IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)
e-ISSN: 2278-1684,p-ISSN: 2320-334X, Volume 15, Issue 4 Ver. II (Jul. - Aug. 2018), PP 56-62
www.iosrjournals.org

A Comparative Evaluation of the Effects of Categorical Factors on the Safety of Multilane Interstate Highways

Aschalew Kassu[1], Mahbub Hasan[1], Redahegn Sileshi[2]
[1]*Department of Mechanical & Civil Engineering and Construction Management*
Alabama A&M University, Normal, AL 35762
[2]*Department of Physics, University of North Georgia, Gainesville, GA 30566*
Corresponding Author: Aschalew Kassu

Abstract:*This article reports the results of comparative analysis of the significance of categorical factors with two to five levels influencing safety on interstate highways. The study used the Highway Safety Information System (HSIS) accident data observed on interstate highways in the State of Ohio for the five years period ranging from 2010-2014. The analysis considered about thirteen potential variablesincluding roadway design elements, environmental conditions, drivers' demographic factors, time of the day, and day of the week. The results suggest that among the variables considered, only about half of the factors appeared to have significant differences in crash rates across their respective categories. The identified explanatory variables are mainly associated with highway design elements such as median type, number of lanes, access control and contour of the roadway; and the drivers' age and gender.*

Keywords: *Highway safety, Interstate highway,transport safety, traffic accidents,traffic safety, categorical variables.*

Date of Submission: 09-07-2018 Date of acceptance: 23-07-2018

I. Introduction

In the United States, the total motor vehicle-related fatalities reported in the three years period from 2014-2016 were 35,398; 37,757 and 40,200 with an increase of about seven and six percent respectively (NSC, 2017).In the State of Ohio, the reported fatalities in the three years period were 1,011; 1,105 and 1,129 respectively with about 12 percent increase from 2014 to 2016. The increase in fatalities from 2015 to 2016 was 2 percent, which is relatively very low as compared with a 34 percent increase reported in the State of New Mexico. Based on the same report (NSC, 2017), while detail breakdown of the estimates were not provided, the estimated gross economic cost of the traffic fatalities, injuries and property damages in the year 2016 were estimated to be more than $432 billion, which again shows a 12 percent increase from the year 2015.In the NHTSA Technical Reports,Blincoe et al. (2002 and 2015), provided a detailed breakdown of the economic costs, property damage, fatalities, lost productivity, medical, emergency services and insurance costs, congestion and lost time, and other associated costs. Based on the reports, in the year 2000 and 2010, the economic costs of the traffic crashes were estimated to be about $230.6 billion (Blincoe et al., 2002), and $242 billion (Blincoe et al., 2015) respectively.

One of the major elements of highway design, planning and research is the safety of the road users and pedestrians(Vogt and Bared, 1998). Noland and Oh (2004)used HSIS accident data for the State of Illinois and the negative binomial regression tool to study the effect of roadway engineering elements and demographic factors on injuries and fatalities. The dependent variable considered were total crashes and total fatalities. Their findings reported that most of the highway geometric variables were not significantly associated with the crashes; whereas the segments with larger outside shoulders were reported to have lower crash records.Both traffic accidents and fatalities increase with an increase in the number and widths of the segment lanes. This part of their findings complies with the earlier report by Milton and Mannering (1998), which used two years crash data (1992-1993) in State of Washington and negative binomial model, and reported an increase in traffic accidents with increase in the number of lanes of the segment and a decrease in crash frequency for roadway lane width less than 3.5 meter (11.5 ft). The study also reported that shoulder widths less than 1.5 meters were associated with an increase in crash frequency. Regarding the effect of traffic volume on the segments, the study reported that a higher Annual Average Daily Traffic (AADT) during the normal traffic flow condition on the principal arterials studied is associated with an increase in crash frequency. However, during peak hour traffic, an increase in AADT results in the opposite effect. To study the effect of traffic congestion on the probability of occurrences of traffic crashes on Interstate highway, Zhou and Sisiopiku (1997) examined two years data collected on 16 miles segment of Interstate highway I-94 in Detroit, Michigan. The study concluded that both injury and fatal crash frequencies decrease with an increase in volume to capacity ratio.

Using negative binomial regression model on HSIS data for two-lane rural highways in the states of Minnesota and Washington, Vogt and Bared (1998) found that reduction in horizontal curvature,andincrease in the segment lane widths and shoulder widthsof the roadway studied both in Minnesota and Washington, were associated with reduction in accident frequencies. Abdulhafedh (2016) also used HSIS crash data for Interstate I-90 within the State of Minnesota for the five years period ranging from 2008 to 2012 and compared the performance of Poisson, negative binomial and Artificial Neural Network (ANN) models. For the selected roadway section, the study suggested that the model developed using the ANN approach outperforms the widely used negative binomial regression model.Anderson and Dong (2017) used HSIS heavy vehicle crash data for the state of Minnesota for the eleven years period ranging from 2004 to 2014. The focus of the study was to analyze the contributing factors to weekend versus week-day injury severity of heavy truck drivers in Minnesota. The results indicated that the contributing factors for weekday crashes were different from weekend crashes, and identified 24 and 17 key contributing factors respectively.

From the above studies, we can learn that there are no such a single and universal model and explanatory factors which influence traffic accidents for all roadway geometric condition, time of the day, environmental condition and the driver demographic factor, physical and behavioral condition applicable to all geographic locations. The likelihood of the factors to be significant or not vary depending on the jurisdiction, combination of engineering, environmental effects and human factors such as the drivers' behavior and condition, use of alcohol and drug. Depending on the location, time period, and the interactions of these and other potential factors, may or may not have adverse effects on the likelihood of traffic crash warranting individual and a case by case analysis of crashes data.

II. Methodology

Based on the HSIS Guidebook for State Data (Nujjetty et al., 2015), in the year 2011, accident information collected in the State of Ohio covers nearly 19,500 miles of all functional classification of roadways, consisting of 14,000 miles of State routes, 4,000 miles of U.S. routes, and 1,500 miles of Interstate highways. The data provided by the HSIS is believed to be fairly accurate (Noland and Oh, 2004).The Ohio crash data received from HSIS consists of five separate files. These data include accident, vehicle, occupant, grade, and roadway databy the crash year. The accident data sub-file includes crash location (milepost), roadway classification, day and time of the crash, weather and lighting condition, road surface condition (dry, wet, and snow), the accident type (head on, rear end, etc.,), the crash severity which is classified into five (fatal injury,A injury or incapacitating, B injury or non-incapacitating, C injury or possible injury, and property damage only or no injury), and accident case number. The first three subfiles (accident, vehicle, and occupant) are mainly about accident data. The curve and grade files include mainly the directions of the curves and grades respectively. The occupant file includes the drivers' age, gender, seating position and the occupant injury. The roadway file consists of the classification of the pavement surface of the segment, the width of the pavement, access controlled or not, AADT, median type, pavement width, and number of lanes. The vehicle data file contains vehicle damage scale and contributing factors of the vehicle. To form a single dataset containing the required variables, the data within the separate sub files were linked and merged in accordance with the guideline provided by HSIS Guidebook (Nujjetty et al., 2015) using their respective common variables including accident case number, and vehicle number. Once the variables were properly merged, the crash data corresponding to four and six-lane urban and rural Interstate highwayswerefiltered, and data points with unknown values were also removed.

Table 1 shows the list of the explanatory variables and the corresponding categories evaluatedin the study. Theseinclude highway design elements (number of lanes, roadway contour, median, access control), environmental conditions (weather, lighting, and pavement surface), drivers' demographics (age and gender), time of the day, the day of the week, and accident years. Initially, the dependent variable, crash rate, was categorized under five severity levels (fatal crashes, A-injury (incapacitating injury), B-injury (non-incapacitating injury), C-injury (possible injury), and property damage only accidents. Consistent with Cerwick et al. (2014), in this article, except the property damage only accidents, the other four accident severity levels are reduced into two major injury categories called fatal and injury crashes (Kassu and Anderson, 2018).Thisisaccomplished by merging the fatal and incapacitating injury, and non-incapacitating injury and possible injury together to form fatal and injury crashes respectively. Consistent with earlier works by Oňa et al. (2011), and Yasmin and Eluru (2013), the drivers are classified into three groups as younger drivers (16 to 24), middle-aged drivers (25 to 64), and older drivers (65 and over).

To determine whether the three crash severity levels are statistically significantly different or not across the three crash severity categories, tests of equality of mean crash rateswas performed using Welch's test (Welch, 1951)followed by Games-Howell *post-hoc*pairwise comparison tests. Games-Howell test also provides grouping information for the crash rates across the categories of the factors as well as multiple comparisons of the categories. Homogeneity of the variances and similarity in the distribution of the data were tested using Levene'stest, and non-parametric Kruskal-Wallis independent samples tests (Kruskal and Wallis, 1952)respectively.In general, null hypotheses for the three tests are, (1) the mean crash rates for each categoryof the individual variables are equal, (2) the variances across the categories of the factors is homogenous, and (3) the crash data across the categories of the explanatory factors are drawn from the same distribution.

To control the effect of time series on traffic safety (Noland and Oh, 2004), the five years dataset is categorized with respect to the accident year and used as a variable.Using the individual segment length (SL) in miles, and AADT (Annual Average Daily Traffic), the crash counts observed within each roadway segments are converted to crash rates per MVMT (million vehicles miles of travel) using the Federal Highway Administration's guideline (FHWA, 1990). The AADT data is based on the traffic count information collected by The State of Ohio using the 180 automatic traffic recorders operating 24-hours throughout the year (Nujjetty et al., 2015).

$$MVMT = \frac{(SL)x(AADT)x365}{1,000,000}$$

Table 1. Description of variables used for comparative evaluation of the crash rates across the categories.

Description of Variable	Category and Coding
Accident year	2010, 2011, 2012, 2013, 2014
Day of week	Weekday, Weekend
Hour of day	Morning (6:00 AM -11:59 AM), Afternoon (Noon - 5:59 PM), Evening (6:00 PM - 11:59 PM), Night (12:00 AM - 5:59 AM)
Lighting condition	Light, Dark
Contour of roadway	Straight-Level, Straight-Grade, Curve-Level, Curve-Grade
Weather	Normal, Rain, Snow
Road surface condition	Dry, Wet, Snow
Drivers' age	Younger, Middle Age, Older
Drivers' gender	Female, Male
Access control	No Control, Limited Control, Full Control
Median type	Protected, Unprotected
Number of lanes	Four Lane, Six Lane
Roadway classification	Urban, Rural

III. Results and Discussions

Table 2shows the results of tests of equality of means, homogeneity of variances, and similarity in distribution of the crash data across two or more levels of categorical variables, including accident severity. The HSIS data provided classified the dependent variable based on five crash severity levels. These are fatal crashes, A-injury (incapacitating injury), B-injury (non-incapacitating injury), C-injury (possible injury), and property damage only accidents. To determine whether the three crash severity levels are statistically significantly different or not, across the three crash severity categories, tests of equality of means followed by *post-hoc* tests, homogeneity of variances, and similarity in the distribution of the data were tested. The results of Welch's test (F-value =0.96, p-value = 0.383), and Levene's test of homogeneity of variances (test stat. = 1.06, and p-value = 0.348) suggest that there is no statistically significant difference between the mean values and the variances of fatal, injury and property damage accidents observed between 2010-2014 on Interstate highways in the State of Ohio (Table 2). The results of Games-Howell's post hoc pairwise comparison test also showed that there is no enough evidence to suggest that the three crash severity levels fall under different groups. From Kruskal-Wallis independent samples test (test stat. = 3.07, p-value = 0.215), we can see that, the crash data across the severity levels also follow a similar distribution.

As can be seen in the table, the variations in Interstate crash rate across the five years period ranging from 2010-2014 is also found to be insignificant. Welch's test (F-value= 0.88, p-value = 0.477), and Levene's test (Test stat. = 0.83, p-value = 0.509), suggest that there is no reason to believe that the mean and variances of the Interstate crash rates across the study period is different. Games-Howell's post hoc pairwise comparison test also showed that the five years crash data belong to the same grouping (Table 3(a)) , and pairwise comparison of the mean crash rates also showed that the mean crash rate across individual pairs of the data do not show sizable differences with a level of significance < 0.05 (Table 3(b)).

Drivers'age and gender, the roadway median type, access control, contour of the roadway and number of lanes of the interstate highways resulted in statistically significant difference across their respective categories at a significance level of < 0.05. This implies that there is no enough evidence to suggest that the mean, variances and the distributions of the traffic accidents observed on the Interstate highways across the categories of these factors within the study period is the same. That is, the crash rate on Interstate segments with protected median type is different from the crashes occurred on highway sections with unprotected median type. Similarly, for three categories of a variable such as drivers' age and factor with four categories such as the contour of the roadway; the mean, the variance and the distribution of at least one of the categories under their respective factor is different from the others.Roadway contour is a factor considered as one of the possible factors having an impact on traffic safety. To determine the impacts of the roadway contour, the accidents observed on four different roadway segments categorized as straight-level, straight-grade, curve-level, and curve-grade were tested for equality of means, homogeneity of variances and whether the crash data across these categories were drawn from the same distribution or not. From Table 2, we can see that, Welch's test (F-value = 29.80, p-value = 0.001), Levene's test (Test stat. = 39.33, p-value = 0.001), and Kruskal-Wallis test (Test. stat. = 429.00. p-value = 0.0.001) suggest that we have no evidence to reject the hypothesis that the mean, the variances and the distribution of the accidents observed on one of the roadway contoursis not significantly different from the crashes observed on the other roadway contours. The results of Games-Howell's*post-hoc* test presented in Table 4(a) reveals that the roadway contours characterized as curve-level and curve-grade belong to the same grouping and the straight segments (straight-level and straight-grade) belong to the same group. The pairwise comparison of the mean crash records on curve-level and curve-grade, as well as straight-level and straight-grade roadway segments shown in Table 4 (b), confirms that mean crash rates across these pairs is not significantly different with a level of significance< 0.05.

However, the results for the day of the week (weekday versus weekend), lighting (dark versus light), and roadway classification (urban versus rural) do not show strong evidenceto support the notion that the mean and the variances of the crash rates across their respective two categories aredifferent. That is, in terms of the mean and variances, the crash rates during weekdays are not significantly different from weekends, and the crash rates during dark is not significantly different from crashes observed during day light. The crashes observed on the urban section of the Interstate highways are not significantly different from the crashes occurred on the rural sections of the highways.With regard to the distribution of the crash data, the results of Kruskal-Wallis tests (Table 2) suggest that we do have enough evidence to reject the null hypothesis which states that the distributions of the crash rates across the categories of lighting, and roadway classification are the same. Hours of the day (morning, afternoon, evening, and night), did not also turned out to be significant factors influencing the likelihood of occurrences of traffic accidents. That is, the mean, variances and distribution of the crash data across the four categories of the hours of the day do not appear to be significantly different.Regarding the weather and pavement surface condition, except the distribution of the data which is significant, the mean, and the variances of the traffic accidents observed during normal weather and dry pavement surface are not significantly different from crashes occurred during adverse weather conditions such as rainy and snowy condition. These variables with insignificant differences in crash rates across the categories appear to have no appreciable impact on traffic safety on Interstate highways in the study area within the study period under consideration.

Traffic volume and congestion vary as a function of hours of the day and hence it has been used as one of the variables used to predict traffic accidents (Ivan et al., 2000). To determine if that is true in this particular case as well, the traffic accidents occurred during the four categories of the time of day described in Table 1, were tested for equality of the mean crash rates, homogeneity of variances, and similarity of the distribution of the crashes is tested. Based on the Welch's test (F-value = 1.86, p-value = 0.135), Levene's test (Test Stat. = 2.09, p-value = 0.099), and Kruskal-Wallis test (Test. Stat. = 2.82. p-value = 0.421) shown in Table 2, we have no reason to believe that the mean, the variances and the distribution of the accidents observed during one segment of the hours of the day is different from the others. Games-Howell's post hoc test also suggested that the accident rates observed during the morning, afternoon, evening and night time driving belongs to the same grouping of data (Table 5 (a) and (b)). These results indicate that hours of the day is not a potential predictive factor to estimate the likelihood of occurrences of accidents on interstate highways in the study area. Table 6(a) and (b) summarized the results of Games-Howell's*post-hoc* tests for grouping and pairwise comparison of mean crash rates across the three-level categorical variables.

Table 2. Tests of equality of mean and variances, and similarity in distributions across categories of the variables.

Variable		Welch's Test	Levene's Test	K-W Test	G-H Test
Variables with two categories					
Day of Week	Test Stat.	1.090	0.880	0.210	-1.050
	P-Value	0.296	0.345	0.643	0.296
Lighting	Test Stat.	2.660	2.600	3.870	-1.630
	P-Value	0.103	0.107	0.049	0.103
Drivers' Gender	Test Stat.	9.710	10.350	5.460	-3.120
	P-Value	0.002	0.001	0.019	0.002
Median Type	Test Stat.	557.000	546.000	232.000	23.600
	P-Value	0.001	0.001	0.001	0.001
Number of Lanes	Test Stat.	1226.000	1457.000	3498.000	-39.060
	P-Value	0.001	0.001	0.001	0.001

Roadway Classification	Test Stat.	0.470	3.410	3165.000	1.850
	P-Value	0.429	0.065	0.001	0.065
Variables with three or more categories					
Severity	Test Stat.	0.960	1.060	3.070	
	P-Value	0.383	0.348	0.215	
Accident Year	Test Stat.	0.880	0.830	2.370	
	P-Value	0.477	0.509	0.668	
Hour of Day	Test Stat.	1.860	2.090	2.820	
	P-Value	0.135	0.099	0.421	
Contour of Roadway	Test Stat.	29.800	39.330	429.000	
	P-Value	0.001	0.001	0.001	
weather	Test Stat.	0.350	0.220	12.580	
	P-Value	0.704	0.804	0.002	
Surface Condition	Test Stat.	2.320	1.970	21.520	
	P-Value	0.098	0.140	0.001	
Drivers' Age	Test Stat.	10.230	13.390	6.150	
	P-Value	0.001	0.001	0.046	
Access Control	Test Stat.	75.630	55.310	1724.000	
	P-Value	0.001	0.001	0.001	

Table 3 (a). Games-Howell's*post-hoc* tests for grouping of the categories across the accident years.

Accident Year	N	Mean	St. Dev.	95% CI	G-H Grouping
10	19039	0.205	0.763	(0.19427, 0.21594)	A
11	21370	0.208	0.788	(0.19705, 0.21819)	A
12	20258	0.199	0.708	(0.18943, 0.20892)	A
13	20166	0.196	0.689	(0.18624, 0.20525)	A
14	20956	0.204	0.740	(0.19370, 0.21373)	A

Table 3(b). Games-Howell's pairwise comparison of mean crash rates across the accident years.

Difference of Levels	Diff. of Means	SE of Diff.	95% CI	T-Value	P-Value
11 - 10	0.003	0.008	(-0.01856, 0.02359)	0.33	0.998
12 - 10	-0.006	0.007	(-0.02623, 0.01436)	-0.8	0.931
13 - 10	-0.009	0.007	(-0.02944, 0.01070)	-1.27	0.707
14 - 10	-0.001	0.008	(-0.02194, 0.01915)	-0.19	1.000
12 - 10	-0.008	0.007	(-0.02847, 0.01157)	-1.15	0.778
13 - 11	-0.012	0.007	(-0.03167, 0.00791)	-1.64	0.473
14 - 11	-0.004	0.007	(-0.02418, 0.01636)	-0.53	0.985
13 - 12	-0.003	0.007	(-0.02239, 0.01553)	-0.49	0.988
14 - 12	0.005	0.007	(-0.01492, 0.02400)	0.64	0.969
14 - 13	0.008	0.007	(-0.01125, 0.02720)	1.13	0.790

Table 4(a). Games-Howell's*post-hoc* tests for grouping of the categories across the roadway contour.

Contour of Roadway	N	Mean	St. Dev.	95% CI	G-H Grouping
Curve Grade	8549	0.274	0.895	(0.25475, 0.29270)	A
Curve Level	5423	0.262	1.021	(0.2345, 0.2889)	A
Straight Grade	18249	0.192	0.654	(0.18300, 0.20197)	B
Straight Level	69568	0.191	0.711	(0.18620, 0.19677)	B

Table 4(b). Games-Howell's pairwise comparison of mean crash rates across the roadway contours.

Difference of Levels	Diff. of Means	SE of Diff.	95% CI	T-Value	P-Value
Curve Level - Curve Grade	-0.012	0.017	(-0.0554, 0.0314)	-0.71	0.892
Straight Grade - Curve Grade	-0.081	0.011	(-0.1090, -0.0535)	-7.51	0.001
Straight Level - Curve Grade	-0.082	0.010	(-0.1080, -0.0564)	-8.18	0.001
Straight Grade - Curve Level	-0.069	0.015	(-0.1069, -0.0315)	-4.71	0.001
Straight Level - Curve Level	-0.070	0.014	(-0.1065, -0.0340)	-4.97	0.001
Straight Level - Straight Grade	-0.001	0.006	(-0.01522, 0.01321)	-0.18	0.998

Table 5(a). Games-Howell's*post-hoc* tests for grouping of the categories across the hours of the day.

Hour of Day	N	Mean	St.Dev.	95% CI	G-H Grouping
Afternoon	40691	0.199	0.738	(0.19159, 0.20593)	A
Evening	22439	0.213	0.812	(0.20230, 0.22354)	A
Morning	29627	0.199	0.693	(0.19081, 0.20659)	A

Night	9032	0.204	0.695	(0.18945, 0.21811)	A

Table 5(b). Games-Howell's pairwise comparison of mean crash rates across the hours of the day.

Difference of Levels	Diff. of Means	SE of Diff.	95% CI	T-Value	P-Value
Evening - Afternoon	0.014	0.007	(-0.00263, 0.03093)	2.17	0.133
Morning - Afternoon	0.000	0.005	(-0.01402, 0.01390)	-0.01	1.000
Night - Afternoon	0.005	0.008	(-0.01596, 0.02600)	0.61	0.928
Morning - Evening	-0.014	0.007	(-0.03154, 0.00311)	-2.11	0.151
Night - Evening	-0.009	0.009	(-0.03249, 0.01422)	-1.00	0.747
Night - Morning	0.005	0.008	(-0.01634, 0.02650)	0.61	0.929

Table 6(a). Games-Howell's*post-hoc* tests for grouping of the categories across the three-level categorical variables.

Variable		N	Mean	St.Dev.	95% CI	*G-H Grouping
Weather	Normal	77298	0.201	0.740	(0.19626, 0.20669)	A
	Rain	13926	0.207	0.698	(0.19532, 0.21850)	A
	Snow	10565	0.202	0.779	(0.18749, 0.21720)	A
Surf. Cond.	Dry	68806	0.199	0.733	(0.19347, 0.20442)	A
	Snow	9132	0.206	0.795	(0.19015, 0.22275)	A
	Wet	23851	0.210	0.732	(0.20115, 0.21973)	A
Age	Middle Age	72614	0.196	0.691	(0.19088, 0.20093)	B
	Older	6488	0.196	0.728	(0.17870, 0.21415)	B
	Younger	22687	0.224	0.875	(0.21310, 0.23588)	A
Access	Full Control	96411	0.195	0.732	(0.19043, 0.19967)	C
	Limited Cont.	5331	0.330	0.838	(0.3072, 0.3522)	B
	No Control	47	0.655	0.696	(0.451, 0.860)	A
Severity	Fatal	2873	0.213	0.892	(0.1808, 0.2460)	A
	Injury	20771	0.208	0.772	(0.19715, 0.21814)	A
	PDO	78145	0.200	0.723	(0.19542, 0.20556)	A

*Means that do not share a letter are not significantly different

Table 6(b). Games-Howell's pairwise comparison of mean crash rates across the three-level categorical variables.

Variable	Difference of Levels	Diff. of Means	SE of Diff.	95% CI	T-Value	P-Value
Weather	Rain - Normal	0.005	0.006	(-0.00974, 0.02061)	0.84	0.679
	Snow - Normal	0.001	0.008	(-0.01793, 0.01967)	0.11	0.994
	Snow - Rain	-0.005	0.010	(-0.02706, 0.01793)	-0.47	0.883
Surf. Cond.	Snow - Dry	0.008	0.009	(-0.01302, 0.02804)	0.86	0.668
	Wet - Dry	0.011	0.006	(-0.00139, 0.02437)	2.09	0.092
	Wet - Snow	0.004	0.010	(-0.01842, 0.02639)	0.42	0.909
Age	Older - Middle Age	0.001	0.009	(-0.02148, 0.02252)	0.06	0.998
	Young - Middle Age	0.029	0.006	(0.01372, 0.04345)	4.50	0.001
	Young - Older	0.028	0.011	(0.0029, 0.0532)	2.61	0.025
Access	Limited Cont. - Full Cont.	0.135	0.012	(0.1072, 0.1621)	11.49	0.001
	No Control - Full Control	0.460	0.102	(0.215, 0.706)	4.53	0.001
	No Control - Limited Cont.	0.326	0.102	(0.079, 0.573)	3.19	0.007
Severity	Injury-Fatal	-0.006	0.018	(-0.0466, 0.0352)	-0.33	0.943
	PDO - Fatal	-0.013	0.017	(-0.0523, 0.0265)	-0.77	0.724
	PDO - Injury	-0.007	0.010	(-0.02108, 0.00676)	-1.20	0.451

IV. Conclusions

The effects of thirteen categorical variables commonly used in the analysis and modeling of traffic accidents were evaluated using comparative analysis of the variation of the crash rates across the categories of the explanatory factors. The results indicate that about half of the factors, mainly related to highway design elements and drivers demographics appeared to have significant differences in mean, variances,and distributions of the crash rates across their respective categories. The crash rates across the categories of the environmental conditions (weather, pavement surface, and lighting), roadway designation (rural, urban), the day of the week, and hours of the day do not show remarkable differences in the mean and variances. The variables with insignificant differences in crash rates across their respective categories are not associated with an appreciable effect on the likelihood of occurrences of traffic accidents on Interstate highways. The approach used in the studyprovides valuable information on merits of the expected explanatory factors early on and reduces a substantial number of variables with little or no contribution to the likelihood of occurrences of crashes, and in formulating mathematical traffic accident prediction models.

Acknowledgments

The author would like to thank The U.S. Department of Transportation, Federal Highway Administration (FHWA); Highway Safety Information System (HSIS) for providing the data.

References

[1]. Abdulhafedh, A. (2016). Crash frequency analysis. *Journal of Transp.Techn.* 6, 169-180.
[2]. Anderson, J. C. and Dong, S. (2017). Heavy-vehicle driver injury severity analysis by time of week: A mixed logit approach using HSIS crash data. *ITE Journal*, 41-49.
[3]. Blincoe, L., Seay, A. G.,Zaloshnja, E., Miller, T. R., Romano, E. O.,Luchter, S. and Spicer, R. S. (2002). The economic impact of motor vehicle crashes, 2000," NHTSA Technical Report No. DOT HS 809 446 Washington, DC: National Highway Traffic Safety Administration (NHTSA) (May, 2002).
[4]. Blincoe, L., Miller, T. R.,Zaloshnja, E. and Lawrence, B. A. (2015). The economic and societal impact of motor vehicle crashes, 2010 (Revised). NHTSA Technical Report No. DOT HS 812 013Washington, DC: National Highway Traffic Safety Administration (NHTSA) (May, 2015).
[5]. Cerwick, D. M., Gkritza, K., Shaheed, M. S. and Hans, Z. (2014). A Comparison of the Mixed Logit and Latent Class Methods for crash severity analysis. Analytic Methods in Accident Research 3-4, 11-27.
[6]. Highway Performance Monitoring System, Field Manual. (1990). U.S. Department of Transportation, FHWA Publication 5600.1A, Washington, D.C.
[7]. Ivan, J. N., Wang, C. and Bernardo, N. R. (2000). Explaining two-lane highway crash rates using land use and hourly exposure. Accident Analysis and Prevention 32, 787-795.
[8]. Kassu, A. and Anderson, M. (2018). Determinants of severe injury and fatal traffic accidents on urban and rural highways. Int. J. Traffic and Transp. Engineering 8(3), 294-308.
[9]. Kruskal W. H. and Wallis, A. W. (1952). Use of ranks in one-criterion variance analysis.Journal of the American Stat. Ass. 47(260), 583-621.
[10]. Milton, J. and Mannering, F. (1998). The relationship among highway geometries, traffic-related elements and motor-vehicle accident frequencies. Transportation 25, 395-413.
[11]. Noland, R. B. and Oh, L. (2004). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. Accident Analysis and Prevention 36, 525-532.
[12]. Nujjetty, A. P., Sharma, S. and Council, F. M. (2015). Guidebook for State Data Files: Ohio. Federal Highway Administration, Highway Safety Information System (HSIS), Office of Safety and Office of Safety Research & Development, Washington, D.C., June 2015.
[13]. NSC Motor Vehicle Fatality Estimate. (2017). National Safety Council (NSC) http://www.nsc.org/NewsDocuments/2017/12-month-estimates.pdf, Accessed on Oct 30, 2017).
[14]. Oña, J., Mujalli, R.O. andCalvo, F.J. (2011). Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accident Analysis and Prevention 43, 402-411.
[15]. Vogt, A. and Bared, J. (1998). Accident models for two-lane rural segments and intersections.Transp. Res. Rec.: J Trans. Res. Board1635, 18-29.
[16]. Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. Biometrika 38(3/4), 330-336.
[17]. Yasmin, S. andEluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis and Prevention 59, 506-521.
[18]. Zhou, M. and Sisiopiku, V. (1997). Relationship between volume-to-capacity ratios and accident rates. Transp. Res. Rec.: J Trans. Res. Board 1581, 47-52.

Aschalew Kassu " A Comparative Evaluation of the Effects of Categorical Factors on the Safety of Multilane Interstate Highways." IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE) , vol. 15, no. 4, 2018, pp. 56-62