© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Accident Severity Analysis On the North-South Expressway Using Binomial Logistic Regression

Sudesh Nair Baskara[1*], Haryati Yaacob[2], Sitti Asmah Hassan[2], Mohd Rosli Hainin[2], Mohd Shahrir Amin Ahmad[3]

[1]Faculty of Engineering and Quantity Surveying,
INTI International University, 71800 Nilai, Negeri Sembilan, MALAYSIA

[2]School of Civil Engineering, Faculty of Engineering,
Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, MALAYSIA

[3]Southern Region, Malaysian Highway Authority, 81400 Senai, Johor, MALAYSIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2023.15.06.028
Received 16 July 2023; Accepted 30 September 2023; Available online 28 December 2023

Abstract: Accidents on Malaysian expressways must be monitored on a regular basis since severe accidents occur on expressways due to higher posted speeds than on other roadways. This study aims to analyse accident severities based on pavement conditions using binomial logistic regression. Data on accident severities and pavement conditions were gathered from the Malaysian Highway Authority (MHA) and the The Royal Malaysian Police (PDRM) respectively. Three binomial logistic regression models were constructed based on four accident severity categories of death, serious injury, minor injury and damage. The accident severity was grouped into different classifications in which Model 1, Model 2 and Model 3 were developed. To assess the model's predictive capabilities, predicted accident severity levels were distinguished with actual accident severity levels. Based on the results, Model 1 and Model 2 except Model 3 have significant pavement conditions and are viable for accident severity predictions whereby both models exhibited high prediction accuracy, has good fit and good in differentiating between two classifications. The model's classifier is better at classifying accident severity classes with more samples than categorizing accident severity classes with fewer data. The odds ratio of both Model 1 and Model 2 revealed that International Roughness Index (IRI) has greater influence in predicting accident severity compared to rut depth (RD) and mean texture depth (MTD) particularly on predicting death. This study suggested that a comparable study be conducted on other road classifications.

Keywords: Accident severity models, pavement conditions, binomial logistic regression

1. Introduction

Road safety is a concern, particularly with the increasing number of traffic accidents. It is critical to examine accidents involving various accident severity levels to reduce accidents that contribute to severe accidents, particularly deaths. This paper aims to investigate pavement conditions that influence accident severities on North-South expressway. Expressway was chosen for this study because severe accidents mostly occurred on expressways, and the posted speed on expressways is higher than on federal and state roads. Expressway has lower accident rate compared to federal and state roads, yet it is frequently ranked first among all road types in terms of deaths [1]. According to Darma [2], the rate of deaths per kilometer on expressways in Malaysia is 0.404, which is the highest compared to federal and state roads. This study focused on functional pavement condition which considers the surface of the road that was obtained through

*Corresponding author: sudesh.baskara@newinti.edu.my

the Multi Laser Profiler. The Multi Laser Profiler includes the International Roughness Index (IRI), rut depth (RD) and mean texture depth (MTD) data. Several researchers have linked accidents with pavement conditions [3]-[12]. The performance of accident severity model shown to have good accuracies and acceptable area under curve of receive operating curve [13]-[15]. Since this study examines two categories of dependent variables, the accident severities were grouped into two categories, thus binomial logistic regression was used. Binomial logistic regression estimates parameters by maximum likelihood estimation (MLE), which maximizes the probability of obtaining the observed values in a data set [16], classifies models and uses odds ratio to interpret the coefficients analyses probabilities [17] and avoids conflicting effects when associating variables altogether [18]. The applicability of binomial logistic regression in accident severity analysis was used and verified by a few researchers [19]-[21].

2. Methodology

Accident data were collected from PDRM and the pavement condition data were provided by the Malaysia Highway Authority (MHA). The International Roughness Index (IRI), rut depth (RD), and mean texture depth (MTD) employed in this study covered the general surface features of a road. Accident data and pavement condition data were linked based on the location chainage. The data was collected over a two-year period, totaling 1,789 data points, and was analysed using the R statistical programme. Accident severity being the dependent variable and the continuous values of IRI, RD and MTD were referred as the independent variables for model development. Accident severity data were categorized with binary code of 0 and 1 for fatalities, injuries and damages attributes as shown in Table 1. The linked data is split into 1431 training data and 358 validation data. Three accident severity models were constructed with regard to the pavement conditions. Model 1 investigate probability of death against serious injury, minor injury and damage. Model 2 assess probability of death and serious injuries against minor injury and damage. Model 3 evaluates the probability of death, serious injury and minor injury against damage. Accident severity classifications, classification matrices, the Hosmer-Lemeshow test, and the area under the curve (AUC) of Receiving Operating Characteristics (ROC) were used to evaluate the performance of the accident severity model. The influence of a unit increase in pavement condition on accident severities was obtained using the odds ratio in the binomial logistic regression.

Table 1 - Accident severity categorization

Attributes	Data Type	Category	Description
Fatalities	Binary	0	Serious Injury, Minor Injury and Damage
		1	Death
Injuries	Binary	0	Minor Injury and Damage
		1	Death and Serious Injury
Damages	Binary	0	Damage
		1	Death, Serious Injury and Minor Injury

3. Data Analysis

3.1 Outlier Analysis

Fig. 1 to Fig. 3 shows the standardized residuals plots for the corresponding accident severities according to case numbers. The standardized residuals for all data were within the limit between -3.0 and 3.0 implying that there was no outlier. Therefore, no data needed to be removed for the analysis.

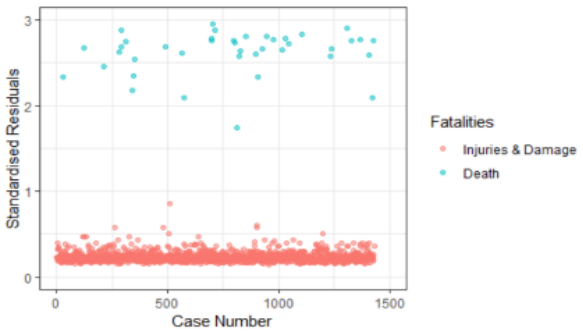

Fig. 1 - Residual plot Model 1

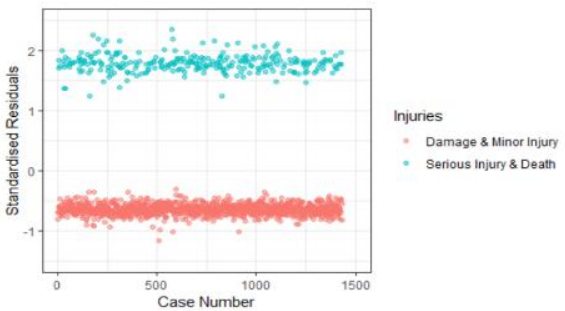

Fig. 2 - Residual plot Model 2

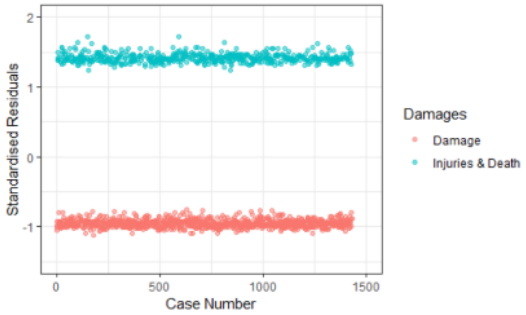

Fig. 3 - Residual plot Model 3

3.2 Assumption of Linearity

Based on the graphs shown in Fig. 4 to Fig. 12, a linear relationship exists between the log odds of accident severities in relation to the pavement conditions except for Fig. 6. The MTD values were transformed to a higher order polynomial with MTD power of four which is labelled as MTD2.

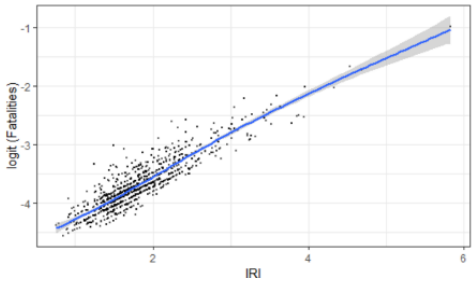

Fig. 4 - Linearity of IRI for Model 1

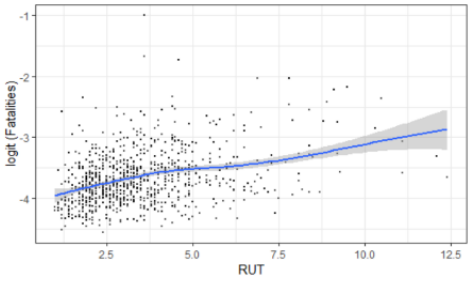

Fig. 5 - Linearity of RD for Model 1

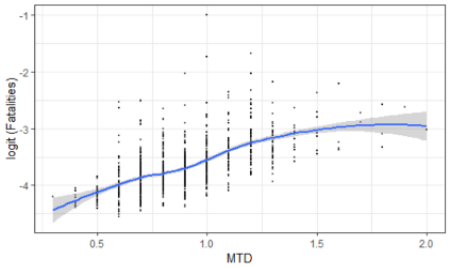

Fig. 6 - Linearity of MTD for Model 1

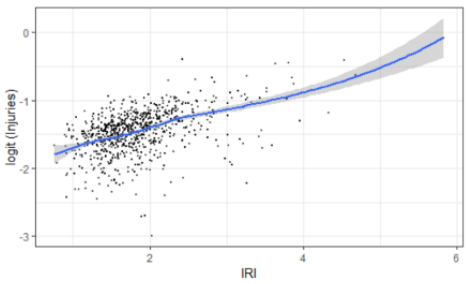

Fig. 7 - Linearity of IRI for Model 2

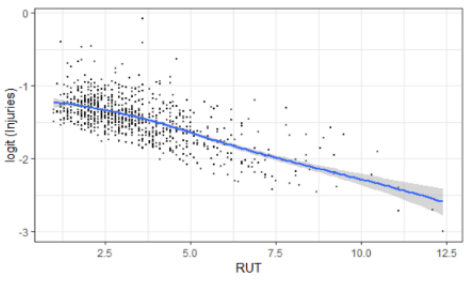

Fig. 8 - Linearity of RD for Model 2

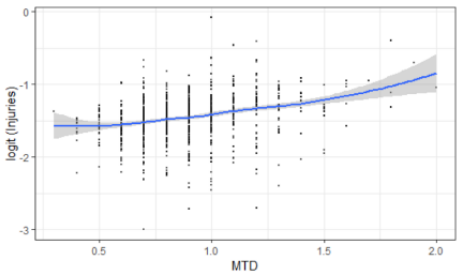

Fig. 9 - Linearity of MTD for Model 2

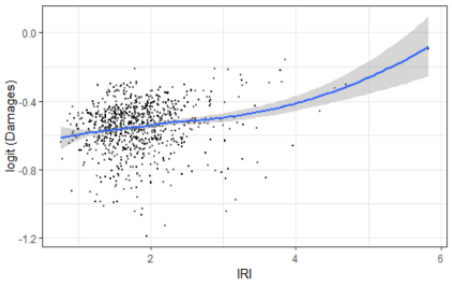

Fig. 10 - Linearity of IRI for Model 3

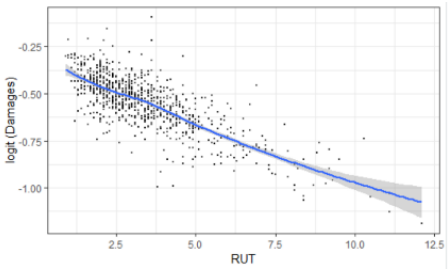

Fig. 11 - Linearity of RD for Model 3

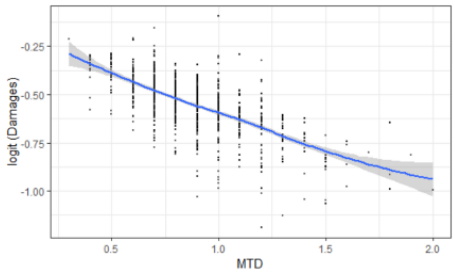

Fig. 12 - Linearity of MTD for Model 3

3.3 Multicollinearity Test

The variable inflation factor (VIF) results shown in Table 2 to Table 4 was below the value of 10 indicating that there was no multicollinearity among the pavement condition variables.

Table 2 - Multicollinearity test results for Model 1

Independent Variable	VIF
IRI	1.10
RD	1.11
MTD2	1.02

Table 3 - Multicollinearity test results for Model 2

Independent Variable	VIF
IRI	1.09
RD	1.13
MTD2	1.10

Table 4 - Multicollinearity test results for Model 3

Independent Variable	VIF
IRI	1.08
RD	1.13
MTD2	1.11

3.4 Binomial Logistic Regression Model

Table 5 and Table 6 show the binomial logistic regression analysis for Model 1 and Model 2 respectively. The IRI and MTD are significant variables for Model 1 while IRI and RD are significant variables for Model 2. There is no any significant variables noted for Model 3, therefore Model 3 was not used for further analysis.

Table 5 - Logistic regression results for Model 1

Model 1	Estimate	Standard Error	z value	p-value
Intercept	-5.088	0.476	-10.695	0.000
IRI	0.679	0.204	3.322	0.000
RD	0.010	0.087	0.116	0.908
MTD	0.152	0.065	2.337	0.019

Table 6 - Logistic regression results for Model 2

Model 2	Estimate	Standard Error	z value	p-value
Intercept	-1.614	0.240	-6.736	0.000
IRI	0.366	0.115	3.175	0.001
RD	-0.148	0.046	-3.313	0.000
MTD	0.532	0.287	1.853	0.064

3.5 Classification Table

The actual accident severities and the predicted accident severities were classified as shown in Table 7 for Model 1 and Table 8 for Model 2. Referring to Table 7, the model correctly predicted the classifications of serious injury, minor injury and damage with 1391 cases for Model 1. However, the model did not predict death correctly with 40 deaths predicted at the classification of serious injury, minor injury and damage. In Table 8, model correctly predicted 1153 accident cases into the minor injury and damage classifications but the model failed to predict the death and serious injury with 278 deaths were predicted at the classifications of minor injury and damage.

Table 7 - Classification table of training data for Model 1

Model 1		Actual Accident Severity	
	Classification	Serious Injury, Minor Injury, Damage	Death
Predicted Accident Severity	Serious Injury, Minor Injury, Damage	1391	40
	Death	0	0

Table 8 - Classification table of training data for Model 2

Model 2		Actual Accident Severity	
	Classification	Minor Injury, Damage	Death, Serious Injury
Predicted Accident Severity	Minor Injury, Damage	1153	278
	Death, Serious Injury	0	0

Classification matrices for Model 1 and Model 2 are shown in Table 9 and Table 10. Based on the classification in Table 7 for Model 1, since the model correctly predicted the classifications of serious injury, minor injury and damage with 1391 accident cases out of total 1431, therefore the accuracy is 97.20% which the misclassification is 2.8% as shown in Table 9. The accuracy for Model 2 accounted for 80.57% with 1153 accident cases out of total 1431, therefore the misclassification error is 19.43%.

Table 9 - Classification matrices of training data for Model 1

Accuracy	Misclassification Error	Sensitivity (TPR)	FPR	Specificity (1-FPR)
97.20%	2.8%	0%	0%	100%

Table 10 - Classification matrices of training data for Model 2

Accuracy	Misclassification Error	Sensitivity (TPR)	FPR	Specificity (1-FPR)
80.57%	19.43%	0%	0%	100%

3.6 Goodness of Fit Test

The Hosmer-Lemeshow test in Table 11 produces p-value of 0.793 for Model 1 and 0.174 for Model 2. Both p-values are above 0.05 indicating both models have good model fit. There is no difference between the actual accident severity and predicted accident severity.

Table 11 - Hosmer-Lemeshow test results

Model	Chi-Square	p-value	Outcome
Model 1	4.667	0.793	Good model fit
Model 2	11.514	0.174	Good model fit

3.7 Receiving Operating Characteristic (ROC)

Fig. 13 shows Model 1 has area under the curve of 0.625 (62.5%) while Fig. 14 shows Model 2 with area under the curve of 0.582 (58.2%) in differentiating between accident severity classifications.

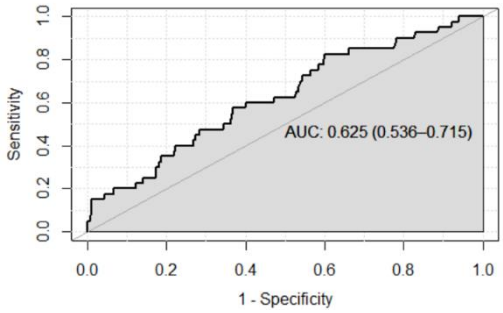

Fig. 13 - AUC for Model 1

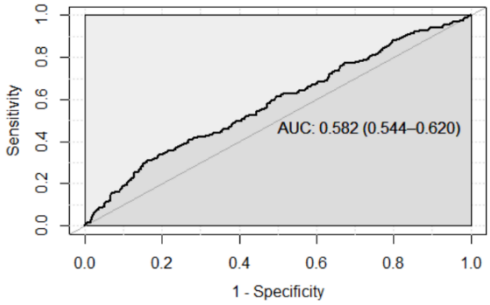

Fig. 14 - AUC for Model 2

3.8 Model Validation

Table 12 and Table 13 classified the actual accident severities and predicted accident severities using validation data. Referring to Table 12, the model correctly predicted the classifications of serious injury, minor injury and damage with 344 cases for Model 1 but did not predict death correctly with 14 deaths predicted at the classification of serious injury, minor injury and damage. In Table 13, model correctly predicted 293 accident cases into the minor injury and damage classifications, but the model failed to predict the death and serious injury with 65 death cases were predicted at the classifications of minor injury and damage.

Table 12 - Classification table of validation data for Model 1

Model 1		Actual Accident Severity	
	Classification	Serious Injury, Minor Injury, Damage	Death
Predicted Accident Severity	Serious Injury, Minor Injury, Damage	344	14
	Death	0	0

Table 13 - Classification table of validation data for Model 2

Model 2		Actual Accident Severity	
	Classification	Minor Injury, Damage	Death, Serious Injury
Predicted Accident Severity	Minor Injury, Damage	293	65
	Death, Serious Injury	0	0

Classification matrices for Model 1 and Model 2 based on validation data are shown in Table 14 and Table 15. Based on the classification in Table 12 for Model 1, since the model correctly predicted the classifications of serious injury, minor injury and damage with 344 accident cases out of total 358, therefore the accuracy is 96.09% which the

misclassification is 3.91% as shown in Table 14. The accuracy for Model 2 accounted for 81.84% with 293 accident cases out of total 358, therefore the misclassification error is 18.16%.

Table 14 - Classification matrices of validation data for Model 1

Accuracy	Misclassification Error	Sensitivity (TPR)	FPR	Specificity (1-FPR)
96.09%	3.91%	0%	0%	100%

Table 15 - Classification matrices of validation data for Model 2

Accuracy	Misclassification Error	Sensitivity (TPR)	FPR	Specificity (1-FPR)
81.84%	18.16%	0%	0%	100%

3.9 Comparison Between Training Model and Validation Model

A comparison of the training and validation models for Model 1 and Model 2 revealed that the results were comparable in accuracy and misclassification error. This demonstrated that the training model had been validated and that the accident severity classification was correctly predicted.

3.10 Odds Ratio and Probability of Accident Severities

Table 16 depicted the results of odds ratio and probability of Model 1. The odds of resulting in death is greater than the odds of major injury, minor injury, and damage by 1.972 for IRI. A unit increase in IRI increases the probability of death by 97.2% compared to major injury, minor injury, and damage. Meanwhile, the MTD coefficient suggested that the odds of death are greater than the odds of major injury, minor injury, and damages by 1.164 with a unit increase in MTD increases the probability of death by 16.4% in comparison to major injury, minor injury, and damage.

Table 16 - Odds ratio for Model 1

Independent Variables	Odds Ratio
IRI	1.972 (97.2%)
MTD	1.164 (16.4%)

The odds ratio for IRI and RD in Model 2 are shown in Table 17. The findings showed that IRI had an odds ratio of 1.442 and RD had an odds ratio of 0.863. The odds of death and serious injury over the odds of minor injury and damages was 1.442 for IRI. For every unit increase in IRI value, the probability of death and serious injury increased by 44.2% compared to the probability of minor injury and damage. Meanwhile, for RD, the odds ratio is 0.863 with each unit increase in RD value reduces the chances of death and serious injury by 0.137 (13.7%) but increased the probability of damage and minor injury by 15.9% (1/0.863 = 1.159).

Table 17 - Odds ratio for Model 2

Independent Variables	Odds Ratio
IRI	1.442 (44.2%)
RD	0.863 (-13.7%)

4. Conclusion

This study has met the outliers, linearity and multicollinearity outcomes in order to produce logistic regression models. Model 1 and Model 2 have proven to show a relationship between accident severities and pavement conditions while Model 3 failed to show any significant relationship. Both Model 1 and Model 2 were found to be high in predicting accuracy. However, both models did not predict well on the severe accident severity involving death. The model's inability to forecast death is primarily due to an imbalanced data set. The data acquired revealed that there were more incidents involving accidents with damage compared to minor injuries, major injuries, and death. This makes the classifier less inclined towards death due to the smaller number of accident cases. However, both models had an AUC greater than 50%, showing that the models could distinguish between accident severity classifications [22]. The validation model results were relatively similar to the training model results. Based on the results of the analyses, the models were appropriate for forecasting accident severity probabilities without issues on overfitting or underfitting. The odds ratio results for Model 1 indicated that the IRI had a greater impact than MTD in predicting the probabilities of accident severities while the impact of IRI was higher than RD in predicting accident severity probabilities for Model 2. Since

there is a relationship between IRI, RD, MTD with accident severities, this study recommends similar research to be conducted on federal roads, state roads and rural roads so that a comparison can be made.

Acknowledgement

The authors would like to acknowledge and appreciate the PDRM and Malaysia Highway Authority for providing useful information for this study.

References

[1] Ye F., Cheng W., Wang C., Liu H. & Bai J. (2021). Investigating the severity of expressway crash based on the random parameter logit model accounting for unobserved heterogeneity. Advances in Mechanical Engineering, doi:10.1177/16878140211067278.
[2] Darma Y., Karim M. R. & Abdullah S. (2017). An analysis of Malaysia road traffic death distribution by road environment. Sādhanā, 42(9), 1605-1615.
[3] Li Y., Liu C. & Ding L. (2013). Impact of pavement conditions on crash severity. Accident Analysis and Prevention, 59, 399-406.
[4] Lee J., Nam B. & Abdel-Aty M. (2015). Effects of pavement surface conditions on traffic crash severity. Journal of Transportation Engineering, doi: https://doi.org/10.1061/(ASCE)TE.1943-5436.0000785.
[5] Chan C. Y., Huang B., Yan X. & Richards S. (2010). Investigating effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS). Journal of Advanced Transportation, 44(3), 150-161.
[6] Chan C. Y., Huang B., Yan X. & Richards S. (2009). Relationship between highway pavement condition, crash frequency, and crash type. Journal of Transportation Safety and Security, 1(4), 268-281.
[7] Buddhavarapu P., Banerjee A. & Prozzi J. A. (2013). Influence of pavement condition on horizontal curve safety. Accident Analysis and Prevention, 52, 9-18.
[8] Jiang X., Huang B., Yan X., Zaretzki R. L. & Richards S. (2013). Two-vehicle injury severity models based on integration of pavement management and traffic engineering factors. Traffic Injury Prevention, 14(5), 544-553.
[9] Hussein N. & Hassan R. (2017). Surface condition and safety at signalised intersections. International Journal of Pavement Engineering, 18(11), 1016-1026.
[10] Alhasan A., Nlenanya I., Smadi O. & MacKenzie C. A. (2018). Impact of pavement surface condition on roadway departure crash risk in Iowa. Infrastructures, 3(2), 14.
[11] Kassu A. & Anderson M. (2018). Analysis of severe and non-severe traffic crashes on wet and dry highways. Transportation Research Interdisciplinary Perspectives, 2, 100043.
[12] Mamlouk M., Vinayakamurthy M., Underwood B. S. & Kaloush K. E. (2018). Effects of the international roughness index and rut depth on crash rates. Transportation Research Record, 2672(40), 418-429.
[13] Abdelwahab H. T. & Abdel-Aty M. A. (2001). Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transportation Research Record, 1746(1), 6-13.
[14] Ratanavaraha V. & Suangka S. (2014). Impacts of accident severity factors and loss values of crashes on expressways in Thailand. IATSS Research, 37(2), 130-136.
[15] De Oña J., Mujalli R. O. & Calvo F. J. (2011). Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accident Analysis and Prevention, 43(1), 402-411.
[16] Moomen M., Rezapour M. & Ksaibati K. (2018). An investigation of influential factors of downgrade truck crashes: A logistic regression approach. Journal of Traffic and Transportation Engineering, 6(2), 185-195.
[17] Agresti A. (2007). An Introduction to Categorical Data Analysis. John Wiley & Sons, pp. 1-372.
[18] Sperandei S. (2014). Understanding logistic regression analysis. Biochemia Medica, 24(1), 12-18.
[19] Xi J. F., Liu H. Z., Cheng W., Zhao Z. H. & Ding T. Q. (2014). The model of severity prediction of traffic crash on the curve. Mathematical Problems in Engineering, 1, 1-5.
[20] Al-Taweel H. M. H., Young W. & Sobhani A. (2016). A binary logistic regression model of the driver avoidance manoeuvers in two passenger vehicle crashes. Australasian Transport Research Forum (ATRF), 38, 1-9.
[21] Hsu Y. T., Chang S. C. & Hsu T. H. (2020). Analysis of traffic accident severity at intersection using logistic regression model. Journal of Engineering Research and Reports, 13(4), 1-9.
[22] Hosmer Jr D. W., Lemeshow S. & Sturdivant R. X. (2013). Applied Logistic Regression. John Wiley & Sons, pp. 398.