Statistical Analysis and Development of Accident Prediction Model of Road Safety Conditions in Hisar City

Keshav Bamel1, Sachin Dass2, Saurabh Jaglan3 and Manju Suthar4
1PG Student, M.Tech. in Highway Safety and Engineering, 2,3,4Assistant Professor.
1,2,3Department of Civil Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat, Haryana, India.
4Department of Civil Engineering, Chandigarh University, Chandigarh, India
1keshavbamel72@gmail.com, 2sachindass@gmail.com, 3saurabhjaglan86@gmail.com, 4manju.e10702@cumail.in

Abstract: The severity of road accidents is a big problem around the world, particularly in developing countries. Recognizing the major contributing variables can help reduce the severity of traffic accidents. This research uncovered new information as well as the most substantial target-specific factors related to the severity of road accidents. T-stat, P-value, Significance and other test values are determined to check the dependency of dependent variable on independent variable in order to obtain the most significant road accident variables. In this research, a comparative analysis of accident data from Hisar and Haryana are compared. According to the findings, Haryana's accident severity index (46.20) was higher in 2019 than Hisar's (36.01), while Hisar had fewer accidents per lakh population (33.34) than Haryana (38.40). The outcomes of the study were used to develop an effective and precise accident predicting model is developed for Hisar city and state Haryana using a statistical method. Four models were created using linear regression analysis, two each for Hisar and Haryana. These models produce good results with a margin of error that is within acceptable bounds (0-5%), allowing them to be used to predict future traffic accidents and deaths.

Key words: Accident severity index, linear regression analysis and Road accidents
1. INTRODUCTION

Transportation system is a vital component of a country's economic growth. It has an impact on the rate, consistency and pattern of growth. National Highways' capacity to handle traffic must keep up with economic expansion. With a total length of 62.16 lakh kilometers India possesses the world's second-largest road network. National Highways (NH), Expressways, State Highways (S.H), Major District Roads (M.D.R), Other District Roads (O.D.R) and Village Roads (V.R) are included in this category. Apart from economic growth road transport has a bigger importance in both passenger and freight traffic than other forms of transportation because of its ease of access, versatility of operations, door-to-door access, and reliability [1–3].

Road safety continues to be an issue of worry because of its magnitude, severity and indirect impact on national economy, human health and local economies. It is a serious nationwide issue. Road traffic injuries now account for a significant portion of all deaths and hospitalizations worldwide, with significant socioeconomic implications. Deaths and injuries caused by traffic accidents result in huge financial losses as well as severe emotional suffering. Road accidents, which is based on data provided by law enforcement agencies across the country, in 2019 India has reported 449,002 accidents in which 151,113 people got killed [4–6].

Hisar city in state Haryana has mixed traffic flow and the city is witnessing an increase in the number of accidents with the growth in economy, as the city houses a large number of industries. In 1990, road accidents laid in ninth place out of over hundred separately identified causes of death and disability. It is forecasted that by the year 2020, road accidents will move up to sixth place. Including Jindal Stainless (Hisar) Limited (JSHL) placed in the city, Hisar is also having 2nd largest cropped area (645,000 hectares) in Haryana state [7], which itself reflects the high use of agricultural vehicles such as tractor, field cultivators, combine harvesters and many others including animal drawn vehicles like bullock carts. As the city is gifted with historical monuments and places combined with industrialization and agricultural activities, Hisar attracts a large traffic volume due to which the road accidents are increasing with the economic development of the city. Hisar has a total road length of 2380 km (metalled road length) highest in the state Haryana and accounting for 8.8% of total road length of the state itself. In , Hisar was the second-largest city in terms of population after Faridabad. The study will help in developing a proper model and predictions of road accidents in the area selected. There can be the upgradation in the traffic pattern of city to prevent the road accidents and provide good traffic movements. In Hisar city accidents per lakh population 33.34 can be prevented and the overall accident phenomenon can be improved from the present study. There can be a great impact of the study in society for reduction of the accidents and the death ratio of selected city [8–10].
2. LITERATURE REVIEW

The statistics from 20 countries in 1938 on road casualties, automobiles and population. The research discovered that ten of the death values predicted by this technique were within 15% of their true values, 19 were within 40%, and one value was in error up to 67% [11–13]. Analysis for its lack of model reliability. He claimed that the Smeed formula could not be applied to all countries. The inability of Smeed's model to forecast deaths in many industrialised and developing nations prompted the authors to refine it by including other variables [14–16].

Traffic crash models for India's major megacities. The study's main finding was that, in order to reduce accidents, key policies and rules must be changed to reduce the rise of individual vehicle population and urge people to use public transport vehicles. To estimate road fatalities in Ghana, Ofosu and Lamptey directly applied Smeed's law of deaths per population to data. The amended Smeed law has a 7.8% average error, while the Original Smeed law has a 17 percent average error. In the United States, Zlatoper studied at the problems observed of motor vehicle deaths. "Income, the ratio of urban to rural driving, spending on highway police and safety, vehicle inspection, and adult seat belt use regulations with secondary enforcement policies are all inversely correlated to motor vehicle death rates," he said. Koren and Boros made a comparative investigation of whether Smeed's law is still applicable in today's traffic problem. So according Smeed's Law, even if the number of fatalities increases as the number of vehicles increases, the number of fatalities per vehicle decreases as the number of vehicles increases. This pattern is still being followed, according to their 2007 research of 139 nations. According to Smeed's law, fatalities per 10,000 population should grow as ownership rates increase, however the real data fits better with a curve that increases initially and then reduces [17,18], see Table 1 and figure 1.

| Model Name | Equation | Remarks |
|-------------|----------|---------|
| Smeed’s fatality prediction model (1949) | \( C/R = 0.0003 \left( \frac{R}{P} \right)^{-0.67} \) | C= Total deaths, R= No. of registered vehicles, P= population |
| Andreassen accident, fatality and injury prediction Model (1991) | \( A = e^{11.26 \times R^{0.495} \times P^{-0.83}} \) \( C = e^{13.47 \times R^{0.613} \times P^{-0.63}} \) \( I = e^{13.47 \times R^{0.604} \times P^{-0.54}} \) | C= Total deaths, R= No. of registered vehicles, P= population |
| P. Valli’s accident, fatality and injury prediction Model (2004) | \( A/R = 0.0008(R/P)^{0.75} \) \( C/R = 0.0003(R/P)^{0.58} \) \( I/R = 0.0014(R/P)^{0.57} \) | A= No. of accidents, I= No. of Injuries |
3. METHODOLOGY AND DATA COLLECTION

![Flowchart for process of study](image)

**Figure 1**: Flowchart for process of study

After collecting the data for year 2011 to 2019, that is further evaluated on the basis of various characteristics of the accidents. The information collected is then sorted and organized into tables for analysis. Various analysis and comparisons are conducted to show the change in accident rates over time and to track the pattern. For Hisar and Haryana accident severity index, comparative analysis and modeling of data using linear regression analysis is carried out to predict accidents and fatalities [19–21].

4. ACCIDENT SEVERITY INDEX (ASI)
Accident severity is defined as the number of people died per 100 accidents.
Figure 2: A.S.I of different cities of Haryana in 2020

It can be clearly seen from figure-2 that Hisar's ASI is lower than Haryana's because it is ranked 19th in state Haryana, cities like Jhajjar (62.25) and Yamunanagar (56.05) must follow policies used by administration of Hisar because the A.S.I in these cities is approaching dangerous levels, forcing Haryana to rank among the top A.S.I causing states, see table 2 and 3.

Figure 3: ASI comparison from year 2011-19
Figure-3 demonstrates that Haryana's ASI is quickly increasing; according to government records, first-aid is provided to 0.648 person per accident in 2019 and 2020, compared to 0.7614 person per accident in 2013 and 2014, when the A.S.I was at its lowest of 41.99. Hisar's A.S.I is lower than the state average because is state Haryana, Hisar ranks 5th and 8th in terms of accidents and deaths, respectively, while 18 cities have higher value of A.S.I when compared to Hisar in 2020.

5. COMPARATIVE ANALYSIS
After evaluating all the data of Hisar, Haryana and India graphs are plotted between Hisar and Haryana to see the trends throughout the analysis period. All of the graphs are shown in per lakh figures to make it easier to understand the values.

**Formula: Accident/Injury/Death * 10000/Registered vehicles/Population/Road length**

6. ANALYSIS ON THE BASIS OF POPULATION

Figure-4 shows that Hisar has less number of accidents, injury and death per one lakh population because Hisar (14.78%) has more population rise in comparison to Haryana (13.89%) Throughout the period of analysis (2011-19) and in addition, Hisar's A.S.I. has diminished, causing the graph to decline.

7. ANALYSIS ON THE BASIS OF NUMBER OF REGISTERED VEHICLES
It can be seen from figure-5 that Haryana (110.01) has the lowest number of accidents per lakh registered vehicles in year 2019, Haryana (46.48 percent) has made more progress in reducing the numbers during the analysis period when compared to Hisar (33.30). The reason for this is that there has been very little increase in the number of accidents, while the number of registered vehicles has increased in the following order: Haryana (101.64%) > Hisar (80.31%).

8. ANALYSIS ON THE BASIS OF ROAD LENGTH

As of 2020, Hisar is the third largest city in the state of Haryana, with the highest road length 2380Km (8.8%) of the state’s road length. Because of the low share of Hisar in accidents (accident 6.01%, injury 6.65% and death 4.68% in 2019) and large share in road length, all the plots are dropping in figure-6, resulting in a lower value of accident, injury, and death per lakh Km road length.
9. MODEL ANALYSIS

The term "linear regression analysis" refers to any approach for constructing a model and assessing various variables with the goal of focusing on the relationship between the dependent and independent variables.

\[ Y = a + b \times X \]

\( Y \) = Dependent variable, \( X \) = Independent variable (Explanatory variable),
\( a \) = \( Y \)- intercept (value of \( Y \) when, \( X = 0 \)), \( b \) = slope of line

Following outcomes are possible in model analysis:-
- Null Hypothesis: We assume that there is no statistical relation between dependent and independent variable.
- Hypothesis (Non-zero correlation): There is a strong correlation between dependent and independent variable.

10. MODEL FOR HARYANA

**Table 2**: Accident prediction model of Haryana

| Regression Statistics |  |
|-----------------------|--|
| Multiple R            | 0.836718 |
| R Square              | 0.700097 |
| Adjusted R Square     | 0.650113 |
| Standard Error        | 259.8639 |

**ANOVA**

| Df | SS      | MS      | F       | Significance F |
|----|---------|---------|---------|----------------|
| Regression | 1 | 945845.5 | 945845.5 | 14.0664 | 0.009594 |
| Residual | 6 | 405175.3 | 67529.22 |         |         |
| Total   | 7 | 1351021 |         |         |         |

**Coefficients**

| Coefficients | Standard Error | t Stat | P-Value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|--------------|----------------|--------|---------|-----------|-----------|-------------|-------------|
| Intercept    | -321.53        | -0.107 | 0.9180  | -7651.23  | 7008.16   | -7651.23    | 7008.16     |
| X Variable 1 | 0.000414       | 0.000111 | 3.742  | 0.000143 | 0.00068  | 0.000143    | 0.00068     |

**Table-3**: Death prediction model of Haryana

| Regression Statistics |  |
|-----------------------|--|
| Multiple R            | 0.939043 |
| R Square              | 0.881801 |
| Adjusted R Square     | 0.862101 |
| Standard Error        | 110.9117 |
The total number of accidents is the ‘Y’ variable, whereas the total population of Haryana is the ‘X’ variable. While developing the model, confidence level is set at 95%, and both significance-F (0.009594 = M₁ and 0.00054 = M₂) and P-value are less than 0.05, indicating that the model is good and matches the hypothesis (non zero correlation). High R² value (0.70=M₁ and 0.88=M₂) of both the models shows their accuracy, see table 2 and 3.

![Graph](image)

**Figure-7**: Validation of models for state Haryana

From figure-7 we can conclude that the models are highly accurate and the error is in range of 0-5% only which indicate that models can be used for prediction of accidents and deaths in Haryana.

### 11. MODEL’S FOR HISAR

**Table-4**: Accident prediction model for Hisar

| Regression Statistics |
|-----------------------|
| Multiple R            | 0.72721 |
| R Square              | 0.528834 |
| Adjusted R Square     | 0.469939 |
| Standard Error        | 26.14016 |

| ANOVA                  |
|------------------------|
| Df | SS  | MS  | F   | Significance F |
|-----|-----|-----|-----|----------------|
| Regressor | 1   | 550633.6 | 550634 | 44.7619 | 0.00054 |
| Residual | 6   | 73808.39 | 12301.4 | 0.00054 |
| Total | 7   | 624442 |          |          |

| Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% |
|--------------|----------------|--------|---------|-----------|-----------|
| Intercept -3719.16 | 1278.495 | 2.9090 | 0.0270 | -6847.52 | -590.795 |
| Variable 1 0.000316 | 4.72E-5 | 6.6904 | 0.0005 | 0.0002 | 0.000431 |

The total number of accidents is the ‘Y’ variable, whereas the total population of Haryana is the ‘X’ variable. While developing the model, confidence level is set at 95%, and both significance-F (0.009594 = M₁ and 0.00054 = M₂) and P-value are less than 0.05, indicating that the model is good and matches the hypothesis (non zero correlation). High R² value (0.70=M₁ and 0.88=M₂) of both the models shows their accuracy, see table 2 and 3.

From figure-7 we can conclude that the models are highly accurate and the error is in range of 0-5% only which indicate that models can be used for prediction of accidents and deaths in Haryana.

### 11. MODEL’S FOR HISAR

**Table-4**: Accident prediction model for Hisar

| Regression Statistics |
|-----------------------|
| Multiple R            | 0.72721 |
| R Square              | 0.528834 |
| Adjusted R Square     | 0.469939 |
| Standard Error        | 26.14016 |

| ANOVA                  |
|------------------------|
| Df | SS  | MS  | F   | Significance F |
|-----|-----|-----|-----|----------------|
| Regressor | 1   | 550633.6 | 550634 | 44.7619 | 0.00054 |
| Residual | 6   | 73808.39 | 12301.4 | 0.00054 |
| Total | 7   | 624442 |          |          |

| Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% |
|--------------|----------------|--------|---------|-----------|-----------|
| Intercept -3719.16 | 1278.495 | 2.9090 | 0.0270 | -6847.52 | -590.795 |
| Variable 1 0.000316 | 4.72E-5 | 6.6904 | 0.0005 | 0.0002 | 0.000431 |
Regression | 1 | 6135.537 | 6135.537 | 8.979168 | 0.017162
Residual | 8 | 5466.463 | 683.307 | |
Total | 9 | 11602 |
| Coeff. | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 52.886 | 187.7705 | 0.28 | 0.785358 | -380.113 | 485.885 | -380.113 | 485.8858 |
| X Variable 1 | 0.0003 | 0.0001 | 2.99 | 0.017162 | 6.93E-05 | 0.00053 | 6.93E-05 | 0.000532 |

Table-5: Death prediction model of Hisar

| Regression Statistics |
|---|
| Multiple R | 0.670076 |
| R Square | 0.449002 |
| Adjusted R Square | 0.380127 |
| Standard Error | 6.744759 |

| ANOVA |
|---|
| Df | SS | MS | F | Significance F |
| Regression | 1 | 296.5658 | 296.5658 | 6.519107 | 0.034002 |
| Residual | 8 | 363.9342 | 45.49178 |
| Total | 9 | 660.5 |

| Coeff. | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|---|
| Intercept | 376.08 | 48.449 | 7.762 | 5.42E-05 | 264.3591 | 487.806 | 264.359 | 487.8067 |
| X Variable 1 | -6.6E-05 | 2.59E-05 | -2.55 | 0.034002 | -0.00013 | -6.4E-06 | -0.00013 | -6.4E-06 |

Significance-F value ($M_1 = 0.017162$ and $M_2 = 0.034002$) of both the model’s is far below 0.05 indicating that the model is good and rejects the null hypothesis. R square value (0.52 = $M_1$ and $M_2 = 0.449$) and the t-stat value (2.996 = $M_1$ and $M_2 = -2.553$), indicating that both variables have a solid association and good accuracy, see table 4 and 5.
From figure-8 we can conclude that the models are highly accurate and the error is in range of 0-5% only which indicate that models can be used for prediction of accident and deaths in Hisar.

12. SUMMARY OF MODEL ANALYSIS

| Model Name                  | R^2   | Intercept | X-variable | Significance-F | Model Equation       | Error in predicted values |
|-----------------------------|-------|-----------|------------|----------------|----------------------|---------------------------|
| Accident model for Haryana  | 0.7   | 2995.489  | 0.00011    | 0.0095         | Y=0.000414X - 321.53 | 2.14%                     |
| Accident model for Hisar    | 0.528 | 187.771   | 0.0001     | 0.0171         | Y=0.0003X + 52.8862  | 3.22%                     |
| Death model for Haryana     | 0.8818| 1278.495  | 4.72E-05   | 0.000541       | Y=0.000316X - 3719.16| 2.29%                     |
| Death model for Hisar       | 0.449 | 48.4491   | 2.60E-05   | 0.034          | Y=-6.6E-05X + 376.083| 2.18%                     |

The current study was carried out in the city Hisar of state Haryana. Over the study period, the number of people killed or injured in road accidents has increased or remained steady in the city. As a result, an attempt has been made to address this issue in the field of accident forecasting. The study focuses on a statistical strategy for predicting accidents based on current circumstances, using death and number of accident as dependent variables.

The study introduced four new models which are Precise and accurate for predicting traffic accidents and deaths, see table 6.

13. Conclusion, Limitation and Future Scope

The present study has come up with following findings:

1. In 2019, accident severity index of Haryana (46.20) is higher than Hisar (36.01).
2. Hisar contributed for 6.01% accidents taking place in the state during 2018-2019.
3. Accidents per lakh population are higher in Haryana (38.40) when compared to Hisar (33.34) in year 2019.
4. Accident density (number of accidents per 10000 km of roads) was highest for Haryana (4270.243) in 2018, which is approximately 1.5 times of Hisar (3036.697).
5. In 2019, Haryana (110.01) have lower value of accidents per lakh registered vehicles when compared to Hisar (121.26).
6. The study's findings were used to create a model for predicting road accidents and death in Hisar and Haryana. Linear regression was used to create four models, two each for Hisar and Haryana. These models provide good results with an error of about 0-5 percent and can thus be used to predict future road accident and deaths.
7. Error in forecasting is in the following manner: Haryana (2.21% error) < Hisar (2.70% error) and death models (66.54%) are more accurate when compared to accident model (61.4%).

13.1. limitation of study

The present study is totally based upon the data available on the website and other offices of government/private sectors. There may be a chance of missing of complete information of minor accidents as well as the less important road accidents. Sometimes various types of accidents are not counted in the government/private sectors, but it exists and have significant effects on the overall accident behavior. Accident model and death model discussed in the study are totally based on available data, but there may be a chance of difference in actual conditions as per site of the study area of paper.

13.2. Future Scope of the Study

The data analyses in the present study in only for Hisar city and Haryana state. The statistics of paper can be used for different district of Haryana as well as other states. The comparison can be done for two or more cities with India for more accurate analysis and results. The analysis used in the present study on the stretch can be done on different stretches and different states, where maximum data is available for accident prediction models. Accident prediction of different cities and roads can be increased by adding more area of cities and states under study.

REFERENCES

[1] Agrawal R, Chhadva Y, Addagarla S and Chaudhari S 2021 Road surface classification and subsequent pothole detection using deep learning 2021 2nd International Conference for Emerging Technology, INCET 2021 (Institute of Electrical and Electronics Engineers Inc.)

[2] Ganji M R, Ghelmani A, Golroo A and Sheikhzadeh H 2021 Asphalt pavement macrotexture monitoring in cracked surfaces by using an acoustical low-cost continuous method Autom. Constr. 132

[3] Andarini D, Camelia A and Ibrahim M M 2021 Factors related to road accidents in palembang, south sumatera, indonesia Int. J. Public Heal. Sci. 10 638–45

[4] M Selveindran S, Samarutilake G D N, Rao K M N, Pattisapu J V, Hill C, Kolas A G, Pathi R, Hutchinson P J A and Vijaya Sekhar M V 2021 An exploratory qualitative study of the prevention of road traffic collisions and neurotrauma in India: perspectives from key informants in an Indian industrial city (Visakhapatnam) BMC Public Health 21

[5] Coronell G, Arellana J and Cantillo V 2021 Location of speed control cameras on highways: A geospatial analysis Transport 36 199–212

[6] Jun Y, Park J and Yeom C 2021 The evaluation of experimental variables for sustainable virtual road safety audits Sustain. 13

[7] Lapegna M and Stranieri S 2022 DClu: A Direction-Based Clustering Algorithm for VANETs
Management ed C H Barolli L. Yim K. Lect. Notes Networks Syst. 279 253–62

[8] Scorza F and Fortunato G 2021 Cyclable Cities: Building Feasible Scenario through Urban Space Morphology Assessment J. Urban Plan. Dev. 147

[9] Nazif-Munoz J I, Martinez P, Williams A and Spengler J 2021 The risks of warm nights and wet days in the context of climate change: assessing road safety outcomes in Boston, USA and Santo Domingo, Dominican Republic Inj. Epidemiol. 8

[10] Burghardt T E, Babić D and Pashkevich A 2021 Performance and environmental assessment of prefabricated retroreflective spots for road marking Case Stud. Constr. Mater. 15

[11] Lei X, Chen X and Rhee S H 2021 A hybrid access method for broadcasting of safety messages in IEEE 802.11p VANETs Eurasisp J. Wirel. Commun. Netw. 2021

[12] Ekroll V, Svalebjørg M, Pirrone A, Böhm G, Jentschke S, van Lier R, Wagemans J and Høye A 2021 The illusion of absence: how a common feature of magic shows can explain a class of road accidents Cogn. Res. Princ. Implic. 6

[13] Michalec O, O’Donovan C and Sobhani M 2021 What is robotics made of? The interdisciplinary politics of robotics research Humanit. Soc. Sci. Commun. 8

[14] Feiss R, Hautmann A, Asa N, Hamann C, Peek-Asa C and Yang J 2021 Balancing safety on the road with risk from COVID-19: A content analysis of policy adaptations by Divisions of Motor Vehicles Accid. Anal. Prev. 162

[15] Lioi A, De Santos-Berbel C, Bassani M and Castro M 2021 3D Modeling of Highway Guardrails for Sight Distance Assessment J. Transp. Eng. Part A Syst. 147

[16] Singh R, Sharma R, Vaseem Akram S, Gehlot A, Buddhi D, Malik P K and Arya R 2021 Highway 4.0: Digitalization of highways for vulnerable road safety development with intelligent IoT sensors and machine learning Saf. Sci. 143

[17] Choi D-A and Ewing R 2021 Effect of street network design on traffic congestion and traffic safety J. Transp. Geogr. 96

[18] Orsini F, Gecchele G, Rossi R and Gastaldi M 2021 A conflict-based approach for real-time road safety analysis: Comparative evaluation with crash-based models Accid. Anal. Prev. 161

[19] Khan A M, Jamil M, Mia M, Pimenov D Y, Gasiyarov V R, Gupta M K and He N 2018 Multi-objective optimization for grinding of AISI D2 steel with Al2O3 wheel under MQL Materials (Basel). 11

[20] Singh G, Pruncu C I, Gupta M K, Mia M, Khan A M, Jamil M, Pimenov D Y, Sen B and Sharma V S 2019 Investigations of machining characteristics in the upgraded MQL-assisted turning of pure titanium alloys using evolutionary algorithms Materials (Basel). 12

[21] Abbas A T, Gupta M K, Soliman M S, Mia M, Hegab H, Luqman M and Pimenov D Y 2019 Sustainability assessment associated with surface roughness and power consumption characteristics in nanofluid MQL-assisted turning of AISI 1045 steel Int. J. Adv. Manuf. Technol. 105 1311–27