PROSPECTS FOR USING CORRESPONDENCE ANALYSIS TO CHARACTERISE TRAVEL DEMAND FOR AUTHORITIES WITH LIMITED TRANSPORT MODELLING RESOURCES

M NKOSI, P MOKILANE and M MOKONYAMA

Council for Scientific and Industrial Research (CSIR), Built Environment, Meiring Naude Road, Brummeria; Email: mnkosi@csir.co.za; Tel: 012 841 3991

ABSTRACT

The estimation of travel demand for complex passenger transport infrastructure projects such as Bus Rapid Transit (BRT) systems, is fraught with multiple challenges that include the reliability of demand forecasts. The choice of variables used to estimate the demand is often informed by historical perspectives and not a genuine understanding of the said variables. The paper explores the use of Correspondence Analysis (CA) as an alternative method to facilitate the understanding the relationship between demand and supply side variables prior to the application of more intense mode choice modelling techniques. The raw data from the 2014 Gauteng household travel survey was used to develop the required models in areas where BRT services were operational in order to characterise the BRT users, followed by probability estimation of the actual demand. It is shown that CA can be reliably used as a resource efficient alternative for the initial characterisation of travel demand where household travel survey data exist. This is especially useful for planning authorities that do not have large transport planning budgets and human resources to acquire and employ extensive land use-transport interaction models.

Keywords: Correspondence analysis, choice modelling, decision support tools, planning authorities

1. INTRODUCTION

Initial characterisation of travel demand is important to help understand the nature of the demand in an area prior to undertaking more in-depth analyses of mode choice. This further helps reduce assumptions that often compromise the efficacy of choice models e.g. mode, agent or activity. The initial characterisation of demand would especially be useful when an analyst has large dataset(s) from a typical household travel survey, and seeks to understand sensitivity of mode choice to the multiple variables in the survey. This requires an understanding of both the qualitative and quantitative aspects of travel demand. For example, in household travel surveys, safety and security is among the most cited reasons for not using public transport. However, such variables are hardly used in travel demand modelling. This subsequently limits the versatility of such models. It is thus important to have an initial understanding of how the different variables relate even when fewer variables are used for demand modelling.

Multivariate techniques such as Principal Components Analysis (PCA), Correspondence Analysis (CA), and Correlation Analysis (CA) are useful for exploring relationships between variables in big datasets. Basically, these multivariate techniques are able to create associations between the variables, for example, variable(s), e.g. through clustering and interrogating patterns of interest.

The focus of this paper is on Correspondence Analysis. It investigates the use of Correspondence Analysis, supplemented by Logistic Regression (LR), in order to showcase the initial characterisation of travel demand. The 2014 Gauteng household travel survey dataset is used in the analysis, but limited to the City of Johannesburg. Particular focus is placed on the characterisation of users of the Bus Rapid Transit (BRT) mode. The BRT mode is chosen because of its increased popularity as a low-cost mass public transport solution.

2. BACKGROUND

Public transport has a significant role to play in terms of facilitating social inclusion and/or access in line with user constraints of budgeted time and money. This is especially important in South Africa given that a significant proportion of the population is classified as “captive” to public transport. The public transport services are also provided in an environment characterised by complex interaction between trip making and land use, which in turn warrants investment in decision support systems.

Decision support tools currently used in transport planning tend to be resource intensive, especially in relation to the required data inputs. In South Africa, there are issues concerning completeness and quality of input data, which in turn affects the development of robust transport models (Wray, et al., 2013). For example, the actual demand for BRT type services in South African cities have been found to be consistently much lower than modelled values (Scorcia & Ramon, 2017).

While the above problems are common across many parts of South Africa, they are especially pronounced in the under-resourced municipalities. Limited resources to plan and to implement largely result in questionable public transport plans, thus creating a vicious circle of poor public transport service delivery.

The paper investigates the extent to which such shortcomings can begin to be addressed through the use of Correspondence Analysis (CA). The CA technique is capable of profiling individual choices in terms of variables characterising the individuals, and to visually present the information for ease of interpretation. While Diana & Pronello (2010) have previously demonstrated CA in mode choice modelling, in general, both locally and internationally, there has been little research towards the application of CA in transport planning. Nevertheless, CA has been applied more widely in the field of road safety and driver behaviour.

Although the case study adopted in this paper is that of a metropolitan municipality, the principles which are discussed are still applicable to other municipalities that have limited resources. The study area was selected for the following reasons;

- The municipality has newer public transport services e.g. BRT and Gautrain for testing.
- Public transport users are exposed to a wider set of public transport modes.

3. THEORETICAL CONSIDERATIONS ON CORRESPONDENCE ANALYSIS

Hair, et al. (1998) describe CA as a compositional technique because the perceptual map it generates is based on the association between objects and a set of descriptive characteristics or attributes specified by the researcher. Generally, the goal of CA is to uncover the important associations among categories of the response variable with those of predictor variables using graphical representations (Benzécri, 1992).

CA is an attractive classification technique because no underlying distributional assumptions are required, thus accommodating any type of categorical variable such as binary, ordinal or nominal. The technique examines the relationships between categories of binary, ordinal or nominal data in a contingency table. It further uses a Chi-square technique to standardise the response variable and form the basis for establishing the inherent associations. The measure of association used in CA is the chi-square distance between the response categories (Clausen SE, 1998). The Chi-square is a standardised measure of the actual cell frequencies compared to expected cell frequencies. The Chi-square is calculated as:

$$\text{Chi} - \text{square} = \frac{(\text{expected cell frequency} - \text{actual cell frequency})^2}{\text{expected cell frequency}}$$

where, the expected value is given as:

$$\text{expected} = \frac{\text{total column frequency } \mathbf{x} \text{ total row frequency}}{\text{overall frequency}}$$

CA is extendable to Multiple Correspondence Analysis (MCA) where more than one predictor variables are considered. Unfortunately, the limitation of correspondence analysis is that it cannot be used to evaluate or calculate the probabilities of passengers choosing modes of transport. It is a technique used to uncover the associations between the categories of modes of transport and categories of variables characterising passengers such as their age categories or gender and categories of the attributes of modes of transport.

4. CASE STUDY APPLICATION

CA is used in the paper to profile passengers in terms of their choices of modes. In addition, for the purpose of completeness, Logistic Regression is used to estimate the probable number of passengers to use a mode of travel (in this case BRT) when presented with alternatives. The logistic regression uses variables generated from CA to estimate the number of passengers who are more likely to use BRT relative to other modes of transport available to them. The logistics regression is formulated as:

$$p = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} \tag{1}$$

Where p is the predicted output (probability), β_0 is the intercept term and β_1 is the coefficient for input value (x). x could be the age of the passenger or the monthly salary. If the predicted probability (p) is less than 0.5 then the passenger is less likely to use BRT otherwise the passenger is more likely to use BRT.

The mode of transport is used as the dependent variable (1 representing customers using BRT services and 0 representing customers using other modes of transport available to them). The technique is extendable to multinomial logistic regression (MLR) in cases where the dependent variable has three or more categories (polytomous as opposed to dichotomous as in this case).

CA, in this case study, profiles users of different modes of transport (i.e. who uses what mode of transport?) and also profiles the modes of transport themselves (what characterises each mode of transport?) in the area of Johannesburg.

A simple example of correspondence analysis is presented in Figure 1 where the association between modes of transport and level of education is examined. The mode of transport in Figure 1 is considered to be mostly associated with the categories of the "highest level of education" based on their closeness to it on the map.

In Figure 1, passengers who have diplomas and/or degrees as their highest levels of education are more associated with the following modes of transport:

- Gautrain bus
- Gautrain
- Motorcycle, and
- Car as a driver.

Therefore, passengers using these modes of transport have a degree and/or diploma as their highest level qualification. Passengers with a primary school qualification or no qualification are associated with the following modes of transport:

- Car as a passenger
- School bus, and
- Walk all the way.

These are mostly school-going children.

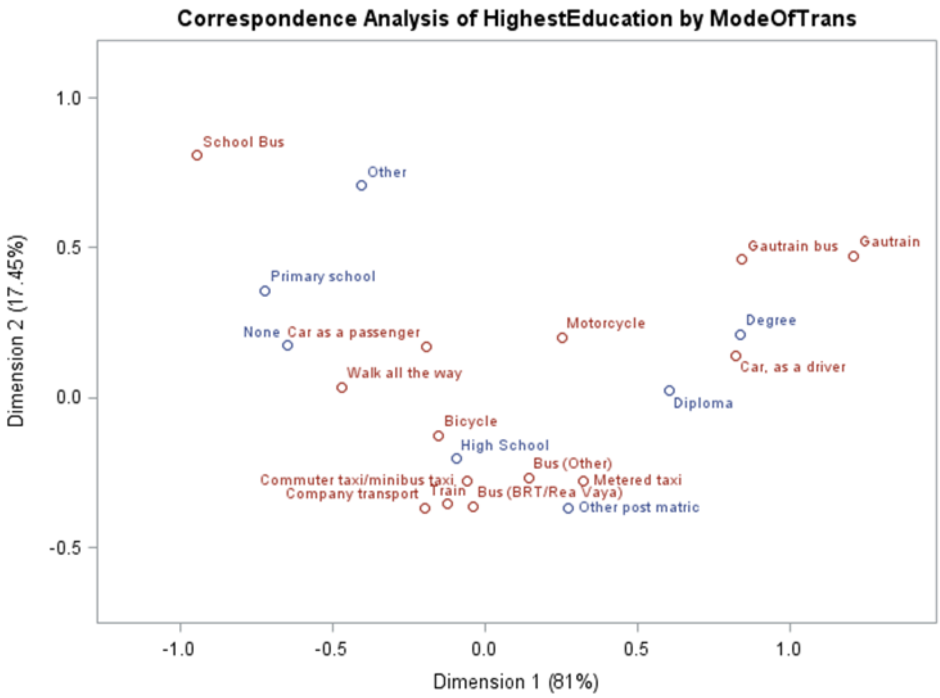

Figure 1: Correspondence analysis of mode of transport by level of education

Figure 2 shows that, passengers whose salaries range between R6 001 and R30 000, aged 40 and above, and have attained degrees and/or diplomas as their highest qualifications are mostly associated with "car as drivers", "motorcycle" and "Gautrain" as their modes of transport.

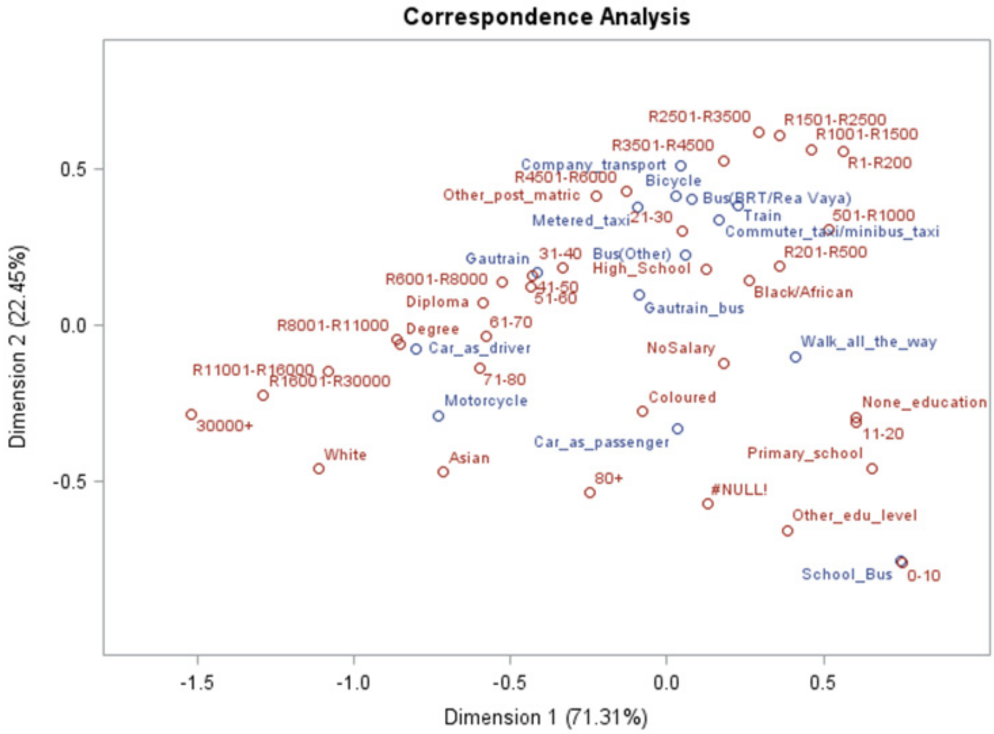

Figure 2: Correspondence analysis of mode of transport by level of education, salary and race

To estimate demand for BRT, the study area was limited to a section of the network where public transport users have reasonable access to all four modes i.e. Bus, Taxi, Train, and BRT using a geographic information system (GIS) platform. Passenger trips in the household travel survey were limited to trips originating and destined within the area of access for the four modes. Given that the resulting sample was low (given that the household travel survey did not specifically target BRT users), the results in the paper are for illustrative purposes.

Table 1 shows the confusion matrix for multiple logistic regression. Both the actual reported modes of travel and the predicted model results are provided. For modes in which the sample was high, the model performed well, for instance, the minibus taxi mode was able to give a positive-positive (96%) and for the train service (82%). In terms of the bus mode (58%) and the Bus Rapid transit (BRT) mode (32%), the model performed moderately.

Table 1: Confusion matrix

		Predicted mode			
Actual mode		Bus (other)	Taxi	Train	Total
	Bus (other)	32 (58%)	21 (38%)	2 (4%)	55
	Taxi	12 (3%)	64 (96%)	7 (2%)	83
	Train	0 (0%)	16 (18%)	74 (82%)	90
	BRT	20 (32%)	38 (61%)	4 (7%)	62

From an analysis of Table 1 and interpretation of study findings, there are observable limitations in the methodology. Firstly, the assumed access may not necessarily be available. For example, there may be a route but the frequency of the service could be such that for specific individuals, services may not be available. Secondly, like in other methods, the technique is highly dependent on the sample size for reliability.

5. CONSIDERATIONS FOR APPLICATION

For transport planning purposes, before constructing expensive infrastructure on which such modes as BRT can be operated, it is prudent for planners to ensure that an adequate number of people reach the threshold ridership passenger population required in that area that would make use and ensure sustainability of the planned service. CA is therefore, an important modelling technique for transport planners because it explores the characteristics of people using different modes of transport in relatively more detail. The question that the transport planner would then need to answer is the extent to which people in the area under consideration possess “characteristics” of typical BRT passengers. Where few people possess such characteristics, then the risk of experiencing low patronage may be high.

6. CONCLUDING REMARKS

The paper aimed to profile the use of Correspondence Analysis to both profile passengers in terms of their choices based on variables characterising both the individual and the service, and to facilitate the forecast demand for specific modes.

Many municipalities in South Africa have limited capacity to carry out planning for public transport services and the currently available techniques and methods are resource intensive i.e. travel demand models. Also, important to note are the long-term financial and social impacts of implementing a public transport service when decisions are not informed by robust tools. To this end, it was important to explore methods that use existing databases such as the National Household Travel Survey to facilitate decision making.

A case study showcasing application was presented. From the case study, albeit with a small sample size, the potential application of the CA technique, coupled with logistic regression, was demonstrated. It is argued that Correspondence Analysis should feature more in the toolkits of transport planners. Although, the discussed results are indicative in nature, they do however provide government officials and planners with the tool to estimate the modal split. This is especially useful when municipalities intend to introduce new services, and want to estimate the likely demand to use the service should adequate access be provided.

7. REFERENCES

Benzécri, JP, 1992. Correspondence Analysis Handbook. New York: Marcel Dekker.

Clausen, SE, 1998. Applied Correspondence Analysis. Thousand Oaks, CA: Sage.

Diana, M and Pronello, C, 2010. Traveler segmentation strategy with nominal variables through correspondence analysis. Transport Policy, 17(3), pp. 183-190.

Hair, JF, Anderson, RE, Tatham, RL and Black, WC, 1998. Multivariate data analysis. Upper Saddle River.

SACN, 2014. How to build Transit Oriented Cities: Exploring possibilities, Johannesburg: South African Cities Network.

Scorcia, H. and Ramon, MR. 2017. Why South African cities are different? Comparing Johannesburg Rea Vaya Bus rapid Transit system with its Latin American Siblings. [Online] Available at: http://pubdocs.worldbank.org/en/900101490196071185/Draft-Manuscript-Why-South-African-BRTs-are-different-ForConnectNote-Feb.pdf. [Accessed 29 June 2018].

StatsSA, 2013. National Household Travel Survey, Pretoria: Statistics South Africa.

Venter, C, 2016. Are we giving BRT passengers what they want? User preference and market segmentation in Johannesburg, Pretoria: Proceedings of the 35th Southern African Transport Conference.

Wray, C, Musango, J, Damon, K and Cheruiyot, K, 2013. Modelling urban spatial change: A review of international and South Africa modelling initiatives, Braamfontein: Gauteng City Region Observatory.