DOCUMENT RESUME

ED 344 926 TM 018 261

AUTHOR	Schumacker, Randall E.
TITLE	Goodness of Fit Criteria in Structural Equation Models.
PUB DATE	Apr 92
NOTE	22p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Chi Square; Comparative Analysis; Equations (Mathematics); Estimation (Mathematics); *Evaluation Criteria; Evaluation Methods; *Goodness of Fit; *Mathematical Models; Multivariate Analysis; Research Methodology; *Structural Equation Models
IDENTIFIERS	*Maximum Internal Correlation Coefficient

ABSTRACT

Several goodness of fit (GOF) criteria have been developed to assist the researcher in interpreting structural equation models. However, the determination of GOF for structural equation models is not as straightforward as that for other statistical approaches in multivariate procedures. The four GOF criteria used across the commonly used statistical packages are: (1) chi square; (2) GOF; (3) adjusted GOF; and (4) root mean square residual. Two other approaches are the Tucker-Lewis Index (1973) and the Normed Fit Index (1980). Other measures that provide model comparisons are the normed chi-square, the parsimonious fit index, and the Akaike Information Criterion. A new measure of model fit, the maximum internal correlation coefficient, is proposed as a GOF criterion in structural equation models. This approach offers another test of model fit. Model testing is suggested for effectiveness and functionality of the specified variable relationships. Two tables summarize GOF criteria, and there is a 32-item list of references. (SLD)

Goodness of Fit Criteria in Structural Equation Models

Randall E. Schumacker, Ph.D.

Educational Foundations, Research, and Special Education

University of North Texas

Paper presented at the annual meeting of the
American Educational Research Association
April 21, 1992
Tuesday, 10:35 - 11:15, Hilton Ballroom A, Lobby Level
San Francisco, California

Goodness of Fit Criteria
in
Structural Equation Models

INTRODUCTION

Various statistical approaches have been proposed for conducting educational research. These statistical approaches are reviewed to provide a background for the development of structural equation modeling. Each statistical approach has a criteria by which statistical significance is indicated. These criteria or significance tests have been given the general name, Goodness of Fit (GOF) criteria.

Multiple regression provided the important multiple correlation coefficient (R) and subsequent R-squared value (R^2) to determine the overall contribution of a set of variables to prediction. Multiple regression also permits full and restricted models to be tested for the significant contribution of each variable in a model. The use of all possible regressions can further delineate the best set of multiple independent predictors. The importance of multiple regression in explanation and prediction as well as the statistical tests have been well established (Pedhazur, 1982).

Path analysis is essentially multiple-multiple regression equations and as such utilizes the R^2 value, t-test, and F-test to examine overall model fit, reduction in paths, and to test separate partial (path) coefficients (Pedhazur, 1982, 607-614; Williams and Klimpel, 1974). Another GOF criteria called a meaningfulness

criteria (Asher, 1976, 32-34), examines the discrepancy between the original correlations in the matrix and those reproduced from the model based on direct, indirect, correlated, and spurious effects (Marascuilo and Levin, 1983, 169-172). The chi-square statistic with df equal to the number of over identified restrictions has also been widely used (Pedhazur, 1982, 98). Another GOF criteria termed Q (Pedhazur, 1982, 619-622) uses the R^2 value and depending upon sample sizes greater than 100 involves an adjustment using W. Boyle (1970) previously introduced path analysis capability using nominal and ordinal variable paths which was of concern in mixed models in multiple regression.

Factor analytic models are often distinquished between exploratory and confirmatory varieties. In either case, the expressions of communality and uniqueness of a variable are just expressions of the general notion of predicted R^2 and residual variance (Loehlin, 1987, 20; Kim and Mueller, 1978a and 1978b; Long, 1983). Factor loadings are equivalent to path coefficients and variables or items which share communality on a factor assist in describing or labeling latent variables.

Latent variable models or structural equation models are unique in that "factors" are related which distinquishes it from relationships and communalaties among observed/measured variables (Duncan, 1975; Bentler & Weeks, 1980; Lomax, 1982; Long, 1983; Plewis, 1985). Structural equation models have become widely used in the social and behavioral sciences (Saris & Stronkhorst, 1984; Anderson, 1987; Fassinger, 1987; Bollen & Ting, 1991).

GOODNESS-OF-FIT CRITERIA

Several GOF criteria have been developed to assist the researcher in interpreting structural equation models. The determination of goodness-of-fit for structural equation models however is not as straightforward as with other statistical approaches in multivariable procedures such as the analysis of variance, multiple regression, path analysis, discriminant analysis, and canonical analysis. These multivariable methods use observed variables which are assumed to be measured without error and have statistical tests with known distributions. Structural equation modeling procedures have no single statistical test of significance which identifies a correct model given the sample data. Consequently, the statistical packages have developed a number of GOF criteria, unfortunately they do not provide all of the same GOF criteria (Table 1). The four GOF criteria common across the statistical packages are: chi-square; goodness-of-fit; adjusted goodness-of-fit; and root mean square residual (except EQS program).

Insert Table 1 Here

GOODNESS-OF-FIT CRITERIA TYPES

The various GOF criteria are typically used in combination to assess model fit, model comparison, and model parsimony (Hair, et al., 1992, 489-496). All of the GOF criteria, except X^2, do not have an associated statistical test of significance. Therefore, most GOF criteria range in value from 0 (no fit) to 1 (perfect fit) (Bentler, 1980; Baldwin, 1989). Table 2 summarizes many of the GOF criteria with an associated level of acceptable fit and interpretation.

Insert Table 2 Here

MODEL FIT

Model fit determines the degree to which the structural equation model fits the data. Model fit GOF criteria commonly used are: chi-square (X^2), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), and root mean square residual (RMR) (Joreskog and Sorbom, 1989, 25). These GOF criteria are based on differences between the observed (original, S) and model implied (reproduced, Σ) correlation/covariance matrix.

Chi-square (X^2)

A large X^2 value relative to the degrees of freedom indicates that the observed and reproduced (estimated) matrices differ. Statistical significance indicates the probability that this difference is due to sampling variation. A low X^2 value with significance levels greater than .05 indicate that the two matrices are not statistically different. The researcher is interested in obtaining a low X^2 value with a non-significant probability level. The X^2 test being non-significant indicates that the model fits the data, however, an uncertainty will always persist because other models are possible that may fit the data. Although the X^2 GOF criteria is the only statistical test procedure, it is sensitive to sample size because as sample size increases (generally above 200), the X^2 test has a tendency to indicate a significant difference between equal models. In contrast, as sample size decreases (generally below 100) the X^2 test indicates non-significant differences between the observed and reproduced matrices in unequal models. The X^2 test is also sensitive to departures from multivariate normality of the observed variables.

Three approaches are commonly used to calculate X^2 in latent variable models (Loehlin, 1987, 54-67). They are maximum likelihood (ML), generalized least squares (GLS), and ordinary least squares (OLS). Each approach estimates a best fitting solution and evaluates the model fit. The ML estimates are consistent, unbiased, efficient, scale invariant, scale free, and

normally distributed if the observed variables meet the multivariate normality assumption. The GLS estimates have the same properties of the MLS approach under a less stringent multivariate normality assumption and provide an approximate chi-square test of model fit to the data. The ULS estimates do not depend on a normality distribution assumption, however, the estimates are not as efficient nor are they scale invariant or scale free. The ML X^2 statistic is: $X^2 = (n-1)\ F_{ml}$; GLS X^2 statistic is: $X^2 = (n-1)\ F_{gls}$; and ULS X^2 statistic is: $X^2 = (n-1)\ F_{uls}$. Where:

$$F_{ml} = tr\ (\ S\ \Sigma^{-1}\ ) - n + \ln |\Sigma| - \ln |S|\ ;$$

$$F_{gls} = .5\ tr\ [\ (\ S - \Sigma\ )\ S^{-1}\ ]^2\ ;$$

$$F_{uls} = .5\ tr\ [\ (\ S - \Sigma\ )^2\ ]\ ;$$

$$df = .5\ (p+q)\ (p+q+1) - t\ ;$$

t = total number of independent parameters estimated ;

n = number of observed variables ; and

(p + q) = number of observed variables analyzed.

Goodness-of-Fit (GFI) and Adjusted Goodness-of-Fit (AGFI) Indices

The GFI index is based on a ratio of the sum of the squared differences between the observed and reproduced matrices to the observed variances, thus allowing for scale. The GFI measures the amount of variance and covariance in S that is predicted by the reproduced matrix Σ. The GFI index can be computed for ML, GLS, or ULS estimates (Bollen, 1989, 276-277). The GFI index for ULS estimates would be computed as:

$$GFI = 1 - \frac{tr\ [(S - \Sigma)^2)]}{tr\ (S^2)}$$

The AGFI adjusts the GFI index for the degrees of freedom of a model relative to the number of variables. The AGFI index is computed as:

$$AGFI = 1 - (k/df)\ (1 - GFI)$$

Where:

k = number of variables in matrix Σ

df = number of degrees of freedom in model

The GFI and AGFI indices can be used to compare the fit of two different models with the same data or compare the fit of models with different data.

Root Mean Square Residual (RMR)

The RMR index uses the square root of the mean squared differences between matrix elements in S and Σ. It is used to compare the fit of two different models with the same data. The RMR index is computed as:

$$RMR = [\ (1/k)\ \Sigma_{ij}\ (\ s_{ij} - \sigma_{ij}\ )^2\ ]^{1/2}$$

MODEL COMPARISON

Given the role chi-square has in model fit of latent variable models, two others have emerged as variants for model comparison:

Tucker-Lewis Index (TLI) and Normed Fit Index (NFI) (Bentler & Bonett, 1980, 1982; Loehlin, 1987, 68). These GOF criteria typically compare a proposed model to a null model. In the EQS program, the null model is indicated by the Independence Model Chi-square value. The null model could also be any model that establishes a base for expecting other models to exceed.

Tucker-Lewis Index (TLI)

Tucker and Lewis (1973) developed this index initially for factor analysis, but later extended it to structural equation modeling. The measure can be used to compare alternative models or a proposed model against a null model. It is computed using the X^2 statistic as follows:

$$TLI = \frac{(X^2_{null} / df_{null}) - (X^2_{proposed} / df_{proposed})}{(X^2_{null} / df_{null}) - 1}$$

Normed Fit Index (NFI)

The NFI is a measure which rescales chi-square into a 0 (no fit) to 1.0 (perfect fit) range (Bentler & Bonett, 1980). It is used to compare a proposed model to a null model as follows:

$$NFI = \frac{(X^2_{null} - X^2_{proposed})}{X^2_{null}}$$

MODEL PARSIMONY

Parsimony refers to the number of estimated coefficients required to achieve a specific level of fit. Basically, an "over identified" model is compared to a "restricted" model. The AGFI measure discussed previously will also provide an index of model parsimony. Others which provide model comparison are: Normed Chi-square (NC); Parsimonious Fit Index (PFI); and Akaike Information Criterion (AIC). These indices indicate parsimonious goodness of fit, taking into account the number of parameters required to achieve a given level of chi-square.

Normed Chi-square (NC)

Jöreskog, K.G. (1969) proposed that X^2 be adjusted by the degrees of freedom to assess model fit. The NC measure can identify inappropriate models in two ways: (1) a model which is "over-identified" and capitalizes on chance; or (2) models that do not fit the observed data and need improvement. The NC measure, like many others is affected by sample size. It is calculated as:

$$X^2 = \frac{X^2}{df}$$

Parsimonious Fit Index (PFI)

The PFI measure is a modification of the NFI measure (James, Muliak, & Brett, 1982). The PFI however takes into account the

number of degrees of freedom used to obtain a given level of fit. Parsimony is achieved with a high degree of fit for fewer degrees of freedom in specifying coefficients to be estimated. The PFI is used to compare models with different degrees of freedom and is calculated as:

$$PFI = (df_{proposed} / df_{null}) \times NFI$$

Akaike Information Criterion (AIC)

The AIC measure is used to compare models with differing number of constructs similar to the PFI (Akaike, 1987). The AIC measure will always be negative, but values close to zero indicate a more parsimonious model. It indicates both model fit (S and Σ elements similar) and a model not "over-identified" (parsimony). The AIC measure is calculated as:

$$AIC = \frac{-X^2}{2}$$

Muliak et al. (1989) evaluated the X^2, NFI, GFI, AGFI, and AIC goodness-of-fit indices. They concluded that these indices fail to assess parsimony and are insensitive to misspecification of causal relationships. Moreover, they recommend an approach that assesses relative fit of the structural equation model among latent variables (model fit) independent of assessing the fit of the hypothesized relations of indicator variables to latent variables

(measurement model). Their rationale is that with few latent variables, most parameter estimates involve relations among observed indicator variables of latent variables and therefore measurement model estimates rather than latent variable relationship estimates determine a greater proportion of the covariance structure explained. They propose the following adjustment to separately estimate the effects of the latent variable estimates from the measurement model estimates:

$$RNFI_j = (F_u - F_j) / [F_u - F_m - (d_j - d_m)]$$

Where:

$F_u = X^2$ of full model

$F_j = X^2$ of structural equation model (latent variable model)

$F_m = X^2$ of confirmatory factor model (measurement model)

d_j = degrees of freedom for structural equation model

d_m = degrees of freedom for measurement model

A corresponding relative parsimony ratio is given by:

$$RP_j = [d_j - d_m] / [d_u - d_m]$$

Where:

d_j = degrees of freedom for structural equation model

d_m = degrees of freedom for measurement model

d_u = degrees of freedom for null model

To compare different models for fit, multiply RP_j x $RNFI_j$ to obtain a relative parsimonious fit index appropriate for assessing how well and to what degree the models explain relationships both in indicator variable measurement of latent variables and among the the latent variables.

NEW MEASURE OF MODEL FIT

Recent developments have offered another approach to assessing model fit (Joe and Mendoza, 1989a). The maximum internal correlation coefficient has been recommended for further development in structural equation models as a GOF criteria (Joe and Mendoza, 1989b, 243). It could be used to test the equality of pairs of eigenvalues from a correlation matrix or test the null hypothesis of independence in a set of variables. It is also useful in resolving "heywood cases" and multicollinearity problems (identifies which variables are causing the problem). There are several potential problems however in that several internal correlations are possible (pairwise combinations of eigenvalues of a correlation matrix) and the distribution of the statistic is unknown. Venables (1980) however has presented the distribution of the internal correlation and a test of significance which approximates an F distribution under the special case of a 2 x 2 covariance matrix. The maximum internal correlation is calculated as:

$$\rho(^{*}) = (\lambda_1 - \lambda_p) / (\lambda_1 + \lambda_p)$$

Where:

λ_1 = largest eigenvalue in the population correlation matrix

λ_p = smallest eigenvalue in the population correlation matrix

The author proposes that the application of the internal correlation coefficient to structural equation modeling would take the form of:

$$\rho(^*) = (\lambda_S - \lambda_\Sigma) / (\lambda_S + \lambda_\Sigma)$$

Where:

λ_S = largest eigenvalue in original correlation matrix

λ_Σ = largest eigenvalue in reproduced correlation matrix

This would provide an overall measure of the comparison of equality in the elements of the original and reproduced correlation /covariance matrices or between two proposed model correlation matrices.

SUMMARY

Structural equation modeling requires the use of various GOF criteria in combination to interpret a model. Additionally, appropriate sample size, standardization of variables, and normality determinations need to be considered. Appropriate sample size can be determined using the Critical N (CN) statistic (Hoelter, 1983) which is: $CN = (X^2 / F) + 1$. CN gives the sample size at which the F value would lead to a rejection of H_o: $S = \Sigma$ at

a specified alpha level. The CN statistic is output by the SAS/CALIS progr m and indicates the recommended sample size. Standardization of variables occurs automatically when inputing a correlation matrix without mean and standard deviation values specified. This however doesn't occur when inputing a covariance matrix therefore caution is advised in selecting and interpreting GOF criteria because the variables may not be on the same scale of measurement. Normality of variables should further be examined prior to inputing a correlation or covariance matrix in structural equation models. As noted previously, several GOF criteria have evolved as adjustments to the X^2 and subjective interpretation has been made easier with scaling from 0 to 1. However, GOF criteria must be interpreted in view of sample size, standardization of variables, and the normality assumption.

The new GOF criteria, internal correlation coefficient, holds the possibility of yet another statistical test of model fit once the distributional properties are better known. However, model fit is of itself a subjective approach since no single "correct" model is being determined. Other models may be equally plausible given the sample data. What then is the solution to this dilemna? The author suggests <u>model testing</u> for effectiveness and functionality of the specified variable relationships. This implies that the variables are chosen such that they can be manipulated to examine (test) the effect upon other variables in the model. Model testing takes the researcher to the next step beyond determining whether the model "fits" the sample data.

REFERENCES

Akaike, H. (1987). Factor Analysis and AIC. Psychometrika, 52, 317-332.

Anderson, J.G. (1987). Structural equation models in the social and behavioral sciences: Model building. Child Development, 58, 49-64.

Asher, H. (1976). Causal Modeling. CA: Sage Publications.

Baldwin, B. (1989). A Primer in the Use and Interpretation of Structural Equation Models. Measurement and Evaluation in Counseling and Development, 22, 100-112.

Bentler, P.M. & D.G. Bonett (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bentler, P.M. & D.G. Bonett (1982). Significance Tests and Goodness of Fit in Analysis of Covariance Structures. In A Second Generation of Multivariate Analysis: Measurement and Evaluation (Vol. 2). Edited by Claes Fornell, NY: Praeger Publishers.

Bollen, K.A. (1989). Structural Equations with Latent Variables. NY: John Wiley & Sons.

Bollen, K.A. and Ting, K. (1991). Statistical Computing Software Reviews. The American Statistician, 45(1), 68-73.

Boyle, R. (1970). Path analysis and ordinal data. American Journal of Sociology, 75(4), 461-480.

Duncan, O.D. (1975). Introduction to structural equation models. New York: Academic Press.

Fassinger, R. (1987). Use of the structural equation modeling in counseling psychology research. Journal of Counseling Psychology, 34, 425-440.

Hair, J.F. Jr., Anderson, R.E., Tatham, R.L., and Black, W.C. (1992). Multivariate Data Analysis with Readings(3rd Ed.). New York: Macmillan Publishing.

Hayduk, L.A. Structural Equation Modeling with LISREL: Essentials and Advances. MD: Johns Hopkins University Press.

Hoelter, J.W. (1983). The analysis of covariance structures: Goodness-of-Fit indices. Sociological Methods and Research, 11, 325-344.

James, L.R., Muliak, S.A., Brett, J.M. (1982). Causal Analysis: Assumptions, Models, and Data. California: Sage Publications.

Joe, G.W. and J.L. Mendoza (1989a). The internal correlation: Its applications in statistics and psychometrics. Journal of Educational Statistics, 14(3), 211-226.

Joe, G.W. and J.L. Mendoza (1989b). Further comments on the internal correlation: A rejoinder. Journal of Educational Statistics, 14(3), 239-244.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183-202.

Jöreskog, K.G. and Sörbom, D. (1989). LISREL 7 User's Reference Guide. Mooresville, IN: Scientific Software Inc.

Kim, J. and C.W. Mueller (1978a). Introduction to Factor Analysis: What it is and how to do it. CA: Sage Publications.

Kim, J. and C.W. Mueller (1978b). Factor Analysis: Statistical Methods and Practical Issues. CA: Sage Publications.

Loehlin, J.C. (1987). Latent Variable Models: An introduction to factor, path, and structural analysis. NJ: Lawrence Erlbaum.

Lomax, R.G. (1982). A guide to LISREL-type structural equation modeling. Behavior Research Methods and Instrumentation, 14, 1-8.

Long, J.S. (1983). Confirmatory Factor Analysis. CA: Sage Publ.

Marascuilo, L. and J. Levin (1983). Multivariate Statistics in the Social Sciences: A Researcher's Guide. CA: Brooks Cole.

Muliak, S.A., James, L.R., Alstine, J.V., Bennett, N., Lind, S., Stilwell, C.D. (1989). Evaluation of Goodness-of-fit Indices for Structural Equation Models. Psychological Bulletin, 105(3), 430-445.

Pedhazur, E.J. (1982). Multiple Regression in Behavioral Research (2nd Ed.). NY: Holt Rinehart Winston.

Plewis, I. (1985). Analyzing Change: Measurement and Explanation Using Longitudinal Data. New York: John Wiley & Sons.

Saris, W. and Stronkhorst, H. (1984). Causal Modelling in Nonexperimental Research. Amsterdam: Sociometric Research Fnd.

Tucker, L.R. and Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Venables, W. (1976). Some implications of the union-intersection principle for tests of sphericity. Journal of Multivariate Analysis, 6, 185-190.

Williams, J.D. and R.J. Klimpel (1974). Path analysis and causal models as regression techniques. Multiple Linear Regression, Viewpoints, 5(3), 1-20.

Table 1: Goodness-of-fit Criteria in Statistical Packages

Package/Program	GOF Criteria
SAS / CALIS	Chi-square
	Goodness-of-Fit Index
	Adjusted Goodness-of-Fit
	Root Mean Square Residual
	*Bentler Comparative Fit Index
	*Normal Theory Reweighted LS Chi-square
	Akaikes Information Criterion
	*Consistent Information Criterion
	*Schwarz Bayesian Criterion
	*McDonald Centrality
	*Bentler & Bonett Non-normed Fit Index
	Bentler & Bonett Normed Fit Index
	James, Muliak, & Brett Parsimonious Fit Index
	*Wilson & Hilferty Z-test
	*Bollen Normed Index Rho1
	*Bollen Non-normed Index Delta2
	Hoelter Critical N
SPSS / LISREL 7	Chi-square
	Goodness-of-Fit Index
	Adjusted Goodness-of-Fit Index
	Root Mean Square Residual
BMDP / EQS	Chi-square
	Goodness-of-Fit Index
	Adjusted Goodness-of-Fit Index
	Akaikes Information Criterion
	*Bozdogan Consistent AIC
	Normal Theory Reweighted LS Chi-square
	*Bentler & Bonett Non-normed Fit Index
	Bentler & Bonett Normed Fit Index
	*Bentler Comparative Fit Index
Scientific Software/ LISREL 7	Chi-square
	Goodness-of-Fit Index
	Adjusted Goodness-of-Fit Index
	Root Mean Square Residual

Note: The following require individual calculation:
Tucker-Lewis
Normed Chi-square
Internal Correlation

* These indices are not covered in the manuscript, consult the respective statistical user guide for a reference.

Table 2: GOF criteria, acceptable fit levels and interpretation.

GOF Criteria	Acceptable Level	Interpretation
Chi-square	Tabled X^2 value	Compare obtained X^2 value to tabled value for given degrees of freedom. ML and GLS estimates preferred for covariance matrix and ULS for correlation matrix.
Goodness-of-Fit (GFI)	0 (no fit) to 1 (perfect fit)	Value close to .9 reflects a good fit.
Adjusted GFI	0 (no fit) to 1 (perfect fit)	Value adjusted for df with .9 reflecting a good model fit.
Root Mean Square (RMR)	Researcher defines level	Relationship to observed S residual variance/covariance
Tucker-Lewis	0 (no fit) to 1 (perfect fit)	Value close to .9 reflects a good model.
Normed Fit Index	0 (no fit) to 1 (perfect fit)	Value close to .9 reflects a good model.
Normed Chi-square	1.0 to 5.0	Less than 1.0 is a poor model fit. More than 5.0 reflects need for improvement. Differences of .06 to .09 when comparing models.
Parsimonious Fit Index	0 (no fit) to 1 (perfect fit)	Compare alternative models.
Akaike Information Criterion	0 (perfect fit) - (poor fit)	Compare alternative models.
Internal Correlation	0 ($S = \Sigma$) to 1 ($S \neq \Sigma$)	Model Fit between S and reproduced Σ.