A NOTE ON NEYMAN'S TEST OF GOODNESS OF FIT USING THE EXTENDED LISREL MODEL WITH MEAN STRUCTURES

Daisy Bertrand, Abdessadek El Ahmadi

▶ **To cite this version:**

Daisy Bertrand, Abdessadek El Ahmadi. A NOTE ON NEYMAN'S TEST OF GOODNESS OF FIT USING THE EXTENDED LISREL MODEL WITH MEAN STRUCTURES. 2023. hal-04322688

HAL Id: hal-04322688

https://amu.hal.science/hal-04322688

Preprint submitted on 5 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

A NOTE ON NEYMAN'S TEST OF GOODNESS OF FIT USING THE EXTENDED LISREL MODEL WITH MEAN STRUCTURES

Daisy BERTRAND, PhD
Aix Marseille Univ, CERGAM, Aix-en-Provence, France

Abdessadek El AHMADI, PhD
Laboratoire de Neurosciences Cognitives (LNC), Centre National de la Recherche Scientifique (CNRS), Aix-Marseille Université, 13007 Marseille, France

Fédération de Recherche 3C, Centre National de la Recherche Scientifique (CNRS), Aix-Marseille Université, 13003 Marseille, France

Address:
Aix-Marseille University Cognitive
Psychology Dpt Avenue R. Schuman, 29
F-13621 Aix-en-Provence France

Tel : ++ (33) 4 42 95 37 25
e-mail : abdessadek.el-ahmadi@univ-amu.fr

A NOTE ON NEYMAN'S TEST OF GOODNESS OF FIT USING THE EXTENDED LISREL MODEL WITH MEAN STRUCTURES

SUMMARY

It is shown in this note that for testing discrepancy between a set of observed proportions $p_1, p_2, ..., p_k$ in k mutually exclusive cells and corresponding expected probabilities , $\pi_1, \pi_2, ..., \pi_k$, Neyman's statistic is n / (n – 1) times the Jöreskog and Sörbom chi-square goodness-of-fit of a factor model with no latent variables but with mean structures for (k – 1) binary indicators coding for the k-cell category. When the empirical data lead to a rejection of hypothesized model, the modification index gives useful information about post hoc tests for categorical data.

A NOTE ON NEYMAN'S TEST OF GOODNESS OF FIT USING THE EXTENDED LISREL MODEL WITH MEAN STRUCTURES

Consider a single k-cell multinomial distribution with cell nonzero probabilities $\pi_1, \pi_2, ..., \pi_k$, with $\sum_{i=1}^{k} \pi_i = 1$. Let $n_1, n_2, ..., n_k$ be the observed cell frequencies with $\sum_{i=1}^{k} n_i = n$ and let $\hat{\pi}_i = p_i = \frac{n_i}{n}$ be a consistent estimator of $\pi_i (i = 1,2,...,k)$. If the k binary variables $u_1, u_2, ..., u_k$ are indicator variables for k possible outcomes of a categorical variable, these binary variables are linearly dependent so that only the first (k-1) of them will be considered. The expectation of each u_i is $E(u_i) = \pi_i$, $(\mathrm{i} = 1, 2, ..., \mathrm{k} - 1)$ and the variance-covariance matrix Vp of the (k-1) indicator variables is given by (Kendall & al., 1987; Bishop & al., 1975)

$$V_p = \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \dots & -\pi_1\pi_{k-1} \\ -\pi_2\pi_1 & \pi_2(1-\pi_2) & \dots & -\pi_2\pi_{k-1} \\ \vdots & \vdots & \vdots & \vdots \\ -\pi_{k-1}\pi_1 & -\pi_{k-1}\pi_2 & \dots & \pi_{k-1}(1-\pi_{k-1}) \end{pmatrix}$$

for which

$$V_p^{-1} = \begin{pmatrix} \frac{1}{\pi_1} + \frac{1}{\pi_k} & \frac{1}{\pi_k} & \dots & \frac{1}{\pi_k} \\ \frac{1}{\pi_k} & \frac{1}{\pi_2} + \frac{1}{\pi_k} & \dots & \frac{1}{\pi_k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\pi_k} & \frac{1}{\pi_k} & \dots & \frac{1}{\pi_{k-1}} + \frac{1}{\pi_k} \end{pmatrix}$$

Let $\vec{\pi} = (\pi_1, \pi_2, ..., \pi_{k-1})$ and $\vec{p} = (p_1, p_2, ..., p_{k-1})$. Asymptotically, $n^{1/2}(\vec{p} - \vec{\pi})$ has a (k-1)-variate normal distribution with zero mean vector and variance-covariance matrix Vp. Thus, the quadratic form

$$Q_p = n(\vec{p} - \vec{\pi})'(V_p^{-1})(\vec{p} - \vec{\pi}) = n\sum_{i=1}^{k} \frac{(p_i - \pi_i)^2}{\pi_i}$$

is distributed like χ^2 with (k-1) degrees of freedom. This result, due to Pearson (1900), is the basis of the so-called Pearson's goodness-of-fit test; this test is used as an index of discrepancy between a set of observed proportions in k mutually exclusive cells and corresponding expected probabilities for a specific hypothesis under investigation.

For situations where all p_i>0, Neyman (1949) suggested an alternative to the method of Pearson's chi-square, where the quadratic form is modified to

$$Q_N = n(\vec{p}-\vec{\pi})'(V_N^{-1})(\vec{p}-\vec{\pi}) = n\sum_{i=1}^{k}\frac{(p_i-\pi_i)^2}{p_i}$$

with

$$V_N = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_{k-1} \\ -p_2p_1 & p_2(1-p_2) & \cdots & -p_2p_{k-1} \\ \vdots & \vdots & \vdots & \vdots \\ -p_{k-1}p_1 & -p_{k-1}p_2 & \cdots & p_{k-1}(1-p_{k-1}) \end{pmatrix}$$

and

$$V_N^{-1} = \begin{pmatrix} \frac{1}{p_1}+\frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \frac{1}{p_k} & \frac{1}{p_2}+\frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_{k-1}}+\frac{1}{p_k} \end{pmatrix}$$

The purpose of this note is to point out that the Neyman statistic Q_N may also be obtained by the Jöreskog and Sörbom (2001) criterion Q_{JS} for the extended Lisrel model with mean structures (see also Bollen, 1989).

$$Q_{JS} = (n-1)\left\{(\vec{s}-\vec{\sigma})'W^{-1}(\vec{s}-\vec{\sigma}) + (\vec{m}-\vec{\mu})'V^{-1}(\vec{m}-\vec{\mu})\right\}$$

The first part contains $\vec{s}$, the vector of the elements in the lower half, including the diagonal, of the sample covariance matrix S used to fit the model to the data. It also contains $\vec{\sigma}$, the vector of corresponding elements of Σ reproduced from the model parameters. The W matrix serves as a correct weight matrix (see Joreskog & Sorbom, 2001). The second term involves the sample mean vector $\vec{m}$, the population mean vector $\vec{\mu}$, and the weight matrix V defined as the sample covariance matrix S for Generalized Least Squares (GLS) estimation. If the model holds and is identified, Q_{JS} is approximately distributed in large samples as χ^2 with (w-v) degrees of freedom, where w is the number of statistics (observed means and covariances) and v is the number of independent parameters estimated.

For obtaining the Neyman statistic, is candidate every model which cancels the first term by fixing $\vec{s} = \vec{\sigma}$, and which lets free the second term which involved the sample mean vector $\vec{m}$ and the population mean vector $\vec{\mu}$. As will be argued, the Neyman goodness-of-fit test can be obtained by constraining an extended confirmatory factor analysis with mean structures. That is, the Neyman goodness-of-fit test is equivalent to the Jöreskog and Sörbom criterion when are specified and estimated by GLS the mean parameters of the confirmatory factor model, with binary indicators and with no latent variables.

The extended confirmatory factor model with mean structures is defined by the following (Jöreskog and Sörbom, 2001)

$$\vec{y} = \vec{\tau}_y + \Lambda_y \vec{\eta} + \vec{\varepsilon}$$

where $\vec{y}$ is a vector of observed binary variables, $\vec{\tau}_y$ a vector of constant intercept terms, $\vec{\eta}$ a vector of latent variables, Λ_y a factor-loading matrix and $\vec{\varepsilon}$ a vector of error terms. It is traditionnally assumed that $\text{cov}(\vec{\eta}, \vec{\varepsilon}) = 0$ and $E(\vec{\varepsilon}) = \vec{0}$ and $E(\vec{\eta}) = \vec{0}$. As a matter of interest, factor analysis of binary variables was initiated by Christoffersson (1975) and later expanded by Muthén (1978, 1983, 1984) and Muthén and Christoffersson (1981).

As alluded to above, the Neyman chi-square can be obtained by specifying $\Lambda_y = 0$, $\vec{\tau}_y$ to be fixed equal to $\vec{\pi}$ and $\text{cov}(\vec{\varepsilon}, \vec{\varepsilon}) = \Theta_\varepsilon$ to be completely free. In this case, the fit function F for the extended Lisrel model with mean structures (Jöreskog and Sörbom, 2001; Jöreskog et al., 2000) reduces to be $F = (\vec{m} - \vec{\mu})' V^{-1} (\vec{m} - \vec{\mu})$ where V is the sample covariance matrix S for GLS estimation. Here, the observed covariance matrix S of the linearly independent indicators is equal to V_N. Hence

$$Q_{JS} = (n-1)F = (n-1)(\vec{m} - \vec{\mu})' V^{-1} (\vec{m} - \vec{\mu}) = (n-1)(\vec{p} - \vec{\pi})' V_N^{-1} (\vec{p} - \vec{\pi}) = \frac{n-1}{n} Q_N$$

which has a distribution like a χ^2 with (w-v) degrees of freedom, where w is the number of statistics and v is the number of independent parameters estimated (see also Christoffersson, 1975, appendix 1). The independent parameters are those of the completely free matrix Θ_ε, and the statistics are from the covariance matrix V_N and the vector $\vec{\text{p}}$. In general, (w–v) equals the number of restrictions imposed to the parameters $\vec{\pi}$ by the hypothetical underlying model, in general (k – 1). Thus, Jöreskog and Sörbom statistic Q_{JS} reduces to the Neyman statistic

$Q_{N,}$ a part from the factor (n-1)/n. In other words, the Neyman statistic Q_N is n times the minimum value of the fit function whereas Q_{JS} is (n-1) times this minimum. If the model holds, both statistics Q_{JS} and Q_N converge on the same value and both have the same asymptotic distribution under Ho.

To illustrate, we consider a single multinominal distribution corresponding to a categorical variable with 4 categories : the observed proportions and the expected probabilities are respectively $\vec{p} = (0.35; 0.10; 0.15; 0.40)$ and $\vec{\pi} = (0.15; 0.30; 0.45; 0.10)$. The sample size is n = 40. The observed covariance matrix S for the three first linearly independent indicator variables is

$$S = \begin{pmatrix} 0.2275 & & \\ -0.0350 & 0.0900 & \\ -0.0525 & -0.0150 & 0.1275 \end{pmatrix}$$

The data analysis was performed on a Sony micro-computer using Proc Calis of the SAS software package (SAS Institute, 2004). A generalized least squares (GLS) estimation method was used. The goodness-of-fit indices described earlier are Q_P (3) = 60.00, Q_{JS} (3) = 52.23 and Q_N (3) = 53.5. Hence, for GLS estimation, the Q_{JS} is (n-1) times the minimum value of the fit function for the specified model, and the Neyman Q_N is n times this minimum. This minimum is equal here to F = 1.3393.

CONFIRMATORY APPROACH AND MODIFICATION INDEX

The empirical example presented above for demonstrating the equivalence between Neyman's test and the Joreskog-Sorbom' test was rather trivial. However, the confirmatory approach to modeling multinomial distribution offers an interesting extension of the classical goodness-of-fit tests. In a confirmatory analysis, the investigator has such knowledge about the problem that he/she is able to specify not only the expected probabilities but also other constraints (equality, order, zeroes ..). Because the pattern of constrains is specified a priori on the basis of theory, the confirmatory approach is particularly well suited to the theoretical questions posed in a study. Whithin this approach, the Neyman's chi-square Q_N is of interest because it has been shown by Bhapkar (1961, 1966) to be algebraically identical to the Wald statistic (1943), as adapted to the categorical data, for testing the hypothesis

$$H_0 = F_i(\vec{\pi}) = 0\,,\ (i = 1,2,...,u)$$

where F's are u prespecified fonctions of $\vec{\pi}$. Test statistics such as the Wald statistics and their application to problems involving the analysis of discrete or categorical data are discussed in Grizzle & al. (1969), Koch & al. (1977) and Kosh & Bhapkar (1982). Details concerning their general theoretical properties are given in Stroud (1971) ans Sen and Singer (1993). Thus, at a practical level, an important question is how can we identify the 'best' model when comparing several alternative well-fitting models, as the finding that a given model provides an adequate fit, provides little information in the absence of alternative models which may fit just as well, or better. Another important question is when the empirical data lead to a rejection of an ill-fitting and presumably incorrect model, how can we modify it or respecify it by either freeing parameters that formerly were fixed or fixing parameters that formerly were free. A modification index was proposed by Sörbom (1989) to serve as a guide in the search for reducing the discrepancy between model and data when one parameter is freed or when equality constraint is relaxed. In our context, relaxing a fixed parameter or an equality constraint simply means moving one parameter from $\vec{\pi}$ to $\vec{p}$. For each fixed parameter, the modification index (MI) is a measure of predicted decrease in χ^2 if a single constraint is relaxed and the model is reestimated. This suggests ways in which the fit may be improved.

Let us illustrate our approach with an example taken from Haberman (1978) who presents data concerning the political views of subjects interviewed during the U.S. 1975 General Social Survey (NORC, 1975).

Insert Table 1 here

Haberman suggests testing the symmetry of political views model whereby categories 1 and 7 should be equally likely, as should categories 2 and 6, and categories 3 and 5. The data, the expected frequencies estimated for this model, the value of Q_{JS}, Q_P and Q_N are given in table 1. These statistics are distributed approximately as χ^2 with degrees of freedom equal to the number of different response patterns minus one minus the number of independent contraints, (here three equality constraints $\pi_1 = \pi_7$, $\pi_2 = \pi_6$ and $\pi_3 = \pi_5$). Other fit indices were given by the Calis Procedure and only a few of them are listed here (see Table 2).

Insert Table 2 about here

It is important to note that good-fitting models should produce consistent results on many different indices (see Bollen, 1989). The Goodness of Fit Index (Jöreskog & Sörbom, 2001), the Comparative Fit Index CFI (Bentler, 1990) and the Steiger and Lind (1980; Steiger, 1998) root mean squared error approximation (RMSEA) coefficient are the most frequently reported overall fit indices. The Akaike Information Criterion AIC (Akaike, 1987) is a helpful index to use when comparing models. These indices are here GFI=0.9985, CFI=0.9968, RMSEA=0.0313, and AIC=1.1090 (see Table 2a). In addition to the observed proportions, Table 2 gives detailed assessment of fit: the fitted probabilities under the symmetry hypothesis, the t value for each parameter estimate, the modification index for each parameter, and the corresponding estimated change when the parameter is freed. All these statistics show a good fit and no need to seek alternative models to the symmetry one (see Table 2b).

Discussion

The emphasis of Pearson or Neyman is on testing rather than estimation of cell probabilities or parameters. The method presented herein represents a particular formulation of the Neyman's test with linear or non-linear constraints. The advantage of this formulation is that it allows one to test hypotheses regarding importance of a subset of parameters within a general confirmatory model. Another advantage is given by the Lisrel approach which is to allow us great latitude in our choice of models. For example, when the analysis of empirical data leads to a rejection of a hypothesized model, the researcher must reformulate the model in some way, even there is, in most instances and at least theoretically, an overwhelming number of possible actions that could be taken for a better model. For instance, the modification index measures how much we will be able to reduce the discrepancy between model and data when one parameter is added or freed or when one equality constraint is relaxed. This also may give useful information about the multiple comparison procedures for categorical data. It is also known that the Lisrel models can also be used to analyze data from several groups with some or all parameters constrained to be equal over groups (see Sörbom, 1974; Jöreskog & Sörbom, 2001). In this case, a homogeneity test a multi-sample of constraints across groups assesses whether paired observations on two variables, expressed in a contingency table, are independent of each other – for example, whether people from different regions differ in the frequency with which they report that they support a political

candidate. When specifying the model, the vectors $\vec{\tau}_y$ and the error covariance matrices Θ_ε may be postulated to be invariant over groups.

REFERENCES

AKAIKE, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317-332.

BENTLER, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

BHAPKAR, V.P. (1961). Some tests for categorical data. *Annals of Mathematical Statistics, 32,* 72-83.

BHAPKAR, V.P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statististical Association, 61,* 228-235.

BISHOP, Y.M.M., FIENBERG, S.E. & HOLLAND, P.W. (1975). *Discrete Multivariate Analysis.* Cambridge, MA: MIT Press.

BOLLEN, K.A. (1989). *Structural equations with latent variables.* New York: Wiley

BROWNE, M. (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, 40,* 5-32.

GRIZZLE, J.E., STARMER, C.P. & KOCH, G.G. (1969). Analysis of categorical data by linear models. *Biometrics, 25,* 489-505.

HABERMAN, S.J. (1978). *Analysis of Qualitative Data, Vol. 1: Introductory Topics.* New York: Academic Press.

JORESKOG, K.G. & SORBOM, D. (2001). *LISREL 8 user's reference guide* [Computer software manual]. Lincolnwood IL: Scientific Software International.

JORESKOG, K.G., SORBOM, D., Du TOIT, S., & Du TOIT, M. (2000). *LISREL 8: New Statistical Features. Second printing with revisions*. Lincolnwood: ScientificSoftware International;

KENDALL, M., STUART, A. & ORD, J.K (1987). *Kendall's Advanced Theory of Statistics*. New York: Oxford University Press, 604 pages.

KOCH, G.C. & BHAPKAR, V.P. (1982). Chi-square tests, in Johnson N.L. and Kotz S. (Eds), *Encyclopedia of Statistical Sciences*. New York: Wiley, pp. 442-457.

KOCH, G.G., LANDIS, J.R., FREEMAN, J.L., FREEMAN, D.H. & LEHNEN, R.G. (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics, 33,* 133-158.

MUTHEN, B. & CHRISTOFFERSSON, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46,* 407-419.

MUTHEN, B. (1978). Contributions to factor analysis of dichotomized variables. *Psychometrika, 43,* 551-560.

MUTHEN, B. (1983). Latent variable structural equation modelling with categorical data. *Journal of Econometrics, 22,* 43-65.

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

NEYMAN, J. (1949). Contributions to the theory of the χ^2 test. In J. Neyman, *Proceedings of the First Berkley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 239-273.

NORC (National Opinion Research Center) (1975). Codebook *for the Spring 1975 General Social Survey.* Chicago: National Opinion Research Center, University of Chicago.

PEARSON, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling. *Philos. Mag., Ser. 5, 50,* 157-75 (reproduced in *Karl Pearson's Early Statistical Papers.* London: Cambridge University Press, 1948).
SAS Institute, Inc. (2004). *SAS/STAT User's Guide, Version 9.* Cary, NC: SAS Institute, Inc
SEN, P.K. & SINGER, J.M. (1993). *Large Sample Methods in Statistics.* London: Chapman and Hall.
SORBOM, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229-239.
SORBOM, D. (1989). Model modification. *Psychometrika, 54,* 371-384.
STEIGER, J.H. & LIND, J.C. (1980). *Statistically based tests for the number of common factors*, paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
STEIGER, J.H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling, 5*, 411-419.
STROUD, T.W.F. (1971). On Obtaining Large-Sample Tests from Asymptotically Normal Estimators. *Annals of Mathematical Statistics, 42,* 1412-1424.
WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54,* 426-482.

Table 1 : Political views : ordinal data (reference : Haberman, 1978, p.85)

	Categories from Extremely liberal (1) to Extremely conservative (7)						
Political views	1	2	3	4	5	6	7
Observed count	46	179	196	559	232	150	35
Fitted count	40.5	164.5	214	559	214	164,5	40,5

Political views : (1) Extremely liberal, (2) Liberal, (3) Slightly liberal, (4) Moderate, middle of the road, (5) Slightly conservative, (6) Conservative, (7) Extremely conservative

Table 2a: overall fit indices

Jöreskog and Sörbom	Q_{JS} (3) = 7.1090	p = 0,0685
Pearson	Q_P (3) = 7.0781	p = 0.0694
Neyman	Q_N (3) = 7.1478	p = 0.0673

Table 2a: detailed assessment of fit

Parameters	π_1	π_2	π_3	π_4	π_5	π_6	π_7
Observed	0.0329	0.1281	0.1403	0.4001	0.1661	0.1074	0.0251
Fitted	0.0286	0.1174	0.1529	0.4021	0.1529	0.1174	0.0286
t-value	9.2266	20.7532	24.8584	30.7079	24.8584	20.7532	9.2266
MI	1.5064	2.5758	3.0501	0.0000	3.0501	2.5758	1.5064
Estimated change	0.0045	0.0111	-0.0122	0.0000	0.0137	-0.0097	-0.0034
p	0.2197	0.1085	0.0807	-	0.0807	0.1085	0.2197