Environment and Ecology 36 (3) : 761—766, July—September 2018
Website: environmentandecology.com ISSN 0970-0420

Accidental Deaths in India : Forecasting with ARIMA Model

A. N. Patowary, B. Dutta, M. P. Barman, S. R. Gadde

Received 27 February 2018; Accepted 29 March 2018; Published on 18 April 2018

Abstract Accident is one of the burning problems for pre-mature end to human lives. Road accident in India is an increasing trouble and has raised one of the country's major problems. This paper outlines development of a conventional time series model viz. Autoregressive Integrated Moving Average (ARIMA) model for the annual total number of deaths due to accident (natural and unnatural) in India covering the period 1967 to 2015 and to forecast the number of annual accidental deaths likely to occur in future. We investigated and found that ARIMA (2, 2, 1) model is suitable for the given data set.

Keywords ARIMA, Forecasting, India, Accidental death.

A. N. Patowary*
College of Fisheries, Assam Agricultural University, Raha 782103, Assam, India
e-mail: a_patowary@yahoo.com

B. Dutta, M. P. Barman
Department of Statistics, Dibrugarh University, Dibrugarh 786004, India
e-mail: bornalidutta75@gmail.com
barman_manash@rediffmail.com

S. R. Gadde
Department of Statistics, Dodoma University, Tanzania
e-mail: gaddesrao@gmail.com
*Corresponding author

Introduction

An accident is an unexpected incident that results in injury, deaths, damage to property or some other losses. It is one of the vital reasons for pre-mature end to human lives. Injuries resulting from the accident make many people handicap and these adverse victims are important members of a family as well as the society. Finally, the lost due to accident is felt by a family and the society. Accidental deaths and injuries influence the economic development of a country as major portion of the victims are economically active which support the family for their livelihood. National Crime Record Bureau classified causes of accidental deaths into three categories viz. natural accidents, un-natural accidents and other causes of accidental death (sudden deaths with no apparent cause of death like heart attacks, death during pregnancy, poisoning). Due to the advantages of modern technology and medical discoveries, the proportion of accidental deaths due to natural causes (exposure to cold, starvation, epidemic, cyclone) is seem to be in decreasing trend. On the other hand, incidence of road accident is increases day by day. In India, every one hour nearly 1 person died due to cause attributable to force of nature while 38 persons killed in unnatural accidents in the year 2015. Out of 4,13, 457 accidental deaths, about 81.3% deaths are due to unnatural causes and 16.2% deaths are due to other causes and the remaining 2.5% are due to causes attributable to force of nature 2015 (National Crime Record Bureau 2015). Most of the victims of un-natural accidental deaths are in the age group of 18 to 45 years, accounting for 61.6% of total persons killed in such accidents in India during the year 2015 (National Crime Record Bureau 2015).

Road Traffic Accident (RTA) which is identified as the third most important cause of overall mortality and the main cause of death among the age group 15-44 year and it represents 12% of global burden of disease (Sivakumar and Krishnaraj 2015). High rate of road accidents and the resulted number of deaths and injuries is a major concern for India which adversely affects the social structure and economy. Considering the gravity of the situation, it is necessary to take appropriate measures for reducing this high level of deaths and injuries. Forecasting of accidents or number of deaths due to it may be very valuable information for taking different safety measures. Forecasting help the policy makers as well as engineers to investigate the safety system to prevent accidents and number of deaths due to it in the up-coming days.

Time series is a sequence of numerical data that generally occurs in uniform intervals over a period of time. The main application of time series analysis is forecasting by analyzed historical data (Monfared et al. 2013). Forecasting techniques is usually applied as an aid in controlling past and present operations in planning for future needs. Among the most effective approaches for analyzing time series data is the ARIMA (Autoregressive Integrated Moving Average) model introduced earlier.

There are extensive literatures available of applying ARIMA model in accidental data (mainly road accidents) in different countries of the world including India. Ofori et al. (2012) used Exponential Smoothing and ARIMA model to forecast road accident injuries in Ghana. Also, Quddus (2008) used a time series model to understand the trend of accidents and ascertain the viability of applying ARIMA models on data from Taybad city. Moreover, Avuglah et al. (2014) applies the ARIMA time series model to study the trends and patterns of road traffic accident cases in Ghana as well as makes a five-year forecast. The results reveal that ARIMA (0, 2, 1) is the best model to analyze the road traffic accident in Ghana. Similarly, Adebola et al. (2015) used ARIMA model of time series, to analyzed yearly data on the consequences of road accident data in Nigeria from 1960-2013. Further, Mutangi (2015) used ARIMA model for the annual Zimbabwe traffic accident statistics from 1997 to 2013 and to forecast the number of annual traffic accidents likely to occur in future.

With this background, an attempt has been made to know the pattern of accidental deaths in India from 1967-2015 and to forecast the number of annual accidents likely to occurs in future using appropriate ARIMA model. Previous study showed that if mean count of a data set is very high, then error may be approximated as normal and ARIMA model may be satisfactory (Quddus 2008). For this reason, we have used ARIMA model instead of using other time series models for count data. The organizations of the paper are as follows: Section 2 deals with source of the data and complete methodology undertaken in this study. In Section 3, we have given results and discussion of the study. Finally, conclusion of the study is given in Section 4.

Materials and Methods

The data are secondary in nature and obtained from National Crime Record Bureau, Ministry of Home Affairs annual Report on Accidental Deaths and Suicides in India covering the period from 1967 to 2015.

The methodology and the theorems propounded earlier called the Autoregressive Integrated Moving Average (ARIMA) has been used . This is an advance technique of forecasting requires long seasonal time series data. This model decomposes historical data into an Autoregressive (AR) process, where there is a memory of past values, an Integrated (1) process, which accounts for stabilizing or making the data stationary plus a Moving-Average (MA) process, which accounts for previous error terms making it easier to forecast.

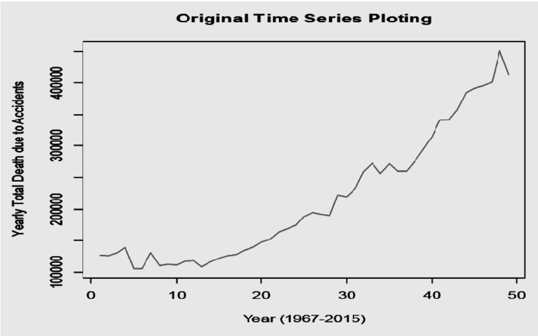

Fig. 1. Plotting of original data.

Results and Discussion

Fig. 1 shows the time plot of total number of deaths due to accident in India from 1967-2015. From Fig. 1, it is observed that the time plot exhibits a significant upward trend. Next, we divide the whole data set into two parts: testing (1967-2005) and validation (2006-2015). First, we develop the ARIMA model for testing part and compare the forecasted values from the model with validation part. If the model fits well and fulfills the all assumptions, then we refit the model for the whole data set i.e. 1967-2015 and forecast accidental deaths for the upcoming 10 years. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of testing data set reveal the significant trend.

By applying augmented Dickey-Fuller test to the first differenced data, it is found that p-value >0.05, so we cannot reject the null hypothesis i.e. the first

Table 1. Parameter estimation of ARIMA (2, 2, 1).

Parameter	Estimate	Std error	Z	p-value
$\hat{\Phi}_1$	-0.5277	0.1679	-3.1421	0.00001
$\hat{\Phi}_2$	-0.3560	0.1809	-1.9684	0.00001
$\hat{\theta}_1$	-0.8017	0.1060	-7.5638	0.00001

differenced series is still behave non stationary (Fig. 2). Therefore, for better result, the researchers go for second differencing of the data (Fig. 3). A clear non stationary behavior of the second differenced data has been observed in (Fig. 4). Similarly, Augmented Dickey-Fuller test tells the same truth (p-value <0.05).

In Fig. 4, it is observed that there is single significant spike at lag 1 in the ACF indicating MA (1) process whereas two significant spikes have been observed at lag 1 and lag 2 in the PACF suggesting AR (2) process might be feasible. Therefore, the researcher acquire a rough idea that ARIMA (2, 2, 1) model, may be tentatively appropriate to total accidental death data which is given by

$$(1-\Phi_1 B - \Phi_2 B^2)(1-B)^2 Y_t = (1-\theta_1 B)\, e_t \qquad (3.1)$$

For the ARIMA (2, 2, 1) model obtained above Eq. (3.1), the parameters of the model are estimated by maximum likelihood method obtained from R software is given in Table 1. In Table 1, it is observed that all the coefficients are significant. The constant term of the model is omitted due to differencing. Then the estimated model can be written as

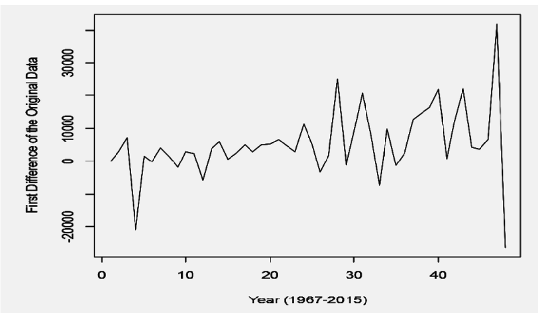

Fig. 2. First difference of the original data.

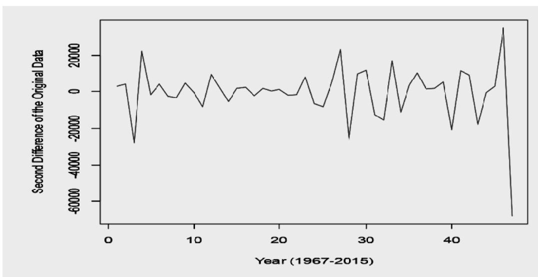

Fig. 3. Second difference of the original data.

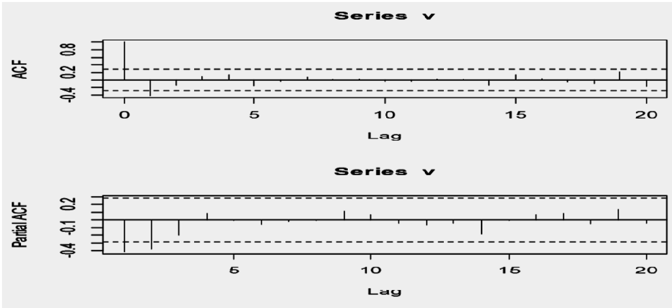

Fig. 4. ACF and PACF of the second difference of the original data.

$$1+0.5277\times B+0.3560\Phi_2\times B^2)(1-B)^2\ Y_t=(1+0.8017\times B)e_t \quad (3.2)$$

If the model ARIMA (2, 2, 1) fits well, the standardized residuals estimated from this model should behave as an i.i.d.(independent and identically distributed) sequence with mean zero and variance σ^2. From standardized plot of residuals, it is observed that residuals are lie within the limit of -3 and +3. Also, (Fig. 5) shows, none of the autocorrelations is individually statistically significant and nor the Ljung-Box-Pierce Q-statistics are statistically significant. Here, the Ljung-Box test statistic is 12.22 and p-value is 0.2011, so the null hypothesis of independence of the residual series cannot be rejected. Using the white noise test (from the normwn.test package in R : Perform a univariate test for white noise), it is obtained that p value is 0.4711 which means that the residuals series is white noise (with mean 0 and variance σ^2). To be sure that the predictive model cannot be improved upon, it is also a good idea to check whether the forecast errors are normally distributed or not, quintile-quintile (q-q) plot of the residuals is drawn which is given in (Fig. 6).

Moreover, by applying Komogorov-Smirnov (K-S) test to the residuals it is found that D=0.0277 and p-value =0.3212 which indicates that the residuals follow normal distribution well. Now, to check the validity of the fitted model, the actual observations are plotted with predicted values from 2006-2015 (10 years) in (Fig. 7) where blue line represents predicted numbers of death from the model and red line repre-

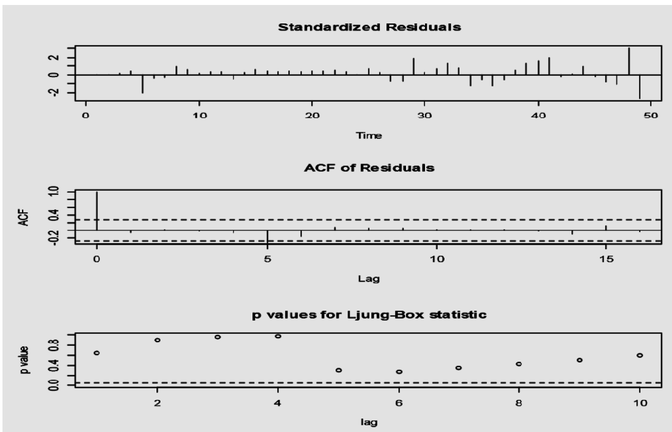

Fig. 5. Standardized residuals, the ACF of the residuals and the p-values of the Q-statistic.

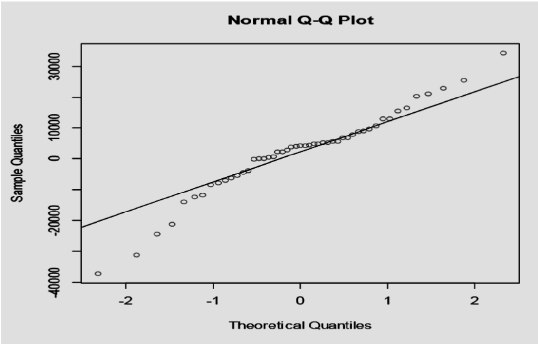

Fig. 6. Q-Q plot of the residuals.

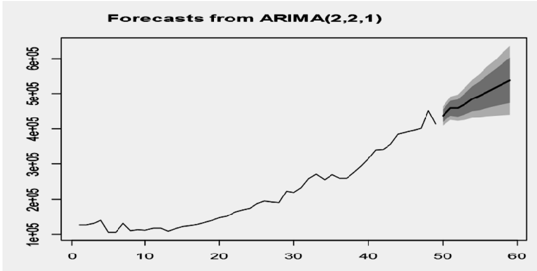

Fig. 8. Trend of accident cases in India for upcoming 10 years.

sents actual observations. From Fig. 7, it is observed that the total number of predicted deaths due to accidents in India from 2006-2015 is almost equal and exact pattern with the actual data. Therefore, it may be concluded that proposed model would be good fitted to the yearly total numbers of death in India due to accident from 1967-2015.

Moreover, the forecasted number of deaths due to accidents in India for the upcoming 10 years which is given in Table 2 along with 95% confidence interval. Also, (Fig. 8) represents forecasted values along with 99% and 95% Confidence Interval (CI).

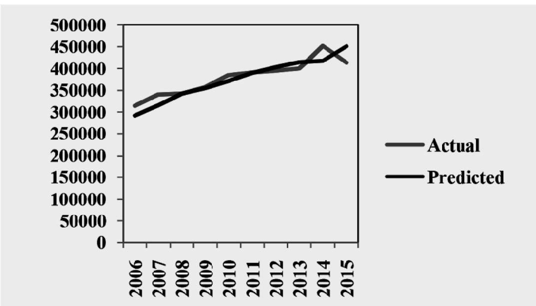

Fig. 7. Plotting of predicted values with actual observations (1976-2015).

Table 2. Forecasting accident cases in India for upcoming 10 years.

Year	Point forecast	95% CI Lower limit	95% CI Upper limit
2016	436031	408836	463225
2017	458363	425619	491106
2018	459150	421455	496844
2019	471391	425009	517773
2020	485259	431155	539363
2021	494190	432405	555976
2022	505148	434678	575617
2023	516793	437461	596126
2024	527355	438935	615774
2025	538243	440289	636197

Conclusion

In this paper, the yearly deaths due to accident in India have been studied using the Box-Jenkins (ARIMA) methodology. The estimation and diagnostic analysis results revealed that the models' are adequately fitted to the historical data. The residual analysis, confirmed that there is no violation of assumptions in relation to model adequacy. Therefore, it may be concluded that, selected model ARIMA (2, 2, 1) would be good fitted to the yearly total number of deaths in India due to accident from 1967-2015. The forecasted number of deaths due to accidents in India for the upcoming 10 years also exhibits an upward trend. Therefore, the policy makers and managers should pay more attention to prevent accidents in the up-coming days. Most of the accidental deaths are un-natural and human error is one of the major causes of this problem. To avoid increase in un-natural accident, the government should create awareness program among the masses. For example, to prevent road accident; Road Safety Agency should create awareness program among the masses so as to notify (or remind) them of the rules and regulations of transportation. Also, the law makers should give severe and immediate penalty for the lawbreakers. Un-natural accidents are control of human hand, only precautions are necessary. The entire nation must aware of this burning problem and should obey the rules and regulations. Natural accidents are beyond the control of human hand but can take preventive measures on time. Furthermore, government as well as masses should not be left alone to solve these burning problems.

References

Adebola FB, Sansi RA, Adegoke NA (2015) Consequence of road traffic accident in Nigeria: Time series approach. Int J Comp Appl Technol and Res 4 (4) : 262—273.

Avuglah RK, Adu-Poku KA, Harris E (2014) Application of ARIMA models to road traffic accident cases in Ghana. Int J Stat and Appl 4 (5) : 233—239.

Monfared AB, Soori H, Mehrabi Y, Hatami H, Delpisheh A (2013) Prediction of Fatal road traffic crashes in Iran using the Box-Jenkins time series model. J Asian Sci Res 3 (4) : 425—430.

Mutangi K (2015) Time series analysis of road traffic accidents in Zimbabwe. Int J Stat and Appl 5 (4) : 141—149.

National Crime Record Bureau. Ministry of Home Affairs annual report on accidental deaths and suicides in India covering the period from 1967 to 2015.

Ofori T, Ackah B, Ephraim L (2012) Statistical models for forecasting road accident injuries in Ghana. Int J Res in Environ Sci and Technol 2 (4) : 143—149.

Quddus MA (2008) Time series count data models : An empirical application to traffic accidents. Accident Analysis and Prevention 40 (5) : 1732—1741.

Sivakumar T, Krishnaraj R (2015) Relationship between road traffic accidents and socio-demographic factors in India. Int J Pharmaceut Sci Review and Res 32 (2) : 137—144.