DOCUMENT RESUME

ED 222 524 TM 820 671

AUTHOR Khattab, Ali-Maher; Hocevar, Dennis
TITLE Significance Testing in Confirmatory Factor Analytic Models.
PUB DATE Mar 82
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Aptitude Tests; Correlation; *Factor Analysis; Goodness of Fit; Hypothesis Testing; *Mathematical Models; Predictor Variables; Secondary Education; *Statistical Significance
IDENTIFIERS Aptitudes Research Project; *Confirmatory Factor Analysis; *Structure of Intellect

ABSTRACT

Traditionally, confirmatory factor analytic models are tested against a null model of total independence. Using randomly generated factors in a matrix of 46 aptitude tests, this approach is shown to be unlikely to reject even random factors. An alternative null model, based on a single general factor, is suggested. In addition, an index of model efficiency is introduced as a useful adjunct to contemporary indices of overall fit. The usefulness of these procedures are demonstrated in a confirmatory factor analysis based on Guilford's Structure of the Intellect model. (Author)

Significance Testing in Confirmatory Factor Analytic Models

Ali-Maher Khattab and Dennis Hocevar
University of Southern California

Paper presented at the
1982 conference of the
American Educational Research Association

Significance Testing in Confirmatory Factor Analytic Models

Ali-Maher Khattab and Dennis Hocevar

University of Southern California

Abstract

Traditionally, confirmatory factor analytic models are tested against a null model of total independence. Using randomly generated factors in a matrix of 46 aptitude tests, this approach is shown to be unlikely to reject even random factors. An alternative null model, based on a single general factor, is suggested. In addition, an index of model efficiency is introduced as a useful adjunct to contemporary indices of overall fit. The usefulness of these procedures are demonstrated in a confirmatory factor analysis based on Guilford's Structure of the Intellect model.

Confirmatory factor analysis (Joreskog, 1969; Joreskog & Sorbom, 1979) is designed to test the fit of a particular model to an observed covariance or correlation matrix. Originally, Joreskog (1969) proposed a X^2 goodness-of-fit test designed to determine whether a model can account for observed correlations between measured variables. Several authors (e.g., Bentler, 1980; Joreskog, 1969, 1979) have pointed out that this test is overly dependent on sample size in that the probability of rejecting a model increases as the sample size increases. In other words, support for a given model would be more likely in a small sample, while in a large sample, a plausible model would be rejected due to minute difference between the observed and predicted intercorrelations.

Because of the sample size problem, several authors (McGaw & Joreskog, 1971; Tucker & Lewis, 1973) have urged the use of a null model in the testing of exploratory factor analytic hypotheses. More recently, Bentler (1980) has extended this idea to confirmatory factor analysis and other multivariate techniques based on maximum likelihood estimation. Typically, the null model is the case where no common factors are assumed. Because the null model is a nested case of a more substantive proposed model, it can be statistically compared with the model of interest.

The purpose of the present study is to demonstrate a potential problem with this procedure and to offer several suggestions for dealing with this problem. The problem associated with the use of a null model is that human abilities are positively correlated (see Brody & Brody, 1978, for review), and one would not expect to find support for such a model regardless of the alternative model that is hypothesized.

To demonstrate this supposition, data from Guilford's Aptitude Research Project were reanalyzed. In this study (Guilford, 1968), 46 tests were given to 197 high school students. Initially, a factor structure based on a random pattern was proposed. Specifically, odd numbered tests were designated as factor 1, and even numbered tests were designated as factor 2. When random factor 1 (R1) was contrasted with the null model of total independence it was highly significant (X^2 difference (23) = 714.54 p < .001). Similar results were obtained for random factor 2 (R2), X^2 difference (23) = 487.58, p < .001. These two models were then combined into an oblique two-factor model which was significantly better than either of the one-factor models, X^2 difference (24) $>$ 662.54, p < .001. The final factor loadings for the random two-factor model were inspected, and it is noteworthy that all factor loadings were significant. At this point conventional procedures have provided no basis for rejecting the random models. Indeed, their tenability has been repeatedly verified.

The analysis above clearly indicates a problem with the uncritical use of confirmatory factor analysis. First, the commonality among correlated tests of mental ability can be subdivided into two or more random factors, and second, traditional ways of evaluating confirmatory models do not provide rigorous tests of the appropriateness of a given model.

A possible solution to this problem lies in a redefinition of the null model. Rather than test theoretical models against the untenable null hypothesis of total independence, models can be tested against the more likely hypothesis that a single common factor accounts for the intercorrelations between measured variables. Thus, in the above example, one might propose that the two-factor random model be compared to a one factor general model.

In the present example, the X^2/df ratios are almost identical (2.348 vs 2.351), suggesting that the more complex two-factor model is no better than the simple one factor model. By the law of parsimony, the one factor model should be the accepted model.

An alternative to a significance testing is to compare the fit of a single factor model to a more complex multifactorial model. Bentler (1980) has proposed a fit index for multivariate models (coefficient delta) which is a slight modification of the more familiar Tucker-Lewis (1973) index for evaluating exploratory models. The index is given by the following formula:

$$\Delta = (F_k - F_1)F_o$$

where F_k and F_1 correspond to minimum function values for two hierarchical (i.e., nested) models. As shown by Bentler (1980), coefficient delta ranges between zero and one, and consequently, provides a normed fit index for competing models. Unfortunately, delta will increase as model complexity increases, or stated another way, delta will increase as more parameters are free to vary.

A slightly modified special instance of Bentler's (1980) coefficient delta avoids the problem of the influence of degrees of freedom on delta. This efficiency index, hereafter called epsilon is given by:

$$\varepsilon = 100\ (F_k - F_o)/((F_o) * (S_k - S_o))$$

where F_k is the minimum fit function evaluated for the model of interest, F_o is the minimum fit function for the null model of complete independence, S_k is the total number of independent parameters in F_k, and S_o is the total number of independent parameters for F_o. The epsilon index adjusts for the number of free parameters, and consequently models which differ in degrees of freedom are more directly comparable. The index may be thought of as the per parameter average increase in fit. Thus, the index is referred to as a measure of model efficiency.

The use of Bentler's delta and the proposed epsilon are shown in the following extension of the previous example. According to Guilford's Structure of Intellect model, one-half of the tests in the present battery are tests which are comprised of semantic content. Thus, two underlying factors, symbolic and semantic content can be hypothesized. Using this conceptualization, a number of alternative models, as well as the two-factor random model discussed earlier, are given in Table 1. As shown in Table 1, when coefficient delta is used, models with a larger number of free parameters generally have a better fit than models with fewer free parameters. To illustrate the two-factor random model (R1 + R2) is better than the theoretically based semantic or symbolic models (.389 > .184 and .389 > .236). Similarly, a randomly generated three factor model (G + R1 + R2) is better than and theoretically based two-factor model (SEM + SYM) using coefficient delta as the bases for comparison (.537 > .437). The epsilon coefficients provide a different picture. In two cases, the theoretically based models (i.e., SYM, SEM + SYM) have larger efficiency indices than the random models (i.e., R1 + R2 and G + R1 + R2).

References

Bentler, P. & Bonnet, . Significance testing and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 1980, 88, 588-607.

Joreskog, K.G. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1978, 43, 443-477.

Joreskog, K.G. & Sorbom, D. LISREL IV: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Chicago: International Fducational Services, 1978.

Joreskog, K.G. & Sorbom, D. Advances in Factor Analysis and Structural Equation Models. Cambridge: ABT, 1979.

Tucker, L.R. & Lewis, C.A. A reliability coefficient for maximum likelihood factor analysis. Psychometrika. 1973, 38, 1=10.

Table 1

Delta and Epsilon Coefficients

	delta	epsilon
1. Semantic only (SEM)	.184	.802
2. Symbolic only (SYM)	.236	1.026
3. R1 + R2	.389	.827
4. SEM + SYM	.437	.929
5. General (G)	.388	.842
6. G + R1 + R2	.537	.577