Indian Journal of Science and Technology, *Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/85089, December 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Backward Elimination Approach of Regression Model to Decide the Road Safety Policies Across the States in India

Neeraj Dhanraj Bokde*, Aditya Gupta and Kishore Kulat

Visvesvaraya National Institute of Technology Nagpur (VNIT), Nagpur - 440010, Maharashtra, India; neerajdhanraj@gmail.com, aditya2590@gmail.com, kdkulat@ece.vnit.ac.in

Abstract

Objectives: To observe the road traffic accidents pattern across the states in India and to conclude whether the expenditure on road and transportation infrastructure is really justified with the traffic safety across the country. To decide the road safety policies across the states in India using backward elimination approach of regression model. **Methods/Analysis:** This study attempts to show statistical relations between road traffic accidents with predictive variables; number of vehicles registered, total length of roads, total number of road accidents and total expenditure on transportation infrastructure in the states of India. Time series data covering the period 2001 to 2013 was used on yearly road traffic accidents and other predictive variables for all 28 states of India. **Findings**: In recent decades, road accidents are becoming one of the most growing concerns in India. As mentioned in report of Ministry of Road Transport & Highways Government of India, road accidents impose significant costs which are about 3% GDP for India since 1999 to 2000. In 2004, India was 9th leading cause of death and expected to be 5th leading cause of death by 2030 worldwide. These facts are enough to catch attention of Government agencies and researchers towards safer traffic solutions. The results from the analysis shows few findings like expenditure on transportation infrastructure may not be an important predictive variable for road traffic accidents for pan India situations. Finally, if all the states of India partitioned in different clusters according to the behavior of states, the study states that there exists a strong positive correlation between road traffic accidents, number of vehicles registered and total length of roads. The relationship found to be statistically significant at 5% level of significance for each cluster of states. **Applications/Improvements:** The results obtained in study can assist transportation departments and Government bodies to implement better strategies across the country to ensure road traffic safety and reduce losses in terms of money and infrastructure costs.

Keywords: Policies in India, Regression Analysis, Road Accidents, Transportation, Road Safety

1. Introduction

Cities in developing countries like India are facing problem of heavy rise in the population. The cities are under burden being people from villages are migrating toward the cities. This surely leads to very high transport demand. The mobility of the people and the goods are the factors to define the growth of the cities. Hence, attention towards transportation system is one of the main issues for the all developing countries. Very few of the cities in India are equipped with good public transport facilities including city buses and railways. Still, these city traffic authorities

**Author for correspondence*

are lacking to provide the service for all people, because of many reasons like overpopulation, routes unavailability, long waiting queues, etc. This ultimately leads to heavy increase in private motor vehicle users and demands for more convenient public transportation in the country.

Since last few decades, India is facing the problems related to traffic accidents which resulted into injuries, fatalities, hospitalization and disabilities[1]. Large number of motor vehicle user and lack of proper infrastructure in India is one of the significant reason for this problem[2]. As mentioned in report Basic road statistics of India[1], in year 2014, total number of road accidents encountered were reached to 4,89,400. Also, due to road accidents in year 2014, for every hour, on an average 56 accidents took place and 16 lives were lost in India. Study also states that, around 86.3 percent if accidents occurred in top thirteen states during 2014. These facts about Indian road traffic scenarios are enough to catch attention of the Government and researchers to think about it. The report of road accident in India[2] also states that traffic junctions are most accidents prone areas, since about 57 percent of total accidents took place at the junctions itself, hence it is very important to use transportation budget efficiently to improve the infrastructure of the country in safer way.

An organization Ministry of Road transport and highways under Central Government of India is working on making policies for road transportation, road infrastructure and research related to transportation to increase mobility and efficiency of road transport in country. The reports[3,4] of the same organization states that, for year 2012-13, planning commission have provided Rs. 32,600 crores as the annual outlay. Out of this outlay, the amount around Rs. 21,515 lakhs are estimated for road safety budget. Also, the large amount of budget is estimated for various tasks like awareness of people, driver training, schemes for highways accidents, equipments for road safety and testing of pollution, maintaining transport database, R&D, formations of inspection authorities and many more. It can be very interesting to check whether the huge expenditure on such traffic facilities like road and bridges infrastructure really improved the road traffic accidents profile of the country. This article also focuses on different patterns of dependency of various predictive variables on the road accidents in different states of India.

Though the whole world is facing problems related to traffic accidents and its harmful results, developing and overpopulated country like India is facing same problem with higher risk of losses related to life and economy. The factors responsible for road accidents may vary from region to region or culture to culture. Many researchers have worked to discover such factors from existing available data and also data generated in simulation environments. The various factors those affecting on road accidents are economic status, development, population, driver behavior, education level and many such parameters of respective countries. Articles[5–10] is the research articles showing relations between economic growth and road accidents. Most of those articles proved the positive relations between economic growths to rise in road accidents. Articles[11,12] are showing relations between urbanization and industrialized development with road accidents. Articles[13–15] discuss about how geographic variations affect on the road accidents in developed countries like USA, Spain and Netherlands. Whereas relation between population, number of vehicles and road accidents is discussed in articles[16–18]. Apart from this, researchers all over the world are trying find relation of various external factors like the effects of dietary habits of a driver on driving performance and its impact on road traffic accidents[19]. There can be many other solutions to avoid such traffic fatalities with help of Intelligent Traffic Systems (ITS)[20] which ensures the more safety on roads.

2. Methodology

In current study, a linear regression model was used to analyze the dynamics of changes and variations in road traffic accidents with respect to number of vehicles registered, total length of roads, total number of road accidents and total expenditure on transportation infrastructure in the states of India through time-series data as described.

2.1 Data Source

The study is relied on data obtained from the Ministry of Road Transport and Highways, Government of India[1–4]. The data specifically comprised of time series data on yearly road traffic accidents and other predictive variables values for all 28 states of India covering the period 2001 to 2013. The original data is treated for missing values and screened outliers. In the current study, the statistical operations on all the variables were examined.

2.2 Method of Data Analyses

Data analyses were performed using R language in R-Studio. Exploratory data analyses were conducted to examine the patterns and relationships among dependent variable number of road accidents and logarithmic values of other predictive variables underlying the use of regression with backward elimination approach. Also, before finalizing the fitted model, the correlation coefficients of all the variables were investigated to ensure the significance of each predictive variables. It is important to state that, to finalize the model was fitted perfectly, for all analyses P value < 0.05 was considered to be statistically significant.

3. Observations

The dependent variable in the study was the total number of road accidents across the country and it was tested with other predictive variables. The regression analysis technique with backward elimination approach is used to find these relations and dependency on other variables. Table 1 shows the regression analysis results between number of road accidents and other predictive variables.

Table 1. Regression analysis results between number of road accidents and predictive variables: Total Expenditure, Total Road length and No. of Vehicles

Regression Model			
Dependent Variable: Total number of Road traffic accidents			
Method: least squares			
Variable	**Coefficient**	**t-Statistic**	**Prob.**
Intercept	5.075686	9.814	0.0102
Total Expenditure	0.107805	1.36	0.3068
Total Road length	0.007989	0.056	0.9607
No. of Vehicles	-0.028006	-0.318	0.7806
Regression Statistics			
F-statistics	2.141		
Prob. (F-statistics)	0.3341		
Adjusted R-squared	0.4064		

The insignificant R squared value in Table 1 states that something is wrong with the model. So, it will be better to observe effect of all the predictive variables on number of road accidents individually.

Many research articles are in favor to conclude that road traffic accidents are directly dependent on the number of vehicles registered in that region. While observing the data for pan India (including UTs), there is strong correlation factor between road accidents and the number of registered vehicles in India. Apart from this, while considering same relationship among all the 28 states in India individually, 14 states shows strong positive correlation coefficients while 4 states shows strong negative correlation coefficient and remaining few states don't significantly strong correlation between number of road accidents and the number of registered vehicles as shown in Figure 1.

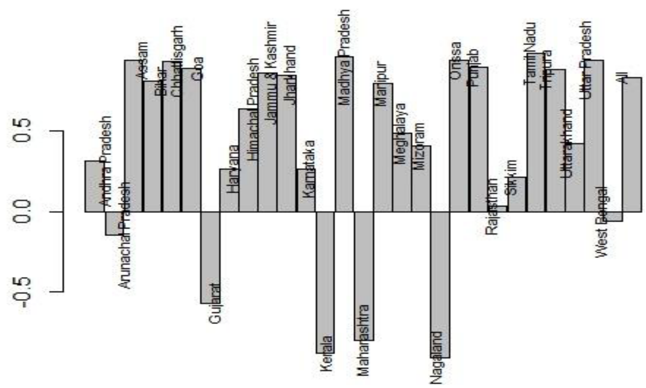

Figure 1. Correlation between road accidents and the number of registered vehicles.

While observing regression analysis between these two factors, 96% of the states, out of those which were showing strong correlation coefficients are also giving very high R squared value and less than 5% significance level of the P value. This indicates that a mathematical regression model is possible to explain the relation between number of road accidents and number of vehicles for those particular states. This regression model may be useful to predict the effect of the number of vehicles on road traffic accidents in future, if expected number of registered vehicles are known. For pan India (including UTs) situation, the same regression coefficients are as mentioned in Table 2(a). The significant P and R squared values indicates the success of the fitted model.

Whereas, while observing relation of number of road accidents with the total length of roads across the states on India, very strong positive correlation coefficients are found among most of the higher population states as shown in Figure 2. About 92% states of India shows strong correlation coefficient for relation of number of road accidents with the total length of roads in the respective states. Though some of the states show the negative correlation, count of such states is negligible. The situation is almost similar for pan India scenarios (including UTs). The correlation factor between road accidents and total

Table 2. Regression analysis results between number of road accidents (a) No. of Vehicles (b) Total Road length (c) Total Expenditure on transportation infrastructure.

Regression Model						
Dependent Variable: Total number of Road traffic accidents						
Method: least squares						
	(a)		(b)		(c)	
Variable	**Intercept**	**No. of Vehicles**	**Intercept**	**Total Road length**	**Intercept**	**Total Expenditure**
Coefficient	4.89116	0.15637	4.44179	0.18975	5.39271	0.04528
t-Statistic	23.021	3.717	9.409	2.626	28.677	1.578
Prob	4.4E-7	0.00989	3.19E-5	0.0341	9E-7	0.175
Regression Statistics						
F-statistics	13.81		6.895		2.491	
Prob. (F-statistics)	0.009891		0.03411		0.1754	
Adjusted R-squared	0.6467		0.6243		0.199	

number of length of roads for pan India is 0.7044, which indicates the strong positive relation between them.

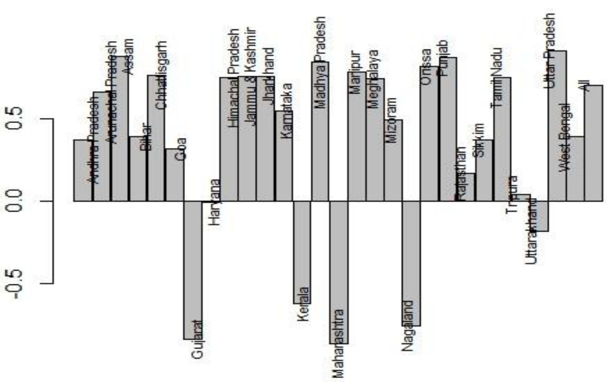

Figure 2. Correlation between road accidents and the total length of roads.

To check the exact mathematical relation between those factors, regression analysis is applied for available data, considering total length of road as predictive variables. In this regression model for each state in India, 18 of total 28 states, especially more populated states give more significant P values and squared R values. Also, the pan India situation is perfectly matched with it. It shows the regression model as shown in Table 2(b). The model seems fitted perfectly, since, the R squared value is 0.6243 and P value is 0.0341, which are significant enough.

The third factor is total expenditure on road and bridges in India. Out of 28 states in India, only 15 states show strong correlation coefficients with the number of road traffic accidents. That is about only 50% of the total number of states. Out of those 15 states, 8 states show strong positive correlation coefficients, whereas 7 states show strong negative coefficients as shown in Figure 3. For pan India, the correlation coefficient between number of road accidents and total expenditure on transportation infrastructure is 0.576, which is not that much significant to state any relation between those factors.

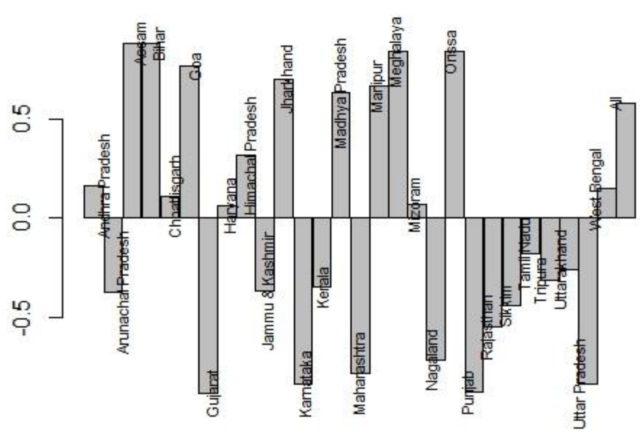

Figure 3. Correlation between road accidents and total expenditure on transportation infrastructure.

Along with this, as shown in Table 2(c), the regression model for number of road accidents versus total expenditure of transport infrastructure as predictive variable does not reflect any attractive and significant relations. Out of 28 states, 19 states in India don't have significant P values and R squared values. Hence, the model associated with factor total expenditure on transport infrastructure is not reliable.

From results observed in Table 2, it is clear that the factor, total expenditure on transport infrastructure is not affecting the number of road traffic accident fatalities as that much regularly and significantly as of other two factors, total number of vehicles registered and road lengths present in respective regions. Hence, backward elimination approach is used to fit the new model for pan India scenario by omitting the insignificant factor from the regression model expressed in Table 3. After fitting the model with number of vehicles registered and total length of roads as predictive variables affecting on road accidents, the regression coefficients with their significance and R squared value are calculated.

Table 3. Regression analysis results between number of road accidents and predictive variables: Total Road length and No. of Vehicles

Regression Model			
Dependent Variable: Total number of Road traffic accidents			
Method: least squares			
Variable	**Coefficient**	**t-Statistic**	**Prob.**
Intercept	5.07089	4.456	0.0066
Total Road length	-0.04754	-0.161	0.87815
No. of Vehicles	0.18216	1.095	0.32348
Regression Statistics			
F-statistics	5.799		
Prob. (F-statistics)	0.04981		
Adjusted R-squared	0.5782		

Table 3 states that model is not perfectly fit, since the regression coefficients are not significant up to 0.05 values of P. Observed value of R squared is very low. Also, condition t>2 is also not satisfied. Hence, for pan India situation, this regression model does not seem to be convenient. This failure of model fitting may be because the states of India may show entirely different behavior from each other. This behavior can be inspected by observing the correlation factors for each states of India. Some states are showing strong positive relations and some are showing strong negative relations with road accident fatalities. Figure 4 shows the relation between the state wise correlation coefficients of number of vehicles registered and length of roads in states with the total number of road accidents in states. Three clusters are formed by K-mean clustering algorithm. It can be observed in Figure 4, all the states in India are divided in three clusters. It might be possible that, states in each cluster can have similar behavior regarding transportation system. Though regression model for pan India criteria was unsuccessful, it may be possible that, each cluster of states may successfully fit in a model. After trying same regression model as in Table 3 to all of the clusters of states, results are obtained as in Table 4.

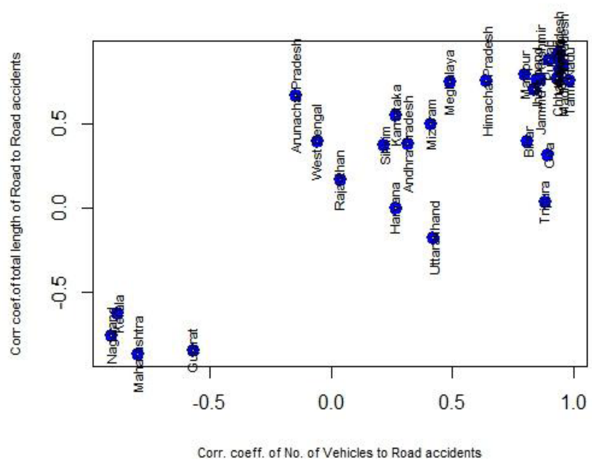

Figure 4. Relation between the state wise correlation coefficients of number of vehicles registered and length of roads with the total number of road accidents in India.

Table 4 reveals that the model is fitted perfectly for each cluster of states, since P values for all predictive variables are lower down significantly and R squared values are large enough.

4. Discussion

In current study, state wise data of road traffic accidents, number of vehicles registered, length of roads

Table 4. Regression analysis results between number of road accidents with Total Road length and No. of Vehicles for (a) states in Cluster 1 (b) states in Cluster 2 (c) states in Cluster 3

Regression Model								
Dependent Variable: Total number of Road traffic accidents								
Method: least squares								
		For States in Cluster 1			For States in Cluster 2			For States in Cluster 3	
Variable		**Total Road length**	**No. of Vehicles**		**Total Road length**	**No. of Vehicles**		**Total Road length**	**No. of Vehicles**
Coefficient		1.2477	0.9258		0.49947	0.75891		0.177431	0.64957
t-Statistic		2.5981	3.448		6.459	14.822		2.127	9.628
Prob		0.01458	0.00175		1.25E-08	< 2e-16		0.0363	2.89E-15
Regression Statistics									
F-statistics		172.6			589.7			117.91	
Prob. (F-statistics)		2.20E-16			<2.20E-16			<2.20E-16	
Adjusted R-squared		0.9172			0.9431			0.7288	

across country and the total expenditure on traffic infrastructure is examined from year 2001 to 2013 and effect of these predictive variables on number of road accidents are examined. Initially, correlation coefficient of all of these predictive variables with respect to road accidents is observed. In most of the states of India, correlation coefficient of number of vehicles registered and total length of roads find strong relation with road traffic accidents, whereas the total expenditure shows very weak relations with it. The regression analysis also shows the similar results in Table 2.

Table 3 discusses about the regression analysis by omitting the predictive variable total expenditure on transportation infrastructure for pan India situation. Still the analysis results don't show any significant relations. That may be because of the drastic change in the behavior of various states. Figure 4 shows the relation between correlation coefficient of the number of vehicles registered, total road length in country and number of road traffic accidents. K mean clustering algorithm leads to three clusters formations and the regression analysis model for each of these clusters reflects the significant result. This concludes that, each cluster of states follows the same traffic behavior according to the observed predictive variables.

5. Conclusion

Following results can be concluded related to study, based on the analyses and interpretations of the available data:

Firstly, expenditure on transportation infrastructure may not be an important predictive variable for road traffic accidents for pan India situations. In other way, it can be said that, government agencies which are spending on expenditure of traffic infrastructure should spend it by considering safety point of view. Secondly, there exists a strong positive correlation between road traffic accidents, number of vehicles registered and total length of roads in India. But these variables are unable to fit a significant regression model while considering pan India situation. Finally, if all the states of India partitioned in 3 different clusters according to the behavior of states, the relationship was found to be statistically significant at 5% level of significance for each cluster of states. The study also revealed that model of each cluster fits the data very well as it shows very high R squared values.

It will be very interesting to find out the factors which may lead to differentiate the states in three different clusters. Out of these clusters, one cluster consists of Maharashtra, Gujarat and Kerala, which are significantly leading in GDP growth, high capita income, high literacy rate and high consumption of alcohol. But, in present study, effect of these variables on number of road accidents is not studied. The usefulness of this study is that, it can be helpful to decide various policies for safer traffic system according to the behavior observed by various states of India. Further in detailed study may investigate the present needs, weakness and failure of the states of India and may be helpful to design the statewide policies for safer transportation to lower down the road traffic accidents across the country.

6. References

1. Basic road statistics of India 2013, Government of India, Ministry of Road Transport and Highways, Transport Research Wing, New Delhi [Internet]. [cited 2015 Dec 01]. Available from: http://www.indiaenvironmentportal.org.in/.
2. Road Accidents in India 2013, Government of India, Ministry of Road Transport and Highways, Transport Research Wing, New Delhi [Internet]. [cited 2015 Dec 01]. Available from: http://morth.nic.in/showfile.asp?lid=1465.
3. Road Transport Year Book 2012, Government of India, Ministry of Road Transport and Highways, Transport Research Wing, New Delhi [Internet]. [cited 2015 Dec 01]. Available from: http://morth.nic.in/showfile.asp?lid=838.
4. Outcome budget India 2015-16, Government of India, Ministry of Road Transport and Highways, Transport Research Wing, New Delhi [Internet]. [cited 2015 Dec 01]. Available from: http://indiabudget.nic.in/.
5. Elvik R. An analysis of official economic valuations of traffic accident fatalities in 20 motorized countries. Accident Analysis and Prevention. 1995 Apr 30; 27(2):237–47.
6. Sehat M, Naieni KH, Asadi-Lari M, Foroushani AR, Malek-Afzali H. Socioeconomic status and incidence of traffic accidents in Metropolitan Tehran: A population-based study. International Journal of Preventive Medicine. 2012 Mar, 3(3):181.
7. Bishai D, Quresh A, James P, Ghaffar A. National road casualties and economic development. Health Economics. 2006 Jan 1; 15(1):65–82.
8. Hasselberg M, Laflamme L. Road traffic injuries among young car drivers by country of origin and socioeconomic position. International Journal of Public Health. 2008 Jan 1; 53(1):40–5.
9. Kopits E, Cropper M. Traffic fatalities and economic growth. Accident Analysis and Prevention. 2005 Jan 31; 37(1):169–78.
10. Agyemang B. Regression analysis of road traffic accidents and population growth in Ghana. International Journal of Business and Social Research. 2013 Nov 23; 3(10):41–7.
11. Bekibele CO, Fawole OI, Bamgboye AE, Adekunle LV, Ajav R, Baiyeroju AM. Risk factors for road traffic accidents among drivers of public institutions in Ibadan, Nigeria. African Journal of Health Sciences. 2008 Jun 20; 14(3): 137–42.
12. Garg N, Hyder AA. Exploring the relationship between development and road traffic injuries: A case study from India. The European Journal of Public Health. 2006 Oct 1;16(5):487–91.
13. van Beeck EF, Mackenbach JP, Looman CW, Kunst AE. Determinants of traffic accident mortality in the Netherlands: A geographical analysis. International Journal of Epidemiology. 1991 Sep 1; 20(3):698–706.
14. Rivas-Ruiz F, Perea-Milla E, Jimenez-Puente A. Geographic variability of fatal road traffic injuries in Spain during the period 2002–2004: An ecological study. BMC Public Health. 2007 Sep 26; 7(1):266.
15. Moudon AV, Lin L, Jiao J, Hurvitz P, Reeves P. The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. Accident Analysis and Prevention. 2011 Jan 31; 43(1):11–24.
16. Sivakumar T, Krishnaraj R. Relationship between road traffic accidents and socio-demographic factors in India. International Journal of Pharmaceutical Sciences Review and Research. 2015; 33(2):137–44.
17. Jadaan KS. A study of accident rates in Kuwait. Journal of the University of Kuwait (Science). 1982; 9:41–50.
18. Enu P. Road traffic accidents and macroeconomic conditions in Ghana. Social and Basic Sciences Research Review. 2014; 2(9):374–93.
19. Mulay P, Mulatu S. What you eat matters road safety: A data mining approach. Indian Journal of Science and Technology. 2016 Apr; 9(15):1–8.
20. Mittal P, Singh Y. Development of intelligent transportation system for improving average moving and waiting time with artificial intelligence. Indian Journal of Science and Technology. 2016 Jan; 9(3):1–7.