Middle-East Journal of Scientific Research 24 (3): 514-517, 2016
ISSN 1990-9233
© IDOSI Publications, 2016
DOI: 10.5829/idosi.mejsr.2016.24.03.23052

Using Structural Equation Models For Visualization of Categorical Data

[1]Cengiz Gazeloğlu and [2]Zerrin Asan Greeencare

[1]Department of Electrical and Electronics Engineering,
Engineering Faculty, Abdullah Gul University, Kayseri, Turkey,
[2]Department of Statistics, Science Faculty, Anadolu University, Eskisehir, Turkey

Abstract: The methods used for categorical data analysis can be listed as correspondence analysis, multiple correspondence analysis and the others. Different visualization techniques are appropriate to the measurement level of the data and special methods have been developed to handle univariate, bivariate and multivariate data. In contingency table analysis we now also have graphical models, familiar from structural equation modeling (SEM) and path analysis. SEM are often visualized by a graphical path diagram. SEM is a comprehensive statistical approach to testing hypotheses about relations among observed and latent variables. SEM analysis is performed using either correlation or covariance matrices derived from raw data. Maximum likelihood and generalized least squares methods are among the mostly used methods in SEM studies. The data used in these methods must consist of only continuous variables and should have a distribution close to normal. In the cases of categorical data or data not normally distributed, the correlation coefficient is altered due to the disruption of normality assumption. In this case, the results traditional guessing methods provide are biased. As a result, level of measurement has a crucial role on the decision of the method used for analysis. We showed SEM for visualization in real data set.

Key words: Categorical Data · Structural Equation Modeling · Path Analysis

INTRODUCTION

Data visualization is a new term. It expresses the idea that it involves more than just representing data in a graphical form. The information behind the data should also be revealed in a good display; the graphic should aid readers or viewers in seeing the structure in the data [1]. Graphic has a great importance for data visualization. They reveal data. Indeed graphics can be more precise and revealing than conventional statistical computations. [2]. Graphics provide an excellent approach for exploring data and are essential for presenting results. Although graphics have been used extensively in statistics for along time, there is not a substantive body of theory about topic. It is tended to be concerned more with statistical properties of results. The closer linking of graphics with statistical modeling can make this more explicit [1].

Structural Equation Modelings (SEMs) are well recognized as the most important statistical method to serve some purpose. SEMs can be applied to many fields [3]. It is a growing family of statistical methods for modeling the relations between variables [4]. SEMs consist of methods that are more powerful than methods like multiple regression, path analysis, factor analysis, time series analysis and covariance analysis [5]. It has also been utilized for categorical data modeling in numerous studies available in literature.

This study focuses on the solution of categorical data and structural equation modeling using LISREL software. The main goal of the study is to analyze the results of SEM when applied on the data obtained from course evaluation forms using likert type scale. The data obtained via likert type scale are non-continuous due to its categorical structure [6]. The application of SEM on non-continuous data has unique aspects. The most important one is the usage of asymptotic covariance matrix and weighted least square method as guessing method.

Corresponding Author: Cengiz Gazeloğlu, Department of Electrical and Electronics Engineering, Engineering Faculty, Abdullah Gul University, Kayseri, Turkey.

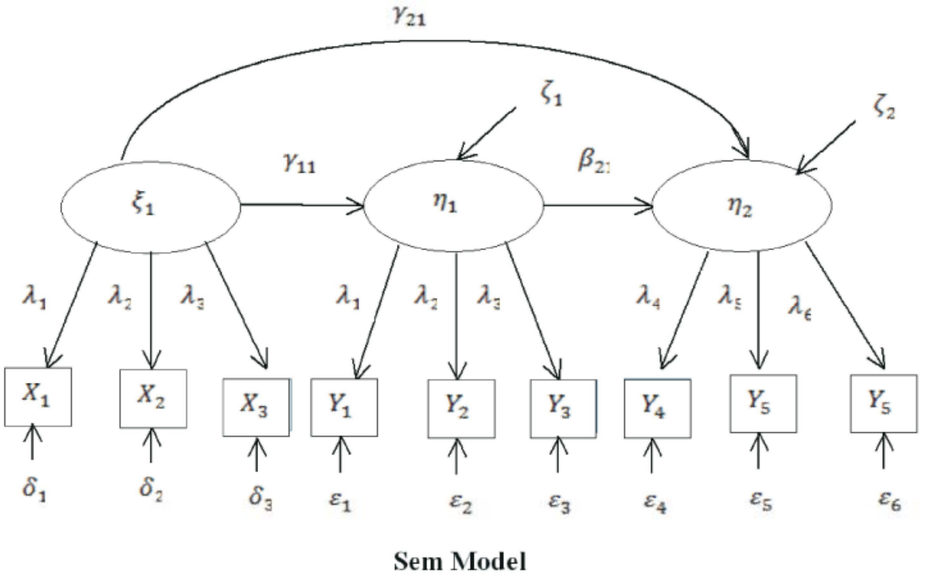

Fig. 1: Structural equation model

Table 1: Correlation coefficients for level of measurement

Correlation Coefficient	Level of Measurement
Pearson product-moment	Both variable interval
Spearman rank, Kendall's tau	Both variable ordinal
Phi, contingecy	Both variable nominal
Point Biserial	One variable interval, one variable dichotomous
Gamma, Rank Biserial	One variable ordinal, one variable nominal
Biserial	One variable interval, one variable artificial
Polyserial	One variable interval, one variable ordinal with underlying continuity
Tetrachoric	Both variables dichotomous (nominal artificial)
Polychoric	Both variables ordinal with underlying continuities

Categorical Data with Structural Equation Modeling: Bollen (1989), refers to the three major information in the historical course of SEM, These: (1) path analysis, conceptual synthesis of the structural model and measurement model and (3) general estimation systems. Causal models showed improvement in historical order, These models; regression models, path analysis, confirmatory factor analysis and SEM. Figure 1 shows the general structure of SEM.

The correlation coefficients calculated varies for level of measurement like pearson for interval, tetrachoric for nominal and so on [7]. Table 1 shows correlations coefficients.

Application: We showed SEM for visualization of categorical data in application. In this study, the dataset contains the results of the study on 5820 subjects performed by Gunduz *et al.* [8]. There is a total of 28 course specific questions and additional 5 attributes in data set. During the study, to maintain the validity of the structural equation model, factorial analysis is applied to the data utilized. The Figure 2 shows model for student's attribute. The KMO level of 0.870 proves that such analysis is doable in Figure 2. In addition, total variance ratio is determined as 69% in Figure 2. The variance ratio, eigenvalues and Cronbach alpha values of each factor and the factor loadings of each question within the factors are listed in Table 2.

The analysis of the above given model shows that there are 4 active factors, namely QU(Quality), SA(Sharing Attitude), USE(Use Behaviour), PF(Profile) and INC(Incentives). Moreover, the QU factor is affected by the PEU endogenous variable with a ratio of 0.63. Also, the most effective variable on the PEU factor is the PEU2 variable with a ratio of 0.83 among 2 other variables. In addition, the USE factor is the most effective one on the QU factor with a ratio of 0.69.

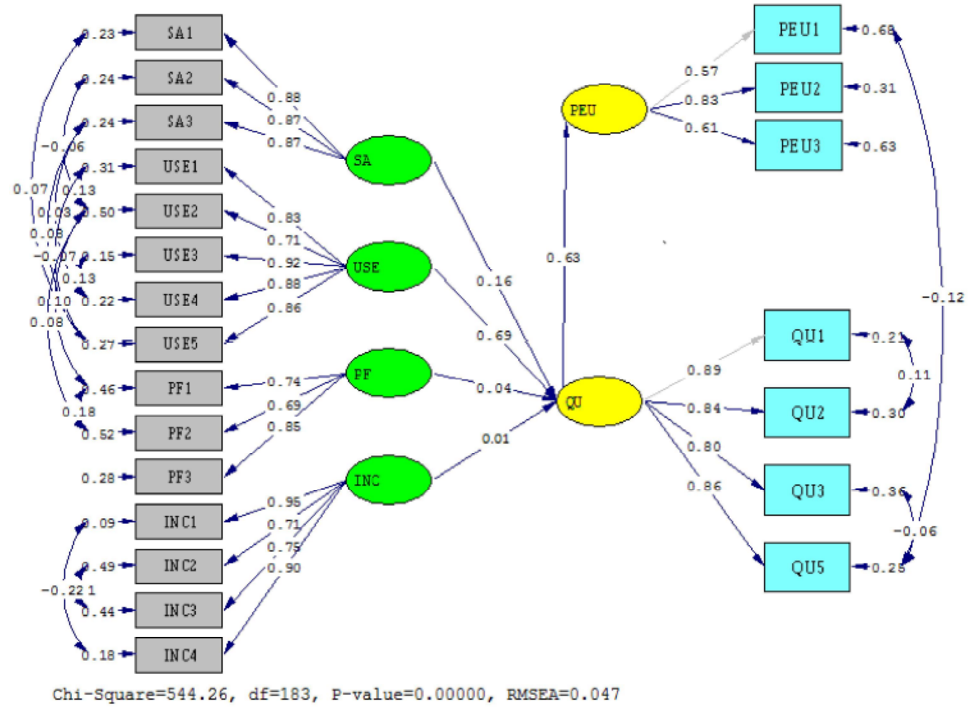

Fig. 2: SEM Model for Turkish student evaluation data

Table 2: Factor loadings

Factors	Factor loading	Eigenvalues	Total variance %	Cronbach alpha
USE		6,307	28,668	0,867
USE3	0,837			
USE4	0,797			
USE1	0,777			
USE2	0,726			
USE5	0,558			
QU		2,461	11,187	0,851
QU1	0,842			
QU2	0,833			
QU3	0,768			
QU5	0,666			
INC		2,090	9,502	0,834
INC3	0,865			
INC2	0,795			
INC1	0,775			
INC4	0,774			
SA		2,025	9,205	0,831
SA3	0,821			
SA1	0,819			
SA2	0,806			
SA		2,025	9,205	0,831
SA3	0,821			
SA1	0,819			
SA2	0,806			
PF		1,175	5,340	0,783
PF1	0,827			
PF2	0,794			
PF3	0,761			
PEU		1,085	4,933	0,669
PEU2	0,715			
PEU1	0,712			
PEU3	0,666			

Table 3: Model evaluation

Fit Measure	Good Fit	Acceptable Fit	Model
RMSEA	0<RMSEA<0.05	0,05 = RMSEA = 0,10	0,047
NFI	0,95 = NFI =1	0,90= NFI = 0,95	0,98
NNFI	0,97 = NNFI =1	0,95= NFI = 0,97	0,98
CFI	0,97 = CFI =1	0,95= CFI = 0,97	0,99
GFI	0,95 = GFI =1	0,90= GFI = 0,95	0,99
AGFI	0,90 = AGFI =1	0,85= AGFI = 0,90	0,98

[10].

On the other hand, the INC factor is the least effective one with 0.01 coefficient. There are 3, 5, 3 and 4 influential variables on the SA, USE, PF and INC factors, respectively. INC1 variable is the most effective variable on INC factor with the ratio of 0.95; while the least influential one is INC2 variable with a ratio of 0.71. Thus, a unit change on the INC2 factor will cause an increase of 0.71 on the INC factor. PF factor is affected by PF1, PF2 and PF3 variables. The variable coefficients are 0.74, 0.69, 0.85, respectively. Similarly, the above model represents the factors and the coefficients on the QU factor. Finally, the QU factor is affected by 4 variables namely QU1, QU2, QU3 and QU5 with respective coefficients of 0.89, 0.84, 0,80 and 0.86.

RESULTS

Various software used for SEM analysis may provide different goodness of fit statistics indices or different names for the same indices in Table 3. For LISREL software, the indices of GFI, AGFI, RMSEA, CFI and NNFI are used in addition to Chi Square values [9].

As a result, the goodness of fit statistics indices are within the boundaries of acceptable limits. Last but not least, the modifications to the model are applied as the software approves. It is showed using structural equation models for visualization categorical data in this study.

The results of the assessment, if requested to increase the quality of the course is necessary to respect the students' opinions about the course's teacher. Because 92% of the students 'opinions of the teachers respect students conduct' seems like an effect caused by SEM. It also affects the quality of education in promoting the students indirectly lessons. When making various activities related to the course for participation in the course will be conducted by teachers, students of class participation will be affected. There is a indirect effect for the quality of course.

The teacher is another factor for quality in education. To have updated information about courses and trainers that have an impact of 85% in order to renew itself constantly, has a lot of influence on students. This affects the quality of education. An overall evaluation of the course, there are some important issues that should be of good quality and can be processed in an understandable way. The teacher to student opinions about the course of the early issues must be extremely respectful. In addition to examining the teacher out and keep himself updated about current issues are the most important issues affecting the quality of education. Explanation of the final plan as early course on how to handle the course on the quality of education, effective using of class hours, giving enough projects and homework to students, making various activities related to the course of the processing time of the course and students are encouraged by teachers to the course affects significantly the quality of course. Such efficiency and directions established for all of these effects is situated in the SEM model.

REFERENCES

1. Chen, C., W. Hardle and A. Unwin, 2008. Handbook of Data Visualization, Springer.
2. Tufte, E.R., 2001. The visual Display of Quantitative Information, Graphic Press.
3. Hoyle, R.H., 2012. Handbook of Structural Equation Modelling, The Guilford Press.
4. Lee, S.Y., 2007. Structural Equation Modeling A Bayesian Approach*,* Wiley and Sons.
5. Şehribanoğlu, S., 2005. Yapısal Eşitlik Modelleri ve Bir Uygulaması, Yüksek Lisan Tezi,Yüzüncü yıl Üniversitesi, Fen Bilimleri Enstitüsü, Van.
6. Agresti, A., 2013. Categorical Data Analysis, Wiley, Inc. Publication.
7. Schumacker, R.E. and R.G. Lomax, 2004. A Beginner's Guide to Structural Equation Modeling, Lawrence Erlbaum Associates.
8. Gunduz, G. and E. Fokoue, 2013. Turkiye Student Evaluation Data Set,https://archive.ics.uci.edu/ml/datasets (15.08.2015).
9. Sümer, N., 2000. Yapısal Eşitlik Modelleri: Temel Kavramlar ve Örnek Uygulama, Türk Psikoloji Yazıları, Cilt. 3, Sayı. 6, ss, pp: 49-73.
10. Engel, S. and H. Moosbrugger, 2003. Evaluating The Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures, Methods of Psychological Research Online, 8(2): 23-74.