Identifying Hidden Patterns of Fatal Pedestrian Traffic Accidents in East Azerbaijan Province of Iran: Application of Categorical Principal Component Analysis (CATPCA)

Milad Jamali-Dolatabad

Tabriz University of Medical Sciences

Parvin Sarbakhsh (✉ p.sarbakhsh@gmail.com)

Tabriz University of Medical Sciences

Homayoun Sadeghi-bazargani

Tabriz University of Medical Sciences

Research Article

Keywords: Pedestrian safety, CATPCA, PCA, pattern, Traffic-related death, Iran

Posted Date: December 14th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-118059/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Identifying patterns and hidden relationships among fatal pedestrian road traffic accident (FPRTA) features can be effective to reduce pedestrian fatalities. This study aimed to detect the patterns of FPRTA in the East Azerbaijan province of Iran.

Methods: The present study is a descriptive-analytic study based on the data of all 1782 FPRTAs that occurred in East Azerbaijan province of Iran during the years 2010 to 2019 collected by the forensic organization. CATPCA (Categorical Principal Component Analysis) was performed to recognize hidden patterns in the data by extracting principal components from the set of 13 features of FPRTA. The importance of each component was assessed by using the variance accounted for (VAF) index.

Results: The optimum number of components to fit the CATPCA model was six with a 71.09% explanation of the total variation. The first and most important component with VAF=22.04% contained the demographic and socioeconomic characteristics of the pedestrians. The second-ranked component with VAF=12.96% was about the injury. The third component with VAF=10.56% was about the severity of the accident. The fourth component with VAF=9.07% was somehow related to the knowledge and observance of the traffic rules. The fifth component with VAF=8.63% was about the quality of medical relief and finally, the sixth component with VAF=7.82% was about environmental conditions.

Conclusion: CATPCA revealed hidden patterns of FPRTA data in the format of six components. The revealed patterns showed that some interactions between correlated features led to a higher mortality rate.

Background

Pedestrians are among the most vulnerable road users, which about 22% of the fatalities due to traffic accidents in the world are related to them[1]. In the case of Iran, it equals 23% of total traffic deaths.[1] Every year, around 400,000 pedestrians are killed in road accidents, which more than half of them belong to low-income countries[2]. In developing countries, injury and death of pedestrians are more severe than in developed countries and 85% of deaths and 90% of disabilities occur in the developing countries[3–7].

According to the WHO (World Health Organization), Iran is one of the developing countries with a high rate of traffic-related death. The estimated road traffic death rate in Iran is 20.5 per 100,000 people in 2016 which 22% of this rate occurs in pedestrians[1]. The proportion of pedestrian deaths is high in populous cities[8]. East Azerbaijan Province is located in the northwest of Iran. According to the 2016 Iranian population census, the population of this province is 3.91 million people, and accounting for 4.89 percent of the country's total population is considered one of the most populous provinces of Iran with a high rate of pedestrian road traffic accidents (FPRTA).

The pattern of fatal traffic accidents can be different from non-fatal accidents. Also, it can be different in various places. So, identifying patterns in the accident data and knowing the importance and priority of the factors contributed to occurring or severity of the accidents is of utmost importance. Recognition of such patterns enables managers and policymakers to estimate the effectiveness of interventions planned to reduce accidents or the severity of injuries.

On the other hand, many variables (in machine learning terminology; features) contribute to the occurrence and severity of traffic accidents. In data with high dimensions, it can be hard to detect and describe patterns manually. So, for accurate pattern extraction, we need a better way to deal with high-dimensional data which leads to simple interpretations. Dimension reduction techniques can be used to reduce a large set of variables to a smaller set so that still contains most of the information in the large set and we can quickly extract patterns and insights from it. One of the most commonly used methods to reduce the dimension of data and reveal hidden patterns is the principal component analysis (PCA) method. An extended version of this method for categorical data (i.e., ordinal and nominal data) is the nonlinear categorical principal component analysis method (CATPCA)[9].

So, the objective of this study is to identify the hidden pattern of the fatal pedestrian road traffic accidents in East Azerbaijan province of Iran from 2010 to 2019 based on the forensic organization data by using the CATPCA method for taking into account the unsupervised nature of the data (unlabeled data) and categorical nature of the variables of this study.

Methods

Data

The present study is a cross-sectional (descriptive-analytic) study based on data collected by the forensic organization of East Azerbaijan province. According to the WHO definition, a fatal traffic accident was defined as the accident in which the person involved in the accident

was killed immediately or within 30 days as a result of the accident[1]. A total of 7,785 deaths due to the road traffic accident has been recorded by the forensic organization of East Azerbaijan province during the years 2010 (March 21) to 2019 (March 21). From them, 139 accidents had occurred in other provinces, and 238 deaths had occurred after the 30th day so they were omitted from the data. Therefore, according to the WHO definition, 7408 traffic accident fatalities have occurred in East Azerbaijan province of Iran which 1782 of them (24.05%) were pedestrian fatal traffic accidents. So, the final number of fatal pedestrian traffic accidents that were included in this study was 1,782.

Collected data of the forensic organization includes information such as demographic and socioeconomic characteristics of pedestrians (age, gender, job, marriage, and education), kind of vehicle involved in the crash, type of transferring injured pedestrian to hospital, injured organs, leading cause of death, location of the death, location of the accident (urban- Suburban roadways), lightness condition and season of the accident. Details of data collection have been published elsewhere[10].

Statistical methods

To describe the data, numbers and percentages were used. Association between variables was assessed via the Chi-Square test. To find the structure and principal components of the data, due to the categorical nature of our variables (ordinal or nominal), the unsupervised CATPCA method was used for data analysis.

The importance of each component was assessed by using the variance accounted for (VAF). Scree plot and Kaiser's criterion which recommends retaining all factors with an eigenvalue above 1[11] were used to determine the optimum number of components. Varimax with Kaiser Normalization rotation method was used to rotate components. The R software version 3.5.1 and Gifi packages were used to fit the CATPCA model.

In the following, a brief introduction to the CATPCA method is mentioned.

The PCA and CATPCA methods

In scientific research, summarizing and extracting information from raw data is very important. Considering that in recent decades the data collected in all areas, especially in the field of medical sciences, have high dimensions, statistical methods in extracting information from these data become problematic.

One of the statistical methods used in this situation is the dimension reduction technique. So that the information in the total data is summarized in some components that are derived from the combination of the main variables so that extracted components still contains most of the information of the original data. In practice, instead of using all the variables in the analysis, extracted components are used.

One of the most commonly used methods for dimension reduction is the principal component analysis (PCA) method that is an unsupervised reduction technique. In unsupervised methods, the users do not need to supervise the model; instead, it allows the model to work on its own to discover patterns and information that was previously undetected. The purpose of PCA as an unsupervised method is to reveal hidden structures and bring out strong patterns in a dataset by converting a set of observations of possibly correlated variables into a smaller number of **uncorrelated** variables called principal components and rank these components based on the score of each component.

The PCA method begins this task by estimating the covariance matrix of the predictor variables. In the PCA method, instead of using all the Eigenvalues, such as the regular regression method, we try to extract and apply a smaller number of Eigenvalues[12].

The non-linear CATPCA method is the nonlinear equivalent of the PCA method to reduce dimensions in categorical data. The most important advantage of non-linear CATPCA over the linear PCA method is that it combines nominal and ordinal variables and can discover nonlinear relationships between variables. Unlike the PCA method, CATPCA does not have high sensitivity to classical statistical assumptions such as normality (multivariate normality) and linear relationships between variables[9, 11]. CATPCA converts every category of the variables to a numeric value, using optimal quantification (also known as optimal scaling). In CATPCA similar to PCA, the overall summary diagnostic value is the percentage of variance accounted for (VAF) by the principal components which equals the sum of the eigenvalues (sum of squared of loading value of constituent variables) of the components divided by the total number of variables[11].

Results

Data description

In this study, data of 1782 fatal pedestrian traffic accidents were investigated. Thirteen features related to these accidents were included in the data analysis.

The descriptive pattern of fatal pedestrian accidents was as follows: Of the 1,782 fatal pedestrian accidents, the majority of victims were male (78.23%). with increasing age, the number of victims increased, so that, most of the victims (35.13%) were over 65 years old.

Most pedestrians killed in traffic accidents (75.08%) were married. Also, most of them were illiterate (44.22%) and self-employed (52.52%). Although there are no major differences in the number of fatal pedestrian accidents in different seasons, the highest number of fatal accidents occurred in the summer (29.24%) and the lowest in the winter (20.15%). In 72.22% of fatal pedestrian accidents, the involved car in the accident was a light vehicle. Most fatal accidents (62.40%) occur in daylight, and most of the victims (79.18%) were taken to the hospital by ambulance. Also, in most of these fatal accidents, injured organs were the head and neck (52.3%) and the most common leading cause of death was head trauma (58.25%). Furthermore, 40.7% of pedestrian fatalities resulting in traffic accidents occurred at the scene of the accident and 50.95% of them occurred in the hospital or after discharge from the hospital and the rest died during transfer. The majority of fatal accidents (59.82%) occurred on suburban roads. The description of the existed features is presented in Table 1.

Table 1
Description of the variables

variable	Attribute	number	percent
Pedestrian gender	male	1394	78.23
	female	388	21.77
Pedestrian Age group (year)	0–4	87	4.88
	5–14	121	6.79
	15–29	187	10.49
	30–44	269	15.10
	45–64	492	27.61
	≥ 65	626	35.13
Season of the accident occurrence	spring	456	25.59
	summer	521	29.24
	Autumn	446	25.03
	winter	359	20.15
Kind of vehicle involved in the crash	Light machinery	1,287	72.22
	Heavy machinery	271	15.21
	Motorcycle or bicycle	99	5.56
	Others	34	1.91
	Missing	91	5.11
Pedestrian occupation	student	171	9.59
	unemployed	508	28.51
	government employee	58	3.25
	Self-employed	936	52.52
	Without any Occupation (children)	87	4.88
	Missing	22	1.23
Pedestrian marital status	Single	434	24.35
	Married	1338	75.08
	Missing	10	0.56
Light condition	Daylight	1,112	62.40
	Night	402	22.56
	Twilight	160	8.98
	Missing	108	6.06
transmission of the injured pedestrian to the hospital	Ambulance	1,411	79.18
	Other	213	11.95
	Missing	158	8.87
Pedestrian education	Illiterate	788	44.22
	Primary education	453	25.42
	Guidance, High school and diploma	441	24.75

variable	Attribute	number	percent
	Academic education	72	4.04
	Missing	28	1.57
Pedestrian injured organs	Head and neck	929	52.13
	Chest, abdomen, and pelvis	183	10.27
	Other organs	27	1.52
	More the one organ	642	36.03
	Missing	1	0.06
leading cause of Pedestrian death	Head Trauma	1,038	58.25
	Bleeding	112	6.29
	Multiple Fractures	293	16.44
	Mixed causes	300	16.84
	Others	35	1.96
	Missing	4	0.22
location of the Pedestrian death	Death at the scene	714	40.07
	Death during the transfer	144	8.08
	Death at the hospital or Death at home	908	50.95
	Missing	16	0.90
Location of accident occurrence	Suburban roadways	1,066	59.82
	Urban roadways	694	38.95
	Missing	22	1.23

The Results Of The Catpca Method

For pattern extraction by CATPCA and finding hidden relationships among variables, ordinal and nominal variables were defined to the model. Given that the maximum number of possible components is the number of the main variables, the initial model was fitted using 13 components to determine the number of optimal components. According to Fig. 1, which shows the scree plot of the data, the point where the slope of the curve is leveling off (the elbow) was six indicates the number of factors that should be generated by the analysis. Also, according to Kaiser's criterion to find the number of the major components (with Eigenvalue higher than 1), the most appropriate number of components in this study was six.

Therefore, the CATPCA model with six components was fitted to the data. All underlying variables remained in CATPCA analysis with considerable loading values (> 0.4). The results have been presented in Table 2.

Table 2
Loading value of the features loaded in the six major components

Features	Comp1 (demographic and socioeconomic characteristic)	Comp2 (injury)	Comp3 (severity of the accident)	Comp4 (knowledge and observance of the traffic rules)	Comp5 (medical relief)	Comp6 (environmental conditions)
Age group	.944					
Marital status	.910					
Pedestrian occupation	.879					
Injured organs		.908				
The leading cause of death		.906				
Type of vehicle involved			.802			
location of the death			.675			
Education				-.770		
Gender				.675		
Location of accident occurrence					.800	
Transmission of the injured					.418	
Season of the accident occurrence						.756
Darkness						.710
Rotation Method: Varimax with Kaiser Normalization.						

According to Table 2, the information in the data can be summarized in six components. These six components together explain 71.09% of the total variation of the data which includes an acceptable amount of total variance existed in the whole data.

According to Table 2, "age", "marital status" and "job" of the pedestrians killed in the fatal traffic accidents formed the first and the most important principle component which had the largest Eigenvalue (2.86) and consequently the largest amount of the explained variance (22.04%) (Table 3). We can name this component as "demographic and socioeconomic factor".

Table 3
Eigenvalues values and variance explained by each component

	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6
Eigenvalues	2.86	1.68	1.37	1.18	1.12	1.06
variance accounted for (VAF)	22.04	12.96	10.56	9.07	8.63	7.82
Cumulative VAF	22.04	35.00	45.56	54.63	63.26	71.09

About the hidden pattern and correlation of these variables with each other that appeared together in the first component and formed demographic and socioeconomic factor in fatal pedestrian data, the results (Table 4) showed that there were significant pairwise associations between all these three variables ($p\text{-value} < .001$). The pattern was that in pedestrians aged ≥ 30 years, in both single and married marital status, the majority of victims were self-employed. This rate in single pedestrians was 62.5% and in married pedestrians was 60.9%.

Table 4
Detected associations between variables

Occupation								
Marital status		Without any occupation (children)	Student	Unemployed	Government employee	Self-employed	Total	
single	Age <30	87(25.9%)	161(47.9%)	19(5.7%)	9(2.7%)	60(17.9%)	336(100%)	
	Age >= 30	0(0.0%)	2(2.1%)	30(31.3%)	4(4.2%)	60(62.5%)	432(100%)	
married	Age <30	0(0.0%)	3 (5.3%)	9 (15.8%)	4(7.0%)	41(71.9%)	57(100%)	
	Age >= 30	0(0.0%)	5(0.4%)	449(35.5%)	41(3.2%)	770(60.9%)	1265(100%)	
Component 1, p-value for all pairwise associations < .001*								
Pedestrian injured organs	The leading cause of death							
			Head Trauma	Bleeding	Multiple Fractures	Mixed causes	Others	total
	Head and neck		811(87.3%)	24(2.6%)	44(4.7%)	4(0.4%)	46(5.0%)	929(100%)
	Chest, abdomen, and pelvis		2(1.1%)	53(29.1%)	84(46.2%)	6(3.3%)	37(20.3%)	182(100%)
	More the one organ		222(34.7%)	30(4.7%)	154(24.1%)	19(3.0%)	214(33.5%)	639(100%)
	Other organs		2(7.4%)	5(18.5%)	11(40.7%)	6(22.2%)	3(11.1%)	27(100%)
	Component 2, p-value < .001*							
Kind of vehicle involved in the crash	location of the Pedestrian death							
			Death at the scene	Death during the transfer		Death at the hospital or Death after discharge		total
	Light machinery		486(38.1%)	111(8.7%)		678(53.2%)		1275(100%)
	Heavy machinery		150(55.6%)	25(9.3%)		95(35.2%)		270(100%)
	Motorcycle or bicycle		6(6.1%)	2(2.0%)		90(91.8%)		98(100%)
	Others		20(60.6%)	2(6.1%)		11(33.3%)		33(100%)
	Component 3, p-value < .001							
Pedestrian gender	Pedestrian education							
			Illiterate	Primary education	Guidance, High school and diploma	Academic education		total
	male		550(40.1%)	372(27.2%)	385(28.1%)	63(4.6%)		1370(100%)
	female		238(62.0%)	81(21.1%)	56(14.6%)	9(2.3%)		384(100%)
	Component 4, p-value < .001							
Location of accident occurrence	transmission of the injured pedestrian to the hospital							
			Ambulance		Other	total		
	Suburban roadways		875(89.8%)		99(10.2%)	974(100%)		

*by chi square test

Occupation						
Light condition	Urban roadways	52(82.4%)		111(17.6%)	631(100%)	
	Component 5, p-value < .001					
	Season of the accident occurrence					
		spring	summer	Autumn	winter	total
	Daylight	313(28.1%)	346(31.1%)	245(22.0%)	208(18.7%)	1112(100%)
	Twilight	22(13.8%)	26(16.3%)	67(41.9%)	45(28.1%)	160(100%)
	Night	87(21.6%)	116(28.9%)	116(28.9%)	83(20.6%)	402(100%)
Component 6, p-value < .001						
*by chi square test						

According to Tables 2 and 3, the second most important component with 12.96% of explained variance includes "the injured organ" and "leading cause of death" variables. We named this component as "injury".

About the hidden relationship and correlation between these two variables, which together have formed injury type component, there was a significant association between them (p-value < .001) and the results show that the rate of head trauma as a leading cause of death is higher in the pedestrians with head injury (87.3%) in compared to other injured organs that is obvious (Table 4).

The constituent variables of the third component were "kind of vehicle" and the "place of the death". This component with VAF 10.56% was in the third rank of importance. We can name this component as accident severity. The hidden relationship between these two variables was statistically significant (p-value < .001). The rate of death at the accident sense was higher in accidents with heavy vehicles (56.6%) compared with light vehicles (38.1%) (Table 4).

The fourth component was related to the variables of "education and gender of pedestrians". This component also explained 9.07% of the total variance of the data and we can say that this component is somehow related to the knowledge and observance of the traffic rules. The hidden relationship between these two variables was statistically significant (p-value < .001) and this showed that they were correlated to each other. About the hidden relationship and correlation of these two variables, which together have formed knowledge and observance of the traffic rules component, the result (Table 4) showed that 38.7% of female victims were literate while this rate in male victims was 60.5%.

The fifth component was related to the variables of "the location of the accident" and "mode of transferring injured pedestrian to the hospital". This component which explained 8.63% of the total variance of the data, was about the quality of medical relief.

About the hidden relationship and correlation of these two variables, which together have formed the quality of the medical relief component, the relationship between these two variables was significant (p-value < .001). The result (Table 4) showed that the rate of transferring injured pedestrians to hospital by ambulance in the suburban road (89.8%) was higher than urban roads (82.4%).

Finally, the last component has consisted of "season" and "lightness condition of the accident". This component with VAF = 7.82% was about environmental conditions. About the hidden pattern, the relationship between these two variables was statistically significant (p-value < .001). Most of the daylight pedestrian fatal accidents have been occurred in summer (31.1%) and spring (28.1%) seasons. Also, among fatal pedestrian accidents occurred in twilight hours, most of them were in autumn (41.9%) and winter (28.1%). About night accidents, there was no significant difference in the rate of fatal accidents between seasons (Table 4).

Discussion

This research was conducted to identify the structures of the traffic accidents that lead to pedestrian's death in East Azerbaijan during the years 2010–2019. With respect to the ordinal and nominal nature of our variables, CATPCA was the most appropriate method to bring out the pattern of this data and hidden relationships among variables. Identifying patterns in the data of the fatal pedestrian crashes and knowing the importance and priority of the factors affecting fatal traffic accidents could be effective to reduce accidents or the severity of injuries. Furthermore, using powerful statistical methods such as CATPCA in traffic data analysis, helps researchers to improve the statistical capacity of studies and accuracy and precision of their findings and bring out hidden aspects of the data.

According to the results of the descriptive pattern, most of the fatal pedestrian road traffic accident victims were male, elderly, self-employed, illiterate, and married. Also, fatal pedestrian accidents that occur in the summer were slightly more common than in other seasons. Furthermore, most of the fatal accidents occurred during the day, on suburban roads, and with light machinery. In most fatal accidents, the injured pedestrians were taken to hospital by ambulance. Also, in most of these fatal accidents, injured organs were the head and neck and the leading cause of death was head trauma. These findings were almost similar to other studies [13].

About the hidden pattern, all underlying variables were remained in CATPCA analysis with considerable loading values and conformed six influential principal components related to fatal pedestrian accidents which explain about 71% of the total variation of the data. This considerable amount of explained variance demonstrates the ability of this model in summarizing information of our high-dimension and complex data.

According to the unsupervised nature of our data, and lack of outcome variables, to assess and discuss influential factors of fatal pedestrian accidents, we can compare the fatal accident features distribution with the distribution of them in the target pedestrian population or pedestrian accidents population. Existing the difference shows that these features have some roles in occurring an accident or its severity or both of them.

The first and most important component contained the demographic and socioeconomic characteristics of the pedestrians (age, job, marriage). We can conclude that the features in the first component focus more on pre-crash characteristics which influencing the incidence and the severity of the crash.

Regarding the age itself, according to the studies, age has an undeniable effect on occurring an accident or its severity or both of them due to the lack of ability of the children and elderly age groups or the risk-taking nature of young age groups. It has been observed in a number of studies that age has a direct relation with the death of pedestrians, and older people are more likely to die[14–17]. In our study, we had an increasing trend so that with increasing age, the number of victims increased, and the majority of victims (35.13%) were over 65 years old.

Regarding occupation and marital status, we can say that these variables are effective in the traffic volume on the road. Pedestrians with a self-employed occupation were more likely to have fatal traffic accidents because of more traffic on the road[18]. Consistent with our results, a study on injured pedestrians showed that the majority of injured pedestrians were self-employed or workers[19]. Regarding pattern, although due to the nature of the marital status, age, and occupation, the existence of associations between these variables were expected, some parts of these associations were different from the usual population.

The pattern was that in pedestrians aged ≥ 30 years, in both single and married marital status, the majority of victims were self-employed. Although most Iranian people are self-employed, the observed percentage is higher than the self-employed people rate in the Azerbaijan population. It shows that being self-employed can lead to a higher chance of having fatal pedestrian traffic accidents may be due to the higher volume of road traffic.

The second-ranked component was about the injured organ (injured organ and the leading cause of death). This component is a post-crash factor related to the severity of the accident. There is no doubt about the injured organs that plays an important role in the death or survival of the person. In most studies, injury to the head was the first and most common reason for hospitalization, and lower limb organs were at the next rank[13, 19–21]. The leading cause of death can be considered as an indicator of the severity of the accident and has a high importance role in the outcomes related to driving accidents. About the hidden relationship, the results show that the rate of head trauma as the leading cause of death is higher in the pedestrians with a head injury that is obvious.

The third component was about the severity of the accident (kind of vehicle, place of death). With regard to vehicle type, in our study, the distribution of the kind of vehicle in fatal accidents is not similar to the distribution of it in non-fatal pedestrian accidents and the proportion of heavy vehicle in assessed fatal accidents is more than its proportion in the non-fatal pedestrian accidents [22]. Consequently, the proportion of light vehicle (76.1%) and motorcycle or bicycle (5.9%) is less than their proportions in non-fatal accidents[22]. So, the kind of vehicle is an influencing factor of both occurring and the severity of an accident. About occurring, it can be due to the higher number of light vehicles and therefore the higher volume of road traffic (which leads to a higher number of traffic accidents). About severity, if an accident occurs, the chance of death is higher for a heavy vehicle, and those who suffer from the crash with light vehicles are likely to survive more than those who are involved in a crash with heavier vehicles. In most studies, the impact of this variable on Traffic-related pedestrian death has been shown, so that heavy vehicles have increased the chance of death, because of the reasons such as the shape, mass, and design of the car that could increase released kinetic energy of the accident that is important in determining an injury severity[14, 15, 23–25]. Also, the studies have shown that crashes involving motorcycles were less likely to die[26, 27].

The location of the death can also be an indication of the severity of the incident which can occur at the accident scene or in the hospital or after discharge, depending on the severity of the incident. In our data, about half of injured pedestrians were died in the hospital or after discharging and about 40% died at the accident scene and others along the way to the hospital.

Between 2006 and 2012, 159227 people died in traffic accidents in Iran. A total of 97336 (61.13%) of them died inside hospitals or other health centers and 61891 (38.87%) died at the accident scene or during the transfer[28] that is consistent with our result.

About the hidden relationship and correlation between kind of vehicle and place of death, which together have formed severity of accident component, the result shows that the rate of death at the accident scene is higher in accidents with heavy vehicle (56.6%) compared with light vehicle (38.1%) that is because of injury severity. The fourth component was somehow related to the knowledge and observance of the traffic rules (gender, education).

Regarding gender, the risk of death in men and women has been reported differently according to the age groups or in-city or out-of-city events[26, 29, 30].

According to our results, the majority of the killed pedestrians are men (78.23%). Because the gender distribution of the under investigation data is not similar to its distribution in East Azerbaijan population (50% male, according to the last census) and also, is not similar to its distribution in non-fatal pedestrian accidents in East Azerbaijan and is similar to its proportion in the fatal accidents[22], consistent with other studies[15, 24, 31, 32] we can conclude that Gender is effective on both occurring and severity of the accidents. About occurring, the rate of traffic accidents in males is higher than females. This can be because of more road traffic volume in males than females (maybe because of the kind of their job) and lower observance of the traffic rules and risk taking behaviors in males than females.

It shows that, after the incidence of an accident, the severity of that accident can be higher for males than females may be because of their cross in an unauthorized hazardous location such as highways. High-speed collisions happen on highways (where the nature of the road allows motorists to reach higher speeds than inner-city roads) leads to higher severity accidents.

On the other hand, education has an important role in the occurring and severity of accidents. According to our results, most of the victims were illiterate (44.9%). This rate is completely different from the rate of illiteracy in East Azerbaijan (15%) according to the last census. so it can be an effective factor in occurring and also the severity of the accident. We can say maybe because of the higher awareness of higher educated people about pedestrian laws, as well as observing the rules that are higher in educated people, can prevent accident occurrence or causes an accident to occur mostly in authorized places and crosswalk that has fewer consequences than other unauthorized places. This finding is consistent with other studies[19]. In our date, the combination of these two risk factor (gender and education) lead to the groups of pedestrians that have the lowest and highest mortality rate. The majority of the victims were illiterate male pedestrians (31.4%) while the minority were females with an academic education (.5%). About the hidden relationship and correlation of these two variables, the result showed that 39.5% of male victims versus 61.3% of female victims were illiterate while the proportion of illiteracy in East Azerbaijan population is 10% of males versus 20% of females according to the last census.

So we can conclude that education is more effective in female's awareness and observance about traffic rules compared to males and we can say that higher educated women usually tend to act more lawfully and safely compared to other pedestrians.

The fifth component was about the quality of medical relief (place of accident, transferring by ambulance). According to our results, most of the fatal accidents occurred on suburban roads.

The Suburban incident is usually more fatal maybe because of the inability in providing relief to injured people as quickly as an urban accident. Similar studies have also shown that mortality rates in rural areas (less than 5,000 people) are higher due to the lack of medical facilities in these areas[26, 31, 33, 34].

Furthermore, high severity of injury due to the high-speed of vehicles on the out-of-town roads or higher traffic of the heavy vehicles on the roads outside of the city rather than inside could be the cause of high mortality rates in out-of-town events. This issue has been assessed in many studies and it has been shown that the chance of death in events occurring in the urban areas and high-density sites is lower than out-of-town roads due to the severity of the accidents[29, 31, 33, 35].

Regarding the transferring of the injured person, it is obvious that transferring an injured person to a hospital by ambulance can reduce fatalities. Between 2006 and 2012, 159227 people died in traffic accidents in Iran. A total of 97336 (61.13%) of them died inside hospitals or other health centers and 61891 (38.87%) died at the accident scene or during the transfer[28]. So, providing rapid and specialized assistance for the injured person by ambulances can decrease the hazard of death outcomes. In our data, the results show that the

majority of injured pedestrians have been transferred to hospital by ambulance (79.18%) which shows the appropriate medical relief facility. However, given that about 21% of victims have been transported by a vehicle other than an ambulance, reducing this number to zero can be effective in reducing pedestrian road traffic deaths.

But about the hidden relationship and correlation of place of accident and transfer by ambulance, which together appeared in the common component have formed quality of medical relief component, the result shows that the rate of transferring injured pedestrians to hospital by ambulance in the suburban road (89.8%) was higher than urban roads (82.4%). It shows that the rate of transferring to hospitals by other cars in inner-cities pedestrian accidents is more than out of cities may be because of existing near hospitals and existing other cars in the sense of accidents that this kind of transfer has increased the mortality rate of the accidents.

The sixth component was about environmental conditions (time of day and season of the accident). The season of the accident can also affect both the occurrence and severity of the accident. Some studies have shown the effect of season on the severity of injuries in traffic accidents due to slippery roads, weak visibility, and other adverse weather conditions[36–38].

The time of the accident (day or night) that could be related to the status of lighting and visibility, is the variable that plays an important role in the pedestrian traffic accident occurrence. In most studies, it has been observed that events occurring at night or with lower lighting conditions have a direct relationship with the severity of the incident, and in low light conditions, the severity of the accident rises[26, 27, 33, 35, 39, 40]. But in our study, results show that most of the fatal accidents have occurred in summer and daylight. The reason may be that the pedestrian's traffic on the roads is higher during the day and summer compared to the other seasons or night. Also, due to favorable weather and light conditions, drivers usually drive at high speed and less carefully, so it is less possible to control the speed of the car to prevent accidents. About the hidden pattern, the results suggest a strong association between the outcome of traffic accidents and environmental conditions. Most of the daylight pedestrian fatal accidents have occurred in the summer and spring seasons. Also, among fatal pedestrian accidents occurred in twilight hours, most of them were in autumn. About night accidents, there was no significant difference between seasons.

It can be due to the more pedestrian volumes during the day in the summer season and also the higher speed of cars in this season due to the favorable weather condition.

About the twilight hours and autumn season may be because of changes in day duration, light, and weather conditions the risk of accidents in twilight times of autumn gets higher.

According to the findings in this study, it seems that to decrease the mortality rate of the pedestrians, at first policymakers and health managers should focus to control factors that are the cause of the crash incidence. Some personal characteristics such as age, sex, job, education, and marriage can affect the occurring and severity of an accident. It seems that high-risk pedestrians such as children and elderly people need more attention and care from their family and the community to avoid a car accident. Also, the awareness and observance of the people about traffic safety and rules can decrease the accident incidence rate and its severity.

It can be possible to prevent occurring an accident by training the citizens to increase their traffic knowledge and observance of the traffic rules. Moreover, improving the lightness of the road, using technology, or some new equipment for drivers could be effective to avoid occurring an accident. After occurring an accident, we can decrease the severity of crashes by improving car body design and safety features of them, especially in heavy vehicles. Also, relief and emergency health services are so important to reduce the mortality rate.

On the other hands, according to the results of hidden pattern and relationships among variables, we can say that in addition to the main effects of risk factors on occurring and severity of the accident, there are some combinations of features such as elderly self-employed pedestrians or illiterate males can increase the risk of occurring or severity of an accident. So, by identifying such combinations, we can focus on more details on such certain groups of pedestrians not all people to have more effective for prevention intervention programs. Also, accidents that involved vehicles were heavy and the injured organ was head, were more fatal. Furthermore, not transferring injured pedestrians by ambulance in urban areas, daylight of summer, and twilights of autumn increased the risk of fatality accidents.

So, improving the lightness of the road, improve relief facilities, improving car body design can

Limitation

The limitations of this study were the low number of variables collected and the lack of links among other relevant databases, such as the police or the hospital databases to collect or complete the related variables more precisely.

Conclusions

CATPCA revealed some hidden patterns and relationships between the features of the fatal pedestrian traffic accident data in the format of six components. It shows that some combinations of features can increase the risk of having a fatal accident.

Thus, identifying and controlling such hidden patterns according to their importance could be effective to reduce the occurrence and severity of pedestrian traffic accidents in East Azerbaijan province of Iran. Following this paper, it is suggested that similar studies be performed on pedestrian accident data of different places to compare their patterns.

List Of Abbreviations

FPRTA: Fatal Pedestrian Road Traffic Accident

CATPCA: Categorical Principal Component Analysis

PCA: Principal Component Analysis

VAF: variance accounted for

Declarations

Ethics approval and consent to participate

The study received ethical approval from the Ethics Committee of Tabriz University of Medical. The data used in this study is a part of registered data of the forensic organization. We got the data of this study from forensics organization. So, we had no direct contact with relatives of dead pedestrian. This organization has obtained informed consent from Next of kin/Legally authorized representatives of dead participants to use their information. Explanations of this data are provided in other articles[10]. So, we confirm that informed consent from Next of kin/Legally authorized representatives of dead participants whose data is used in the study has been taken by forensic organization.

Consent for publication

The authors have agreed on the content of the manuscript.

Availability of data and materials

The datasets analyzed during the current study are not publicly available due used data was registered in the forensic organization and they do not belong to the researchers of this study but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors are grateful to the Research Vice-Chancellor of Tabriz University of Medical Sciences for financing the project.

Authors' contributions

MJ and PS were responsible for the study design. MJ and PS did the analyses. PS and MJ were responsible for data interpretation. MJ helped in the drafting of the manuscript. PS critically reviewed the manuscript and provided the final manuscript. HS helped in the study design and data collection. All authors have read and approved the final manuscript.

we confirm that all methods were carried out in accordance with relevant guidelines and regulations in the manuscript.

Acknowledgements

We appreciate the police and the forensic organization because of the availability of the data. This article has been extracted from the thesis submitted for MSc degree in Biostatistics which has been approved by the ethics committee of Tabriz University of Medical Sciences (Ethics number: IR.TBZMED.REC.1397.081).

References

1. Organization WH: **Global status report on road safety 2018 (2018)**. Geneva, Switzerland, WHO 2019.
2. Naci H, Chisholm D, Baker TD: **Distribution of road traffic deaths by road user group: a global comparison**. *Injury prevention* 2009, **15**(1):55–59.
3. Asiamah G, Mock C, Blantari J: **Understanding the knowledge and attitudes of commercial drivers in Ghana regarding alcohol impaired driving**. *Injury prevention* 2002, **8**(1):53–56.
4. Nantulya VM, Reich MR: **The neglected epidemic: road traffic injuries in developing countries**. *Bmj* 2002, **324**(7346):1139–1141.
5. Mock CN, Gloyd S, Adjei S, Acheampong F, Gish O: **Economic consequences of injury and resulting family coping strategies in Ghana**. *Accident Analysis & Prevention* 2003, **35**(1):81–90.
6. Mock C, Kobusingye O, Anh LV, Afukaar F, Arreola-Risa C: **Human resources for the control of road traffic injury**. *Bulletin of the world Health Organization* 2005, **83**:294–300.
7. Mabunda MM, Swart L-A, Seedat M: **Magnitude and categories of pedestrian fatalities in South Africa**. *Accident Analysis & Prevention* 2008, **40**(2):586–593.
8. Mohan D, Tsimhoni O, Sivak M, Flannagan MJ: **Road safety in India: challenges and opportunities**. In.: University of Michigan, Ann Arbor, Transportation Research Institute; 2009.
9. Linting M, van der Kooij A: **Nonlinear principal components analysis with CATPCA: a tutorial**. *Journal of personality assessment* 2012, **94**(1):12–25.
10. Samadirad B, Khodadoost M, Sadeghi-Bazargani H, Shahedifar N, Heydari ST, Gholamzadeh S: **Epidemiology of Fatal traffic injuries registered in east Azerbaijan and Fars forensic medicine organizations: the research protocol**. *Journal of Clinical Research & Governance* 2016, **5**(1).
11. Linting M, Meulman JJ, Groenen PJ, van der Kooij AJ: **Nonlinear principal components analysis: introduction and application**. *Psychological methods* 2007, **12**(3):336.
12. Wold S, Esbensen K, Geladi P: **Principal component analysis**. *Chemometrics and intelligent laboratory systems* 1987, **2**(1–3):37–52.
13. Haleem K, Alluri P, Gan A: **Analyzing pedestrian crash injury severity at signalized and non-signalized locations**. *Accident Analysis & Prevention* 2015, **81**:14–23.
14. Ma Z, Lu X, Chien SI-J, Hu D: **Investigating factors influencing pedestrian injury severity at intersections**. *Traffic injury prevention* 2018, **19**(2):159–164.
15. Li D, Ranjitkar P, Zhao Y, Yi H, Rashidi S: **Analyzing pedestrian crash injury severity under different weather conditions**. *Traffic injury prevention* 2017, **18**(4):427–430.
16. Olszewski P, Osińska B, Zielińska A: **Pedestrian safety at traffic signals in Warsaw**. *Transportation research procedia* 2016, **14**:1174–1182.
17. Kebria SE, Soori H: **Study of epidemiological pattern of pedestrian's road traffic injuries in 2014 and determination of related risk factors on severity of injury**. *Payesh (Health Monitor)* 2017, **16**(3):293–302.
18. Ghafari Fam S, Sadeghi Bazargani H, Salarilak S: **A survey on epidemiological and demographic aspects of pedestrian road accident victims in Shohada referral university hospital, East Azerbaijan province**. *The Journal of Urmia University of Medical Sciences* 2014, **25**(7):549–560.
19. Hatamabadi H, Vafaei R, Haddadi M, Abdolvand A, Esnaashari H, Soori H: **Epidemiologic study of road traffic injuries by road user type characteristics and road environment in Iran: a community-based approach**. *Traffic injury prevention* 2012, **13**(1):61–64.
20. Peymani P, Heydari ST, Hoseinzadeh A, Sarikhani Y, Moafian G, Aghabeigi MR, Maharlouei N, Foroutan A, Ahmadi SM, Ghaffarpasand F: **Epidemiological characteristics of fatal pedestrian accidents in Fars Province of Iran: a community-based survey**. *Chinese journal of traumatology* 2012, **15**(5):279–283.
21. Zangooei Dovom H, Shafahi Y, Zangooei Dovom M: **Fatal accident distribution by age, gender and head injury, and death probability at accident scene in Mashhad, Iran, 2006–2009**. *International journal of injury control and safety promotion* 2013, **20**(2):121–133.
22. Jamali-Dolatabad M, Sadeghi-Bazargani H, Sarbakhsh P: **Predictors of fatal outcomes in pedestrian accidents in Tabriz Metropolis of Iran: Application of PLS-DA method**. *Traffic injury prevention* 2019, **20**(8):873–879.
23. Matsui Y, Oikawa S, Sorimachi K, Imanishi A, Fujimura T: **Association of impact velocity with risks of serious injuries and fatalities to pedestrians in commercial truck-pedestrian accidents**. In.: SAE Technical Paper; 2016.

24. Verzosa N, Miles R: **Severity of road crashes involving pedestrians in Metro Manila, Philippines.** *Accident Analysis & Prevention* 2016, **94**:216–226.
25. Aziz HA, Ukkusuri SV, Hasan S: **Exploring the determinants of pedestrian–vehicle crash severity in New York City.** *Accident Analysis & Prevention* 2013, **50**:1298–1309.
26. Tay R, Choi J, Kattan L, Khan A: **A multinomial logit model of pedestrian–vehicle crash severity.** *International journal of sustainable transportation* 2011, **5**(4):233–249.
27. Zhang G, Yau KK, Zhang X: **Analyzing fault and severity in pedestrian–motor vehicle accidents in China.** *Accident Analysis & Prevention* 2014, **73**:141–150.
28. Rahmani K, HashemiNazari S, Ghadirzadeh M: **Trend Analysis of Traffic Accidents Deaths in Iran During 2006–2012: Hospital or Pre-Hospital Occurred Deaths.** *Journal of Rafsanjan University of Medical Sciences* 2016, **15**(2):115–128.
29. Moudon AV, Lin L, Jiao J, Hurvitz P, Reeves P: **The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington.** *Accident; analysis and prevention* 2011, **43**(1):11–24.
30. Oikawa S, Matsui Y, Doi T, Sakurai T: **Relation between vehicle travel velocity and pedestrian injury risk in different age groups for the design of a pedestrian detection system.** *Safety Science* 2016, **82**:361–367.
31. Islam S, Hossain AB: **Comparative Analysis of Injury Severity Resulting from Pedestrian–Motor Vehicle and Bicycle-Motor Vehicle Crashes on Roadways in Alabama.** *Transportation Research Record* 2015, **2514**(1):79–87.
32. Olszewski P, Szagała P, Wolański M, Zielinska A: **Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland.** *Accident Analysis & Prevention* 2015, **84**:83–91.
33. Pour-Rouholamin M, Zhou H: **Investigating the risk factors associated with pedestrian injury severity in Illinois.** *Journal of safety research* 2016, **57**:9–17.
34. Islam S, Jones SL: **Pedestrian at-fault crashes on rural and urban roadways in Alabama.** *Accident; analysis and prevention* 2014, **72**:267–276.
35. Mohamed MG, Saunier N, Miranda-Moreno LF, Ukkusuri SV: **A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada.** *Safety Science* 2013, **54**:27–37.
36. Zhang G, Yau KKW, Chen G: **Risk factors associated with traffic violations and accident severity in China.** *Accident; analysis and prevention* 2013, **59**:18–25.
37. Kashani AT, Mohaymany AS: **Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models.** *Safety Science* 2011, **49**(10):1314–1320.
38. Sun Z, Wang J, Chen Y, Lu H: **Influence Factors on Injury Severity of Traffic Accidents and Differences in Urban Functional Zones: The Empirical Analysis of Beijing.** In: *Int J Environ Res Public Health*. vol. 15; 2018.
39. Coate D, Markowitz S: **The effects of daylight and daylight saving time on US pedestrian fatalities and motor vehicle occupant fatalities.** *Accident; analysis and prevention* 2004, **36**(3):351–357.
40. Jang K, Park SH, Kang S, Song KH, Kang S, Chung S: **Evaluation of Pedestrian Safety: Pedestrian Crash Hot Spots and Risk Factors for Injury Severity.** *Transportation Research Record* 2013, **2393**(1):104–116.

Figures

Figure 1
the scree plot

Figure 1
the scree plot