# Determination of Factors Contributing to Road Traffic Accidents in South Africa Using Multiple Correspondence Analysis

Marion ADEBIYI, Ropo OGUNSAKIN, Timothy ADELIYI and Oludayo, O. OLUGBARA

*ICT and Society Research Group, Luban Workshop, Durban University of Technology, P.O box 1334, Durban 4000, South Africa*

Correspondence: mariona@dut.ac.za, +27680200413

**Abstract.** Despite the road safety global sustainable development goal of stabilizing the increased level of road traffic fatalities, road traffic accidents have become an intrinsic problem in the modern society. The problem of road traffic accidents in most developing nations of the world is because of the rapid advancement in social-economic, cultural and resources management. The primary objective of this study was to determine the factors that mostly contribute to the occurrence of road accidents in South Africa using the technique of multiple correspondence analysis on road traffic data samples collected between the years 2016 and 2017. Results of the multiple correspondence analysis on the study data have affirmed that among factors investigated, speed, location, and occasion mostly contribute to traffic road accidents in South Africa between the years 2016 and 2017. The validation of the analysis results has established 95% confidence ellipse for each level of the variable used. This study provides useful information that can aid informed decision making by road traffic authorities and other interested stakeholders to reduce road accident occurrences in South Africa.

**Keywords.** Correspondence analysis, road accident, road fatality, road traffic

## 1. Introduction

Road safety is a global concern and a social-economic issue of importance to many individuals. Global death from a road accident was estimated at 1.35 million in 2016 alone [1, 2] It causes the death of people in all age groups and most surprisingly, one of the leading causes of death of young adult and children from 5–29 years of age [3]. A report from the world health organization (WHO) has established that no reduction has been recorded in the number of road crash deaths in any low-income country from 2013 because death rate in such nations is about three-times higher than in the high-income countries [2]. Human social factors contribute largely to the high percentage of road crashes in a middle-income country, thus, an indication that road traffic accidents can be prevented or even avoided.

The challenges of road traffic accidents in this part of the world are because of the rapid advancement in socioeconomic, cultural and resource management. Hence, there has been studies emphasizing on road traffic accidents towards addressing several problems associated with road crashes. Some studies have analyzed multiple factors that lead to the degree of fatality of a road accident and causes of road accidents. These factors include, condition of a vehicle and type of vehicle in the road accident cases, injury level and death rates of patients of road accidents, condition of drivers and contribution of road users to road accident fatality. In addition, it involves many other interesting aspects that can provide a sustainable solution that would help to significantly reduce or halt road accident fatalities in any country. In South Africa, for example, a total of point estimate deaths of 14,507 have been recorded from road crashes in 2016 alone [1]. However, one of the proclaimed decades of action for road safety of the sustainable development goals was to stabilize and then half the rapid increase in the level of road traffic fatalities around the world by 2020 [4]. Consequently, improving the socioeconomic conditions of citizens and rigorous search for other paramount factors that mostly contribute to road traffic fatalities are fundamental to any nation.

Previous studies have exploited different methods and techniques to investigate cases of road traffic accidents. Descriptive statistics alongside the frequency distributions of incidence of road accident by various vehicle types were analyzed in [5]. Several tables and graphs were used to represent the degree of accident fatalities with respect to vehicle type as the major cause of the accident in four cities in the North-Central region of Nigeria. The study findings provided useful deductions for authorities in charge of policy formulation and decision making [5]. Another study did a law enforcement assessment of factors contributing to road accidents [6] and concluded that recklessness and non-compliance of drivers to traffic rules are the major causes of road crashes. The authors in [7] studied various factors associated with sleepiness-related vehicle crashes in Finland and concluded that short sleep is a major contributor to road vehicle crashes in Finland. In [8], a random sampling technique was applied to assess the magnitude and determinants of road traffic accidents in Ethiopia. Their results revealed that most accidents occurred because of violation of traffic rules and regulations.

The authors in [9] expressed concern about Pedestrian safety, with an increasing number of road traffic accidents that have led to a high percentage of deaths in India. They studied the key factors influencing pedestrian crashes and attempted to identify the association between various contributing factors. They deployed the technique of multiple corresponding analysis (MCA) for the large and complex database from crashes in Tamilnadu state of India over nine years. They concluded in their study that lighting conditions, locations, pedestrian behavior, crossings, and physical separation are the most significant contributing factors to pedestrian crashes. In [10], wrong way driving (WWD) is a traffic safety problem with the rare occurrence but characterized by severe injuries and fatalities. Failure to control the impacts of key contributing factors has generated bias in previous studies, so MCA was used as a distribution-free technique. Five years of WWW crashes in Louisiana of United States of America was used to determine the key associations among the contributing factors of road crashes. They generated sixteen significant clusters, which include various categories that determined key factors such as different locality types, roadways at dark, no lighting at night, roadways with no physical separations, roadways with higher posted speed, roadways with inadequate signage and markings, and older drivers. The authors in [11] provided a novel methodology proposal that supports road traffic stakeholders and professionals to publish technical notes for supporting the activities that involve data collection and pieces of evidence of motor vehicle collision scenarios using low cost unmanned aerial vehicles.

Vorko-Jović et al. [12] presented a simple and bivariate analysis approach aimed at reducing injury incidence related to road traffic accidents by using, odds ratio, and confidence interval of 95% in determining road accident risks. However, because of the unsatisfactory explanation in the detailed causes of a road accident and injury prevention, Kim et al. [13] presented a mixed logit model in exploring the effects of various observable characteristics on driver injury severities conditions on a single-vehicle crash. However, there work focused solely on the heterogeneity hypothesis effects of age and gender. Chen et al, [14] improved on the approach from Kim et al. [13] by presenting a performance measuring technique to measure the multi-dimensional concept of road safety scientifically and systematically that can combine the multilayer safety performance indicators (SPIs) into an overall index using the improved entropy TOPSIS–RSR (Road Safety Risk) method. Bachoo et al. [15] presented a cross-sectional survey design approach in analyzing the influence of anger, impulsivity, sensation-seeking, and driver attitudes on risky driving behavior among postgraduate university students in Durban, South Africa that lead to a road traffic accident.

Dadashova et al. [16] proposed a methodological development approach for model selection of significant predictors, explaining fatal road accidents by addressing both explanatory variables and adequate model selection issues. Their approach was carried out by combining neural network design and statistical approaches. Despite the advantages, it proffers in identifying factors of road

ISSN: 0868-4952

accident fatalities. However, they claimed that their approach is ongoing research with limitations that lead to high computational cost and processing time. In the literature there have been different proposed approaches in analyzing road traffic accidents. However, these research works have emphasized the aftermath of accidents and not on factors that cause the road traffic accident itself. Furthermore, there is no record in the literature to the best of our knowledge addressing the gap in determining road accident causation factors in South Africa bawd on scientifically proven data analysis techniques such as the MCA. Consequently, this study presents the MCA dimension of factors that mostly contribute to road accident in South Africa.

## 2. Material and Methods

*2.1 Data*

The study dataset is a set of categorical data obtained from a public domain of samples of accident crashes collected in an in-depth investigation of road traffic accidents over a period of 2016-2017 across the nine provinces in South Africa (https://www.kaggle.com/velile/car-accidents/data). The sample variables of 360 dataset instances include vehicle type, number of vehicles involved in the accident, number of casualties, speed zone, driver speed, surface condition, time of the day, date and year of accident, occasion of accident occurrence (normal day, public holidays), city and street name, accident location area (residential area, industrial area, highway area), driving against traffic, accident severity. A total of 360 cases was considered as a sample instance of 45 Bakkie, 45 Bicycle, 45 Bus, 45 Minibus, 45 Minibus taxi, 45 motorcars, 45 motorcycles and 45 truck for quantitative MCA analysis.

*2.2 Statistical analysis*

Data were analyzed using SPSS version 26.0 and R software to piggyback on the simplicity of MCA implementation provided in these tools. Descriptive statistical analysis was initially performed aiming to characterize the sample. Then MCA was performed on the study dataset to assess the possible associations among categories of the variables investigated. The MCA is a multivariate statistical technique of interdependence with an exploratory character that is indicated for a situation in which the researchers of this study wish to analyze a categorical dataset with many inherent variables [17]. MCA has been applied in this study because of its strength in analyzing an in-depth investigation described by a set of nominal variables. It provides a graphically visualizable organizational structure for variables and categories in a dimensional space that is useful for identifying patterns in a dataset and associations among the parameters investigated [18]. In MCA, it is possible to graphically represent the most important relationships associations among categories of variables and highlight groups of individuals with specific profiles. Discrimination measures were employed in the analysis to indicate the variables that are most relevant to the formation of each dimension, and centroid coordinates to help locate each category on the perceptual map [17]. In addition, the analysis calculates the inertia and eigenvalue for each dimension to reflect how much of the total data variability is being explained [18]. Hence, the present study considered a two-dimensional solution as most appropriate for the determination of factors contributing to road traffic accidents in South Africa.

## 3. Results and Discussion

This study presents the MCA results graphically to help in simplifying the process of interpreting the relationships that may exist among variables [19]. The list of all the variables investigated along with their frequency and percentage are presented in Table 1. These variables are chosen based on the findings by other researchers that were documented in the literature. MCA dimensions and qualitative variables were represented by the coefficient of determination ($R^2$) and P-value of the overall test performed.

**Table 1: Characteristics of data variables utilized in this study**

| Variable | Category | Frequency | Percentage |
|---|---|---|---|
| Province | Durban | 56 | 15.6 |
| | Eastern Cape | 47 | 13.1 |
| | Free state | 24 | 6.7 |
| | Gauteng | 37 | 10.3 |
| | Limpopo | 10 | 2.8 |
| | Mpumalanga | 49 | 13.6 |
| | Northwest | 33 | 9.2 |
| | Northern Cape | 64 | 17.8 |
| | Western cape | 40 | 11.1 |
| Time of the day | 0:00 to 1:59 | 30 | 8.3 |
| | 2:00 to 3:59 | 30 | 8.3 |
| | 4:00 to 5:59 | 30 | 8.3 |
| | 6:00 to 7:59 | 30 | 8.3 |
| | 8:00 to 9:59 | 30 | 8.3 |
| | 10:00 to 11:59 | 30 | 8.3 |
| | 12:00 to 13:59 | 30 | 8.3 |
| | 14:00 to 15:59 | 30 | 8.3 |
| | 16:00 to 17:59 | 30 | 8.3 |
| | 18:00 to 19:59 | 15 | 4.2 |
| | 20:00 to 21:59 | 30 | 8.3 |
| | 22:00 to 23:59 | 30 | 8.3 |
| Year | 2016 | 120 | 33.3 |
| | 2017 | 240 | 66.7 |
| Speed zone | 60km | 99 | 27.5 |
| | 80km | 168 | 46.7 |
| | 120km | 93 | 25.8 |
| | Easter | 99 | 27.5 |

| | | | |
|---|---|---|---|
| Occasion | Festive | 60 | 16.7 |
| | Normal days | 129 | 35.8 |
| | Weekends | 72 | 20.0 |
| Location | Highway area | 93 | 25.8 |
| | Industrial area | 168 | 46.7 |
| | Residential area | 99 | 27.5 |
| One way | No | 283 | 78.6 |
| | Yes | 77 | 21.4 |
| Speed (km) | 120km/k | 168 | 46.7 |
| | 140km/h | 93 | 25.8 |
| | 90km/h | 99 | 27.5 |
| Accident severity | Bumper accident | 99 | 27.5 |
| | Fatal accidental | 93 | 25.8 |
| | Head-on accident | 168 | 46.7 |
| Surface condition | Curve and Hilcrest | 39 | 10.8 |
| | Curve and level | 79 | 21.9 |
| | Straight and Hilcrest | 49 | 13.6 |
| | Straight and level | 71 | 19.7 |
| | Straight with upgrade | 48 | 13.3 |
| | Straight level and upgrade | 74 | 20.6 |

The categorical variables sharing similar features are located near each other and this is demonstrated in a two-dimensional graphical display of the data as shown in Figure 1 [19, 20]. The magnitude of information related to each dimension is termed eigenvalue [19, 21], whose values range between 0 and 1 to indicate the total variance among the variables. Based on the findings of this study, it was established that the first and second dimensions exhibited greater eigenvalues than the other dimensions. The first and second dimensions as presented in Table 2, have eigenvalues of 0.53 and 0.49 respectively and altogether both explain about 24% of the data variability.

**Table 2.** Eigenvalues and variances of top five dimensions

| | Dimension | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Eigenvalue | 0.532 | 0.494 | 0.267 | 0.224 | 0.204 |
| Percentage of variance (%) | 12.375 | 11.496 | 6.219 | 5.208 | 4.750 |
| Cumulative percentage of variance | 12.375 | 23.871 | 30.090 | 35.298 | 40.048 |

Furthermore, the associations among all the study variables are shown in Figure 1. Based on the findings, it can be established that those variables located to each other exhibited similar contributions, indicating that they show a relative same contribution to the total explained variances.

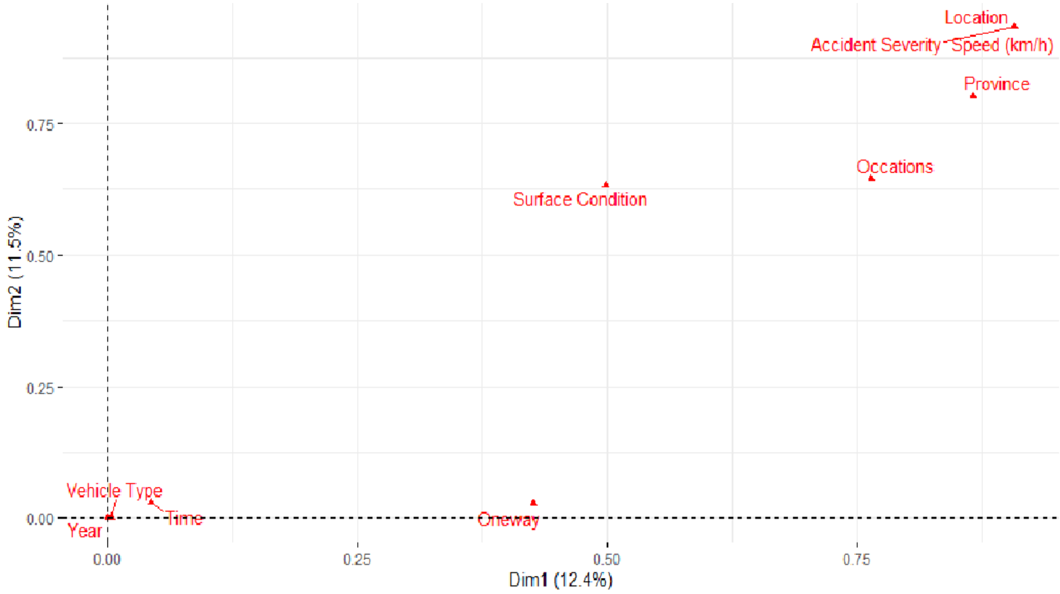


**Figure 1: MCA plot of study variables**

In addition, for each dimension, the farther a point from the center of the map, the greater is its contribution to the eigenvalue of the dimension considered. Hence, the MCA result affirms that the variable that contributed most to the first dimension were "speed", "location" and "accident severity" while that of the second dimension were "province", "occasion" and "accident severity". The contributing factors to road traffic accidents in South Africa as reported in this study are presented in descending order of importance. Table 3 lists the significance of each of the variables in both dimensions for $R^2$ and the P-value of the overall test. As a rule of thumb, the $R^2$ value ranges from 0 to 1 such that 0 implies no association and 1 indicates a very strong association between the MCA dimension and a variable. The risk of accident crashes as shown in Table 3 is not associated with vehicle type, year, speed zone and time.

**Table 3: Significance of key variables on the first plane**

| **Dimension 1** | **$R^2$** | **P-value** |
|---|---|---|
| Speed (km/h) | 0.907 | <.0001 |
| Location | 0.904 | <.0001 |
| Accident severity | 0.899 | <.0001 |
| Province | 0.845 | <.0001 |
| Occasions | 0.763 | <.0001 |
| Surface condition | 0.498 | <.0001 |
| One-way | 0.426 | <.0001 |
| **Dimension 2** | **$R^2$** | **P-value** |
| Province | 0.802 | <.0001 |
| Occasions | 0.756 | <.0001 |

| | | |
|---|---|---|
| Accident severity | 0.634 | <.0001 |
| Location | 0.508 | <.0001 |
| Speed (km/h) | 0.434 | <.0001 |
| Surface condition | 0.366 | <.0001 |

In addition, the contributions of the categories of variables in the top 2 dimensions in descending order of significance are shown in Figure 2, which depicts the category of variables that contributes most to road traffic accidents in South Africa.

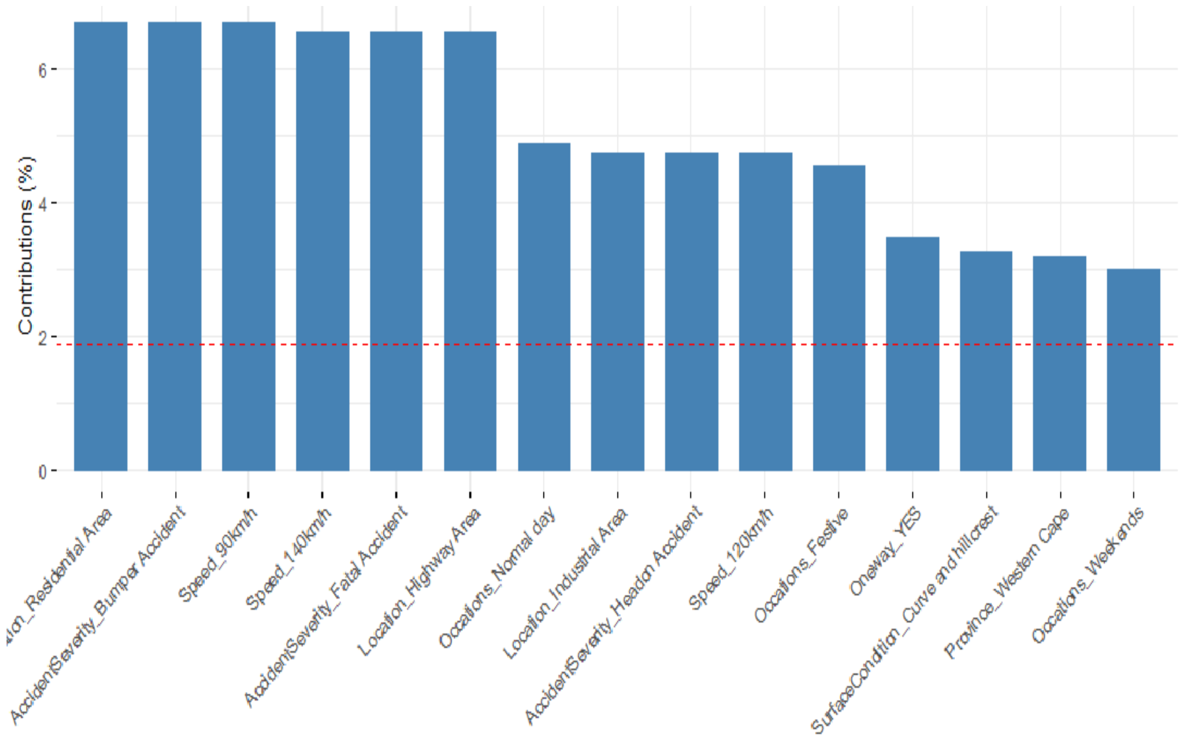


**Figure 2: Contributions of study variables in top two dimensions**.

It can be established from Figure 2 that location and occasion were both linked to dimensions 1 and 2 to signify the importance of these factors to the occurrence of road traffic accidents. The 95% confidence ellipse was established for each level of the variables to validate the categorical variables utilized in this study, as shown in Figure 3. It can be established from the figure that the confidence ellipse gives a better representation of the degree of uncertainty associated with the different point position.

ISSN: 0868-4952

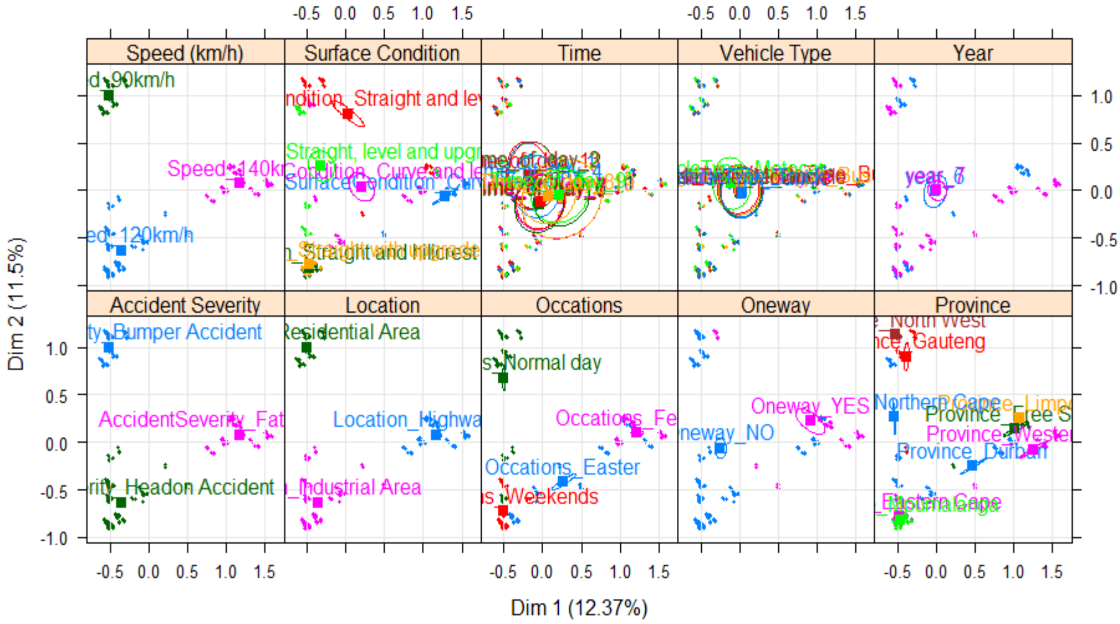


**Figure 3: Biplot of individuals and variables with confidence ellipses.**

Figure 4 shows that all the categories of variables as indicated in Figure 3, such as "time", "vehicle type", "surface condition" and "one-way" differ significantly from one another to indicate no overlapping on the confidence ellipse. Consequently, these variables can be grouped as the underlying factors that mostly contribute to road traffic accidents in South Africa.

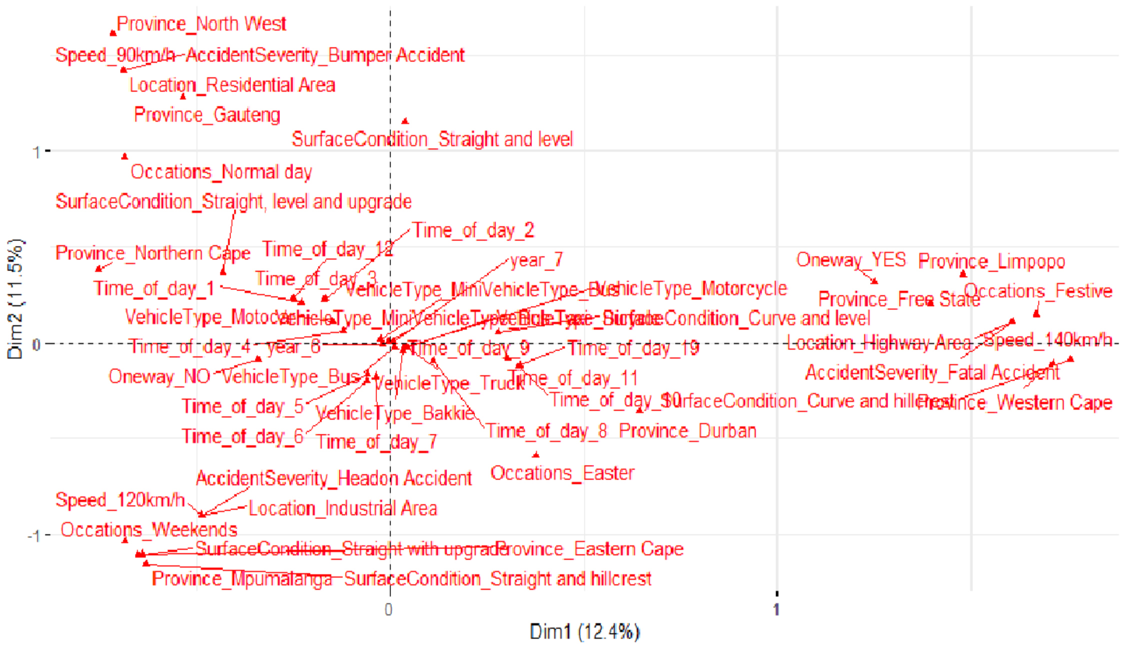


**Figure 4: MCA plot of all study categories**

## 4. Conclusion

This work has contributed significantly to the analysis of one of the most proclaimed decades of action for road safety of the sustainable development goals in South Africa by providing a wealth of information that can be used to reduce road traffic accidents. The exploited methodological technique of MCA has shown a stimulating result for obtaining information that can aid road traffic agencies to reduce the recurring cases of road traffic accidents. The analysis as presented in this study elucidates the factors that mostly contribute to the incidences of road traffic fatalities across the provinces in South Africa between 2016-2017. The analysis specifies less severe but important factors that feature less in the influence of road traffic accidents based on the $R^2$ and p-values of the MCA analysis. Finally, the in-depth analysis of the study dataset is of great importance because it propels the avowal that vehicles that are constantly maintained can yield control at an emergency, whereby traffic road fatalities may be predictable, controllable, reduced or even can be avoided completely.

**Acknowledgement**

The appreciation of the first and second authors goes to the Research and Postgraduate Directorate of Durban University of Technology, Durban in South Africa for hosting the Postdoctoral Fellows and to the ICT and Society Research Focus Area for funding this research.

## References

1. Global status report on road safety 2018. *Geneva: World Health Organization*; (2018). Licence: CC BYNC-SA 3.0 IGO.
2. Sadik-Khan, J. and Solomonow, S., 2020. Mean streets: The global traffic death crisis. *Foreign Aff.*, *99*, p.141.
3. World Health Organization. Disease, injury and causes of death country estimates, 2000–2015 [Internet]. World Health Organization. (2017). Available from https://www.who.int/healthinfo/global_burden_disease/estimates_country_2000_2015/en/.
4. South African National Road Safety Strategy 2016 – 2030, published in 2017. https://www.gov.za/documents/national-road-safety-strategy-2016-2030-22-may-2017-0000.
5. Adeniyi A.E, Olagunju M, Adebiyi M.O, Ogundokun O.R, Lukman A.F, Amusan A.E, Ayegba P.O and Ehiedu P.A, Vehicle-caused road accidents of four major cities in North-Central region of Nigeria. (2010–2015), *International Journal of Civil Engineering and Technology*, **10** (2019), 124-134.

ISSN: 0868-4952

6. Jonathan J.R, Shirley R, Salissou M and Aidan F, what are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records, *Road Accident Analysis*, **115** (2018), 11-24.
7. Juhani K, Timo T, Adel B, Markku P, Sleep versus non−sleep-related fatal road accidents, *Sleep Medicine*, **51**(2018), 148-152.
8. Woldu A.B, Desta A.A, Woldearegay T.W. Magnitude and determinants of road traffic accidents in Northern Ethiopia: a cross-sectional study. *BMJ Open***10 (**2020). doi: 10.1136/bmjopen-2019-034133
9. Sivasankaran, S.K and Balasubramanian, V., 2020. Investigation of pedestrian crashes using multiple correspondence analysis in India. *International journal of injury control and safety promotion*, *27*(2), pp.144-155.
10. Das, S., Avelar, R., Dixon, K. and Sun, X., 2018. Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis & Prevention*, *111*, pp.43-55.
11. Pádua, L., Sousa, J., Vanko, J., Hruška, J., Adão, T., Peres, E., Sousa, A. and Sousa, J.J. Digital reconstitution of road traffic accidents: a flexible methodology relying on UAV surveying and complementary strategies to support multiple scenarios. *International journal of environmental research and public health*, *17*(6), (2020),1868.
12. Vorko-Jović, A., Kern, J. and Biloglav, Z., 2006. Risk factors in urban road traffic accidents. Journal of safety research, 37(1), pp.93-98.
13. Kim, J.K., Ulfarsson, G.F., Kim, S. and Shankar, V.N., 2013. Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. Accident Analysis & Prevention, 50, pp.1073-1081.
14. Chen, F., Wang, J. and Deng, Y., 2015. Road safety risk evaluation by means of improved entropy TOPSIS–RSR. Safety science, 79, pp.39-54.
15. Bachoo, S., Bhagwanjee, A. and Govender, K., 2013. The influence of anger, impulsivity, sensation seeking and driver attitudes on risky driving behaviour among post-graduate university students in Durban, South Africa. Accident Analysis & Prevention, 55, pp.67-76.
16. Dadashova, B., Arenas-Ramírez, B., Mira-McWilliams, J. and Aparicio-Izquierdo, F., 2016. Methodological development for selection of significant predictors explaining fatal road accidents. Accident Analysis & Prevention, 90, pp.82-94.
17. Abdi, H. and Valentin, D., Multiple correspondence analysis, *Encyclopedia of measurement and statistics*, **2** (2007), 651-66.
18. Ayele, D., Zewotir, T. and Mwambi, H., Multiple correspondence analysis as a tool for analysis of large health surveys in African settings. *African health sciences*, **14** (2014), 1036-1045.

19. Das S, Sun X. Factor association using multiple correspondence analysis in vehicle-pedestrian crashes. Transportation Research Record: *Journal of Transportation Research Board*, **2519** (2015), 95–103.
20. LeRoux B, Rouanetm H. Multiple Correspondence Analysis, *SAGE Publications, Inc.: United State*, (2010).
21. Das S, Sun X. Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research*, **39** (2016), 146–155.