Paper 3144-2015

"Why Do People Choose to Drive over Other Travel Modes?" Interpreting Multidimensional Scaling Dimensions with SAS®

Jun Guan Neoh, Centre for Risk Research, University of Southampton

Maxwell Chipulu, Centre for Risk Research, University of Southampton

ABSTRACT

We demonstrated a method of using *SAS®* to ease the interpretation of dimensions of a Multidimensional Scaling (MDS) model, a process which could be difficult without *SAS®*. We used the *PROC LOGISTIC* function of SAS to identify the variables which are strongly associated with each dimension, thus greatly aiding our interpretation procedure. In our demonstration, we examined the reasons that motivate people to choose to drive to work over other travel modes. We were able to extract the drivers' motivation in four dimensions, namely responsibility for family, issues with public transportation, need to use the car for other tasks, and location. The paper is relevant to MDS users, who may find it useful to utilize *SAS®* to analyze MDS results.

INTRODUCTION

Reduced car use is associated with public benefits such as reduction in carbon emission and congestion (Greene and Wegener, 1997). For transportation planners aiming to persuade drivers to use other travel modes, it is in their interest to understand "what motivates people to choose to drive over other travel modes?"

We attempt to answer this question with Multidimensional Scaling (MDS) analysis, a statistical method for measuring similarity/dissimilarity of cases in the dataset. MDS has the advantages of incorporating qualitative data into quantitative analysis and portraying the results on visual maps, thus making the findings more accessible to non-specialists (Neophytou and Molinero, 2004). A key output of the MDS model is the interpretations of the models' dimensions, which provides us with an overview of the data from different angles (Kruskal and Wish, 1978). However, the process of interpreting the dimensions can be challenging; the analyst has to look at the data points with extreme values on each dimension, and try to derive a common theme among the extreme positive and negative variables respectively (Kruskal and Wish, 1978). We will discuss how we use *SAS®*'s *PROC LOGISTIC* procedure to make this step easier.

MULTIDIMENSIONAL SCALING

Our MDS procedure is as follows. For our data, we used a dataset of self-reported reasons for driving to work collected from a bi-annual travel survey of employees of a UK university. In this survey, the respondents were asked to type their reasons for choosing driving as their travel mode to work (i.e. the responses were qualitative). We coded the responses and arranged them into a matrix of cases (respondent) by variables (reasons to drive), where for each "variable" of driving reason, a value of "1" is given if the respondent reports that reason, and "0" if the reason was not reported. **Table 1** shows an excerpt of the matrix.

Case #	CaringElderlyFamilyMember	CarPermit_inc	CarpoolWithColleagues	CarpoolWithFamily	CarryEquipment
7	0	0	0	0	0
8	0	0	0	0	1
9	1	0	0	1	0
10	0	0	0	0	1
11	0	1	0	0	0
12	1	0	0	0	1
13	0	0	0	0	0
14	0	1	1	0	0
15	0	0	1	0	0

Table 1: Matrix of driving reasons (excerpt)

We build our MDS model with *PROC MDS*, using the matrix as our input. Readers are pointed to the *SAS Institute* (2008) user's guide for detailed guidance on *PROC MDS*. To ensure goodness-of-fit of the MDS model, it is important to determine the number of dimensions to produce. This can be determined by an 'elbow test' of normalized raw stress values, followed by a model degeneracy test (see Chipulu *et al.*, 2013). MDS models are considered a "good" fit when stress levels are at 0.05 and a "very good" fit at 0.01 (Kruskal and Wish, 1978), with the caveat that higher dimensions may contain residual variation, making them harder to interpret (Neophytou and Molinero, 2004). Hence the user must decide whether to sacrifice model fit for dimension interpretability. In our case, we extracted six dimensions because stress levels were recorded at 0.01 indicating a very good fit, but decided to interpret the first four dimensions, as adding the 5th dimension brings only a marginal improvement to the model fit (by a mere 0.002 points).

Finally, we run the model: the MDS produced an output of final coordinates for the variables for each dimension, which are plotted on visual maps. Here, we attempt to interpret the meaning of each dimension. Previous MDS studies looked for common themes between variables located on the extreme sides of the dimension map axes (Mar Molinero and Xie, 2007), or examined the variables with coordinates having large (absolute) values (Chipulu *et al.*, 2013); the idea being that these variables are strongly associated to the respective dimension. As mentioned, this can be a difficult process: intuitively, it becomes increasingly harder to draw similarities when the dimensions have too many extreme variables to work with. Making matters more challenging, the analyst is expected to portray the theme of the positive variables and negative variables as opposites.

In our study, the MDS model produced a list of coordinates for 32 variables on six dimensions which could be plotted on three dimension maps, each reflecting two dimensions (**Figure 1, Figure 2** and **Figure 3**). *Dim_1, Dim_3* and *Dim_5* are interpreted by reading the maps vertically while *Dim_2, Dim_4* and *Dim_6* are deduced horizontally. As expected, interpreting the dimensions via the maps proved to be challenging. Take the Dim_1 versus Dim_2 map for example (**Figure 1**): While it was easy to find a common theme among the variables on the top axis (*CarpoolWithFamily*, *ChildrenEmergency* and *CaringElderlyFamilyMember* implies a "driver with family responsibility" theme), it was harder to draw similarities between the variables at the bottom (*Security*, *Weather*, *PublicTransportExpensive, ParkAndRide_inc*). Moreover, in order for the "family" interpretation to work, we would need to depict the bottom variables as opposites to the top ("driver without family responsibility"); this requires some justification. Likewise for *Dim_2*, it was difficult to find any similarities for the variables on the far left (*CarpoolWithColleagues, CarryEquipment*, and *Security*) and far right (*CaringElderlyFamilyMember PublicTransportAvailability_inc* and *ParkAndRide_inc*). This difficulty persists in the remaining dimensions from *Dim_3* to *Dim_6*. Similarly, dimension interpretation was equally difficult when examining variables with large absolute values from **Table 2**.

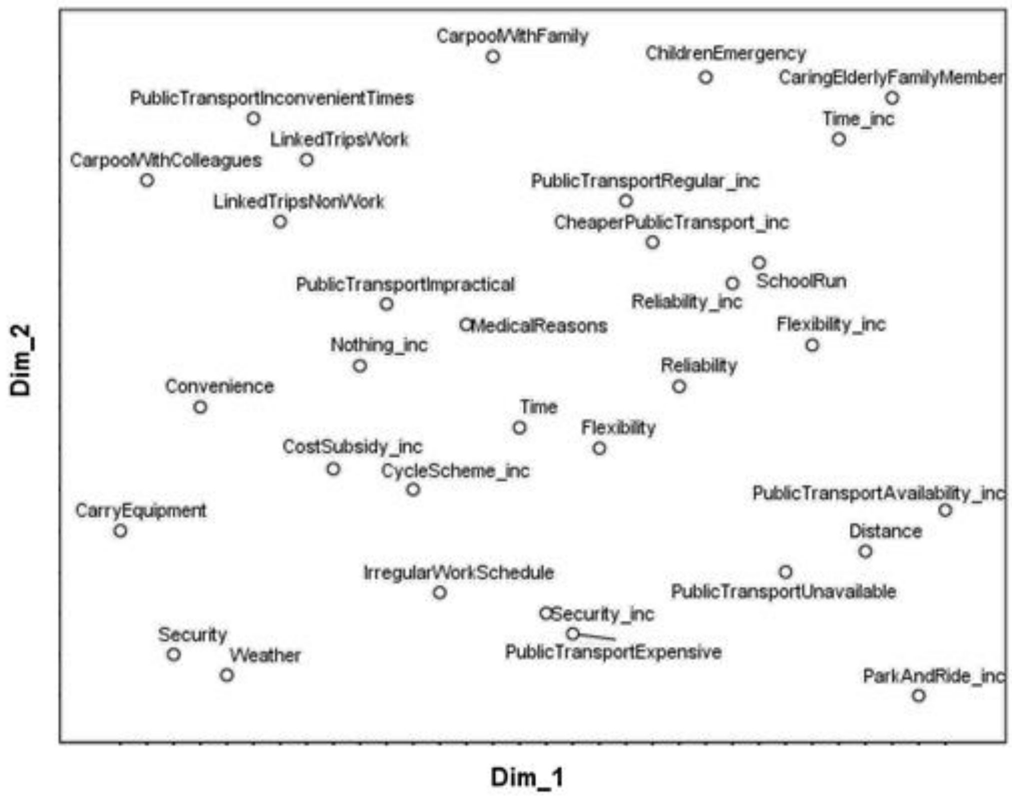

Figure 1: Map of Dim_1 versus Dim_2

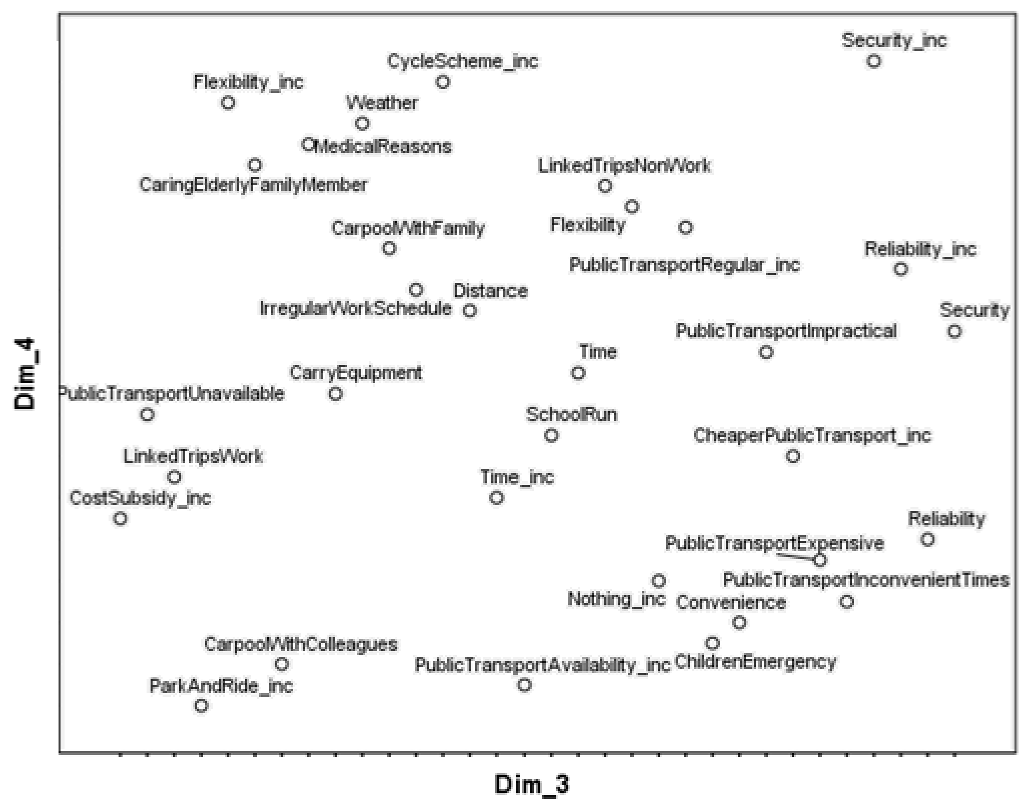

Figure 2: Map of Dim_3 versus Dim_4

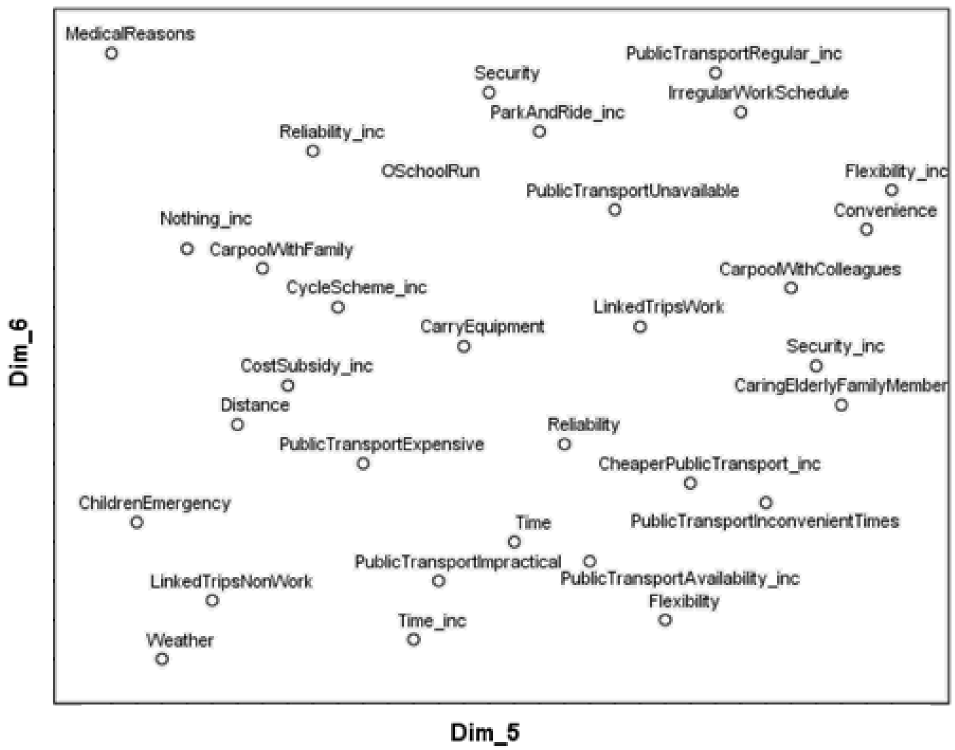

Figure 3: Map of Dim_5 versus Dim_6

	Dimension					
	1	2	3	4	5	6
CaringElderlyFamilyMember	.396	.435	-.317	.236	.362	-.013
CarpoolWithColleagues	-.775	.293	-.266	-.406	.298	.065
CarpoolWithFamily	-.027	.667	-.171	.136	-.208	.086
CarryEquipment	-.817	-.194	-.206	-.037	-.053	.005
CheaperPublicTransport_inc	.214	.211	.223	-.080	.205	-.076
ChildrenEmergency	.219	.440	.131	-.349	-.570	-.095
Convenience	-.341	-.053	.173	-.330	.373	.088
CostSubsidy_inc	-.178	-.118	-.436	-.133	-.208	.004
CycleScheme_inc	-.163	-.121	-.130	.516	-.125	.044
Distance	.395	-.279	-.071	.085	-.229	-.016
FlexiTime_inc	.333	.002	-.351	.516	.498	.121

Independence	.090	-.094	.003	.197	.185	-.259
IrregularWorkSchedule	-.127	-.302	-.133	.094	.232	.216
LinkedTripsNonWork	-.241	.239	-.013	.199	-.239	-.256
LinkedTripsWork	-.205	.307	-.398	-.095	.158	.023
MedicalReasons	-.040	.032	-.210	.250	-.719	.549
Nothing_inc	-.177	-.027	.051	-.163	-.245	.086
ParkAndRide_inc	.419	-.748	-.364	-.611	.049	.167
PublicTransportExpensive	.031	-.324	.233	-.154	-.120	-.059
PublicTransportImpracticalChanges	-.163	.084	.196	.004	-.062	-.186
PublicTransportAvailability_inc	.448	-.146	-.053	-.495	.073	-.149
PublicTransportInconvenientTimes	-.257	.371	.337	-.246	.287	-.091
PublicTransportUnavailable	.323	-.280	-.411	-.054	.092	.099
PublicTransportRegular_inc	.095	.282	.071	.193	.208	.336
Reliability	.216	-.040	.671	-.133	.072	-.045
Reliability_inc	.304	.117	.572	.133	-.196	.152
SchoolRun	.305	.145	-.017	-.076	-.065	.141
Security	-.359	-.369	.721	.038	-.026	.246
Security_inc	.001	-.306	.433	.517	.301	.005
Time	-.022	-.072	-.016	-.017	.000	-.103
Time_inc	.374	.352	-.055	-.122	-.064	-.423
Weather	-.271	-.503	-.196	.389	-.263	-.663

Table 2: Final coordinates of variables

Hence, we took a different approach to determine which variables have strong associations to each dimension: instead of relying on variables with large absolute values, we based our interpretations on variables which are statistically significant to each dimension. We ran a logistic regression model in

SAS®, treating the final coordinates of the dimensions, arranged in cases, as independent variables; and each “1” or “0” values for the driving reason variables as outputs. We ran the following syntax:

```
/*Export output*/
ods html file = "\\Folder\Output.xls";
ods html select RSquare (persist) ParameterEstimates (persist);

/*Macro for logit regression*/
%MACRO logit2 (depend);
Title2 "&Dim";
      PROC LOGISTIC DATA= Libref.MDSCoordinates_Matrix;
      Model &Dim   (event = last)= Dimension_1
      Dimension_2 Dimension_3 Dimension_4 Dimension_5
      Dimension_6/LINK = LOGIT RSq;
      RUN;;;;;
quit;

/*Logit regression for each variable as dependent*/
%MEND logit2;
      %logit2(Variable_1);
      %logit2(Variable_2);
      %logit2(Variable_3);
      %logit2(Variable_4);
      …
      %logit2(Variable_32);
ods html close;
```

The result is a table of maximum likelihood estimates to indicate the association of each variable with the respective dimensions. For demonstration purposes, we produced an excerpt of the results (instead of the full table) in **Table 3**.

	Statistically significant variables (*p< .05; **p< .01; ***p< .001)
Dimension 1	Care for elderly family member (47.9*); school run (10.3*).
Dimension 2	Public transport unavailable (-22.1***); public transport too many changes (10.38***); direct public transport needed (5.7***).
Dimension 3	Linked trips for non-work reasons (-3.1**); independence(-3.1***); Time (-9.1***); Public transport times unsuitable (9.16***).
Dimension 4	Distance (3.1*); time (2.8*); public transport unavailable (12.12***); public transport expensive (4.9***).

Table 3: Maximum likelihood estimates of the dimensions with self-reported driving reasons (excerpt).

With logit regression, we had more success with attaching meaning to the dimensions. As we concentrate on the statistically significant variables, the number of variables we need to pay attention to is reduced considerably, making our inferring task easier. Our interpretations for the first four dimensions are as follows. The strong and positive variables in Dimension 1 suggest a family-responsibility reason for driving. In dimension 2, there were many statistically significant variables ($p< .05$); we used higher critical values ($p<.001$ or $p<.01$) to narrow our scope, focusing on the variables with higher statistical significance. This led us to find that Dimension 2 is dominated by practicality issues with public transportation. For Dimension 3, the negative variables of ‘independence’ and ‘linked trips’ points towards task-based reasons for driving, i.e. these drivers need their car for other tasks during their commute. Dimension 4 describes drivers who have issues with distance, time, and public transport availability; alluding to a residence location-based reason for driving. In short, our MDS produced four dimensions, or themes, of self-reported reasons to drive: responsibility for family, issues with public transportation, need to use the car for other tasks, and location.

Upon completion of the dimension interpretations, we have the option to conduct further analyses in *SAS®*. For instance, as displayed in **Figure 4**, we could test the validity of our interpretations and examine the causal relationships between the dimensions. For our example, it may be useful for transportation planners to not only identify the reasons for driving, but to also understand how the elements of each dimension interact with each other and the commuters' travel choice. Such interactions and causal relationships can be examined via a Structural Equation Model (SEM) in *SAS®* (see Holtzman and Vezzu, 2011 for a guide). In this case, MDS could be used as an exploratory factor analysis (EFA) tool to identify the latent variables for the SEM analysis.

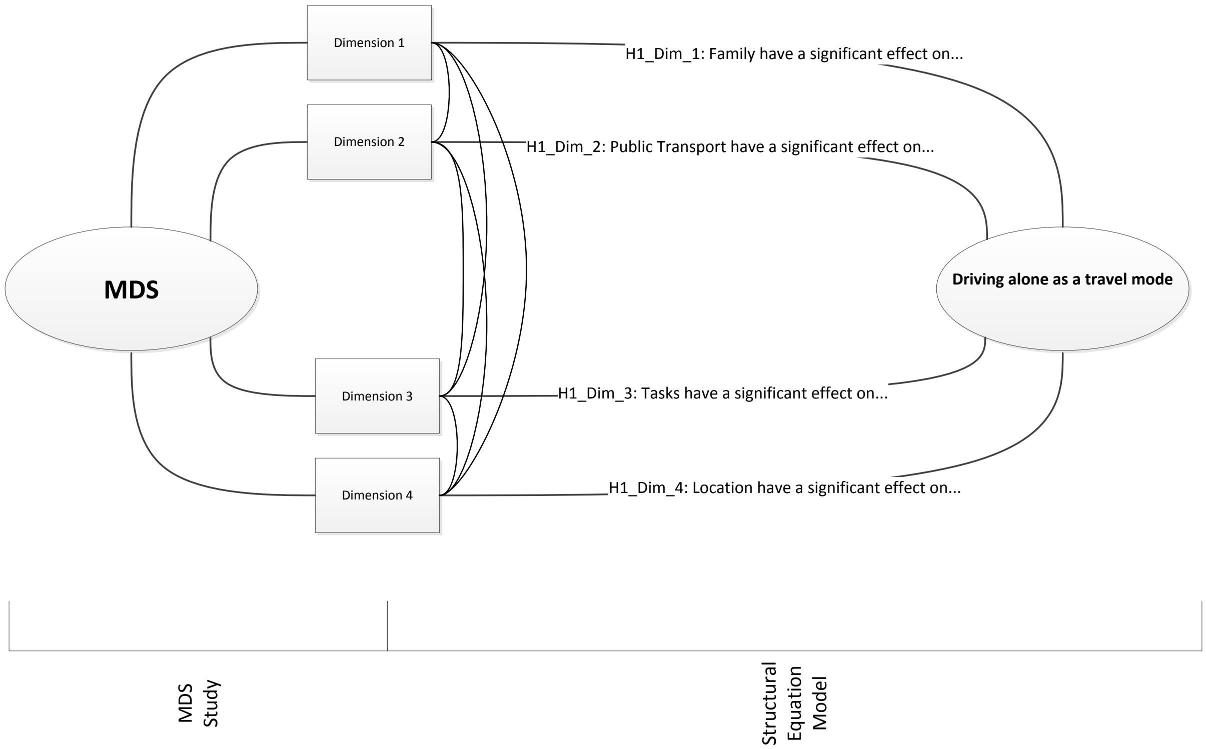

Figure 4: Validating the MDS model with SEM

CONCLUSION

This paper has demonstrated the usefulness of *SAS®* in carrying out multidimensional scaling analysis, particularly to ease the task of dimension interpretation with *PROC LOGISTIC*. We hope that we are able to encourage MDS applicants to use *SAS®*; likewise, we hope we are able to inspire *SAS®* users to consider MDS as a tool in tackling their research problems.

REFERENCES

Chipulu, M., Neoh, J.G., Ojiako, U. and Williams, T. (2013), 'A Multidimensional Analysis of Project Manager Competences', *IEEE Transactions on Engineering Management,* Vol. 60, No. 3, pp. 506-517.

Greene, D.L. and Wegener, M. (1997), 'Sustainable transport', *Journal of Transport Geography,* Vol. 5, No. 3, pp. 177-190.

Holtzman, S. and Vezzu, S. (2011), 'Confirmatory Factor Analysis and Structural Equation Modeling of Noncognitive Assessments using PROC CALIS', *24th Annual Northeast SAS Users Group (NESUG) Proceedings*, 2011, Portland, ME.

Kruskal, J.B. and Wish, M. (1978), *Multidimensional Scaling*, London: SAGE Publications.

Mar Molinero, C. and Xie, A. (2007), 'What do UK employers want from OR/MS?', *Journal of the*

Operational Research Society, Vol. 58, pp. 1543-1553.
Neophytou, E. and Molinero, C.M. (2004), 'Predicting Corporate Failure in the UK: A Multidimensional Scaling Approach', *Journal of Business Finance & Accounting,* Vol. 31, No. 5-6, pp. 677-710.
SAS Institute, Inc. (2008), 'The MDS Procedure', *SAS/STAT® 9.2 User's Guide*, Cary, NC: SAS Institute Inc., pp. 3698-3741.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jun Guan Neoh
University of Southampton
jun.neoh@soton.ac.uk