Macro Impact Factors of Road Traffic Accidents and Prediction Analysis of Accident Fatalities

Runze Li[a], Baohua Guo*[a]

[a]School of energy Science and Engineering, Henan Polytechnic University, Jiaozuo, Henan, 454003, China

*guobaohua@139.com

Abstract. This study aims to analyze and predict the number of deaths in road traffic accidents through multiple linear regression models, with a focus on examining macro influencing factors. The article collects multiple macroeconomic and social variables that affect the mortality rate of traffic accidents, including Gross Domestic Product (GDP), motor vehicle ownership, road mileage, number of motor vehicle drivers, and year-end total population. Firstly, the data was preprocessed, including missing value processing and outlier detection. Subsequently, a multiple linear regression model was used to model the data, and the assumptions of the model were validated. Through stepwise regression analysis, significant influencing factors were screened and the final regression model was constructed. The goodness of fit and predictive performance of the model are evaluated through cross validation and independent test sets. The results show that the number of motor vehicles, the length of highways open to traffic, and the number of motor vehicle drivers are the main macro factors affecting the number of deaths in traffic accidents. Based on this model, we can effectively predict the number of deaths in traffic accidents in the future, providing scientific basis for traffic management departments to formulate corresponding preventive measures. Research has shown that multiple regression models have high application value and accuracy in analyzing traffic accidents at the macro level.

Keywords: multiple linear regression; traffic accident; death toll; macro impact factors

1 INTRODUCTION

With the development of economy and social progress, the number of motor vehicles and road traffic flow continue to increase, and the incidence of road traffic accidents has also increased, which has become an important factor threatening public safety and social stability. Traffic accidents not only cause great harm to individuals and families, but also place a heavy burden on society and the economy. As can be seen from Figure 1, the number of road accidents, fatalities and direct property losses in China from 2005 to 2022. Therefore, it is of great significance to study the influencing factors of road

G. Chen et al. (eds.), *Proceedings of the 2024 International Conference on Rail Transit and Transportation (ICRTT 2024)*, Advances in Engineering Research 254,
https://doi.org/10.2991/978-94-6463-610-9_15

traffic accidents and accurately predict the number of traffic accident fatalities for formulating effective traffic safety policies and measures.

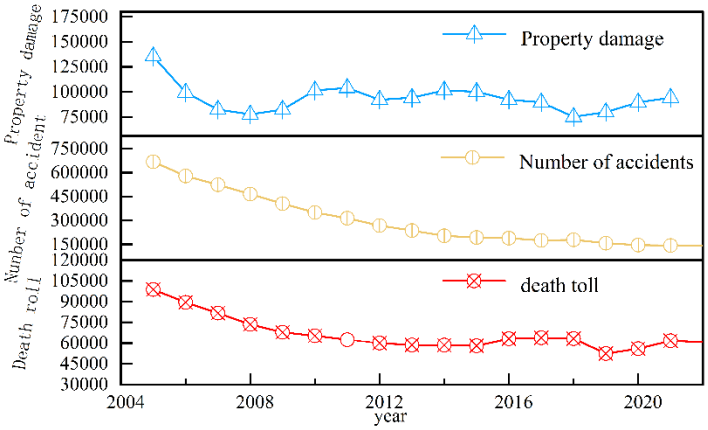

Fig. 1. Index values of road traffic accidents in China from 2005 to 2022

Existing studies have shown that the occurrence and severity of road traffic accidents are influenced by a variety of factors, including road conditions, weather conditions, vehicle technical status, and driver behavior. Scholars' research focuses on the four aspects of people, vehicles, roads and environment, and most of them are analyzed from the level of single or multiple factors; Gao Yinan and others[1], through the analysis of the causes of 1.09 million traffic accidents from 2015 to 2019, concluded that speeding, technical inadequacy, overloading transportation, and lagging standards and regulations are the essential factors, which can be key points for reducing the number of accidents; Shen Kun[2] used the method of multiple linear regression to establish and optimize a prediction model for the number of fatalities in road traffic accidents and the impact factors of road traffic accidents;Xu Dongxing and others[3] used the revised Grey Relational Theory in Deng's Relational Theory and Grey System Theory to analyze the main factors related to traffic accident data and made better predictions of the number of traffic accidents; When Zheng Zhixiao et al[4]. explored the relationship between the driver's driving visual perception time and the accident, they came to the research conclusion that the driver's emergency braking operation behavior during driving is very easy to cause collision accidents in a short time;Based on the traffic accident data of the United Kingdom from 2005 to 2017, Feng et al[5].used the Apriori algorithm of association rules to analyze, and concluded that the regional traffic speed limit has a strong correlation with environmental characteristics and location; Xie Xuebin et al[6] constructed The time series combination model makes a good prediction of the number of traffic accidents;Chien and others[7] used the Generalized Ordered Logistic Regression Model to determine the factors affecting the severity of traffic accidents, and the results showed that time, location, season, tunnel length, and severe weather are significant factors; Karimnezhad and Moradi[8] constructed a Bayesian network model based on Iranian traffic accident data to analyze the

severity of accidents, and the results indicated that the use of seat belts, vehicle and driver's license type, and cultural level are the most significant factors.However, the macro analysis of the influencing factors of traffic accidents, such as gross domestic product (GDP), motor vehicle ownership, road mileage, number of motor vehicle drivers, and total population at the end of the year, has not been fully explored. The impact of these macroeconomic and social factors on traffic accidents may reveal a broader range of socio-economic laws and policy orientations.

Based on the above research, this paper uses mathematical statistical analysis and multiple linear regression analysis to analyze the influencing factors of traffic accidents, and establishes a prediction model for traffic accident fatalities. This study analyzes the above macro influencing factors to explore the degree and mechanism of their impact on the number of road traffic accident fatalities. Through in-depth analysis of data and model validation, the significance and contribution rate of each factor to the number of traffic accident fatalities are evaluated, aiming to provide a scientific decision-making basis for traffic management departments and help formulate more effective prevention and response measures.

2 MACRO INFLUENCING FACTORS OF TRAFFIC ACCIDENTS

In the field of macro research, the occurrence of road traffic accidents is closely related to the level of regional economic development, the objective environment of traffic operation, the level of motor vehicle driving and related human social activities. United States researcher I· Aligar analyzed the number of traffic accident deaths and more than 30 related factors in 48 states in United States, and screened out five factors that have a greater impact on traffic accidents: highway mileage/total mileage, number of cars inspected, road or area area, annual average temperature, and per capita income in the region, and established a regression equation to analyze and predict the death rate per million vehicles in traffic accidents. Based on China's national conditions, this paper collects and collates relevant data, and selects five macro influencing factors, including GDP, motor vehicle ownership, highway mileage, number of motor vehicle drivers, and total population at the end of the year.

3 CONSTRUCTION OF PREDICTIVE MODELS

3.1 Establishment of Multiple Linear Regression Equations

In order to study the impact of five factors, namely, GDP, motor vehicle ownership, highway mileage, number of motor vehicle drivers and total population at the end of the year, on the number of road traffic accident deaths in China, a multiple linear regression equation was established in this study. The advantage of using multiple linear regression is that it can consider the influence of multiple independent variables at the same time, revealing the relative contribution of each factor to the number of traffic fatalities. In addition, the method is well explanatory and can clearly show the specific degree

and direction of influence (positive or negative) of each independent variable. This has an important reference value for traffic management departments when formulating policies, and helps to take more targeted measures to reduce traffic accident fatalities.

Table 1. Statistical data on the number of deaths and main influencing factors in road traffic accidents in China from 2005 to 2022

Year	The number of traffic accident fatalities y /person	GDP x_1 /100 million yuan	The number of motor vehicles x_2 / 10,000units	Highway mileage x_3 / 10,000 km	Number of motor vehicle drivers x_4 /10,000 people	Total population at the end of the year x_5 /10,000 people
2005	98738	187318.9	11755.1	334.52	13069.52	130756
2006	89455	219438.5	12494.8	345.73	14213.87	131448
2007	81649	270092.3	13792.4	358.37	15363.88	132129
2008	73484	319244.6	14856.5	373.02	17336.56	132802
2009	67759	348517.7	16769.6	386.08	19167.58	133450
2010	65225	412119.3	19107.4	400.82	20068.47	134091
2011	62387	487940.2	20905.5	410.64	22817.62	134916
2012	59997	538580.4	22255.4	423.75	25250.83	135922
2013	58539	592963.2	23216.8	435.62	26955.93	136726
2014	58523	643563.1	24450.5	446.39	2989232	137646
2015	58022	688858.2	25854.8	457.73	32858.05	138326
2016	63093	746395.1	27341.6	469.63	35876.98	139232
2017	63772	832035.9	29767.7	477.35	36016.94	140011
2018	63194	919281.1	31735.8	484.65	40971.31	140541
2019	52388	986515.2	33831.4	501.25	43503.56	141008
2020	55950	1013567.6	35215.6	519.81	45618.42	141212
2021	61703	1149237.5	38526.4	528.07	48140.87	141260
2022	60676	1210207.2	40504.8	535.48	50208.53	141175

In the population linear regression function, the regression coefficient of the function variable is unknown and can be estimated by the observed values of the sample data in Table 1, then the linear relationship can be expressed as

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} \tag{1}$$

In the formula:

$Y_i -$ Sample conditional mean of explanatory variable;

$\beta_k -$ the estimation of the overall regression parameter.

and the linear relationship of n sets of observations $(Y_i, X_{2i}, X_{3i}, \cdots, X_{ki})(i = 1, 2, \cdots, n)$ can be obtained by n observations of the explanatory variable Y and multiple explanatory variables, and the multiple linear regression equation can be obtained by regression analysis by SPSS software according to the data in Table 1, that is, $n = 18, k = 6$

$$y = 119752.539 - 0.261x_1 + 1.04x_2 - 405.512x_3 + 3.855x_4 - 8.141x_5 \tag{2}$$

In the formula:

$y -$ the number of traffic accident fatalities (people);

x_1 – GDP (100 million yuan);

x_2 – the number of motor vehicles (10,000 units);

x_3 – the mileage of highways (10,000 km);

x_4 – the number of motor vehicle drivers (10,000 people);

x_5 – the total population at the end of the year (10,000 people).

3.2 Correlation Analysis

Correlation analysis is a statistical method used to measure the linear relationship and strength between two or more variables. Specifically, correlation analysis reveals whether there is an association between variables, and the direction and extent of that association. In this paper, we use correlation analysis to explore the relationship between five factors: gross domestic product (GDP), motor vehicle ownership, road mileage, number of motor vehicle drivers, and total population at the end of the year, and traffic fatalities. In this paper, the correlation analysis of the five element variables affecting the number of traffic accident fatalities in the above regression equation is carried out, and the results are shown in Table 2.

Table 2. Correlation analysis between factor variables

Feature variables	Regression coefficients	Correlation with dependent variable (+/-)
Traffic accident fatalities	/	/
GDP	−0.261	−
Motor vehicle ownership	1.04	+
Highway mileage	−405.512	−
Number of motor vehicle drivers	3.855	+
Total population at the end of the year	−8.141	−

The regression coefficient $\ddot{\beta}_k$ is the coefficient corresponding to the independent variable x_k in the traffic accident regression equation. When the regression coefficient is negative, that is, $\beta_k<0$, the corresponding independent variable is negatively correlated with the dependent variable y, and vice versa. The absolute value of the regression coefficient $|\beta_k|$ represents the magnitude of change in y for each unit change in the independent variable. The larger the value of $|\beta_k|$, the greater the magnitude of change in the dependent variable y.

As shown in Table 2, the analysis results in the table indicate that GDP, highway mileage, and the year-end total population are negatively correlated with the number of traffic accident fatalities. This suggests that an increase in these factors may help reduce the number of traffic accident fatalities, and vice versa.

3.3 Regression Equation Significance Test

Through the operation results of SPSS software, the observed value of the test statistic $F = 17.635$ is obtained, at this time, the given significance level is α=0.05, and $F_{1-0.05}(5,18-5-1) = 3.11$ is obtained by checking the F distribution quantile table , when $F \geq F_{1-n}(m, n-m-1)$ (m is the number of independent variables; n is the sample size of the independent variable), then the regression effect of the equation is significant. and it is obvious that $F_{1-0.05}(5,18-5-1) = 3.11$ meets the significance condition, so the null hypothesis of the significance test of the regression equation is not valid, that is, the linear impact of the five influencing factors on the number of traffic accident fatalities is significant.

4 CONCLUSIONS

The study effectively predicts the number of traffic accident fatalities over a period of time through a multiple linear regression model. The model shows that the number of motor vehicles, the length of roadways, and the number of motor vehicle drivers have a significant impact on the number of traffic accident fatalities. Through in-depth data analysis and model validation, the significance and contribution rate of each factor on traffic accident fatalities are evaluated, providing scientific decision-making references for traffic management departments and aiding in the formulation of more effective preventive and responsive measures. The results indicate that the multiple regression model has high applicability and accuracy in analyzing traffic accidents at the macro level.

REFERENCES

1. GAO Y N(2023).Characteristics and mechanism of road traffic accidents. Journal of Safety and Environment, (11).
2. Shen K (2017). Prediction and Management Strategies for Road Traffic Accident Fatalities in China . Safety and Environmental Engineering, 24(05):138-144.
3. XU D X;YIN X;ZHANG X F;SUN T;YE J(2019). Grey Analysis and Prediction of Maritime Traffic Accidents on the Trunk Line of Yangtze River. Navigation of China, (02).
4. ZHENG Z X;LI P H;HU M H;LI Y B(2020). Influence of Visual Distraction on Driver's Car-following Behavior. Journal of Highway and Transportation Research and Development, 37(01):122-130.
5. Feng M, J Zheng, J Ren, et al(2020). Association Rule Mining for Road Traffic Accident Analysis: A Case Study from UK.
6. XIE X B; KONG L Y(2021).Traffic accident prediction based on ARIMA and XGBoost combined model. Journal of Safety and Environment，21(01): 277-284.
7. Chien, Steven, I-Jy, et al(2016). Exploring factors affecting injury severity of crashes in freeway tunnels. Tunnelling &Underground Space Technology, 59: 100-104.
8. Karimnezhad-Ali, Moradi-Fahimeh(2015).Road accident data analysis using Bayesian networks. Transportation Letters,9(1):12-19.