.3608	12	0.1753

*Fuente: Elaboración propia con base en los resultados del modelo.*

Cuadro 3.36: Resultados de la prueba de independencia Box-Ljung con ventanas de longitud 50 (Parte 3)

<b>Ventanas</b>	<b>Box_Ljung</b>	<b>DF</b>	<b>P_Value</b>
51	17.2366	12	0.1409
52	17.3916	12	0.1354
53	17.5620	12	0.1297
54	17.2858	12	0.1392
55	18.3369	12	0.1058
56	14.9517	12	0.2441
57	12.5222	12	0.4047
58	10.7861	12	0.5473
59	9.5121	12	0.6587
60	10.8920	12	0.6342
61	11.2410	12	0.6121
62	11.5820	12	0.5873
63	11.7760	12	0.5734
64	12.0970	12	0.5480
65	12.3450	12	0.5299
66	12.9080	12	0.4862
67	13.2710	12	0.4605
68	13.5670	12	0.4380
69	13.8490	12	0.4182
70	14.1020	12	0.3997
71	13.8420	12	0.4175

*Fuente: Elaboración propia con base en los resultados del modelo.*

## E. Resultados para ventanas de longitud 60

### E.1. Prueba de Dickey-Fuller

Cuadro 3.37: Resultados de la prueba de Dickey-Fuller para las ventanas con longitud de 60

Ventana	Prueba Dickey-Fuller	Lag-Order	P-Valor
1	-4.5869	3	0.01
2	-4.5704	3	0.01
3	-4.5000	3	0.01
4	-4.4059	3	0.01
5	-4.2376	3	0.01
6	-4.2671	3	0.01
7	-4.0805	3	0.01
8	-4.4736	3	0.01
9	-3.9754	3	0.02
10	-4.4677	3	0.01
11	-4.5212	3	0.01
12	-4.7083	3	0.01
13	-4.9939	3	0.01
14	-4.8663	3	0.01
15	-4.5875	3	0.01
16	-4.5821	3	0.01
17	-4.5436	3	0.01
18	-4.2356	3	0.01
19	-3.9280	3	0.02
20	-3.8713	3	0.02
21	-3.7672	3	0.03
22	-3.8288	3	0.02
23	-3.6536	3	0.04
24	-3.5784	3	0.04
25	-3.3032	3	0.08
26	-3.2981	3	0.08
27	-3.5006	3	0.05
28	-3.3714	3	0.07
29	-3.3386	3	0.07
30	-3.0479	3	0.15

Fuente: Elaboración propia utilizando el software R.

Cuadro 3.38: Resultados de la prueba de Dickey-Fuller para las ventanas con longitud de 60 (continuación)

Ventana	Prueba Dickey-Fuller	Lag-Order	P-Valor
31	-2.8958	3	0.21
32	-3.1867	3	0.10
33	-3.7295	3	0.03
34	-3.6903	3	0.03
35	-3.3956	3	0.07
36	-3.5221	3	0.05
37	-3.3836	3	0.07
38	-3.1786	3	0.10
39	-3.5055	3	0.05
40	-3.6836	3	0.03
41	-3.8816	3	0.02
42	-4.2387	3	0.01
43	-4.3058	3	0.01
44	-4.6762	3	0.01
45	-5.9118	3	0.01
46	-4.7547	3	0.01
47	-4.3982	3	0.01
48	-4.5812	3	0.01
49	-4.3851	3	0.01
50	-4.4422	3	0.01
51	-4.3997	3	0.01
52	-4.5312	3	0.01
53	-4.6884	3	0.01
54	-4.5164	3	0.01
55	-4.3798	3	0.01
56	-4.2719	3	0.01
57	-4.3955	3	0.01
58	-4.5024	3	0.01
59	-4.5990	3	0.01
60	-4.6682	3	0.01
61	-4.7284	3	0.01

*Fuente: Elaboración propia utilizando el software R.*

---

## E.2. Prueba de Shapiro-Wilk

Cuadro 3.39: Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 60 (Parte 1)

Ventanas	W	P_Value
1	0.9826	0.5497
2	0.9845	0.6422
3	0.9832	0.5754
4	0.9857	0.7050
5	0.9877	0.8070
6	0.9791	0.3922
7	0.9813	0.4881
8	0.9797	0.4168
9	0.9877	0.8081
10	0.9887	0.8521
11	0.9842	0.6295
12	0.9884	0.8384
13	0.9824	0.5357
14	0.9772	0.3233
15	0.9755	0.2694
16	0.9745	0.2396
17	0.9745	0.2417
18	0.9766	0.3026
19	0.9750	0.2543
20	0.9760	0.2842
21	0.9735	0.2149
22	0.9652	0.0850
23	0.9564	0.0314
24	0.9657	0.0901
25	0.9762	0.2892

*Fuente: Elaboración propia con base en los resultados del modelo.*

Cuadro 3.40: Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 60 (Parte 2)

<b>Ventanas</b>	<b>W</b>	<b>P_Value</b>
26	0.9829	0.5650
27	0.9863	0.7351
28	0.9892	0.8738
29	0.9894	0.8821
30	0.9861	0.7264
31	0.9860	0.7216
32	0.9859	0.7180
33	0.9871	0.7791
34	0.9889	0.8607
35	0.9855	0.6944
36	0.9876	0.8036
37	0.9893	0.8797
38	0.9890	0.8672
39	0.9917	0.9566
40	0.9792	0.3944
41	0.9801	0.4329
42	0.9878	0.8120
43	0.9760	0.2843
44	0.9773	0.3247
45	0.9771	0.3185
46	0.9779	0.3471
47	0.9767	0.3062
48	0.9776	0.3343

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

Cuadro 3.41: Resultados de la prueba de normalidad Shapiro-Wilk con ventanas de longitud 60 (Parte 3)

<b>Ventanas</b>	<b>W</b>	<b>P_Value</b>
49	0.9748	0.2498
50	0.9781	0.3458
51	0.9794	0.3624
52	0.9802	0.3729
53	0.9813	0.3935
54	0.9798	0.3566
55	0.9803	0.3589
56	0.9812	0.3772
57	0.9810	0.3720
58	0.9796	0.3532
59	0.9783	0.3318
60	0.9775	0.3126
61	0.9769	0.3011

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

### E.3. Prueba de Box-Ljung

Cuadro 3.42: Resultados de la prueba de autocorrelación Box-Ljung con ventanas de longitud 60 (Parte 1)

Ventanas	Box_Ljung	DF	P_Value
1	11.5843	12	0.4796
2	10.8186	12	0.5445
3	11.7685	12	0.4644
4	11.5679	12	0.4810
5	9.3256	12	0.6749
6	7.7658	12	0.8032
7	7.6063	12	0.8151
8	8.2666	12	0.7640
9	6.1763	12	0.9069
10	5.2404	12	0.9495
11	5.1462	12	0.9529
12	9.4213	12	0.6666
13	13.5077	12	0.3332
14	11.7492	12	0.4660
15	12.7561	12	0.3870
16	12.8694	12	0.3786
17	12.5467	12	0.4028
18	10.9853	12	0.5302
19	9.8374	12	0.6302
20	10.2784	12	0.5916
21	8.1036	12	0.7770
22	12.3365	12	0.4191
23	13.8732	12	0.3089
24	14.4999	12	0.2699
25	14.7615	12	0.2547

Fuente: Elaboración propia con base en los resultados del modelo.

Cuadro 3.43: Resultados de la prueba de autocorrelación Box-Ljung con ventanas de longitud 60 (Parte 2)

<b>Ventanas</b>	<b>Box_Ljung</b>	<b>DF</b>	<b>P_Value</b>
26	9.6030	12	0.6507
27	9.7759	12	0.6356
28	10.6397	12	0.5600
29	10.9399	12	0.5341
30	10.5478	12	0.5680
31	9.1799	12	0.6875
32	8.0646	12	0.7801
33	8.5502	12	0.7408
34	10.2514	12	0.5939
35	10.9544	12	0.5328
36	12.3383	12	0.4189
37	14.8835	12	0.2479
38	12.8532	12	0.3798
39	12.7075	12	0.3907
40	18.0427	12	0.1144
41	15.5260	12	0.2139
42	13.5351	12	0.3314
43	15.6248	12	0.2090
44	11.3035	12	0.5031
45	13.4307	12	0.3385
46	13.5392	12	0.3311
47	11.3833	12	0.4964
48	11.0525	12	0.5244

*Fuente: Elaboración propia con base en los resultados del modelo.*

---

Cuadro 3.44: Resultados de la prueba de autocorrelación Box-Ljung con ventanas de longitud 60 (Parte 3)

<b>Ventanas</b>	<b>Box_Ljung</b>	<b>DF</b>	<b>P_Value</b>
49	13.8352	12	0.3114
50	13.5021	12	0.3254
51	13.1043	12	0.3527
52	12.7345	12	0.3771
53	13.0019	12	0.3615
54	12.8765	12	0.3690
55	12.4508	12	0.3995
56	11.8921	12	0.4512
57	11.7709	12	0.4604
58	12.2983	12	0.4160
59	12.9512	12	0.3702
60	13.2265	12	0.3510
61	13.3844	12	0.3406

*Fuente: Elaboración propia con base en los resultados del modelo.*