

Factor Association Using Multiple Correspondence Analysis in Vehicle-Pedestrian Crashes

Subasish Das, PhD Candidate
University of Louisiana

Xiaoduan Sun, PhD & PE, Professor
University of Louisiana

Research Question

What are the key contributing factors for vehicle pedestrian crashes? What's the purpose of using multiple correspondence analysis (MCA) in place of conventional parametric approaches?

Abstract

This research uses **Multiple Correspondence Analysis (MCA)**, a data analysis method used for detecting and representing underlying structures in a categorical data set, to analyze eight years (2004-2011) of vehicle-pedestrian crashes in Louisiana. The findings indicated several non-trivial focus groups such as drivers with high occupancy vehicles, female drivers in bad weather conditions, and drivers distracted by mobile phone use. Other key associated factors were hillcrest roadways, dip/hump aligned roadways, roadways with multiple lanes, and roadways with no lighting at night. Male drivers were seen to be more inclined towards severe and moderate injury crashes. Fatal pedestrian crashes were correlated to two-lane roadways with no lighting at night. This method helped to measure significant contributing factors and degrees of association between the factors by analyzing the systematic patterns of variations with categorical datasets of pedestrian crashes.

Introduction

In 2011, 4,432 pedestrians were killed (14% of total traffic crash fatalities) and 69,000 pedestrians were injured in vehicle-pedestrian crashes in the United States. Improvement of pedestrian safety is one of the topmost priorities in the American Association of State Highway and Transportation (AASHTO) Strategic Highway Safety Plan (SHSP).

The vehicle-pedestrian crash statistics of Louisiana call for instant and advanced solutions to alleviate safety concerns of the pedestrians. The objective of this study was the application of MCA on vehicle-pedestrian crashes to (1) identify the relative closeness of the key association factors, (2) find important nontrivial association between the key factors and (3) provide intuitions to select better countermeasures to improve pedestrian safety. Improving pedestrian safety is crucial to accomplish the state's **Destination Zero Deaths** goal and the MCA method used in this paper will help to find the relative closeness of the key association factors so that necessary actions can be taken to improve the strategy of pedestrian safety.

The research introduced in this paper is unique and different from previous studies and serves as a starting point to demonstrate the use of MCA to determine the significant cloud of crash contributing factors for vehicle pedestrian crashes which could help state agencies develop the most efficient crash countermeasures.

Multiple Correspondence Analysis

Multiple Correspondence Analysis (MCA) is widely used in categorical data analysis, especially in social science and marketing research, and is considered as an extension of correspondence analysis (CA) to more than two variables. It is an exploratory data analysis method that can visualize the patterns of the cluster consisting of key contributing factors.

- For a database with categorical variables, the scheme of the MCA method can be explained by taking an individual row or record i as an example where three variables (represented by three columns) have three different category indicators (a1, b1, and c3).
- The spatial distribution of the points calculated by the dimensions based on these three categories would be generated by MCA. MCA yields two clouds of points as shown in the figure: the cloud of individual records and the cloud of categories.
- A cloud of points is not just a simple "graphical display", it is similar to a geographic map with the same distance scale in all directions.

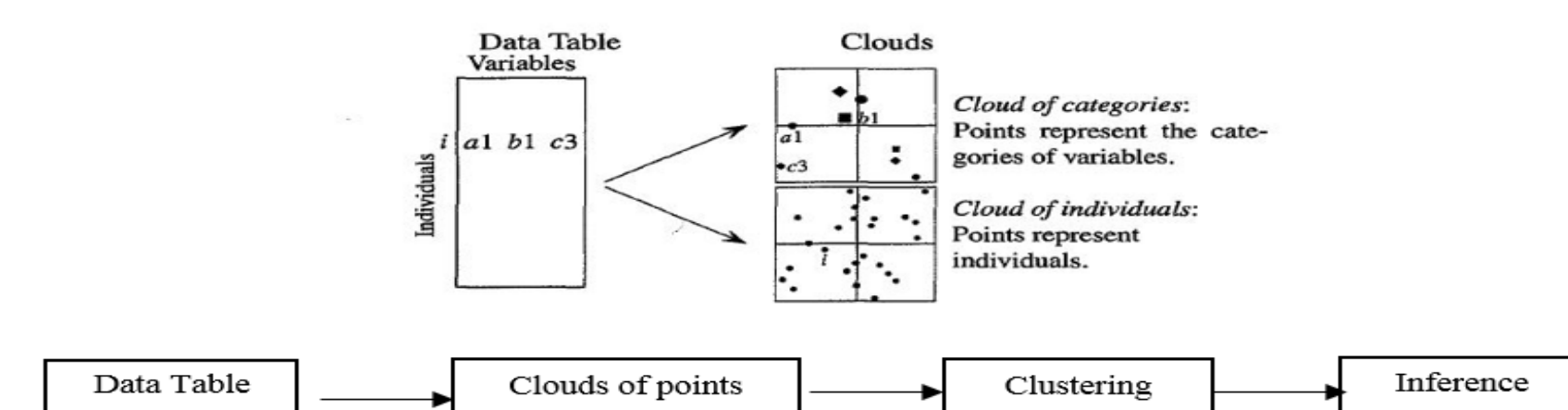


Figure 1: Clouds of points generated by MCA method.

Data Description

Eight years (2004-2011) of Louisiana crash data was used in this study.

- In the crash database, there are numerous variables that are non-pertinent for this research, such as: the VIN, driver 's license number, database manager 's name, police report number etc. To focus on the meaningful analysis, a set of key variables were selected such as the roadway geometrics (alignment and lighting), collision type, environmental factor (weather), driver related factors (driver gender, age and condition), number of occupants and pedestrian related factor (pedestrian gender, age, condition and severity). The variable selection method used the research findings of the previous related research with engineering judgment.
- An initial analysis indicated that some variables are highly skewed which means that a majority of crashes fall into one of the two or more categorical values. For example, 94% of the crashes involved roadways with straight level alignment, 76% of the crashes were single vehicle crashes, and 78% of crashes involved single occupant crashes.
- It is seen that 61% of pedestrians involved in crashes were male which was higher than the general trend (around 50 to 55% of traffic crashes involved male drivers in Louisiana). The not-too-skewed variables include collision type, pedestrian injury, and lighting condition.

Discussion

Discussion on the combination clouds are stated below:

- **Combination cloud 1** combines a wider variety of variable categories: hillcrest aligned four-lane roadways, single vehicle collisions, severe and moderate pedestrian injuries, number of occupants 2 and 3, and male pedestrians. It indicates that hillcrest aligned four-lane roadways were crash prone towards moderate and severe pedestrian injury. It also indicates that larger occupancy vehicles were responsible for single vehicle-pedestrian crashes on this specific type of roadway.
- **Combination cloud 2** associates male pedestrians with moderate injury crashes while the number of occupants in the vehicles was two. It indicates that car occupancy had some role in pedestrian related crashes.
- **Combination cloud 3** associated several factors: complaint injury of female pedestrians, dawn or dusk, abnormal weather and night time crashes in roadways with lighting. **Combination cloud 4** combines a few factors: clear weather, single occupant, six-lane straight-level aligned roadways, head-on collisions and driver distraction due to outside events. Both of the findings are non-trivial in nature.
- **Combination cloud 5** also associates different variable categories: driver distraction from mobile or inside equipment, daytime right angle and sideswipe crashes, dip/hump roadways with unknown information on lanes, and PDO pedestrian crashes. This combination indicates cell phone involvement in dip/hump aligned roadways.
- **Combination cloud 6** associates three variable categories: fatal pedestrian crash, nighttime crash, and two-lane roadways with no lighting. It indicates that absence of lighting at night is a significant factor for pedestrian traffic severity. This cloud clearly indicates one major focus group on roadway geometrics.

Conclusion

In particular the ability of MCA to deal with multidimensional data makes it particularly useful for exploring the factors influencing traffic crash occurrences. The findings from this research shed light on the pattern recognition of vehicle-pedestrian crashes and expose new aspects in pedestrian safety and also point to potential future research considering more variables and large dataset from multiple states.



Major Findings

The results of the MCA detects several interesting combination clouds. The future work on the degree of association of the crash contributing factors can help safety management systems identify the most effective crash reduction strategies.

Principle MCA Plot

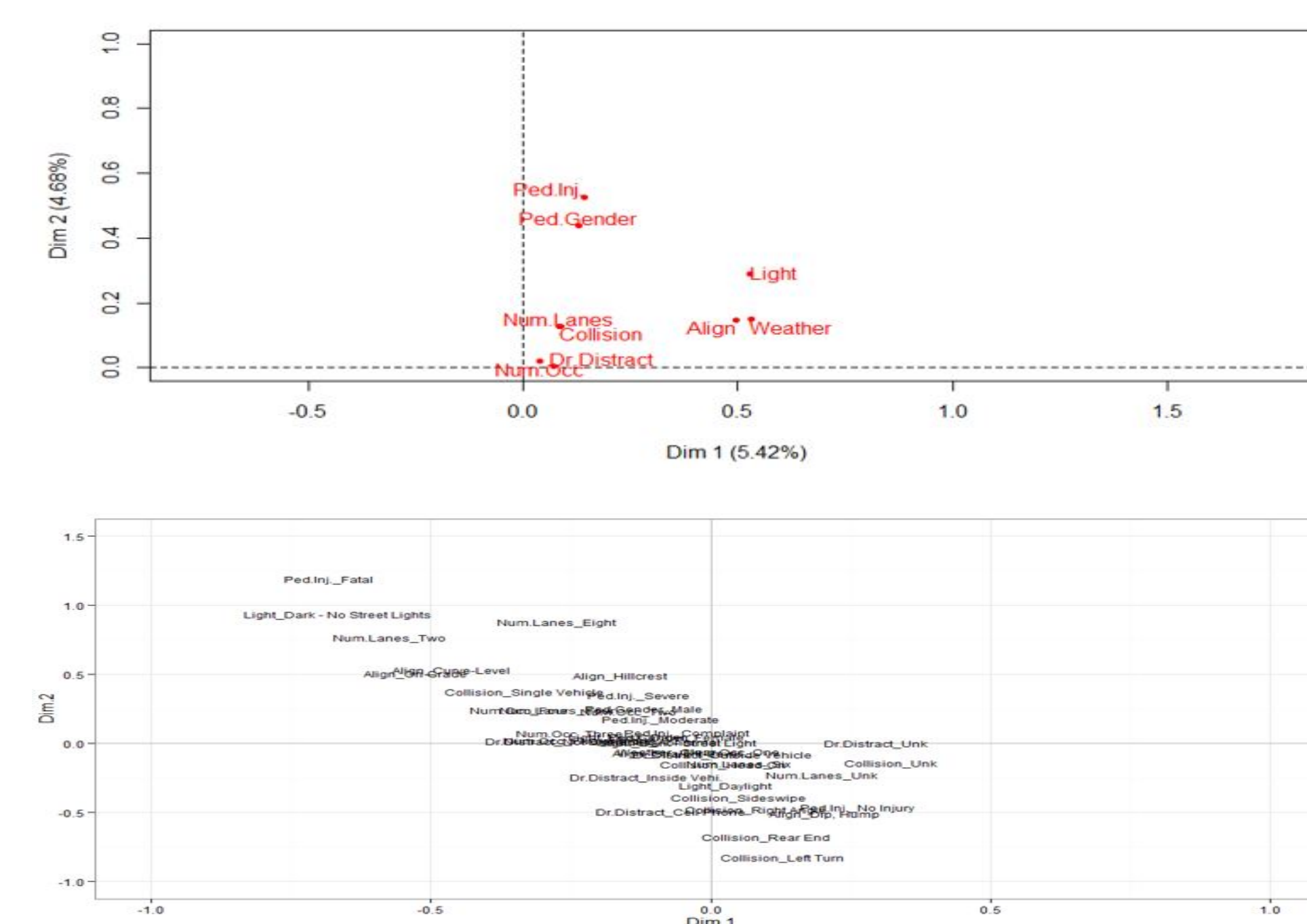


Figure 2: MCA plot for the variable categories.

Combination Clouds

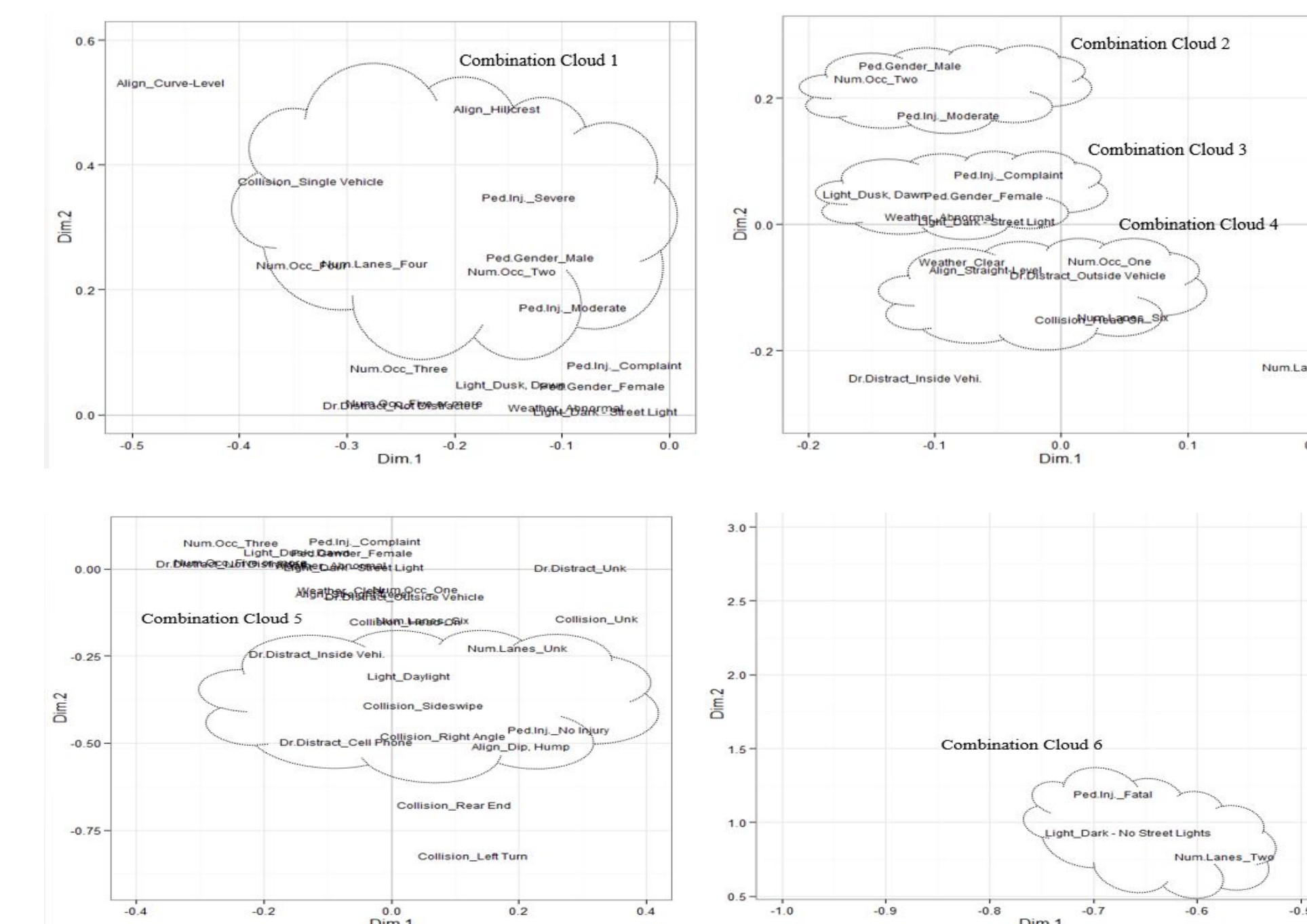


Figure 3: Combination Clouds.