Chapter 5
Using WarpPLS in E–Collaboration Studies: Descriptive Statistics, Settings, and Key Analysis Results

Ned Kock
Texas A&M International University, USA

ABSTRACT

This is a follow-up on a previous article (Kock, 2010b) discussing the five main steps through which a nonlinear structural equation modeling analysis could be conducted with the software WarpPLS (warppls.com). Both this and the previous article use data from the same E-collaboration study as a basis for the discussion of important WarpPLS features. The focus of this article is on specific features related to saving and analyzing grouped descriptive statistics, viewing and changing analysis algorithm and resampling settings, and viewing and saving the various minor and major results of the analysis. Even though its focus is on an E-collaboration study, this article contributes to the broad literature on multivariate analysis methods, in addition to the more specific research literature on E-collaboration. The vast majority of relationships between variables, in investigations of both natural and behavioral phenomena, are nonlinear; usually taking the form of U and S curves. Structural equation modeling software tools, whether variance- or covariance-based, typically do not estimate coefficients of association based on nonlinear analysis algorithms. WarpPLS is an exception in this respect. Without taking nonlinearity into consideration, the results can be misleading; especially in complex and multi-factorial situations such as those stemming from E-collaboration in virtual teams.

DOI: 10.4018/978-1-4666-2020-9.ch005
INTRODUCTION

The vast majority of the statistical methods used in the behavioral sciences, and many of those used in the natural sciences, can be seen as special cases of structural equation modeling (SEM). This applies to both univariate (a.k.a. bivariate) and multivariate statistical analysis methods (Hair et al., 1987). This can be demonstrated through a sequential logical inference process. In short, it can be shown that most of these methods are instances of multiple regression analysis, which is itself an instance of path analysis, which in turn is an instance of SEM.

Methods like ANOVA, ANCOVA, MANOVA and MANCOVA can be shown to be special cases of multiple regression analysis (Hair et al., 1987; Rencher, 1998). In multiple regression analysis, hypothesis testing is typically conducted through the calculation of coefficients of association between multiple independent variables and one main dependent variable. These coefficients of association normally take the form of standardized partial regression coefficients (Rencher, 1998; Rosenthal & Rosnow, 1991). The corresponding P values are the probabilities that the relationships reflected in the coefficients are “real”.

Path analysis is a method developed by Sewall Wright in the 1930s (Wolfe, 1999; Wright, 1934) and later “rediscovered” by statisticians and social scientists. Sewall Wright was an evolutionary biologist and animal breeder. He was also one of the founders of the field of population genetics. Population genetics unified Darwin’s theory of evolution with Mendel’s theory of genetics. Another co-founder of the field of population genetics was Ronald A. Fisher, who also has made many contributions to the field of statistics (Hair et al., 1987; Kock, 2009).

Any path analysis model can be decomposed into one or more multiple regression models (Gefen et al., 2000; Kline, 1998). Each of the multiple regression models can then be solved separately, and the solution combined into one main solution to the path analysis model. In this sense, multiple regression analysis can be seen as a special case of path analysis. Since SEM is essentially path analysis with latent variables (LVs), then path analysis can be seen as a special case of SEM (Maruyama, 1998). As a corollary, all of the methods discussed above can also be seen as special cases of the SEM.

In SEM, LV scores are calculated as weighted averages of their respective indicators. Usually there are two or more indicators for each LV, although that is not always the case. Once LV scores are calculated, the SEM solution problem is reduced to the solution of a path analysis model. That is achieved through the calculation of path coefficients and respective P values, as well as several other ancillary statistical coefficients. The path coefficients are standardized partial regression coefficients, which are mathematically identical to those obtained through multiple regression analyses.

The calculation of weights linking indicators to LVs is one of the key aspects that differentiate SEM approaches. Those approaches can be divided into two with main types: covariance-based (Chin et al., 2003; Gefen et al., 2000; Haenlein & Kaplan, 2004; Kline, 1998). One of the main advantages of covariance-based SEM is that it employs robust statistics to calculate P values, and thus can be seen as a nonparametric equivalent to covariance-based SEM. That is, unlike covariance-based SEM, variance-based SEM typically yields robust results even in the presence of small samples and multivariate deviations from normality (Chin et al., 2003; Gefen et al., 2000). Variance-based SEM is often referred to as PLS-based SEM, where “PLS” stands for “partial least squares” or “projection to latent structures”. The term “PLS-based SEM” is actually more commonly found in the literature than the term “variance-based SEM” (Chin et al., 2003; Haenlein & Kaplan, 2004).