Road Accident Analysis and Severity prediction Model On State Highway-5 (Halol-Shamlaji Section)

[1]Shah Dhruvit, [2]Dr. H.R.Varia, [3]Prof. P.M.Shah

[1]M.E. Civil(Transportation),[2]Phd (Transportation),[3]M.E. Civil, PhD (Pursuing)
Tatva institute of technological studies, Modasa

Abstract— **According to MORTH-2014 India has the highest no of accidents in the World. Accident Severity has been increasing year by year. A road accident is an unplanned and uncontrolled occasion, which happened on a road open to an open activity bringing about individual harm, harms to the property and death toll in which no less than one moving vehicle was included.Rapid growth of population coupled with increased economic activities has favored in tremendous growth of motor vehicles. This is one of the primary factors responsible for road accidents. It is observed that few works have been carried out on statistical analysis of accidents particularly on Four-lane State Highway. The essential objective of this paper is to gather the road accident data of chosen stretch from Halol CH 335.800 km (SH-5) to Shamlaji CH 501.800 km (SH-5) from the L&T Toll Plaza and GSRDC office Gandhinagar and to workout purposeful examination of road accident of it and propose remedial measures too. Total 548 road crash recorded in last 24 months, 2013-2014. The collected data were analyzed to evaluate the effect of influencing parameters on accident rate.**

IndexTerms—road accident, blackspot analysis, prediction model

---

I. INTRODUCTION

Road safety is one of the most important problems in our society. Every year 1.2 million of people are killed and between 20 and 50 million people are injured in road accidents. If current trends continue road traffic accidents are predicted to be third leading contributor to the global burden of Disease and injury by 2020 (Torregrosa et al.,2012)India had earned the dubious distinction of having more number of fatalities due to road accidents in the world. Road safety is emerging as a major social concern around the world especially in India. India had earned the dubious distinction of having more number of fatalities due to road accidents in the world. Road safety is emerging as a major social concern around the world especially in India (Shiv kumar and Krishnaraj,2012).

II. Objective of the Study

The primary objectives of the study are:

(i) to identify the engineering factors that may contribute to the cause of accident,
(ii) to identify the black spot and give preventions
(iii) to propose improvements at the location,and
(iv) to develop the severity prediction model using SPSS software.

III. STUDY STRETCH AND DATA COLLECTION

The study stretch was selected from Chainage Km 335.170 to Km 380.00 on SH-5 which connects Halol to Shamlaji in Gujarat state. The accidents data were collected from L&T, IDPL Toll Plaza with prior permission. The Study Stretch is shown in Fig.

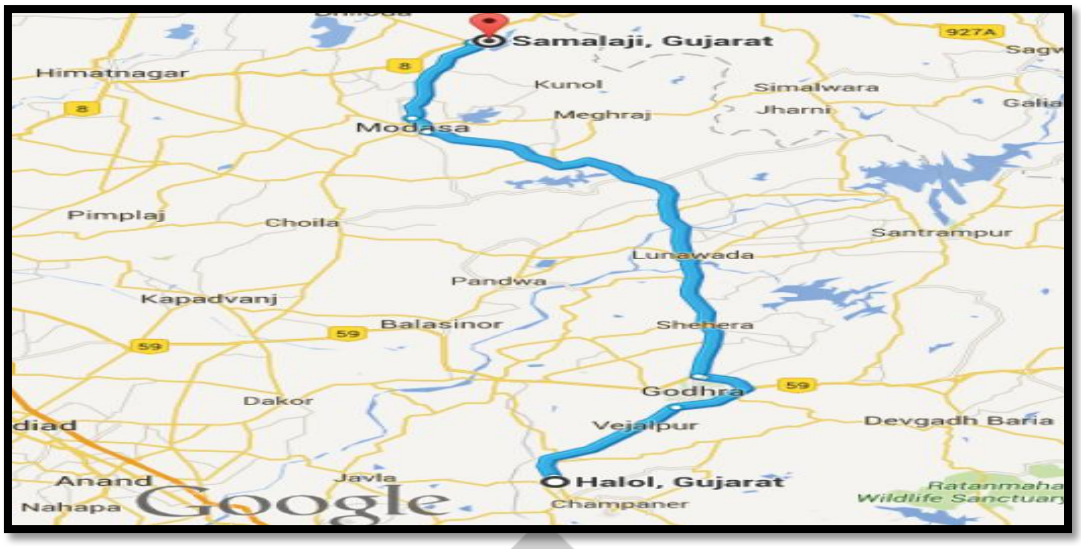

Fig-1. Study Stretch (Source: Google Map)

IV. Accident Data analysis

The accident data collected during last 2 years were plotted with MS excel.Total no of accidents, fatalities; injuries are shown in Table 1.

Table no. 1

Year	2013	2014
Total no. Of accidents	443	471
Person killed	51	30
Person injured	274	392
Non-Injury	118	49

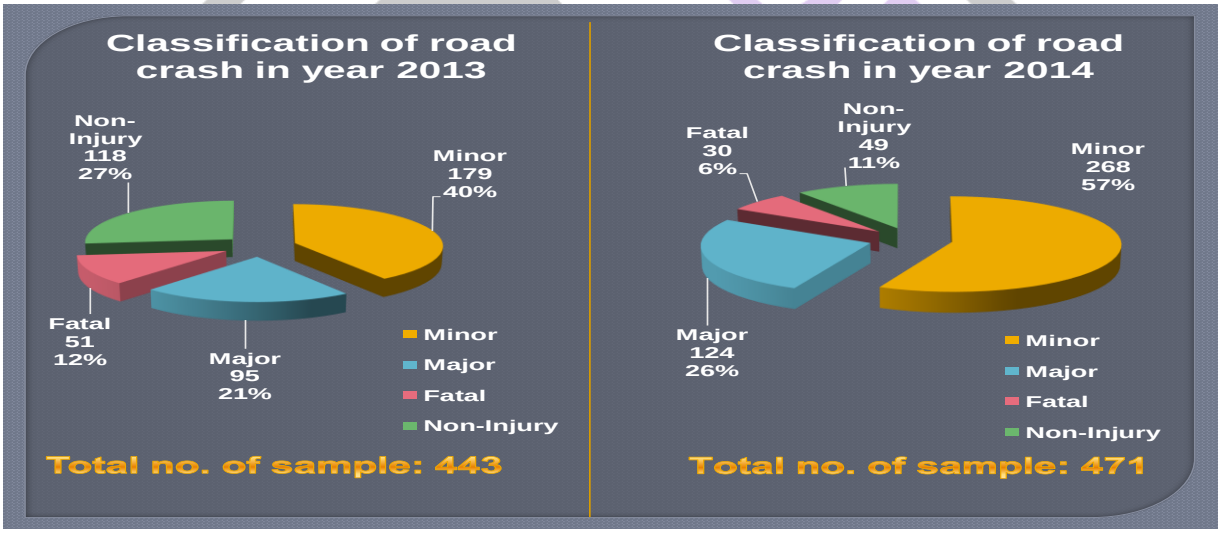

Fig-2. Classification of road crash

V. Black spot Identification

A "*black spot*" is defined as any location that exhibits a collision potential that is significantly high when compared with some normal collision potential derived from a group of similar location. Normally the number of road crashes at a particular site will vary widely from year to year, even if there are no changes in traffic or in the road layout. In statistical terms, road crashes at individual sites are rare, random, multifactor events. This means that comparison between the numbers of road crash at particular site must be made with respect to a fixed time period, typically one year. Furthermore, a single year data will be subjected to considerable statistical variation. Ideally, several years data are required, from which a mean, annual road crash rate can be calculated. Three years is generally regarded as a practical minimum period for which a reasonably reliable annual average rate can be calculated. Due to non availability of the data for three years, in this analysis one year data has been considered to discover the road crash prone locations.

Location of Black-Spot

Based on the exploratory analysis of the data and employing four different methods for the detection of the road crash prone

locations the outcome of the absolute investigation has been presented on the map of country.

As a prelude to the model development, it was felt essential to understand the correlation amongst the variables being considered for model development for that purpose correlation analysis has been performed and same is elaborated in next chapter.

Fig-3. Black spot location on section wise

VI. Linear Correlation - Categorical Regression Analysis (CATREG)

CATREG extends the standard approach by simultaneously scaling nominal, ordinal and numerical variables. The procedure quantifies categorical variables so that the quantifications reflect characteristics of the original categories. The procedure treats quantified categorical variables in the same way as numerical variables. The goal of categorical regression with optimal scaling is to describe the relationship between a response variable and a set of predictors. By quantifying this relationship, values of the response can be predicted for any combination of predictors. The variables considered for this analysis include Time of Road crash, Nature of Road crash, Cause of Road crash, Vehicle involved in accident, Classification of Road crash. The analysis performed considering classification of road crash as response variable and all other variables as predictor variable. Table-2 below shows matrix obtained from analysis.

	TIMEOFCRASH	NATUREOFACCIDENT	CAUSES	VEHICLEINVOLVED
TIMEOFCRASH	1.000	.013	-.057	-.224
NATUREOFACCIDENT	.013	1.000	.073	.137
CAUSES	-.057	.073	1.000	.108
VEHICLEINVOLVED	-.224	.137	.108	1.000
Dimension	1	2	3	4
Eigenvalue	1.317	1.023	.929	.731

A close look at the above table reveals that there is no linear correlation amongst the independent variables ranging between -0.224 and 0.137 indicating the absence of multi co linearity amongst the chosen variables.

Model Summary

Multiple R	R Square	Adjusted R Square	Apparent Prediction Error
.485	.235	.226	.765

Model Summary

Multiple R	R Square	Adjusted R Square	Apparent Prediction Error
.485	.235	.226	.765

Dependent Variable: CLASSIFICATIONOFROADCRASH
Predictors: TIMEOFCRASH NATUREOFACCIDENT CAUSES VEHICLEINVOLVED

By studying above table-2 it has been observed that no linear correlation exists amongst variables, indicating an absence of multi co linearity amongst the chosen variables. By performing linear correlation and analysing the results it can be concluded that partial linear correlation amongst the independent variables are very low and hence all the variables can be considered for model building.
The Correlation matrixes derived in this study to assess the linear correlation amongst the variables are indicative of the fact that the variables chosen for the development of the model are each an unique entity and hence can be applied for model development in its existing form and the developed model is explained in next section.

VII. Road Crash Severity Prediction Model

The development of a severity model begins by conditioning on the event, to use a phrase from statistics. The models focus on the injury severity of observed Road crash. The data was analyzed to predict the severity of the road crashes using various identified independent variables. The goodness of fit of the developed models was analyzed and their applicability to other datasets was tested through validation. The percentage accuracy of the models during calibration as well as validation was also tested.

Models popularity is due to the fact that the formula for the choice probabilities takes a closed form and is readily interpretable. A limitation of the Multinomial Logistic regression model is that it has the "Independence from Irrelevant Alternative"- IIA property. The IIA property implies that the relative probability of choosing between any two alternatives is independent of all other alternatives. Correlation among unobserved factors across alternatives makes the MNL regression model ineffective under these conditions. In addition, the MNL regression model does not account for the ordinal nature of discrete data and thus the ordering information for accident severity (ranked for e.g. fatal, major, minor and non injury) is lost. The Multinomial Logistic regression model is often used to predict the Road crash severity. Although MNL regression models do not recognize order in injury levels, they do avoid certain restrictions posed by standard ordered models. They allow variables to have opposing effects regardless of injury order. Thus, MNL regression models are still applied in many studies for the Road crash severity analysis.

To calibrate the Road crash Severity Prediction Model considering MNL regression, Statistical Packages for Social Studies (SPSS) software has been used. The road crash severity prediction model was calibrated by considering 70 per cent of the dataset and the remaining 30 per cent of the data was used for validation purpose as described in next section. The statistical output generated from the MNL regression model by using SPSS is presented in below Table 3 and 4.

Table-3: Multinomial Logistic regression Model Output of Model fitting Information for Road Crash Severity Prediction

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	Degree of freedom	Sig.
Intercept Only	1.207E3			
Final	933.736	272.848	75	1.000

Table-4: Multinomial Logistic regression Model Output of Goodness of fit for Road Crash Severity Prediction

	Chi-Square	df	Sig.
Pearson	1031.638	972	.090
Deviance	773.207	972	1.000

From Table 3, it can be inferred that the presence of a relationship between the dependent and independent variables. The significance value for the model is 1.000 which is more than 0.05 indicating that the null hypothesis that there is no difference between the model having the set of independent variables and the model having no independent variable cannot rejected. The existence of a relationship between the independent variable and the dependent variable is hence established. From Table 4, it can be concluded that the high accuracy of goodness of fit wherein the significance values of Pearson is lesser than 0.05 indicating that the proposed model fits the present data set with accuracy. The reference category for calibration was chosen to be 'non injuries' as they form the largest category amongst accident severity.

By analyzing the above data, it can be seen that all the considered parameters are significant in predicting accident severity and the standard error is also quire less signifying the accuracy of the model. From the developed model, prediction table based on the parameter estimates for the calibration sample is shown in Table 5.

Table-5: Prediction Table for Road Crash Severity Prediction Model during Calibration

Observed	Predicted				
	Fatal	Major	Minor	Non-injury	Percent Correct
Fatal	6	6	22	4	15.8%
Major	2	57	69	14	40.1%
Minor	5	46	144	60	56.5%
Non-injury	0	0	59	95	61.7%
Overall Percentage	2.2%	18.5%	49.9%	29.4%	51.3%

From Table 6.5, it can be seen that an overall prediction accuracy of 51.3 per cent with 15.8 percent prediction accuracy for fatal accidents, 40.1 per cent prediction accuracy for major accidents and 56.5 per cent prediction accuracy for minor injuries and 61.7 per cent prediction accuracy for non injury crash has been observed.

VIII. Validation of prediction model

External Validation- 70-30 Cross Validation Strategy

In this validation strategy, the cases are randomly divided into two subsets: a training sample containing 70% of the cases and a holdout sample containing the remaining 30% of the cases. The training sample is used to derive the multinomial logistic regression model. The holdout sample is classified using the coefficients for the training model. The classification accuracy for the holdout sample is used to estimate how well the model based on the training sample will perform for the population represented by the dataset.

If the classification accuracy rate of the holdout sample that is no less than 10% lower than the accuracy rate for the training sample (greater than 0.90 * training accuracy rate), it is deemed sufficient evidence of the utility of the logistic regression model. The probabilities were then used to estimate the accuracy of the model in terms of accident severity for the validation sample. The prediction table for the validated sample is presented in Table-6.

Table-6: Prediction Table for Road crash Severity Prediction Model during Validation

Observed	Predicted				
	Fatal	Major	Minor	Non-injury	Percent Correct
Fatal	1	8	4	1	7.1%
Major	0	15	26	1	35.7%
Minor	1	12	127	9	85.2%
Non-injury	1	0	52	24	31.2%
Overall Percentage	1.1%	12.4%	74.1%	12.4%	59.2%

From Table-6, it can be seen that an overall prediction accuracy of 59.2% with 7.1% prediction accuracy for fatal, 35.7% prediction accuracy for Major, 85.2% prediction accuracy for minor injuries and 31.2% prediction accuracy for non injury accident is observed. From all these calibration and validation results, it can be concluded that the developed model would be able to predict the non accidents with reasonable accuracy, but the fatal accident prediction rate is quite poor.

Conclusion:

- The preventive measures brought through this study further control us to control or cut down these rates by utilizing diverse new safety measures, infrastructural configuration fatalities and most recent vehicle engineering.
- The central purpose of mishap aversion and control methodology is depending on 4 E's, vis. (i) Education, (ii) Enforcement, (iii) Engineering and (iv) Environment and Emergency consideration of road accident exploited people.
- A significant observation from this analysis is more than 75 per cent of road crash has occurred when the road condition is straight and weather condition is sunny day. Straight road may improve speed but monotony resulting from the straight alignment is hazardous for road safety. Straight road also reduces the alertness of driver.
- Multinomial logit model for severity prediction has yielded encouraging results concerning the prediction of the grievous injury and non injury type road crash. Taking into consideration the complex structure of database and presence of anomalies in the dataset, the prediction accuracy of the combined model for grievous injury (40 percent accuracy) is reasonable and for non injury (61.7 percent accuracy) it is satisfactory.
- As the validation of combined severity model is not exhibiting fair level of prediction accuracy the improvement in prediction can be made by adopting other analytical techniques.

References

[1] Agresti, A. (2007). "An Introduction to categorical data analysis." 2nd Edition, John Wiley and Sons, Inc., New York.
[2] Bobade and kulkarni "Identification of Accident Black Spots on National Highway 4 (New Katraj Tunnel to Chandani Chowk)".

[3] Dinu R.R and Veeraragavan A (2011). "Random parameter models for accident prediction on two-lane undivided highways in India", Journal of safety Research 42(2011) 39-42, 2011.
[4] Meshram K. and Goliya H.S (2013) "Accident Analysis on National Highway-3 between Indore to Dhamnod" International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 2, Issue 7, July 2013.
[5] Parikh V.A., Jain A.M.,(2014) Development of an accidental model for Urban area © 2014 IJEDR | Volume 2, Issue 2 | ISSN: 2321-9939
[6] Shah p.m, S. Padma, Dr. E.Madhu, S. M. Dave (2011). "Probabilistic Models for the Prediction of Road Crash Severity on High Speed Corridors."
[7] Shankar, V., J. Milton, and F. Mannering (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. Accident Analysis and Prevention, Vol. 29, No.6, pp. 829-837, 1997.
[8] Williamson M. and Zhou H. (2012). "Develop Calibration Factors For Crash Prediction Models For Rural Two-Lane Roadways In Illinois"Procedia - Social and Behavioral Sciences 43, 2012.