52. Porkaasgari, Ahmad, Ajam, Ahmad and Maryam (2010), Investigating the relationship between the quality of accruals and the lack of information symmetry in listed companies in Tehran Stock Exchange, Tehran Stock Exchange Journal, issue 11, pp. 12-4-107.
53. Qodratollah, Imam Verdi, Karimi, Mojtaba, Najarzadeh, Negin and Shah Karim Oghli, Masoumeh (2011), Effect of Electronic Commerce and ICT on Economic Growth in the G8 Countries and Selected Countries in the Islamic Conference Group.
54. Raei, Reza and Ahmad Pouyanfar (2012), Advanced Investment Management, Tehran, SAMT Publication.
55. Rajan, R.G. and Zingales, D. (1998a). Financial dependence and growth. American Economic Review, 88(3), 559-86.
56. Robert Solow , "A Contribution to the Theory of Economic Growth". Quarterly Journal of Economics, 1956, 70(1)
57. Ronald I.McKinnon , Money and Capital in Economic Development, Washington , DC: Brookings Institution,1973.
58. Sakaran, Uma (2005), Research Methods in Management, Translated by Mohammad Saee and Mohammad Shirazi, Management Training Center Publication.
59. Samadi, Saeed, Khadijeh Nasrallahi and Morteza Karamalian Sichani (2007), The Relationship between Development of Financial Markets and Economic Growth, Quarterly Journal of Economic Research, Vol. 6, No. 3, Autumn 2007, pp. 16-1.
60. Sanginjan, Ali, (2004), The stock Market in Iran, Challenges and Prospects; Proceedings of the Exchange Conference, The Forgotten Opportunity of the Iranian Economy, Faculty of Economics, Allameh Tabatabaei University.
61. Sarmad, Zohreh, Abbas Bazargan and Elaheh Hejazi (2012), Research Methods in Behavioral Sciences, Tehran, Aagah Publication.
62. Schumpeter, J.(1911)The Theory of Economic Development ,Translated by Redvers Opie, Cambridge , Mass.:Harvard University Press, 1934.
63. Shaw,E. (1973), Financial deepening in economic development, New York: Oxford University Press.
64. Sherbaaf, Samira, Sayyed Kamil Tayyibi and Mostafa Rajabi (2013), Foreign Direct Investment, Developing a Financial Market for Economic Growth in Selected meadle east Countries, The First National Electronic Conference on the Perspective of Iran's Economy with the Approach to Supporting National Production, December 28, 2013.
65. Stiglitz, J. and A. Weiss (1983) Incentive Effects of Terminations: Applications to Credit and Labor Markets; American Economic Review, 73(5): 912-927.
66. Tehran Stock Exchange, Transparent Trading Networks, Deputy Economic Studies and Market Development, November 2009.
67. Wolf, Susanna .(2001). Developments and Imapct of ICT use for African SMEs: Implications for rural South Africa, Paper prepared for TIPS Forum.
68. Zamanzadeh, Hamid (2012), Five Decades of Production Ups and Dows (A Study of the Causes of a Low and Unstable Economic Growth in Iran), Journal of Economics, Summer 2012, No. 136, pp. 204-202.

EFFECT OF TOBACCO SMOKING ON DRIVING ACCIDENTS BY USING CLUSTER ANALYSIS

Ahmet Tortum,
Ataturk University department Transportation of Faculty Engineering Civil Engineering
Hojat Nikakhtar,
Ph.D. student Ataturk University of transportation
Civil Engineering of Faculty Engineering
Hamid Shirmohammadi,
Urmia University

Abstract. The aim of the study is to investigate the influence of tobacco smoking as a risky behavior in driving accidents. In this study, the relationship between consumption and non-consumption of tobacco with the behavior and the information on vehicle drivers from the aspect of traffic safety has been investigated. This relationship was confirmed as a hypothesis using different statistical tests of Chi-square, Mann–Whitney U, and Kruskal–Wallis. Our review has shown that in these tests, there is a high probability of a significant relationship between tobacco smoking and the duality of driving accidents and traffic safety. The traffic safety parameters have been considered with accident status, accident type, average distance, and driving time, brake time length, brake distances, length of overtaking, stop length, maximum speed, and the high bright beam of the car's headlights. A cluster analysis has been used to classify drivers as per this relationship, which resulted in the introduction of three groups of drivers: the first cluster include non-smoker drivers who have a very low accident rate, the second cluster include drivers who do not smoke while driving, and the third cluster include drivers who tend to smoke and their car accident rate is higher. For non-smokers of the first and the second clusters, the probability of an accident type comprises collision with another vehicle, and for drivers who do smoke while driving, probability includes other types besides the following: collision with another vehicle, pedestrian accident, overtaking or exit from the road, and collision with a fixed object.

Keywords: tobacco smoking; driving accidents; Chi-square; Mann–Whitney U; Kruskal–Wallis and clustering analysis

In this study, by a questionnaire directly asked by drivers about smoking in driving?

Introduction. Tobacco smoking while driving is one of the human factors that causes driving accidents. There is a lot of evidence to show that smoking while driving has a negative impact on traffic safety. The most obvious is the distraction caused with the lighting the cigarette while driving. Besides the act of cigarette smoking itself causing distraction, there is the risk of the cigarette ash falling on the driver's body, clothes or vicinity, causing a potential reaction or burn, thus creating a risk factor during driving. In this research, by a questionnaire directly asked for drivers, about smoking in driving? (Q 25) The research was conducted in 20 provinces of Iran. The period of presentation and receipt of the first six months in 2016. The research community of people with driving license were over 18 years old. The method of selecting a statistical is randomly simple compared to the population living in each provinc. Smoking as an independent variable has been assumed and the dependent variables have been in the following order, which have been categorized from one to five levels as per the drivers' response. These qu estions are mainly based on the traffic safety information for drivers and the same questions and options have been given in the Table (2-2):

- Exposure by accident? (Q 10)
- Type of accident? (Q 46)
- Average driving distance per day? (Q 14)
- Average driving time in one day? (Q 11)
- Time taken for braking on encountering a danger? (Q 47)
- When driving at 90 Km/h speed and you encounter a dangerous situation and brake, how do you estimate the distance needed to bring the car to a halt? (Q 48)
- You are on the road at 90 Km/h, to overtake a vehicle that is moving at a speed of 80 Km/h, as estimated how many meters do you need along the road, generally? (Q 51)
- When driving at a speed of 50 km/h, how many meters distance do you need for braking to prevent the collision of your car with the next vehicle (in normal weather conditions)? (Q 49)
- Do you know your car's speed is higher than the acceptable speed limit for driving inter-city roads in the day? (Q 50)
- The headlamp light of your car "turns how many meters, approximately? (Q 52)

In the following, the statistical tests of Chi-square, Kruskal–Wallis, and Mann–Whitney U test have been used in statistical analysis of the obtained data and to describe the frequency and histogram. Finally, cluster analysis has been used as a branch of pattern recognition and artificial intelligence has been used in categorizing variables. For researchers and users of clustering, and in a more general form, segmentation and clustering are not the end goal, but can also aid as the starting point on other tasks. For example, by categorizing smokers (variables) according to traffic safety information for drivers, one can plan for traffic control affairs, vehicle design, road design and service, and servicing as more specialized subjects. This paper attempts to use how clusters communicate together in order to extract theories from it. It is clear that these relationships and theories can be exploited in practical applications such as planning, management, and traffic engineering.

Research background. Reason et al. (1990) divides risky driving behaviors into four types: errors, common violations, aggressive violations, and landslides [1]. A driving error is the result of failure in planned driving practice, to achieve its goal. The errors are the result of a failure in the judiciary or inferential process involved in choosing the goal, either in the path to the goal or both. Errors greatly increase the potential for accidents. Violations are a deliberate defiance of behaviors that are necessary for safe driving (such as exceeding the speed limit or lack of maintaining an adequate distance from other vehicles). An aggressive violation is the deliberate defiance of the rules or targeting another individual [2].Landslides are an unwanted deviation from the action that they intend to do. These behaviors will often not have a detrimental effect on other road users (for example, turning on the car's light instead of switching on the windscreen wiper). Research shows that a driver who exhibits risky driving is more likely to be involved in an accident. Also, various studies show that the type of risky driving behaviors varies from one group to another. For example, men commit more violations (not the slip and mistake type) than women [3], older drivers contributed in accidents related to errors and slips [4], violations and mistakes are more relevant to accidents. [5]

Research Methodology. Questionnaire. A self-explanatory questionnaire was prepared to assess the structure of drivers' behavior formation. For validation, this questionnaire was provided to experienced traffic experts to match the results published in prestigious journals and their executive experiences. These specialists included university professors and top-level military officers. A preliminary questionnaire was used to review and make necessary corrections for the validity of the questionnaire.

The questionnaire consists of 58 questions, which can be categorized into five sections:

1. Specifications of drivers such as place of residence, gender, marriage status, age, education, occupation, income, and driving record;

2. Specifications of the vehicle used, such as vehicle type, vehicle facilities such as fire extinguisher, first aid box, ABS system, ice-breaker, and air bag, regular and standard services in main dealership and pre-travel service;

3. Traffic safety information for drivers such as accident situation, crash type, mileage, driving time per day, brake time and brake distance in the face of danger, required span for overtaking, high speed on the road, and beam range of high-intensity headlights;
4. The habits of the driver such as smoking and alcohol, fasten seat belt , pleasure derived from high speed, observance of traffic regulations, behavior in difficult situations, using the mobile phone while driving, the amount of sleep at night, observance of the rules of overtaking, respect for pedestrians crossing, accidental measures, and driving habits according to the area of the road;
5. The driver's tendency to be at risk or risk-taking, behavior in crossing traffic light, overtaking on downhill, attending traffic signs, and traffic regulations.
The questions raised with the subject of paragraph 5 above in Table 3 are exactly in line with the questions presented with statistics. These questions are the basis for risk-taking behavior and driving risk appetite.
Questioning and statistical sample. In this study, the statistical sample was selected from citizens of centers in 20 provinces of Iran. These cities include Tabriz, Urmia, Isfahan, Karaj, Tehran, Mashhad, Ahvaz, Zanjan, Semnan, Shiraz, Qazvin, Qom, Sanandaj, Kermanshah, Gorgan, Rasht, Khorramabad, Sari, Hamedan, and Yazd. A total of 20000 people in each /4112 questionnaires were completed by citizens with a driving license, with the proportion of 1 province in these cities. Among them, 3709 were men and 403 women, 2758 were married, 1110 were single, and 244 were divorced, other specifications have been given in Table 1-2 of the drivers' specifications.

Table1-2: Descriptive statistics of driver specifications questionnaire data

No	driver specifications	Description and number of levels	Level Values	Level Percentage
1	Gender (Q3)	1. Male	3709	90.20
		2. Female	403	9.80
2	Marriage Status (Q2)	1. Married	2758	67.07
		2. Single	1110	26.99
		3. Divorced	244	5.93
3	Age (Q4)	1. 18-24 years	746	18.14
		2. 25-34 years	1770	43.04
		3. 35-44 years	1428	34.73
		4. 45-54 years	130	3.16
		5. Over 55 years	38	0.92
4	Education Level (Q5)	1. illiterate	333	8.10
		2. Elementary	1480	35.99
		3. High school	978	23.78
		4. Diploma	889	21.62
		5. University	432	10.51
5	Occupation (Q6)	1. Guilds and merchants	501	12.18
		2. High income jobs	410	9.97
		3. Employee	700	17.02
		4. Driver	772	18.77
		5. Student	294	7.15
		6. Retired	442	10.75
		7. Unemployed	606	14.74
		8. Other	387	9.41
6	Income (Q7)	1. 0	503	12.23
		2. Less than 300 $	1904	46.30
		3. 300-600 $	1136	27.63
		4. 600-1500 $	330	8.03
		5. Over 1500 $	239	5.81
7	Driving Record (Q9)	1. Less than 1 year	371	9.02
		2. 1-5 years	939	22.84
		3. 6-10 years	1243	30.23
		4. 11-20 years	1084	26.36
		5. Over 20 years	475	11.55

Information obtained with regards to the traffic safety of drivers has been included in Table 2-2. The sequential independent variables from responses gained from the 10 questions that have been presented in five sequences. The

number of each answer as the values of the levels and percentages for each one has been presented in columns 4 and 5, respectively.

Table 2-2: Traffic Safety Information for Drivers				
Row	**Drivers' behavior and information for traffic safety**	**Number and description of levels**	**Levels values**	**Percentage of levels**
1	**Situation of encounter with accident? (Q10)**	**1. Seven times and above**	**493**	**12**
		2. 5 to 6 times	**1071**	**26**
		3. 3-4 times	**947**	**23**
		4. 1 to 2 times	**862**	**21**
		5. I have not accidented	**738**	**18**
2	**Type of accident? (Q46)**	**1. collision with another vehicle**	**1648**	**40**
		2. collision with pedestrian	**659**	**16**
		3. Overturning or leaving the road	**698**	**17**
		4. Collision with a fixed object	**370**	**9**
		5. Other types	**738**	**18**
3	**Average driving distance per day? (Q14)**	**1. Less than 20 km**	**1113**	**27**
		2.20-50 km	**1275**	**31**
		3. 50-150 km	**1190**	**29**
		4.150-300 km	**328**	**8**
		5. More than 300	**206**	**5**
4	**Average driving time per day? (Q11)**	**1.Less than 5 hours**	**1072**	**26**
		2. 5-7 hours	**1399**	**34**
		3. 8-10 hours	**1107**	**27**
		4. 11-13 hours	**350**	**9**
		5. More than 13 hours	**184**	**4**
5	**Braking time in encountering by danger? (Q47)**	**1. Less than one second**	**1234**	**30**
		2.1-1 second	**1892**	**46**
		3. 2-2 seconds	**698**	**17**
		4. 3-4 seconds	**205**	**5**
		5. More than 4 seconds	**83**	**2**
6	**When driving at 90 Km / h speed in encounter a dangerous situation and brake, how do you think about distance of stopping your car? (Q48)**	**1. Less than 25 meters**	**945**	**23**
		2. Between 25 and 50 meters	**1975**	**48**
		3. between 50 to 70 meters	**741**	**18**
		4. between 70 to 90 meters	**287**	**7**
		5. More than 90 meters	**164**	**40**
7	**You are on the road at 90 Km / h to overtake a device that is moving at a speed of 80 Km / h. as estimated how many meters do you need along the road, generally? (Q51)**	**1- Less than 200 meters**	**1317**	**32**
		2- between 200 and 300 m	**1559**	**37.9**
		3- between 300 and 450 m	**578**	**14.1**
		4- between 450 and 550 m	**371**	**9**
		5. More than 550 meters	**287**	**7**
8	**When driving at a speed of 50 km / h and braking the next vehicle, How many meters distance do you need to praccident the collision of your car with the next vehicle?(in**	**1- Less than 5 meters**	**740**	**18**
		2- Between 5 and 15 meters	**1438**	**35**
		3- between 15 and 25 meters	**1380**	**36**
		4- Between 25 and 30 meters	**289**	**7**

	normal weather conditions.)? (Q49)	**5. More than 30 meters**	**165**	**4**
9	**You know your car's speed higher than what speed as high speed in driving inter-city roads in the day? (Q50)**	**1. Between 50 and 60 km/h**	**82**	**2**
		2. Between 60 and 75 km / h	**287**	**7**
		3. Between 75 and 90 km/h	**740**	**18**
		4. between 90 and 110 km/h	**1316**	**32**
		5. More than 110 km/h	**1687**	**41**
10	**The light of headlamp of your car turns how many meters, approximately? (Q52)**	**1. Less than 50 meters**	**1027**	**25**
		2. between 50 and 70 meters	**1233**	**30**
		3. Between 70 and 90 meters	**616**	**15**
		4. between 90 and 100 meters	**824**	**20**
		5. More than 100 meters	**412**	**10**

Tobacco consumption during driving as a dependent variable from the statistical population of the question and results of the number of consumers and their percentage has been listed in Table 2-3.

Table 2-3: Statistics of smoking while driving

Row	Drivers' behavior and information for traffic safety	Number and description of levels	Levels values	Percentage of levels
11	Do you smoke while driving ? (Q25)	1.No	2471	60.09
		2.Yes	1641	39.91

Theory of descriptive analysis of data. Frequency of variables. A frequency table has been created as per the variables, and indicates the number of people who have chosen a certain response. These frequencies have been shown in (2-2) and (2-3). The histogram chart has been created to display the frequency of data, (as in column 6 of Table 1-4). In these charts, the distribution of the variable has been compared with the normal distribution and a normal curve has been plotted on the histogram. The variable data is aggregated and divided by the number of observations. The mean of the variable value in the data group has been obtained.

Kruskal–Wallis Test. The Kruskal–Wallis test compares the mean values with each other and is a series of tests of variance analysis. In this test, more than three groups are compared. The hypothesis is that the k sample groups are extracted from a statistical population and their means are compared. In this way, using ratings, one can decide if the k sample group is independent of the statistical population or if they have come from a different statistical population? It is natural that there are differences between them. However, through this test, whether the differences observed in the samples represent differences in the population or are due to chance and accident is determined. According to the means, the likeness of k-sample has been considered as of a common basic population and two hypotheses are planned as null and one:

$_0 : \mu_1 = \mu_2 = ... = \mu$ $_1 : \mu \neq \mu$ $\neq$

In this test, k samples are added to each other for N observations. Next, for each N, a rank is extracted. The lowest rank is one, and the highest is N. Then for each of the k groups, the sum of the ranks is calculated. This test determines that the sum of ranks are so different that we cannot determine if they have been extracted from a common statistical population. [7] The Kruskal–Wallis statistical index is obtained from the following equation:

$$= \frac{12}{(\ + 1)} \sum_{=1} \frac{^2}{} - 3(\ + 1)$$

In this relation:

k = number of groups, nj = number of individuals in each group, N = total number of individuals in all groups, Rj = sum of ranks in each group

3-3 Man–Whitney U-test

The application of this test is to make a comparison on the basis of qualitative variables in order to rank the variables. This test is a nonparametric test that has no sensitivity to population parameters such as variance. It is used to determine the difference between the two populations using random samples that have been selected from the same population, and define if the measurement scale of the dependent variable studied is in a sequential order. In fact, the goal is to find out if the two relevant populations have a meaningful difference in terms of focusing on each other or if there is a difference between chance and accident? Thus, the size of the two samples, k1 and k2, are ranked in an ascending order, or vice versa, regardless of which variable belongs to which group. After ranking, the total rank of

each group is calculated separately. The total rank of the first group is R1 and the sum of the rankings of the larger group is R2. [8]

$$= {}_1 \cdot {}_2 + \frac{{}_1({}_1 + 1)}{2} - {}_1$$
$$' = {}_1 \cdot {}_2 + \frac{{}_2({}_2 + 1)}{2} - {}_2$$
$$+ ' = {}_1 + {}_2$$

For samples larger than 20, in accordance with the following formula, Z (normal variable) is replaced:

and mean Standard deviation $= \frac{| - |}{}$

3-4.Chi-square test

In general, the impact of tobacco consumption on traffic accidents and traffic safety has been studied considerably. Here, the variable, i.e., the consumption or non- consumption of cigarettes with the subject of accident and traffic safety of drivers, is examined in a two by two table. The goal is to determine if there is a relationship between the two criteria or are independent of each other. In order to obtain the chi-square statistics, the need to know the expected frequency is as follows:

$$= \frac{\cdot \cdot}{}$$

Where the $\cdot$ is the sum of the observed frequencies in i-th row and $\cdot$ is the sum of the observed frequencies in the j-th column, and $$ is the sum of the total abundances.

$$ = \sum \sum$$

The Chi-square test is also called the Goodness of Fit Test.

$$^2 = \sum_{=1} \frac{(-)^2}{}$$

A hierarchical cluster analysis

Cluster analysis is a descriptive data analysis tool for organizing observed data into meaningful categories or clusters based on family information that maximizes the similarity of cases within each cluster. In summary, the statistical methodology of cluster analysis involves participating in a homogeneous class for creating a practical classification.

In this method, a two-step arrangement is created as follows:

1. Using Ward's method, which selects the squared Euclidean Distance as the initial cluster distance between two object points. Hierarchical analysis helps to determine the number of clusters that we need to work with.
2. The next step is the transformation of a hierarchical cluster analysis with a suitable number of cluster selections, in which each item in our cluster is assigned a special cluster.

Hierarchical cluster analysis is a major statistical method for finding relatively homogeneous clusters of cases that are based on measured characteristics. This analysis begins with each case as a separate cluster, for example, there is a cluster for each person's response regarding the dependence of cigarette smoking on traffic safety. It then combines the clusters in sequence, and reduces the number of clusters in each step, until only one cluster remains. The clustering method uses dissimilarity or spacing between objects during the formation of clusters. The most common is the Euclidean distance, which is as follows:

$$(,) = \sqrt{\sum_{=1} (-)^2}$$

Where in $(,)$ is the Euclidean distance between two points *X* and *Y* in a space with d dimension (index). Euclidean intervals are the starting point for many clustering methods. The cluster analysis methods used on the same data set often result in diverse answers. In fact, many methods are only suitable when clusters are almost of a spherical form [9, 10]

Research findings

Statistical tests Results

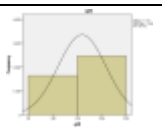
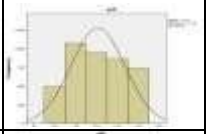
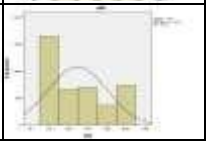
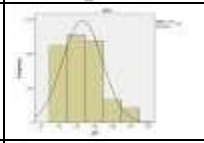
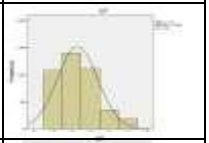
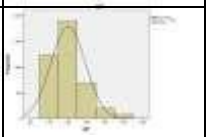
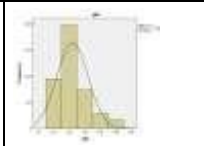
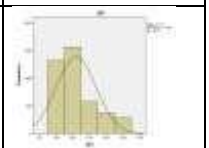
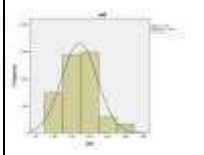
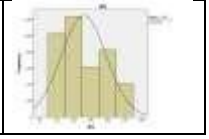
In Table 4-1, the results of the two-variable Chi-square test is shown. Since the results of the survey show the relationship between tobacco consumption and the ranking variables of traffic safety information with a significant level of less than 0.01, the research hypothesis has been confirmed.

The Kruskal–Wallis test is a nonparametric test that evaluates the ranking variables to determine if the average ranking difference is significant or not. This test demonstrates the importance of the difference in average rankings, which is then followed by the Chi-square test formula. As shown in Table 4-1, the two by two relationship between cigarette consumption and traffic safety items is found in all cases, with an Asymp.Sig. value of 0.000, which means that the assumption of a lack of relationship between them is rejected. Furthermore, this test also confirms that there is no significant difference between cigarette consumption and any of the items in the first column of Table 4-1 (Traffic Safety).

variable is nominal and the The Mann–Whitney U test is used for sequential statistical data. When the independent dependent variable is ranked, this nonparametric test is used to distinguish between two populations using random samples which are used by the same population In fact, the goal is to determine whether the two relevant societies have a meaningful difference in terms of their focus or if the differences have arisen by chance or accident. The results of the Mann–Whitney U test have been given in Table 4.1 for the absolute Z value. In each test, the absolute value of Z score has been greater than 1.96.The assumption of the relationship between tobacco consumption with traffic safety items is rejected. The Asymp.Sig. value is 0.000, since it is less than 0.05 and even less than 0.01, then the Mann–Whitney U test is rejected with an error of less than 1% and the assumption of the relationship is accepted. But in three cases, the relationship with tobacco consumption has been rejected. These three items, namely (Q 10) an accident situation, (Q 47) Braking time in the face of danger, and (Q 51) the desired length of overtaking, are shown in the fourth column of Table 4-1.

As seen in Tables 2-2 and 2-3, the frequency of questionnaire variables is displayed. It shows the proportion of all individuals that have chosen a particular answer. We have also identified how many people have selected a particular response and we compared the percentages of answers. These tables represent a frequency distribution for a variable, which means that the frequency of observations in each group identifies the desired variable. In the fifth column of Table 4-1, the mean value of the answers has been indicated.

Table 4-1: evaluation of the relationship of tobacco consumption with traffic safety with descriptive analysis and statistical tests

Do you smoke while driving ? (Q25)				Mean=1.6	
statistical tests / Traffic Safety Information for Drivers	Chi-Square	Kruskal-Wallis H	Mann Whithney U	Ferquency	Histogram
Situation of encounter with accident? (Q10)	df=4 Sig=0.000	df=4 Sig=0.000	| |=0.00	Mean=3.07	
Type of accident? (Q46)	df=4 Sig=0.000	df=4 Sig=0.000	| |=13.00 Sig=0.000	Mean=2.49	
Average driving distance per day? (Q14)	df=4 Sig=0.000	df=4 Sig=0.000	| |=7.00 Sig=0.000	Mean=2.33	
Average driving time per day? (Q11)	df=4 Sig=0.000	df=4 Sig=0.000	| |=5.00 Sig=0.000	Mean=2.31	
Braking time in encountering by danger? (Q47)	df=4 Sig=0.000	df=4 Sig=0.000	| |=1.00 Sig=0.060	Mean=2.03	
When driving at 90 Km/h speed in encounter a dangerous situation and brake, how do you think about distance of stopping your car? (Q48)	df=4 Sig=0.000	df=4 Sig=0.000	| |=3.061 Sig=0.002	Mean=2.21 0	
You are on the road at 90 Km / h to overtake a device that is moving at a speed of 80 Km / h. as estimated how many meters do you need along the road, generally? (Q51	df=4 Sig=0.000	df=4 Sig=0.000	| |=0.000 Sig=0.000	Mean=2.21	
/When driving at a speed of 50 km ,h and braking the next vehicle How many meters distance do you need to praccident the collision of your car with the next vehicle?(in normal weather)conditions.)? (Q49	df=4 Sig=0.000	df=4 Sig=0.000	| |=14.018 Sig=0.000	Mean=2.44	
You know your car's speed higher than what speed as high speed in driving inter-city roads in the day? (Q50)	df=4 Sig=0.000	df=4 Sig=0.000	| |=2.00 Sig=0.010	Mean=4.03	
The light of headlamp of your car turns how many meters, approximately? (Q52)	df=4 Sig=0.000	df=4 Sig=0.000	| |=6.00 Sig=0.000	Mean=2.60	

Cluster analysis

Analyze the results of the WARD hierarchical clustering method

The cluster analysis results of SPSS start with the agglomeration table. This chart provides a solution for each number of clusters from 1 to 4112 (total number of cases). The central column entitled coefficient when read from bottom to

top, for a cluster shows the amount of agglomeration of 45220, for two clusters of 38676 for the three 35234 clusters and thelast. These coefficients have been extracted from Table 4-2 for six clusters and are listed in Table 4-3.

Table 4-2: Output obtained by running Ward clustering

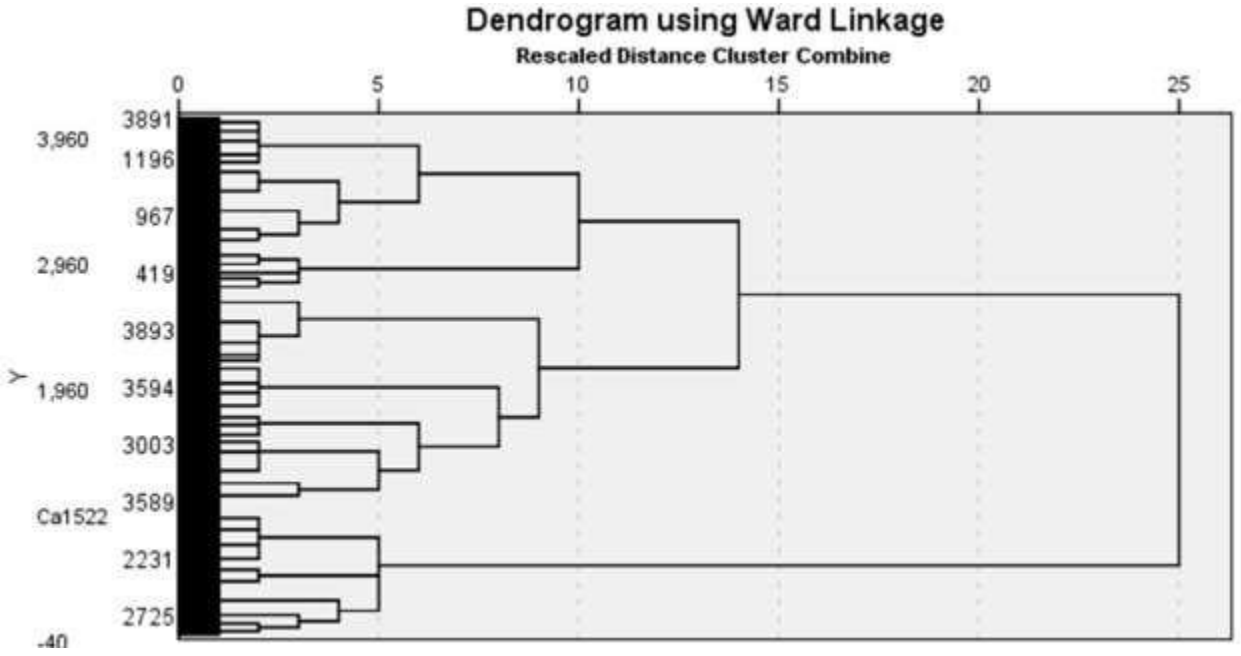
Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4012	4112	.000	0	0	3935
2	4011	4111	.000	0	0	3946
3	4010	4110	.000	0	0	24
			.			
			.			
			.			
4104	1	2	25557.000	4093	4103	4111
4105	3	55	26946.000	4102	4088	4107
4106	4	57	28516.000	4100	4090	4109
4107	3	8	30430.000	4105	4091	4108
4108	3	39	32756.000	4107	4096	4110
4109	4	17	35234.000	4106	4098	4110
4110	3	4	38676.000	4108	4109	4111
4111	1	3	45220.000	4104	4110	0

Table 4-3: Cluster analysis coefficients

Number of clusters	Last steps of agglomeration	Coefficients of the final steps	change
2	45220	38676	6544
3	38676	35234	3442
4	35234	32756	2478
5	32756	30430	2326
6	30430	28516	1914

4-1. Dendrograms derived from Ward cluster analysis



By reviewing the dendrogram diagram 4-1, the validation for the agglomeration table is obtained. By using this graph and the last column as the change of cluster analysis coefficient, Table 3-4 allows us to select the fourth number as the optimal number of clusters.

Analysis of the results of the two-step clustering process

From the analysis of the results' classification received from the questionnaires' data, three clusters can be distinguished as follows. The first cluster with 2144 members (52.2%), the second cluster with 1024 members (24.9%), and the third cluster with 944 members (32.0%). The characteristics are listed in Table 3-5 as follows. Due to the nature of this study, it cannot be claimed that each of the variables is the direct cause of driving accidents, but that the probability of accidents in a cluster is greater with the characteristics of that group.

The first cluster is of non-smoker drivers who have a very low chance of accidents. The characteristics of these drivers in traffic safety have been obtained as follows:

- 32.5% of the members of this cluster have had a crash once or twice until now.
- For 51.8%, average driving time per day is five to seven hours.
- For 55.6%, average mileage per day is likely to be 20 to 50 kilometers.
- 38.4% of the members in the cluster have had an accident type of a collision with the other vehicle.
- 40.4% in this cluster, while driving at speeds of 90 km/h and if encountering a dangerous situation and braking, think their vehicle will come to a halt after a distance of 15 to 25 meters.
- 40.2% of the members in the cluster, when driving at a speed of 90 km/ h and overtaking a vehicle moving at a speed of 80 km/ h, estimate the length of the road needed for overtaking as 200 to 300 meters.
- 71.2% of these drivers do not smoke while driving.
- 40.3% of these drivers consider the speeds between 90 and 110 km/h on inter-city roads to be high.
- 40.3% of this cluster when moving at speeds of 90 km/h and if encountering risky conditions and braking, think their vehicle will stop after 25 to 50 meters.
- 33.6% of the members in this cluster consider one to two seconds as the braking duration time upon encountering danger.
- 30.7% of these drivers believe that the headlamp light of their vehicle illuminates between 50 and 70 meters.

The second cluster are drivers who do not smoke while driving. The characteristics of these drivers in traffic safety are as follows:

- 47.8% of the members in the cluster have had an accident three to four times.
-For 91.8%, the average driving time per day is eight to ten hours.
- For 95.8%, the average driving distance per day is probably 50 to 150 km.
- 68.3% of the cluster's members have had a collision with another vehicle as a type of accident.
- 60.0% of this cluster, while moving at a speed of 90 km/h and braking after encountering a dangerous situation think the vehicle will stop after a distance of 5 to 15 meters.
- 44.0% of the cluster's members when driving at speeds of 90 km/h and overtaking a vehicle moving at a speed of 80 km/h, estimate that the length of the road needed for overtaking is 200 to 300 meters.
- 59.9% of these drivers do not smoke during driving.
- 44.2% of these drivers consider the speeds of 110 km/h on inter-city roads to be high.

The third cluster is the driver who smokes while driving. The characteristics of these drivers in traffic safety are as follows:

- 67.3% of the cluster's members have crashed more than seven times.
of these cluster probably less than five hours drive. - 100% of the drivers
- For 100%, average driving distance per day is probably less than 20 km.
- 65.1% of the members have had other accidents type, which exclude following accident types:
- Collision with another vehicle, pedestrian accident, overtaking or exit from the road, collision with a stationary body
- 43.3% of this cluster, when moving at a speed of 90 km/h, on encountering of a dangerous situation and braking, think the vehicle will stop after a distance of 15 to 25 meters.
- 30.8% of this cluster, while moving on the road at speeds of 90 km/h and to overtake a vehicle moving at a speed of 80 km/h, estimate the required length of the road for overtaking to be 450 to 500 meters.
- 65.1% of these drivers smoke while driving.
- 39.1% of these drivers consider the speeds over 110 km/h on inter-city roads to be high.

Conclusions and recommendations. The human factor has the most important role in driving accidents. At the same time, the impact of risky behaviors is more highlighted. Tobacco consumption, as a human behavior, is the most obvious reason for distraction, especially during driving. As a result of tobacco consumption, the following cases for drivers can arise:

- Turn on the cigarette and engage both hands with the subject of cigarette smoking and driving the driver off the main issue of driving.
- Cigarette smoke causes eye irritation and disturbs the driver's vision.
- The production of carbon monoxide within the vehicle creates a toxic environment for the driver and occupants.
- Increased levels of carboxyhemoglobin causes a reduction in night vision for the driver.
- Physiological problems for the driver, especially cardiovascular complications, are caused by cigarette smoking.

- Tobacco consumption disrupts cognitive thought and judgment and it affects the performance of sensitive tasks such as driving and other tasks that require mental concentration.

Through statistical analysis, this research showed that the number of accidents among individuals who smoke during driving is more than those who do not smoke. The other thing is that these people drive more. The relationship between traffic accident and traffic safety with smoking has been proven by statistical tests. The analysis of the causes of accidents should lead to an improvement in road safety. There is a need for measures to improve human performance, service roads, and vehicles to improve the safety of traffic. These include the following cases: Notification, training programs, regulation of rules, crimes and punishments, installation of warning signs, improving the path margin and the adequacy of guardrails, construction of walking facilities, and warning systems both inside the vehicle and externally. Further research is needed to evaluate the other parameters caused by human behavior that affect traffic safety. These can include the act of lighting cigarettes, eating and drinking, mobile phone use, and talking with passengers while driving.

References

1. Reason, J., Manstead, A., Stradling, S. Baxter, J., Campbell, K. (1990). Errors and violations on the roads: a real distinction? Ergonomics, 33, 10 & 11, 1315 – 1332.
2. Lawton, R., Parker, D., Manstead, A. S. R., Stradling, S. G. (1997). The role of affect in predicting social behaviours: the case of road traffic violations. Journal of Applied Social Psychology, 27, 1258–1276
3. Ozkan, T., & Lajunen, T. (2006). Why are there sex differences in risky driving? The relationship between sex and gender-role on aggressive driving, traffic offences, and accident involvement among young Turkish drivers, Aggressive Behavior, 31(6), 547-558.
4. Parker, D.L., McDonald, L., Rabbitt, P. & Sutcliffe, P. (2000). Elderly drivers and their accidents: the aging driver questionnaire. Accident Analysis & Prevention, 32, 751-759.
5. Elliott, M.A., Baughan, C.J., Sexton, B.F. (2007). Errors and violations in relation to motorcyclists' crash risk. .Accident Analysis & Prevention. 39(3):491-9
6. Gerber b.susan, Finn voelkl , Using SPSS for Windows : Data Analysise and Graphics, Springer, Second Edition.
7. Research Methods in the social Sciences Fourth Edition by chava Frankfort-Nachmias and David Nachmias 1992 – London-Melbourne-Auckland
8. Alan, C.E.(2007). Statistical Analysis Quick Reference Guidebook With spss
9. Jolliffe, I.(2002) Principle component analysis , 2nd edition, Springer-Verlag.
10. Kaufman, L. and Rousseuw P.J. (2005) Finding groups in data : an introduction to cluster analysis, Wiley.
11. Rasouli, M.R., Nouri, M., Zarie, M.R., Sadat, S., Rahimi-Movaghar, V. (2008). Comparison of road traffic mortalities and injuries in Iran with other countries, Chinese Journal of Traumatology, 11(3),131-134.
12. Stough, C. & King, R. (2010). The role of alcohol and other drugs in road deaths and serious injuries, Prevention Research Quarterly, 12, 1-23.
13. Leistikow BN, Martin DC, Jacobs J, Rocke DM, Noderer K. Smoking as a risk factor for accident death: a meta-analysis of cohort studies. Accid Anal Prev 2000;32:397-405.
14. Global status report on road safety 2013: supporting a decade of action.www.who.int/violence_injury_prevention
15. Harvey A. Data systems: a road safety manual for decision-makers and practitioners. Geneva: World Health Organization; 2013; Available from:
16. Pei X, Wong S C, Sze N.N. The Roles of Exposure and Speed in Road Safety Analysis. Accid Anal Prev. 2012; 48: 464-71.
17. Peden M, Scurfield R, Sleet D, Mohan D, Hyder AA, Jarawan E, et al. World report on road traffic injury prevention. World Health Organization Geneva; 2004.