LINEAR REGRESSION ANALYSIS ON THE CONTRIBUTORY FACTOR OF ACCIDENT IDENTIFIED IN ROAD SAFETY AUDIT USING PYTHON

Sreeparvathy CM¹, Chitturi Sravanti¹
¹Assistant Professor, Mahatma Gandhi Institute of Technology, Hyderabad
Corresponding Author Email: cmsreeparvathy_civil@mgit.ac.in

ABSTRACT

The rising number of road accidents in Indian Scenario is quite alarming. The analysis of cause of these accidents will help to adopt remedial measures for the same. In this paper based on the data collected using Road Safety Audit (RSA), an attempt is made to analyze and understand the root cause for accident. The linear equation and relationship is developed using Python programming which is a competent tool for analysis. In this paper, the analysis is carried out on data collected along a stretch of 12 km of length of road from Langerhouz (suburban Road) to MGIT. RSA was conducted for the same stretch that has six intersections falling along the stretch. The basic methodology used for data collections were Prompt list, volume and accident data collection. The accident data were analyzed statistically and a linear regression model was generated with various causes of the accident as the independent variable. For the purpose of analysis, the stretch was divided into four sub stretch based on the similarities in the geometric properties of the road. RSA was also performed for the six intersections falling in this stretch. The data collected on accident has reflected on a number of contributory factors from which the main causative factors were identified and recommendations were made to improve safety with the best possible solutions. Regression analysis was conducted for the accident data obtained from the police station for the years 2014 – 2018. The solution of the regression equation is developed using Python programming language. The relation to dependent variable, accident was formulated linearly using the independent variable based on the factors contributing to accident.

Keywords: Accident, RSA, Linear Regression Analysis, causative factors, data, Python Programming.

1. INTRODUCTION:

Road traffic endures to be a chief progressive concern, a communal well-being concern and is a principal cause of death and injury across the world. Road accidents in India take the life of almost 1.5 lakh people per annum. Consequently, our country accounts for almost 11% of the accident related deaths in the World. At The third Global Ministerial Conference held in Stockholm, Sweden on 19 and 20 February, 2020; all the contributors including India, endorsed their strong obligation for achieving the goals of bringing down road accident related deaths by at least 50% by 2030[10]. But the big question here is how we can reduce the number of accidents with the scenario of growing vehicular traffic and with an imbalance in supply of the road infrastructure [4-7].

A serious analysis of the causative factors of the accident is a crucial measure in preventing the accidents [8-10]. Road Safety Audit (RSA), a recognized technique for assessing accident...
potential and safety performance to ensure safety on roads[11-16]. The RSA pointers were created by CRRI (central road analysis institute) under the MOSRTH (Ministry of Shipping, Road, Transport, Highways) and were adopted by IRC in 2010 as a manual for road safety audit IRC-SP-88: 2010.

The objective of this work is to evaluate existing design of roads and ensuring that safety related design criteria have been met, and also to meet

A. To identify and report on the crash potential and safety problems of a road stretch.
B. To study the cause of the accidents and to suggest corrective treatment measures at potential locations
C. To analyze traffic accident data and establish relationship between dependent and independent using regression models.
D. Identifying the best solution to reduce accidents.

2. Linear Regression in Python

The analysis of the accident data to identify the main cause for accident is carried out using Python coding. The number of independent variables is quite high and finding relationship between the parameters will be a cumbersome process. This challenge could be reduced by solving the equation using python algorithm[17].

3. LITERATURE REVIEW and BACKGROUND:

A Detailed Analysis On Accident Case Studies For NH1, India: carried out by Rajat Sharma and O.P. Mittal, revealed that the study was to lessen the number of accidents on the road and deliver a safer journey to the road users[1].

Analysis of Road Accident Data Of Stretch From Radhanpur Junction to Chanasma Junction:

Patel Sivamkumar conducted investigation of road accident aiming to analyze the traffic accidents happening in a selected stretch by statistical method which is facing stress as reduced level of administration and increment in numerous extents of accidents because of massive number of road user clients, especially LMV. It attains investigative review of the mishaps information and proposes remedial measures for attaining safety. Accident data was gathered from numerous police stations along the stipulated stretch[2].

Francis John Gichaga Et Al. Road Safety Audit In India: A Review ISSN: 2347 –4718 [3].
The authors suggested conduct road safety audits. They considered the road client learning, day, night, wet, and dry conditions etc., While doing road safety audit for their better results.
Road Accidents Study Based on Regression Model: a Case Study of Vijayawada city to Hanuman Junction[11]

The Paper debates the importance of an accident prediction model based on regression analysis. It is tried to develop accident model with Vijayawada city as the area of study. The data set which is mainly lethal accidents and total accidents for Vijayawada city is fetched from “Traffic Police Department”. This data were related to hourly classified traffic volume per lane extracted from classified traffic volume count survey of Vijayawada city.

4. STUDY AREA:

The stretch selected begins from Langerhouz and ends at MGIT. The zone under study is about 12 kilometers long and it contained six major intersections as shown in Figure 1. The Figure 2 depicts the nasingi rotary intersection. The identified road stretch passes through the areas of Langerhouz, Ramdev Guda, Narsingi and Gandipet. The six intersections surveyed include Tippu khan bridge intersection, Ramdev Guda intersection, Narsingi Rotary intersection, NCC Urban rotary intersection (shown in Figure 3), Kokapet intersection and Gandipet intersection. The selected stretch of about 12 Km from Langer Houz to MGIT was selected for this study. Base on the similarities, the selected route was divided into four stretches. The vehicular population has amplified drastically in this stretch. Also, the accident count has increased from 94 in 2014 to 126, 135, 193 in the successive years till 2017. This is because the road stretch is connected to the outer ring road of the Hyderabad city. And also, a lot of educational institutions have been in the area in the past few decades which also played a prominent role in increasing the number of vehicles being rode on this stretch.

The stretches selected for data collection are given as follows. The collected data are represented in the tables below.

1. Stretch 1- Langer Houz to Ramdev Guda (3 km)
2. Stretch 2- Ramdev Guda to Kokapet junction (6 km)
3. Stretch 3 - Kokapet junction to Gandipet junction (1.6 km)
4. Stretch 4 - Gandipet junction to MGIT college (1.4km)

Figure 1: Study Area
Figure 2: Narsingi Rotary Intersection
### Table 1: Tipu Khan Bridge Intersection

| FEATURE                  | REMARK              |
|--------------------------|---------------------|
| Type of junction         | T Junction          |
| Types of markings        | Centre line & Edge  |
|                          | markings            |
| Width of road            | 7.28 m              |
| Stopping sight distance  | Appropriate         |
| Lighting at the junction | Available           |
| Is the intersection      | Yes                 |
| lay out obvious to all   |                     |
| road users               |                     |

### Table 2: Ramdevguda Intersection

| FEATURE                  | REMARK              |
|--------------------------|---------------------|
| Type of junction         | T Junction          |
| Types of markings        | Centre line & Edge  |
|                          | markings            |
| Width of road            | 6.42 m              |
| Stopping sight distance  | Appropriate         |
| Lighting at the junction | Available           |
| Is the intersection      | Yes                 |
| lay out obvious to all   |                     |
| road users               |                     |

### Figure 3: NCC Urban Rotary Intersection

### Table 3: Narsingi Rotary Intersection

| FEATURE                  | REMARK              |
|--------------------------|---------------------|
| Type of junction         | Round about         |
| Types of markings        | Border/ Edge        |
|                          | markings            |
| Width of road            | 6.48 m              |
| Stopping sight distance  | Appropriate         |
| Lighting at the junction | Available           |
| Is the intersection      | Yes                 |
| lay out obvious to all   |                     |
| road users               |                     |

### Table 4: NCC Urban Rotary Intersection

| FEATURE                  | REMARK              |
|--------------------------|---------------------|
| Type of junction         | Round about         |
| Types of markings        | Border/ Edge        |
|                          | markings            |
| Width of road            | 6.48 m              |
| Stopping sight distance  | Appropriate         |
| Lighting at the junction | Available           |
| Is the intersection      | Yes                 |
| lay out obvious to all   |                     |
| road users               |                     |
| FEATURE                  | REMARK            |
|-------------------------|-------------------|
| Type of junction        | Y Junction        |
| Types of markings       | Border/ Edge markings |
| Width of road           | 6.48 m            |
| Stopping sight distance | Appropriate       |
| Lighting at the junction| Available         |
| Is the intersection lay out obvious to all road users | Yes |

**Table 6: Gandipet Intersection**

| FEATURE                  | REMARK            |
|-------------------------|-------------------|
| Type of junction        | T Junction        |
| Types of markings       | No markings       |
| Width of road           | 6.42 m            |
| Stopping sight distance | Appropriate       |
| Lighting at the junction| Available         |
| Is the intersection lay out obvious to all road users | Yes |

**Table 7: Stretch Langerhouz to Ramdevaguda**

| FEATURE                  | REMARK            |
|-------------------------|-------------------|
| Type of junction        | Divided           |
| Types of markings       | Faded             |
| Width of road           | 7.28 m            |
| Stopping sight distance | Appropriate       |
| Lighting at the junction| Available         |
| Footpaths               | Not appropriate   |

**Table 8: Ramdevaguda to Kokapet**

| FEATURE                  | REMARK            |
|-------------------------|-------------------|
| Type of junction        | Divided           |
| Types of markings       | Faded             |
| Width of road           | 7.28 m            |
| Stopping sight distance | Appropriate       |
| Lighting at the junction| Available         |
| Footpaths               | Appropriate       |

**Table 9: Kokapet to Gandipet**

| FEATURE                  | REMARK            |
|-------------------------|-------------------|
| Type of junction        | Undivided         |
| Types of markings       | Faded             |
| Width of road           | 6.42 m            |
| Stopping sight distance | Appropriate       |
| Lighting at the junction| Not available     |
| Footpaths               | No footpaths      |

**Table 10: Gandipet to MGIT**

| FEATURE                  | REMARK            |
|-------------------------|-------------------|
| Type of junction        | Divided           |
| Types of markings       | Faded             |
| Width of road           | 7.28 m            |
| Stopping sight distance | Appropriate       |
| Lighting at the junction| Not available     |
| Footpaths               | No footpaths      |
5. METHODOLOGY:

The methodology followed in Road Safety Audit (RSA) is as per on the guidelines given in IRC: SP: 88-2010. The outline of the procedure followed is as given in Figure 4.

![Figure 4 RSA Steps](image)

**DATA COLLECTION:** The data required for the analysis is collected from the stretch M.G.I.T to Langer Houz and they are as follows:

**5.1 SPOT SPEED ANALYSIS:**

Spot speed is the speed or instantaneous speed of a particular vehicle when it passes a section over a given period of time. Spot speed is determined by calculating the time taken by the vehicle to cross the section at a particular spot. These Spot speed studies are generally carried out to frame out the distribution curve for the road system and to analyses the median speed over the specific location on the cumulative frequency graph. Manually the speed spot study has been performed in order to carry out the whole process for our data and the data has been analyzed and depicted in Figure 5.
Figure 5: Speed curve between % cumulative frequency and spot speed

The traffic and number of accidents had increased due to surge in number of vehicles according to the collected data. As the numbers of increasing vehicles cannot be controlled, so speed controlling is the only way to reduces the accidents on the stretch.

General formula for calculating percentile speed is given in Eq(1)[15].

\[ SD = \frac{PD - P_{min}}{P_{max} - P_{min}} \times (S_{max} - S_{min}) + S_{min} \]  \hspace{1cm} (1)

Where,
- \( SD \) = Speed at desired percentile.
- \( PD \) = Desired percentile.
- \( P_{max} \) = Higher cumulative percent.
- \( P_{min} \) = Lower cumulative percent.
- \( S_{max} \) = Higher speed
- \( S_{min} \) = Lower speed

For 85% percentile speed (\( S_{85\%} \))
- \( PD \) = Desired percentile =85%
- \( P_{max} \) = Higher cumulative percent =88%
- \( P_{min} \) = Lower cumulative percent =60%
- \( S_{max} \) = Higher speed =35kmph
- \( S_{min} \) = Lower speed =25kmph

85% percentile of speed, \( S_{85\%} =\frac{85\% - 60\%}{88\% - 60\%} \times (35kmph - 25kmph) + 25kmph \)
$S_{85\%} = 33.92 \text{ kmph.}$

The 85% percentile speed is taken as the safe speed. It is observed from the data that the safe speed is 33.92 kmph

For 50% percentile speed ($S_{50\%}$)

$P_D = \text{Desired percentile} = 50\%$
$P_{\text{max}} = \text{Higher cumulative percent} = 60\%$
$P_{\text{min}} = \text{Lower cumulative percent} = 24\%$
$S_{\text{max}} = \text{Higher speed} = 25\text{kmph}$
$S_{\text{min}} = \text{Lower speed} = 15\text{kmph}$

$50\% \text{ percentile of speed, } S_{50\%} = \frac{50\% - 24\%}{60\% - 24\%} \times (25\text{kmph} - 15\text{kmph}) + 15\text{kmph}$

$S_{50\%} = 22.2 \text{ kmph}$

The 50% percentile speed is taken as the median speed. Therefore, median speed from the observed data is 22.2 kmph.

5.2 ACCIDENT DATA FROM POLICE STATION:

The information accessible for accident studies is the FIR (First Information Report) registered in the police stations. The data from these records of last five years (2014-2018) were extracted from the FIR record filed under IPC no.279/337/338/304. The yearly variation of accidents and monthly variation of accidents are given in Figure 6 and 7. The classification of accidents based on type of vehicle depicted in Figure 8. The variations in number of accident with time are depicted in Figure 9. The reasons for accident yearly for 2015-2018 are depicted in Figure 10. This was made possible with the approval of the concerned Superintendent of Police (S.P), Narsingi, Hyderabad. The following data was collected:

1. Date and Time of occurrence of accidents

2. Details of accident i.e., fatal, non-fatal and damage.

3. Reason for accidents.

4. Type of Vehicles involved

5. Details of the victim such as name, gender, age, caste etc.,
Figure 6: Yearly variation of accident

Figure 7: Monthly variation of accident for year 2014-2018

Figure 8: Type of vehicle involved in accident
Accidents do tend to display a specific trend based on the reason of accidents. From the graph, 22% of the accidents are due to drunken driving and the possible reason may be due to presence of three wine shops and two bar shops on the 12 km study stretch. Other main reason for accident is over speeding of vehicles with 15% of accidents.

**ACCIDENT ANALYSIS USING REGRESSION**

Regression analysis is used as a tool to predict the causes of the accidents. The factors which are responsible for an accident are organized in the form of a dataset. This dataset is given as an input to the regression model. Regression is a highly sophisticated technique used to form equations from the data given. Regression analysis helps to comprehend which among the independent variables are related to the dependent variable. This also gives an opportunity to explore the forms of these relationships. In this project, two-wheelers, three-wheelers, four wheelers and pedestrian accidents are dependent variables whereas over speeding, overtaking, wrong route, unmarked cross walks, rash driving, drunken driving, left hand turns, distracted
driving, negligence and failure of vehicle are independent variables. Regression analysis is performed in python (a powerful programming language) using the dataset as an input.

The flow chart of the python programming is given below in figure 11.

**ALGORITHM**

START

Read CSV File using Pandas

Print first 5 lines of dataset

Plot Linearity analysis between dependent and independent variable.

Fit the plot and form the equation

Is the fit acceptable?

NO

YES

Display output as cause of accident with respect to vehicle type

Figure: 11. Flow Chart for Python Programming

The Figure 12 Graphs are plotted between the dependent and independent variables. The independent variables are taken on the Y axis and the dependent variables are taken on the X axis. These graphs give the linearity analysis between the variables. It is found whether the variables are directly proportional or indirectly proportional with each other. Regression package is then imported from sklearn, preprocessing header file. The sklearn preprocessing
consists of many mathematical packages, from which regression package is imported. The independent variables are given as an input to the regression model and dependent variables are the output.

This iteration process is repeated four times to get four equations for two-wheelers, three-wheelers, four wheelers and pedestrian accidents. In order to show the relation between the dependent and independent variables in the form of integers, a correlation matrix is displayed. The one with highest correlation for each dependent variable is considered as the one for the main cause of accident.

By using python, the final output is

OUTPUT:
HIGHEST CAUSE OF ACCIDENT TWO-WHEELER DRIVING: DRUNKEN DRIVING
HIGHEST CAUSE OF ACCIDENT THREE-WHEELER DRIVING: FAILURE OF VEHICLE
HIGHEST CAUSE OF ACCIDENT FOUR-WHEELER DRIVING: OVERTAKING
HIGHEST CAUSE OF ACCIDENT PEDESTRIANS: UNMARKED CROSS_WALKS
6. Linear Regression Model Output:

Following are the linear regression models for the above said output. This output was developed using the Python Programming.

TWO_WHEELER_ACCIDENT_PREDICTION = 60.52659366835968 + (0.8835489180301376)*(Overspeeding) + (-0.013903430201835145)*(Overtaking) +(-0.4746813528889487)*(Wrong route)+ (-0.08788011564534434)*(Rash Driving)+(0.9663859345978911)*(DISTRACTED DRIVING)+(0.6661544460556678)*(DRUNKEN_DRIVING) + (0.19158696195195613)*(NEGLIGENCE) +(-0.29103800105356953)*(FAILURE_OF_VEHICLE)

THREE_WHEELER_ACCIDENT_PREDICTION = -1.3446508772739705 +(-0.6791804683191274)*(Overspeeding)+( 0.10638585421992523)*(Overtaking) + (0.6175515387640259)*(Wrong route)+(0.022439570196344686)*(Rash Driving) +(1.0366055684206095)*(DISTRACTED DRIVING)+( 0.06163588534062545)*(DRUNKEN_DRIVING)+(-0.1211628280348451)*(NEGLIGENCE)+( 0.269510463150262)*(FAILURE_OF_VEHICLE)

FOUR_WHEELER_ACCIDENT_PREDICTION = -38.67083306197179 + (0.541033638632203)* (Overspeeding)+ ( 1.0364571783294403)*(Overtaking) +(0.4846882035648246)*(Wrong route)+(0.31519003412147684)*(Rash Driving) +(1.5648254479113677)*(DISTRACTED DRIVING)+(0.14405380294950065)* (DRUNKEN_DRIVING) + ( -0.3171766064670485)*(NEGLIGENCE) + ( 0.17030091861710325)*(FAILURE_OF_VEHICLE)

PEDESTRIAN_ACCIDENT_PREDICTION =5.499999999999989+ (0.5384615384615391)*(Negligence)+(1.2692307692307694)*(Unmarked cross walks) + ( 0.19230769230769257)*(Left hand turn)

Table 11: Causes of accident

| S.NO | VEHICLES/ PEDESTRIANS INVOLVED IN ACCIDENTS | REASONS FOR ACCIDENTS |
|------|------------------------------------------|-----------------------|
| 1    | Two-wheelers                             | Drunken Driving       |
| 2    | Three-wheelers                           | Failure of Vehicle    |
| 3    | Four-wheelers                            | Over taking           |
| 4    | Pedestrians                              | Unmarked Cross Walks  |
7. CONCLUSIONS:

Based on the present study of RSA and accident analysis for Langerhouz to M.G.I.T the following suppositions have been concluded:

1) Problems were identified on the basis of prompt lists at the Intersections. Most of the problems at the intersections were found to be Improper delineation, Improper road markings, Lack of traffic Islands, Unauthorized parking, Unevenness of pavement profile, Sign board obstruction, Lack of reflectors etc.

2) Based on the prompt lists at the stretch, the complications identified at the stretch were insufficient street lights, Absence of traffic lights, Inappropriate sign boards, Improper lane markings, Bifurcation of lanes, Absence of zebra crossings, Lack of provision for cyclists and pedestrians, Improper reflectors etc.,

3) On the basis of accident data given by Narsingi Police station, the major reasons for accidents were identified with the help of regression analysis. The causes of accidents in the stretch were found out to be:

Acknowledgement:
This paper and the work and data collection involved would not have been probable without the exceptional support of my student T. Keerthi, students of Mahatma Gandhi Institute of Technology.

REFERENCES

1. Rajat Sharma, O.P. Mittal; Review Paper on “A Detailed Analysis on Accident Case Studies for NH 1, India”; International Journal of Advanced Research in Education & Technology (IJARET) 226 Vol. 3, Issue 2 (April - June 2016).

2. Patel Savan Kumar, “Analysis of road accident data of stretch from Radhanpur junction (Mahesana) CH 00.00 km (SH-55) to Chanasma junction CH 32.80 km (SH-55), IJAIEM, Volume 3, Issue 12, December 2014, ISSN 2319 – 4847.

3. Francis John Gichaga, “The Impact of Road Improvements on Road Safety and Related Characteristics”; IATSS Research (2016), University of Nairobi, Kenya.

4. Chetan R. Mankar, N.H.Pitale; RSA: An Accident Studies of selected stretch road;ISSN 2278-0211, Vol 3 Issue 5, May 2014.

5. Ashutosh Kotkar; Road Safety Audit; International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181; IJERTV7 IS0 4086, Vol. 7 Issue 04, April-2018.

6. M.Priyanka, B.H.Nagendra Rao; Road Safety Audit; IJARIIE-ISSN(O)-2395-4396; Vol-3 Issue-6 ,2017.
7. Dr. S. S. Jain, P. K. Singh, Dr. M Parida; Road Safety Audit for four lanes national highways; 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA.

8. N. Naveen, Dr. T. Ilango; Road safety audit of Ananthagiri hills ghat road at Vikarabad district; International Journal of Civil Engineering and Technology (IJCIET) Volume 9, Issue 4, April 2018, pp. 816–825.

9. Road Accidents 2019-MoRTH

10. Global status report on road safety 2018, WHO, 17 June 2018 | Report

11. Nishnanth.Dulam, Ganesh.Burila, B G Rahul, Dola Sanjay S; “Road Accidents Study Based on Regression Model: a Case Study of Vijayawada city to Hanuman Junction”, International Journal of Scientific & Engineering Research Volume 9, Issue 5, May 2018

12. IRC: SP:88-2010 Manual on road safety audit.

13. L.R. Kadiyali, Principles and Practices of Highway Engineering; Published May 1st 2017 by Khanna Publishers.

14. Nicholas J. Garber, Lester A. Hoel, Traffic and Highway Engineering; Published by Cengage Learning; 005 edition, January 1, 2014.

15. C.S. Papacostas, P.D. Prevedouros, Transportation Engineering and Planning; Published by Pearson; 3 edition, June 5, 2000.

16. https://morth.nic.in/road-safety-engineering-cell - Ministry of Road Transport and Highways.

17. https://en.wikipedia.org/wiki/Python_(programming_language)