Asian Journal of Probability and Statistics

11(2): 12-20, 2021; Article no.AJPAS.65335

ISSN: 2582-0230

Time Series Analysis of Road Traffic Accidents in Ghana

Frank B. K. Twenefour[1*], Emmanuel Ayitey[2], Justice Kangah[3] and Lewis Brew[4]

[1]Department of Mathematics, Statistics and Actuarial Science, Takoradi Technical University, P.O.Box 256, Takoradi, Ghana.
[2]Department of Mathematics, Adiembra Senior High School, P.O.Box 16, Sekondi, Western Region, Ghana.
[3]Department of Mathematics, Shama Senior High School, P.O.Box 30, Shama, Western Region, Ghana.
[4]Department of Mathematics, University of Mines and Technology, P.O.Box 237, Tarkwa, Ghana.

Authors' contributions

This article was carried out in collaboration among all authors. Authors JK, EA and FBKT designed the study, performed the statistical analysis and wrote the protocol and the first draft of the manuscript. Author LB managed the analysis and proof read the manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v11i230262
Editor(s):
(1) Dr. Belkacem Chaouchi, Khemis Miliana University, Algeria.
Reviewers:
(1) Nancy Awadallah Awad, Sadat Academy for Management Sciences, Egypt.
(2) Revati Shriram, Cummins College of Engineering for Women, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/65335

Review Article

Received: 05 December 2020
Accepted: 11 February 2021
Published: 04 March 2021

Abstract

This study uses Time Series models to predict the annual traffic accidents in Ghana. The traffic accidents data spanning from January 1990 to December 2019 was used. The Box-Jenkins model building strategy was used. The Augmented Dickey Fuller (ADF) test showed that the accident data was stationary. Three ARMA models were suggested based on the ACF and PACF plots of the differenced series, these were ARMA (0,0), ARIMA (1,0), and ARMA (2,0). The model with the smallest corrected Akaike Information Criteria (AICs) and Bayesian Information Criteria (BIC) was chosen as the best model. The Ljung-Box statistics among others were used in assessing the quality of the model. ARMA (1,0) was the best model for the Ghana annual Traffic Accident data. The results showed that, from January to July, it would be difficult to make accurate estimates of the number of road incidents for the years leading up to 2020. This was due to the fact that the white noise process values were statistically independent at various times.

Keywords: ARMA; Box-Jenkins model; ACF; ADF; AIC; BIC; PACF; Ghana annual traffic accident data; time series model.

**Corresponding author: E-mail: frank.twenefour@ttu.edu.gh;*

1 Introduction

Road accident has always been an issue of concern in Ghana, Africa and the world at large. According to [1], road accidents account for more than one million deaths per year in the world. The number of vehicles purchased by Ghanaians has increased annually since 2009 due to the dollarization of the economy, which has made it easy for individuals to purchase second-hand vehicles mainly from other countries, in this publication, road traffic accidents were listed as one of the world's top ten causes of death [2]. According to [1,3] due to the casual attitude of the Government of Ghana to the construction of roads and the violation of basic traffic rules by motor users in the region, the number of road accident cases has increased in recent years. Generally, there are many causes of road traffic accidents, among them include human error, vehicle conditions, road environment, over-speeding and road users [4]. Ghana does not have specific laws governing driving experience of our public roads for motor drivers [5]. Policies on reckless driving are often seldom applied. The implementation of such measures against drivers of public motor vehicles has been jeopardized by corruption in the police force as well as in the numerous drivers' unions in the country [6,7].

Until recently, when some state governments banned their use on major roads in Ghana, the use of motorcycles had inflated the number of road traffic accident cases [8,9]. If governments do not pay attention to road construction and have strict punishments against road traffic violators, these cases will escalate to an alarming level [10,9].

Abane has estimated a rise in road traffic accidents in Ghana [8,11]. According to his report, if proper changes or tougher regulations were not enforced in the near future, then the death rate on our roads will rise from 21,548 to 31,404 within two years: from 2011 to 2013 [8,9]. [4,12] stated that the rise in casualty death in road accident can be attributed to a number of factors like: tiredness on the part of drivers, neglect of road safety protocols, break failures on roads, etc. According to [13,14], the box - Jenkins Analysis refers to a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average (ARIMA) time series models. The method applies to the autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models which are basically used to find the best fit of a time-series model. It is usually appropriate for time series of medium to long length (at least 50 observations). The Box-Jenkins model has been shown to be efficient for analyzing data and forecasting by many authors. The Box-Jenkins model provides the sure-to-accurate approach as opposed to other forecasting models [15,16].

The Box-Jenkins model will be used in this paper to evaluate Ghana's monthly road accident data from January 1990 to December 2019 to assess the trends of road accident incidents, injuries and fatalities along the Accra-Tema motorway. The number of cases will be the priority.

2 Materials and Methods

2.1 Source of data

From January 1990 to December 2019, the data was obtained from the National Road and Safety Commission. The information focuses on recorded traffic accidents. SPSS and R applications are used in the analysis.

2.2 The Box-Jenkins model

The Box-Jenkins time series model was employed. Unit root tests were used to verify the stationary nature of the data and, after the first test, the data proved to be a stationary series (Z_t). The Box-Jenkins ARMA models and a mixture of AR and MA were utilized in this analysis. The model ARMA (p, q) conforms to the relationship below [6].

$$\emptyset(B)(1-B)^d Z_t = \theta(B)a_t$$

Where

$$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - - - \emptyset_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - - - \theta_p B^p$$

Where B is the backshift operator, p is the number of autoregressive parameters in the model and q represents the number of moving average parameters in the same model. The autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to determine the trial time series models (model formulation). The parameters were estimated by the R software. The trial model(s) were then subjected to rigorous diagnostic checks using residuals: ACF and PACF plots of residuals [15,16]. The Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) were used to select the best model [15]. The final model was then used to make forecasts for the year 2020.

2.3 The ARMA model

According to [16,6,10], The ARMA (autoregressive, moving average) model is defined as follows:

$$X_t = \emptyset_1 X_{t-1} + \cdots + \emptyset_p X_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Where the autoregressive parameters to be calculated are the $\emptyset's$ (thetas), the moving average parameters to be estimated are the $X's$, and the $a's$ are a series of unknown random errors (or residuals) that are believed to obey the normal distribution of probability [15,16].

The backshift operator is used by Box-Jenkins to promote the writing of these models. The reverse shift operator, B, has the effect of shifting the time duration from t to t-1. Thus $BX_t = X_{t-1}$ and $B^2 X_t = X_{t-2}$. Using this backshift notation, the above model may be rewritten as:

$$\left(1 - \emptyset_1 B - \cdots - \emptyset_p B^p\right)X_t = \left(1 - \theta_1 B - \cdots - \theta_p B^q\right)a_t$$

This may be abbreviated even further by writing

$$\emptyset_p(B)X_t = \theta_q(B)a_t$$

Where,

$$\emptyset_p(B) = \left(1 - \emptyset_1 B - \cdots - \emptyset_p B^p\right)$$

and

$$\theta_q(B) = \left(1 - \theta_1 B - \cdots - \theta_p B^q\right)$$

These formulas show that the operators $\emptyset_p(B)$ and $\theta_q(B)$ are polynomials of B of orders p and q respectively.

3 Results and Discussion

Data Analysis and Tests for Stationarity: Total Reported Accidents.

Fig. 1 displays an Augmented Dickey Fuller (ADF) test for stationarity was performed on the data and showed a p-value of 0.01 suggesting further that the data is stationary. This means that the mean is constant and seems not to be changing with time. The ACF and PACF plots of the time series data is presented in Figs. 2 and 3 respectively.

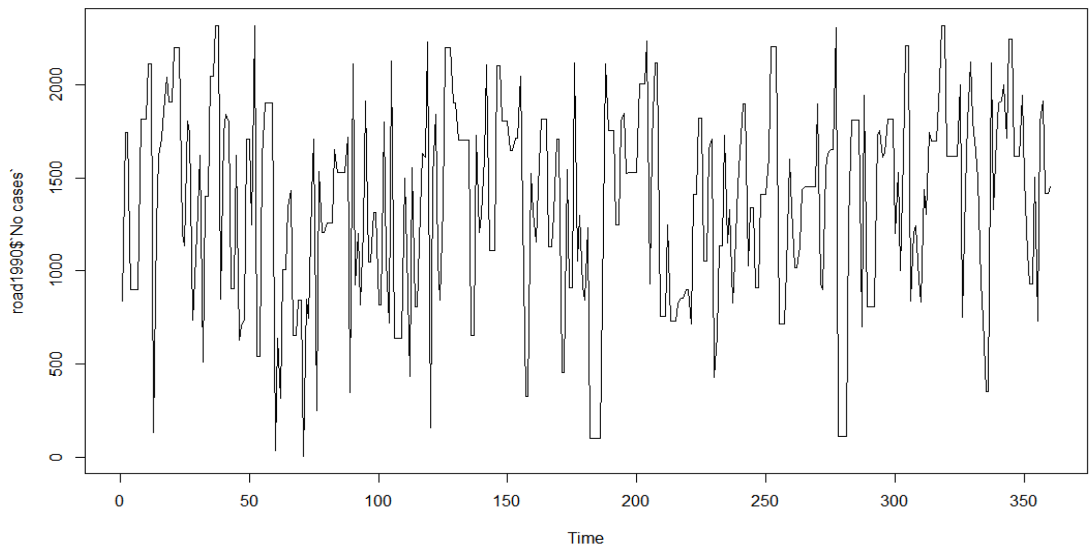

Fig. 1. Time series data of number of reported cases

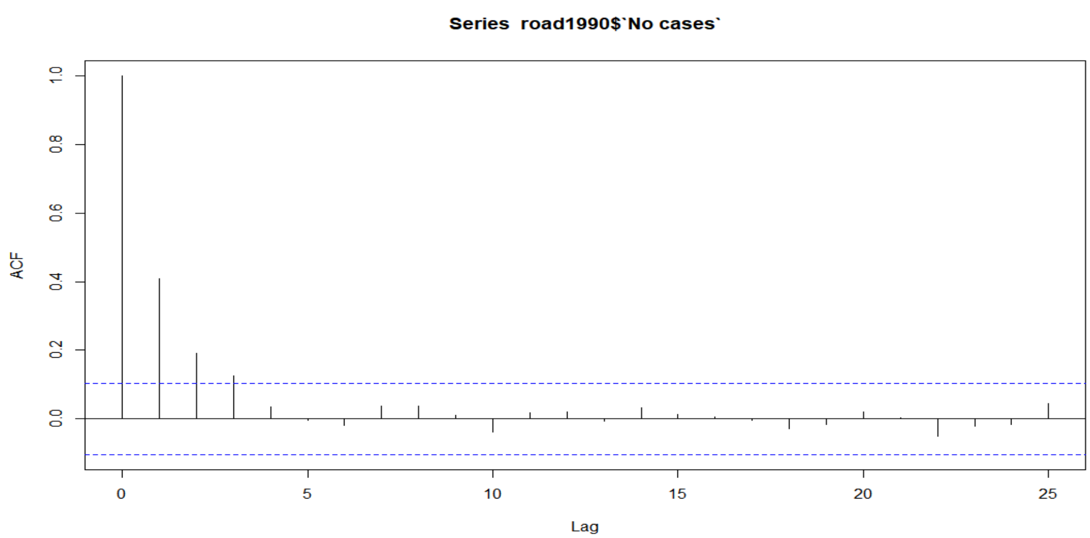

Fig. 2. ACF of number of reported cases

The ACF shows an exponential curve and a large lag 1 spike is seen in the PACF plots, indicating that the time series is not a white noise process, but an AR (1) model. Three different models have been tested, such as ARMA (0,0), ARMA (1,0) and lastly ARMA (2,0).

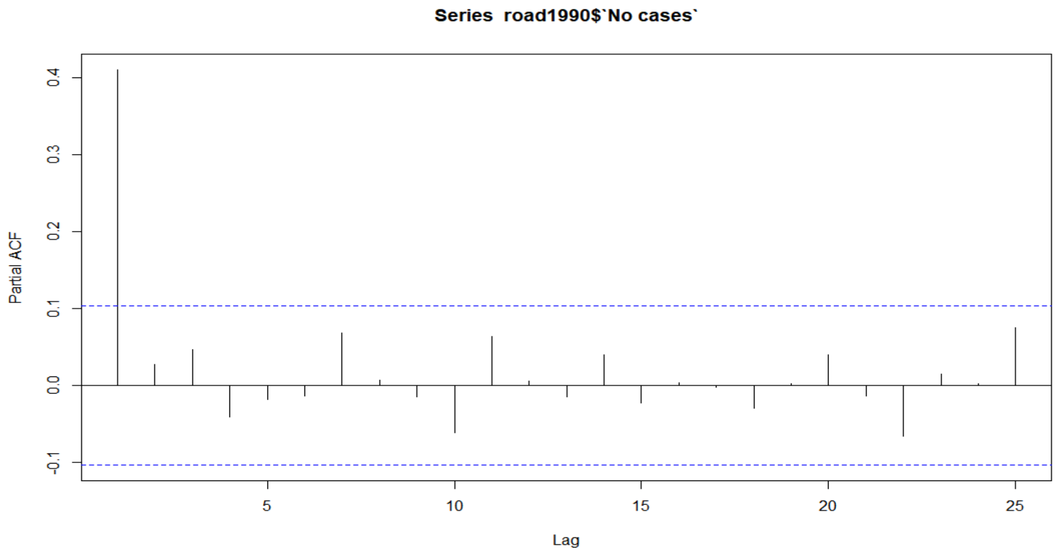

Fig. 3. PACF of number of reported cases

Table 1 below summarizes the outcomes of these models.

Table 1. Results of the analysis for the three models

Model	AICs
ARMA (0,0)	5,560.68
ARMA (1,0)	5,496.36
ARMA (2,0)	5,498.12

From Table 1 above, based on the corrected Akaike Information criteria (AICs), among the three proposed models, an ARMA (1, 0) has the lowest AICs value; that is 5,496.36. Since the lowest AICs is considered the best fit, ARMA (1,0) was thus chosen as the tentative model for the data. The Ljung Box test for ARMA residuals (1, 0) shows evidence for residual associations with p = 0.0052. An ADF test showing p= 0.01 shows that the residuals are stationary. The ACF and PACF both display a substantial lag spike of 1. ACFs or PACFs are relevant if they are more than $m = \frac{1.96}{\sqrt{k}}$ where k is the sample size.

In this study, the sample size $(k) = 361$

$$\therefore m = \frac{1.96}{\sqrt{361}} = 0.10315$$

Evidence on Table 2 shows that since the result of ACF at lag 1 exceeds 0.1031, it means that the ACFs are significant at lag 1. The ARMA (1, 0) model was thus confirmed as the final model for the data. The ACF plots of residuals for ARMA (1, 0) showing significant spike at lag 1 are presented in Figs. 4 and 5.

Table 2. ACF result at different lags

LAG	ACF
1	1.0
2	-0.01
3	0.00

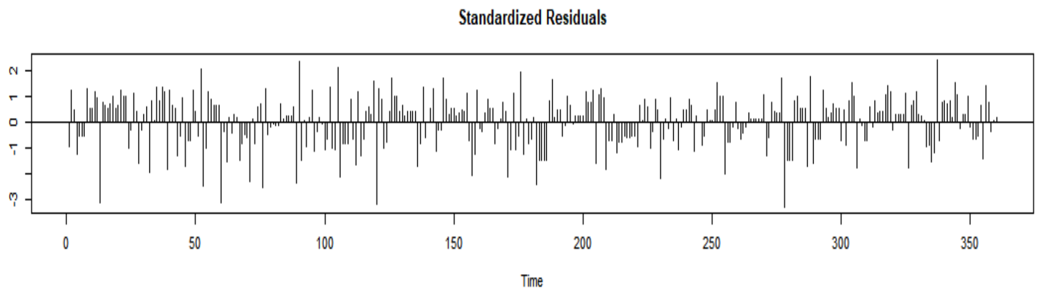

Fig. 4. Standardized Residual plot of ARMA (1, 0)

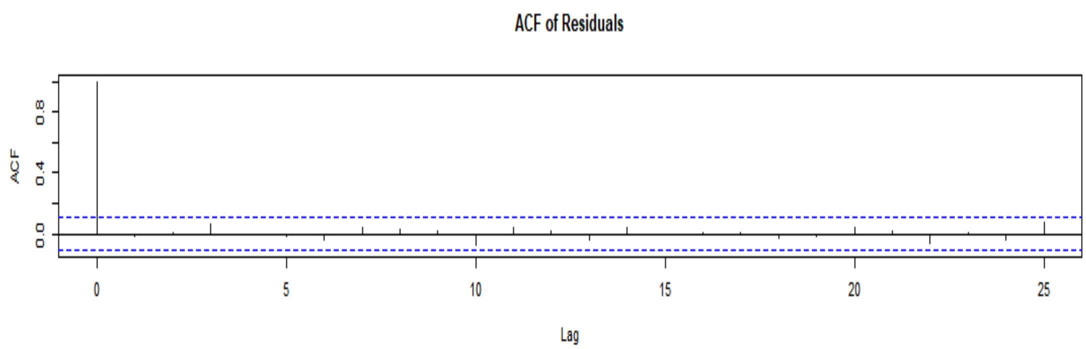

Fig. 5. ACF Residual plot of ARMA (1, 0)

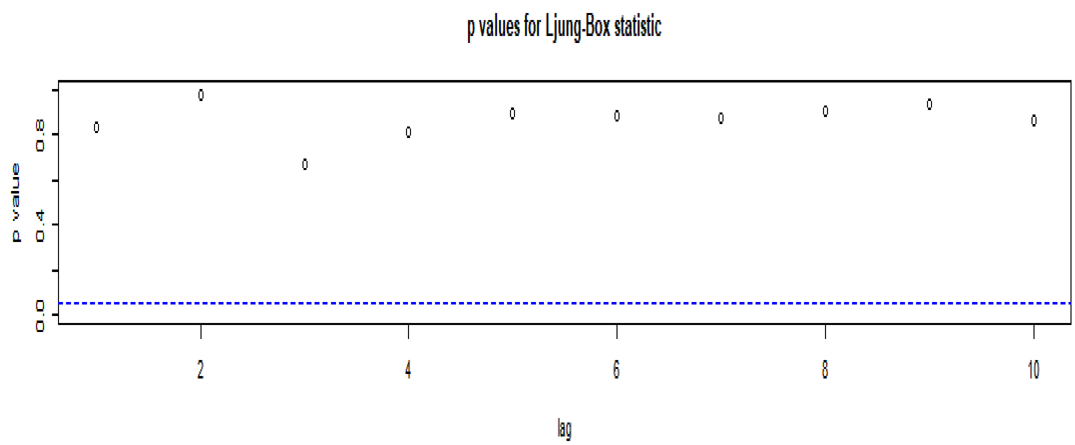

Fig. 6. The p-value for Ljung-Box statistic

4 Forecasting

Suppose Z_t is the accident data series, which in this case was found to be white noise implying that $Z_t = \alpha$. Fig. 7 and Table 3 shows the forecasts for Jan-2020 to July 2020 and confirms the above theory.

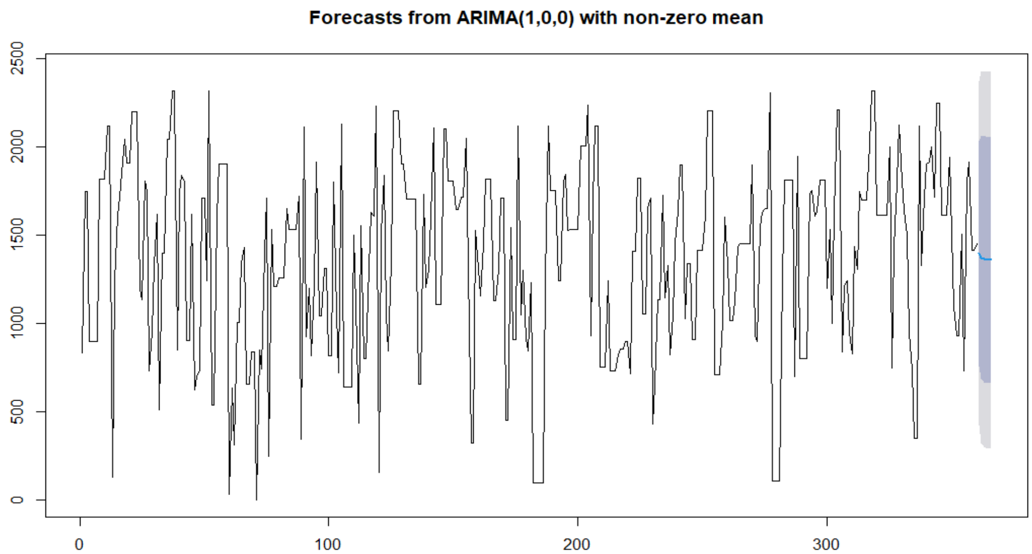

Fig. 7. Forecasting road accident data

Table 3. Forecasts and their corresponding 80% and 95% confidence limits

Year	Point forecast	80% LB	80%UB	95%LB	95%UB
Jan-2020	1396.935	761.4604	2032.409	425.0606	2368.809
Feb-2020	1396.935	761.4604	2032.409	425.0606	2368.809
Mar-2020	1396.935	761.4604	2032.409	425.0606	2368.809
Apr-2020	1396.935	761.4604	2032.409	425.0606	2368.809
May-2020	1396.935	761.4604	2032.409	425.0606	2368.809
June-2020	1396.935	761.4604	2032.409	425.0606	2368.809
July-2020	1396.935	761.4604	2032.409	425.0606	2368.809

Table 4. Model accuracy

MSE	RMSE	MAE
0.723	495.86	402.44

Tables 3 and 4 presents the forecasts and their corresponding confidence limits and model accuracy respectively. The point prediction errors appear to have a constant variance over time. Fig. 7 shows the ACF plot of the forecast errors with a large lag spike, implying a lag connection. It also means the model may be improved any further. The Box-Ljung statistic for the forecast errors is 0.723 with a p-value of 0.052 implying that there is evidence of autocorrelations in the point forecast.

5 Conclusion

From January 1990 to July 2019, this analysis used monthly road accident data from Ghana's national road safety commission. The data consisted of one group: total cases registered. Based on results from the Augmented Dickey Fuller (ADF) test, the knowledge was unit root stationary. The Box-Jenkins model building strategy (1976) was applied on series data. The ACF and PACF of the series data gave a guideline on tentative models to be used. A spike in both plots (ACF and PACF) was considered significant if it was greater than 0.1032. There were significant spikes in the ACF and PACF plots for the data sets at lag 1.

ARMA (0,0), ARMA (1,0), and ARMA (2,0) were suggested as three models. The ARMA model with the smallest AICs was chosen as the best model. In evaluating the consistency of the model, Ljung-Box statistics were also used. ARMA (1, 0) has emerged as the best model for Ghana's overall recorded cases of road accidents. Due to the nature of the model used which uses the value at the forecast origin as a forecast, forecasts for Jan-2020 to July-2020 yielded the same result.

The results of these findings are that it is difficult to create predictions using a white noise method due to the fact that the values are statistically independent at different times.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] WHO, Global Status Report on Road Safety 2018, WHO, Geneva, Switzerland; 2018. Available:https://apps.who.int/iris/bitstream/handle/10665/276462/9789241565684-eng.pdf?ua=1

[2] Corden M, Neary PJ. Booming sector and de-industrialisation in a small open economy. The Economic Journal. 1982;92(368):825–848.

[3] Avuglah RK, Adu-Poku KA, Harris E. Application of ARIMA models to road traffic accident cases in Ghana. International Journal of Statistics and Applications. 2014;4:233-239.

[4] Awialie Akaateba M, Amoh-Gyimah R. Driver attitude towards traffic safety violations and risk-taking behavior in Kumasi: The gender and age dimension. International Journal for Traffic & Transport Engineering. 2013;3(4).

[5] National Road Safety Commission. Road safety dialogue: Newsletter of the National Road Safety Commission (2010, Jan-Dec). Accra, Ghana; 2010.

[6] Ho SL, Xie M, Goh TN. A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. Computers & Industrial Engineering. 2002;42(2-4):371-375.

[7] Abdel-Aty M. Investigating the relationship between cellular phone use and traffic safety. ITE Journal. 2003;73(10):38.

[8] Abane AM. Travel behaviour in Ghana: Empirical observations from four metropolitan areas. Journal of Transport Geography. 2011;19(2):313–322.

[9] Odero W. Motorcycle injuries in east Africa: Magnitude, risk factors and prevention. In Proceedings of the RTIRN Regional Workshop, Accra, Ghana, December 2009.

[10] Harvey AC, Todd PHJ. Forecasting economic time series with structural and Box-Jenkins models: A case study. Journal of Business & Economic Statistics. 1983;1(4):299-307.

[11] Abane AM. Background characteristics and accident risk of commercial vehicle drivers in the Cape Coast- Elmina area of the Central Region of Ghana. The Oguaa Journal of Social Sciences. 2010; 5(1):64-86.

[12] Oppong RA. Statistical analysis of road accidents fatality in Ghana using Poisson regression. Doctoral dissertation.

[13] Makridakis S, Hibon M. ARMA models and the Box–Jenkins methodology. Journal of Forecasting. 1997;16(3):147-163.

[14] Balasubramanian V, Jagannath M. Detecting motorcycle rider local physical fatigue and discomfort using surface electromyography and seat interface pressure. Transportation Research Part F: Traffic Psychology and Behaviour. 2014;22:150–158.

[15] Lu Y, AbouRizk SM. Automated box–jenkins forecasting modelling. Automation in Construction. 2009;18(5):547-558.

[16] Weiss AA. ARMA models with ARCH errors. Journal of Time Series Analysis. 1984;5(2):129-143.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/65335