INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES
HRMARS MANAGEMENT ACADEMIC RESEARCH SOCIETY EXPLORING INTELLECTUAL CAPITAL
www.hrmars.com
ISSN: 2222-6990

A Statistical Analysis of Road Accident Fatalities in Malaysia

Ahmad Husam Azhari, Farah Athirah Md Zaidi, Muhammad Haziq Ahmad Anuar, Jamilah Othman

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v12-i5/13058 DOI:10.6007/IJARBSS/v12-i5/13058

Received: *08* March 2022***, Revised:*** *12* April 2022***, Accepted:*** *29* April 2022

Published Online: 16 May 2022

In-Text Citation: (Azhari et al., 2022)

To Cite this Article: Azhari, A. H., Zaidi, F. A. M., Anuar, M. H. A., & Othman, J. (2022). A Statistical Analysis of Road Accident Fatalities in Malaysia. *International Journal of Academic Research in Business and Social Sciences*, *12*(5), 1224 – 1230.

Vol. 12, No. 5, 2022, Pg. 1224 – 1230

http://hrmars.com/index.php/pages/detail/IJARBSS | **JOURNAL HOMEPAGE**

Full Terms & Conditions of access and use can be found at
http://hrmars.com/index.php/pages/detail/publication-ethics

INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES
HRMARS MANAGEMENT ACADEMIC RESEARCH SOCIETY
EXPLORING INTELLECTUAL CAPITAL
www.hrmars.com
ISSN: 2222-6990

A Statistical Analysis of Road Accident Fatalities in Malaysia

Ahmad Husam Azhari[1], Farah Athirah Md Zaidi[2], Muhammad Haziq Ahmad Anuar[3], Jamilah Othman[1]

[1]Center of Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Persiaran Seremban Tiga/1, Seremban 3, 70300 Seremban, Negeri Sembilan, Malaysia.
Email: jamilahothman5@gmail.com

Abstract

Road fatalities are the number of deaths caused by road accident intentionally or unintentionally. Previous studies show that the number of road fatalities in Malaysia is still concerning and increasing from year to year. This unresolved issue was not only life threatening but it also potentially gives major problems to the economic growth in the country. This study aims to determine the best model in forecasting road accident fatalities in Malaysia. There are two methods used which are Autoregressive Integrated Moving Average (ARIMA) model and Auto-Regressive Poisson model. Each of the model was constructed and analysed by using Statistical Package for Social Science (SPSS) and R Studio, then the models were evaluated and compared to each other based on their error value. The error measure that used in this study are Relative Fit Error (RFE), Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) by selecting the lowest value of error measure. The lowest error measure considered as the best model and will be used to forecasts the number of road deaths in five years ahead that is 2021 until 2025. The results indicate that the best model is ARIMA (0,2,1) from the Box-Jenkins method as it has the lowest error measure as compared to Auto-Regressive Poisson model. On this basis, it is recommended that future research to implement ARIMA model in forecasting and to include other type of model to be compared with in measuring the error measure.

Keywords: Road Fatalities, Autoregressive Integrated Moving Average (Arima), Auto-Regressive Poisson, Forecast

Introduction

According to the World Health Organization's (WHO's) in the global status report (2018), approximately 1.35 million deaths each year as a result of road traffic crashes and 54% of it were caused by vulnerable road users. WHO also stated that even though most of the countries have approximately 60% of the world's vehicle, more than 90% of the world's fatalities on the roads were tend to take place in the low- and middle-income countries which also a leading cause of death for children and young adults age of 5-29 years old. In Malaysia,

there was 15,044 total of road traffic accidents which consists of 6,167 deaths, 3,022 serious injuries and 5,855 minor injuries in 2019. According to Malaysian Institute of Road Safety Research (MIROS), registered vehicle in 1995 was 6,802,375 and had increased to 22,702,221 in 2012 that holds three times the number on 1995 This statistic shows that people were depending primarily on vehicles in their daily lives which lead to the increment in road fatalities, traffic congestion, and pollution (Urie et al., 2016). Figure 1.1 below shows the distribution of road death in Malaysia from 1980 to 2020 as reported by MIROS.

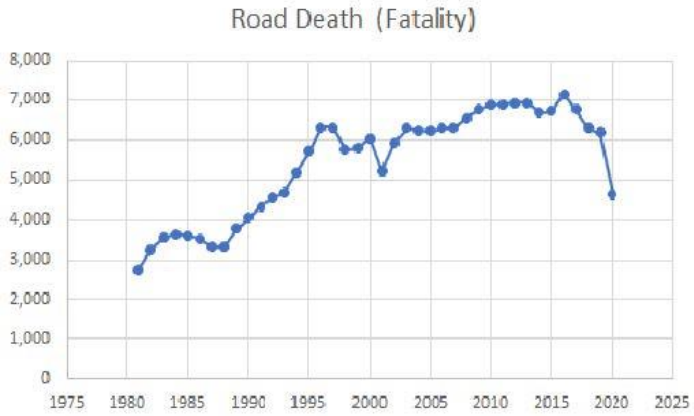

Figure 1.1: Distribution of Road Death in Malaysia

The number of road deaths continues have been alarming, and it was recognized as a global issue in every country (Uzumcuoglu et al., 2016). This problem was not only life-threatening but indirectly could had a huge impact on the country's economic growth if not curbed immediately. Malaysia was also experiencing the same problem with no exception, according to the Department of Statistics Malaysia (2019), they stated that there were 548598 road accidents cases that occurred during 2018 in Malaysia. DOSM (2019) also reported that transport accidents were one of the top ten causes of death in 2018 in Malaysia.

Based on the current situation, reducing accidents to reduce the number of deaths had become an urgent issue that the government must address. The number of fatalities in road accidents must be forecasted to develop a road safety plan in terms of personnel, budgets, and policies. Therefore, forecasting road fatalities is a major step in determining the target for road safety (Wittenberg et al., 2013). Hence, the Malaysian Road Safety Plan 2021-2030 was currently under development. Malaysia's Road Safety Plan 2021-2030 was also a rolling planned that had been adjusted as needed to meet the goal of reducing road accident deaths and injuries (Road Safety Plan 4 (2021-2030).

Methodology

Description of Data

The secondary data from Royal Malaysia Police (PDRM) was used with the main variable was the number of fatalities (death) in road accidents for every year starting from 1972 until 2020. The next variable was the total population of Malaysians by year starting from 1972 to 2020 from Department of Statistics Malaysia Official Portal (DOSM). Besides the data on the number of registered vehicles annually starting from 1972 until 2020 from Road Transport Department Malaysia (JPJ).

ARIMA Time Series Model

ARIMA modelling was commonly applied to time-series analysis, forecasting and control. The term ARIMA was an acronym for the combination that comprises of Auto Regressive (AR) Integrated (I) Moving Average (MA) models. In 1970, Box and Jenkins introduced the ARIMA model. It was also known as Box-Jenkins's methodology, which contained a set of operations for recognizing, estimating, and diagnosing ARIMA models using time series data. ARIMA models have demonstrated their capacity to deliver accurate short-term prediction. In terms of short-term prediction, it consistently defeated complicated structural models. The future value of a variable in an ARIMA model was a linear combination of previous values and past errors, expressed as follows:

$$Y_t = \emptyset_0 + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + ... + \emptyset_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} \quad (1.0)$$

where, y_t was the actual value and ε_t was the random error at t, $\emptyset_i$ and θ_j were the coefficients, p and q were integers that were often referred to as autoregressive and moving average, respectively.

Poisson Regression Model

For modelling count data, the Poisson regression model is frequently referred to as a common model. The Poisson regression model usually was used for most count data in road safety modelling because it fits the statistical features of count data and flexible enough to be reparametrized into other types of distributional functions (Cameron & Trivedi, 2001). The purpose of Poisson regression is to predict a dependent variable made up of count data when there is one or more independent variables. Poisson regression will be able to determine which of the independent variables (population, number of registered vehicle and road length in Malaysia) have a statistically significant effect on the dependent variable (number of road accident death).

Results and Discussions

ARIMA Time Series Model

This study starts by plotting the data, auto-correlation function (ACF) and partial autocorrelation function (PACF) in the collection to ensure more reliable stationary condition of the data. Based on the plot we can see that the data has an increment pattern from 1972 to 2020. This has not met the stationarity assumption, thus another assumption checking was done by checking the auto-correlation function (ACF) and partial auto-correlation function (PACF). A statistical test that also known as a unit root test is one of an alternative way in determining the stationary status of a data. The result shows that the time-series data is non-stationary. A further analysis was done to ensure the data is stationary. This study conducted model identification in determining the value for ARIMA (p,d,q) where p: autoregressive terms, q: the number of non-seasonal differences needed for it to be stationary, and q: the number of lagged forecast errors in the predicted equation. The result shows the best model was ARIMA (0,2,1) in order to fulfil stationary assumption and to forecast the number of road fatality in Malaysia for the next five years. The model then was checked again to ensure the stationary status with ACF and PACF. Finally the result shows the number of road fatality was stationary-based status. Since the data is stationary, therefore ARIMA (0,2,1) is the best value in predicting the number of road fatalities in Malaysia.

Poisson Regression Model

The Omnibust Test shows that the Poisson regression model is statistically significant since its p-value is 0.000.

Table 1: Parameter of Estimates of Poisson

Parameter	B	Sig.
Intercept	6.215	0.000
Population	0.0000001326	0.000
Number of registered vehicles	-0.00000006791	0.000
Road length	0.000001127	0.000

The model also shows that the independent variables which are population, number of registered vehicles and road length have a significant relationship towards the model as shown in the above table.

From the table, the equation of this regression model is as follows:

$$\mu_i = exp^{6.215} exp^{0.0000001326\text{Population} - 0.000000006791\text{Number of registered vehicles} + 0.000001127\text{Road length}} \quad (2.0)$$

Conclusion

ARIMA model has been considered the best model in forecasting road accident fatalities in Malaysia. According to model identification for ARIMA time series model, it shows that ARIMA (0,2,1) is the best model to fulfil the assumptions and forecast the number of road death in Malaysia for upcoming five years.

As the recommendation for the future research, the sample size or number of observations for data collected either primary data or secondary data should have been larger. Since the number of observations are small, the finding for certain model could not be achieved. In this study, the sample size is small which is 49 years for annual data. As a result, one of the models analyses that has been conducted in the study which is negative binomial regression model could not be used to predict or forecast the number of road accident fatality because all of the variables used were found not significant to the model. Therefore, by increasing the sample size and including more independent variables, other models could be used in forecasting road accident fatalities in Malaysia.

References

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), 199– 217.

Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the arima model. In 2014 uksim-amss 16th international conference on computer modelling and simulation (pp. 106–112).

Cameron, A. C., & Trivedi, P. K. (2001). Essentials of count data regression. A companion to theoretical econometrics, 331.

Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (Vol. 53). Cambridge university press.

Chengye, P., & Ranjitkar, P. (2013). Modelling motorway accidents using negative binomial regression. Journal of the Eastern Asia Society for Transportation Studies, 10, 1946–1963.

Commandeur, J. J., Bijleveld, F. D., Bergel-Hayat, R., Antoniou, C., Yannis, G., & Papadimitriou, E. (2013). On statistical inference in time series analysis of the evolution of road safety. Accident Analysis & Prevention, 60, 424–434.

Elvik, R., Sagberg, F., & Langeland, P. A. (2019). An analysis of factors influencing accidents on road bridges in norway. Accident Analysis & Prevention, 129, 1–6.

Eusofe, Z., & Evdorides, H. (2017). Assessment of road safety management at institutional level in malaysia: A case study. IATSS research, 41(4), 172–181.

Hassouna, F., & Pringle, I. (2019). Analysis and prediction of crash fatalities in australia. The Open Transportation Journal, 13(1).

Ismail, N., & Jemain, A. A. (2007). Handling overdispersion with negative binomial and 45 generalized poisson regression models. In Casualty actuarial society forum (Vol. 2007, pp. 103–58).

Jacobs, G., Aeron-Thomas, A., & Astrop, A. (2000). Estimating global road fatalities.

Jomnonkwao, S., Uttra, S., & Ratanavaraha, V. (2020). Forecasting road traffic deaths in thailand: Applications of time-series, curve estimation, multiple linear regression, and path analysis models. Sustainability, 12(1), 395.

Karlaftis, M. G., & Golias, I. (2002). Effects of road geometry and traffic volumes on rural roadway accident rates. Accident Analysis & Prevention, 34(3), 357–365.

Law, T. H., Umar, R. R., & Wong, S. (2005). The malaysian government's road accident death reduction target for year 2010. IATSS research, 29(1), 42–49.

Maxwell, O., Mayowa, B. A., Chinedu, I. U., & Peace, A. E. (2018). Modelling count data; a generalized linear model framework. Am J Math Stat, 8(6), 179–183.

Mohd. (2005). Introductory business forecasting : a practical approach. Shah Alam: Pusat Penerbitan Universiti (UPENA).

Mustafa, M. (2006). Management and privatization of road network in bangladesh. In Road engineering association of Asia and Australasia (reaaa) conference, 12th, 2006, philippines.

Naghawi, H. (2018). Negative binomial regression model for road crash severity prediction. Modern Applied Science, 12(4), 38.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. Journal of the Royal Statistical Society: Series A (General), 135(3), 370–384.

Oppong, R. A. (2012). Statistical analysis of road accidents fatality in ghana using poisson regression (Unpublished doctoral dissertation).

Page, Y. (2001). A statistical model to compare road mortality in oecd countries. Accident Analysis & Prevention, 33(3), 371–385.

Prasetijo, J., & Musa, W. Z. (2016). Modeling zero–inflated regression of road accidents at johor federal road f001. In Matec web of conferences (Vol. 47, p. 03001).

Quddus, M. A. (2008). Time series count data models: an empirical application to traffic accidents. Accident Analysis & Prevention, 40(5), 1732–1741.

Radin, U. (2005). Updates of road safety status in malaysia. IATSS research, 29, 78–80.

Razali, M. S. A., & Zamzuri, Z. (2016). Modeling motorcycle road accidents with traffic offenses at several potential locations using negative binomial regression model in malaysia.

Road safety plan (2021-2030). (n.d.). https://www.mot.gov.my/en/land/safety/ road-safety-plan-2021-2030. (Accessed: 2010-09-30)

Rohayu, S., Sharifah Allyana, S., Jamilah, M., & Wong, S. (2012). Predicting malaysian road fatalities for year 2020.

Sharma, A., & Landge, V. (2013). Zero inflated negative binomial for modeling heavy vehicle crash rate on indian rural highway. International Journal of Advances in Engineering & Technology, 5(2), 292.

Urie, Y., Velaga, N. R., & Maji, A. (2016). Cross-sectional study of road accidents and related law enforcement efficiency for 10 countries: a gap coherence analysis. Traffic injury prevention, 17(7), 686–691.

Wai, A. H. C., Seng, S. Y., & Fei, J. L. W. (2019). Fatality involving road accidents in malaysia: A comparison between three statistical models. In Proceedings of the 2019 2nd international conference on mathematics and statistics (pp. 101–105).

Wittenberg, P., Sever, K., Knoth, S., Sahin, N., & Bondarenko, J. (2013). Determining quantitative road safety targets by applying statistical prediction techniques and a 47 multi-stage adjustment procedure. Accident Analysis & Prevention, 50, 566–577.

Yannis, G., Papadimitriou, E., & Folla, K. (2014). Effect of gdp changes on road traffic fatalities. Safety science, 63, 42–49.

Yu, L. (2012). Using negative binomial regression analysis to predict software faults: A study of apache ant.

Yusria, D. (2017). A time series analysis of road traffic fatalities in malaysia/yusria darma (Unpublished doctoral dissertation). University of Malaya.

Zakaria, R., & Ahmad, Y. (2007). Pemodelan regresi antara kemalangan jalan raya dengan isipadu trafik, persimpangan, lampu isyarat dan waktu kejadian. International Journal of Management Studies, 14(2), 113–125.