IATSS Research 39 (2016) 146–155

Contents lists available at ScienceDirect

IATSS Research

Association knowledge for fatal run-off-road crashes by Multiple Correspondence Analysis

Subasish Das [a,*], Xiaoduan Sun [b]

[a] Systems Engineering, University of Louisiana at Lafayette, Lafayette, LA 70504, United States
[b] Civil Engineering Department, University of Louisiana at Lafayette, Lafayette, LA 70504, United States

ARTICLE INFO

Available online 13 July 2015

Keywords:
ROR crashes
Multiple Correspondence Analysis
Dimensionality reduction
Cloud of combination groups

ABSTRACT

In 2013, 346 out of 616 fatal crashes in Louisiana were single vehicle crashes with Run-Off-Road (ROR) crashes being the most common type of single vehicle crash. In order to create effective countermeasures for reducing the number of fatal single vehicle ROR crashes, it is important to identify any associated key factors that can quantitatively assess the performance of roads, vehicles and humans. This research uses Multiple Correspondence Analysis (MCA), a multidimensional descriptive data analysis method that associates a combination of factors based on their relative distance in a two dimensional plane, to analyze eight years (2004–2011) of fatal ROR crashes in Louisiana. This method measures important contributing factors and their degree of association. The results revealed that drivers of lightweight trucks, drivers on undivided state highways, male drivers in passenger-vehicles at dawn, older female (65–74) drivers in non-passenger vehicles, older drivers facing hardship to yield in partial access control zones, and drivers with poor reaction time due to impaired driving were closely associated with fatal ROR crashes.

Results of the MCA method can help researchers select the most effective crash countermeasures. Further work on the degree of association between the identified crash contributing factors can help safety management systems develop the most efficient crash reduction strategies.

1. Introduction

Most single vehicle crashes are ROR crashes, which are more likely to result in fatalities and severe injuries than typical vehicle crashes are [1]. In 2012 and 2013, respectively 384 and 346 out of a total of 652 and 616 fatal crashes in Louisiana were ROR crashes [2]. From prior studies, we know that single vehicle ROR crashes are usually caused by a combination of factors such as inadequate roadway design, mechanical problems, environmental conditions and/or drivers' poor performance [3–5]. The combination of factors could be spatially different (i.e. crashes occurring on highways versus intersections) and temporally different (i.e. crashes occurring in December versus those in May). Failure to recognize the spatial and temporal differences of those factors may lead to insufficient or ineffective actions taken to reduce the number of ROR crashes.

Identifying crash-prone factors and combinations of factors by analyzing a large dataset is not a trivial task. The commonly used statistical inferential methods, i.e. ANOVA, and safety performance models cannot identify the combination of factors simultaneously. Multiple Correspondence Analysis (MCA) is an extension of Correspondence Analysis (CA) for more than two variables and is widely used in categorical data analyses, especially in social sciences and marketing research [6]. By using this technique we can visualize the patterns of combined crash contributing factors. MCA helps researchers discover the structure of categorical data by presenting complicated relationships in a simple chart that demonstrates a combination of significant variables through the reduced data dimension analysis. This method presents the correlation between the variables and their relationship to the interested resultant variable by creating combination clouds.

The persistently high rate of fatal ROR crashes in Louisiana and the United States indicates continuous need for research. Reducing ROR crashes is critical in fulfilling state and national safety goal and MCA will help determine the association between key factors of fatal ROR crashes so that transportation authorities can take necessary actions to reduce crash frequencies and severities.

2. Literature review

J.P. Benzécri developed Multiple Correspondence Analysis (MCA), a statistical approach based on Correspondence Analysis (CA). MCA is usually considered to be one of the main standards of geometric data

* Corresponding author. Tel.: +1 225 288 9875; fax: +1 337 739 6688.
E-mail addresses: subasishsn@gmail.com (S. Das), xsun@louisiana.edu (X. Sun).
Peer review under responsibility of International Association of Traffic and Safety Sciences.

http://dx.doi.org/10.1016/j.iatssr.2015.07.001

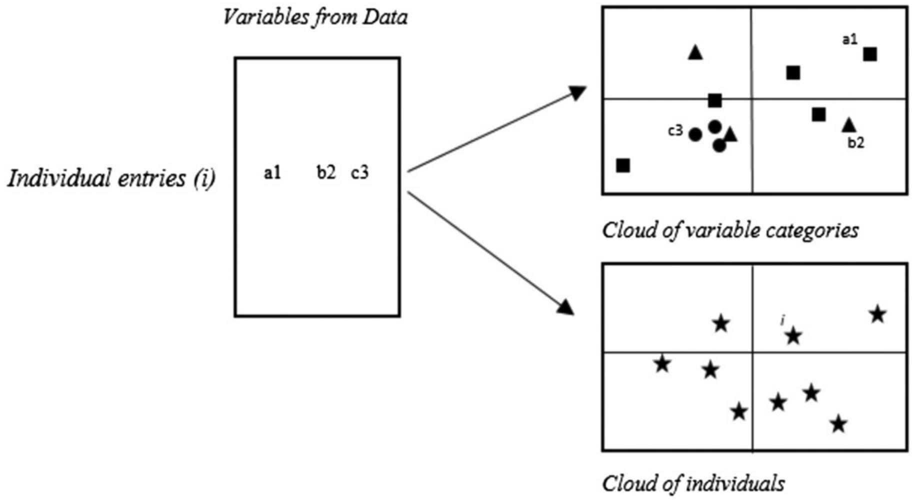

Fig. 1. Data table and the two clouds of points generated by MCA with the flowchart [6].

analysis (GDA) in the fields of social science and marketing research. GDA is also referred to as the pattern recognition method that treats arbitrary data sets as clouds of points in *n*-dimensional space. However, in the field of multivariate transportation data analysis, researchers rarely use geometric methods. Roux and Rouanet pointed out that this method, though it is a powerful tool for analyzing a full-scale research database, is still rarely discussed and therefore under-used in many promising fields [6].

Hoffmann and De Leeuw used MCA as a multidimensional scaling method to show how questions posed of categorical marketing research data can be answered with MCA in terms of significant and meaningful results [7]. Fontaine was the first to use MCA for a typological analysis of vehicle-pedestrian crashes [8]. For Fontaine's research, the classification of pedestrians involved in crashes was divided into four major groups. The typology produced by this analysis reveals correlations between criteria without necessarily indicating a causal link with the crashes. The resulting typological breakdown served as a basis for in-depth analysis to improve the understanding of these crashes and propose necessary strategies. Golob and Hensher utilized MCA to establish causality of nonlinear and non-trivial relationships between socioeconomic descriptors and outcomes of travel behavior [9]. Factor et al. used MCA to conduct a systematic exploration of the connection between drivers' characteristics and their involvement in collision types [10]. There is a vast amount of literature on accident research and model development that, for the sake of brevity, cannot be covered in this article. The research team has compiled an extensive list of this literature in a webpage for the convenience of any interested readers [11].

The research introduced in this paper serves as a starting point to demonstrate the application of MCA to determine the significant clouds of crash contributing factors for fatal ROR crashes. The findings will help state agencies determine effective crash countermeasures.

3. Methodology

3.1. Theory

For a database or table with categorical variables, the scheme of MCA can be explained by taking an individual record (in row), *i*, where three variables (represented by three columns) have three different category indicators (a_1, b_2, and c_3). The spatial distribution of the points calculated by the dimensions based on these three categories would be generated by MCA. As shown in Fig. 1, MCA yields two clouds of points: the cloud of individual records and the cloud of categories [6]. A cloud of points is not just a *simple graphical* display; it can be compared to a geographic map with the same scale in all directions. A geometric diagram cannot be strained or contracted along one specific dimension. Thus, a

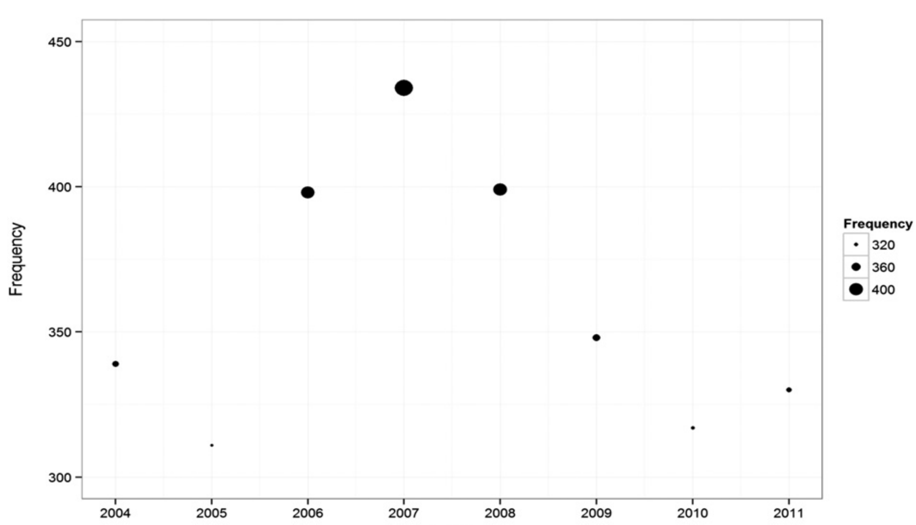

Fig. 2. Fatal ROR crashes.