DOCUMENT RESUME

ED 351 389 TM 019 227

AUTHOR Fraas, John W.; Newman, Isadore

TITLE A Binomial Test of Model Fit.

PUB DATE Apr 92

NOTE 23p.; 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 Career Development; Estimation (Mathematics); *Evaluation Methods; *Goodness of Fit; *Heuristics; *Hypothesis Testing; *Models; Statistical Significance

IDENTIFIERS *Binomial Test; T Values

ABSTRACT

A new method for evaluating model fit that is easy to use and interpret is presented. The new method, which uses a binomial test of the number of hypotheses (paths) in a model that are supported by the data, has heuristic value when considering problems associated with other goodness-of-fit measures. An application of the binomial test as a goodness-of-fit measure is presented. The procedure is applied to a career deveiopment model studied by K. K. Sidhu (1988). A hypothesis (path) is judged to be supported by the data when the parameter estimate possesses the hypothesized sign and is statistically significant. For the model, the parameter estimates were expected to be positive and T values were expected to exceed two before the coefficient would be considered statistically significant. Since all seven of the parameter estimates had positive signs and T values in excess of two, all seven hypotheses are judged to be supported by the data. In the final step, the binomial probability value is calculated. Two tables and one figure illustrate the application. (SLD)

A Binomial Test of Model Fit

John W. Fraas

Ashland University

Isadore Newman

The University of Akron

A Paper Presented at the 1992 Annual Meeting of the American Educational Research Association, San Francisco, California

April 20-24, 1992

Running head: A BINOMIAL TEST

Abstract

A new method for evaluating model fit that is easy to use and interpret is presented. The new method, which utilizes a binomial test of the number of hypotheses (paths) in a model that are supported by the data, has heuristic value when considering problems associated with other goodness-of-fit measures. An application of the binomial test as a goodness-of-fit measure is presented.

A Binomial Test of Model Fit

There has been a substantial increase in the number of studies by social scientists during the past twenty years that use path analytic models. A major advantage of the modeling technique is that it involves an investigation of a theoretical framework. That is, the theory is represented in the path diagram. A researcher using the modeling approach generally determines the degree to which the data fit the theoretical model being investigated. A variety of goodness-of-fit (GOF) measures have been proposed to test model fit (Bentler & Bonett, 1980; Bentler & Weeks, 1980; Tanka & Huba, 1985; Joreskog & Sorbom, 1986; Mendoza & Mueller, 1990; Schumacker, 1992).

The most commonly used procedures is the chi-square goodness-of-fit test (Bentler & Weeks, 1980). This chi-square test is designed to test whether a given model provides an acceptable fit of the observed data. The degree of fit of the model and observed data is evaluated by testing the null hypothesis with a chi-square test. In this application of the chi-square test, the chi-square value will increase as the differences between the unconstrained covariance matrix and the constrained covariance matrix increases. Thus, if the null hypothesis is rejected, the researchers will conclude that the model does not adequately reproduce the observed covariance matrix.

A number of writers have questioned the use of the chi-square goodness-of-fit test for various reasons (Bentler & Bonett, 1980; Long, 1983; Joreskog & Sorbom, 1986). Joreskog and Sorbom (1986) expressed the concern that:

> ... in most empirical work the model is only tentative and is only regarded as an approximation to reality. From this point of view the statistical problem is not one of testing a given hypothesis (which a priori may be considered false) but rather one of fitting the model to the data and to decide whether the fit is adequate or not (p. I.38).

A second concern with the use of the chi-square statistic as a goodness of fit test centers on the fact that its level of significance is related to the sample size. That is, as the sample size increases, even small differences between the constrained covariance matrix and unconstrained covariance matrix may be statistically significant. Bentler and Bonett (1980) referred to such a model as being minimally false.

A third concern addresses the inability of the chi-square goodness-of-fit test to indicate where the model does not match the theory. That is, the paths that are not supported by the data will not be identified by this test.

Based on these and other concerns, Joreskog and Sorbom (1986) suggest that the chi-square test not be used as a test statistic, but rather the researcher should interpret a large chi-square value as indicating a bad fit between the data and the model. In a similar manner, a small chi-square value would indicate a good fit. Degrees of freedom could be used to determine whether the chi-square value was large or small (See Carmins & McIver, 1981, and Hoelter, 1983, for discussions of the use of degrees of freedom as the standard by which to judge the size of the chi-square value.)

Two other goodness-of-fit tests used with structural equation models are the goodness-of-fit index (GFI) and the adjusted goodness-of-fit index (AGFI). The GFI indicates the relative amount of variance and covariance jointly accounted for by the model. The AGFI is the GFI adjusted for the model's degrees of freedom. The values for both the GFI and the AGFI normally range between 0 and 1, although Joreskog and Sorbom (1986) state that it is theoretically possible for these values to become negative.

The major problem with these two indexes is that their statistical distributions are unknown; thus, there is no standard to which one can compare them (Joreskog & Sorbom, 1986). Volkan (1991) suggests, however, that GFI values of .9 or higher are generally considered to indicate an adequate fit of the model.

Another problem noted by other researchers (March, Balla & McDonald, 1988; Mendoza & Mueller, 1990) is that the GFI and AGFI are related to the sample size. Joreskog and Sorbom (1986) discussed two other problems with the GFI and the AGFI. First, these measures may indicate that the overall fit of the model is adequate, but one or more of the relationships in the model may be poorly determined. Second, if these measures indicate that the model does not fit the data well, they do not provide information regarding what is wrong with the model. That is, the paths of the model not supported by the data are not revealed by these measures.

This paper presents a goodness-of-fit measure that conceptually differs from the chi-square, GFI and AGFI goodness-of-fit measures. The proposed goodness-of-fit measure uses a binomial test to determine whether the number of hypotheses supported by the model is due to chance (see Newman, 1991, for an initial discussion of this concept).

The Binomial Test Approach

A binomial test requires that each event be classified into one of two categories (Seigel and Castellan, 1988). Thus, the first step in implementing the binomial test as a goodness-of-fit measurement of a theoretical model is to determine the criteria that will be used to judge whether a given hypothesis (path) is supported by the data. The researcher could use various criteria to judge whether a hypothesis (path) is supported by the data: (a) The parameter estimate exceeds an a priori effect size, (b) the parameter estimate is statistically significant, (c) the parameter estimate reflects the hypothesized sign, or (d) a combination of these criteria could be used. To illustrate how different criteria could be used, consider studies in which directional or nondirectional hypotheses are used.

If directional hypotheses are the only type of hypotheses used in the model, the researcher could determine the number of hypotheses supported by the data in two ways. In one procedure, the researcher would count a hypothesis as being supported by the data simply by determining if the sign of the parameter estimate matched the hypothesized sign. This technique would produce a binomial test of the model's goodness of fit in which the power of the test would be independent of the sample size. Instead, the power of the test would be dependent on the number of hypotheses derived from the model. That is, as the number of hypotheses derived from the model increased, the power of the test would increase.

With the other criterion that could be used with directional hypotheses, the researcher would classify a given hypothesis as being supported by the data when the corresponding parameter estimate is statistically significant and it possesses the hypothesized sign. This technique of determining the number of hypotheses supported by the data for use in the binomial test would produce a goodness-of-fit estimate that is related to the sample size as well as the number of hypotheses derived from the model. That is, as the sample size increased, the power of the test of any given hypothesis derived from the model would increase. The power of the binomial test would still be dependent on the number of hypotheses derived from the model. As that number increases, the power of the binomial test increases.

If a researcher uses only nondirectional hypotheses to represent a model, the sign of the estimated parameter for any given hypothesis could not be used to determining if the data supported the hypothesis. The number of hypotheses supported by the data could only be determined by the examining the statistical significance of each hypothesis. If the parameter estimate was statistically significant, the hypothesis would be counted as being supported by the data. When the number of hypotheses supported by the data is determined in this manner, the power of the statistical test of each hypothesis is related to both the sample size and the number of hypotheses derived from the model.

If a researcher has a model that incorporates both directional and nondirectional hypotheses, the number of hypotheses supported by the data could be counted in various ways. In the one method, the directional hypotheses would be counted as being supported by the data when the signs of the parameter estimates matched the hypothesized signs; and the nondirectional hypotheses would be counted as being supported by the data when the parameter estimates were statistically significant. In another method, only hypotheses that were statistically significant, regardless of whether they were directional or nondirectional, would be considered supported by the data. Using either of these two methods of counting the number of hypotheses would produce a goodness-of-fit estimate that was related to the total number of hypotheses derived from the model and the sample size.

Once the criteria for determining whether a given hypothesis (path) is supported by the data, the second step in the binomial test approach requires the researcher to determine which parameter estimates support the research hypotheses. An examination of which hypotheses are not supported by the data may provide some insight into where the model is weak. In the final step, the researcher uses a binomial test to provide a conservative and robust probability estimate of obtaining no more than the number of hypotheses identified by the researcher as not being supported by the data. If the probability produced by the binomial test is less than the alpha level, the researcher can conclude that the data are supportive of the model.

An Application of the Binomial Test

To illustrate the use of the binomial test of model fit, the procedure is applied to a career development model studied by Sidhu (1988). The model is presented in Figure 1, and the variables in the model are listed in Table 1.

Insert Figure 1 about here

Insert Table 1 about here

As previously discussed, in the first step, the researcher must determine what criterion will be used to judge whether the data support a given hypothesis (path) of the model. In this example, a hypothesis (path) was judged to be supported by the data when the parameter estimate possessed the hypothesized sign and it was statistically significant. For the model proposed by Sidhu (1988), the parameters were expected to possess positive signs. The criterion suggested by Joreskop and Sorbom (1986, p. III.12) that the *t* value of a parameter estimate should exceed 2 before the coefficient is deemed to be statistically significant was used.

The seven equations that represent the paths in the model are as follows:

1: SCOPE $= \beta_{21}$ LMX + e_1

2: SUPSAT $= \beta_{31}$ LMX + e_2

3: CRE $= \beta_{42}$ SCOPE + e_3

4: WS $= \beta_{54}$ CRE + e_4

5: KAO $= \beta_{64}$ CRE + e_5

6: KAS $= \beta_{74}$ CRE + e_6

7: ABILITY $= \beta_{84}$ CRE + e_7,

where the β values represent path coefficients and e equals the error terms. The parameter estimates and their t values are listed in Table 2. Since all seven of the parameter estimates had positive signs and t values in excess of 2, all seven hypotheses (paths) were judged to be supported by the data in Step 2 of the binomial test procedure. In the final step, the following formula was used to calculate the binomial probability value:

$$p(x) = \frac{n!}{x!(n-x)!} (.5)^{x} (.5)^{(n-x)}$$

where:

1. p is equal to the probability of obtaining x hypotheses (paths) not

supported by the data out of *n* number of hypotheses (paths).

2. *x* is equal to the number of hypotheses (paths) not supported by the data.

3. *n* is equal to the number of hypotheses (paths) in the model.

The calculation is as follows:

$$p(x \leq 0) = \frac{7!}{0!7!}(.5)^0(.5)^7 = .008$$

Since the probability (.008) of having 0 out of 7 hypotheses being supported by the data is less than the alpha level of .05, the researcher would conclude that the number of hypotheses supported by the data is greater than one would expect by chance. In addition, it should be noted that in the second step of the binomial test procedure, all seven paths were supported by the data. If that had not been the case, the paths not supposted by the data could have been identified for possible further study.

Conclusion

The binomial test of model fit has heuristic value since it is applied to the number of hypotheses (paths) supported by the data. In the binomial test approach, determining whether each hypothesis (path) is or is not supported by the data is important. Thus, the criteria used to determine whether a given hypothesis (path) is supported by the data is the crucial first step in this GOF measurement.

In this procedure a researcher could select one or more criteria to judge whether a given hypothesis is supported by the data. The criteria are: (a) The parameter estimate exceeds an a priori effect size; (b) the parameter estimate is statistically significant; (c) the sign of the parameter estimate matches the hypothesized sign or (d) any combination of these criteria. Another criteria, which was not examined in this paper, that might be appropriate is differential weighting of each hypothesis (path). In our example, all hypotheses had equal importance in the model. This criteria might not be appropriate for some models. Whether one should give equal weight to all hypotheses is not a statistical decision, but rather a logical and theoretical one. Although we did not address how one could differentially weight hypotheses based on their relative theoretical importance to the model, we do think that this approach needs to be investigated. We believe that the differential weighting of hypotheses would be consistent with a scientific framework that requires a statistical test to reflect a researcher's question of interest. This would avoid a Type VI error which is the inconsistency between the statistical procedure used by a researcher and the research question (Newman, Deitchman, Burkholder, Sanders & Ervin, 1976).

It should be noted that the decision regarding which criteria to use is not based on statistical manipulation, rather the logical framework of a theory. The binomial test of model fit involves the number of hypotheses (paths) supported by the data; and, therefore, it does not matter whether the variables are observed or latent. Thus, the binomial test can be applied in structural equation/covariance models. It may also prove useful in the two-step approach to structural equation modeling where the structural model is evaluated separate from the measurement model (Anderson & Gerbing, 1988; Muliak, 1989).

References

Anderson, J.C. & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103(3), 411-423.

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

Bentler, P. M. & Weeks, D. G. (1980). Linear structural equations with latent variables. Psychometrika, 45, 289-308.

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

Joreskog, K. C., & Sorbom, D. (1986). LISREL VI: analysis of linear structural relationships by the method of maximum likelihood (3rd ed.). Moorsville, IN: Scientific Software Inc.

Long, J.S. (1983). Covariance structure models: An introduction to LISREL. Beverly Hills, CA: SAGE.

Marsh, H., Balla, J., & McDonald, R. P. (1988).

Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size. Psychological Bulletin, 103, 391-410.

Mendoza, T. R., & Mueller, R. (1990, October). Comparative study of selected fit indices in structural equation modeling. A paper presented at the meeting Mid-Western Educational Researcher Association, Chicago, IL.

Muliak, S.A., James, L.R., Alstine, J.V., Bennett, N., Lind, S., and Stilwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105(3), 430-445.

Newman, I. (1990, April). To path analyze or not to path analyze: is there an alternative approach?. A paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Newman, I. Deitchman, R., Burkholder, J., Sanders, R., & Ervin, L. (1976). Type VI error: inconsistency between the statistical procedure and the research question. Multiple Linear Regression Viewpoints, 6(4), 1-19.

Schumacker, R.E. (1992, April). Goodness-of-fit criteria in structural equation models: A summary with recommendations. A paper presented at the American Educational Research Association, San Francisco, CA.

Seigel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.). New York: McGraw-Hill.

Sidhu, K. K. (1988). The role of a leader-member exchange and work experience in the early career development of professionals. Unpublished doctoral dissertation, University of Cincinnati, Cincinnati, OH.

Tanka, J. S., & Huba, G. (1985). A fit index for covariance structure models under arbitrary GLS estimation. British Journal of Mathematical and Statistical Psychology, 38 , 197-201.

Volkan, K. (1990, April). Using LISREL: some notes on technique and interpretation. A paper presented at the meeting of the American Education Research Association, Chicago, IL.

Table 1

Variables Included In the Model[a]

Variable	Variable Label	Variable Number
Leader-Member Exchange	LMX	1
Job Scope	SCOPE	2
Supervision Satisfaction	SUPSAT	3
Career Related Experiences	CRE	4
Occupation Related Abilities	ABILITY	5
Knowledge About Occupation	KAO	6
Knowledge About Self-In-Occupation	KAS	7
Work Satisfaction	WS	8

[a]Career development model (Sidhu, 1988).

Table 2

Path Coefficients and t-test values for Model

Path	β	t^a
p_{21}	0.63	7.41
p_{31}	0.47	15.42
p_{42}	0.86	10.00
p_{54}	0.63	7.84
p_{64}	0.76	9.41
p_{74}	1.14	12.82
p_{84}	0.81	10.85

[a]The criteria of $t > 2.00$ was used for testing the statistical significance of a path coefficient (Joreskog and Sorbom, 1986, p. III.12).

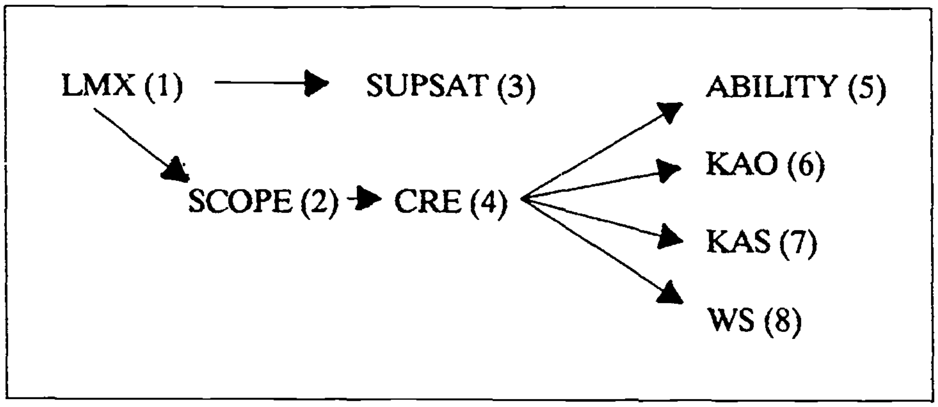

Figure 1. Career Development Model (Sidhu, 1988).