

MODELOS DE SERIES TEMPORALES APLICADOS AL NÚMERO MENSUAL DE ACCIDENTES DE TRÁNSITO EN EL PARAGUAY.

MODELS OF TEMPORARY SERIES APPLIED TO THE MONTHLY NUMBER OF TRANSIT ACCIDENTS IN PARAGUAY.

Autor: MARÍA PABLÍN ACOSTA DE ENCISO

Institución: Facultad de Ciencias Exactas y Naturales

Dirección de Correo: mapablinacosta@gmail.com

RESUMEN

Los accidentes de tránsito constituyen una problemática social compleja en todo el mundo. La Organización Mundial de la Salud (OMS) designa los peligros del tráfico vial como una epidemia oculta y una de las tres epidemias mundiales que se han descuidado en los últimos tiempos. El Ministerio de Obras Públicas y Comunicaciones (MOPC) y el Plan Nacional de Seguridad Vial en el Paraguay, lo presentan como un inaceptable problema humano, social y económico que causa gravísimos daños físicos, psicológicos y materiales a las víctimas y a sus familias, así como daños socio-económicos irreparables a la sociedad en su conjunto. En este sentido, para el estudio de esta problemática social, en primer lugar, mediante el empleo de datos referentes al número mensual de accidentes de tránsito en el Paraguay en el período 2000 al 2012 cedidos por la Patrulla Caminera de este país. Esta investigación busca, a través de los modelos ARIMA, propios del análisis de series temporales, identificados a partir de la metodología Box-Jenkins, obtener buenas estimaciones de la tendencia evolutiva del número mensual de accidentes de tránsito, registrados en

las diferentes carreteras en el Paraguay, mediante un proceso de diagnóstico y validación que cumpla las condiciones para que se convierta en una de las principales herramientas estadísticas para la estimación del número mensual de accidentes de tránsito en el Paraguay.

Palabras clave: Series temporales, Población y Muestra, Modelo ARIMA, Metodología de Box-Jenkins, Número mensual de accidentes de tránsito.

ABSTRACT

Traffic accidents are a complex social problem worldwide. The World Health Organization (WHO) designated the dangers of road traffic as a hidden epidemic and one of the three global epidemics that have been neglected in recent times. The Ministry of Public Works and Communications (MOPC) and the National Road Safety Plan of Paraguay, presenting it as an unacceptable human, social and economic problem that causes grave physical, psychological and material to the victims and their families damage and damage socio-economic irreparable damage to society as a whole. In this sense, the study of this social problem, first, by using data concerning the monthly number of traffic accidents in

Paraguay in the period 2000 to 2012 assigned by the Highway Patrol in this country. This research seeks, through the ARIMA own analysis of time series models identified from the Box-Jenkins methodology, get good estimates of the evolutionary trend of the monthly number of traffic accidents recorded on different roads of Paraguay through a validation process of diagnosis and that meets the conditions for it to become one of the main statistical tools to estimate the monthly number of traffic accidents in Paraguay.

Keywords: Time series, Population and Sample, ARIMA model, Box-Jenkins, monthly number of traffic accident

1. INTRODUCCION

En los últimos años, debido a los múltiples factores que lo ocasionan, los accidentes de tránsito constituyen una problemática social compleja en todo el mundo.

La Organización Mundial de la Salud (OMS, 2013), en su informe sobre. La salud en el Mundo 2003 publicado en la segunda quincena de diciembre, expresa que aunque más de 3000 personas mueren cada día, aun no se ha logrado reconocer cabalmente la magnitud del problema ya que los accidentes de tránsito suelen causar pocas muertes a la vez y los medios de comunicación les dan poca cobertura, lo que contribuye a la naturaleza oculta de la epidemia de lesiones ocasionadas por el tráfico vial. En efecto, la OMS designa los peligros del tráfico vial como una epidemia oculta y una de las tres epidemias mundiales que se han descuidado, las otras dos son enfermedades cardiovasculares y la globalización de las enfermedades relacionadas con el tabaquismo.

En el Paraguay, la Dirección Nacional de Transporte (DINATRA)

es la principal responsable del ordenamiento y expansión del sector transporte. Cada año, para tener agrupadas las informaciones referentes al sector que sirvan de herramienta útil a todas las personas e instituciones interesadas en el tema, presenta anuarios estadísticos que constituyen trabajos de compilación. Por su parte, la Patrulla Caminera es la encargada de velar por la tranquilidad y seguridad de la ciudadanía, además de organizar la información referente a los accidentes de tránsito registrados en las carreteras nacionales.

Antecedentes

A nivel nacional no se han realizado investigaciones sobre el número mensual de accidentes de tránsito aplicando análisis de series temporales. Se presentan a continuación, algunas investigaciones realizadas sobre el número de accidentes de tránsito en otros países.

Análisis Lineal del Número de Accidentes de Tránsito en la Provincia de Guayas} (Manzo Peláez y Pino Chauca, 2004)

Esta investigación se fundamentó en el análisis de dependencia de las variables existentes que inciden notablemente en el número de accidentes de tránsito y fue realizada por Nathaly Manzo Peláez y Delia Pino Chauca en el año 2004. El objetivo fue la de analizar la dependencia existente en el número de accidentes de tránsito en relación al aumento de velocidad y el estado de embriaguez en los individuos. Se aplicó un modelo econométrico que ayudó a predecir el comportamiento del número de accidentes de tránsito, conociendo el número de conductores que excedían la velocidad permitida y además, la cantidad de personas alcoholizadas al momento del accidente. Análisis económico de los factores explicativos de la evolución de los

accidentes de tráfico en España. Una aplicación empírica (Fumero Lázaro, 2014)

Este trabajo fue realizado por Daniel Fumero Lázaro en el año 2014. En esta investigación, se ha analizado los factores explicativos de la evolución de las tasas de siniestralidad de tráfico vial en España, haciendo especial énfasis en el papel del permiso por puntos y los efectos de la crisis económica. Para ello, primero se realizó un análisis descriptivo de la evolución de los principales indicadores de movilidad. Posteriormente, se diseñaron varios modelos de regresión múltiple que permitieron analizar el impacto del permiso por puntos y otros factores como la composición del parque automovilístico, la movilidad, las infraestructuras viarias, la crisis económica, entre otros aspectos.

Análisis exploratorio espacial de los accidentes de tránsito en Ciudad Juárez, México (Hernández Hernández, 2012)

Esta Investigación fue realizado por Vladimir Hernández Hernández. Tuvo como principal objetivo elaborar una herramienta para el estudio exploratorio de los accidentes viales en Ciudad Juárez, Chihuahua, México, que aplique de forma exclusiva la variable geográfica espacial (ubicación). El método utilizado fue de un estudio observacional y transversal que utiliza un sistema de información geográfica para explorar la naturaleza espacial de 13305 accidentes viales registrados durante 2008 y 2009 en Ciudad Juárez. Se construyeron indicadores aproximados al flujo de tránsito y se incluyeron dos variables: índice de jerarquía urbana y densidad de población. Además, el análisis exploratorio de datos espaciales constituyó en una fase previa al uso de técnicas multivariadas de mayor alcance. La aplicación de las técnicas del análisis exploratorio por sí misma permitió tipificar agrupamientos

espaciales, identificar autocorrelación global e indicar la dirección de las variables en estudio.

2. MATERIALES Y METODOS

La información la constituyen los datos referentes al número mensual de accidentes de tránsito en el Paraguay en el período 2000 al 2012 registrados por la Patrulla Caminera, en todo el territorio Paraguayo, que está a cargo del Ministerio de Obras Publicas y Comunicaciones.

La investigación es del tipo cuantitativo. Por otro lado, es retrospectiva debido a que el número mensual de accidentes de tránsito en el Paraguay es estudiado utilizando datos tomados de los registros históricos de la Patrulla Caminera.

Debido a que las informaciones utilizadas son secuenciales en el tiempo, este estudio es transversal, longitudinal y retrospectivo. Además, el diseño es apropiado bajo un enfoque no experimental debido a que no hay manipulación de la variable en estudio.

Software utilizado en el análisis de datos

El trabajo completo se lleva a cabo con el software RStudio 0.97.9551 (Crawley, M.I, 2013).

RStudio es un conjunto de herramientas integradas diseñadas para ayudar a ser más productivo el trabajo con R-project 2.15.0. Incluye una consola, editor de resultados que permite la ejecución directa de códigos, así como herramientas para el trazado, la historia, la depuración y la gestión del espacio de trabajo. En resumen, RStudio es un entorno de programación en R disponible en la mayor parte de los sistemas operativos, que, también trabaja como compilador de latex y es muy útil para juntar en una sola hoja de trabajo las sintaxis de R y Latex.

Para el Análisis de series temporales existen varios paquetes en R que permiten acceder a los cálculos necesarios. De entre los ofrecidos para el ajuste de modelos de series temporales, en este trabajo se utilizan los paquetes "tseries" y "forecast".

Serie de tiempo

En la actualidad existen innumerables aplicaciones de la teoría de series de tiempo con el propósito de comprender y descubrir las influencias y estructuras considerando el hecho de que los datos son tomados en diversos periodos, como así también es necesario diversa literatura y teorías que tratan esta técnica estadística.

Una serie de tiempo es definida por Chatfield (2000), como un conjunto de observaciones medidas secuencialmente a través del tiempo. Estas mediciones se pueden realizar de forma continua a través del tiempo o tomarse en un conjunto discreto de puntos de tiempo.

Componentes intervinientes y tipos de modelos en series temporales

Según Cryer y Chan (2008), en una serie temporal intervienen cuatro componentes teóricas. Dichas componentes se detallan a continuación:

Tendencia (T_t): Movimiento de larga duración que se mantienen durante todo el periodo de observación.

Ciclo (C_t): Oscilaciones alrededor de la tendencia producidos por periodos alternativos de prosperidad y depresión.

Estacionalidad (S_t): Movimiento que se produce, dentro de un periodo anual, por motivos ya sea (económicos, climáticos, sociales, etc.)

Residuos (ϵ_t): Movimientos erráticos generados por causas ajenas al fenómeno y no repetidos en el tiempo.

Entonces, un modelo de serie temporal es una función compuesta por estas cuatro componentes teóricas. Por lo tanto, según sea la forma de esta función, un modelo de serie temporal

puede ser aditivo, multiplicativo o mixto. Los modelos aditivos y multiplicativos se construyen respectivamente a partir de la suma y multiplicación de las distintas componentes, es decir:

$$Y_t = T_t + C_t + S_t + \epsilon_t$$

$$Y_t = T_t * C_t * S_t * \epsilon_t$$

Mientras que un modelo mixto se expresa como:

$$Y_t = T_t * (1 + C_t) * (1 + S_t) + \epsilon_t$$

Según Pepió Viñals (2001), la identificación de los tipos de modelos que se ajustan a una serie temporal es a través del patrón de comportamiento de la componente estacional de la serie; es aditivo si dicho patrón se mantiene con amplitud constante y es multiplicativo si se va ampliando con el tiempo.

Modelo Autorregresivo

Un modelo autorregresivo, en una serie temporal Y_t , constituye un modelo de regresión lineal en el cual la actual observación en la serie está en función de observaciones anteriores y cumple la condicional de ser un proceso estacionario.

Los modelos autorregresivos son los modelos más simples estudiados al momento de analizar una serie temporal.

Modelo Autorregresivo de orden $p > 1$, AR(p)

Según Parra Rodríguez (2005), este modelo se basa en que el valor actual de la serie Y_t , se explica en función de p valores pasados $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ donde p determina el número de rezagos necesarios para estimar un valor actual. El modelo autorregresivo de orden p , AR(p), puede escribirse como:

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t$$

donde $\phi_1, \dots, \phi_p$ son los parámetros del modelo, c es una constante y ϵ_t es un término de error. Muchos autores

omiten el término constante, con fines de simplificación.

Modelo Medias Móviles

Un modelo de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Estos modelos se denotan normalmente con las siglas M A seguidos, como en el caso de los modelos autorregresivos, del orden entre paréntesis. Así, un modelo con 1 término de error respondería a la expresión M A (1).

Medias Móviles de orden $q > 1$

Un modelo de medias móviles de orden q , M A (q) se define como:

$$Y_t = \varepsilon_t + \sum_{i=1}^q \phi_i \varepsilon_{t-i}$$

donde $\phi_1, \phi_2, \dots, \phi_p$ son los parámetros del modelo $\varepsilon_t, \varepsilon_{t-1}, \dots$ son, de nuevo, los términos de error.

El orden de los modelos de medias móviles suele ser escogidos de acuerdo a la periodicidad detectada mediante los datos analizados, por ejemplo: M A (4) para series trimestrales, M A (12) para series mensuales, etc.

Modelo Autorregresivo de Media Móvil, ARMA(p, q)

Un modelo autorregresivo de media móvil, ARMA (p, q), es un modelo con p términos autorregresivos y q términos de media móvil. Este modelo es una combinación entre un modelo AR(p) y un modelo MA(q). Entonces, un modelo ARMA (p, q) se define como:

$$Y_t = \varepsilon_t + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \phi_i \varepsilon_{t-i}$$

Habitualmente se asume que los términos de error ε_t constituyen procesos ruido blanco.

Proceso Autorregresivo Integrado de Media Móvil, ARIMA(p, d, q)

Si la serie Y_t no fuera estacionaria, tomando d diferencias simbolizado por Δ^d , se logra que lo sea. Si la serie $W_t = \Delta^d Y_t$, es estacionaria, entonces se puede decir que Y_t sigue un proceso autorregresivo integrado de media móvil de orden (p, d, q), lo que suele expresarse como ARIMA(p, d, q) donde los parámetros p, d y q son números enteros no negativos que indican el orden de las distintas componentes del modelo respectivamente, las componentes autorregresiva, integrada y de media móvil. Cuando alguno de los tres parámetros es cero, es común omitir las letras correspondientes del acrónimo-AR para la componente autorregresiva, I para la integrada y MA para la media móvil. Por ejemplo, ARIMA (0,1,0) se puede expresar como I(1) y ARIMA(0,0,1) como MA(1). El modelo ARIMA (p, d, q) se puede representar como:

$$Y_t = \phi_0 + \sum_{i=1}^p \phi_i \Delta^d Y_{t-1} - \sum_{i=1}^q \phi_i \varepsilon_{t-i} + \varepsilon_t$$

en donde:

d es el número de diferencias que son necesarias para convertir la serie original en estacionaria,

$\phi_1, \dots, \phi_p$ son los parámetros de la parte autorregresiva del modelo,

$\phi_1, \dots, \phi_p$ los parámetros pertenecientes a la parte medias móviles del modelo,

ϕ_0 es una constante, y

ε_t es el término de error.

Metodología de Box-Jenkins

Seguendo a Quesada (2011), la metodología que sugiere Box-Jenkins, consta de cuatro etapas:

Identificación: Consiste en elegir uno o más modelos ARIMA, como candidatos que pueden representar adecuadamente

el comportamiento de la serie. En ésta etapa deben determinarse las transformaciones necesarias para conseguir estacionariedad, contraste de inclusión de un término de tendencia determinística θ_0 y elegir los órdenes p y q a través de las funciones de autocorrelación para cada modelo.

Estimación: Consiste en estimar los parámetros de cada uno de los modelos identificados en la fase anterior.

Diagnosis (Validación): Trata de determinar si los modelos identificados y estimados son adecuados para representar a los datos. Las deficiencias encontradas en ésta etapa pueden utilizarse como información para reformular los modelos.

Predicción: Con los modelos que han sido diagnosticados favorablemente, se pueden realizar predicciones. Esta etapa también puede poner de manifiesto qué modelos poseen deficiencias a la hora de predecir, y puede utilizarse como herramienta de validación de los modelos.

Criterios de Información de Akaike

Los estadísticos de información están basados en la función de verosimilitud que incluye una penalización vinculada con los parámetros estimados (Quesada, 2011).

Akaike (1974), propone el criterio basado en la función verosímil que incluye una penalización que aumenta

con el número de parámetros estimados en el modelo. Premia los modelos que dan un buen ajuste en términos de verosimilitud y a la vez tienen pocos parámetros (o sea son parsimoniosos).

Test para los parámetros

Cada uno de los parámetros de un modelo pueden ser contrastados por la prueba t de Student para determinar su significancia en el modelo. La prueba de Dickey-Fuller para identificar si la serie es un proceso estacionario. También el test de Ljung-Box que indica si existe dependencia entre los m primeros residuos estimados, es decir, si estos residuos presentan correlación no nula.

Validación Cruzada para Series Temporales

La validación cruzada no es una técnica estadística básica, más bien es una forma de medir el rendimiento predictivo de un modelo estadístico, para ello se deja un conjunto de datos no utilizados, en la estimación, que comúnmente se le llama conjunto de pruebas. Teniendo en cuenta que la precisión predictiva del modelo se mide por el error medio cuadrado.

Arlot y Celisse (2010), describen en su trabajo que cuando los datos no son de validación cruzada independiente se vuelve más difícil, ya que dejando de lado una observación no quita toda la información asociada, debido a las correlaciones con otras observaciones.

3. RESULTADOS Y DISCUSIÓN

Modelado del número mensual de accidentes de tránsito en el Paraguay por descomposición ARIMA

El modelo que permita realizar pronósticos del número mensual de accidentes de tránsito a través de ARIMA, contruidos a partir de la metodología de Box-Jenking se analizará en esta sección.

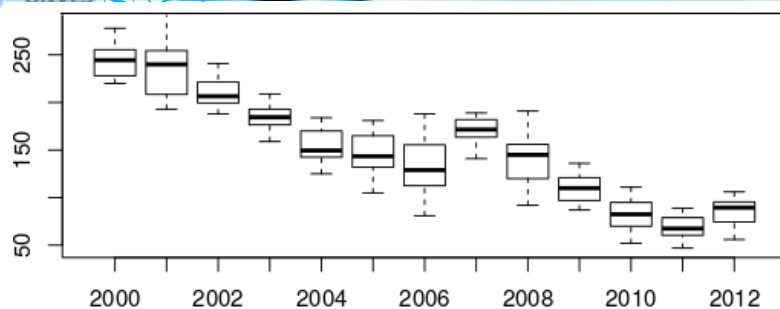


Figura 1: Diseño de cajas del número de accidentes de tránsito anuales en Paraguay, periodo 2000 al 2012.

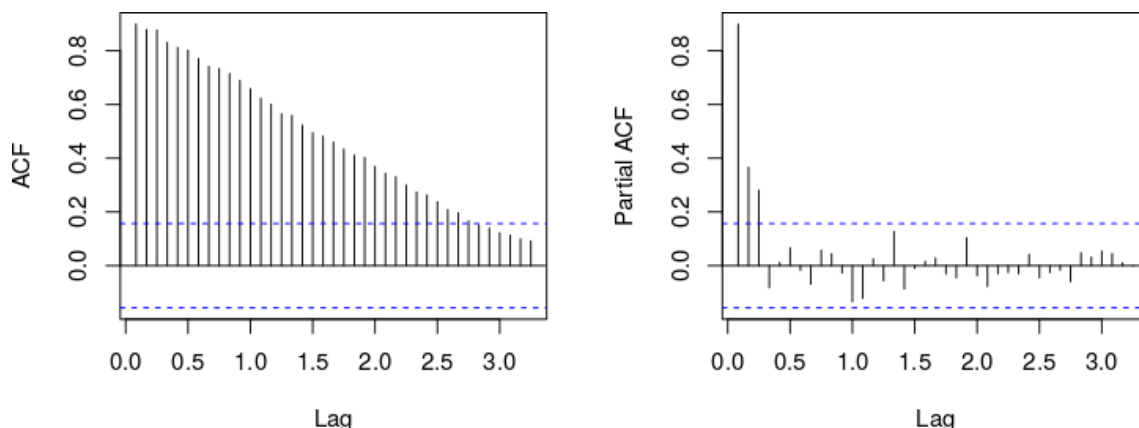


Figura 2: Correlograma Simple y Parcial de la serie original del número mensual de accidentes de tránsito en el Paraguay.

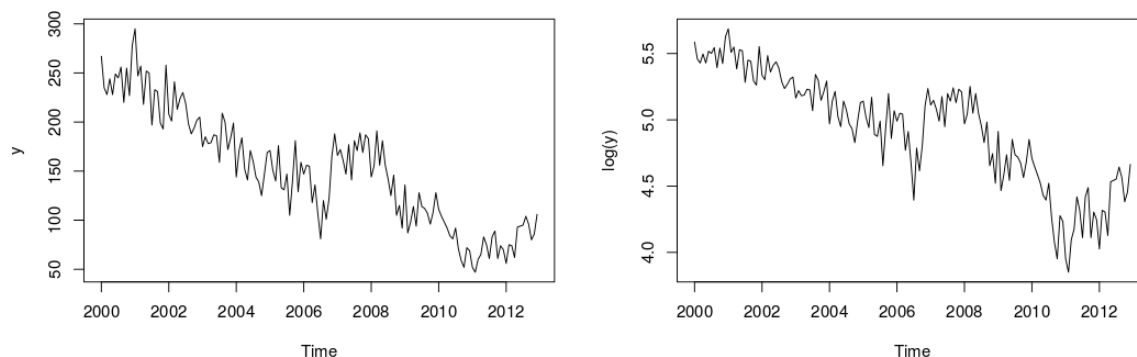


Figura 3: Serie original y la serie con transformación logarítmica del número mensual de accidentes de tránsito en el Paraguay.

Según la Figura 1 se observa que la serie no es estacionaria en media debido a que muestra una marcada tendencia decreciente. Además, la Figura 2 muestra que los coeficientes de autocorrelación tienen un comportamiento exponencial decreciente muy lento, patrón característico de las series no estacionarias. Por su parte, la Figura 3 muestra que al tomar logaritmo de la serie original Y_t se obtiene una serie que muestra una estabilidad en varianza.

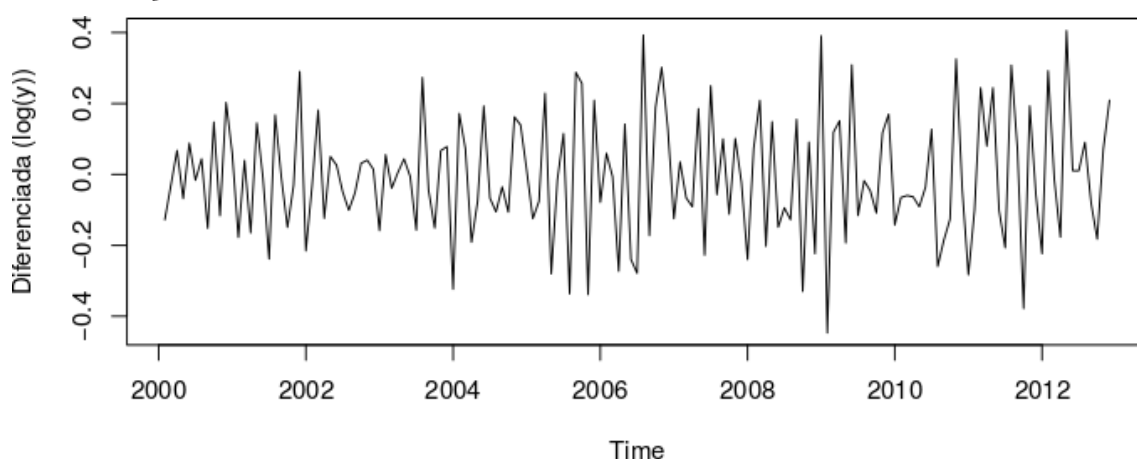


Figura 4: Serie con transformación logarítmica y una diferenciación del número mensual de accidentes de tránsito en el Paraguay.

Como siguiente paso para lograr que la serie $\log(Y_t)$ sea estacionario en media se procede a realizar una diferencia de orden uno en dicha serie. La Figura 4 muestra el comportamiento de la serie transformada:

$$Y_t^* = \log(Y_t) - \log(Y_{t-1})$$

Para probar la independencia de los términos de la serie Y_t^* se realiza el test de Dickey-Fuller. Para dicho contraste, se supone que la serie Y_t^* es un modelo AR(1) de la forma:

$$Y_t^* = \rho Y_{t-1}^* + \epsilon_t$$

y se realiza el siguiente contraste

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

Tabla 1: Test de Dickey-Fuller para la serie transformada

Test de Dickey-Fuller		
Serie	t - valor	$Pr(> t)$
Y_t^*	-6.025	$1.22e - 08 * **$

La Tabla 1 presenta el resumen de la prueba de Dickey-Fuller sobre la serie Y_t^* . Según los resultados, existen evidencias estadísticas para rechazar H_0 debido a que el p-valor es muy inferior a 0.05. En conclusión, podemos decir que la serie Y_t^* es estacionaria.

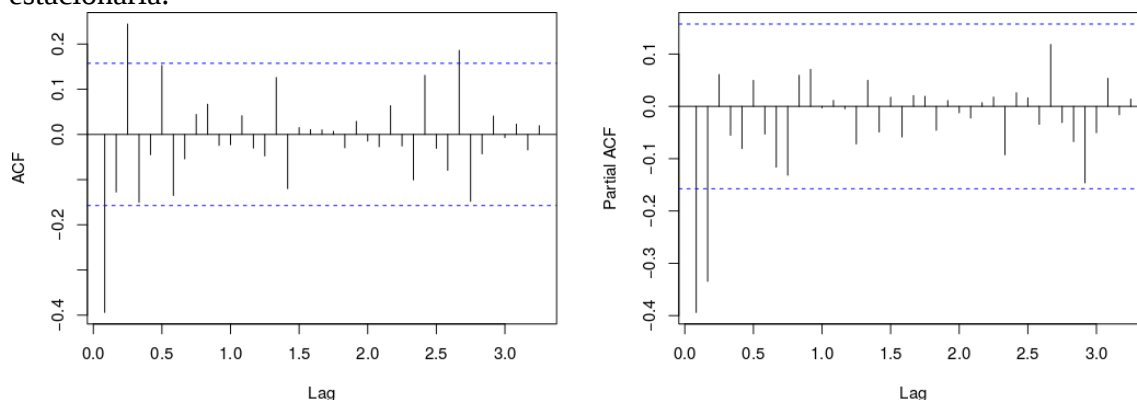


Figura 5: Serie y correlograma con transformación logarítmica y una diferenciación del número mensual de accidentes de tránsito en el Paraguay.

Según los correlogramas simple y parcial, Figura 5, se pueden proponer los siguientes modelos ARIMA:

Un modelo ARIMA $(p,1,0)$, con $0 < p \leq 2$, debido a que el gráfico de autocorrelación simple muestra un patrón de decrecimiento rápido en los primeros retardos, mientras que el parcial PACF muestra que a partir del retardo dos todos los coeficientes ya no son significativos.

Un modelo ARIMA $(0,1,q)$, con $0 < q \leq 3$, debido a que el gráfico de autocorrelación simple muestra que a partir del retardo tres todos los coeficientes ya no son significativos, mientras que el parcial PACF muestra un patrón de decrecimiento rápido en los primeros retardos.

Un modelo ARIMA $(p,1,q)$, con $0 < p \leq 2$ y $0 < q \leq 3$, debido a que los autocorrelogramas muestran patrones de decrecimientos rápidos en los primeros retardos.

Tabla 2: Modelos ARIMA(p,1,q) propuestos para la estimación del número mensual de accidentes de tránsito.

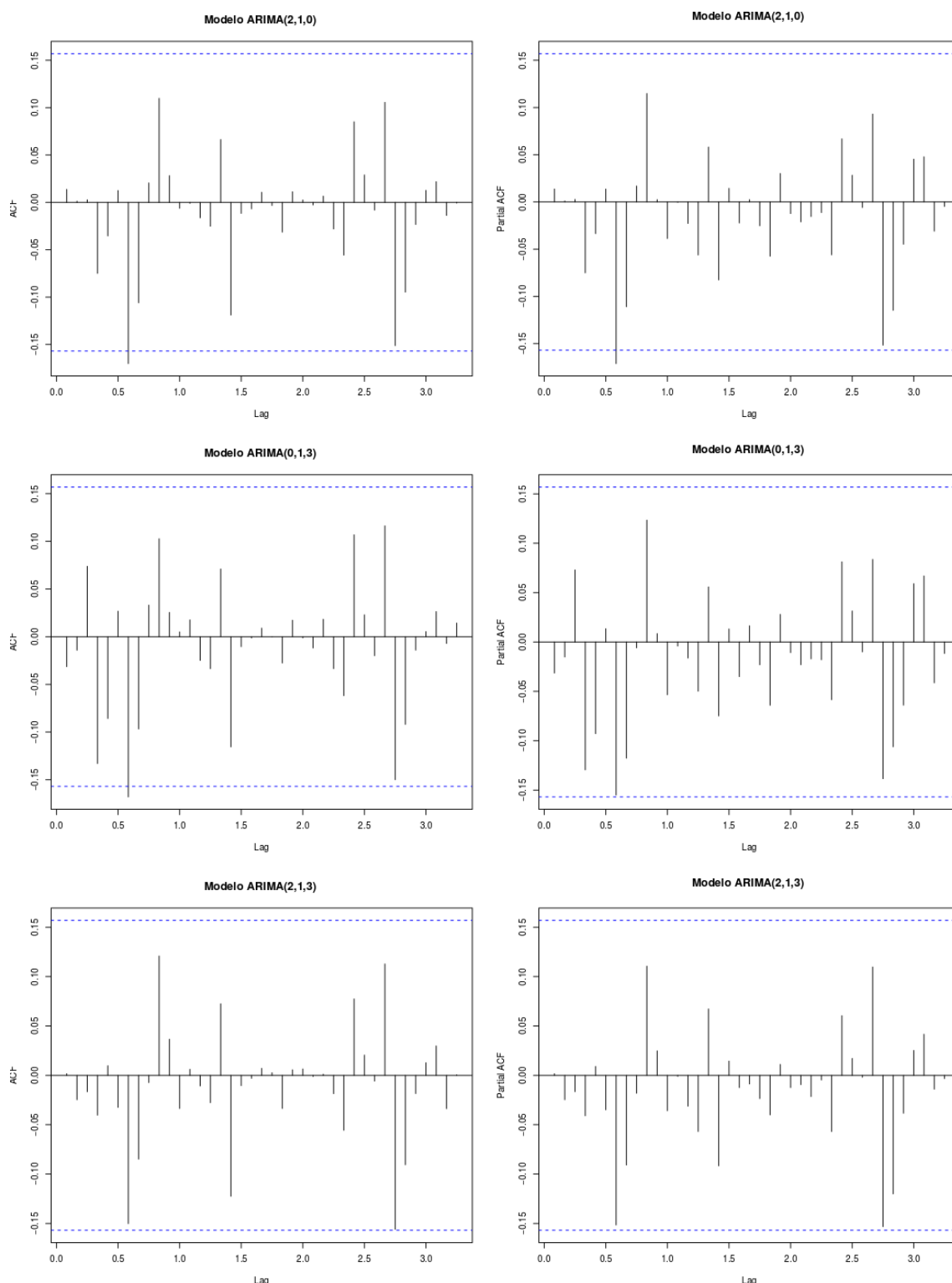
ARIMA(p,1,0)	AIC	RMSE	ARIMA(0,1,q)	AIC	RMSE
ARIMA(1,1,0)	-126.75	0.158	ARIMA(0,1,1)	-139.18	0.152
ARIMA(2,1,0)	-143.43	0.149	ARIMA(0,1,2)	-137.24	0.152
			ARIMA(0,1,3)	-137.91	0.150

Según la Tabla 2, los modelos autorregresivo y de medias móviles que muestran los mejores ajustes son: el ARIMA (2,1,0) y ARIMA (0,1,3) debido a que muestran valores bajos en los estadísticos AIC y RMSE.

Tabla 3: Resumen de las estimaciones de los estadísticos a partir de los modelos ARIMA(2,1,0), ARIMA(0,1,3) y ARIMA(2,1,3).

Modelos	AIC	RMSE	χ^2 Ljung-Box	p – valor
ARIMA(2,1,0)	-143.43	0.149	10.2441	0.8041
ARIMA(0,1,3)	-137.91	0.150	13.9251	0.5312
ARIMA(2,1,3)	-139.48	0.147	9.9005	0.826

La Tabla 3 resume los valores de los estadísticos AIC, RMSE y χ^2 de Ljung-Box, además de los p-valores, obtenidos a partir de los modelos ARIMA (2,1,0), ARIMA(0,1,3) y ARIMA(2,1,3). Todos los p-valores son muy superiores a 0.05, indicando que existen evidencias estadísticamente significativas para no rechazar la hipótesis de que todos coeficientes de autocorrelaciones, en los tres modelos, sean iguales a cero.



Figuran 6: Correlogramas simples y parciales construidos a partir de los tres modelos ARIMA propuestos.

En la Figura 6 se presentan los correlogramas simples y parciales construidos a partir de los modelos ARIMA (2,1,0), ARIMA(0,1,3) y ARIMA(2,1,3). Para los tres casos, los gráficos ACF y PACF muestran que, los coeficientes de autocorrelaciones caen dentro o están muy poco alejados de la franja de confianza de Bartlett, indicando que los errores constituyen un proceso ruido blanco.

Finalmente, los tres modelos ARIMA presentados en la Tabla 3 sirven para realizar pronósticos del número mensual de accidentes de tránsito. Dado que, los modelos ARIMA (2,1,0) y ARIMA(0,1,3) toman parte del modelo ARIMA(2,1,3), este último será el modelo seleccionado para su posterior diagnóstico y validación.

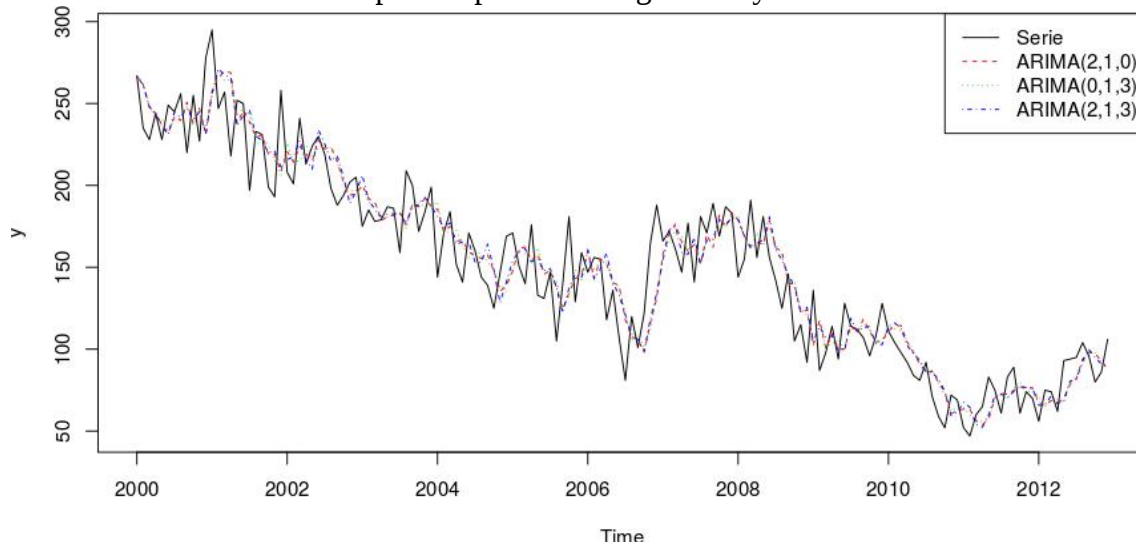


Figura 7: Ajustes conseguidos con los tres modelos ARIMA seleccionados.

Validación del modelo de serie temporal utilizado para la estimación del número mensual de accidentes de tránsito en el Paraguay

Se procede ahora a diagnosticar y validar el modelo de mejor ajuste propuesto en la sección anterior mediante el estudio de los residuos y la validación cruzada de k iteraciones. Para el diagnóstico, como primer paso, se realiza contrastes t sobre los parámetros que componen el modelo ARIMA (2,1,3). Para ello, se plantean las siguientes hipótesis:

$$H_0: \beta = 0, \text{ con } \beta \in \{c, \phi_1, \phi_2, \theta_1, \theta_2, \theta_3\}$$

$$H_1: \beta \neq 0$$

Tabla 4: Contrastes t realizados sobre los parámetros del modelo ARIMA(2,1,3).

Parámetros	estimación	SE	t – valor	p – valor
constante	-0.0063	0.0065	-0.973	0.322
ϕ_1	-0.8182	0.2625	-3.753	2.5 e-4
ϕ_2	-0.7040	0.3583	-3.181	1.78 e-3
θ_1	0.3006	0.3057	-1.307	0.193
θ_2	0.2550	0.5537	-1.107	0.270
θ_3	-0.1850	0.2236	1.120	0.265

La tabla 4 ofrece el resumen de los contrastes t realizados sobre cada uno de los parámetros intervinientes en el modelo ARIMA (2,1,3). Los p -valores correspondientes a la constante y a los parámetros que conforman $\theta_1, \theta_2, \theta_3$ son muy superiores a 0.05, y por lo tanto, no contribuyen de manera significativa en el modelo. Entonces el modelo ARIMA queda expresado como:

$$Y_t^* = \phi_1 Y_{t-1}^* + \phi_2 Y_{t-2}^* + \epsilon_t$$

Tabla 5: Contrastes t realizados sobre los parámetros intervinientes en el modelo ARIMA(2,1,0).

Parámetros	estimación	SE	$t - valor$	$p - valor$
ϕ_1	-0.5314	0.0761	-6.939	2.56 e-8
ϕ_2	-0.3382	0.0758	-4.419	1.87 e-5

En la Tabla 5 se presenta el resumen de las estimaciones de los parámetros ϕ_1 y ϕ_2 los cuales son significativos. Como siguiente paso, en el proceso de diagnóstico se realiza el análisis de los residuales. Para tal fin, en primer lugar, se realizan los siguientes contrastes:

$$\begin{aligned}
 H_0: E(\epsilon_t) &= 0, \quad \forall t \\
 H_1: E(\epsilon_t) &\neq 0, \text{ para alg } t \\
 H_0: \hat{\rho}(k) &= 0, \quad \forall k = 1, 2, \dots, 39 \\
 H_1: \hat{\rho}(k) &\neq 0, \text{ para alg } k
 \end{aligned}$$

Tabla 6: Contrastes t y prueba de Ljung-Box sobre la media y los coeficientes de autocorrelación de los errores.

Modelo	Prueba t		Prueba de Ljung-Box	
	t	$p - valor$	χ^2	$p - valor$
ARIMA(2,1,0)	-1.0166	0.3109	25.029	0.9596

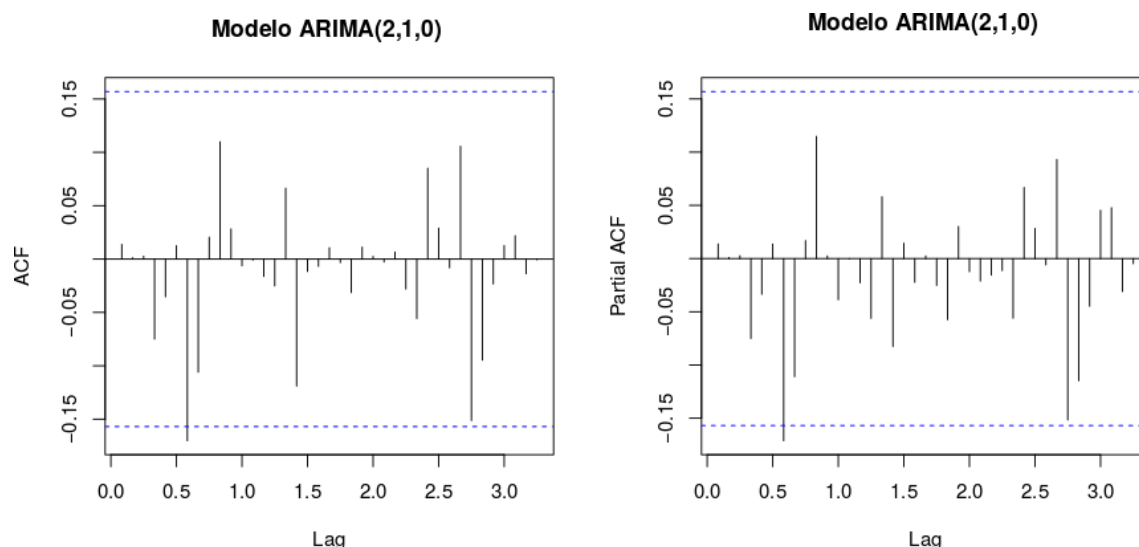


Figura8: Ajustes conseguidos con los modelos presentados en la Tabla 6

Según la Tabla 6, los p -valores obtenidos en los contraste t y en la prueba de Ljung-Box son muy superiores a 0.05. Por lo tanto, no existen evidencias estadísticamente significativas para rechazar H_0 en ambos casos. Entonces, podemos asumir que la media y todos los coeficientes de autocorrelación de los errores son iguales a cero. Además, la Figura 8 muestra que todos los coeficientes no son significativos. En conclusión, los errores constituyen un proceso ruido blanco.

Entonces, se procede a la realización del proceso de validación cruzada de $k=8$ iteraciones. Para ello, en cada iteración, se divide la muestra en dos subconjuntos, una que será la muestra de entrenamiento n_e y la otra de prueba n_p .

Tabla 7: Estimaciones de los parámetros del modelo ARIMA(2,1,0) a través de las muestras de entrenamiento, en 8 iteraciones

Iteración	Muestra de entrenamiento		Estimaciones de		
k	Años abarcados	n_e	ϕ_1	ϕ_2	RMSE
1	2000-2005	72	-0.6695	-0.5777	0.1098
2	2001-2006	72	-0.5341	-0.4214	0.1412
3	2002-2007	72	-0.5328	-0.3232	0.1401
4	2003-2008	72	-0.5381	-0.2753	0.1524
5	2004-2009	72	-0.5927	-0.2448	0.1610
6	2005-2010	72	-0.5466	-0.1934	0.1670
7	2006-2011	72	-0.4535	-0.1834	0.1724
8	2007-2012	72	-0.5194	-0.2847	0.1621
Promedio			-0.5483	-0.3129	0.1507

Tabla 8: Validaciones de las muestras de entrenamiento a través de las muestras de prueba para el ARIMA(2,1,0), en 8 iteraciones

Iteración	Muestra de prueba	Estimaciones de		
k	Años abarcados	n_p	$MBE\%$	$MABE\%$
1	2006-2012	84	-3.0524	15.0928
2	2000 y 2007-2012	84	-2.4243	13.3082
3	2000-2001 y 2008-2012	84	-3.0010	14.1935
4	2000-2002 y 2009-2012	84	-1.6974	12.8337
5	2000-2003 y 2010-2012	84	-2.1057	12.4169
6	2000-2004 y 2011-2012	84	-1.2160	11.9367
7	2000-2005 y 2012	84	-1.2160	11.4920
8	2000-2006	84	-1.8532	11.6841

En las Tablas 7 y 8 se resumen las estimaciones a través de las muestras de entrenamiento y de prueba en 8 iteraciones diferentes. La Tabla 7 muestra que las estimaciones en promedio de los parámetros, utilizando las muestras de entrenamiento, presentan leves diferencias de una iteración a otra, indicando de esta manera una muy buena estabilidad dentro de la muestra total considerada. La Tabla 8 muestra leves subestimaciones en todos los casos. Los valores del MBE en valor absoluto son inferiores a 3.1% en todos los casos. Los valores del MABE son inferiores a 16% en todos los casos, indicando que las estimaciones del número mensual de accidentes de tránsito fueron estimados con errores inferiores a ese porcentaje en cada iteración.

Con todo lo expuesto, el modelo ARIMA estimado y validado en esta investigación está dado por:

$$\hat{Y}_t^* = -0.5314Y_{t-1}^* - 0.33382Y_{t-2}^*$$

Por lo tanto, si se mantienen las condiciones actuales, el número mensual de accidentes de tránsito en el Paraguay pueden ser estimados mediante la expresión:

$$\hat{Y}_t = Y_{t-1}^{0.4686} \cdot Y_{t-2}^{0.1932} \cdot Y_{t-3}^{0.3382}$$

CONCLUSIONES

Debido a la problemática social ocasionada por los accidentes de tránsito, dentro del territorio paraguayo, y por ende la importancia de estimar su evolución a través de los diferentes meses del año, este trabajo revela resultados satisfactorios desde el punto de vista estadístico.

Utilizando la técnica del método univariado de series temporales, esta investigación permite obtener estimaciones razonables del número mensual de accidentes de tránsito en el Paraguay, a través del modelo de mejor ajuste ARIMA (2,1,0) según los estadísticos AIC y RMSE.

Este modelo fue seleccionado después de realizado un proceso de diagnóstico sobre los coeficientes intervinientes en el modelo, mediante diversos contrastes de hipótesis, que demostraron sus significancias en el modelo ARIMA (2,1,0). Luego, el proceso de validación cruzada de 8 iteraciones, con lo que se pudo constatar que las estimaciones de los parámetros del modelo ARIMA estimado muestren leves diferencias de una iteración a otra, indicando una muy buena estabilidad dentro de la muestra total considerada. También es utilizada para validar los modelos estimados mediante la muestra de entrenamiento. Que reveló muy buena estabilidad en la muestra tomada y en las estimaciones de los parámetros intervinientes.

Por otro lado, es importante mencionar ciertas limitaciones que presenta el modelo ARIMA estimado. Según las bibliografías consultadas los modelos ARIMA son muy sensibles a la inclusión de nuevos datos y exigen que se mantengan las condiciones en las cuales fue tomada la muestra en el transcurso del tiempo, por esta razón es importante destacar que el modelo

ARIMA seleccionado no será útil para realizar estimaciones a largo plazo.

Mediante este modelo, que constituye una fuerte herramienta estadística, se puede decir que existen condiciones para que se convierta en una de las principales herramientas para la estimación del número mensual de accidentes de tránsito en el Paraguay.

Este trabajo espera dar incapié para estudios posteriores y de mayor envergadura sobre la problemática en estudio, por lo que se recomienda posibles mejoras al modelo propuesto en investigaciones futuras que puedan surgir, por ejemplo considerar un modelo más completo como SARIMA o ARCH (Autorregresive Conditional Heteroskedsticity), que sean capaces de procesar de mejor forma la información existente. También es factible e importante aplicar una Regresión Múltiple que pueda medir las asociaciones de factores influyentes en el número mensual de accidentes de tránsito en el Paraguay.

AGRADECIMIENTOS

A Dios.

Quien en su infinita bondad y amor, me ha dado salud y fuerzas para llegar hasta estas instancias y cumplir mis metas.

A mi madre.

Por haberme apoyado en todo momento, por sus consejos, sus valores, por la motivación constante, pero más aún le agradezco por sus oraciones incansables que me han ayudado a seguir adelante con fuerza y confianza.

A mi familia.

Por darme amor, cariño y comprensión en los momentos más difíciles y sobre todo, les agradezco por darme siempre confianza para alcanzar mis anhelos.

A mi tutor.

El Prof. Dr. Xavier Bardina, por la paciencia mostrada y las orientaciones precisas en todo el proceso.

A la Facultad de Ciencias Exactas y Naturales (FACEN) y a la CONACYT por darme la oportunidad de formar parte de esta maestría y aumentar mis conocimientos profesionales en el área de la investigación

REFERENCIAS

- Andrade, M. G., Sánchez, C.E., (2011): “Determinación de zonas susceptibles a accidentes de tránsito en el Cantón Rumiñahui, mediante el desarrollo de implementación de un sistema de información geográfica para la Policía Nacional”. Proyecto de grado para la obtención del título de Ingeniería, Escuela Politécnica del Ejercito, Sangolquí, Ecuador. Disponible en, <http://repositorio.espe.edu.ec/bitstream/21000/4613/1/T-ESPE-032743.pdf>
- Akaike, H., (1974): “A New Look at the Statistical Model Identification”. IEEE Transactions on Automatic Control.
- Arlot, S., Celisse, A., (2010): “A survey of cross-validation procedures for model selection”. Statist. Surv. 4 (2010), 40–79. doi:10.1214/09-SS054. Disponible en, <http://projecteuclid.org/euclid.ssu/1268143839>
- Bowerman, B. L., koehler A., O’Connell. R., (2009): “Pronósticos, series de tiempo y regresión. Un enfoque aplicado”. CENGAGE Learning(original), Progreso No.10 Col. Centro Ixtapaluca Edo. De México, 4 edition. Impreso en Cosegraf la versión en Español. Disponible en, https://seriesdetiempo.files.wordpress.com/2013/02/lectura_descomposic3b3n-multiplicativa-y-mc3a9todos-de-suavizamiento-12-feb-2013.pdf
- Brockwell, Peter J., R. A. D. (1987): “Time Series: Theory and Methods”. Series in Statistics. Springer, Colorado, U.S.A., first edition.
- Brockwell P. J., Davis R. A., (2002): “Introduction to time series and forecastin”. Series in Statistics. Springer, United States of America, 2nd edition. Disponible en, <http://www.masys.url.tw/Download/2002-BrockwellIntroduction%20Time%20Series%20and%20Forecasting.pdf>
- Chatfield, C., (2000): “Time-series forecastin”. CHAPMAN & HAL-L/CRC Boca Raton London New York Washington, D.C.. Disponible en, [https://is.muni.cz/el/1456/jaro2009/PMAPEM/NotSortedYet/Time-Series-Forecasting Chris Chatfield 1584880635.pdf](https://is.muni.cz/el/1456/jaro2009/PMAPEM/NotSortedYet/Time-Series-Forecasting%20Chris%20Chatfield%201584880635.pdf)
- Crawley, M. J., (2013): “The R Book, Second Edition”. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, United Kingdom. Disponible en, <http://www.Bookarchive.org>
- Cryer J. D., Chan K., (2008): “Time Series Analysis with applications in R”. Second Edition, Springer Science Business Media LLC, USA.
- Fumero Lázaro D. D., (2014): “Análisis económico de los factores explicativos de la evolución de los accidentes de tráfico en España. Una aplicación empírica”. Tesis de grado Grado en Economía Facultad de Ciencias Económicas y Empresariales, Universidad La Laguna, España. Disponible en, <http://riull.ull.es/xmlui/handle/915/219>
- Giraldo G., N. D., (2006): “Notas de Clase Series de Tiempo con R”. Escuela de Estadística Universidad Nacional de Colombia Medellín.
- Gonzalez C, M.P., (2009): “Análisis de Series temporales: Modelos ARIMA”. Facultad de Ciencias Económicas y Empresariales, Universidad de Vasco (UPV-EHU), ISBN: 978-84-692-3814-1.

- González C, M.P., (2009): “Técnicas de predicción económica”. Facultad de Ciencias Económicas y Empresariales, Universidad de Vasco (UPV-EHU), ISBN: 978-84-692-3815-8.
- Gonzalo R., (2008): “Series de Tiempo”. Universidad De Chile Facultad De Ciencias Físicas y Matemáticas.
- Hernández Hernández V., (2012): “Análisis exploratorio espacial de los accidentes de tránsito en Ciudad Juárez, México”. Rev Panam Salud Pública. 2012;31(5):396–402. Disponible en, <http://www.scielo.org/pdf/rpsp/v31n5/a07v31n5.pdf>
- Manzo Pelaez, N., Pino Chauca, D., (2006): “Análisis Lineal del Número de Accidentes de Tránsito en la Provincia de Guayas”. Trabajo de grado de la Carrera de Economía, Departamento de Investigación, Desarrollo Tecnológico e Innovaciones, UNEMI, Ecuador. Disponible en, <http://es.scribd.com/doc/48241845/Analisis-Lineal-Del-Nmero-de-Accidentes-en-La-Provincia-Del-Guayas-PDFscribd>.
- Ministerio de Obras Públicas y Comunicaciones (MOPC), (2008): “Plan Nacional de Seguridad Vial 2008-2013. Documento Ejecutivo”. Consultado el 25 de abril 2015. Disponible en, [http://www.mopc.gov.py/mopcweb/pdf/pnsv 2008 2013.pdf](http://www.mopc.gov.py/mopcweb/pdf/pnsv%2008%202013.pdf)
- Mogni, A. P.,(2013): “Modelos de Series de Tiempo con aplicaciones en la industria aerocomercial”. Tesis de Licenciatura, Universidadde Buenos Aires, Facultad de Ciencias Exactas y Naturales. Disponible en, [http://cms.dm.uba.ar/academico /carreras/licenciatura/tesis/2013/Andres-Mogni.pdf](http://cms.dm.uba.ar/academico/carreras/licenciatura/tesis/2013/Andres-Mogni.pdf)
- Montgomery, D.C. & Runger, G.C (2003):“Applied Statistics and Probability for Engineers”. Third Edition, Arizona State University, John Wiley & Sons, Inc, 976 páginas.