PSYCHOMETRIKA—VOL. 44, NO. 4,
DECEMBER, 1979

MAXIMUM LIKELIHOOD ESTIMATION OF THE POLYCHORIC CORRELATION COEFFICIENT

ULF OLSSON

UNIVERSITY OF UPPSALA

The polychoric correlation is discussed as a generalization of the tetrachoric correlation coefficient to more than two classes. Two estimation methods are discussed: Maximum likelihood estimation, and what may be called "two-step maximum likelihood" estimation. For the latter method, the thresholds are estimated in the first step. For both methods, asymptotic covariance matrices for estimates are derived, and the methods are illustrated and compared with artificial and real data.

Key words: ordinal data, polychoric correlation.

1. *Introduction*

Especially in the behavioral sciences, data are often recorded as ordinal variables with only a few scale steps. Examples of such variables are attitude items, rating scales, Likert items and the like. Typical cases are when a subject is asked to report some attitude on scales like

(a)	approve	don't know		disapprove	or
(b)	approve strongly	approve	don't know	disapprove	disapprove strongly.

When analyzing this kind of data, a common approach is to assign integer values to each category [for example 1, 2 and 3 in Example (a) and 1 through 5 in Example (b)] and proceed in the analysis as if the data had been measured on an interval scale with desired distributional properties. To quote Wainer and Thissen [1976], in such cases "a quick and easy approach is to assume normality and be on your merry way".

Although many statistical methods seem to be fairly robust against this kind of deviation from the distributional assumptions—at least in not-so-extreme cases—there are instances when this approach may lead to erroneous results.

For example, Olsson [1979] showed that application of factor analysis to discrete data may lead to incorrect conclusions regarding the number of factors, and to biased estimates of the factor loadings, especially when the distributions of the observed variables are skewed in opposite directions. This is mainly due to biased estimates of the correlations. Thus, there seems to be some need for correlation estimates which are more viable when the observed data are ordinal with only a few scale steps.

This paper was read at the 1978 European Meeting on Psychometrics and Mathematical Psychology in Uppsala, Sweden, June 1978.

Research reported in this paper has been supported by the Bank of Sweden Tercentenary Foundation under project "Structural Equation Models in the Social Sciences", project director Karl G. Jöreskog.

The author's present affiliation and the address for reprint requests is: Ulf Olsson, The Swedish University of Agricultural Sciences, Department of Economics and Statistics, S-750 07 Uppsala 7, SWEDEN.

0033-3123/79/1200-2795$00.75/0

In this paper, we shall discuss the maximum likelihood estimation of correlation coefficients from ordinal data. In short, our problem may be summarized as follows: We observe two ordinal variables, x and y. These are classified into s and r categories, respectively. A cross-tabulation of x by y gives the observed frequencies as denoted in Table 1. We further assume that underlying x and y there are some latent variables, ξ and η, which are bivariate normally distributed. The relation between x and ξ may be written

$$
\begin{aligned}
x &= 1 \quad \text{if } \xi < a_1 \\
x &= 2 \quad \text{if } a_1 \leqq \xi < a_2 \\
x &= 3 \quad \text{if } a_2 \leqq \xi < a_3 \\
&\vdots \\
x &= s \quad \text{if } a_{s-1} \leqq \xi,
\end{aligned}
$$

TABLE 1

The General Form of the Raw Data: a Cross-tabulation of x by y.

x \ y		b_1	b_2		b_{r-1}
	1	2	3	...	r
1	n_{11}	n_{12}	n_{13}		n_{1r}
a_1 2	n_{21}	n_{22}			n_{2r}
a_2 3	n_{31}				n_{3r}
⋮					
a_{s-1} s	n_{s1}				n_{sr}

a_i and b_j denote thresholds, where

$a_0 = b_0 = -\infty$ and $a_s = b_r = +\infty$

and correspondingly for y. The parameters a_i are usually termed thresholds. The problem is to estimate the correlation ρ between ξ and η, given data in the form of Table 1. This is a generalization of the arguments behind the familiar tetrachoric correlation coefficient to polytomous variables.

We shall discuss estimation of ρ by means of the maximum likelihood method. Even given the method of estimation, the problem may be solved in at least two different ways. One way is to estimate ρ and the thresholds simultaneously. Alternatively, the thresholds are first estimated as the inverse of the normal distribution function, evaluated at the cumulative marginal proportions of the table, and the maximum likelihood estimate of ρ is then computed given the thresholds. This may be called a "two-step maximum likelihood" procedure. The latter approach has the advantage of greater ease in the numerical computations, although the former is formally more correct. One of the points discussed in the paper is the differences in results from the two methods in some reasonable cases.

In Section 2 we summarize the results of some earlier writers in the area. In Section 3 the likelihood equations are derived, and in Section 4 we derive asymptotic standard errors of the estimates. Section 5 contains a numerical computer study, based on both true data and on Monte Carlo simulations. There we also analyze some real data. Finally our results are summarized in Section 6.

2. *Earlier Research*

2 × 2 Tables: The Tetrachoric Correlation

The tetrachoric correlation coefficient was suggested by Pearson [1901] as a measure of bivariate normal correlation, when only data from a 2 × 2 cross-classification of the data are available. Pearson also supplied formulae for the standard error of the correlation.

The tetrachoric correlation may be obtained by solving ρ from

$$\Phi(h, k; \rho) = [2\pi(1-\rho^2)^{1/2}]^{-1} \int_{-\infty}^{h}\int_{-\infty}^{k} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right] dx\, dy. \tag{1}$$

The method suggested by Pearson [1901] was to expand the right-hand side of (1) into a series expansion in ρ, the so called tetrachoric series. This leads to a polynomial equation in ρ, where the degree of the polynomial depends on the number of terms in the series expansion.

Hamdan [1970] showed that the tetrachoric r is equivalent to the maximum likelihood estimate of ρ from a 2 by 2 table. This is not unexpected, since the estimation problem involves three parameters (ρ, h and k where h and k are the thresholds), and the data supplies three independent statistics in the table, for example $n_{1.}$, $n_{.1}$ and n_{11}. Thus, the estimation problem is just identified. The advantage of the maximum likelihood approach is that asymptotic standard errors are easily obtained through the inverse of the second order derivative of the log likelihood.

The tetrachoric correlation is a biased estimate of the corresponding true correlation. Brown and Bendetti [1977] showed that the expected cell frequencies are critical for the bias. If no expected frequencies are less than 5, the bias is negligible. This applies also for the bias of the standard error if $\rho = 0$. If $\rho \neq 0$, Brown and Bendetti recommended the standard error based on ML-estimation, but even this converges fairly slowly towards true $\sigma_{\hat\rho}$ with increasing n. It seems that expected frequencies of at least 10 are necessary in order to obtain reasonable estimates of the standard error.

Computer routines for the numerical solution of (1) are now available as packed programs at many computer installations. Froemel [1971] made a comparison of some routines. The best routine to date is probably that by Kirk [1973], who attacks the problem by Gaussian quadrature and Newton-Raphson iteration.

$r \times s$ Tables: The Polychoric Correlation

Several coefficients have been suggested as measures of association in contingency tables; we shall here confine ourselves to those which, like the tetrachoric correlation, use an assumption of an underlying bivariate normal distribution, for which the correlation is to be estimated.

Pearson [Note 1] suggested that the mean square contingency, which is based on the usual χ^2, could be used to estimate ρ from a polychoric table. Ritchie-Scott [1918] developed a coefficient based on a weighted mean of all possible tetrachoric correlations which may be computed from the table. Pearson and Pearson [1922], influenced by the result of Ritchie-Scott, suggested simpler methods for larger tables.

Lancaster and Hamdan [1964] showed that the mean square contingency does not work well when the classification is crude. Instead, using the theory of orthonormal functions, they generalized the tetrachoric series expansion to the polychoric case. Pearson's corrected ϕ^2 was equated to a series which they called the polychoric series and the corresponding polynomial equation was solved for ρ. The method does not produce standard errors of the estimates; confidence intervals are instead obtained by inserting 2.5 and 97.5 percent limits for χ^2 in ϕ^2 and solving the equations. Hamdan [1968] showed that Pearson's [1901] tetrachoric series is a special case of the polychoric one. Hamdan [1971] gave computing formulae for smaller tables, and Martinson and Hamdan [1975] designed a computer program for the method. Martinson and Hamdan [1971] used a Maximum Likelihood approach to estimate ρ given the thresholds, and did also present formulae for the asymptotic standard error. The latter formulae, however, do not take into account that the thresholds are estimated from the data, and that they are thus also subject to random errors.

In all the above methods, the thresholds are regarded as fixed, although already Pearson and Pearson [1922] were aware that this might not be ideal. For the tetrachoric case this does not matter, since, as noted above, the estimation problem is just identified. For 3 by 3 tables, for example, we have 5 parameters (ρ and two thresholds for each variable) but 8 independent proportions, i.e. the problem is over-identified. Tallis [1962] attacked the problem via maximum likelihood estimation of ρ and thresholds simultaneously in 2×2 and 3×3 tables. For 2×2 tables the results should be the tetrachoric correlation [cf. Hamdan, 1970], but for 3×3 tables the results may differ from Lancaster and Hamdan's [1964] polychoric correlation.

3. Derivation of the Likelihood Equations

The data consists of an array of observed frequencies $n_{ij}: i = 1, 2, \cdots, s; j = 1, 2, \cdots, r$, as given in Table 1.

If we denote by π_{ij} the probability that an observation falls into cell (i, j), the likelihood of the sample is

$$L = C \cdot \prod_{i}^{s} \prod_{j}^{r} \pi_{ij}^{n_{ij}} \tag{2}$$

where C is a constant. Taking logarithms,

$$l = \ln L = \ln C + \sum_{i=1}^{s} \sum_{j=1}^{r} n_{ij} \ln \pi_{ij}. \tag{3}$$

The thresholds for x are denoted by a_i, $i = 0 \cdots s$ and the thresholds for y by b_j, $j = 0 \cdots r$, where $a_0 = b_0 = -\infty$ and $a_s = b_r = +\infty$. It follows that

$$\pi_{ij} = \Phi_2(a_i, b_j) - \Phi_2(a_{i-1}, b_j) - \Phi_2(a_i, b_{j-1}) + \Phi_2(a_{i-1}, b_{j-1}) \tag{4}$$

where Φ_2 is the bivariate normal distribution function with correlation ρ.

Case 1: All Parameters Are Estimated Simultaneously

The parameters to be estimated are ρ, $a_1 \cdots a_{s-1}$, $b_1 \cdots b_{r-1}$. Partial differentiation of l with respect to these parameters yields

$$\frac{\partial l}{\partial \rho} = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{n_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \rho} \tag{5}$$

$$\frac{\partial l}{\partial a_k} = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{n_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial a_k} \tag{6}$$

$$\frac{\partial l}{\partial b_m} = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{n_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial b_m}. \tag{7}$$

Since $\partial \Phi_2(u, v)/\partial \rho = \phi_2(u, v)$ where ϕ_2 is the bivariate normal density function, (see Tallis, 1962, p 344; see also Johnson & Kotz, 1972, p 44), it follows that

$$\frac{\partial \pi_{ij}}{\partial \rho} = \phi_2(a_i, b_j) - \phi_2(a_{i-1}, b_j) - \phi_2(a_i, b_{j-1}) + \phi_2(a_{i-1}, b_{j-1}). \tag{8}$$

Therefore, (5) may be written

$$\frac{\partial l}{\partial \rho} = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{n_{ij}}{\pi_{ij}} \{\phi_2(a_i, b_j) - \phi_2(a_{i-1}, b_j) - \phi_2(a_i, b_{j-1}) + \phi_2(a_{i-1}, b_{j-1})\}. \tag{9}$$

In (6), it is evident that

$$\frac{\partial \pi_{ij}}{\partial a_k} = \begin{cases} 0 \text{ if } i \neq k \text{ and } i \neq k+1 \text{, i.e. if the formula for } \pi_{ij} \text{ does not contain } a_k \\ \dfrac{\partial \Phi_2(a_k, b_j)}{\partial a_k} - \dfrac{\partial \Phi_2(a_k, b_{j-1})}{\partial a_k} & \text{if } k = i \\ -\dfrac{\partial \Phi_2(a_k, b_j)}{\partial a_k} + \dfrac{\partial \Phi_2(a_k, b_{j-1})}{\partial a_k} & \text{if } k = i-1. \end{cases} \tag{10}$$

Thus, in (6) it suffices to let i go from k to $k+1$. Therefore, (6) may be written

$$(11)\qquad \frac{\partial l}{\partial a_k} = \sum_{j=1}^{r} \frac{n_{kj}}{\pi_{kj}} \left\{ \frac{\partial \Phi_2(a_k, b_j)}{\partial a_k} - \frac{\partial \Phi_2(a_k, b_{j-1})}{\partial a_k} \right\} + \frac{n_{k+1,j}}{\pi_{k+1,j}} \left\{ -\frac{\partial \Phi_2(a_k, b_j)}{\partial a_k} + \frac{\partial \Phi_2(a_k, b_{j-1})}{\partial a_k} \right\} = \sum_{j=1}^{r} \left(\frac{n_{kj}}{\pi_{kj}} - \frac{n_{k+1,j}}{\pi_{k+1,j}} \right) \left\{ \frac{\partial \Phi_2(a_k, b_j)}{\partial a_k} - \frac{\partial \Phi_2(a_k, b_{j-1})}{\partial a_k} \right\}.$$

Also, if we let ϕ_1 and Φ_1 denote univariate normal density and distribution function, respectively,

$$(12)\qquad \frac{\partial \Phi_2(u, v)}{\partial u} = \phi_1(u) \cdot \Phi_1 \left\{ \frac{(v - \rho u)}{(1 - \rho^2)^{1/2}} \right\}$$

[Tallis, 1962, p 346]. Equation (6) may now be written as

$$(13)\qquad \frac{\partial l}{\partial a_k} = \sum_{j=1}^{r} \left(\frac{n_{kj}}{\pi_{kj}} - \frac{n_{k+1,j}}{\pi_{k+1,j}} \right) \cdot \phi_1(a_k) \cdot \left[\Phi_1 \left\{ \frac{(b_j - \rho a_k)}{(1 - \rho^2)^{1/2}} \right\} - \Phi_1 \left\{ \frac{(b_{j-1} - \rho a_k)}{(1 - \rho^2)^{1/2}} \right\} \right].$$

From the symmetry it also follows that

$$(14)\qquad \frac{\partial l}{\partial b_m} = \sum_{i=1}^{s} \left(\frac{n_{im}}{\pi_{im}} - \frac{n_{i,m+1}}{\pi_{i,m+1}} \right) \cdot \phi_1(b_m) \cdot \left[\Phi_1 \left\{ \frac{(a_i - \rho b_m)}{(1 - \rho^2)^{1/2}} \right\} - \Phi_1 \left\{ \frac{(a_{i-1} - \rho b_m)}{(1 - \rho^2)^{1/2}} \right\} \right].$$

Equations (9), (13) and (14) constitute the set of first order derivatives of the log likelihood.

Case 2: The Thresholds are Computed from the Marginals

Most earlier researchers in this area [Pearson & Pearson, 1922; Lancaster & Hamdan, 1964; Hamdan, 1971; Martinson & Hamdan, 1971, 1975] have regarded the thresholds as given from the cumulative marginal proportions of the table. Although theoretically non-optimal, this approach has the advantage of reducing the computational labor.

In this case, the equation system to be solved is

$$(9)\qquad \frac{\partial l}{\partial \rho} = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{n_{ij}}{\pi_{ij}} [\phi_2(a_i, b_j) - \phi_2(a_{i-1}, b_j) - \phi_2(a_i, b_{j-1}) + \phi_2(a_{i-1}, b_{j-1})] = 0$$

$$(15)\qquad a_i = \Phi_1^{-1}(P_{i\cdot})$$

$$(16)\qquad b_j = \Phi_1^{-1}(P_{\cdot j}),$$

where P_{ij} is the observed proportion in cell (i, j), and where $P_{i\cdot}$ and $P_{\cdot j}$ are observed cumulative marginal proportions of the table, i.e.,

$$(17)\qquad P_{i\cdot} = \sum_{k=1}^{i} \sum_{j=1}^{r} P_{kj}$$

and

$$(18)\qquad P_{\cdot j} = \sum_{i=1}^{s} \sum_{k=1}^{j} P_{ik}.$$

4. *Variance/Covariance Estimates*

Case 1

We denote the sample size by N, and introduce the notation $\boldsymbol{\theta}' = (\rho, a_1, a_2, \cdots, a_{s-1}, b_1, b_2 \cdots, b_{r-1})$. The matrix $\mathbf{I}_{(\theta)}$ of expected second order derivatives of l with respect to $\boldsymbol{\theta}$ is obtained from

$$(19) \qquad [\mathbf{I}_{(\theta)}]_{m,n} = N \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{1}{\pi_{ij}} \left(\frac{\partial \pi_{ij}}{\partial \theta_m}\right)\left(\frac{\partial \pi_{ij}}{\partial \theta_n}\right)$$

[Tallis, 1962, p 348]. The derivatives within parenthesis in (19) are obtained from (8) and (10).

A large-sample estimate of the covariance matrix of $\hat{\boldsymbol{\theta}}$ is therefore

$$(20) \qquad \mathbf{V} = \mathbf{I}_{(\hat{\theta})}^{-1}$$

Case 2

We define $\mathbf{P}' = (P_{11}, P_{12}, \cdots, P_{1r}, P_{21}, \cdots, P_{2r} \cdots, P_{s1}, \cdots, P_{sr})$. The estimators are defined as the solution to the system

$$(21) \qquad \begin{aligned} F_1(\boldsymbol{\theta}, \mathbf{P}) &= \frac{\partial l}{\partial \rho} \cdot \frac{1}{N} = 0 \\ [F_2(\boldsymbol{\theta}, \mathbf{P})]_i &= a_i - \Phi_1^{-1}(P_{i.}) = 0 \\ [F_3(\boldsymbol{\theta}, \mathbf{P})]_j &= b_j - \Phi_1^{-1}(P_{.j}) = 0 \end{aligned}$$

Let $F = (F_1 \quad F_2' \quad F_3')'$.

The estimator $\hat{\boldsymbol{\theta}}$ is implicitly defined from $F(\hat{\boldsymbol{\theta}}, \mathbf{P}) = \mathbf{0}$.

Since

$$(22) \qquad \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{P}} = -\left(\frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}}\right)^{-1} \frac{\partial \mathbf{F}}{\partial \mathbf{P}},$$

it follows from asymptotic theory that the large-sample covariance matrix for $\hat{\boldsymbol{\theta}}$ is,

$$(23) \qquad \operatorname{cov}(\hat{\boldsymbol{\theta}}) = \left(\frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}}\right)^{-1} \left(\frac{\partial \mathbf{F}}{\partial \mathbf{P}}\right) \boldsymbol{\Sigma} \left(\frac{\partial \mathbf{F}}{\partial \mathbf{P}}\right)' \left(\frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}}\right)^{-1\prime}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{P}$, and where $\partial \mathbf{F}/\partial \mathbf{P}$ and $\partial \mathbf{F}/\partial \boldsymbol{\theta}$ are computed in the probability limit.

In (23), $\partial \mathbf{F}/\partial \boldsymbol{\theta}$ is structured as

$$(24) \qquad \begin{bmatrix} \dfrac{\partial^2 l}{\partial \rho^2} & \dfrac{\partial^2 l}{\partial \rho \partial \mathbf{a}} & \dfrac{\partial^2 l}{\partial \rho \partial \mathbf{b}} \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & \mathbf{I} \end{bmatrix}$$

The second order derivatives in (24) are given in the Appendix. Since $n_{ij} = N \cdot P_{ij}$, we get for the first row of $\partial \mathbf{F}/\partial \mathbf{P}$:

$$(25) \qquad \frac{\partial F_1}{\partial P_{ij}} = \frac{1}{\pi_{ij}} [\phi_2(a_i, b_j) - \phi_2(a_{i-1}, b_j) - \phi_2(a_i, b_{j-1}) + \phi_2(a_{i-1}, b_{j-1})].$$

Also, since $D[f^{-1}(y_0)] = 1/D(f(x_0))$, and since $\partial P_{k.}/\partial P_{ij} = 1$, we get

(26)
$$\frac{\partial F_{2k}}{\partial P_{ij}} = \begin{cases} 0 \text{ if } P_k \text{ does not contain } P_{ij}, \text{ i.e. if } k < i \\ -\dfrac{\partial}{\partial P_{ij}}[\Phi^{-1}(P_{k.})] = -\dfrac{\partial \Phi_1^{-1}}{\partial P_{k.}} \cdot \dfrac{\partial P_{k.}}{\partial P_{ij}} \\ = -\dfrac{1}{\dfrac{\partial}{\partial a_k}[\Phi_1(a_k)]} \\ = -\dfrac{1}{\phi_1(a_k)} \quad \text{otherwise.} \end{cases}$$

Similarly,

(27)
$$\frac{\partial F_{3m}}{\partial P_{ij}} = \begin{cases} 0 \quad \text{if } m < j \\ -\dfrac{1}{\phi_1(b_m)} \quad \text{otherwise.} \end{cases}$$

In Σ, finally, the general element σ_{mn} is obtained from

(28)
$$N\sigma_{mn} = \begin{cases} P_m(1 - P_m) & \text{if } m = n \\ -P_m P_n & \text{otherwise.} \end{cases}$$

5. *Some Numerical Results*

The formulae of the preceeding section have been used in a computer program for maximum likelihood estimation of polychoric correlations, written by the author. The program reads a cross-table (or, at the user's option, raw data from which the cross-table is prepared). Starting values for the thresholds and the correlation are computed, and the program iteratively solves the likelihood equations by a Newton-Raphson algorithm. The program uses subroutines from the IMSL [1975] library, except for the bivariate normal distribution function, which was adopted from Kirk [1973]. The output consists of final threshold estimates, correlation estimate, variance/covariance matrix of the estimates, and the function value at the maximum. The latter may be used to test the fit of the model. Empty rows and/or columns of the table are deleted from the computations. The numerical results consist of the following steps:

(i) Some runs were made with the program where the cell frequencies are expected values, taken from some division of the bivariate normal surface. Thus, these runs are made with artificial population data. These runs may be used partly to check the numerical accuracy of the program, but first of all they give the expected variances and covariances of the estimates, in large samples. If the expected correlations between $\hat{\rho}$ and the thresholds are small this may serve as an indication that simultaneous estimation of ρ and thresholds is unnecessary.
(ii) Some runs were made with Monte Carlo data, generated from the population data. These simulations were performed in order to check the theoretical large-sample results derived in Section 4.
(iii) Finally, some empirical data were analyzed.

Artificial Population Data

Design. The data were chosen according to a factorial design with the following parameters. The true correlation, ρ, was chosen as .15, .50 and .85. The number of classes in the table, $r = s$, was 2, 3, 5 and 7. The thresholds were placed so that, for each variable, the probabilities of the classes $1, \cdots, s$ were the same as those of a binomial distribution with parameters P and $n = s - 1$. Parameter P was chosen so that the skewness of the distribution attained specified values γ. This procedure was the same as that of Olsson [1979]. The combinations of skewnesses used were

$$(0,0), (1, -1), (1, 0) \text{ and } (1, 1).$$

Results. The absolute difference between true correlation and the correlation computed by the program was in no case larger than 10^{-5}, which implies reasonable accuracy of the program. For reasons of space, we do not exhibit all the variance/covariance matrices of the parameter estimates here. In Table 2 we only give one typical example.

For low values of ρ, the expected correlations between $\hat{\rho}$ and the thresholds are low; for $\rho = 0$ they are zero. For higher values of ρ, however, these expected correlations are in some cases more substantial, with a typical peak value of .20. This implies an increasing degree of dependence between correlation estimate and threshold estimates with increasing ρ. In the Monte Carlo data we shall study how much this means in practice.

In Table 3 we present the expected variance for $\hat{\rho}$ for the different combinations of parameters, given $N = 500$. For other values of N, the variances are proportional. The values are the same for both methods, to the given accuracy.

The case $r = s = 2$ was included in the computations with the following problem in mind: Given a set of skewed ordinal data, should the correlations be computed as (a) a polychoric correlation, or (b) a tetrachoric correlation from data dichotomized near the median?

It can be seen in Table 3, that the variances are uniformly smaller for the polychoric than for the tetrachoric correlation. Given highly skewed trichotomous data ($\gamma_1/\gamma_2 = 1./-1.$ or $1./1.$) the loss in efficiency if it is possible to dichotomize at the median is moderate, or may even be a small gain. In general, however, there seems to be a loss in efficiency of alternative (b) as compared to alternative (a) above.

TABLE 2

An Example of the Expected Covariance (in and below the diagonal) and Correlation (above the diagonal) Matrix for the Parameter Estimates. Artificial Population Data with $r=s=5$, $\gamma_1=1.$, $\gamma_2=0.$ and $\rho=.15$

	ρ	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4
ρ	.00281	.01895	.04680	.05151	-.00368	.05022	.02250	-.02482	-.04738
a_1	.00059	.33913	.20104	.06310	.02311	.00822	.00972	.00761	.00379
a_2	.00040	.01873	.02558	.31104	.11377	.02717	.03449	.02914	.01598
a_3	.00021	.00277	.00375	.00569	.36431	.05186	.07101	.06473	.03880
a_4	-.00001	.00075	.00102	.00154	.00315	.06111	.09513	.09126	.06003
b_1	.00023	.00042	.00038	.00034	.00030	.00774	.38355	.17451	.06619
b_2	.00007	.00033	.00032	.00031	.00030	.00198	.00343	.45465	.17455
b_3	-.00008	.00026	.00027	.00029	.00030	.00090	.00156	.00343	.38352
b_4	-.00022	.00019	.00022	.00026	.00030	.00051	.00090	.00198	.00774

TABLE 3

Expected Variances for r for the Artificial Population Data. The Values are the Same for Both Methods.

		γ_1/γ_2			
ρ	r	0./0.	1./-1.	1./0.	1./1.
.15	2	.00478	.00635	.00545	.00607
	3	.00295	.00429	.00354	.00417
	5	.00231	.00344	.00281	.00339
	7	.00215	.00318	.00261	.00314
.50	2	.00329	.00494	.00382	.00405
	3	.00194	.00313	.00238	.00272
	5	.00144	.00238	.00182	.00218
	7	.00131	.00215	.00166	.00201
.85	2	.00080	.00295	.00108	.00096
	3	.00046	.00104	.00064	.00063
	5	.00028	.00069	.00040	.00047
	7	.00023	.00058	.00034	.00041

Monte Carlo Data

Design. The parameters for the Monte Carlo data were chosen using the same factorial design as above, except that $r = 2$ was excluded. For each combination of parameters, 10 samples of size 500 were generated, using the multinomial routine GGMUL of the IMSL [1975] library. Each sample was analyzed using both methods.

Results. The estimates of the correlation coefficients in the different samples are displayed in Tables 4 to 6, for $s = 3$, 5 and 7, respectively, along with the mean value and the variance in each cell. In most cases, the two methods produce estimates which are very similar. A closer inspection of Tables 4 to 6 reveals, that the differences between the methods does increase with increasing ρ. For $\rho = .15$ the largest difference is 8×10^{-4}, for $\rho = .50$ it is 31×10^{-4} and for $\rho = .85$ it is 43×10^{-4}. This is in agreement with the results given above. These values might give an indication of magnitude of the difference between the two methods.

The question of bias of the estimates may be studied in several ways. We have performed the following comparisons:

The mean value of the 10 sample values were tested against the corresponding true value. Both when the theoretical and when the sample variance were used, 4 of the 36 sample means were significantly different from the true value at the 5% level. Since all the significant values were higher than the corresponding true values, this might indicate a

TABLE 4

Estimated Correlations in Samples Generated with the Indicated Values of True Correlations and Skewnesses
r=s=3. n=500
The left column in each cell is the two-step estimate, the right column is the full maximum likelihood estimate.

ρ		γ_1/γ_2							
		0./0.		1./-1.		1./0.		1./1.	
.15		.1277	.1277	.0342	.0343	.1647	.1648	.2021	.2023
		.2245	.2245	.1414	.1414	.1319	.1319	.0208	.0208
		.2007	.2007	.1754	.1755	.0346	.0346	.1637	.1638
		.1018	.1018	.1912	.1912	.1859	.1859	.1703	.1702
		.0767	.0767	.2324	.2324	-.0025	-.0026 *	.1780	.1781
		.2032	.2033	.1649	.1649	.1792	.1791	.2122	.2125
		.0740	.0740	.1560	.1561	.1566	.1567	.1439	.1441
		.1983	.1983	.1700	.1701	.0643	.0643	.1020	.1020
		.1718	.1718	.1458	.1458	.1551	.1551	.2000	.2000
		.1546	.1546	.1418	.1419	.0622	.0622	.0193	.0193 *
	$\bar{x}$	.1533	.1533	.1553	.1554	.1132	.1132	.1412	.1413
	s^2	.0028	.0028	.0023	.0023	.0041	.0041	.0046	.0046
.50		.4840	.4846	.5327	.5325	.4583	.4586	.5285	.5257
		.5344	.5345	.5107	.5105	.4714	.4713	.5513	.5487
		.5197	.5199	.4851	.4857	.5594	.5599	.4699	.4699
		.5326	.5329	.3915	.3915	.4450	.4447	.6112	.6102 *
		.4892	.4896	.4162	.4161	.4914	.4915	.5604	.5622
		.5399	.5392	.4317	.4319	.5002	.5002	.4205	.4197
		.4681	.4678	.4023	.4025	.4748	.4748	.5474	.5484
		.5063	.5058	.5023	.5023	.4803	.4803	.4379	.4385
		.5156	.5152	.5412	.5420	.5877	.5876 *	.4638	.4650
		.5427	.5424	.6187	.6188 *	.5115	.5117	.5327	.5329
	$\bar{x}$	.5133	.5132	.4832	.4833	.4980	.4981	.5124	.5121
	s^2	.0006	.0006	.0047	.0047	.0018	.0018	.0034	.0033
.85		.8326	.8321	.8440	.8440	.8484	.8484	.8544	.8555
		.8446	.8449	.8781	.8781	.8400	.8402	.8968	.8958 *
		.8681	.8677	.8249	.8249	.8828	.8830	.8730	.8714
		.8527	.8532	.8791	.8791	.8306	.8307	.8570	.8574
		.8739	.8737	.8775	.8775	.8761	.8762	.8656	.8665
		.8307	.8301	.8506	.8506	.8573	.8574	.8373	.8360
		.8491	.8485	.8764	.8764	.8592	.8594	.8414	.8399
		.8715	.8728	.8427	.8427	.8645	.8647	.8680	.8697
		.8389	.8393	.8231	.8231	.8452	.8454	.8630	.8619
		.8679	.8677	.8433	.8433	.8654	.8657	.8431	.8460
	$\bar{x}$	.8530	.8530	.8540	.8540	.8570	.8571	.8600	.8600
	s^2	.0002	.0002	.0004	.0004	.0002	.0002	.0003	.0003

Note: Underlined entries are significantly different from the true value (5% level) when the true variance is used. Starred (*) entries are significantly different from the true value (5% level) when the variance estimate is taken from the sample. The stars apply to both entries in each pair.

TABLE 5

Estimated Correlations in Samples Generated with the Indicated Values of True Correlations and Skewnesses. r=s=5. n=500
The left column in each cell is the two-step estimate, the right column is the full maximum likelihood estimate.

ρ		γ_1/γ_2 0./0.	1./-1.	1./0.	1./1.
.15		.2540 .2540*	.0656 .0657	.1762 .1763	.0325 .0326
		.1449 .1449	.1865 .1866	.2676 .2678*	.1329 .1329
		.1909 .1909	.1738 .1738	.1575 .1575	.0767 .0767
		.1700 .1700	.1523 .1523	.1399 .1403	.1306 .1307
		.2247 .2247	.0941 .0940	.2306 .2305	.1457 .1457
		.1760 .1760	.1303 .1304	.2312 .2313	.1773 .1773
		.1426 .1426	.0462 .0462	.1927 .1935	.1469 .1471
		.1456 .1456	.2315 .2316	.1419 .1419	.2483 .2483
		.1636 .1636	.1306 .1307	.1612 .1612	.2187 .2188
		.2072 .2072	.1010 .1010	.1589 .1590	.1603 .1605
	$\bar{x}$	.1820 .1820*	.1312 .1312	.1858 .1859*	.1470 .1471
	s^2	.0012 .0012	.0029 .0029	.0017 .0017	.0035 .0035
.50		.5574 .5587	.5080 .5085	.4949 .4956	.4008 .4020
		.5161 .5159	.5026 .5029	.5040 .5044	.4922 .4927
		.5588 .5581	.5532 .5535	.5487 .5486	.4311 .4320
		.5187 .5174	.4987 .4988	.5708 .5696	.4799 .4804
		.5522 .5524	.4027 .4024	.4904 .4899	.5107 .5098
		.4844 .4841	.5244 .5251	.4569 .4563	.5023 .5028
		.4782 .4780	.4348 .4352	.5050 .5044	.4790 .4785
		.4690 .4689	.6096 .6095*	.5421 .5413	.5404 .5412
		.5420 .5433	.5222 .5223	.5292 .5292	.5102 .5113
		.5400 .5398	.5119 .5123	.5435 .5442	.5372 .5377
	$\bar{x}$	.5216 .5217	.5068 .5071	.5186 .5184	.4884 .4888
	s^2	.0010 .0011	.0029 .0029	.0010 .0010	.0017 .0017
.85		.8738 .8734	.8370 .8373	.8352 .8350	.8582 .8574
		.8613 .8615	.8414 .8416	.8462 .8458	.8741 .8730
		.8782 .8775	.8600 .8599	.8557 .8570	.8605 .8567
		.8717 .8711	.8262 .8261	.8921 .8932*	.8321 .8284
		.8652 .8654	.8183 .8180	.8599 .8599	.8511 .8507
		.8429 .8429	.8497 .8496	.8548 .8560	.8754 .8765
		.8445 .8455	.8468 .8471	.8153 .8158	.8558 .8600
		.8560 .8562	.8727 .8726	.8552 .8566	.8499 .8513
		.8851 .8859*	.8306 .8305	.8659 .8675	.8212 .8208
		.8751 .8749	.8297 .8296	.8337 .8335	.8446 .8465
	$\bar{x}$	.8654 .8654*	.8412 .8412	.8514 .8520	.8523 .8521
	s^2	.0002 .0002	.0002 .0002	.0004 .0004	.0003 .0003

Note: Underlined entries are significantly different from the true value (5% level) when the true variance is used. Starred (*) entries are significantly different from the true value (5% level) when the variance estimate is taken from the sample. The stars apply to both entries in each pair.

TABLE 6

Estimated Correlations in Samples Generated with the Indicated Values of True Correlations and Skewnesses.
r=s=7 n=500
The left column in each cell is the two-step estimate, the right column is the full maximum-likelihood estimate.

ρ		γ_1/γ_2 0./0.		1./-1.		1./0.		1./1.	
.15		.1428	.1433	.1429	.1430	.2075	.2078	.0702	.0702
		.1168	.1169	.2287	.2290	.0984	.0986	.1703	.1703
		.1690	.1693	.1676	.1676	.1601	.1601	.1399	.1399
		.1031	.1033	.2381	.2385	.1058	.1058	.1872	.1870
		.1543	.1546	.2229	.2228	.1334	.1336	.2351	.2353
		.2256	.2262	.1140	.1141	.1963	.1964	.2274	.2275
		.1778	.1779	.1557	.1557	.1292	.1294	.1645	.1644
		.1426	.1428	.2274	.2274	.1779	.1779	.1572	.1573
		.1257	.1259	.1927	.1931	.1676	.1679	.2023	.2026
		.2409	.2413*	.1105	.1105	.0999	.1001	.0295	.0296 *
	$\bar{x}$	.1599	.1602	.1801	.1802	.1476	.1478	.1584	.1584
	s^2	.0018	.0018	.0021	.0021	.0014	.0014	.0038	.0038
.50		.4542	.4555	.5216	.5216	.5022	.5049	.5289	.5287
		.5037	.5041	.4829	.4829	.5083	.5097	.5072	.5066
		.4793	.4802	.3976	.3984 *	.5126	.5148	.5429	.5399
		.5283	.5288	.4994	.4991	.5698	.5697	.4958	.4956
		.5607	.5611	.4809	.4807	.4982	.4990	.6026	.6051 *
		.5314	.5335	.4220	.4222	.4688	.4692	.4949	.4947
		.4905	.4908	.4707	.4712	.4978	.4983	.5658	.5653
		.4223	.4226*	.4916	.4918	.4948	.4949	.5566	.5583
		.4678	.4686	.4804	.4806	.4424	.4417	.5670	.5680
		.5415	.5430	.4820	.4824	.5384	.5408	.4700	.4722
	$\bar{x}$	.4980	.4988	.4729	.4731	.5033	.5043	.5332	.5334 *
	s^2	.0017	.0017	.0012	.0012	.0011	.0011	.0015	.0015
.85		.8470	.8476	.8493	.8493	.8362	.8388	.8313	.8286
		.8718	.8728	.8418	.8421	.8523	.8528	.8438	.8434
		.8323	.8334	.8916	.8915*	.8633	.8632	.8713	.8700
		.8267	.8270	.8621	.8621	.8414	.8427	.8579	.8596
		.8745	.8750	.8131	.8130	.8730	.8727	.8743	.8786
		.8445	.8457	.8859	.8859	.8671	.8678	.8390	.8388
		.8106	.8105*	.8545	.8549	.8237	.8279	.8548	.8555
		.8772	.8769*	.9210	.9210*	.8719	.8712	.8321	.8326
		.8215	.8228	.8502	.8503	.8713	.8711	.8568	.8582
		.8506	.8520	.8585	.8585	.8285	.8289	.7953	.7967*
	$\bar{x}$	.8457	.8464	.8628	.8629	.8529	.8537	.8457	.8462
	s^2	.0005	.0005	.0008	.0008	.0003	.0003	.0005	.0005

Note: Underlined entries are significantly different from the true value (5% level) when the true variance is used. Starred (*) entries are significantly different from the true value (5% level) when the variance estimate is taken from the sample. The stars apply to both entries in each pair.

tendency towards a slight positive bias. The results were the same for both methods. The significant values have been indicated in Tables 4–6.

The variances among the 10 observations in each sample were tested against the corresponding theoretical variances, using χ^2 with 9 d.f. Only one value was significant at the 5% level (see Tables 4–6).

Each single correlation estimate was tested against the corresponding true value, both using the theoretical variance and the estimate from the sample. The significant values at the 5% level are indicated in Tables 4–6. Using the theoretical variance, 20 samples out of 360 (5.6%) were significant, of which 11 were over-, and 9 under-estimates. For the empirical variance estimates, 21 were significant (5.8%) for the full ML estimate, and 21 (5.8%) for the two-step estimate. It may be noted that the differences in variance estimates between the two methods are very small. An example of a generated crosstable, including population parameter values and estimates, is given in Table 7.

Analysis of Some Real Data

As a complement to the Monte Carlo results, we have also analyzed some real data with each of the two methods. The data consist of answers to nine five-step attitude items

TABLE 7

A Generated Cross-table, and the Corresponding Parameter Values and Sample Estimates.

x \ y	1	2	3
1	13	6	0
2	69	113	22
3	41	132	104

	ρ	a_1	a_2	b_1	b_2
Population value	.50	-1.70	-.31	-.67	.67
Sample estimate	.49	-1.77	-.14	-.69	.67
Estimated standard error	.048	.103	.056	.061	.061

TABLE 8

Some Real Data Analyzed with the Two Methods.

Variables		Raw correlation	ML-estimate	TS-estimate
2	1	.470	.7277	.7215
3	1	.413	.5979	.5958
3	2	.390	.6267	.6240
4	1	-.084	-.2444	-.2412
4	2	-.001	-.0995	-.0990
4	3	-.459	-.6074	-.5899
5	1	.448	.7026	.6878
5	2	.364	.6202	.6140
5	3	.455	.6496	.6453
5	4	-.128	-.2752	-.2712
6	1	.414	.5898	.5880
6	2	.323	.5362	.5348
6	3	.540	.6583	.6523
6	4	-.313	-.4187	-.4116
6	5	.479	.7056	.6978
7	1	.355	.4912	.4908
7	2	.329	.5083	.5065
7	3	.431	.5353	.5319
7	4	-.172	-.2511	-.2478
7	5	.395	.6075	.6016
7	6	.556	.6531	.6470
8	1	.187	.2767	.2760
8	2	.081	.2021	.2013
8	3	.169	.2308	.2301
8	4	.015	.0170	.0170
8	5	.210	.3366	.3351
8	6	.202	.2533	.2525
8	7	.112	.1639	.1628
9	1	-.129	-.2519	-.2508
9	2	.049	-.0270	-.0269
9	3	-.340	-.4426	-.4385
9	4	.427	.5299	.5246
9	5	-.139	-.2751	-.2728
9	6	-.254	-.3312	-.3284
9	7	-.205	-.2775	-.2742
9	8	.087	.1055	.1050

of very varying skewness (about from −3. to +2.). The data was kindly supplied by Dr. Thorleif Pettersson. The sample size was 329. In Table 8 we give the raw correlations, as well as correlations estimated with the two-step (TS) and full maximum likelihood (ML) methods.

The results do support the conclusions arrived at earlier, that the differences between the methods are quite small, but that they increase with increasing ρ. The differences between the correlations estimated by our methods and the correlations computed as if the data had been on an interval scale (in the table denoted "Raw correlation") are seen to be large.

6. Discussion

To summarize the results of the preceding section, it seems that the bias of the estimates is small, and that the variances are close to the theoretically derived ones for the data analyzed so far. Inspection of the cross-tables for the few extreme cases does suggest that bad estimates are more likely in tables where some marginals are small, i.e. where some expected cell frequencies are low. Since we have not studied this problem systematically, this statement must be taken in a loose sense, however.

For practical purposes, the differences between the two methods discussed here seem to be small, especially when the true correlation is small. This also applies for the variance estimates. It may be noted here, that the simpler variance estimates used by Martinson and Hamdan [1971] also turn out to be sufficiently accurate for practical work.

One problem which may have some practical importance is, that when several correlations are estimated, the full maximum likelihood estimate may lead to different threshold estimates for variable x when ρ_{xy} is estimated than when ρ_{xz} is estimated. From a theoretical point of view, this is not entirely satisfactory. One solution is to use the two-step estimate, for which the thresholds are estimated from the marginals. A second solution would be to estimate the correlations for all variables simultaneously, including all the thresholds.

Another problem concerns the robustness of the methods. We have assumed here that underlying each response there is some latent variable which is normally distributed; an assumption, by the way, which is testable. In applications, such distributional assumptions are seldom exactly met. It might be a worthwhile task to examine to what degree departures from the assumption of normality has any effect on the correlation estimates.

Appendix: Some Details of the Derivations

We shall here derive the derivative of F_1 with respect to ρ, a_i and b_j, computed at the probability limit.

$$\text{(A1)} \qquad F_1 = \frac{1}{N}\sum\sum \frac{n_{ij}}{\pi_{ij}}[\phi(a_i, b_j) - \phi(a_{i-1}, b_j) - \phi(a_i, b_{j-1}) + \phi(a_{i-1}, b_{j-1})].$$

Since

$$\text{(A2)} \qquad \frac{\delta}{\delta\rho}[\phi(u, v)] = \phi(u, v)\cdot\left\{\frac{uv(1-\rho^2) + \rho(u^2 - 2\rho uv + v^2) - 2\rho(1-\rho^2)}{(1-\rho^2)^2}\right\} = g(u, v),$$

say, it follows that the derivative of F_1 with respect to ρ may be written

$$\text{(A3)} \qquad \frac{\delta F_1}{\delta\rho} = \frac{1}{N}\sum\sum \frac{n_{ij}}{\pi_{ij}}[g(a_i, b_j) - g(a_{i-1}, b_j) - g(a_i, b_{j-1}) + g(a_{i-1}, b_{j-1})]$$
$$- \frac{1}{N}\sum\sum \frac{n_{ij}}{\pi_{ij}^2}\cdot\frac{\delta\pi_{ij}}{\delta\rho}\cdot[\phi(a_i, b_j) - \phi(a_{i-1}, b_j) - \phi(a_i, b_{j-1}) + \phi(a_{i-1}, b_{j-1})].$$

Since, in the probability limit, $n_{ij}/N\pi_{ij} = 1$, this reduces to

(A4)
$$\operatorname*{plim}_{N\to\infty} \frac{\partial F_1}{\partial \rho} = \sum\sum [g(a_i, b_j) - g(a_{i-1}, b_j) - g(a_i, b_{j-1}) + g(a_{i-1}, b_{j-1})] - \sum\sum \frac{1}{P_{ij}} \left(\frac{\partial \pi_{ij}}{\partial \rho}\right)^2 .$$

In (A4), it is easily seen that the first sum is reduced to

(A5)
$$g(a_s, b_r) - g(a_s, b_0) - g(a_0, b_r) + g(a_0, b_0) = 0.$$

This is so because all other terms cancel, and because g is zero in all points containing α_0, α_s, b_0 and/or b_r. Thus,

(A6)
$$\operatorname*{plim}_{N\to\infty} \frac{\partial F_1}{\partial \rho} = -\sum\sum \frac{1}{\pi_{ij}} \left(\frac{\partial \pi_{ij}}{\partial \rho}\right)^2 .$$

For the derivatives with regard to the thresholds, we may write

(A7)
$$\frac{\delta F_1}{\delta a_k} = \frac{1}{N} \sum \left\{ \frac{n_{ij}}{\pi_{ij}} \left[\frac{\delta \phi(a_k, b_j)}{\delta a_k} - \frac{\delta \phi(a_k, b_{j-1})}{\delta a_k} \right] - \frac{n_{i+1,j}}{\pi_{i+1,j}} \left[\frac{\delta \phi(a_k, b_j)}{\delta a_k} - \frac{\delta \phi(a_k, b_{j-1})}{\delta a_k} \right] \right\} - \frac{1}{N} \sum\sum \frac{n_{ij}}{\pi_{ij}^2} \cdot \frac{\delta \pi_{ij}}{\delta a_k} [\phi(a_i, b_j) - \phi(a_{i-1}, b_j) - \phi(a_i, b_{j-1}) + \phi(a_{i-1}, b_{j-1})].$$

Since, as noted before,

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \frac{n_{ij}}{\pi_{ij}} = 1,$$

the two terms in the first sum cancel. Therefore,

(A8)
$$\operatorname*{plim}_{N\to\infty} \frac{\partial F_1}{\partial a_k} = -\sum\sum \frac{1}{\pi_{ij}} \left(\frac{\partial \pi_{ij}}{\partial \rho}\right) \cdot \left(\frac{\partial \pi_{ij}}{\partial a_k}\right) .$$

The result for $\partial F_1/\partial b_m$ is analogous.

REFERENCE NOTE

1. Pearson, K. Mathematical contributions to the theory of evolution. XIII: On the theory of contingency and its relation to association and normal correlation. *Drapers Company Research Memoirs, Biometrics Series,* No. 1, 1904.

REFERENCES

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

Froemel, E. C. A comparison of computer routines for the calculation of the tetrachoric correlation coefficient. *Psychometrika,* 1971, *36,* 165–174.

Hamdan, M. A. On the structure of the tetrachoric series. *Biometrika,* 1968, *55,* 261–262.

Hamdan, M. A. The equivalence of tetrachoric and maximum likelihood estimates of ρ in 2×2 tables. *Biometrika,* 1970, *57,* 212–215.

Hamdan, M. A. On the polychoric series method for estimating ρ in contingency tables. *Psychometrika,* 1971, *36,* 253–259.

IMSL Library 1 (Ed. 5). Houston, Texas: International Mathematical and Statistical Libraries, 1975.

Johnson, N. L. & Kotz, S. *Distributions in statistics: Continuous multivariate distributions.* New York: Wiley, 1972.

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

Lancaster, H. O. & Hamdan, M. A. Estimation of the correlation coefficient in contingency tables with possibly nonmetrical characters. *Psychometrika,* 1964, *29,* 383–391.

Martinson, E. O. & Hamdan, M. A. Maximum Likelihood and some other asymptotically efficient estimators of correlation in two way contingency tables. *Journal of Statistical Computation and Simulation,* 1971, *1,* 45–54.

Martinson, E. O. & Hamdan, M. A. Algorithm AS 87: Calculation of the polychoric estimate of correlation in contingency tables. *Applied Statistics,* 1975, *24,* 272–278.

Olsson, U. On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research,* 1979 (in press).

Pearson, K. Mathematical contributions to the theory of evolution, VII: On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, Series A,* 1901, *195,* 1–47.

Pearson, K. & Pearson E. S. On polychoric coefficients of correlation. *Biometrika,* 1922, *14,* 127–156.

Ritchie-Scott, A. The correlation coefficient of a polychoric table. *Biometrika,* 1918, *12,* 93–133.

Tallis, G. M. The maximum likelihood estimation of correlation from contingency tables. *Biometrics,* 1962, *18,* 342–353.

Wainer, H. & Thissen, D. Three steps towards robust regression. *Psychometrika,* 1976, *41,* 9–34.

Manuscript received 10/16/78
Final version received 6/21/79.