STRUCTURAL EQUATION MODELING, 7(4), 629–636

TEACHER'S CORNER

Confirmatory Factor Analysis With Different Correlation Types and Estimation Methods

Randall E. Schumacker
Educational Research
Department of Technology and Cognition
University of North Texas

Susan T. Beyerlein
Center for Public Management
Department of Public Administration
University of North Texas

Structural equation modeling techniques can use different correlation coefficients and different estimation methods in confirmatory factor analysis (CFA). The rationale for examining correlation types and estimation methods is related to their effect on the weight matrix (W^{-1}) in the CFA formula for determining the fit function statistics. The results of this study help us to understand that the type of correlation matrix and estimation method effects factor loadings and fit functions. Some suggested alternatives are to use either a limited information estimator for categorical variable analysis or multinomial full information estimators based on modern item response theory.

Recent work in the item response theory (IRT) method of test construction has sparked a renewed interest in the use of confirmatory factor analysis (CFA) with dichotomous data (Takane, Yoshio, & de Leeuw, 1987; Wise & Tatsuoka, 1986). In the past, underlying constructs or latent variables were determined using factor-an-

Requests for reprints should be sent to Randall E. Schumacker, College of Education, Matthews Hall 304, P.O. Box 311337, University of North Texas, Denton, TX 76203–1337. E-mail: rschumacker@unt.edu

alytic techniques that often overlooked the variable scale of measurement (Gorsuch, 1983). The controversy over scales of measurement is not new; for example, Stevens (1968) argued that psychological tests yielded scores that were ordinal not interval in nature, hence violating assumptions of parametric statistics. Today, there is a realization that data of the true–false, correct–incorrect, and satisfied–dissatisfied type are prevalent and need to be appropriately analyzed. Factor analysis, which was typically conducted under the premise that interval level (continuous) measurement of variables occurred, is being questioned when using dichotomously scored test data. No longer is the Pearson correlation matrix the only type of matrix used in factor analysis (Jöreskog & Sörbom, 1993).

PROBLEMS WITH CORRELATION COEFFICIENTS

Mislevy (1986) contended that the use of the phi coefficient with dichotomous data in factor analysis is problematic on several counts. First, the values of the phi coefficient are dependent not only on the strength of the relation between the variables, but also on the means of the individual variables. If two dichotomous variables are perfectly ordered on the Guttman scale, the Pearson r attains a value of 1 only if the means of the two variables are equal. Second, because of the bounded nature of the dichotomous variable, the standard linear factor analysis model is misspecified from the onset. In other words, regressing a dichotomous variable on any continuous latent variable that is unbounded cannot be linear. Third, a factor model for underlying continuous variables and functions of observed discrete variables depends on the skewness of the discrete variables, that is, their mean values. Finally, the choice of cutoff values for binary variables affects the values of the expected phi coefficients. According to Mislevy (1986), the factor analysis of phi coefficients on binary variables produced by the same underlying correlation structure, but dichotomized at different points, can conform to factor models with different structures and possibly different numbers of factors.

The Pearson correlation coefficient computed with dichotomous data has also been unfavorably reviewed. Ethington (1987), in reviewing the literature on the use of observed dichotomous variables in LISREL analyses, determined that the Pearson correlation coefficient underestimates the factor loadings of the categorical variables and overestimates the chi-square goodness-of-fit values, and therefore is not recommended for use with dichotomous data. Jöreskog and Sörbom (1988), in two different Monte Carlo studies, investigated the various types of correlation coefficients to determine which one is the "best." The Phi, Spearman rank, and Kendall tau-b correlations performed poorly, whereas the tetrachoric correlations with ordinal data produced robust parameter estimates and better fitting models. For these reasons, they recommended the use of the tetrachoric correlation with dichotomous data.

Muthén (1984) examined the advantages and disadvantages of tetrachoric correlations in the context of the assumption of symmetric normal distributions for underlying latent variables when using dichotomous variables. One problem associated with their use was that tetrachoric correlation matrices are not ensured positive definiteness as are Pearson correlation matrices. A matrix is nonpositive definite when at least one eigenvalue is negative. This could be caused by sampling, outliers, or variable collinearity. One approach to correcting this problem is to smooth the matrix using ridge or principal component methods (Wothke, 1992).

A nonpositive definite tetrachoric correlation matrix may indicate violation of the assumption of underlying normality, or it may arise from sampling variability. In addition, tetrachoric correlation matrices generally provide extremely inflated chi-square values and underestimated standard errors of estimates due to larger variability than Pearson correlations. One solution is to specify a factor model that assumes the observed variables are related to a set of latent variables through a nonlinear model, thereby avoiding the problems generated by using dichotomous variables with the linear factor analysis model (nonlinear models are specified for use with logit and probit regression as well as IRT modeling). One can also assess the differences in the methods and choose what appears to be the best, given the nature of the particular data and research goals.

It should be noted that factor analysis and principal component methods are different (Jöreskog & Sörbom, 1979). Factor analysis is a correlation-oriented approach that aims to reproduce the intercorrelation among the variables, whereas principal components is a variance-oriented technique that aims to reproduce the total variable variance with all components required to always reproduce the correlations exactly. This study implemented CFA, hypothesized factors that were less than the number of variables, and examined how well the intercorrelations were reproduced. In general, the number of hypothesized factors in CFA will not account for as much variance as the same number of principal components.

PROBLEMS WITH ESTIMATION METHODS

According to Ethington's (1987) review, various studies indicated that several estimation procedures perform equally well in factor analysis models containing categorical data when they did not deviate from normality. These were maximum likelihood (ML), generalized least squares (GLS), categorical variable methodology (CVM), and asymptotically distribution free (ADF) procedures. However, where extreme skewness existed, ML and GLS chi-square values tended to be inflated, and standard errors tended to be underestimated. Distortions were not evident when sample size was at least 400. The results suggested that the choice of estimation procedure may not be as important as the choice of correlation type when the concern is with the robustness of parameter estimates. However, estimation procedures impact hypothesis testing and assessment of fit.

Mislevy (1986) examined the ULS, GLS, and ML estimation procedures for factor models utilizing dichotomous responses. He characterized ULS and GLS as "limited information" solutions, and the marginal ML estimation procedure as a full information solution. The advantages of unweighted least squares (ULS) for factor models with dichotomous data included a superiority over the use of phi coefficients and relative economy. The disadvantages included a loss of information, lack of standard error, goodness-of-fit indexes, and problems in estimating S (the correlation matrix among the latent ηs when extreme values are present).

Both GLS and ML provided better estimates of standard errors of estimation and statistical tests of model fit. Because of constraints in their respective algorithms, Mislevy (1986) suggested using ML for long tests with few factors and GLS for short tests with many factors. Both appeared acceptable for short tests and few factors, but at present, neither was very good for long tests and many factors. In summary, the estimation methods provide different results in the presence of variable measurement scale, variance–covariance, and normality assumptions, leaving one to sort out when to use which estimation method.

METHOD AND PROCEDURE

Participants and Instrument

Two sections of an undergraduate class at the University of North Texas participated in the study. A total of 100 students in the two sections of the course were given a 35-item multiple choice test over a single content domain. Each item had four choices and was scored 1 for a correct answer and 0 for an incorrect answer. The Cronbach alpha internal consistency reliability for the item responses was .63. It should be noted that reliability does not imply that items measure only one factor (Nunnally & Bernstein, 1994). A factor analysis using the principal axis extraction method yielded a single factor. Factor loadings ranged from .35 to .57.

Design

The 35 dichotomous item responses for the 100 students were converted into two types of correlation matrices: Pearson and tetrachoric. A CFA with a single unidimensional factor was hypothesized. The UL, GLS, and ML estimation methods were then computed using the PRELIS and LISREL procedures available in LISREL8 (Jöreskog & Sörbom, 1996a) and PRELIS2 (Jöreskog & Sörbom, 1996b). The factor loadings and fit functions were compared given these two types of matrices and three estimation methods.

The generally weighted least squares (WLS) method was not employed in this study due to the large sample size requirement. Similarly, the diagonally weighted least squares (DWLS) method was not included because it doesn't lead to asymptotically efficient estimates of model parameters (factor loadings), and it only uses the asymptotic variances rather than the covariance of the estimated coefficients in the weight matrix (W^{-1}).

RESULTS

Several types of correlations can be computed depending on the scale of measurement; for example, tetrachoric, phi, Spearman, Pearson, point-biserial, and biserial. However, only the Pearson, tetrachoric, and biserial correlations can be used in current computer programs to conduct CFA. In this study, only Pearson and tetrachoric were compared under ULS, GLS, and ML estimation methods. The factor loadings and fit functions were compared to examine the impact of using these different correlation matrices with dichotomous data under different estimation methods.

The LISREL8 and PRELIS2 computer programs that generated the factor loadings and fit functions for both types of correlation matrices read in a raw data file that contained an identification number followed by 35 item responses dichotomously coded for 100 students. The programs created either a Pearson or tetrachoric correlation matrix. The /OU or output statement included either the ULS, GL, or ML command option to produce the factor loading comparisons in Table 1 and the fit function comparisons in Table 2.

The factor loadings varied depending on the correlation type and estimation method. The fit functions differed significantly depending on the type of correlation matrix and estimation method. The tetrachoric matrix produced a nonpositive definite matrix in the ULS and GLS estimation methods. The maximum likelihood method did not yield a nonpositive definite matrix, but did contain larger average standard errors (root mean squared residual). These results are consistent with expectations given prior research and concerns over the correlation type and estimation method used.

SUMMARY

The results of this heuristic example indicated that the type of correlation coefficient and estimation method yielded different factor loadings and fit functions. Prior research indicated that the tetrachoric correlation using the ML estimation method was a consistent estimator with small variances not being a problem if a sufficiently large sample size was used. However, the normal-theory chi-square fit function and the standard errors are not robust (Muthén & Kaplan, 1985), plus the problem of nonpositive definite matrices plaques the use of polychoric correlations in factor analysis.

The use of dichotomous data in CFA presents a unique set of problems. The problems are compounded when considering a mixed correlation matrix where the variables are measured on different scales. In this case, the researcher is warned against conducting the analysis. A guideline to follow is one in which the correlated variables are measured on the same scale of measurement. PRELIS2 itself provides a restriction in that only the Pearson, tetrachoric, and biserial correlation types are permitted. The Pearson correlation matrix with interval data is preferred because of the CFA requirement of variance/covariance among variables and normality assumption.

TABLE 1
Comparison of Factor Loadings by Correlation Type and Estimation Method

	Pearson			Tetrachoric		
Variable	*ULS*	*GLS*	*ML*	*ULS*	*GLS*	*ML*
V1	0.241	0.216	0.235	0.535	0.061	0.523
V2	0.173	0.208	0.177	0.379	0.050	0.372
V3	–0.583	–0.426	–0.614	–0.665	0.004	–0.675
V4	0.016	–0.033	0.034	0.074	0.061	0.052
V5	0.173	0.157	0.149	0.115	0.071	0.084
V6	–0.297	–0.169	–0.332	–0.376	0.084	–0.395
V7	0.119	0.132	0.106	–0.017	0.073	–0.042
V8	0.404	0.436	0.355	0.339	–0.030	0.336
V9	–0.067	–0.014	–0.106	–0.192	–0.083	–0.187
V10	0.004	0.078	–0.011	–0.009	–0.013	0.002
V11	–0.274	–0.160	–0.288	–0.247	–0.005	–0.246
V12	–0.105	–0.002	–0.122	–0.191	–0.075	–0.172
V13	–0.603	–0.471	–0.636	–0.696	0.016	–0.703
V14	–0.136	–0.008	–0.167	–0.016	0.067	–0.021
V15	0.154	0.250	0.113	0.233	0.066	0.220
V16	–0.024	0.093	–0.066	0.081	0.060	0.068
V17	0.307	0.459	0.282	0.368	–0.020	0.380
V18	0.134	0.250	0.111	0.260	0.066	0.254
V19	–0.293	–0.184	–0.341	–0.419	–0.025	–0.433
V20	0.273	0.411	0.215	0.265	–0.028	0.268
V21	–0.201	–0.138	–0.238	–0.424	0.077	–0.445
V22	0.148	0.167	0.106	0.121	0.055	0.086
V23	–0.305	–0.188	–0.317	–0.336	–0.020	–0.333
V24	–0.458	–0.322	–0.494	–0.606	0.006	–0.613
V25	0.135	0.235	0.090	0.037	–0.041	0.030
V26	0.178	0.366	0.119	0.251	0.053	0.242
V27	–0.252	–0.101	–0.286	–0.275	–0.010	–0.269
V28	0.088	0.050	0.090	0.181	0.072	0.169
V29	0.086	0.266	0.041	0.239	0.058	0.233
V30	–0.087	–0.075	–0.077	–0.691	–11.369	–0.647
V31	0.424	0.708	0.402	1.109	0.064	1.149
V32	0.567	0.616	0.512	0.880	0.048	0.860
V33	0.283	0.299	0.259	0.428	0.060	0.424
V34	0.311	0.483	0.269	0.537	0.048	0.535
V35	–0.265	–0.117	–0.311	–0.662	–0.026	–0.665

Note. ULS = unweighted least squares; GLS = generalized least squares; ML = maximum likelihood.

DISCUSSION

The rationale for examining correlation types and estimation methods is related to their effect on the weight matrix (W^{-1}) in the CFA formula for determining the fit function statistics (Browne, 1984). The fit function formula is $F(\theta) = (s - \sigma)' W^{-1} (s - \sigma)$, where s is the diagonal and lower half elements of the correlation/covariance

TABLE 2
Comparison of Fit Functions by Correlation Type and Estimation Method

	Pearson			*Tetrachoric*		
Fit Function	*ULS*	*GLS*	*ML*	*ULS*[a]	*GLS*[a,b]	*ML*
Chi-square[c]	827.07	618.60	863.22	0.00	34.40	39.92
Coefficient *R*	.777	.925	.781	—	.292	.387
GFI	.745	.643	.679	.438	.980	.977
AGFI	.713	.598	.638	.368	.978	.974
RMSR	.107	.175	.107	—	.319	.276

[a]Theta delta matrix (error variance) not positive definite. [b]Ridge option invoked because matrix not positive definite. [c]Assumes multivariate normality.

matrix S used to fit the model to the data, and σ is a vector of corresponding elements of $\Sigma(\theta)$ reproduced from the model parameters θ. Elements of the W^{-1} matrix are supposed to be nonpositive definite of the order $p \times p$, where $p = k(k+1)/2$ or the number of elements in the matrix. The elements of the matrix W, before inverting, are chosen by the ULS, GLS, or ML methods to be consistent estimators of the asymptotic covariance between elements of the S and Σ matrices. The basic assumption is that if the model holds in the population and the variance–covariance in S converge in probability to the corresponding elements in Σ, as the sample size increases, the fit function gives a consistent estimate of θ. In practice, there is not just one fit function that satisfies this condition; hence other factor models may be plausible. Here is the basic problem of determining the best model, the parameter estimates themselves, and the statistical test of fit.

RECOMMENDATIONS

The CFA of dichotomous data based on Pearson correlation or covariance often produces biased estimates. Several authors therefore introduced the alternate method of tetrachoric and biserial correlations for analyzing dichotomous data, which are less biased estimates of the population correlations (Christoffersson, 1975; Muthén, 1984). The latest commercially available program M*plus* (Muthén & Muthén, 1998) incorporates newer dichotomous data handling routines in a user-friendly environment.

Although dichotomous data analysis in PRELIS2 is possible, it is now known that the non-Pearson matrices are often nonpositive definite, even with large sample sizes, and model estimation using GLS or ML estimation methods. Jöreskog (1990) recently developed a limited information estimator for CVM based on Browne's (1984) ADF estimator. It is one recommended alternative.

An alternative to CVM is the multinomial full information estimators based on modern IRT (Bock, Gibbons, & Muraki, 1988). Because solutions generated by this method are estimated directly from the multivariate response patterns, the item correlation matrix is not needed except for possibly determining an approximate start value. Such an IRT implementation of factor analysis for dichotomous data is available in the TESTFACT program (Wilson, Wood, & Gibbons, 1984), although limited at this time to four-factor models.

REFERENCES

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261–280.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 10*, 1–19.

Ethington, C. A. (1987). The robustness of LISREL estimates in structural equation models with categorical variables. *Journal of Experimental Education, 55*(2), 80–88.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality & Quantity, 24*, 387–404.

Jöreskog, K. G., & Sörbom, D. (1979). *Chapter 1: Basic ideas of factor and component analysis*. In J. Magidson (Ed.), *Advances in factor analysis and structural equation models* (pp. 5–20). Cambridge, MA: Abt Books.

Jöreskog, K. G., & Sörbom, D. (1988). *PRELIS. A program for multivariate data screening and data summarization. User's Guide* (2nd Ed.). Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jöreskog, K. G., & Sörbom, D. (1996a). *LISREL8: User's reference guide*. Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1996b). *PRELIS2: User's reference guide*. Chicago: Scientific Software International.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics, 11*, 3–31.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered, categorical and continuous latent variable indicators. *Psychometrika, 49*, 115–132.

Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of nonnormal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171–189.

Muthén, L., & Muthén, B. O. (1998). *M-Plus user's guide*. Los Angeles: Author.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Stevens, S. (1968). Measurement, statistics, and the schemapiric view. *Science, 161*, 849–856.

Takane, Y., Yoshio, S., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393–408.

Wilson, D. T., Wood, R., & Gibbons, R. T. (1984). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Chicago: Scientific Software International.

Wise, S. L., & Tatsuoka, M. M. (1986). Assessing the dimensionality of dichotomous data using modified order analysis. *Educational and Psychological Measurement, 46*, 295–301.

Wothke, W. (1992). Nonpositive definite matrices in structural modeling. *Sociological Methods and Research*. Beverly Hills, CA: Sage.