PSYCHOMETRIKA—VOL. 47, NO. 3.
SEPTEMBER, 1982

THE POLYSERIAL CORRELATION COEFFICIENT

ULF OLSSON

THE SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

FRITZ DRASGOW

YALE UNIVERSITY

NEIL J. DORANS

EDUCATIONAL TESTING SERVICE

The polyserial and point polyserial correlations are discussed as generalizations of the biserial and point biserial correlations. The relationship between the polyserial and point polyserial correlation is derived. The maximum likelihood estimator of the polyserial correlation is compared with a two-step estimator and with a computationally convenient ad hoc estimator. All three estimators perform reasonably well in a Monte Carlo simulation. Some practical applications of the polyserial correlation are described.

Key words: point polyserial correlation, dichotomous variables, polychotomous variables, latent variables.

Introduction

A categorical variable Y is often the result of coarse-grained measurement of an underlying continuous variable η. For example, a dichotomous variable is observed as $Y = 1$ when η exceeds some threshold value τ, and as $Y = 0$ otherwise. In psychology [Lazarsfeld, 1959; Lord & Novick, 1968], biometrics [Finney, 1971] and econometrics [Nerlove & Press, Note 1] there are many examples for which it is reasonable to assume that a continuous variable underlies a dichotomous or polychotomous observed variable.

Table 1 presents several of the correlational measures that have been developed to assess the relationship between two variables. Although there are a number of special names for correlations between observed variables having various scale properties, all of these correlations can be computed by the standard formula for a product moment correlation.

The tetrachoric correlation has been generalized to the case where the observed variables X and Y have r and s ordinal categories, respectively. This correlation is called the polychoric correlation coefficient. Tallis [1962] derived a maximum likelihood estimator for the polychoric correlation that may be used when $r = s = 3$. For the general case, Lancaster and Hamdan [1964] derived an estimator based on a series expansion, Martinson and Hamdan [1971] used a two-step estimator, and Olsson [1979] investigated the full maximum likelihood estimator, and compared it to the two-step estimator.

In this paper we consider the case where one observed variable is polychotomous and ordinal, and the other observed variable is continuous. The product moment correlation

By coincidence, the first author and the second and third authors learned that they were working independently on closely related problems and, consequently, decided to write a jointly authored paper.

Address reprint requests to Ulf Olsson, The Swedish University of Agricultural Sciences, Department of Economics and Statistics, S-750 07 Uppsala 7, SWEDEN, or, in North America, to Fritz Drasgow, Department of Psychology, University of Illinois, 603 E. Daniel Street, Champaign, IL 61820, U.S.A.

0033-3123/82/0900-4006$00.75/0

TABLE 1

Types of Correlation Coefficients as a Function of Scale Properties of the Observed Variables X and Y

	Scale of X		
Scale of Y	Dichotomous	Polychotomous-Ordinal Categories	Continuous-Interval
Dichotomous	Observed: Phi Inferred: Tetrachoric	Observed: No special term Inferred: Polychoric (Special Case)	Observed: Point Biserial Inferred: Biserial
Polychotomous-Ordinal Categories		Observed: No special term Inferred: Polychoric	Observed: Point Polyserial Inferred: Polyserial
Continuous-Interval			Observed and Inferred: Product Moment

Note: Latent variables are assumed to be normally distributed.

between these observed variables is called the *point polyserial correlation*, which is an obvious generalization of the point biserial correlation. Similarly, the biserial correlation [Pearson, 1909; Tate, 1955a, b] has been generalized to the *polyserial correlation.* Pearson [1913] and Jaspen [1946] studied the polyserial correlation under a very restrictive type of scoring for the categorical variable. The maximum likelihood estimator of the polyserial correlation has been derived by Cox [1974].

In the next section we derive the relationship between the point polyserial correlation and the polyserial correlation. The only assumption made about the scoring of the categorical variable is that numbers are assigned to categories in a strictly monotonic fashion. A maximum likelihood estimator (MLE), a two-step approximation to the MLE, and a computationally convenient ad hoc estimator of the polyserial correlation are then discussed. A Monte Carlo study is used to compare the three estimators. Finally, our results and their implications are summarized in the Discussion section.

The Relation Between the Polyserial and Point Polyserial Correlations

Model

The joint distribution of the observed continuous variable X and the latent variable η is assumed to be bivariate normal, with parameters $\mu_x = \mu$, $\sigma_x^2 = \sigma^2$, $\mu_\eta = 0$, $\sigma_\eta^2 = 1$ and $\rho_{x\eta} = \rho$. The categorical variable Y is assumed to be related to η by the step function

$$Y = y_j \quad \text{if} \quad \tau_{j-1} \leq \eta < \tau_j, \qquad j = 1, 2, \ldots, r, \tag{1}$$

where for convenience we define $\tau_0 = -\infty$ and $\tau_r = +\infty$. In (1),

$$y_{j-1} < y_j, \qquad j = 2, \ldots, r,$$

and

$$\tau_{j-1} < \tau_j, \qquad j = 2, \ldots, r - 1.$$

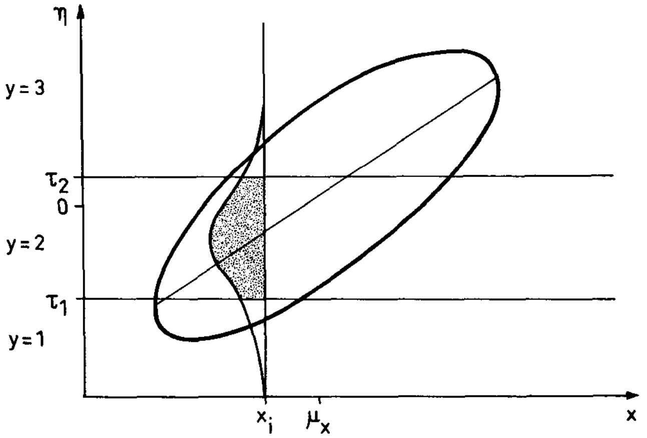

FIGURE 1
An example of a categorical variable Y formed by imposing thresholds upon an underlying continuous variable η.

Note that (1) allows unequal steps $(y_{j+1} - y_j)$ at the threshold values τ_j. The relations between X, Y, and η are illustrated in Figure 1.

Derivation

We now derive the relation of the *polyserial correlation* ρ between X and η, to the *point polyserial correlation* $\tilde{\rho}$ between X and Y. This is accomplished by using the relation between Y and η to determine the variance of Y and to determine the covariance between X and Y. In this subsection we assume without loss of generality that $\mu_x = 0$ and $\sigma_x^2 = 1$.

From (1), it is clear that the distribution of Y is

$$\text{Prob}(Y = y_j) = p_j = \Phi(\tau_j) - \Phi(\tau_{j-1}), \tag{2}$$

where

$$\Phi(\tau) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\tau} \exp\left(-\frac{t^2}{2}\right) dt. \tag{3}$$

The mean and variance of Y are

$$\mu_y = \sum_{j=1}^{r} y_j p_j, \tag{4}$$

and

$$\sigma_y^2 = \sum_{j=1}^{r} y_j^2 p_j - \mu_y^2. \tag{5}$$

Because $E(X) = 0$ and Y is a function of η, the covariance between X and Y is

$$E(XY) = \sum_{j=1}^{r} \int_{\tau_{j-1}}^{\tau_j} \int_{-\infty}^{\infty} x y_j \phi(x, u)\, dx\, du, \tag{6}$$

where

$$\phi(x, u) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-\frac{x^2 - 2\rho x u + u^2}{2(1-\rho^2)}\right\} \tag{7}$$

is the bivariate normal density function of X and η.

The j-th term on the right hand side of (6) is

$$\begin{aligned}
&\int_{\tau_{j-1}}^{\tau_j}\int_{-\infty}^{\infty} xy_j\,\phi(x, u)\,dx\,du \\
&= \frac{y_j}{2\pi(1-\rho^2)^{1/2}}\int_{\tau_{j-1}}^{\tau_j}\int_{-\infty}^{\infty} x\exp\left\{\frac{-(x^2-2\rho xu+u^2)}{2(1-\rho^2)}\right\}dx\,du \\
&= \frac{y_j}{(2\pi)^{1/2}}\int_{\tau_{j-1}}^{\tau_j}\exp\left\{-\frac{u^2}{2}\right\}\int_{-\infty}^{\infty}\frac{x}{[2\pi(1-\rho^2)]^{1/2}}\exp\left\{\frac{-(x-\rho u)^2}{2(1-\rho^2)}\right\}dx\,du \\
&= \frac{y_j}{(2\pi)^{1/2}}\int_{\tau_{j-1}}^{\tau_j}\rho u\exp\left\{-\frac{u^2}{2}\right\}du \\
&= -y_j\rho\{\phi(\tau_{j-1})-\phi(\tau_j)\}, \qquad (8)
\end{aligned}$$

where

$$\phi(\tau) = \frac{1}{(2\pi)^{1/2}}\exp\left\{-\frac{\tau^2}{2}\right\}. \qquad (9)$$

Substituting (8) into (6) and collecting terms yields

$$E(XY) = \sigma_{xy} = \rho\sum_{j=1}^{r-1}\phi(\tau_j)(y_{j+1}-y_j). \qquad (10)$$

Finally, the correlation between X and Y is

$$\tilde{\rho} = \rho\cdot\frac{1}{\sigma_y}\sum_{j=1}^{r-1}\phi(\tau_j)(y_{j+1}-y_j). \qquad (11)$$

Note that (11) presents the relation between the point polyserial correlation, $\tilde{\rho}$, and the polyserial correlation, ρ, in a convenient closed form. It is apparent that $\tilde{\rho}$ depends on the number of Y values, the location of the thresholds, τ_j, and the scoring system for the categories of Y.

Rating Scales

Consecutive integers are often assigned to categories of Y. A seven point rating scale, with $y_j = j$ for $j = 1, \ldots, 7$, provides one example. If consecutive integers are used, the relation in (11) becomes

$$\tilde{\rho} = \frac{\rho\,\Sigma\,\phi(\tau_j)}{\sigma_y}. \qquad (12)$$

As noted by Cox [1974], the relation between the point biserial and biserial correlations used by Tate [1955a] is a special case of (12).

Serial Correlation

The scoring procedure employed by Pearson [1913] and Jaspen [1946] in their derivations of the "serial correlation" defines the category score y_j as the mean of the latent variable η between thresholds τ_{j-1} and τ_j:

$$y_j = \frac{1}{p_j(2\pi)^{1/2}}\int_{\tau_{j-1}}^{\tau_j} z\exp\left\{-\frac{z^2}{2}\right\}dz = \frac{[\phi(\tau_{j-1})-\phi(\tau_j)]}{p_j}, \qquad (13)$$

where τ_0 and τ_r are taken as $-\infty$ and ∞, respectively, and p_j is defined in (2).

To derive the serial correlation from (11) and (13), first note that

$$\tilde{\rho} = \frac{1}{\sigma_y} \sum_{j=1}^{r} p_j y_j \mu_j, \tag{14}$$

where μ_j is the expectation of the observed continuous variable X given $Y = y_j$. From the relations in (11) and (14),

$$\rho = \frac{\sum_{j=1}^{r} p_j y_j \mu_j}{\sum_{j=1}^{r-1} \phi(\tau_j)(y_{j+1} - y_j)}. \tag{15}$$

Substituting the expressions for y_{j+1} and y_j from (13) into the numerator and denominator of (15), and then rearranging the summation in the denominator yields

$$\rho = \frac{\sum_{j=1}^{r} \mu_j[\phi(\tau_{j-1}) - \phi(\tau_j)]}{\sum_{j=1}^{r} \frac{[\phi(\tau_{j-1}) - \phi(\tau_j)]^2}{p_j}}. \tag{16}$$

Equation (16) is Jaspen's [1946] equation for the polyserial correlation.

Estimation of the Polyserial Correlation

In this section, three different methods for estimation of the polyserial correlation are discussed. We no longer assume that $\mu_x = 0$ and $\sigma_x^2 = 1$; instead they are considered to be unknown parameters that must be estimated. For simplicity, we assume that consecutive integers are used to score the categories of Y. Note that other scoring systems may affect the point polyserial correlation, but the correlation between X and η is unaffected by the choice of scoring for Y.

The estimation methods are:

(i) All model parameters are estimated simultaneously by the method of maximum likelihood.

(ii) The mean and variance of X are estimated by the sample mean $\bar{X}$ and the sample variance s_x^2, and inverse values of the normal distribution function evaluated at the cumulative marginal proportions of Y are taken as estimates of the thresholds $\tau_1, \tau_2, \ldots, \tau_{r-1}$. A conditional maximum likelihood estimate of ρ is then computed, given the other parameter estimates. This procedure, termed the *two-step method*, is computationally simpler than maximum likelihood estimation and is therefore appealing if its estimates have properties similar to those of maximum likelihood estimates.

(iii) Sample estimates of $\tilde{\rho}$, σ_y and the thresholds τ_j are inserted into (12), thus providing a simple ad hoc estimator.

The Maximum Likelihood Estimator (MLE)

The likelihood of a sample of N observations (x_i, y_i) is

$$L = \prod_{i=1}^{N} p(x_i, y_i) = \prod_{i=1}^{N} p(x_i)p(y_i \mid x_i), \tag{17}$$

where

$$p(x_i) = (2\pi\sigma)^{-1/2} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right]. \tag{18}$$

Let $Z = (X - \mu)/\sigma$; then the conditional probability in (17) can be obtained by noting that the conditional distribution of η given X is normal with mean ρZ and variance $(1 - \rho^2)$. Thus, the conditional distribution of Y is

$$\text{Prob}(Y = j \mid X) = \Phi(\tau_j^*) - \Phi(\tau_{j-1}^*), \qquad j = 1, 2, \ldots, r \tag{19}$$

where

$$\tau_j^* = \frac{\tau_j - \rho Z}{(1 - \rho^2)^{1/2}}.$$

The MLEs are obtained by maximizing

$$\ell = \log L = \sum_{i=1}^{N} [\log p(x_i) + \log p(y_i \mid x_i)], \tag{20}$$

i.e., by differentiating (20) with respect to the parameters, setting the partial derivatives equal to zero and solving the resulting simultaneous equation system for ρ, μ, σ^2 and τ_j, $j = 1, 2, \ldots, r - 1$.

The partial derivatives of ℓ with respect to the parameters are:

$$\frac{\partial \ell}{\partial \rho} = \sum \frac{1}{p(y_i \mid x_i)} \frac{\partial p(y_i \mid x_i)}{\partial \rho} \tag{21}$$

$$\frac{\partial \ell}{\partial \mu} = \sum \left(\frac{1}{p(x_i)} \cdot \frac{\partial p(x_i)}{\partial \mu} + \frac{1}{p(y_i \mid x_i)} \cdot \frac{\partial p(y_i \mid x_i)}{\partial \mu} \right) \tag{22}$$

$$\frac{\partial \ell}{\partial \sigma^2} = \sum \left(\frac{1}{p(x_i)} \cdot \frac{\partial p(x_i)}{\partial \sigma^2} + \frac{1}{p(y_i \mid x_i)} \cdot \frac{\partial p(y_i \mid x_i)}{\partial \sigma^2} \right) \tag{23}$$

$$\frac{\partial \ell}{\partial \tau_j} = \sum \frac{1}{p(y_i \mid x_i)} \cdot \frac{\partial p(y_i \mid x_i)}{\partial \tau_j}, \qquad j = 1, 2, \ldots, r - 1 \tag{24}$$

Let $z_i = (x_i - \mu)/\sigma$. Then the partial derivative of the conditional probability in (21) is

$$\begin{aligned} \frac{\partial p(y_i \mid x_i)}{\partial \rho} &= \frac{\partial \Phi(\tau_j^*)}{\partial \rho} - \frac{\partial \Phi(\tau_{j-1}^*)}{\partial \rho} \\ &= \phi(\tau_j^*) \frac{\partial \tau_j^*}{\partial \rho} - \phi(\tau_{j-1}^*) \frac{\partial \tau_{j-1}^*}{\partial \rho} \\ &= (1 - \rho^2)^{-3/2} [\phi(\tau_j^*)(\tau_j \rho - z_i) - \phi(\tau_{j-1}^*)(\tau_{j-1} \rho - z_i)]. \end{aligned} \tag{25}$$

Thus, (21) is

$$\frac{\partial \ell}{\partial \rho} = \sum_{i=1}^{N} \left\{ \frac{1}{p(y_i \mid x_i)} \frac{1}{(1 - \rho^2)^{3/2}} [\phi(\tau_j^*)(\tau_j \rho - z_i) - \phi(\tau_{j-1}^*)(\tau_{j-1} \rho - z_i)] \right\}. \tag{26}$$

In (26) and throughout this section, τ_{j-1}^* and τ_j^* are the thresholds surrounding y_i.

To obtain an explicit form for (22), note that

$$\frac{\partial p(x_i)}{\partial \mu} = p(x_i) \cdot \frac{z_i}{\sigma} \tag{27}$$

and

$$\frac{\partial p(y_i \mid x_i)}{\partial \mu} = \phi(\tau_j^*) \cdot \frac{1}{(1 - \rho^2)^{1/2}} \cdot \frac{\rho}{\sigma} - \phi(\tau_{j-1}^*) \cdot \frac{1}{(1 - \rho^2)^{1/2}} \cdot \frac{\rho}{\sigma}. \tag{28}$$

Then

$$\frac{\partial \ell}{\partial \mu} = \sum_{i=1}^{N} \frac{z_i}{\sigma} + \frac{1}{p(y_i \mid x_i)} \frac{\rho}{\sigma(1-\rho^2)^{1/2}} [\phi(\tau_j^*) - \phi(\tau_{j-1}^*)]. \tag{29}$$

To evaluate (23), note that

$$\frac{\partial p(x_i)}{\partial \sigma^2} = \frac{p(x_i)}{2\sigma^2}(z_i^2 - 1) \tag{30}$$

and

$$\frac{\partial p(y_i \mid x_i)}{\partial \sigma^2} = \frac{\rho}{(1-\rho^2)^{1/2}} \frac{z_i}{2\sigma^2} [\phi(\tau_j^*) - \phi(\tau_{j-1}^*)]. \tag{31}$$

Consequently,

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left\{ (z_i^2 - 1) + \frac{\rho z_i}{(1-\rho^2)^{1/2}} [\phi(\tau_j^*) - \phi(\tau_{j-1}^*)] \right\}. \tag{32}$$

Finally

$$\frac{\partial p(y_i = j \mid x_i)}{\partial \tau_k} = \begin{cases} 0 & \text{if } k \neq j, \quad k \neq j-1 \\ \dfrac{\phi(\tau_j^*)}{(1-\rho^2)^{1/2}}, & \text{if } k = j \\ -\dfrac{\phi(\tau_{j-1}^*)}{(1-\rho^2)^{1/2}}, & \text{if } k = j-1. \end{cases} \tag{33}$$

Thus,

$$\frac{\partial \ell}{\partial \tau_k} = \frac{1}{(1-\rho^2)^{1/2}} \sum_{i=1}^{N} \frac{1}{p(y_i \mid x_i)} [\delta_{k,\, y_i} \phi(\tau_j^*) - \delta_{k,\, (y_i - 1)} \phi(\tau_{j-1}^*)], \tag{34}$$

where $\delta_{m,\, m'}$ is Kronecker's delta.

The maximum likelihood estimate of ρ is obtained by solving the nonlinear equation system composed of (26), (29), (32) and (34) set equal to zero.

Variance estimates for the MLEs. Denote by $\boldsymbol{\theta}$ the vector containing all the parameters, i.e., $\boldsymbol{\theta}' = (\rho, \mu, \sigma^2, \tau_1, \tau_2, \ldots, \tau_{r-1})$. It follows from large-sample theory (e.g., Silvey, 1970) that the covariance matrix of $\hat{\boldsymbol{\theta}}$ may be estimated by $\text{Cov}(\hat{\boldsymbol{\theta}}) = \mathbf{I}_{(\hat{\theta})}^{-1}$. In large samples $\mathbf{I}_{(\theta)}$ is the information matrix, whose (m, n)th element is

$$I_{(\theta)}(m, n) = N \sum_{y=1}^{r} \int_{-\infty}^{\infty} \frac{\partial p(x, y)}{\partial \theta_m} \frac{\partial p(x, y)}{\partial \theta_n} \cdot \frac{1}{p(x, y)} \, dx. \tag{35}$$

Explicit expressions for individual elements of $\mathbf{I}_{(\theta)}$ may be found in Cox [1974]. Due to the complexity of these expressions, numerically approximating the information matrix (e.g., see Gruvaeus and Jöreskog, Note 2) may be preferable.

The Two-Step Estimator

The two-step estimator is obtained by estimating μ by $\bar{X}$, σ^2 by s^2 and by estimating the thresholds by the inverse normal distribution function evaluated at the cumulative marginal proportions, P_j, of y. For example, if $N = 200$, and if $y = 1$ is observed 40 times, then $P_1 = .2$ and $\hat{\tau}_1 = \Phi^{-1}(.2) = -.8416$. The estimate of ρ is obtained by maximizing (20) with respect to ρ only, i.e., by setting (26) equal to zero and solving for ρ. Formally, the equation system for the two-step estimator may be written as

$$
\begin{aligned}
F_1 &= \frac{\partial \ell}{\partial \rho} \cdot \frac{1}{N} = 0 \\
F_2 &= \mu - \frac{1}{N} \sum x_i = 0 \\
F_3 &= \sigma^2 - \frac{1}{N} \sum (x_i - \bar{X})^2 = 0 \\
\mathbf{F}_4 &= \tau_j - \Phi^{-1}(P_j) = 0, \qquad j = 1, \ldots, r-1.
\end{aligned}
\tag{36}
$$

The variance of the two-step estimator. Let $\mathbf{F}' = (F_1, F_2, F_3, \mathbf{F}_4')$ and $\mathbf{Z}' = (\bar{X}, s^2, P_1, \ldots, P_{r-1})$. It follows from asymptotic theory that the large-sample covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$
\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \left(\frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}}\right)^{-1} \left(\frac{\partial \mathbf{F}}{\partial \mathbf{Z}}\right) \boldsymbol{\Sigma} \left(\frac{\partial \mathbf{F}}{\partial \mathbf{Z}}\right)' \left[\left(\frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}}\right)^{-1}\right]',
\tag{37}
$$

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

The Ad Hoc Estimator

This estimator is obtained by substituting sample statistics into (12),

$$
\hat{\rho} = \frac{r_{xy} s_y}{\sum \phi(\hat{\tau}_j)},
\tag{38}
$$

where $\hat{\tau}_j = \Phi^{-1}(P_j)$ and r_{xy} is the sample product moment correlation of X and Y. A convenient estimate of its variance may be obtained by jackknifing [Mosteller, 1968].

Monte Carlo Simulation

A Fortran IV computer program was written to examine the effectiveness of the three estimators of ρ. Sample size (N), skewness (γ) of the polychotomous variable in the population, correlation (ρ) and number of categories (r) of the polychotomous variable are input to the program. The program then generates bivariate normal data by the IMSL [1975] GGNSM routine. Thresholds are determined that produce the proper amount of skewness in the polychotomous variable (see Olsson, 1979) and are then applied to one of the normal variables to yield the polychotomous variable.

For each sample generated in this way, all three estimators of ρ were computed. The ML estimator was obtained by numerically solving the equation system composed of (26), (29), (32) and (34), set equal to zero, using a subroutine developed by Gruvaeus and Jöreskog [Note 2]. The routine E04BBF from the NAG [1979] library was used for univariate maximization of (20) with respect to ρ. The ad hoc estimator, finally, was obtained by substituting sample statistics into (12).

Design of Simulation

To examine the behavior of the three estimators, parameters were manipulated in a four-way factorial design. The factors are:

(i) Sample size—samples of $N = 100$ and $N = 500$ were generated;
(ii) Skewness—thresholds that yield symmetric ($\gamma = 0$) and asymmetric ($\gamma = 1$) distributions were used;

(iii) Polyserial correlation—data with low ($\rho = .25$), moderate ($\rho = .50$) and high ($\rho = .75$) polyserial correlations were generated;

(iv) Number of thresholds—polychotomous variables with $r = 3$ and $r = 7$ categories were simulated.

Fifty replications in each of the 24 cells of the design were generated.

Results

The bias of the estimators can be studied by examining mean estimates of the polyserial correlation in each cell of the design. These means are presented in Table 2. Mean point polyserial correlations are also presented in Table 2 because the point polyserial correlation is frequently used by practitioners as if it were an estimate of the polyserial correlation.

There is very little bias evident in Table 2 for the ad hoc, ML and two-step estimators. The mean estimates for the ML and two-step estimators are virtually identical, and tend to

TABLE 2

Means and Root Mean Squared Errors of Estimates of the Polyserial Correlation

			Point			
ρ	r	γ	Polyserial	Ad Hoc	ML	Two-Step
N=100						
.25	3	0	.222(.0781)	.249(.0817)	.247(.0810)	.248(.0813)
	3	1	.207(.1060)	.253(.1181)	.252(.1169)	.253(.1172)
	7	0	.250(.1034)	.260(.1078)	.260(.1072)	.260(.1073)
	7	1	.206(.1057)	.237(.1122)	.237(.1100)	.237(.1103)
.50	3	0	.452(.0814)	.505(.0728)	.503(.0709)	.504(.0710)
	3	1	.399(.1357)	.491(.1122)	.490(.1087)	.490(.1083)
	7	0	.485(.0796)	.506(.0815)	.502(.0815)	.503(.0812)
	7	1	.435(.0971)	.500(.0835)	.497(.0777)	.498(.0777)
.75	3	0	.680(.0897)	.763(.0634)	.756(.0611)	.757(.0609)
	3	1	.619(.1392)	.755(.0552)	.755(.0551)	.755(.0546)
	7	0	.727(.0448)	.758(.0395)	.754(.0391)	.753(.0387)
	7	1	.650(.1034)	.759(.0573)	.756(.0552)	.755(.0543)
N=500						
.25	3	0	.224(.0502)	.250(.0479)	.249(.0478)	.250(.0478)
	3	1	.207(.0625)	.253(.0565)	.252(.0562)	.252(.0563)
	7	0	.232(.0453)	.240(.0443)	.240(.0442)	.240(.0442)
	7	1	.217(.0546)	.248(.0493)	.248(.0500)	.248(.0500)
.50	3	0	.453(.0581)	.505(.0382)	.504(.0380)	.504(.0380)
	3	1	.404(.1021)	.493(.0417)	.493(.0385)	.493(.0385)
	7	0	.485(.0365)	.501(.0342)	.501(.0341)	.501(.0340)
	7	1	.440(.0730)	.502(.0477)	.501(.0466)	.501(.0466)
.75	3	0	.673(.0793)	.751(.0229)	.750(.0230)	.750(.0228)
	3	1	.617(.1355)	.754(.0285)	.751(.0245)	.752(.0246)
	7	0	.729(.0271)	.753(.0177)	.752(.0175)	.752(.0173)
	7	1	.659(.0931)	.751(.0228)	.749(.0209)	.749(.0212)

Note: All means are based on 50 replications. Root mean squared errors are given within parentheses.

be quite close to the means for the ad hoc estimator. The ad hoc estimator does not appear substantially more biased than the ML estimator.

Table 2 illustrates the substantial bias that can result if point polyserial correlations are used as estimates of polyserial correlations. The bias increases as the number of categories decreases, as skewness increases and as the polyserial correlation increases. Consideration of (12) explains all three of these results.

The accuracy of an estimator $\hat{\rho}$ of ρ can also be examined by considering the root mean squared error,

$$\text{RMSE} = \left(\frac{1}{50}\sum_{i=1}^{50}(\hat{\rho}_i - \rho)^2\right)^{1/2}. \tag{39}$$

RMSEs for the various estimators are presented within parentheses in Table 2.

When studying the RMSEs in Table 2, a number of general trends are evident. Note that increasing the sample size decreases the RMSE, increasing the number of categories usually decreases the RMSE, increasing the skew tends to increase the RMSE, and increasing ρ generally decreases the RMSE.

The ML and two-step estimators yield nearly equivalent RMSEs in Table 2. Of course we would expect to see more accurate estimation by the ML estimator if the number of replications was increased to a sufficiently large number. However, Table 2 shows that the two-step estimator, which requires less computer time to calculate, is almost as accurate as the asymptotically optimal ML method. This is particularly evident in samples of $N = 500$, where the differences in RMSEs for the ML and two-step methods are all less than .0003.

The ad hoc estimator performs surprisingly well: its RMSE values exceed the ML estimator's RMSEs by less than .0058 when $N = 100$ and by less than .0040 when $N = 500$. The point polyserial correlation has reasonably small RMSE values when $\rho = .25$, over the range of values of r and γ that we studied. The small RMSEs occur despite the bias seen in Table 2 because the sampling variance of the point polyserial is usually smaller than the sampling variances of the other estimators.* As ρ increases, however, the bias of the point polyserial increases and the sampling variance of the other estimator decreases. When $\rho = .75$, $N = 500$, $r = 3$ and $\gamma = 1$, the RMSE for the point polyserial is substantially larger than RMSEs for the other estimators.

Discussion

The simulation described in the preceding section indicates that the ML, two-step, and ad hoc estimators all have relatively small biases. This result was expected for the maximum likelihood methods, but was not completely expected for the ad hoc method. At least with respect to the ad hoc estimator, generalizations beyond the scope of our simulation parameters are not warranted. The RMSEs of the estimators were surprising only in that the two-step and ad hoc estimators performed nearly as well as the ML estimator.

The results presented here, in conjunction with Olsson's [1979] Table 8 indicate that product moment correlations computed from categorical variables can be seriously misleading. This is particularly evident when rating scale items, with varying numbers of categories and varying degrees of skewness, are used to assess substantively important

* Sampling variances can be obtained using the relation

$$s^2 = \text{RMSE}^2 - (\bar{\rho} - \rho)^2,$$

where RMSE is a root mean squared error and $\bar{\rho}$ is the corresponding mean estimate of the polyserial correlation in Table 2.

constructs. Interitem correlations and item-total correlations can be seriously distorted by discrete measurement of latent continuous variables.

The ad hoc estimator appears to be sufficiently unbiased and accurate for applied research. Because Olsson's [1979] methods for estimating the polychoric correlation also appear sufficiently accurate for practical work, there are now estimators for the various types of "inferred" correlations in Table 1. These estimators should be quite useful to substantive researchers who wish to construct scales. For example, a correlation matrix for factor analysis may be constructed by computing product moment correlations between continuous variables, polychoric correlations between categorical variables and polyserial correlations between continuous-categorical variable combinations. Such a correlation matrix may not be Gramian, but might reduce problems caused by difficulty factors. The most recent version of the LISREL [Jöreskog & Sörbom, 1982] computer program contains options that allow calculations of the polyserial and polychoric correlations.

REFERENCE NOTES

1. Nerlove, M. & Press, S. J. Univariate and multivariate log-linear and logistic models. Santa Monica, The Rand Corporation, R: 1306-EDA/NIH, 1973.
2. Gruvaeus, G. T. & Jöreskog, K. G. *A computer program for minimizing a function of several variables* (E.T.S. Res. Bull. RB70-14). Princeton, NJ: Educational Testing Service, 1970.

REFERENCES

Cox, N. R. Estimation of the correlation between a continuous and a discrete variable. *Biometrics*, 1974, *30*, 171–178.

Finney, D. J. *Probit analysis*. Cambridge: Cambridge University Press, 1971.

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

Jaspen, N. Serial correlation. *Psychometrika*, 1946, *11*, 23–30.

Jöreskog, K. G., & Sörbom, D. *Lisrel V user's guide*. Chicago: National Educational Resources, 1981.

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

Lazarsfeld, P. F. Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science*, Vol. 3. New York: McGraw-Hill, 1959.

Lord, F. M., & Novick, M. R. *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley, 1968.

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.

Mosteller, E. Nonsampling errors. In W. H. Kruskal & J. M. Tanur (Eds): *International encyclopedia of statistics*. New York: The Free Press, 1968.

NAG Fortran Library Manual, Mark 7, Oxford, NAG Ltd, 1979.

Olsson, U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 1979, *44*, 443–460.

Pearson, K. On a new method for determining the correlation between a measured character A and a character B. *Biometrika*, 1909, *7*, 96.

Pearson, K. On the measurement of the influence of "broad categories" on correlation. *Biometrika*, 1913, *9*, 116–139.

Silvey, S. D. *Statistical inference*. Harmondsworth: Penguin, 1970.

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

Tate, R. F. The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*, 1955, *42*, 205–216.(a)

Tate, R. F. Applications of correlation models for biserial data. *Journal of the American Statistical Association*, 1955, *50*, 1078–1095.(b)

Manuscript received 7/10/81
Final version received 4/19/82