KSCE Journal of Civil Engineering (2019) 23(8):3706-3713

DOI 10.1007/s12205-019-1707-5

Transportation Engineering

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Developing Targeted Safety Strategies Based on Traffic Safety Culture Indexes Identified in Stratified Fatality Prediction Models

Soyoung Jung*, Xiao Qin, and Cheol Oh*****

Received August 26, 2018/Revised February 7, 2019/Accepted May 28, 2019/Published Online July 8, 2019

Abstract

The Korea Transportation Safety Authority (KTSA) conducts the special traffic safety culture investigation (STSCI) every year to assist local governments in promoting traffic safety. To address the issue of diversity, the local agencies were grouped into four regions by administrative district unit and offered region-specific safety promotion strategies. However, it is unclear if such a classification truly reflects the underlying differences that contribute to traffic safety. The goal of this study is to identify the most relevant attributes that affect the safety performance of local agencies (called traffic safety culture indexes in the current study) so that targeted safety promotion strategies can be recommended. To accomplish the goal, latent class cluster-based negative binomial regressions were applied for a comprehensive list of factors such as demographics, socio-economic features, roadway conditions, traffic violations and road user driver behavior; resulting in seven clusters of local governments. The following indexes were found to significantly and strongly affect crash fatalities in the clusters: rate of wearing helmet, rate of pedestrian's signal compliance, the number of unlicensed driving violations, total paved road length, province, ratio of male to female, and population density. Further, stratified negative binomial regression models were developed to identify statistically significant factors for predicting fatal crashes within each cluster. These cluster-specific features allow the KTSA to design targeted strategies for effective safety promotion.

Keywords: *local agencies, safety promotion, traffic safety culture indexes, cluster-specific features, targeted strategies*

1. Introduction

Safety culture is the socially constructed abstract system of meaning, norms, beliefs, and values (Myers *et al.*, 2014; Reiman and Rollenhagen, 2014). In the field of traffic, the safety culture varies widely by regional demographics, socioeconomic features, road conditions, and driver behaviors. Accordingly, region-specific and targeted traffic safety promotion strategies should be considered and developed.

Many previous studies have considered a variety of factors as the region-specific traffic safety performance indicators as crashes are complex events. To develop a spatial crash proneness measurement in Florida, Kocatage et al. considered demographics and socio-economic characteristics including ethnicity, poverty, vehicle ownership, and education levels of residents as the socio-economic characteristics (Kocatage *et al*., 2017). Road user behavior and traffic violation characteristics were also identified as traffic safety performance indicators in several US studies. Two survey studies considered driver behavior such as speeding, impaired and drowsy driving, red-light running, cell-phone use, and wearing a seatbelt and helmet as traffic safety performance indicators (Mizenko *et al*., 2014; AAA Foundation for Traffic Safety, 2015). In a Minnesota study, Ward *et al.* identified driving under the influence of alcohol as a traffic safety performance indicator (Ward *et al.*, 2015).

Another US study recommended roadway related features such as traffic volume, number of road lanes, and lane width as predictive safety performance measures for freeway or arterial systems (Islam *et al*., 2016). A California study also recommended roadway related features including mode share for work travel, delay, vehicle miles traveled, mode share of work travel, distressed lane miles, and pavement condition as safety management performance measures in rural areas (Reinke *et al*., 2017).

Weather conditions are commonly considered as a traffic safety performance indicator, reflecting a regional characteristic. Addressing large spatial and temporal variation of winter events, a Canadian study considered winter road maintenance as an effective treatment to improve safety and mobility (Fu *et al*., 2017).

Similarly, local governments in South Korea (Korea, hereafter) have been making efforts to identify their own traffic safety performance features for traffic safety promotion. However, the Korean local governments face budget limits and lack the manpower to address their own traffic safety issues. In this regard, the Korea

*Assistant Professor, School of Safety Engineering, Dongyang University, Yeongju 11307, Korea (Corresponding Author, E-mail: jung2@dyu.ac.kr)
**Associate Professor, Dept. of Civil and Environmental Engineering, University of Wisconsin-Milwaukee, WI 53201, USA (E-mail: qinx@uwm.edu)
***Member, Professor, Dept. of Transportation and Logistics Engineering, Hanyang University Erica Campus, Ansan 15588, Korea (E-mail: cheolo@hanyang.ac.kr)

Transportation Safety Authority (KTSA) has conducted the special traffic safety culture investigation (STSCI) every year. With the STSCI, the KTSA supports local governments by providing consulting services on crash fatality reduction and traffic safety promotion.

During the STSCI, local governments are grouped into four regions by administrative district unit. In each region, the KTSA collected diverse regional information such as crash fatalities and traffic safety performance indicators (called traffic safety culture indexes in the current study) including traffic violations, roadway conditions, road user behavior, demographics and socio-economic features. Then, the KTSA assigned a weight for every safety culture indexes based on the available information and calculated the total score for every local government. A focus of the KTSA's consulting support will be placed on the local governments with low scores.

The problem of this arbitrarily grouping a diverse area by a single variable (i.e., administrative district) is apparent. With the current classification, the safety culture indexes may not have a strong association with the safety outcome (i.e., crash fatalities). Globally, few studies about grouping regions of homogeneous traffic safety culture have been conducted. Therefore, the need for a better rationale to classify the regions into homogenous groups based on their traffic safety characteristics prompts this research. It is also important for the central government to consult each local government group with more specific and targeted safety advice following their unique problems.

Correspondingly, the objectives of this study are to identify new critical traffic safety culture indexes to effectively group local governments and to develop relevant and practical traffic safety promotion strategies. In this study, each Korean local government was characterized by four-year traffic safety culture related information using a latent class cluster-based analysis. This data-driven approach is the first attempt to quantitatively address a methodological process that identified critical classifiers for local government groups and developed traffic safety promotion strategies at the local level.

2. Data Collection and Processing

A local government district in Korea is consistent with an administrative division unit. According to the Local Government Act in Korea, there are three administrative division units at local level, which are called: "gun", "si", and "gu."

The "gun" indicates a fundamental administrative division unit with a population of less than 50,000, which is functionally similar to a "county" in U.S. The "si" is defined as the city with a population of 50,000 or more. Specifically, a city (si) with more than a population of 1,000,000 is called a metropolitan city. If a si has a population of more than 500,000, the si then includes the sub-district, "gu". Counting the entire gun, si, and gu in Korea, a total of 208 local administrative districts in 2014 were considered in this study (Korean Statistical Information Service, 2017). Note that Seoul, the capital city of Korea, was eliminated from a list of local governments that need policy consulting in this study because Seoul is not under the influence of central transportation safety authority but runs its own policy research institute. All of the 208 local government districts are shown in Fig. 1.

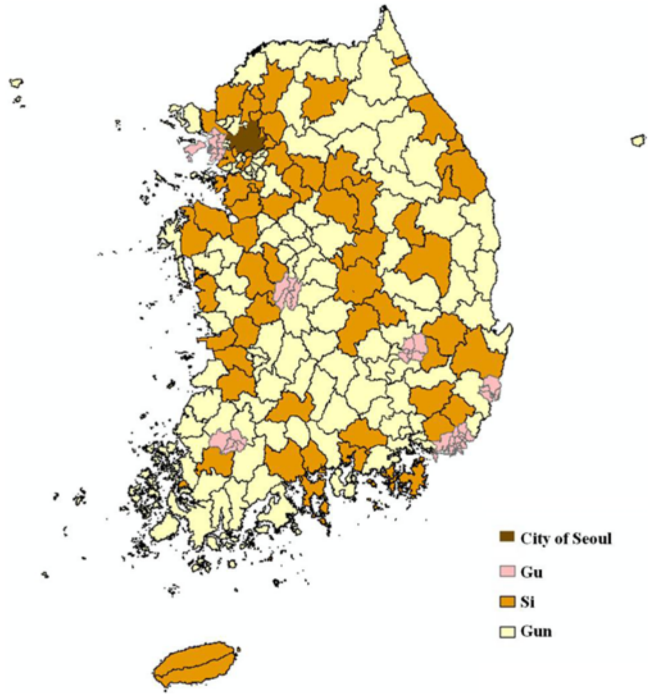

Fig. 1. Boundaries of Local Government Districts in Korea

For the 208 local governments, regional characteristics have been collected by the KTSA from 2011 to 2014. The KTSA dataset have 28 data fields in total including an administrative district, crash fatalities, demographics, socio-economic features, roadway conditions, traffic violations, pedestrian and driver behavior. The KTSA presumed all 28 data fields reflect regional traffic safety culture in local government districts. Four cases included missing data fields, which were omitted from the KTSA dataset. The final KTSA dataset includes 518 cases with 28 variables, which is summarized in Table 1.

In Table 1, there are four local government groups according to the KTSA: sub-divisions of metropolitan cities in particular (gu, Region 1), local government districts with a population of less than 50,000 (gun, Region 2), cities (si) with a population of less than 300,000 (Region 3), and cities (si) with a population of 300,000 or more (Region 4).

3. Latent Class Cluster-based Analysis

The clustering method allows separating groups with critical classifiers so that the variation within the group can be minimized and the differences between groups can be maximized. The classification also mitigates the influence of confounding factors within each homogeneous group.

Typical clustering methods are partitioning-based (such as K-means), hierarchical-based (such as Ward's method) and density-based (such as latent class clustering) approaches (Mohamed *et al.*, 2013). Particularly, the latent class cluster (LCC) method is

Table 1. Average Values in Characteristics of Current Local Government Groups

Variable (unit)		Region 1	Region 2	Region 3	Region 4	Full
Sample cases		68	224	142	84	518
Local-level administrative districts	The number of gun	0	81	0	0	81
	The number of si	0	0	57	26	83
	The number of gu	18	0	0	0	18
Traffic safety performance measure						
The number of crash fatalities (per 10^5 populations)		9	32	18	10	21
Demographics						
Province	East northern province of Korea (cases)	0	36	22	4	62
	Capital province of Korea (cases)	4	13	36	52	105
	North central province of Korea (cases)	0	24	4	4	32
	South central area of Korea (cases)	0	13	11	2	26
	North southern area of Korea (cases)	25	33	24	2	84
	South southern area of Korea (cases)	19	39	14	12	84
	North western area of Korea (cases)	0	13	7	4	24
	South western area of Korea (cases)	20	53	20	0	93
	Jeju island (cases)	0	0	4	4	8
Area of local government district (km^2)		72	699	560	497	546
Population (10^5 people)		2.8	0.6	1.7	6	2.1
Population density (10^3 people/km^2)		7.1	0.1	1.0	3.2	1.8
Ratio of male to female		1.02	1.03	1.04	1.13	1.05
The number of households (10^3)		105	24	68	229	80
Socio-economic features						
Gross district product (10^9 dollars)		6.4	2.1	15	30	11
Total municipal budget (10^6 dollars)		280	342	655	1040	533
Budget of transportation sector (10^6 dollars)		9	20	61	150	51
Ratio of transportation budget (%)		3.5	6	10	14	8
The number of registered vehicles (10^3 vehicles)		103	25.6	69.1	233.8	81.4
Roadway condition						
Total road length (km)		328	432	640	874	548
Total paved road length (km)		272	328	445	717	416
Rate of road pavement (%)		88	77	73	84	79
Total number of parking zones (10^3)		6	2	5	16	5
Total number of parking lots (10^3)		98	12	54	230	71
Traffic violations						
Rate of illegal parking in school zone (%)		13	12	10	14	12
The number of speeding violations (10^3)		28	21	36	64	33
The number of unlicensed driving violations		319	109	274	652	270
The number of driving under alcohol effect (10^3)		2.0	0.4	1.1	3.3	1.2
Pedestrian and driver behavior						
Rate of pedestrian's signal compliance (%)		88	78	85	86	82
Rate of stopline compliance (%)		69	72	71	66	71
Rate of wearing seatbelt (%)		80	61	68	71	67
Rate of traffic signal compliance (%)		95	92	94	94	93
Rate of turning vehicle directional signal (%)		62	73	67	63	69
Rate of wearing helmet (%)		76	67	73	69	70

Note: Regions 1 to 4 indicate the current groups of local governments classified by the KTSA; All values in each cell were averaged over cases; A single parking zone includes several parking lots; The sum of sample cases for all four regions is placed on the "Full" column.

considered to be one of the most advanced methods because its posterior membership probabilities are estimated directly from model parameters and cases are assigned to probability-based classification. In addition, the LCC method allows flexible model formulation that does not imply any assumptions regarding the nature of variables, their underlying distributions, and the correlation patterns across cases and variables. Finally, the LCC method addresses several model selection criteria including akaike information criterion (AIC), bayesian information criterion (BIC), or consistent akaike information criterion (CAIC) to select the optimum number of clusters (Depaire *et al*., 2008; Magidson and Vermunt, 2002).

In this study, the LCC model building was conducted with Latent GOLD 5.0 in which each element of LCC membership was computed from the estimated model parameters. The basic LCC model is formulated as follows (Vermunt and Magidson, 2002).

$$P(Y_i|\theta) = \sum_{j=1}^{J} f_j P(Y_i|C_j, \theta_j) \tag{1}$$

where, P = probability, Y_i = a vector of variables observed from ith case, j (j = 1 to J) = a latent class number, f_j = prior probability of data in latent class C_j, θ_j = a vector of cluster model parameters to be estimated, and P ($Y_i|C_j$, θ_j) = mixture probability density.

The number of clusters was determined in the current study by considering model selection criteria such as AIC, BIC, and CAIC, but BIC was emphasized due to better consistency and accuracy (Nylund *et al.*, 2007). Lower score in the model selection criteria indicates better number of clusters.

Stratified negative binomial (NB) regression models were developed to estimate the impacts of factors on crash fatalities within each LCC-based group of local governments. The number of crash fatalities is an over-dispersed count variable that can be effectively handled by the NB regression (Jung *et al*., 2011):

$$\ln(Y(r,t)) = \sum \beta X(r,t) + \varepsilon \tag{2}$$

where, $Y(r, t)$ = random variable in region r during fixed period of time t, β = coefficients of explanatory variables, $X(r, t)$ = explanatory variables, and exp(ε) follows gamma distribution.

4. Results and Discussion

4.1 Clustering

The use of a large number of variables increases the odds that the variables are no longer dissimilar in the LCC approach. In this regard, absolute Pearson correlations above 0.90 are always problematic (Mooi and Sarstedt, 2011). From all the variables

Table 2. Variance of Each Variable Explained by 7-LCC Model

Variable	R^2 value
Demographics	
Province	0.08
Area of local government district	0.49
Population	0.57
Population density	0.61
The number of households	0.55
Socio-economic features	
Gross district product	0.15
Total municipal budget	0.40
Ratio of transportation budget	0.29
Roadway condition	
Total road length	0.49
Total length of paved road	0.46
Rate of paved road	0.17
Total number of parking zones	0.61
Traffic violations	
Rate of illegal parking in school zone	0.02
The number of speeding violations	0.50
The number of unlicensed driving violations	0.70
Pedestrian/driver behavior	
Rate of pedestrian's signal compliance	0.11
Rate of stop line compliance	0.04
Rate of wearing seatbelt	0.10
Rate of traffic signal compliance	0.05
Rate of turning vehicle directional signal	0.11
Rate of wearing helmet	0.05

Note: Shaded cell in R^2 value indicates high R^2 values of variables to cluster.

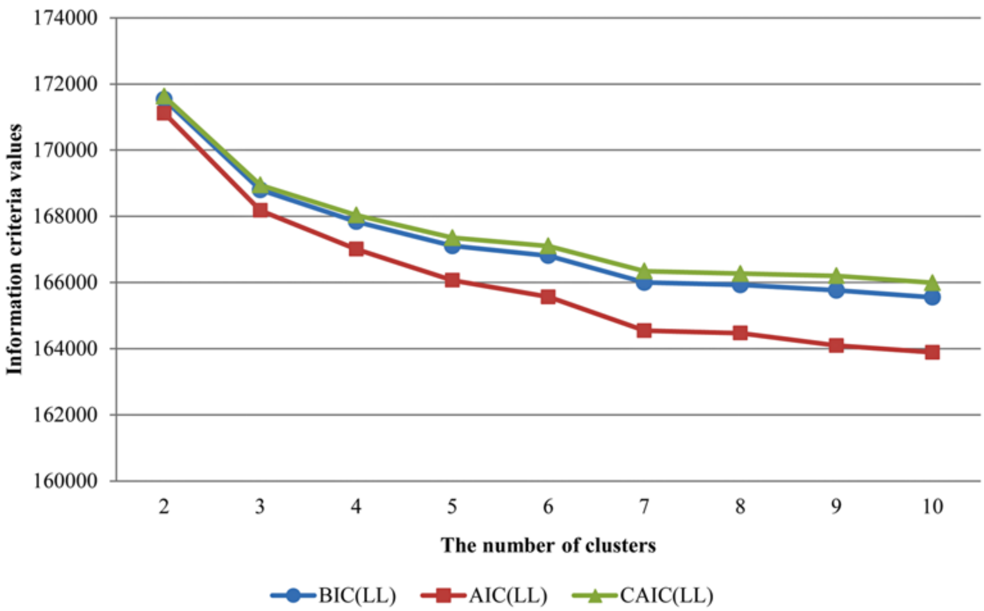

Fig. 2. Information Criteria to Determine the Number of Clusters

shown in Table 1, this study eliminated highly correlated variables in the LCC based on the Pearson correlation value of 0.9 within each group, including: budget of transportation sector, total area of parking lots, ratio of male to female, the number of registered vehicles, and the number of driving under the alcohol influence. The eliminated variables were highly correlated with two or more variables.

As an important traffic safety performance measure, crash fatalities per 10^5 populations that occurred in all Korean local government districts from 2011 to 2014 were considered in the LCC model. During the LCC model building, AIC, BIC, CAIC were compared to determine the number of clusters, as provided in Fig. 2.

According to Fig. 2, AIC, BIC, and CAIC values commonly decreased to seven and they did not show apparent differences after the seven clusters. Especially, the BIC values, which are known to be more reliable (Vermunt and Magidson, 2002), were almost the same after seven clusters. Therefore, seven clusters were finally determined to group crash fatalities that occurred in local governments.

In the seven LCC models, the significance associated with each variable was measured by parameter estimates. For each variable, the p-value for the test of significance in the parameter effect estimates is less than .05, indicating that the null hypothesis stating that all of the effects associated with that variable that are zero would be rejected. In Table 2, the R^2 value of each variable measures how much of the data variation is explained, which is used for evaluating the capability of each variable to discriminate between clusters.

In Table 2, 16 variables with the R^2 values of 10% or greater were considered as critical classifiers that discriminate the variation in the seven LCC models. They are four demographic variables, four roadway variables, three socio-economic variables, two traffic violation variables, and three pedestrian/driver behavior variables. These variables are the top safety indicators of regional diversity that is pertinent to the number of crash fatalities. The resultant LCC profile is provided in Table 3.

Compared to the existing group of local governments (Table 1), the new numbers of clusters and local governments assigned to clusters in Table 3 were evidently different. Most clusters included more than one administrative division unit (gun, si, or gu). This finding suggests that organizing local governments by their safety-related features may be more effective in analyzing safety than grouping them by administrative division unit.

According to Table 3, special features to discriminate Cluster 1 were associated with road user behavior (pedestrian and driver)

Table 3. LCC Profiles

Response		CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7	Full
Cluster sample size (cases)		193	91	64	63	56	40	11	518
Local-level administrative division	The number of gun	65	13	1	1	0	1	0	81
	The number of si	3	22	23	9	9	13	4	83
	The number of gu	0	1	0	9	8	0	0	18
Crash fatalities per 10^5 population		33	25	15	7	7	15	8	21
Cluster classifier (Traffic safety culture index)									
Demographic index									
Area of local administrative district (km^2)		693	706	566	43	109	876	442	546
Population (10^5)		0.5	1.0	2.8	1.7	5.5	5.0	7.7	2.1
Population density (populations/km^2)		82	183	802	6951	6154	658	1971	1762
The number of households (10^3)		22	44	107	68	204	190	283	80
Socio-economic index									
Gross district product per year (10^9 dollars)		1.3	3.3	50.3	3.1	11.7	15.9	17.6	10.6
Total municipal budget per year (10^6 dollars)		333	445	622	258	612	1721	1143	533
Ratio of transportation sector budget per year (%)		5	8	13	8	7	12	21	8
Roadway condition index									
Total road length (10^3 km)		415	649	662	188	535	1388	461	548
Total road length paved (10^3 km)		318	459	487	140	474	1031	423	416
Rate of road pavement (%)		77	73	77	83	90	73	86	79
Total number of parking zones (10^3)		0.9	2.7	8.6	2.9	11.3	19.9	14.3	5.3
Traffic violation index									
The number of speeding per year (10^3)		14	34	38	18	41	99	106	33
The number of unlicensed driving per year		78	168	347	205	523	977	568	270
Pedestrian/driver behavior index									
Pedestrian's signal compliance rate (%)		77	85	83	88	87	87	87	82
Rate of wearing seatbelt (%)		61	67	73	75	74	68	68	67
Rate of turning vehicle directional signal (%)		71	73	67	64	61	60	66	69

Note: All values in each index were averaged; CL indicates cluster; Shaded cell indicates the value of indexes that contribute to a certain cluster separation.

related indexes. The rates of pedestrian's signal compliance (77%) and seatbelt wearing (61%) in Cluster 1 were comparatively lower than those in any other clusters. Cluster 2 distinguished itself by the road user behavior index as well. The rate of turning vehicle directional signal was found to be 73%, which was the highest of any other clusters.

Clusters 3 to 5 also distinguished themselves by a single traffic safety culture index. A lot of yearly gross district product (50.3 billion dollars) was found to be a critical classifier to discriminate Cluster 3. Comparatively high population density (6951 population/km^2) was identified as a single classifier to separate Cluster 4. The rate of road pavement in Cluster 5 was identified highest among seven clusters, which was the special feature to separate Cluster 5.

Cluster 6 was characterized by six traffic safety culture indexes, most evidently by roadway condition indexes, which were: long total road length (1388 thousand km), long total paved road length (1031 thousand km), and large number of parking zones (19.9 thousand), large area of local administrative district (876 km^2), big total municipal budget (1721 million dollars), and large number of unlicensed driving violations.

Similarly, Cluster 7 distinguished itself by several traffic safety indexes. However, Cluster 7 contained only 11 cases and it is hard to generalize cluster characteristics with small sample size.

To sum up, seven clusters are labeled as follows. Note that values in parenthesis for each cluster are averaged over all sample cases:

1. Cluster 1: Local governments with low rates of pedestrian's signal compliance (77%) and wearing seatbelt (61%).
2. Cluster 2: Local governments with high rate of turning vehicle directional signal (73%).
3. Cluster 3: Local governments with lots of yearly gross district product (50 billion dollars).
4. Cluster 4: Local governments with high population density (6961 population/km^2).
5. Cluster 5: Local governments with high rate of road pavement (90%).
6. Cluster 6: Local governments with large district area (876 km^2), big total municipal budget (1721 million dollars per year), large numbers of unlicensed driving violations (977 per year) and parking zones (19.9 thousands), long total road length (1388 thousand km) and paved road length (1031 thousand km).
7. Cluster 7: Local governments with many populations (770 thousands), large number of households (283 thousands), high ratio of transportation sector budget (21% per year),

Table 4. LCC-based NB Regression Models

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6		Cluster 7	
Cluster characteristics	– Low rate of pedestrian's signal compliance – Low rate of wearing seatbelt		– High rate of turning vehicle directional signal		– Lots of gross district product per year		– High population density		– High rate of road pavement		– Large area – Big total budget – Many unlicensed driving violations – Many parking zones – Long road length – Long paved road length		– Many populations – Large number of households – High ratio of transportation budget – Many speeding violations	
Deviance/D.F.	1.03		1.05		1.07		0.94		0.79		0.93		0.88	
Pearson chi-square/D.F.	1.04		1.06		1.06		1.02		0.86		0.94		0.82	
Average fatalities per 10^5 populations (total sample size)	33 (193)		25 (91)		15 (64)		7 (63)		7 (56)		15 (40)		8 (11)	
Explanatory variable	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Intercept	4.152	<.0001	3.416	<.0001	4.681	<.0001	-1.285	0.065	2.888	<.0001	3.060	<.0001		
North southern province							**0.436**	0.005						
Administrative district area											**0.003**	0.002		
Population					-0.002	<.0001			-0.001	0.004				
Population density					-0.001	0.001					-0.001	<.0001		
Male/female ratio							**3.596**	0.004						
The number of households	-0.002	<.0001												
Gross district product per year									0.001	0.080				
Rate of transportation sector budget					**-0.028**	0.001								
Registered vehicle number							-0.010	0.001						
Total paved road length													**1.514**	0.002
Rate of road pavement									**-0.008**	0.050				
Parking lot number					0.003	0.017								
Rate of wearing helmet	-0.002	0.031	-0.007	0.013	-0.002	<.0001								
The number of speeding violations			0.007	0.001										
The number of unlicensed driving violations	0.002	<.0001							**0.006**	0.100				
Pedestrian's signal compliance rate	**-0.003**	0.015	**-1.059**	0.022										
Dispersion	0.072		0.075		0.050		0.063		0.015		0.015		0.010	

Note: Only significant variables were presented.; D.F. and Coeff. indicate the degrees of freedom and coefficients of parameter estimates, respectively.; Bold means parameter estimates of variables that showed comparatively stronger effect on crash fatalities than other variables identified in each cluster.

and large number of speeding violations per year (106 thousands).

4.2 Stratified Crash Fatality Prediction Models

As a primary traffic safety performance measure, fatalities involved in crashes are common interests of Korean local government agencies for traffic safety promotion. The number of crash fatalities per 10^5 populations, a count variable, is the response of LCC-based NB regression model. To build the LCC-based NB regression models, all variables except for administrative division in Table 1 were considered as explanatory variables, which included 25 continuous variables and one nominal variable (province). The relation between a response and each of the explanatory variables was tested so that a single significant predictor to the response could be individually selected. Then, backward elimination was conducted to select the best multiple regression model with a parameter significance level for the response, which removed multicollinearity between highly correlated explanatory variables.

A goodness of model fit, deviance and Pearson chi-squared statistic values were used. The deviance and Pearson chi-square statistic value divided by degrees of freedom should be approximately one, indicating a good model fit for the data set. The parameter significance level used in this study was .01.

The resultant LCC-based NB regression models are provided in Table 4.

4.3 Safety Promotion Strategies

A set of NB models were developed for the dataset based on the current classification (i.e., the four regions shown in Table 1). The results did not provide useful insights on specific and practical traffic safety strategies. Only two types of variables were statistically significant within each region. They were province and population. On the other hand, the LCC-based NB regression based on the new classification offers specific details to develop targeted traffic safety promotion strategies for local governments.

According to Table 4, the impacts of demographic factors (administrative district area, population, population density, the ratio of male to female, and the number of households) were significantly identified in some LCC-based NB regression models, but their impacts on crash fatalities were weak. Demographics related policies need significant time to take effect because they hardly show dramatic change in a certain area.

However, policies from learning the impacts of factors related to roadway condition, traffic violation, road user behavior, take comparatively immediate effect on traffic safety promotion. Hence, we will further emphasize these aforementioned factors that have strong and significant impacts on crash fatalities for the following discussion.

As presented in Table 4, all NB regression models were appropriate in terms of goodness of fit (deviance and Pearson's chi-squared values divided by D.F.) close to one (i.e. close to the perfect fit) and dispersion parameters greater than zero.

In Clusters 1, high rates of wearing a helmet and pedestrian's signal compliance were likely to reduce crash fatalities while a large number of unlicensed driving violations was likely to increase crash fatalities. In Cluster 2, the impact of high pedestrian's signal compliance rate on reducing crash fatalities was stronger than any other factors. These findings in Clusters 1 and 2 imply that education, campaign or law enforcement strategies would be effective to reduce crash fatalities targeting particularly pedestrians and two-wheel vehicle users. More crash fatalities were observed in Clusters 1 (33 fatalities) and 2 (25 fatalities) than any other clusters. This finding also implies that it is necessary for local governments assigned to Clusters 1 and 2 to immediately accept the aforementioned strategies.

Socio-economic features, that is, high rate of transportation sector budget and small number of registered vehicles were found to significantly reduce crash fatalities in Clusters 3 and 4, respectively. Accordingly, an increase of the transportation sector budget and vehicle registration management strategies would be useful for traffic safety promotion in Cluster 3 and 4. Large gross district product that characterizes Cluster 3 could use a substantial municipality budget and transportation sector budget. Additionally, the north southern province and high ratio of male to female significantly increased the likelihood of crash fatalities in Cluster 4. This finding gives a hint on geographical location and target gender of education strategy for traffic safety promotion.

For Clusters 5 and 7, the impacts of roadway condition and traffic violation factors on crash fatalities were comparatively strong. In Cluster 5, high road pavement rate, the characteristics of Cluster 5 as well, and less unlicensed driving violations were likely to reduce crash fatalities. The findings imply that strict law enforcement against unlicensed driving would be helpful to promote traffic safety as well as further improvement of road pavement rate in Cluster 5. In Cluster 7, long paved road length was likely to increase crash fatalities. However, the result is doubtful due to the small sample size in Cluster 7.

For Cluster 6, large district area and low population density were found to reduce the likelihood of crash fatalities. From the finding, administrative district area adjustment or population distribution strategies can be considered to reduce crash fatalities in Cluster 6, which would be a long-term traffic safety promotion strategy.

5. Conclusions

The KTSA consult about traffic safety promotion strategy implementation to support local governments that struggle with a tight budget and shortage of staff. Considering the variation across local agencies, the KTSA currently grouped them into four regions by administrative district unit. However, the current classification of local governments is unable to adequately reflect the diverse traffic safety cultures. Therefore, this study proposed new methods for classifying local agencies by traffic safety culture indexes using LCC-based NB regressions. The key findings are as follows:

1. Seven LCC clusters were determined by sixteen traffic safety culture indexes that provide specifications to develop targeted traffic safety promotion strategies.
2. Targeting pedestrians or two-wheel vehicle users, education, campaign or law enforcement strategies would promote traffic safety in local governments with low rates of pedestrian's signal compliance and wearing a seatbelt (Cluster 1) and high rate of turning vehicle directional signal (Cluster 2).
3. Increasing transportation sector budget and strengthening vehicle registration management would help traffic safety promotion in local governments with great gross district product (Cluster 3) and high population density (Cluster 4), respectively.
4. Strict law enforcement against unlicensed driving would help to promote traffic safety in local governments with high rate of paved roads (Cluster 5).
5. Long-term traffic safety planning such as district adjustment or population distribution is needed for promoting traffic safety in Cluster 6.
6. The result for Cluster 7 implies the need of more sample cases for reliable recommendations.

This data-driven approach equipped decision-makers with insightful information extracted from diverse local information. The methodologies used in this study to identify critical traffic safety culture indexes and to recommend regional specific safety strategies can be applied in other countries with a nature of diversity. However, there is a limitation of data collection in the current study. The KTSA dataset involved neither certain local governments' information nor the recent year (2015 to 2016) information. Therefore, more data collection involving all data records in the entire local governments will provide more comprehensive and reliable results. The analysis in the current study was only based on fatality as the injury level. Future research will combine more injury levels to provide comprehensive safety strategies. An evaluation about fulfilling the KTSA traffic safety promotion strategy recommendations would be also further research issue by before-after study.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1A09083905).

References

AAA Foundation for Traffic Safety (2016). *2015 traffic safety culture index*, AAA Foundation for Traffic Safety, Washington, D.C., USA, pp. 1-35.

Depaire, B., Wets, G., and Vanhoof, K. (2008). "Traffic accident segmentation by means of latent class clustering." *Accident Analysis and Prevention,* Vol. 40, pp. 1257-1266, DOI: 10.1016/j.aap.2008.01.007.

Fu, L., Usman, T., and Kwon, T. (2017). *Development of a model-based performance measurement tool for winter road maintenance management*, Report No. 17-0309, 96th Transportation Research Board Annual Meeting, Washington, D.C., USA.

Islam, M., Perez-Bravo, D., and Silverman, K. (2016). *Performance-based assessment to transportation safety planning for metropolitan travel improvement study*, Report No. 17-06667, 96th Transportation Research Board Annual Meeting, Washington, D.C., USA.

Jung, S., Qin, X., and Noyce, D. A. (2011). "Modeling highway safety and simulation model in rainy weather." *Transportation Research Record*, Vol. 2237, No. 1, DOI: 10.3141/2237-15.

Kocatage, A., Ulak, M., Ozguven, E., Horner, M., and Arghandeh, R. (2017). *Socioeconomic characteristics and crash proneness: A case study in florida using two-step floating catchement area method*, Report No. 17-06321, 96th Transportation Research Board Annual Meeting, Washington D.C., USA., 96th Transportation Research Board Annual Meeting, Washington D.C., USA.

Korean Statistical Information Service (2017). *Regional statistics*. Korean Statistical Information Service, http://kosis.kr/index/index.do [Accessed on July 17, 2017].

Magidson, J. and Vermunt, J. (2002). "Latent class models for clustering: A comparison with K-means." *Canadian Journal of Marketing Research*, Vol. 20, pp. 7-44, DOI: 10.1037/met0000095.

Mizenko, A., Tefft, B., Arnold, L., and Grabowski, J. (2014). *Older American drivers and traffic safety culture: A long road study*, AAA Foundation for Traffic Safety, Washington, D.C., USA, pp. 1-14.

Mohamed, M., Saunier, N., Miranda-Moreno, L., and Ukkusuri, S. (2013). "A clustering regression approach: A comprehensive injury severity analysis of pedestrian-vehicle crashes in New York." *Safety Science*, Vol. 54, No. 1, pp. 37-44, DOI: 10.1016/j.ssci.2012.11.001.

Mooi, E. and Sarstedt, M. (2011). *Concise guide to market research,* Springer, Heidelberg, Germany.

Myers, D. J., Nyce, J. M., and Dekker, S. W. A. (2014). "Setting culture apart: Distinguishing culture from behavior and social structure in safety and injury research." *Accident Analysis and Prevention*, Vol. 68, pp. 25-29, DOI: 10.1016/j.aap.2013.12.010.

Nylund, K., Asparouhov, T., and Muthén, B. (2007). "Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study." *Structural Equation Modeling*, Vol. 14, pp. 535-569, DOI: 10.1080/10705510701575396.

Reinke, D., Damkowich, J., and Landon, D. (2017). *Development of performance measures for rural counties in California*, Report No. 17-06203, 96th Transportation Research Board Annual Meeting, Washington D.C., USA.

Reiman, T. and Rollenhagen, C. (2014). "Does the concept of safety culture help of hinder systems thinking in safety?" *Accident Analysis and Prevention*, Vol. 68, pp. 5-15, DOI: 10.1016/j.aap. 2013.10.033.

Vermunt, J. and Magidson, J. (2002). *Applied latent class analysis*, Cambridge University Press, Cambridge, UK.

Ward, N., Otto, J., Swinford, S., and Borkowski, J. (2015). *Measuring Minnesota's traffic safety culture*, Report No. MN/RC 2015-13, Minnesota Department of Transportation, MN, USA.