Building SARIMA Models to Analysis and Forecast Time Series Data of Road Traffic Accidents in the Kurdistan Region of Iraq

ID No. 364

(PP 382 - 393)

https://doi.org/10.21271/zjhs.27.1.23

Mryam Mohammed Ahmed
Department of Statistics and Information
College of Administration and Economics
Salahaddin University-Erbil
Mryam.Ahmed@su.edu.krd

Saman Hussein Mahmood
Department of Statistics and Information
College of Administration and Economics
Salahaddin University-Erbil
Saman.Mahmood@su.edu.krd

Received: 17/06/2022
Accepted: 22/08/2022
Published: 25/02/2023

Abstract:

Traffic accidents cause injuries and deaths, but also cause material damage to society. Therefore, predicting and identifying the causes of traffic accidents is necessary and important to reduce these losses. The main objective of this study is to develop an ARIMA time series model to investigate and analyze the number of traffic accidents and the number of deaths in traffic accidents in the Iraqi Kurdistan Region according to monthly during the years (2014-2021) and monthly forecast for 2022. Data were obtained from Erbil General Directorate of Traffic. The outcomes demonstrated that the series had seasonal traits. Several models were tested, and the best findings were chosen based on the minimal statistical standards (RMSE, MAE, and MAPE) used for comparison. The best results were we found SARIMA (1,1,1) $(0,1,1)_{12}$ models for the number of accidents and SARIMA (0,1,1) $(1,1,2)_{12}$ models for the number of deaths.

Finally, using the best models, we made a monthly forecast for the number of accidents and deaths. We can say that the rate has not significantly decreased for the forecast period, so the government should develop better and more detailed plans to reduce traffic accidents.

Keywords: SARIMA, Forecasting, Traffic Accidents.

1.Introduction:

Time series forecasting uses knowledge of past values and patterns to predict future behavior. Typically, trend analysis, periodic fluctuation analysis, and seasonality issues are included. As with any other way of forecasting [16]. Traffic collisions are the leading cause of death and injury worldwide. Highway-related crashes kill more than 1.2 million people per year and injure up to 50 million more. By 2030, highway-related crashes are expected to be the world's fifth biggest cause of death. In addition to the death and injury figures, highway crashes cause incalculable agony and suffering, as well as billions of dollars in medical bills and missed productivity. Disorderly driving, unlawful speeding, drunk driving, bad weather, mis driving, and so on are the most common causes of road accidents [10]. Forecasting traffic accidents, damages, and fatalities is a crucial responsibility for traffic safety planners. These forecasts are typically helpful for understanding accident patterns and the effectiveness of current safety actions. That is, safety planners are interested in evaluating present policies and safety measures by examining future accident trends and applying corrective steps [2]. Finding a suitable forecasting model for a traffic accident demand is not an easy process because there are different approaches and models for studying time-series data. Despite the fact that various forecasting methods have been created, one of the most well-liked and frequently used models is the Box-Jenkins model. One method for anticipating data ranges based on inputs from a particular time series is the Box-Jenkins Model. Moving averages, auto regression, and independent component analysis are the three concepts it employs to predict data. The letters p, d, and q stand for these three concepts. Each principle is used in the Box-Jenkins analysis, and the output is an ARIMA (p, d, q) [16]. A seasonal pattern occurs when seasonal factors, such as the time of year or the day of the week, have an impact on a time series. Seasonality occurs at set, definite intervals [6].

Two forecasting models for traffic accidents, the number of accidents, and the number of deaths from road accidents in the Kurdistan Region were developed in this study using information from the General Directorate of Traffic in the governorate of Erbil (2014-2021). The Statistical analysis using program (Statgraphics V. 19).

Methodology:

2.1 Time series:

A time series is a collection of observations made over a period of items in a specific arrangement. The statistical process for analyzing suc a series of data is known as "time series analysis" [5]. Components of the Time Series: Long-term movements in the mean are referred to as a trend (Tt). Seasonal effects (It), or calendar-related cyclical changes, Cycles (Ct): other cyclical fluctuations (such as business cycles), and I—the value of the irregular component (Residuals): other random or systematic fluctuations [3].

$$Y_t = T_t + S_t + C_t + I_t \qquad \ldots\ldots (2.1)$$

Analytical Objectives

The main goals of time series analysis are as follows:

1. Constructing input-output models that outline the processes that underlie the time series' corresponding difference equation.

2.Using the developed models, predicting time series values for the future from the past values.

3.Design of control systems based on the findings of the analysis.

According on where the observed results came from, predicting future values of time series to use a time-domain or bandwidth method can also help with effective operations and production monitoring, failure detection, product quality control, and other related tasks.

After been built and tested, the time series model may be used to predict future time series values at different time intervals d. obviously, forecasting only provides an estimate of the future data values that the input time series will still have [13].

2.2 Stationary Time Series Models:

Stationary stochastic processes are a significant subset of stochastic processes. A time series is considered stable if there is no event of a system failure in the mean (no trend), no systematic change in the variance, and all precisely periodic fluctuations have been eliminated [4].

In theoretical investigations of time series models, stationary is extremely important. There are two types of stationary: strictly stationary and weakly stationary. Random processes that are just first- and second-order probability distribution time stable are referred to be "weakly stationary". Random processes that are strictly stationary have probability distributions that are consistent over time. Some time series exhibit non-stationary, which occurs in linear or nonlinear systems, as fluctuations in the system's representation across time. When there is a smoothly shifting trend component with shifts in the mean as well as fluctuations in the variance of the process, non-stationarities occur [7]. If a time series, y_t, is not stationary but has a first difference, we will term it homogeneous non stationary. $W_t = y_t - y_{t-1} = (1-B)y_t$ or higher-order differences, $W_t = (1-B)^d y_t$,produce a stationary time series [12].

2.3 Autocorrelation Function (ACF):

The coefficient of correlation between z_t and z_{t-k} is:

$\rho_k = \frac{cov\,(z_t\,, z_{t-k})}{\sqrt{Var(z_t)\ \ Var(z_{t-k})}} = \frac{\gamma_k}{\gamma_0}$ ..... (2.2)

It's named the autocorrelation function since it's viewed as a function of lag k [9].

2.4 Partial Autocorrelation Function (PACF):

The partial autocorrelation function is the correlation function between x_t and x_{t+k} after the intervening variables x_{t+1}, x_{t+2},......, x_{t+k-1} have been removed, . This is frequently followed by a conditional autocorrelation as follows:

Corr $(x_t, x_{t+k} /\ x_{t+1}, \ldots\ldots\ldots, x_{t+k-1})$ ..... (2.3)

In time series analysis, this correlation function is known as the (PACF) [1].

2.5 Some Time Series Models:

I.Autoregressive of order (P) Model:

Suppose that $\{a_t\}$ is a purely random process with mean zero and variance σ_a^2 . Then a process $\{Z_t\}$ is said to be an autoregressive process of order p (abbreviated to an AR(p) process) if

$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + a_t$ ..... (2.4)

Where $\{\emptyset_1, \emptyset_2, \ldots, \emptyset_p\}$ are unknown parameters. a_t : White noise purely random variable [4].

II.Moving Average Model of order (q)

The current white noise term and the q most recent previous white noise terms are combined linearly to form a moving average (MA) process of order q.

$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$ ..... (2.5)

Where $(\theta_1, \theta_2, \ldots, \theta_q)$ are unknown parameters. a_t; white noise purely random variable [11].

III.Mixed Autoregressive Moving Average Models

A useful class of time series models is produced when MA and AR procedures are combined. With p AR terms and q MA terms, an ARMA process of order is a mixed autoregressive/moving-average process (p, q). It is supplied by

$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots \emptyset_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$ ..... (2.6)

where $\{a_t\}$ is a purely random process with mean zero and variance σ_a^2 .
The importance of ARMA processes comes from the fact that a stationary time series can often be accurately simulated by an ARMA model with fewer parameters than a pure MA or AR process. This is a classic illustration of the Parsimony Principle. As a result, we seek a model with the fewest number of parameters that can well describe the data [4].

IV.ARIMA model:

Both types of models are fitted to time series data in order to offer generalized information and predict upcoming points in the series. The Auto Regressive Integrated Moving Average (ARIMA) is a generalization of the ARMA model. The ARIMA parameters (p, d, and q) are all positive numbers that represent the autoregressive model (number of time lags), the degree of differencing (number of times past values were subtracted from the data), and the order of the moving average model, respectively [8].

2.6 The Box-Jenkins Procedures

Box and Jenkins made a significant contribution by presenting a generic technique for time-series forecasting that highlights the significance of iteratively choosing an effective model. Indeed, in many fields of statistics, the iterative method to model development that they proposed has subsequently become conventional. Box and Jenkins have shown how differencing may be used to convert ARMA models to ARIMA models, allowing them to deal with non-stationary data. Box and Jenkins also demonstrate how to include seasonal components in seasonal ARIMA (SARIMA) models. ARIMA models are sometimes referred to as Box–Jenkins models because of all of these essential contributions.

In a nutshell, the following are the main steps in creating a Box–Jenkins forecasting model:

1-Model identification: Examine the data to determine which ARIMA process class looks to be the most appropriate.

2-Estimation: Calculate the parameters of the model you've chosen.

3-Diagnostic checking: Check the residuals from the fitted model to verify if they are sufficient.

4-Consideration of alternative models if necessary: Alternative ARIMA models may be explored until a good model is identified if the initial model appears to be unsatisfactory for some reason. When such a model is discovered, calculating forecasts as conditional expectancies is usually quite simple [4].

2.7 SARIMA model:

SARIMA model is the product of seasonal and non-seasonal polynomials and is designated by SARIMA (p, d, q) x$(P, D, Q)_s$, where (p, d, q) and (P, D, Q) are non-seasonal and seasonal components, respectively with a seasonality‘s’. SARIMA model was defined at Equation

$\Phi(B^s)\ \emptyset(B)(1-B^s)^D\ (1-B)^d\ y_t = \Theta(B^s)\ \theta(B)\varepsilon_t$ ….. (2.7)

where: Φ and ∅ = autoregressive (AR) parameters of seasonal and non-seasonal components, correspondingly; Θ and θ = moving average (MA) parameters of seasonal and non-seasonal components, respectively; B = backward operator, B (y_t) = y_{t-1}; (1-$B^s)^D$ = D^{th} seasonal modification of season s; $(1\text{-B})^d$ = d^{th} non-seasonal difference; ε_t = an individualistically distributed random variable; P and p = the orders of the AR components; Q and q = the orders of MA components; D and d are difference terms [15].

2.8 Forecasting:

One of the most fascinating and useful aspects of time-series analysis is the extrapolation of the model beyond the sample size T to give point and interval estimates of values that will be seen at a later period. This is referred to as forecasting, a division of the more general concept of prediction. One method for anticipating data ranges based on inputs from a particular time series is the Box-Jenkins Model. Moving averages, differencing, and auto regression are the three methods it employs to forecast data. These three principles are denoted by the letters p, d, and q. Each principle is employed in the Box-Jenkins analysis, and the results are displayed as an autoregressive integrated moving average, or ARIMA (p, d, q) [14].

$\hat{Z}_t(L) = E(Z_{t+l}\ / Z_t, Z_{t-1}, \ldots, Z_1)$ ….. (2.8)

Where: $\{Z_{t+l}\}$ is the forecasting values with lead time (l) and actual time (t).

3. Applications

A= Number of traffic Accidents

B= Number of Death Accidents

Accidents represent Number of traffic accidents and number of deaths taken from the records of the State Traffic Administration General Directorate of Traffic in Erbil Governorate where the series according to the study how much it is.

Table (1) and (2) shows the data number of traffic accidents and number of deaths January 2014 to December 2021 and 96 months long.

Table 1: monthly average Traffic Accidents in the Kurdistan Region for the period 1/1/2014 – 31/12/2021

Months	2014	2015	2016	2017	2018	2019	2020	2021
1	447	311	341	336	301	372	355	266
2	410	329	262	304	323	292	311	252
3	427	383	334	347	316	337	155	316
4	405	397	318	355	356	334	124	325
5	432	429	370	381	281	320	198	350
6	430	437	385	330	393	370	200	401
7	330	408	377	431	393	426	231	377
8	378	458	474	440	440	410	249	392
9	376	377	469	441	410	467	276	428
10	363	282	408	396	371	475	327	408
11	349	319	380	333	349	414	280	353
12	312	355	316	291	275	262	198	252

Table 2: monthly average Death Accidents in the Kurdistan Region for the period 1/1/2014 – 31/12/2021

Months	2014	2015	2016	2017	2018	2019	2020	2021
1	43	60	53	49	50	44	31	41
2	62	40	51	40	34	39	23	43
3	69	65	58	42	41	37	26	42
4	82	76	66	66	55	27	37	48
5	97	86	49	68	55	46	44	42
6	72	91	83	62	55	64	35	57
7	67	101	76	49	61	56	35	43
8	71	90	81	76	59	51	38	71
9	87	83	100	85	63	63	44	89
10	51	66	71	57	73	57	50	54
11	50	53	68	45	65	63	46	67
12	50	47	61	51	44	28	23	26

3.1 Data Description:

When we look at the table (3), we see that the highest number of traffic accidents is (475) in October 2019 and the lowest number is (124) in April 2020. And the highest mortality rate (101) in July 2015 and the lowest (23) people in February 2020, the annual average number of traffic accidents is (350.771) and the number of deaths is (56.7813).

Table 3: Descriptive data for number of Traffic accidents and number of Death Accidents

	Average	Standard deviation	Maximum	Minimum
Traffic Accidents	350.771	71.2121	475.0	124.0
Death Accidents	56.7813	18.0363	101.0	23.0

3.2 Analyzing Time Series:

Analyzing of time series for Traffic Accidents and Death Accidents in order to obtain the best model. There are several steps of achieving it:

3.3 Plotting the time series data

The first step in time series analysis is drawing the series in order to identify some of the initial characteristics over time, such as oscillations, fluctuations, and note if there is a general trend or not. Every year, these variations occur on a regular basis but at a different rate. This pattern can also be recognized through the variations that suggest the existence of a seasonal component that repeats itself every 12 months Figure (1).

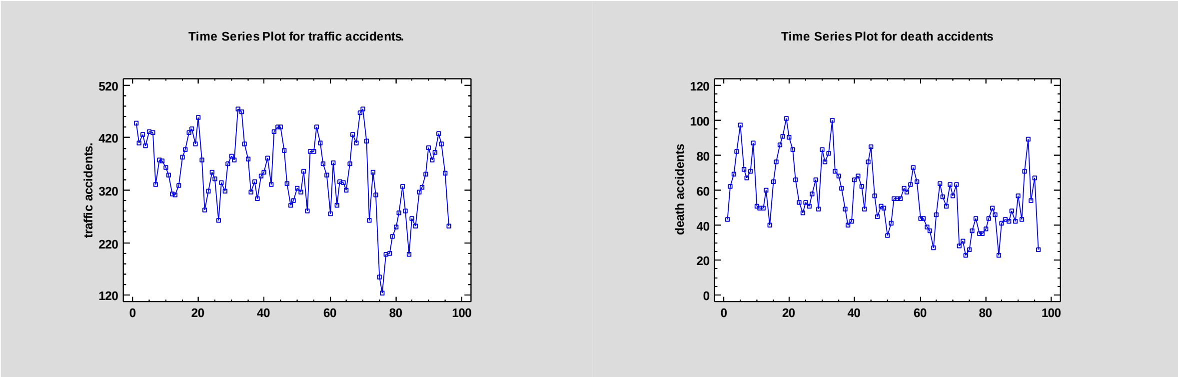

Figure 1: Time series plot of the Original Data of Traffic Accidents and Death Accidents in Kurdistan

To prove of ascertaining about the stationary of the time series, we have drawn the Autocorrelation and partial Autocorrelation functions, as shown in Figure (2) and (3):

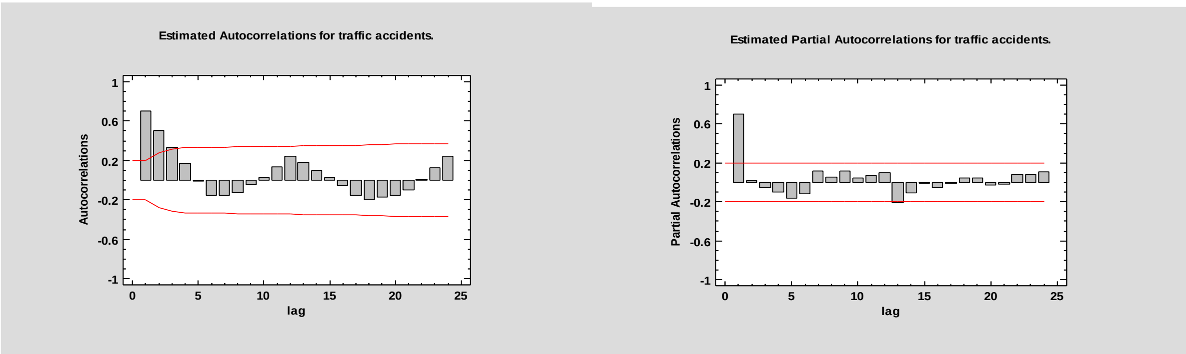

Figure 2: ACF and PACF of the Original Data Traffic Accidents of Monthly Time Series

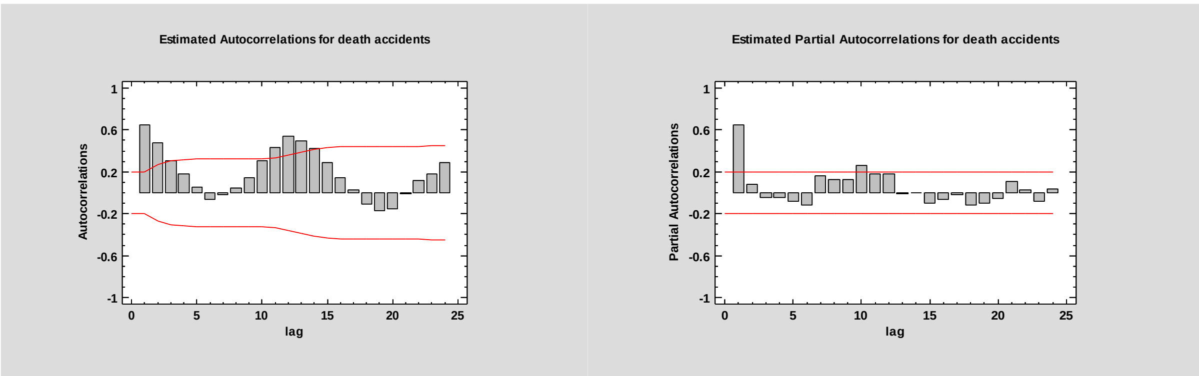

Figure 3: ACF and PACF of the Original Data Death Accidents of Monthly Time Series

In Figures (2) and (3), we can see highs and lows in the value of ACF contributions. For traffic accidents in this case, 3 of the 24 autocorrelation coefficients are statistically significant at the 95.0% confidence level. On the other hand, the PACF shows a large peak at first lag with a rapid decline thereafter, at the 95.0% percent level of confidence, 2 of the 24 auto correlation coefficients in this instance are statistically significant. At the 95.0 % percent confidence level, 7 of the 24 auto - correlation coefficients in this case are statistically significant for death accidents. On the other hand, the PACF shows a large peak at first lag with a rapid decline thereafter, In this instance, the 95.0% percent level of confidence identifies 2 of the 24 partial autocorrelation coefficients as statistically significant. It implies that the time series might not be fully random (white noise). This suggests a non-stationary It can be plainly seen that the data is non-stationary and contains seasonal by the behavior of autocorrelation functions and partial autocorrelation functions. Given that there is a seasonal effect, the Box-Pierce test rejects the null hypothesis, according to which all autocorrelation function coefficients are equal to zero (P-value = 0.0).

3.4 Verifying stationary:

By modifying data until it seems stationary, stationary data is the first stage of the Box-Jenkins strategy to building models.to get stationary of the series and after doing many trials, we conclude that.

a-For Traffic Accidents, the best procedure is to take the first difference for non-seasonal and The removal of a non-stationary about mean (trend) and variance is the first difference between seasonality and Periodic log transform, as presented in Figures (4) and (5) respectively. The value of the Box-Pierce statistic was 32.3182 (p-value = 0.119295) confirming the stationarity and the randomness of the series. Since the P-value for this test is greater than to 0.05, we cannot reject the hypothesis that the series is random at the 95.0% or higher confidence level.

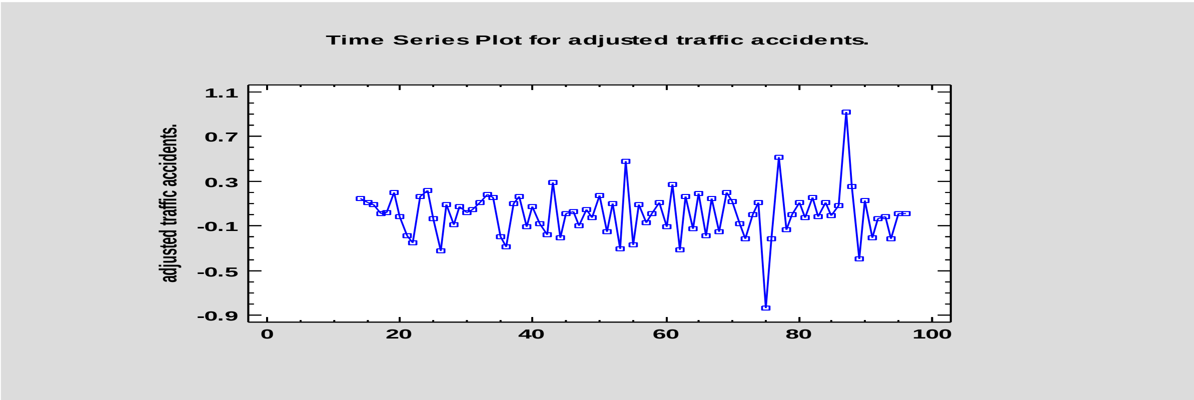

Figure 4: The Time Series After First Non-Seasonal Difference and the Natural Logarithm of Monthly Traffic Accidents

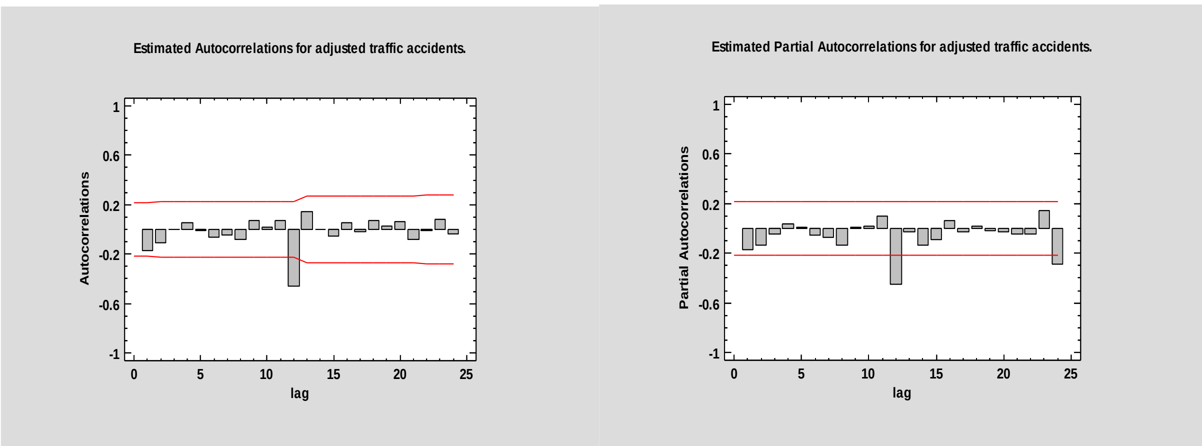

Figure 5: ACF PACF of the Monthly Traffic Accidents After First Non-Seasonal Difference and the Natural Logarithm

b. For Death Accidents, the ideal method is to remove a non-stationary about mean (trend) and variance by taking the initial difference for non-seasonal and power transformations. As presented in Figures (6) and (7) respectively. The value of the Box-Pierce statistic was 34.3257 (p-value = 0.0790443) confirming the stationarity and the randomness of the series.

We cannot rule out the null hypothesis that all autocorrelation function coefficients are equal to zero because the P-value for this test is larger than or equal to 0.05.

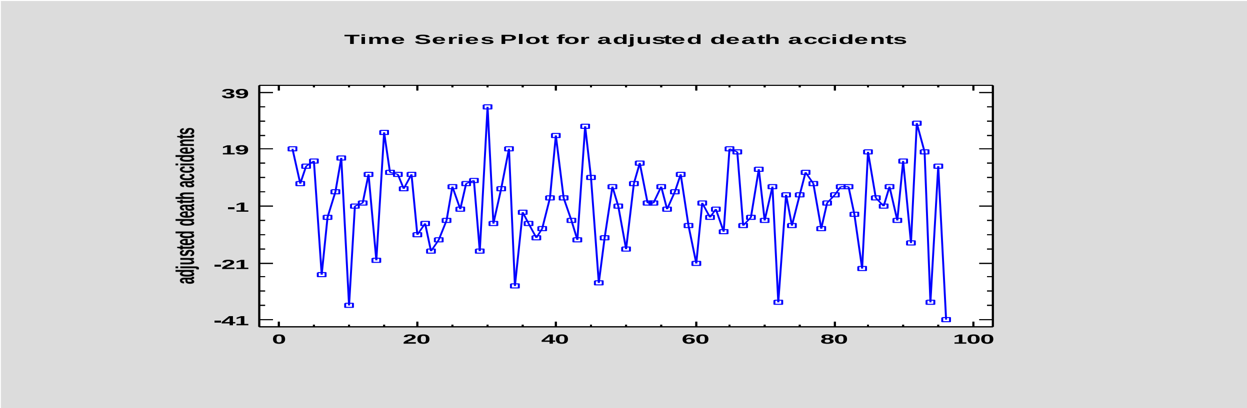

Figure 6: The Time Series After First Difference and the power transformation of Monthly Death Accidents

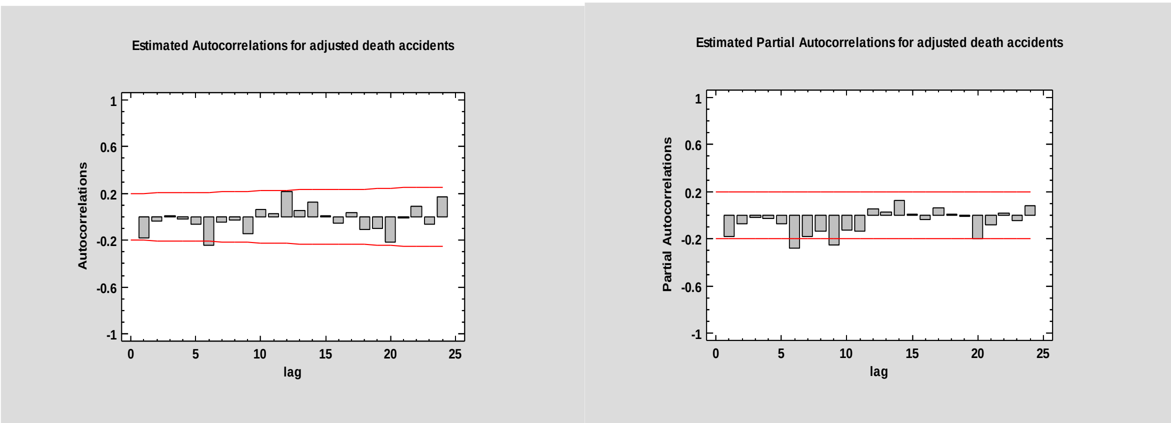

Figure 7: ACF PACF of the Monthly Death Accidents After First Non-Seasonal Difference and the power transformation

3.5 Choosing Fitting Model:

For fitting ARIMA models to time series, Box and Jenkins describe a development has been widely. After obtaining stationary, we proceed to build a suitable model for the adjusted series. Model selection was made by depending on partial autocorrelation and the regression coefficient. We apply the three performance measurements; RMSE, MAE, and MAPE as mentioned in the theoretical section to select the best model from a set of adequate models. Table (4.1) and (5.1) shows different models of SARIMA and the values of the estimated criteria.

a.For Traffic Accidents, From Table 4.1, it is clear that the best and adequate model is SARIMA(1,1,1)(0,1,1)$_{12}$ having the smallest values of criteria compared with the others. The parameters estimation of the specified model is presented in table 4.2 shows that all parameters of non-seasonal and seasonal components are statistically significant.

Table (4.1) Proposed models for monthly Traffic Accidents

Model	RMSE	MAE	MAPE	Significant Parameters
SARIMA(1,1,0)x(1,1,2)$_{12}$	48.3026	33.4041	11.2234	No
SARIMA(1,1,1)x(0,1,1)$_{12}$	**46.4507**	**32.7771**	**11.1746**	**Yes**
SARIMA(1,1,1)x(1,1,1)$_{12}$	46.8564	33.0435	11.2779	No

SARIMA(1,1,1)x(2,1,1)$_{12}$	47.0616	33.0184	11.2968	No
SARIMA(1,1,2)x(1,1,1)$_{12}$	47.4745	33.4857	11.225	No
SARIMA(0,1,1)x(1,1,2)$_{12}$	47.7304	32.9448	11.1628	No

Table 4.2: Parameter Estimation values of the model SARIMA (1,1,1)(0,1,1)$_{12}$ Model
Estimate Model Coefficients

Parameter	Estimate	Standard. Error	t	P-value
AR(1)	0.707606	0.111269	6.35942	0.000000
MA(1)	0.924092	0.0510585	18.0987	0.000000
SMA(1)	0.852648	0.0483174	17.6468	0.000000

b. For Death Accidents, From Table 5.1, it is clear that the best and adequate model is SARIMA(0,1,1)(1,1,2)$_{12}$ having the smallest values of criteria compared with the others. The parameters estimation of the specified model is presented in table 5.2 shows that all parameters of non-seasonal and seasonal components are statistically significant. Positive numbers mean an increase in the death rate and negative numbers mean a decrease in the death rate due to traffic accidents according to the seasons of the year. In parameter estimation, positive numbers mean increased mortality and negative numbers mean decreased mortality in a given season.

Table (5.1) Proposed models for monthly Death Accidents

Model	RMSE	MAE	MAPE	Significant Parameters
SARIMA(1,1,0)x(1,1,2)$_{12}$	11.3118	9.01799	17.6966	Yes
SARIMA(1,1,1)x(0,1,2)$_{12}$	11.3253	8.89415	18.0937	No
SARIMA(1,1,1)x(1,1,1)$_{12}$	11.5441	9.18538	19.1122	No
SARIMA(1,1,1)x(2,1,2)$_{12}$	10.6099	8.50378	17.2767	Yes
SARIMA(1,1,2)x(1,1,1)$_{12}$	11.4022	8.90565	18.0572	No
SARIMA(0,1,1)x(1,1,2)$_{12}$	**10.5262**	**8.41637**	**16.4533**	**Yes**

Table 5.2: Parameter Estimation values of the model SARIMA (0,1,1)$(1,1,2)_{12}$ Model
Estimate Model Coefficients

Parameter	Estimate	Standard Error	t	P-value
MA(1)	0.618383	0.0872545	7.08712	0.000000
SAR(1)	-0.840888	0.129209	-6.50796	0.000000
SMA(1)	-0.185271	0.0777796	-2.382	0.019624
SMA(2)	0.868699	0.0525907	16.5181	0.000000

3.6 Model Diagnostic Checking:

After identifying and estimating the potential SARIMA models, we want to evaluate how well the chosen models fit the data. Analyzing parameters and residuals is part of the model diagnostic checking process.

From the residual plots of ACF and PACF, as shown in Figures (8) and (9), for traffic accidents and death accidents it is clear that all values of residuals at all lags were placed within the tolerance interval at 95% confidence limits. This explains why there is no discernible association between the residuals and why each residual has a minor relationship to its standard error. On the other hand, the cost of a traffic collision is the Box-Pierce test was 15.952 (p-value = 0.77234). And for death accidents the value of the Box-Pierce test was 21.0193 (p-value = 0.395996), for this tests are greater than (0.05), concluding that the hypothesis of white noise at the 95% or higher confidence level cannot be rejected.

It is a suitable model for these data because this indicates that residuals are randomized white noise.

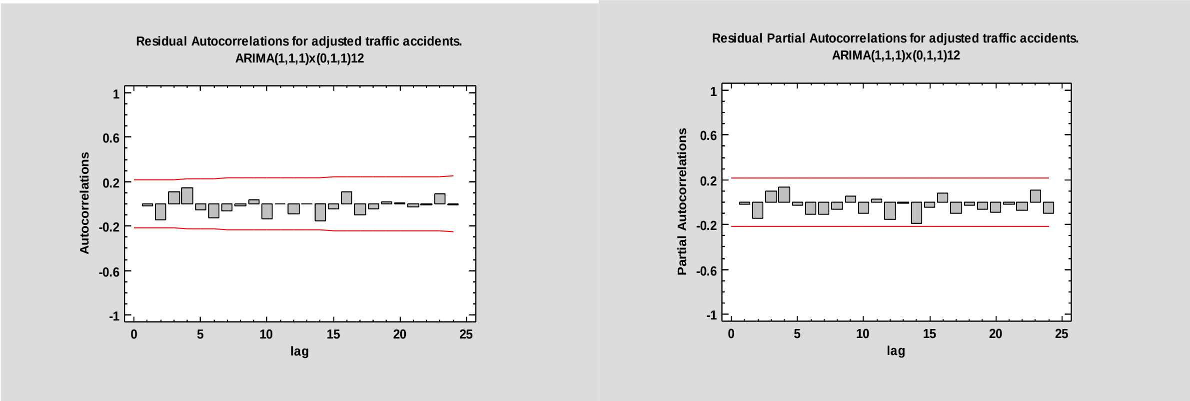

Figure 8: The residuals ACF and PACF for SARIMA (1,1,1)(0,1,1)12 for Traffic Accidents

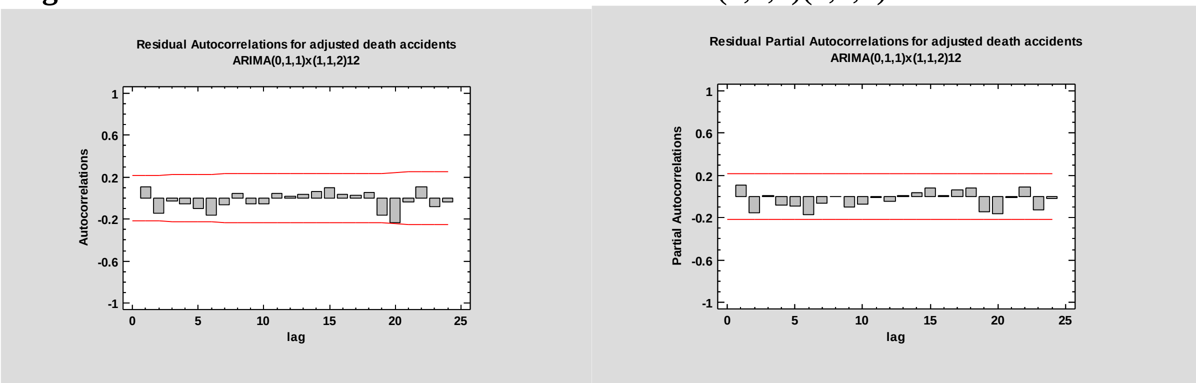

Figure 9: The residuals ACF and PACF for SARIMA (0,1,1) (1,1,2)12 for Death Accidents

3.7 Forecasting:

After going through the steps of identifying the appropriate models of data series traffic accidents and death accidents, assessment of its features and examination of the models we use the models to predict future values (Traffic incidents and death accidents for the coming months), forecasting starting from January 2022 until December 2022. The lower forecast limits and upper forecast for the forecasting under 95% confidence limits. All of the data that was available between January 1, 2014, and December 31, 2021, was used to develop the SARIMA model for forecasting purposes. The predicting monthly is shown in Table (6) and Figure (10) (Traffic incidents and death accidents for the coming months) in Kurdistan Region.

Table 6: Forecasting Values Using

SARIMA (1,1,1)(0,1,1)12 for traffic accidents — SARIMA (0,1,1)(1,1,2)12 for death accidents

Period	Forecast	Lower 95% Limit	Upper 95% Limit		Period	Forecast	Lower 95% Limit	Upper 95% Limit
1/2022	302.129	217.794	419.12		1\ 2022	41.8692	17.951	65.7875
2/2022	276.06	182.147	418.394		2\ 2022	36.035	10.4343	61.6357
3/2022	283.713	178.358	451.302		3\ 2022	40.7071	13.5279	67.8863
4/2022	279.477	170.427	458.304		4\ 2022	49.5878	20.9169	78.2587
5/2022	305.07	182.165	510.9		5\ 2022	51.7182	21.6295	81.807
6/2022	330.93	194.539	562.943		6\ 2022	58.06	26.6173	89.5028
7/2022	337.629	196.034	581.498		7\ 2022	52.8136	20.0728	85.5544
8/2022	364.328	209.361	633.998		8\ 2022	63.0948	29.1055	97.0841
9/2022	374.081	213.043	656.849		9\ 2022	73.024	37.8304	108.218
10/2022	356.264	201.262	630.642		10\ 2022	55.6174	19.2595	91.9754
11/2022	324.054	181.706	577.92		11\ 2022	55.5958	18.1096	93.082
12/2022	254.448	141.678	456.979		12\ 2022	33.6009	0.000	72.1823

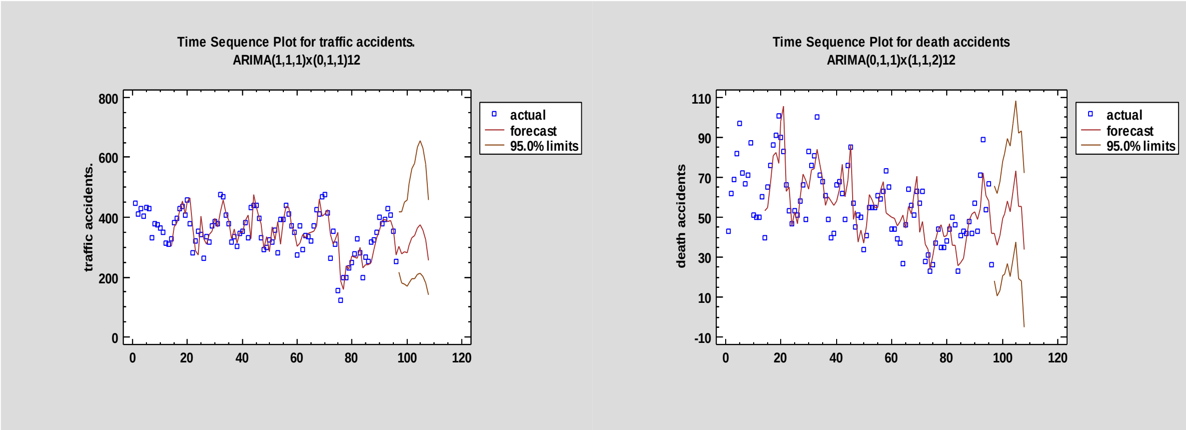

Figure 10: The Actual, Predicted and Forecasted Monthly Data of traffic accident People Using SARIMA (1,1,1)(0,1,1)12 Model and death accident People Using SARIMA (0,1,1)(1,1,2)12 Model in 2022

Conclusion:

Based on the results obtained in this study, we might come to the conclusion that there are seasonal patterns in the time series of traffic fatalities and accidents in the Kurdistan Region of Iraq. Also, after testing several models, we selected the two best models, SARIMA (1,1,1)(0,1,1)12 for traffic accidents and SARIMA (0,1,1)(1,1,2)12 for the number of deaths according to some statistical criteria such as (RMSE,MAE,MAPE).

Forecasting the number of traffic accidents is an important part of traffic management in a given area, can have several positive effects on the area and its citizens. The problematic months and seasons of the year can be identified to prompt the relevant authorities to take action to combat the problems through changes in the road safety strategy.

Using the selected models, we forecast for the next 12 months and found that the number of accidents and deaths did not decrease significantly during the forecast period, and we found that most of the traffic accidents occur in spring and summer because people go to the beaches in spring and in summer because it is school holidays people travel more between cities and towns because the roads in the Kurdistan Region are bad and between some cities the roads are one-side, so there are more traffic accidents in these two seasons than in the other two seasons.

Finally, our suggestion to the Traffic Directorate is to pay more attention to traffic guidelines and conditions and our suggestion to the Ministry of Municipalities is to make the roads two-sided and repair the roads that exist between the provinces and major cities to reduce congestion and reduce traffic accidents.

References:

1- Ahmed, N.M., 2007. *Forecasting Water Inflow and Electric Power Generation in Darbandi-khan Station By Using FFNN, and Adaptive Method In Time Series* (Doctoral dissertation, Phd. Thesis of Philosophy in Statistics, University of Sulaimani, Kurdistan Region–Iraq).
2- Al-Ghamdi, A.S., 1995. Time series forecasts for traffic accidents, injuries, and fatalities in Saudi Arabia. *Journal of King Saud University-Engineering Sciences*, *7*(2), pp.199-217.
3- Brockwell, P.J. and Davis, R.A., 2009. *Time series: theory and methods*. Springer Science & Business Media.
4- Chatfield, C., 2003. *The analysis of time series: an introduction*. Chapman and hall/CRC.
5- Chatfield, C. and Xing, H., 2019. *The analysis of time series: an introduction with R*. Chapman and hall/CRC.
6- Hyndman, R.J. and Athanasopoulos, G., 2018. *Forecasting: principles and practice*. OTexts.
7- Kan, M.S., Tan, A.C. and Mathew, J., 2015. A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing*, *62*, pp.1-20.
8- Kassem, A.A., Raheem, A.M. and Khidir, K.M., 2020. Daily Streamflow Prediction for Khazir River Basin Using ARIMA and ANN Models. *Zanco Journal of Pure and Applied Sciences*, *32*(3), pp.30-39.
9- Kitagawa, G., 2010. *Introduction to time series modeling*. Chapman and Hall/CRC.

10- Mannering, F.L. and Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. *Analytic methods in accident research*, *1*, pp.1-22. Depending on the data of.

11- Metcalfe, A.V. and Cowpertwait, P.S., 2009. *Introductory time series with R* (p. 2). Springer-Verlag New York.

12- Montgomery, D.C., Jennings, C.L. and Kulahci, M., 2015. *Introduction to time series analysis and forecasting*. John Wiley & Sons.

13- Palit, A.K. and Popovic, D., 2006. *Computational intelligence in time series forecasting: theory and engineering applications*. Springer Science & Business Media.

14- Paolella, M.S., 2018. Linear models and time-series analysis: regression, ANOVA, ARMA and GARCH. John Wiley & Sons.

15- Tadesse, K. and Dinka, M., 2017. Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa. *Journal of water and land development*.

16- Wei, W.W., 2006. Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.

دروستكردنی مۆدێلی SARIMA بۆ شیكردنەوە و پێشبینیكردنی داتای زنجیرە كاتییەكانی ڕووداوەكانی هاتوچۆی ڕێگاوبان لە هەرێمی كوردستان

مريم محمد احمد
بەشی ئامار و زانیاریەكان، كۆلێژی كارگێڕی و ئابووری
زانكۆی سەڵاحەدین-هەولێر
Mryam.Ahmed@su.edu.krd

سامان حسين محمود
بەشی ئامار و زانیاریەكان، كۆلێژی كارگێڕی و ئابووری
زانكۆی سەڵاحەدین-هەولێر
Saman.Mahmood@su.edu.krd

پوختە

ڕووداوەكانی هاتووچۆ دەبێتە هۆی بریندابوون و مردن، هەروەها زیانی ماددیش بە كۆمەڵگا دەگەیەنێت. بۆیە پێشبینیكردن و دەستنیشانكردنی هۆكارەكانی ڕووداوی هاتووچۆ پێویست و گرنگە بۆ كەمكردنەوەی ئە و زیانانە. ئامانجی سەرەكی ئە م توێژینەوەیە بریتییە لە پەرەپێدانی مۆدێلی زنجیرە كاتییەكەی ئاریما بۆ بەدواداچوون و شیكردنەوەی ژمارەی ڕووداوەكانی هاتووچۆ و ژمارەی مردنی خەڵك بە هۆی ڕووداوی هاتووچۆوە لە هەرێمی كوردستانی عراق بەپێی مانگانە لە ماوەی ساڵانی (٢٠١٤-٢٠٢١) و پێشبینیكردنی مانگانە بۆ ساڵی ٢٠٢٢. داتاكان لە بەڕێوەبەرایاتی گشتی هاتووچۆی هەولێر وەرگیراون. ئەنجامەكان دەریانخستووە كە زنجیرەكە تایبەتمەندی وەرزی هەیە. دوای تاقیكردنەوەی چەند مۆدێلێك و هەڵبژاردنی باشترین ئەنجام بە پشتبەستن بە كەمترین پێوەرە ئاماریەكان (RMSE, MAE, MAPE) كە بۆ بەراوردكردن بەكارهێنراوە. باشترین ئەنجامەكان ئەوە بوو كە ئێمە مۆدێلی SARIMA (1,1,1)(0,1,1)12مان دۆزیەوە بۆ ژمارەی ڕووداوەكان و مۆدێلی SARIMA (0,1,1)(1,1,2)12 بۆ ژمارەی مردن.

لە كۆتاییدا بە بەكارهێنانی باشترین مۆدێلەكان، مانگانە پێشبینیمان كرد بۆ ژمارەی ڕووداوەكان و ژمارەی مردنەكان. دەتوانین بڵێین بۆ ماوەی پێشبینیكراو ڕێژەكە بە شێوەیەكی بەرچاو كەمی نەكردووە، بۆیە پێویستە حكومەت پلانی باشتر و وردتر دابڕێژێت بۆ كەمكردنەوەی ڕووداوەكانی هاتووچۆ.

وشە گرنگەكان: SARIMA، پێشبینیكردن، ڕووداوی هاتوچۆ .

بناء نماذج SARIMA لتحليل وتوقع بيانات السلاسل الزمنية لحوادث الطرق في إقليم كردستان العراق

مريم محمد احمد
قسم الاحصاء والمعلوماتية، كلية الادرة والاقتصاد
جامعة صلاح الدين-اربيل
Mryam.Ahmed@su.edu.krd

سامان حسين محمود
قسم الاحصاء والمعلوماتية، كلية الادرة والاقتصاد
جامعة صلاح الدين-اربيل
Saman.Mahmood@su.edu.krd

ملخص

تتسبب حوادث المرور في حدوث إصابات ووفيات ، ولكنها تسبب أيضًا أضرارًا مادية للمجتمع. لذلك ، فإن توقع وتحديد أسباب الحوادث المرورية أمر ضروري ومهم للحد من هذه الخسائر. الهدف الرئيسي من هذه الدراسة هو تطوير نموذج سلسلة زمنية ARIMA للتحقيق في عدد الحوادث المرورية وعدد الوفيات في حوادث المرور في إقليم كردستان العراق وتحليلها وفقًا للاشهر خلال السنوات (2014-2021) والتنبؤ بالمعدل الشهرى لسنة 2022. تم الحصول على البيانات من المديرية العامة للمرور أربيل. أظهرت النتائج أن السلسلة الزمنية ليس له خصائص موسمية. بعد اختبار عدة نماذج واختيار أفضل النتائج بالأعتماد على المعايير الإحصائية (RMSE، MAE، MAPE) المستخدمة للمقارنة. تم اختيار أفضل نموذج SARIMA (1، 1، 0) (1، 1، 1)$_{12}$ لعدد الحوادث ونموذج SARIMA (0، 1، 1) (1، 1، 2)$_{12}$ لعدد حالات الوفاة.

أخيراً ، باستخدام أفضل النماذج ، قمنا بعمل تنبؤات شهرية لعدد الحوادث وعدد الوفيات. يمكننا القول أن المعدل لم ينخفض بشكل كبير لفترة التنبؤ ، لذلك يجب على الحكومة وضع خطط أفضل و أكثر تفصيلاً لتقليل حوادث المرور.

الكلمات المفتاحية:SARIMA ، التنبؤ ، الحوادث المرورية.