JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020

A Generalized Linear Regression Model for Accident Prediction

Donta Sandhya Rani[1], SK Swathi[1]

[1]*Department of Civil Engineering, Sree Dattha Institute of Engineering and Science,Hyderabad, Telangana, India*

ABSTRACT

The purpose of this paper is to analyze the relationship between the number of road traffic accidents and road length, traffic conditions and other factors. Taking the number of road traffic accidents subject to Poisson regression, negative binomial (NB) regression and Zero Inflated Negative Binomial (NINB) regression as response variables, we construct a generalized linear model by introducing a joint function. We construct the Traffic Accident Prediction Model Based on Random Forest (RF) Regression. The defect models are compared, and based on the predictive model, selecting the significant factors, and determining the degree of influence factors of road traffic accidents, reducing the number of traffic accidents and improve the overall security of the road.

Keywords: *Accident prediction, traffic conditions, random forest regression.*

1. INTRODUCTION

A traffic collision, also called a motor vehicle collision (MVC) among other terms, occurs when a vehicle collides with another vehicle, pedestrian, animal, road debris, or other stationary obstruction, such as a tree, pole, or building. Traffic collisions often result in injury, death, and property damage.Several factors contribute to the risk of collision, including vehicle design, speed of operation, road design, road environment, and driver skill, impairment due to alcohol or drugs, and behavior, notably speeding and street racing. Worldwide, motor vehicle collisions lead to death and disability as well as financial costs to both society and the individuals involved.In 2013, 54 million people sustained injuries from traffic collisions. This resulted in 1.4 million deaths in 2013, up from 1.1 million deaths in 1990. About 68,000 of these occurred in children less than five years old. Almost all high-income countries have decreasing death rates, while most low-income countries have increasing death rates due to traffic collisions. Middle-income countries have the highest rate with 20 deaths per 100,000 inhabitants, 80% of all road fatalities by only 52% of all vehicles. While the death rate in Africa is the highest (24.1 per 100,000 inhabitants), the lowest rate is to be found in Europe (10.3 per 100,000 inhabitants).Traffic collisions can be classified by general type. Types of collision include head-on, road departure, rear-end, side collisions, and rollovers.Many different terms are commonly used to describe vehicle collisions.

The World Health Organization use the term road traffic injury, while the U.S. Census Bureau uses the term motor vehicle accidents (MVA), and Transport Canada uses the term "motor vehicle traffic collision" (MVTC). Other common terms include auto accident, car accident, car crash, car smash, car wreck, motor vehicle collision(MVC), personal injury collision (PIC), road accident, road traffic accident (RTA), road traffic collision (RTC), and road traffic incident (RTI) as well as more unofficial terms including smash-up, pile-up, and fender bender.Some organizations have begun to avoid the term "accident". Although auto collisions are rare in terms of the number of vehicles on the road and the distance they travel, addressing the contributing factors can reduce their likelihood. For example, proper signage can decrease driver error and thereby reduce crash frequency by a third or more. That is why these organizations prefer the term "collision" to "accident". In the UK the term "incident" is displacing "accident" in official and quasi-official use.

1.1. Causes of Road Accidents

A 1985 study by R. Kumar, using British and American crash reports as data, suggested 57% of crashes were due solely to driver factors, 27% to combined roadway and driver factors, 6%

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020

to combined vehicle and driver factors, 3% solely to roadway factors, 3% to combined roadway, driver, and vehicle factors, 2% solely to vehicle factors, and 1% to combined roadway and vehicle factors. Reducing the severity of injury in crashes is more important than reducing incidence and ranking incidence by broad categories of causes is misleading regarding severe injury reduction. Vehicle and road modifications are generally more effective than behavioral change efforts apart from certain laws such as required use of seat belts, motorcycle helmets and graduated licensing of teenagers.The main objective of the project is to completely study the traffic, accidents, causes of accidents and the reason for the accident and hence make a suitable mathematical model using regression method between the time and accident and hence we shall predict the accidents in the future and make sure we prevent the accidents by suitable preventive measures.

2. TRAFFIC STUDY

Traffic Data Collection and projections thereof of traffic volumes are basic requirements for planning of road development and management schemes. Traffic Data forms an integral part in the science of descriptive national economics and such knowledge is essential in drawing up a rational transport policy for movement of passengers and goods by both government and the private sectors.This Guideline considers the fact that traffic flow data is important in planning of a section of the road network and for its subsequent maintenance. Traffic flow pattern appears to be random in distribution, as it reflects people's motivation in terms of different composition of vehicles on different types of roads under varying environmental conditions. It follows then that data being collected is a methodological statistic, because traffic flow pattern follows a random distribution. Despite such complexities, it does follow fairly and clearly defined patterns that are possible to classify and analyze. Thus, traffic data collection and analysis follow varying trends and plays an important role in the evaluation and management of road network schemes.

2.1. Resources Required for Collection of Traffic Data

Assessment of available resources prior to commencement of any activity is critical to any assignment at hand. For traffic data collection, it is important that proper assessment of the extent or scope of the envisaged counting (quality level of data required) is undertaken. This is aimed at ensuring that the planned and organized exercise is achieved at optimal cost and with the expected accuracy.

Straight Roads: Traffic counting on a straight road is done by traffic enumerators who stand by the roadside, counting and classifying the vehicles as they pass. The enumerator thus record vehicles moving in one direction. In this case there is no complexity if the level traffic is less than 1000 vehicles per day.

Urban Roads: In the context of this Guideline, an urban road is a road located and/ traversing a developed or built-up environment. This type of road may serve as a main arterial or transit route within the urban area, local connector, tertiary, access or even a local street.As a result, traffic counting for these types of roads can be complex as the function of the road and/or its level of service in the road hierarchy as measured by the traffic flow level dictate it. Further complexity could be presented by the proximity of the access intersections associated with the built environment. On this basis, both manual and automatic counting systems are suitable for traffic data collection along these roads

Rural Roads: These are roads ranging from inter-urban main trunk roads to local minor access roads within a rural set up. However, the emphasis within the confines of these guidelines are placed on the higher order type of roads, such as inter-urban trunk roads, tertiary, connector and main access roads within a rural built up area or between the rural built environment.These roads could be counted using both manual and automatic counting systems, depending on the level of traffic flow, capacity of the road and resources required to undertake the counts. If counting of these roads is not intended to include intersection or is not undertaken within a built environment the sites should be planned and sighted in an area free of disturbance.

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020

Dual Carriageways: Dual carriageways are roads consisting of more than one driving lane in each direction irrespective of its location. This is whether the road is within an urban or rural environment and it can range from inter-urbanfreeways to low volume rural connectors, depending on the level of traffic to be served. Functionally, upgrading of single carriageway roads to dual carriageways is a direct result of increasing traffic demand, and it is therefore provided to cater for capacity expansion and improve level of service.Being a high traffic volume road, it is not always easy to efficiently conduct manual traffic counts on these roads. For efficient collection of traffic flow data on dual carriageways, automatic counters are the most appropriate. This considers the volume of traffic and the speed with which vehicles are passing a counting point. However, enumerators could be assigned for manual counting on dual carriageways by allocating each enumerator a lane per direction of flow or just by the direction of traffic flow. This approach will require more enumerators than it is the case with single carriageway roads.

3. TYPICAL CONVERSION OF TRAFFIC COUNT

Main input parameters for design of the road are the Annual Average Daily Traffic (AADT) and the cumulative loading over the design life of the road (normally 20 years), that is the number of vehicles passing a point in both directions per day taking into account the variation inthe traffic flow throughout the year and the total number of axles for the same traffic volume. Determination of the AADT from 12-hour traffic count is achieved by converting to 16-hour flow (the volume of traffic flow counted in hours) by using applicable conversion factors. Having obtained the 16-hour counts, a further conversion to 24-hour flow may be carried out to obtain an Average Daily Traffic flow, and subsequently to Annual Average Daily Traffic. For illustration, the following conversion actors have been used in the calculations.

Scenario	Urban	Inter-urban	Recreation
High	.016	.115	.27
MediumLow	.000	.060	.14
	0.989	.016	0.96

Conversion of Average Daily Traffic to Annual

Average Daily Traffic: Annual Average Daily Traffic is the average traffic that is expected to usea particular road over a year (365 days). The Average Daily Traffic, con- version to Annual Average Daily Traffic is determined from the following expression:

AADT = T-ADT /365

Where: AADT = Average Annual Daily Traffic.

T-ADT = Total Average Daily Traffic

Conversion of Peak Hour Traffic to Average Daily Traffic (ADT)

Peak hour traffic used for design is the traffic, which passes a point during the severest peak hour(s) of the counting period. To convert peak hour traffic to Average Daily Traffic (ADT), the peak hour traffic should first be converted to 12 hour or 16-hour traffic flow and then to 24-hour traffic flow. For instance, if peak hour flow is 10% of 16-hour counts, then for any given number of vehicles, ADT is given by the following:

Peak hour flow * Conversion factor = ADT (16-hour)

Then, ADT (16-hour) * Conversion factor = ADT (24-hour)

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020

The conversion factor is the proportion of traffic flow over a given peak time as it relates to that prevailing traffic counted under same traffic conditions and over a specific counting period.

Conversion of Day Time Traffic to Average Daily Traffic

To convert Day Time Traffic to Average Daily Traffic and subsequently to Annual Average Daily Traffic, derived factors based on the duration of counts shall be used. Toillustration, the following has been assumed:

- Seven (7) day counts is conducted on a busy rural main road.
- Constant 16-hour traffic flow counts from Monday to Friday of 10 000 vehicles each has been obtained
- A further 16-hour constant traffic flow for Saturday and Sunday of 8 000 vehicles each was also obtained
- Calculation-day 16-hour traffic flow =(5 x 10 000) + (2 x 8 000)=66,000 vehicles
- Using a 95% confidence limit for the 24-hour traffic flow with 5% tolerance.
- Then, 16-hour traffic flow is 95% of 24-hour traffic flow, therefore.
- 7 days 24-hour traffic flow = 66 000 / 0.95 = 69 474 vehicles
- Average Daily Traffic (ADT) = 69 474/7 = 9925 vehicles

As for the Annual Average Daily Traffic (AADT), the derived Day Traffic s converted as follows:

AADT =9925 x conversion factor

=9925 x 1.141 (considering medium scenario)

=11 324 vehicles

4. DATA COLLECTION

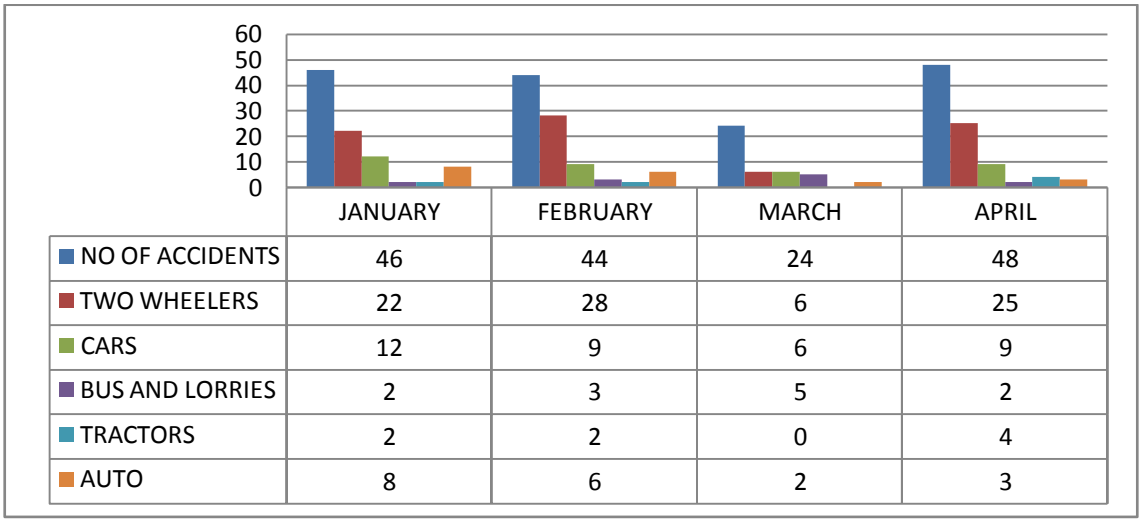
ACCIDENT DATA COLLECTION

MONTH	NO OF ACCIDENTS
JANUARY	46
FEBRUARY	44
MARCH	24
APRIL	48
MAY	51
JUNE	40
JULY	38
AUGUST	28
SEPTEMBER	28
OCTOBER	31
NOVEMBER	25
DECEMBER	33

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020

EFFECT OF ACCIDENTS

	No of persons involved in accidents	Death	Grievous	Minor injuries
JANUARY	69	6	38	25
FEBRUARY	63	3	40	20
MARCH	45	9	25	11
APRIL	72	12	33	27
MAY	80	3	40	37
JUNE	57	6	34	17
JULY	58	3	28	27
AUGUST	46	2	24	20
SEPTEMBER	45	1	25	19
OCTOBER	47	2	23	22
NOVEMBER	35	3	20	12
DECEMBER	53	3	20	30



5. RESULTS AND CONCLUSIONS

5.1. ANALYSIS

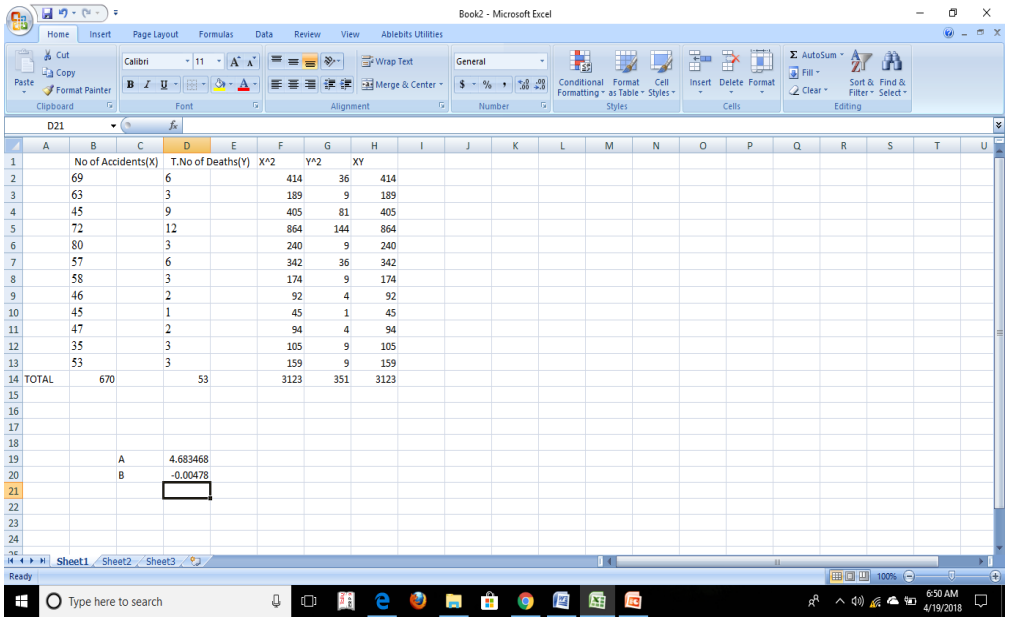
In this case we consider the variables X and Y as total no of accident and total no of deaths respectively and the constants as A and B

Let us assume the equation will be of the form Y=AX+B

$$A=((\sum Y*\sum X^2)-(\sum X*\sum XY))/((n*\sum X^2)-(\sum X*\sum X))$$

$$B=((n*\sum XY)-(\sum X*\sum Y))/((n*\sum X^2)-(\sum X*\sum X))$$

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020



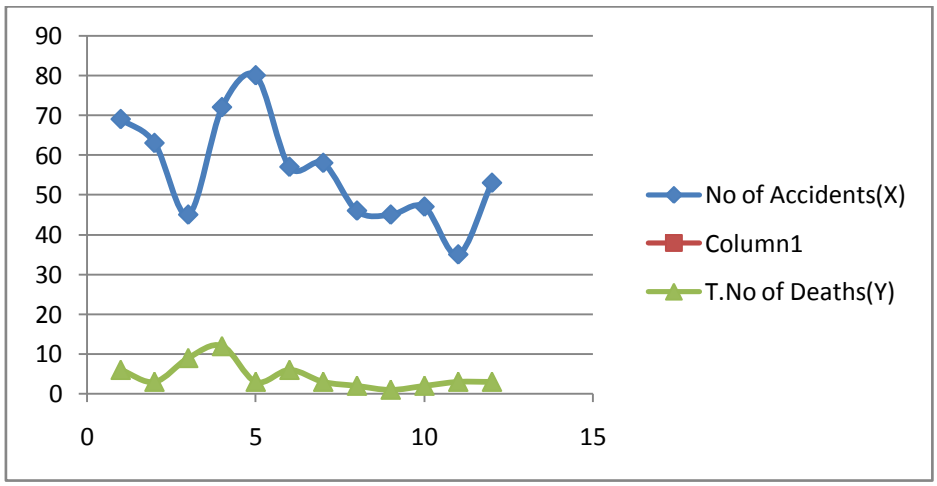
From the above calculations we got

A= 4.683468

B= -0.00478

Hence the regression equation for the accidents is

Y=4.683468X-0.00478



R^2 Values of Variables

S.No	Variable (X)	Expression	Relationship	R^2
1	INJURIES	$Y=0.0024X^3+0.1483X^2-2.3587X+19.481$	POLYNOMIAL	0.7007
2	DEATH	$Y=0.0959X^3+1.1919X^2-2.993X+14.431$	POLYNOMIAL	0.5021
3	2W	$Y=0.0601X^2+1.7711X-7.68$	POLYNOMIAL	0.7039
4	AUTO	$Y=0.0695X^3+1.2273X^2-5.8668X+20.245$	POLYNOMIAL	0.3262
5	LORRY	$Y=1.8833X^3+10.5X^2-16.217X+20$	POLYNOMIAL	0.2499
6	TRACTOR	$Y=0.8833X^2-3.5X+16.867$	POLYNOMIAL	0.0534
7	CAR	$Y=1.5833X^3+8.25X^2-7.6667X+19.481$	POLYNOMIAL	0.6169

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020

5.2. Multiple Linear Regression Analysis

Multiple Regression analysis (MLR) is a system for examining the relationship of a collection of independent variables to a single dependent variable. Multiple linear regression analysis was one of the first and simple methods of analysis taken into consideration for the model development, which gave satisfactory performance. This technique is still being used in the development of simple models. The data used for the model development of no of accidents per year is

No of Accidents(Y)	INJURIES(X1)	DEATHS(X2)	2W(X3)	CAR(X4)
46	25	6	22	12
44	20	3	28	9
24	11	9	6	6
48	27	12	25	9
51	37	3	24	10
40	17	6	18	12
38	27	3	20	8
28	20	2	15	11
28	19	1	24	9
31	22	2	19	5
25	12	3	10	9
33	30	3	22	15

From the matrix, the constant values which we have got are as follows,

a_0=43.60122582
a_1=0.010192044
a_2=7.23865E-13
a_3=0.008516712
a_4=3.26086E-07

5.3. Accident Prediction Model

We develop an accident prediction model for the road length which we have selected by using linear regression analysis. The primary objective of the study was to develop a model to predict any future accidents. For this study, we considered the variables like injuries, deaths and type of vehicles. Road width is not considered since the width is same along the stretch. The accident prediction model developed is as given below:

$$Y_g = a_0 + a_1 X_1^2 + a_2 X_2^2 + a_3 X_3^2 + a_4 X_4^2$$

Y	Y_g	SSR	SSE
46	4.460462	92.92069981	133.1609456
44	25.56121	131.35953	73.294330307
24	18.06479	15. 719523768	101.2999417
48	5.688706	70.74986064	53.45501455

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020

51	5.054613	81.81902085	24.45684977
40	1.683114	154.1790627	86.80436832
38	0.373439	188.4184769	58.1644327
28	18.47593	19.1487294	2.17835791
28	31.60012	306.254575	134.5628221
31	30.03762	254.0077352	81.67857747
25	17.55656	17.45665	6.216465
33	31.1654	8.54651	65.4684

Y = No of Accidents Occurred

Y_g = Generated Accidents

SSR = Sum of Squres of Regression

SSE = Sum of Squres of Error

SST = Total Sum of Squares

SSR = 1340.579894

SSE = 820.740505

SST = SSR + SSE = 2161.320399

R^2 = Co-efficient of Determination = SSR/SST = 0.62025

6. CONCLUSION

Accident Prediction Model is developed using Multiple Linear Regression Analysis for this part of roas is based on the factors influencing road accidents. The dependent variable using in thr model is Number of Accidents (Y). The independent variables used in the model are:

1. Injuries (X_1)
2. Deaths(X_2)
3. Two Wheelers (X_3)
4. Cars(X_4)

The model development in the research using the above variables is:

$$Y_g = a_0 + a_1X_1^2 + a_2 X_2^2 + a_3 X_3^2 + a_4 X_4^2$$

The Coefficient of Determination (R^2) obtained is 0.62025

Accident data from different police station suggests that there is a lack of proper enforcement and education to roadway safety. This weakness can be minimized through comprehensive corrective measures. Local community initiatives to improve the conditions are very sparse. Importantly, such efforts would require considerable resources particularly trained local personnel, safety specialists and researchers so as to build up indigenous capacity and attain sustainable safety program.It is suggested to further refine the model reported in this study using more number if variables to get a more realistic picture in the predicting or forecasting accidents, though accidents occurrence is random phenomenon and therefore we cannot exactly predict future trends by using any model or theory, but it is a very handy tool in the hands of planners and decision makers to take remedial measures in advance by studying future trends using such models, to take mitigation measures to minimize the accident rate to certain extent and to take other safety measures.

ISSN- 2394-5125 VOL 7, ISSUE 12 , 2020

REFERENCES

[1] S. Al-Ghamdi, "Using logistic regression to estimate the influence of accident factors on accident severity," *Accident Analysis and Prevention*, vol. 34, no. 6, pp. 729–741, 2002.
[2] J. C. Milton, V. N. Shankar, and F. L. Mannering, "Highway accident severities and the mixed logit model: an exploratory empirical analysis," *Accident Analysis and Prevention*, vol. 40, no. 1, pp. 260–266, 2008.
[3] M. Bédard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities," *Accident Analysis and Prevention*, vol. 34, no. 6, pp. 717–727, 2002.
[4] K. K. W. Yau, H. P. Lo, and S. H. H. Fung, "Multiple-vehicle traffic accidents in Hong Kong," *Accident Analysis and Prevention*, vol. 38, no. 6, pp. 1157–1161, 2006.
[5] T. Yamamoto and V. N. Shankar, "Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects," *Accident Analysis and Prevention*, vol. 36, no. 5, pp. 869–876, 2004.
[6] K. M. Kockelman and Y. Kweon, "Driver injury severity: an application of ordered probit models," *Accident Analysis and Prevention*, vol. 34, no. 3, pp. 313–321, 2002.
[7] L. Chang and H. Wang, "Analysis of traffic injury severity: an application of non-parametric classification tree techniques," *Accident Analysis and Prevention*, vol. 38, no. 5, pp. 1019–1027, 2006.
[8] J. de Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accident Analysis and Prevention*, vol. 43, no. 1, pp. 402–411, 2011.
[9] M. Simoncic, "A Bayesian network model of two-car accidents," *Journal of Transportation and Statistics*, vol. 7, no. 2-3, pp. 13–25, 2004.
[10] K. Ozbay and N. Noyan, "Estimation of incident clearance times using Bayesian Networks approach," *Accident Analysis and Prevention*, vol. 38, no. 3, pp. 542–555, 2006.