The Practice of Structural Equation Modeling

Ho Yin Wong
Deakin University, Australia

Jason Sit
Bournemouth University, United Kingdom

Jia-Yi Hung
Tzu Chi College of Technology, Taiwan

INTRODUCTION

Business analytics is defined as "the general process of exploration and analysis of data to discover new and meaningful patterns in data" (Kohavi, Rothleder, & Simoudis, 2002, p. 45). Similarly, analytics can be considered as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions" (Davenport & Harris, 2007, p. 7). Business analytics endeavors to develop new insights and understand various business scenarios based on data and statistical methods. There are four main types of business analytics: i) statistical analysis, ii) forecasting/extrapolation, iii) predictive modeling, and iv) optimization (Davenport & Harris, 2007). Statistical analysis is concerned with why a particular scenario is happening, while forecasting/extrapolation focuses on what if these trends continue. Predictive modeling is used to understand what will happen next, and optimization is related to what's the best that can happen. SEM is a useful tool in performing the first three types of business analytics. It possesses explanatory power (Hunt, 2002), forecasting/extrapolation and predictive powers (Kaplan, 2008). With the assist of SEM, business managers are better equipped to perform business analytics.

SEM is a comprehensive statistical modeling technique that can be applied in various business areas. SEM as a statistical technique for measuring relationships with latent variables has been around since the early 20th century (Bollen, 1989). It was the works of Joerskog (1969) and Bagozzi (1980) that popularize this technique. SEM can be performed to examine linear regression models. It can also use the correlation coefficient to determine which items correlated or went together to create the factor model, path model with observed variables, and combining path models and confirmatory factor models (Schumacker & Lomax, 2010). Its modeling capacities, estimation techniques, and extensiveness of applications have led to greater applications of SEM (Lei & Wu, 2007). These advantages have made SEM a popular multivariate data analysis technique across many disciplines. The availability of SEM software packages such as Amos, LISREL, EQS, and MPlus for computers has made SEM readily accessible and popularized the use of it. The purpose of this chapter is to illustrate the use of SEM analysis. This chapter covers: i) background of SEM, ii) the process of performing SEM analysis, iii) conclusion, and iv) definitions.

BACKGROUND

SEM is a combination of multivariate statistical analyses such as factor analysis and multiple regression analysis. It differs to, and indeed improves upon, traditional factor analysis and regression analysis. Traditional factor analysis

DOI: 10.4018/978-1-4666-5202-6.ch225

is exploratory in nature, whereas factor analysis in SEM is confirmatory. Hence, they are termed exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), respectively. CFA possesses a major advantage over EFA, in that it is driven by theory. In EFA, all variables of factors are entered into statistical software. Based on a chosen mathematical computation method, the statistical software decides the numbers of factors and which variables belong to which factors. Contrary to EFA, the approach of CFA is that researchers decide which variables are designated to which factors. After mathematical computations, various fit indices confirm if the designated variables fit into the theorized factors. SEM also goes beyond multiple regression analysis, which computes regression equations separately. SEM computes variance-covariance equations simultaneously. Hence, it allows holistic estimations of inter-relationships among the latent variables. Furthermore, in SEM, causal modeling can be done with latent variables, as opposed to the composite variable used in multiple regression analysis, with the consequence of being able to specify an error term for each variable. In summary, SEM can be used to identify and test a theoretical model and to fit the model to the data in a parsimonious manner (Kline, 2005; Schumacker & Lomax, 2010).

MAIN FOCUS OF THE CHAPTER

Assumptions of SEM

Data need to meet three major assumptions of SEM analysis before they can be analyzed. Firstly, a relatively large sample size is required to maintain power and obtain stable parameter estimates and standard errors. A general rule of thumb for SEM analysis is that at least five cases per parameter estimate, including error terms and path coefficients are required (Bentler & Chou, 1987; Hair, Black, Babin, & Anderson, 2009). Others recommend having a sample size of at least one hundred cases, preferably two hundred when using the Maximum Likelihood estimation procedure (Hair et al., 2009; Loehlin, 1992). The second assumption is multivariate normal distribution of variables. Both inferential statistics and the variance-covariance structure among variables can be affected by skewed data. To its credit, maximum likelihood estimation, as used in SEM analysis, is fairly robust to violations of normality (Chou & Bentler, 1995). Another approach to dealing with non-normally distributed data is the transformation of data. A probit data transformation works best in dealing with skewed data (Schumacker & Lomax, 2010). The third assumption is concerned with incomplete data. Missing data is a problem in SEM, because a loss of information decreases the efficiency of parameter estimates and reduces the sensitivity of the analysis technique, even if the loss of data occurs only randomly (Gold & Bentler, 2000; Gold, Bentler, & Kim, 2003). Listwise and pairwise deletions that represent deleting subjects with missing data on any variable and on each pair of variables used, respectively, lack theoretical justification and tend to introduce biased results when small sample size is used (Allison, 2003). Mean substitution is considered to be appropriate when less than 10% of the data are missing (Donner, 1982; Olinsky, Chen, & Harlow, 2003). Regression imputation works well when there is only a moderate amount of missing data in the data set (Schumacker & Lomax, 2010). Expectation maximization is probably the best approach to deal with miss data (Gallagher, Ting, & Palmer, 2008). Expectation maximization replaces missing values with expected values found on the basis of an expectation maximization algorithm.

Model Specification

Model specification is the step of formally stating a conceptual model. It involves mathematically or diagrammatically indicating hypothesized relationships amongst a set of variable (Kline, 2005). A good conceptual model is driven by theories that can be found in the literature, knowledge in the field, and one's educated guesses (Lei & Wu,

Related Content

A Study of Business Intelligence Strategy Development by Large Organizations
Carmine Carmine Sellittoand Paul Hawking (2019). *Applying Business Intelligence Initiatives in Healthcare and Organizational Settings (pp. 326-339).*
www.irma-international.org/chapter/a-study-of-business-intelligence-strategy-development-by-large-organizations/208106

Loss Profit Estimation Using Temporal Association Rule Mining
Reshu Agarwal, Mandeep Mittaland Sarla Pareek (2016). *International Journal of Business Analytics (pp. 45-57).*
www.irma-international.org/article/loss-profit-estimation-using-temporal/142780

An Empirical Analysis of Delhi - Mumbai Sector Flight Fares
T. Godwin (2017). *International Journal of Business Analytics (pp. 60-78).*
www.irma-international.org/article/an-empirical-analysis-of-delhi---mumbai-sector-flight-fares/187209

Estimating Risk with Copulas
Iva Mihaylova (2014). *Encyclopedia of Business Analytics and Optimization (pp. 838-851).*
www.irma-international.org/chapter/estimating-risk-with-copulas/107287

Semantic Document Networks to Support Concept Retrieval
Simon Boese, Torsten Reinersand Lincoln C. Wood (2014). *Encyclopedia of Business Analytics and Optimization (pp. 2135-2146).*
www.irma-international.org/chapter/semantic-document-networks-to-support-concept-retrieval/107400