# IDENTIFICATION OF ROAD TRAFFIC FATAL CRASHES LEADING FACTORS USING PRINCIPAL COMPONENTS ANALYSIS

**Mohammad M Molla**[1]

[1]*Graduate Research Assistant, Department of Civil and Environmental Engineering, North Dakota State University, North Dakota, USA*

## Abstract

*Traffic crash fatalities create primary safety concern beyond the traffic congestion and delay. Therefore, the purpose of this study is to identify the principal components/factors associated with road traffic crash in the U.S. through retrospective reviewing based on more than two million records of fatal crashes and 38 years (1975-2012) of National Highway Traffic Safety Administration official's Fatal Accident Reporting System (FARS) database. This study portrays an integrated geographic information system and SAS application in order to find the major factors forcing traffic crashes. The resulting geospatial analysis and principal components analysis yielded critical significant factors causing fatal traffic crashes. The outcomes of this research could be used in transportation safety policy making and planning significantly.*

***Key Words:*** *Accident Analysis Prevention, Clustering, Crash Hot Spot, Geographic Information Systems, Principal Components Analysis, and Traffic Crash*

---------------------------------------------------------------------***---------------------------------------------------------------------

## 1. INTRODUCTION

According to World Health Organization (WHO), by the year 2030, the fifth most prevalent reason for deaths in the world will be road traffic fatalities (Washington Post, 2013). The United States has become third for traffic crash deaths globally (WHO, 2013). The consequences of traffic crash fatalities have a larger impact on global economy. In the year 2012-13, the WHO, World Bank, U.S. Census Bureau, Washington Post, and Forbes addressed the global economic impact due to traffic crash fatalities and the economic growth.

Traffic accident or crash factors are random and varied from state to state. In order to understand the necessity of factors associated with traffic crash and accident, a literature review has been done. Traffic accidents and crime occurrence are well defined threats to public safety (Kuo *et al.*, 2013). Kue et al. (2013) said that using data-driven procedures, police departments would assigned constraint resources efficiently in order to help crime and traffic crash safety, which substantially reduce the crime and crashes in the hot spots areas. U.S. DOT (2013) said "Some of life's greatest lessons come when we learn from our mistakes, and in transportation, where safety is of paramount importance, that maxim is all the more true." Despite the reality, U.S. Department of Transportation and Federal Highway Administration (FHWA) are strongly bound to reduce the highway fatalities and injuries. According to FHWA (2012), highway fatalities and injuries has been reduced substantially from year 2007 to 2010 because of highway safety programs influenced an important effects over the nations. According to the traffic safety facts, it might be inferred that after starting gathering traffic fatality crash data since 1975, the traffic fatalities become declining for the consecutive seven year (NHTSA, 2010).

States and municipalities must have accurate accident modification factors so that maximum benefit of the capital investment can be determined (NCHRP, 2005). According to NCHRP (2005), "crash or accident reduction factors (CRFs or ARFs) provide a quick way of estimating crash reductions associated with highway safety improvements and are used by many states and local jurisdictions in program planning to decide whether to implement a specific treatment and/or to quickly determine the costs and benefits of selected alternatives". Accident modification factors (AMF) is facilitating AASHTO and NCHRP in order to develop the strategic safety policy and guidelines so that states and local users can be benefited. NCHRP also discussed that the accident modification factors are always obscured to the end user. In this regards, state agencies are using their own developed system for depending on their states crash data for the accident modification factors planning. Thus the need of in-depth traffic crash/accident analysis is an important issue is becoming more and more important for the interest of federal, states, local agencies.

In this regards, since 1975 NHTSA maintaining the related factors associated with fatal accident in to their structured FARS databases. The FARS data element identifies factors related to the crash expressed by the investigating officer and associated in to three categories of related factors: 1) Crash Related 2) Driver Related, and 3) Person Related. The factors can be broadly categorized into several major categories such as vision obscured; avoiding, swerving, or sliding; roadway features; physical/mental condition; distractions; non-motor-vehicle-operated related; and miscellaneous causes.

Traffic crash fatalities are geospatially related and enclosed by clustered patterns for its geographic location. Statistics revealed that the exponential increase of traffic fatalities are concurrent with the economic growth in the nation's history, aggregated with widely spreading traffic accident hot spots. The amount of traffic crash fatalities may eclipse the all-time high set in the year 1976. Therefore, aggregated with widely spreading traffic accident hot spots, the state transportation agencies and law enforcement agencies are working strategically towards implementing zero fatalities on roadways.

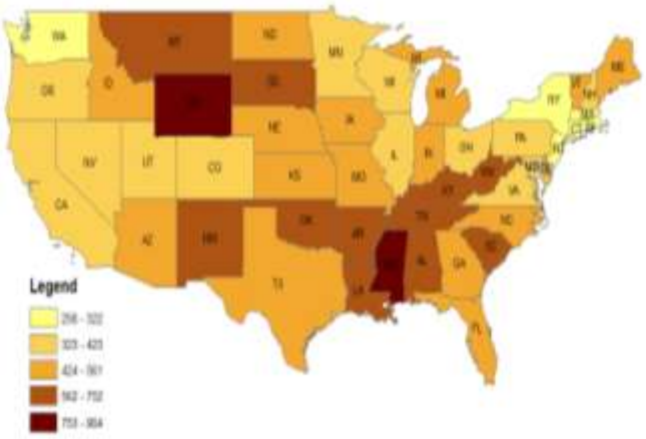


**Fig -1**: Total Number of Fatalities per 100 Thousand People, 1980-2011

Figure 1 illustrates that the total number of traffic crash fatalities per 100,000 people for states from 1980 to 2011. Total number of fatalities was aggregated according to the FARS published data. Based on 100,000 people, the highest number of fatalities per population is in Wyoming and Mississippi, which ranges from 753 to 904 fatalities per 100,000 populations in 38 years. The lowest number of fatalities per 100,000 peoples includes the states of Washington, New York, and New Jersey.

The fuzzy road network has multivariate factors which are directly associated with traffic crash accident, which can be overseen through the related factors of Fatal Accident Reporting System (FARS) database from 1975-2012. FARS database maintained different level of dataset such Vehicle, Accident, and Person (Driver). In this study, our goal is to identify the person (driver) related factors and detect any possible clustering among the states. More specifically, the primary goal of this research can be listed as to:

- Determine the leading person (driver) related factors causing fatal traffic crash in the nation's highway.
- Detect any clustering among the U.S. States based on the identified factors.

## 2. DATA AGGREGATION AND ANALYSIS METHODOLOGY

The study area considered 48 states except District of Colombia and Alaska. Accident data collected from National Highway Traffic Safety Administration were aggregated into the ArcGIS environment. Noted that the data element identified factors related to the driver expressed by an investigating officer. In the database, related factors were stored in a different variable coded varied with year. Thus, data was further processed and cleaned up in a statistical software SAS. After cleaning up the unnecessary data and clean up, final aggregated database showed around two million crash events and 99 driver related factors, which is presented in Table 1.

After processing the data, principal components analysis was performed to identify the leading fatal crash factor and based on the identified factors a single linkage cluster analysis was performed in order to see if is there any clustering among the 48 states. The cubic clustering criterion (CCC), Pseudo F (PSF), and $t^2$ (PST2) statistics are used to determine the number of cluster for the data. SAS (2008) discussed that the CCC and PSF are not suitable in order to identify the number of clusters in a single linkage method because the method has a tendency to shear the tails of the distribution. In that case, PST2 could be utilized for this purpose. Initially, any cluster with a large PST2 value should be selected; the number of clusters used in the analysis could be one greater than the initial number of clusters.

**Table -1**: Driver Related Factors

| SL Factors Description |
|---|
| 00 None |
| 01 Drowsy, Sleepy, Asleep, Fatigued |
| 02 Ill, Passed Out/Blackout |
| 03 Emotional (e.g., Depression, Angry, Disturbed) |
| 04 Reaction to or Failure to Take Drugs/Medication |
| 05 Under the Influence of Alcohol, Drugs, or Medication |
| 06 Inattentive/Careless (Talking, Eating, Car Phones, etc.) |
| 07 Restricted to Wheelchair |
| 08 Road Rage/Aggressive Driving |
| 09 Impaired Due to Previous Injury |
| 10 Deaf |
| 11 Other Physical Impairment |
| 12 Mother of Dead Fetus/Mother of Infant Born Post Crash |
| 13 Mentally Challenged |
| 14 Failure to Take Drugs/Medication |
| 15 Seat Back Not in Normal Position, Seat Back Reclined |
| 16 Police or Law Enforcement Officer |
| 17 Running off Road |
| 18 Traveling on Prohibited Traffic ways |
| 19 Legally Driving on Suspended or Revoked License |
| 20 Leaving Vehicle Unattended with Engine Running; Leaving Vehicle Unattended in Roadway |
| 21 Overloading or Improper Loading of Vehicle with Passenger or Cargo |

22 Towing or Pushing Vehicle Improperly
23 Failing to Dim Lights or to Have Lights on When Required
24 Operating Without Required Equipment
25 Creating Unlawful Noise or Using Equipment Prohibited by Law
26 Following Improperly
27 Improper or Erratic Lane Changing
28 Failure to Keep in Proper Lane
29 Illegal Driving on Road Shoulder, in Ditch, or Sidewalk, or on Median
30 Making Improper Entry to or Exit from Traffic way
31 Starting or Backing Improperly
32 Opening Vehicle Closure into Moving Traffic or Vehicle is in Motion
33 Passing Where Prohibited by Posted Signs, Pavement Markings, Hill or Curve, or School Bus Displaying Warning Not to Pass
34 Passing on Wrong Side
35 Passing with Insufficient Distance or Inadequate Visibility or Failing to Yield to Overtaking Vehicle
36 Operating the Vehicle in an Erratic, Reckless, Careless or Negligent Manner or Operating at Erratic or Suddenly Changing Speeds
37 Police Pursuing this Driver or Police Officer in Pursuit
38 Failure to Yield Right of Way
39 Failure to Obey Actual Traffic Signs, Traffic Control Devices or Traffic Officers, Failure to Observe Safety Zone Traffic Laws
40 Passing Through or Around Barrier
41 Failure to Observe Warnings or Instructions on Vehicle Displaying Them
42 Failure to Signal Intentions
43 Driving too Fast for Conditions
44 Driving Too Fast for Conditions or in Excess of Posted Speed Limit
45 Driving Less Than Posted Maximum
46 Operating at Erratic or Suddenly Changing Speeds
47 Making Right Turn from Left-Turn Lane or Making Left Turn from Right-Turn Lane
48 Making Improper Turn
49 Failure to Comply With Physical Restrictions of License
50 Driving Wrong Way on One-Way Traffic way
51 Driving on Wrong Side of Road (Intentionally or Unintentionally)
52 Operator Inexperience
53 Unfamiliar With Roadway
54 Stopping in Roadway (Vehicle Not Abandoned)
55 Under riding a Parked Truck
56 Improper Tire Pressure
57 Locked Wheel
58 Over Correcting
59 Getting Off/Out of or On/In to Moving Vehicle
60 Getting Off/Out of or On/In to Non-Moving Vehicle
61 Rain, Snow, Fog, Smoke, Sand, Dust
62 Reflected Glare, Bright Sunlight, Headlights
73 Driver Has Not Complied with Learners Permit or Intermediate Driver License Restrictions (GDL Restrictions, Since 2004)
74 Driver Has Not Complied With Physical or Other Imposed Restrictions (Since 2004)
75 Broken or Improperly Cleaned Windshield
76 Other Obstruction
77 Severe Crosswind
78 Wind from Passing Truck
79 Slippery or Loose Surface
80 Tire Blow-Out or Flat
81 Debris or Objects in Road
82 Ruts, Holes, Bumps in Road
83 Live Animals in Road
84 Vehicle in Road
85 Phantom Vehicle
86 Pedestrian, Pedal cyclist, or Other Non-Motorist in Road
87 Ice, Water, Snow, Slush, Sand, Dirt, Oil, Wet Leaves on Road
88 Trailer Fishtailing or Swaying
89 Carrying Hazardous Cargo Improperly
90 Hit-and-Run Vehicle Driver
91 Non-Traffic Violation Charged – Manslaughter or Homicide or Other Assault
92 Other Non-Moving Traffic Violation
93 Cellular Telephone
94 Cellular Telephone in Use in Vehicle
95 Computer Fax Machines/Printers
96 On-Board Navigation System
97 Two-Way Radio
98 Head-Up Display
99 Unknown

## 3. RESULTS AND DISCUSSIONS

### 3.1. Principal Components Analysis

The Principal Components Analysis was completed using the statistical SAS programming. The first eigenvalue of 236.995, accounts for approximately 39.57% of the standardized variance. A total of 99 factors were used during PCA analysis. It was revealed that only thirteen eigenvalues explained the 80% of the variance, which is presented in Table 2 and Figure 2. PCA analysis suggests that the first thirteen principal components provide an adequate summary of the data for most purposes. This thirteen principal components discovered for 45 factors, which is substantially influence on the traffic crash accidents.

**Table -2**: Eigenvalues of Principal Components Analysis

Eigenvalues of the correlation matrix: total=599, average=1

| SL | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| 1 | 236.995207 | 186.713767 | 0.3957 | 0.3957 |
| 2 | 50.281440 | 15.719126 | 0.0839 | 0.4796 |
| 3 | 34.562314 | 9.715977 | 0.0577 | 0.5373 |
| 4 | 24.846337 | 3.905710 | 0.0415 | 0.5788 |
| 5 | 20.940627 | 1.910206 | 0.0350 | 0.6138 |
| 6 | 19.030421 | 1.445640 | 0.0318 | 0.6456 |
| 7 | 17.584781 | 2.975183 | 0.0294 | 0.6750 |

| | | | | |
|---|---|---|---|---|
| 8 | 14.609598 | 1.740300 | 0.0244 | 0.6994 |
| 9 | 12.869298 | 0.028081 | 0.0215 | 0.7209 |
| 10 | 12.841217 | 0.790838 | 0.0214 | 0.7423 |
| 11 | 12.050379 | 0.250035 | 0.0201 | 0.7624 |
| 12 | 11.800344 | 0.911742 | 0.0197 | 0.7821 |
| 13 | 10.888602 | 1.248457 | 0.0182 | 0.8003 |

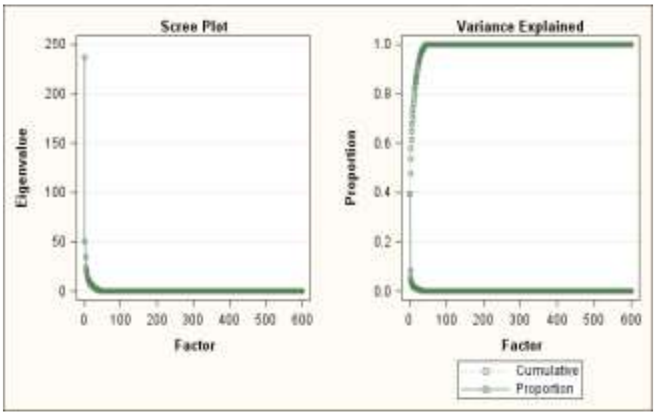


**Fig -2**: Scree Plot of Factors Selection

Interestingly in the first principal component, out of 45 factors, only 23 factors contributing significantly which is presented in the Table 3. Result showed that high speed is the number one cause for fatal traffic accident in the 48 states. High speed, Starting and backing improperly, failure to observe warnings or instructions, passing with insufficient distance, and driving less than posted speed limit is the top five fatal traffic accident factor. Result also showed that there are other factors identified in the top 23 factors which is related to unconsciousness of driver such as depression and inattentive.

**Table -3**: Top Influential Factors

| Factors | Position |
|---|---|
| High-Speed Chase with Police in Pursuit | 1 |
| Starting and Backing Improperly | 2 |
| Failure to observe Warnings or Instructions | 3 |
| Passing with insufficient distance | 4 |
| Driving less than posted speed limit | 5 |
| Towing and pushing vehicle improperly | 6 |
| Operating the vehicle in an erratic, reckless, careless and negligent manner | 7 |
| Creating unlawful noise | 8 |
| Failure to Take Drugs/Medication | 9 |
| Mother of Dead Fetus | 10 |
| Running off Road | 11 |
| Other Drugs(Marijuana, Cocaine) | 12 |
| Deaf | 13 |
| Operating without required equipment | 14 |
| Depression | 15 |
| Passing on wrong side | 16 |
| Seat Back Not in Normal Position | 17 |
| Mentally Challenged | 18 |
| Leaving vehicle unattended | 19 |
| passing where prohibited by posted signs | 20 |
| Inattentive | 21 |
| Driving in excess of posted speed limit | 22 |
| Illegal driving in road shoulder, ditch and sidewalk or median | 23 |

### 3.2. Selection of Number of Clusters

Using single linkage cluster procedure, hierarchical cluster of observation was initially determined. Using the top 45 factors which may support to explain the 80% of the variance of the data was selected in order to select the number of cluster. Table 4 provides the results portraying the last 30 generations of the cluster history for the single linkage method.

**Table -4**: Cluster History

| Number of Clusters | Cubic Clustering Criterion | Pseudo Statistic | FPseudo t-Squared | Norm Minimum Distance | Tie |
|---|---|---|---|---|---|
| 30 | . | 29.4 | 3.0 | 0.0556 | T |
| 29 | . | 26.1 | 3.0 | 0.0556 | T |
| 28 | . | 23.4 | 3.0 | 0.0556 | T |
| 27 | . | 21.2 | 3.0 | 0.0556 | T |
| 26 | . | 19.3 | 3.0 | 0.0556 | T |
| 25 | . | 17.8 | 3.0 | 0.0556 | T |
| 24 | . | 16.4 | 3.0 | 0.0556 | T |
| 23 | . | 15.3 | 3.0 | 0.0556 | T |
| 22 | . | 14.4 | 3.0 | 0.0556 | T |
| 21 | . | 13.5 | 3.0 | 0.0556 | T |
| 20 | . | 12.8 | 3.0 | 0.0556 | T |
| 19 | . | 12.2 | 3.0 | 0.0556 | T |
| 18 | . | 11.7 | 3.0 | 0.0556 | T |
| 17 | . | 11.3 | 3.0 | 0.0556 | T |
| 16 | . | 12.4 | . | 0.0556 | T |
| 15 | . | 13.7 | 3.0 | 0.0556 | T |
| 14 | . | 15.2 | 3.0 | 0.0556 | T |
| 13 | . | 16.9 | 3.0 | 0.0556 | T |
| 12 | . | 18.8 | 3.0 | 0.0556 | T |
| 11 | . | 21.1 | 3.0 | 0.0556 | T |
| 10 | . | 23.9 | 3.0 | 0.0556 | T |
| 9 | -15 | 27.6 | . | 0.0556 | T |
| 8 | -13 | 32.3 | 3.0 | 0.0556 | T |
| 7 | -12 | 38.5 | 3.0 | 0.0556 | |
| 6 | -10 | 46.6 | 24.0 | 0.1112 | T |
| 5 | -8.2 | 57.8 | 9.6 | 0.1112 | T |
| 4 | -9.8 | 27.8 | 41.5 | 0.1112 | T |
| 3 | -8.8 | 23.4 | 15.8 | 0.1112 | |
| 2 | -4.9 | 47.1 | 4.6 | 0.1668 | T |
| 1 | 0.00 | . | 47.1 | 0.1668 | |

Figure 3 plots all three statistics based on the number of clusters. From Table 4 and Figure 3, it is evident that peak values of PST2 found at level 4 and 6. Thus the number of cluster will be 4+1=5, or 6+1=7. But in the last column of Table 4, we can see that there is ties exist for cluster 5, which means at this level, estimation of number of clusters would be indeterminate. Therefore, at level 6 the number of clusters might be suitable as 7. Finally, seven clusters were selected based on the 45 factors. The results of the hierarchical clustering procedure have been presented visually using a tree diagram/dendogram in Figure 4. The dendogram displays all the stages of the hierarchical procedure and the distances at which clusters were merged. Dendogram also suggest for only seven clusters that be considered. The computed clustered states have been presented in choropleth map in Figure 5. It was also revealed that Texas, California, Mississippi, Florida, Pennsylvania, and Ohio has extreme scenario in case of traffic fatalities.

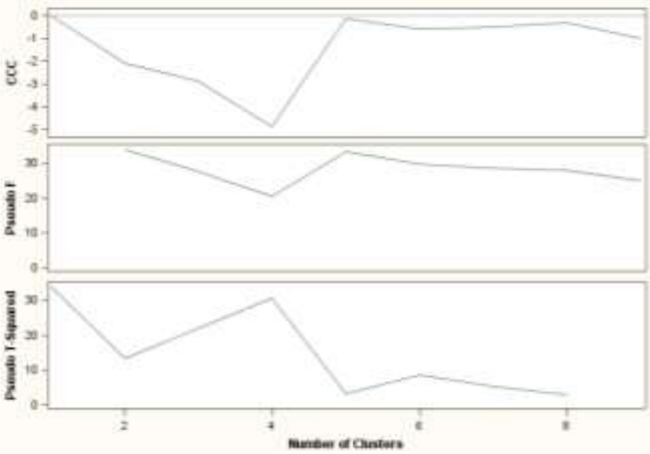


**Fig -3**: Criteria for the Number of Clusters

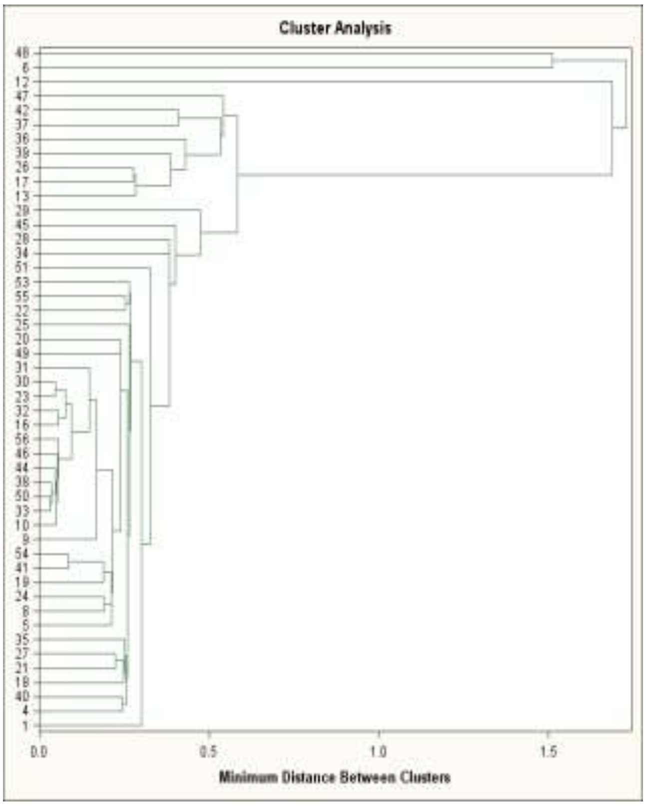


**Fig -4**: Dendogram Showing Minimum Distance between Clusters


**Fig -5**: Clusters of States

## 4. CONCLUSIONS

It might be inferred that based on 38 years of FARS data only 23 drivers related factors contributing significantly or has substantial influence on the nationwide traffic crash accidents/incidents, this has been listed in the Table 3. Out of these top 23 leading factors, top ten factors are high-speed chase with police in pursuit; starting and backing improperly; failure to observe warnings or instructions; passing with insufficient distance; driving less that posted speed limit; towing and pushing vehicle improperly; operating the vehicle in an erratic, reckless; careless and negligent manner; creating unlawful noise; failure to take drugs/medication; mother of dead fetus. Interestingly high-speed chase with police in pursuit; driving less than posted speed limit; driving in excess of posted speed limit has been revealed in the top 23 leading driver related factors causing traffic accident in the United States. Study also discovered that Drugs related to marijuana, cocaine, etc. are the twelfth most reason for the traffic accident in the United States. Inattentive while driving is 21th reason for traffic accident.

The main objective of this research was to see if is there any clustering relationship exist in 48 states considering the driver related most important 45 factors. The computed clusters based on single linkage method revealed that considering 45 driver related factors which causing accident, may classify the 48 states into seven clusters. Number of clusters found technically acceptable but not providing enough information about the remaining 42 states. Since Texas, California, Mississippi, Florida, Pennsylvania, and Ohio has large number of traffic fatalities, this six states are forming six different clusters. These six states are influencing the clustering analysis. Remaining 42 states may be analyzed further to see how the cluster forms without the effect of the larger six states. Due to resources constraints, the goal of the study was limited as specified earlier; however, many cause and effect relationships can be established based on these 38 years data. Thus, further research can be performed considering combining the driver, vehicle and crash related factors.

## REFERENCES


[1] FHWA (Federal Highway Administration-U.S. Department of Transportation). (2012). <http://safety.fhwa.dot.gov/provencountermeasure/pc_memo.htm> (Accessed December, 2013).

[2] Kuo, P., Lord, D. and Walden, T. D. (2013). "Using geographical information systems to organize police patrol routes effectively by grouping hotspots of crash and crime data." Journal of Transport Geography, 30, 138-148.

[3] Molla, M. M., Stone, M. L., Lee, E. (2014). "Geostatistical Approach to Detect Traffic Accident Hot Spots and Clusters in North Dakota". Upper Great Plains Transportation Institute.

[4] NCHRP (National Cooperative Highway Research Program). (2005). "Crash Reduction Factors For Traffic Engineering and Intelligent Transportation System (Its) Improvements: State-Of-Knowledge Report." Research Results Digest 299.

[5] NHTSA (National Highway Traffic Safety Administration). (2010). <http://wwwnrd.nhtsa.dot.gov/Pubs/811363.pdf> (accessed December, 2013).

[6] Rencher, A. C. (2001). "Methods of Multivariate Analysis." Second Edition, John Wiley & Sons, Inc., New York, NY. SAS (SAS Institute Inc.). (2008). "SAS/STAT® 9.2 User's Guide." Cary, NY.

[7] USDOT (United States Department of Transportation Research and Innovative Technology Administration). (2013). <http://www.rita.dot.gov/sites/default/files/rita_archives/rita_publications/horizons/2008_11_12/html/improving_operational_safety.html> (Accessed December, 2013).

[8] Washington Post. (2013). <http://www.washingtonpost.com/blogs/worldviews/wp/2013/01/18/a-surprising-map-of-countries-that-have-the-most-traffic-deaths/> (Accessed December, 2013).

[9] WHO (World Health Organization). (2013). <http://www.who.int/violence_injury_prevention/road_traffic/en> (Accessed December, 2013).