Title
Testing the Assumptions Underlying Tetrachoric Correlations

Authors
Muthen, Bengt O.
Hofacker, Charlie

Publication Date
1988-08-08

Peer reviewed

PSYCHOMETRIKA—VOL. 53, NO. 4, 563–578
DECEMBER 1988

TESTING THE ASSUMPTIONS UNDERLYING TETRACHORIC CORRELATIONS

BENGT MUTHÉN

GRADUATE SCHOOL OF EDUCATION
UNIVERSITY OF CALIFORNIA, LOS ANGELES

CHARLES HOFACKER

COLLEGE OF BUSINESS AND MARKETING
FLORIDA STATE UNIVERSITY

A method is proposed for empirically testing the appropriateness of using tetrachoric correlations for a set of dichotomous variables. Trivariate marginal information is used to get a set of one-degree of freedom chi-square tests of the underlying normality. It is argued that such tests should preferrably preceed further modeling of tetrachorics, for example, modeling by factor analysis. The assumptions are tested in some real and simulated data.

Key words: normality, dichotomous variables, LISCOMP, factor analysis.

1. Introduction

Recent discussions of the tetrachoric correlation coefficient include the topics of improved efficiency and accuracy of computation (see e.g., Divgi, 1979; Kirk, 1973), the statistical properties of the estimate and its standard error (Brown & Benedetti, 1977), and the fitting of factor analysis models to tetrachorics (see e.g., Bock & Lieberman, 1970; Christoffersson, 1975; Muthen, 1984; Muthén & Christoffersson, 1981). The appropriateness of using the tetrachoric correlation coefficient, however, has not been given recent attention. The tetrachoric assumes bivariate normality for continuous latent response variables underlying a pair of dichotomous measures. Today's textbook treatment of coefficients of association for dichotomous variables offers rather vague guidance about the realism of this. Little conciliation seems to have been reached since the heated turn of the century debate between Pearson and Yule, where in arguing for considering only the observed variables, Yule remarked (1912):

> ... all those who have died of smallpox are equally dead: no one is more dead or less dead than another, and the dead are quite distinct from the survivors (pp. 611–612),

with the familiar response:

> ... if Mr. Yule's views are accepted, irreparable damage will be done to the growth of modern statistical theory (Pearson & Heron, 1913, p. 159).

This paper builds on the view that the appropriateness of tetrachorics need not be a philosophical matter but one of statistical inference. There is a hypothesis of underlying

Requests for reprints should be sent to Bengt Muthén, Graduate School of Education, University of California, Los Angeles, Los Angeles, CA 90024.

Presented at the Psychometric Society meeting in Santa Barbara, California, June 25–26, 1984. The research of the first author was supported by Grant No. SES-8312583 from the National Science Foundation.

0033-3123/88/1200-7090$00.75/0

normality given which sample estimates of population tetrachorics are obtained. As usual, we should seek to test the hypothesis by confronting it with data. The estimates should only be used if the hypothesis is not rejected.

One issue, which may have distracted from this approach, is that the estimation of tetrachoric correlations has always been performed using information from only bivariate distributions, that is, 2×2 tables. Here, the 3 independent probabilities describing the distribution are not restricted by the three parameters of the "tetrachoric model": The 2 latent response variable thresholds and the single tetrachoric correlation. The model is "just-identified" and cannot be tested since it fits perfectly. We may note that given this bivariate information, the usual parameter estimates are maximum-likelihood estimates (MLEs), since there is a one-to-one relationship between these parameters and the probabilities, which have the observed proportions as MLEs.

In the type of application we will study, a set of p dichotomous items is usually considered. Here, the bivariate normality specification may be extended to multivariate normality for p latent response variables. For simplicity we will still call this a tetrachoric model, referring to underlying normality, despite the fact that "tetra" refers to the 2×2 feature. The latent response variables have a population covariance matrix with unit diagonal elements and $p(p-1)/2$ population tetrachoric correlations to be estimated. In addition, the tetrachoric model has p population thresholds to be estimated. The p-variate tetrachoric model hence imposes $2^p - 1 - (p + p(p-1)/2)$ restrictions on the multinominal distribution. With multivariate normality the sample tetrachoric correlation matrix will be positive definite in large samples. In any given sample, sample fluctuations or deviations from underlying multivariate normality may prevent this.

In factor analysis contexts, the $p(p-1)/2$ population tetrachorics are further restricted. Generalized least-squares factor analysis of tetrachoric correlations was discussed in Muthén (1978). A large-sample chi-square test is obtained for testing the restrictions on the tetrachorics. It may be noted that the method of Muthén does not break down if the sample tetrachoric correlation matrix is not positive definite. In this paper, however, we are interested in testing the tetrachoric model itself, namely the multivariate normality assumption, without further restricting the tetrachorics in terms of a smaller set of parameters.

Optimally, the full p-variate information from all 2^p cells of the multiway table should be used to estimate the threshold and tetrachoric parameters and test the fit of the p-variate tetrachoric model. Straight-forward ML estimation would, however, involve the computation of the p-variate normal distribution function which is intractable. Even when efficient estimates are computable, reliable testing is prevented by sparsity of data, since already with $p = 10$ there are 1,024 cells. This is the case in the EM approach of Bock and Aitkin (1981), where ML estimation becomes tractable by considering the tetrachoric model *with* further restrictions on the correlations using a small number of uncorrelated factors. Instead of a model test against the unrestricted multinomial alternative, Bock and Aitkin consider chi-square difference tests of improvement in fit when adding factors. On the other hand, the traditional approach of estimating the tetrachoric model parameters by considering the $p(p-1)/2$ marginal, bivariate tables prevents a test of model fit due to just-identification.

In this paper we will propose a testing method that utilizes information from trivariate marginal tables, noting that in this case the tetrachoric model describes the 7 independent probabilities of each triplet in terms of 3 thresholds and 3 tetrachoric correlations. This yields one restriction to be tested against sample cell frequencies that are more likely to be large enough. While for a large value of p, this approach is obviously intractable due to a rapidly increasing number of "triplets", certain special cases of the

population tetrachoric correlation matrix make for relatively simple computations for small to moderate sized problems, say $p \leq 20$.

The inference for triplet testing of trivariate normality is straight-forward when there are three variables. Triplet testing for $p > 3$ variables is less clearcut. We note that such testing cannot lead to support for multivariate normality, and that this is not needed for the use of tetrachorics, only bivariate normality for each pair. Furthermore, rejection in a certain triplet does not necessitate that any of the three pairs involved lacks bivariate normality. However, nonrejection in a certain triplet lends support for bivariate normality for all three pairs involved. Given this, we note that for $p > 3$ variables the set of triplet tests provide an ad hoc, but practically useful decision tool for the use of tetrachorics. Consider the two outcomes (i) no rejection is obtained, and (ii) a certain variable is involved in all rejected triplets. In case (i), we would be using a conservative test procedure to get support for bivariate normality for all pairs involved. In case (ii), the certain variable may be excluded from the analysis of tetrachorics (see also the example in section 3.2).

The possibility of using trivariate information for testing was also considered by Vaswani (1950). (We are grateful to an anonymous reviewer for making us aware of this.) ML estimation of the parameters involved in a triplet was at that time deemed to be too cumbersome. Utilizing a simpler consistent estimator it was suggested that the Pearson chi-square formula could be used to get an indication of whether the sample would lead to an insignificant chi-square value, while it was recognized that significant values had to be deemed inconclusive due to nonefficient estimation of the cell frequencies.

Section 2 describes our proposed "triplet" approach, which uses numerical integration to carry out ML estimation by a simplified way of expressing the trivariate probabilities. In section 3 this approach is used to test the tetrachoric model with different data sets. Section 4 concludes.

2. Estimation and Testing in Triplets

Consider a $2 \times 2 \times 2$ table for the items y_i, y_j, y_k, with corresponding latent response variables y_i^*, y_j^*, y_k^*. Given the specification of p-variate normality, the three y^*'s are trivariate normal with population correlations ρ_{ij}, ρ_{ik}, ρ_{jk}, say. We will consider the special cases of a population tetrachoric correlation matrix for p variables for which each triplet has a "triad" obeying the restriction

$$0 < \frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} < 1. \tag{1}$$

In particular, this includes sets of items with positive correlations, as is often found among a set of items defining a scale, and sets of items where negative correlations can obtain a change of signs by turning the directions of items around. Although not necessary for the chi-square testing, this restriction allows a particularly simple representation of the trivariate probabilities, reducing the computational work of the testing considerably.

Under (1), we may describe the trivariate normal distribution of y_i^*, y_j^*, y_k^* by a just-identified, single-factor model,

$$y_s^* = \lambda_s \eta + \varepsilon_s; \qquad s = i, j, k. \tag{2}$$

where $V(\eta) = 1$, $V(\varepsilon_s) = 1 - \lambda_s^2$, such that the y^* variables have zero means, unit variances, and correlations

$$\rho_{ss'} = \lambda_s \lambda_{s'}; \qquad s \neq s'. \tag{3}$$

In (2), we also assume that the factor η and the residuals ε's are independent and (multivariate) normally distributed. The ε's are assumed to be uncorrelated.

For all trivariate standard normal y^* distributions obeying (1), we may let

$$\left(\frac{\rho_{st}\,\rho_{su}}{\rho_{tu}}\right)^{1/2} = \lambda_s. \tag{4}$$

Then the correlations for the y^*'s may be written as $\rho_{st} = \lambda_s \lambda_t$, which is the correlation structure in (2). This shows that all trivariate standard normal y^* distributions for which (1) holds have a one-factor representation as in (2). There is a one-to-one relationship between the 3 λ's and the 3 ρ's. Hence, this single-factor model is just-identified in terms of the 3 correlations and can be used to test the normality marginally in each triplet.

Let τ_s be the threshold parameter for y_s^* and

$$\Pr(y_s = 1 \mid \eta) = \Pr(y_s^* > \tau_s \mid \eta) = \Phi[(1 - \lambda_s^2)^{-1/2}(-\tau_s + \lambda_s \eta)], \tag{5}$$

where $\Phi(\cdot)$ is the standard normal distribution function. For convenience we may use the equivalent "probit" parameterization of the right-hand-side of (5),

$$\Pr(y_s = 1 \mid \eta) = \Phi(\alpha_s + \beta_s \eta), \tag{6}$$

with

$$\alpha_s = -\tau_s(1 - \lambda_s^2)^{-1/2}, \tag{7}$$

$$\beta_s = \lambda_s(1 - \lambda_s^2)^{-1/2}. \tag{8}$$

Further, let the probability of an observation falling in cell c be denoted π_c ; $c = 1, 2, \ldots, 8$. The computational advantage of our single-factor representation is obtained as in Bock and Lieberman (1970). We may consider the joint trivariate normal density of the y^*'s as the product of the density of y^*'s given η, multiplied by the marginal density of η. The former density simplifies due to conditional independence, and we obtain

$$\pi_c = \int_{-\infty}^{\infty} \left\{ \prod_{s=1}^{3} [\Phi(\alpha_s + \beta_s \eta)]^{y_{cs}} [1 - \Phi(\alpha_s + \beta_s \eta)]^{1 - y_{cs}} \right\} \phi(\eta)\, d\eta, \tag{9}$$

where y_{cs} is 0/1 indicator variables for cell c and variable s. It can be seen from (9) that the trivariate probabilities can be expressed by integration over a single variable η, which may be approximated by Gauss-Hermite quadrature as in Bock and Lieberman (1970). A 40 point quadrature in used in the applications below, but this may be unnecessarily ambitious.

For a given triplet, MLEs of the 6 parameters of the tetrachoric model may be obtained by minimizing

$$F = -2(F_0 - F_1), \tag{10}$$

where

$$F_0 = \sum_{c=1}^{8} f_c \log \pi_c, \tag{11}$$

and

$$F_1 = \sum_{c=1}^{8} f_c \log\left(\frac{f_c}{N}\right). \tag{12}$$

Here, f_c is the sample frequency in cell c and N is the sample size. In (10), F is scaled such that when calculated at the ML values, it gives the one degree of freedom likelihood-ratio

chi-square. In (12), F_1 represents the negative of the log likelihood calculated for the alternative, unrestricted model (see also Bock & Lieberman, 1970).

To further improve the speed of the calculations, we may use the traditional sample thresholds and sample tetrachorics to give initial estimates (starting values) for the α's and β's via (4), (7), and (8). When such initial estimates obtain an F value less than the critical chi-square value at a certain significance level, that triplet needs no further iteration towards MLEs; the normality of the tetrachoric model cannot be rejected. In examples studied, this event occurred most frequently. The calculations may be carried out with the LISCOMP computer program (Muthén, 1987).

The question of choice of significance level is a difficult one. The set of one-degree-of-freedom triplet tests are not independent and we may observe mass significance effects. Also, the power of the triplet tests needs further research. In practice, we may in fact want to use the triplet testing less as a rigorous inferential procedure, but rather as a tool for suggesting "abberrant" items that are frequently involved in triplets with rejection of the model.

3. Examples

3.1 Eye Color

The first example concerns eye color (light-eyed vs. not light-eyed individuals) observed in three generations. The data come from Galton's *Natural Inheritance*, as reported in Kendall and Stuart (1979, p. 572). The sample size is 5,008. The question of interest was whether the resemblance between grandparent and grandchild is mediated by the middle generation. We may hypothesize the underlying trivariate normality of the tetrachoric model. A further modeling step might be to hypothesize that the tetrachoric correlation between grandparent and grandchild is the product of the correlations between grandparent and parent and parent and grandchild. As a first step, however, we may test the appropriateness of the tetrachoric model itself. The data are given in Table 1. The usual sample tetrachorics are all positive and yield a positive definite correlation matrix.

While the 1% critical value is 6.635 with one degree of freedom, the likelihood ratio chi-square for the tetrachoric model was 11.407 (the Pearson chi-square was 11.500). Hence, the tetrachoric model is rejected and it would be inappropriate to proceed to interpretation and further modeling of the correlations. Since the sample size is very large we may consider whether the rejection is due to rather small residuals. To this aim the estimated frequencies under the tetrachoric model are given in Table 1.

3.2 Abortion Attitudes

As a second example, consider a set of six attitudinal items, all related to abortion. The wording is given in Table 2. The data were obtained from the National Opinion Research Center's General Social Surveys. A sample size of 3,921 was obtained by combining the responses for 1973, 1974, 1975. The response pattern frequencies are given in Table 3.

It is relevant to hypothesize one, or possibly two, underlying dimension(s) of abortion attitude(s). We note that there are two "hard" (medical) reasons Defect, Health, three "soft" (social) reasons Nomore, Poor, Single, while the Rape item is not easily classified.

Considering the three soft items alone, the one-degree of freedom tetrachoric model test obtains a likelihood ratio chi-square value of 2.714 (the Pearson chi-square was 2.711). This value corresponds to a probability level of about .10, so the tetrachoric model cannot be rejected despite the fairly large sample. The estimated frequencies are given in Table 4.

Table 1

Three Generations of Eye Color

(N = 5,008)

		GRANDPARENTS	
		Light	Dark
PARENTS - Light			
	Child - Light	1928 1908.1 .21	596 619.8 .91
	Child - Dark	303 326.2 1.65	225 200.9 2.89
PARENTS - Dark			
	Child - Light	552 576.9 1.02	508 480.9 1.53
	Child - Dark	395 368.3 1.94	501 527.0 1.28

Entries are: Observed frequency

Estimated frequency

Cell contribution to likelihood ratio chi-square.

Considering all six items, the usual sample tetrachoric correlation matrix is positive definite with positive elements. There are

$$\binom{6}{3} = 20$$

triplets to be tested. The testing of these triplets is displayed in Table 5. For 12 out of 20 triplets no iterations would have been needed since the tetrachoric model could not be rejected using the (inefficient) starting value estimates derived from the usual sample tetrachorics. Four of the 20 triplets are rejected at the 5% level. We note that the Rape item is involved in all four of these rejected triplets (each item appears in 10 triplets). We

Table 2

Abortion data: Wording of Six Items

	Should it be possible for a pregnant woman to obtain a legal abortion if ...
DEFECT	(1) There is a strong chance of a defect?
NOMORE	(2) She is married and wants no more children?
HEALTH	(3) Her health is endangered by the pregancy?
POOR	(4) She cannot afford more children due to low income?
RAPE	(5) She was raped?
SINGLE	(6) She is not married?

conclude that we have no evidence to reject multivariate normality for the five items excluding Rape, while the Rape item may warrant particular considerations.

The triplet with the highest chi-square value is Nomore, Poor, Rape (chi-square of 11.964). In Table 6 the observed and estimated frequencies are given for this triplet. We note that the largest contributions to the misfit come from the category "No" on RAPE. Similar results are found for the other rejected Rape triplets. Hence, No respondents for Rape may be viewed as a different sub-group requiring special modeling.

Another interesting question concerns the estimation of the tetrachoric correlations. How do the usual, pair-based correlations compare to the triplet based ones in well-fitting and mis-fitting triplets? Inspecting results for the 21 abortion item correlations we found little difference between the pair- and triplet-based correlations. Using bivariate information seems sufficient. Interestingly, the estimates are close also in misfitting triplets.

The final aspect of the abortion example concerns the inappropriateness of using tetrachorics involving the Rape item when further modeling the tetrachorics in a factor analysis. Using the methodology of Muthén (1978) we performed exploratory and confirmatory factor analyses of the abortion items with and without the Rape item. While the results do not differ much between the five and six item solutions, the five item two-factor solutions do fit somewhat better than the six item two-factor solutions. Excluding the abberant Rape item seems to "polish" the results.

3.3 Generated Data

In the final example we will consider the triplet testing approach in some generated data with 12 variables, so that there are 4,096 possible response patterns. Hence, even with very large samples, many empty cells are to be expected, prohibiting a test against the unrestricted multinomial alternative. In line with (2), the population model is as follows. A single normal factor η and 12 uncorrelated normal residuals ε generate 12

Table 3

Abortion data: Frequency of Reponse Patterns

(N = 3,921)

Pattern	Frequency	Pattern	Frequency	Pattern	Frequency	Pattern	Frequency
000000	229	010000	0	100000	13	110000	0
000001	0	010001	0	100001	0	110001	0
000010	33	010010	1	100010	20	110010	2
000011	3	010011	1	100011	0	110011	1
000100	0	010100	1	100100	4	110100	0
000101	0	010101	0	100101	1	110101	0
000110	2	010110	2	100110	2	110110	0
000111	1	010111	5	100111	1	110111	3
001000	130	011000	4	101000	186	111000	12
001001	0	011001	2	101001	4	111001	2
001010	128	011010	5	101010	744	111010	55
001011	5	011011	3	101011	103	111011	48
001100	4	011100	1	101100	12	111100	6
001101	1	011101	0	101101	4	111101	2
001110	15	011110	4	101110	174	111110	123
001111	10	011111	12	101111	152	111111	1645

2, 4, 6: 1483 for (0, 0, 0)

multivariate normal latent response variables y^*. The factor and the y^*'s have unit population variances and zero population means. The loadings λ are all 0.7 which would result in population tetrachoric correlations among the y^*'s of 0.49. All y's give $y = 1$ as opposed to $y = 0$ observations when exceeding a population threshold of zero, except y^*_{10} and y^*_{12} for which the population values are -1.0 and 1.0, respectively. We will distort this model so that the specification of underlying multivariate normal y^*'s no longer holds for all y variables. This can be done by violating the conditional independence assumptions for y responses given η, in a way that only affects a few select cells in the multiway table of response patterns. Specifically, if $y_{10} = 1$ and $y_{11} = 0$ as predicted by the single factor model (i.e., $y^*_{10} > -1.0$ and $y^*_{11} \leq 0.0$), we let $y_{11} = 1$ with a 50% chance (randomly determined). Once this is determined, we let a response of $y_{10} = 1$ and $y_{11} = 1$ result in changing a predicted $y_{12} = 0$ response ($y^*_{12} \leq 1.0$) to $y_{12} = 1$ with a 75% chance (randomly determined). This can illustrate items responded to in order 1–12 with the last three items exhibiting a response consistency effect in attitudinal studies or the learning of a problem-solving technique in achievement testing, in both cases due to an inappropriate choice of the last three items. For example, a person may not have a strong enough attitudinal propensity y^*_{12} to warrant the "agree" response $y_{12} = 1$ (i.e., $y^*_{12} \leq 1.0$), but the wording of the last three items is such that he/she feels a need to be consistent in agreement.

While the first ten items still obey the normality assumption of the tetrachoric model,

Table 4

Abortion data: Observed and Estimated Frequencies

For NOMORE, POOR, SINGLE

		SINGLE	
		No	Yes
NOMORE - No			
	POOR - No	1483	115
		1490.1	108.1
		.03	.44
	POOR - Yes	213	170
		206.2	176.5
		.22	.24
NOMORE - Yes			
	POOR - No	79	57
		71.9	63.9
		.70	.75
	POOR - Yes	137	1667
		143.7	1660.5
		.31	.03

Entries are

Observed frequency

Estimated frequency

Cell contribution to likelihood ratio chi-square.

adding the last two items will violate multivariate normality of all y^*'s. The violation of normality can for instance be seen in the trivariate distribution of the last three items. Using (i, j, k) to denote the pattern of 0's and 1's for items 10, 11, 12, adding the conditional independence distortion shifts response probability units from (1, 0, 0) to (1, 1, 0), from (1, 0, 1) to (1, 1, 1), and from (1, 1, 0) to (1, 1, 1), so that cells for which $y_{10} = 0$ are never affected. The trivariate normal model could be perfectly fitted to the four modified cells with $y_{10} = 1$, but that would clearly involve changes in parameter values, such that the four unaffected cells with $y_{10} = 0$ no longer fit perfectly. This is shown in Table 7, where expected frequencies under the true model and predicted frequencies under the tetrachoric model are given for a sample of 2,000. The tetrachoric model will give varying degrees of misfit for all triplets involving item(s) 11 and/or 12. We note that in (unsuccess-

Table 5

Tests of Triplets for Abortion Data

(N = 3,921)

Triplet	LR Chi-Square	Triplet	LR Chi-Square
DEFECT,NOMORE,HEALTH	0.085	NOMORE,HEALTH,POOR	0.003
DEFECT,NOMORE,POOR	3.237	NOMORE,HEALTH,RAPE	7.326**
DEFECT,NOMORE,RAPE	0.016	NOMORE,HEALTH,SINGLE	0.292
DEFECT,NOMORE,SINGLE	0.503	NOMORE,POOR,RAPE	11.964**
DEFECT,HEALTH,POOR	1.544	NOMORE,POOR,SINGLE	2.639
DEFECT,HEALTH,RAPE	5.444*	NOMORE,RAPE,SINGLE	5.748*
DEFECT,HEALTH,SINGLE	0.086	HEALTH,POOR,RAPE	0.098
DEFECT,POOR,RAPE	2.539	HEALTH,POOR,SINGLE	0.035
DEFECT,POOR,SINGLE	0.041	HEALTH,RAPE,SINGLE	2.778
DEFECT,RAPE,SINGLE	1.147	POOR,RAPE,SINGLE	1.953

*Significant at 5% level (critical value = 3.841).

**Significant at 1% level (critical value = 6.635).

fully) attempting to find a well-fitting tetrachoric model to the 10, 11, 12 triplet, the three y^* correlations of 0.49 between items 10, 11, 12 will clearly be overestimated. As we will see, this leads to a distortion of the single factor correlational structure for the 12 items.

Ten data sets were generated according to the above model, using a sample size of 2,000 in each. Let us first consider exploratory factor analysis (Muthén, 1978) of the usual tetrachoric correlation matrix for all twelve variables. In Table 8, the average chi-square values are given for the tests of one and two factors. We see that the single factor model is clearly rejected in favor of a two-factor model, using either the absolute values or a chi-square difference test of adding a second factor. Hence, we would incorrectly use tetrachorics to conclude that two content factors underlie these variables, whereas in reality one of the factors is a methodological artifact. The average PROMAX rotated factor loadings are also given. The last three items form a factor of their own, whereas factor one seems to represent the real content factor. The loadings of the first nine items on the first factor are reasonably close to the correct value of 0.7, while the last three items' loadings on the first factor do not come close to this value. We note that although Item 10 itself does not violate the normality assumption, its loading on the first factor is only about half of the correct value. Item 12 exhibits a minor Heywood problem.

Table 6

Abortion Data: Observed and Estimated

Frequencies for NOMORE, RAPE, POOR

	RAPE	
	No	Yes
NOMORE - No		
POOR - No	562	1036
	569.2	1027
	.10	.06
POOR - Yes	26	357
	17.2	365.7
	4.38	.21
NOMORE - Yes		
POOR - No	20	116
	13.7	121.4
	2.90	.24
POOR - Yes	10	1794
	18.7	1786.7
	4.05	.03

Entries are

Observed frequency

Estimated frequency

Cell contribution to likelihood ratio chi-square.

Instead we may apply our triplet testing approach as a preliminary step, preceding the factor analysis. The correct decision would be to discard items 11 and 12 due to violations of the normality specification. With twelve variables there are 220 triplets to be tested. In the ten replications there was an average of 29.4 triplets involving rejection of trivariate normality at the 5% level. An average of 6.6 of these were incorrect in the sense of not involving either Item 11 or 12. The number of rejected triplets at the 5% level that each variable was involved in are as follows in order 1–12: 5.3, 4.0, 5.2, 4.6., 4.8, 5.4, 4.0, 4.5, 4.6, 12.5, 14.9, 17.3. Hence, the last two items are clearly singled out, with item 10 coming a close third since its presence together with either item 11 or 12 enhances the nonnormality. At the 5% level, the two variables most frequently involved in rejected triplets were 11 and 12 in eight of the ten cases, and 10 and 12 in two of the ten cases. The number of rejected triplets at the 1% level that Variables 10, 11, and 12 were involved in

Table 7

Generated Data: Predicted Frequencies Under the True Model and the Best Fitting Tetrachoric Model (N = 2,000) for the Triplet 10, 11, 12

Response Pattern	000	001	010	011	100	101	110	111
True Model	248.872	4.035	60.368	4.035	343.362	30.184	257.522	1051.621
Best Fitting Tetrachoric Model	259.871	1.571	46.653	10.595	329.792	35.808	278.493	1037.217

Likelihood ratio/Pearson χ^2 with 1 degree of freedom for the tetrachoric model: 15.477/15.647.

Estimated correlations under the tetrachoric model: 0.760 (10, 11), 0.832 (10, 12), 0.910 (11, 12).

are: 7.0, 10.3, 12.0. At the 1% level, the two variables most frequently involved in rejected triplets were 11 and 12 in nine out of the ten cases, and 10 and 12 in one of the ten cases. We conclude that the triplet testing approach is quite successful in singling out the two "nonnormal items" 11 and 12. We can then proceed with a factor analysis of tetrachorics for the first ten items, whereupon correct results will be obtained.

4. Conclusion

The issue of using tetrachoric correlations has long been surrounded by a certain amount of mystique and a separation of researchers into two camps, believers and nonbelievers in underlying continuous normal variables. Typical discussions center around the "existence" versus "nonexistence" of such variables in general, their nonexistence for dichotomous variables called "true" dichotomies (sex is often mentioned as such a variable), the purportedly strong assumption of normality, and the possibility of a nonpositive definite correlation matrix.

Our view is that the use of tetrachorics implies the hypothesizing of a model, a model which optimally should pass testing before its parameters' estimates (the sample tetrachorics) are interpreted and further used. In this sense, tetrachoric correlations are a bit different from ordinary Pearson-product-moment correlations. The latter can be a priori rejected for use on noninterval or nonratio-scaled variables (such as in the case of phi coefficients for dichotomous variables), but cannot be rejected in data due to nonnorma-

Table 8

Generated Data: Factor Analysis Solution

as an Average of Ten Replications*

χ^2 - value for 1 factor: 378.622 (d.f. = 54)

χ^2 - value for 2 factor: 43.151 (d.f. = 43)

Variable	Factor 1	Factor 2
1	.816	.001
2	.713	-.004
3	.716	-.006
4	.718	-.007
5	.717	-.013
6	.703	.010
7	.712	.007
8	.699	.017
9	.730	-.013
10	.364	.596
11	.204	.736
12	-.081	1.131
	Factor correlation	
	0.570	

*PROMAX rotations

lity of interval scaled variables, since the use of these correlations do not imply any statements about higher order moments (such as zero third-order moments of y^*'s in triplets). This observation points to the fact that one could try to generalize tetrachoric correlations to "latent correlations" for underlying latent response variables that are interval scaled but not necessarily multivariate normal.

The underlying latent response variables y^* should be viewed as convenient statistical constructs, whose existence need not be addressed. An equivalent representation is given without such variables in the conditional probability, probit, Equation (6). As in bioassay, where the "factors" η are observed, the latent response variable concept is reasonable as long as the model of (6) is reasonable. For instance, the covariation of the variable sex can very well be described by tetrachoric correlations as long as it is relevant to view the probability of observation of a certain gender to be an s-shaped function of some "factors" η that are normal. If the s-shaped functions are normal distribution functions and conditional independence holds, the data can then be described by multivariate normal y^*'s. With a large enough sample, the usual tetrachoric correlation matrix will be positive definite. But given such an hypothesis, confronting it with data as was done in this paper, is important in order to establish its plausibility. We expect testing of underlying normality to yield different results in different applications. If the tetrachoric model cannot be rejected, we certainly have a very parsimonious and convenient description of our data. If normality is rejected for some variables, investigating the reason why may teach us something new about the data.

The alternative hypothesis of non-normal latent response variables may be correct in several different instances and we think that our testing procedure has a power to reject the null hypothesis that may vary greatly over these instances. We think that nonnormality commonly comes about as in the generated data example, due to direct dependencies among items over and above factor influence (note that a second "methods" factor would not be able to perfectly account for the data since it would also affect the $y_{10} = 0$ cells). The abortion example is perhaps of a related kind. Here one may expect reasonably good power of rejecting normality. Other violations may come about due to nonnormal residuals, that is, items having conditional probability curves (given the factors) that do not follow a normal distribution function, for example, having a nonzero lower asymptote (as may be the case with guessing in achievement testing), or a nonmonotonic function. Also, the factors may themselves be nonnormal. These instances lead to cell frequencies which can not be perfectly represented by multivariate normal latent response variables. However, it may be very hard to choose among these alternative hypotheses and between them and the null hypothesis, since they may give rise to relatively small differences in cell frequencies, particulary when considering trivariate tables. Although difficult, it would be of interest to further study the power of testing the normality assumptions and also the misestimation of underlying correlations when the hypothesis of normality is incorrectly retained.

References

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*, 179–197.

Brown, M. B., & Benedetti, J. K. (1977). On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika, 42*, 347–355.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5–32.

Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika, 44*, 169–172.

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics* (Vol. 2). New York: McMillan.

Kirk, D. B. (1973). On the numerical approximation of the bivariate normal (tetrachoric) correlation coefficient. *Psychometrika, 38*, 259–268.

Muthén, B. (1978). Contributions to factor analysis of dichotomous dependent variables. *Psychometrika, 43*, 551–560.

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

Muthén, B. (1987). *LISCOMP. Analysis of linear structural equations with a comprehensive measurement model. User's guide.* Mooresville, IN: Scientific Software.

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 407–419.

Pearson, K., & Heron, D. (1913). On theories of association. *Biometrika, 9*, 159–315.

Vaswani, S. (1950). Assumptions underlying the use of the tetrachoric correlation coefficient. *Sankhya, 10*, 269–276.

Yule, G. U. (1912), On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society, 75*, 579–652.

Manuscript received 5/2/85
Final version received 9/3/87