AD NUMBER

AD029741

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 02 APR 1954. Other requests shall be referred to Office of Naval Research, Arlington, VA 22203.

AUTHORITY

ONR ltr dtd 26 Oct 1977

AD 29741

THE THEORY OF CORRELATION BETWEEN TWO CONTINUOUS VARIABLES WHEN ONE IS DICHOTOMIZED

By

Robert F. Tate

University of Washington

Technical Report No. 14

April 2, 1954

Contract N8onr-520 Task Order II
Project Number NR-042-038

Laboratory of Statistical Research
Department of Mathematics
University of Washington
Seattle, Washington

1. Introduction

The problem of biserial correlation arises when one is sampling from a bivariate normal population in which one of the variables has been dichotomized, giving rise to only two observable values, say 0 and 1, and one wishes to use this dichotomized sample to estimate, or test hypotheses concerning the correlation coefficient ρ of the original bivariate normal distribution. ρ is sometimes given the name biserial correlation coefficient. This name reflects the former confusion between sample statistics and population parameters, referring of course to the fact that a sample drawn from the observable bivariate population just described may be thought of as two separate series of observations, those in which the dichotomized variable has the value 0 and those in which it has the value 1. It is apparent that the numbers of observations in the two series are dependent binomial random variables whose sum is the sample size.

The term biserial correlation was introduced in 1909 by Karl Pearson [6], who was the first to perceive the statistical importance of this particular type of problem. He proposed as an estimator the sample biserial correlation coefficient. The asymptotic variance of this estimator was derived in 1913 by Soper [8]. Since that time, with the exception of certain references to discriminatory analysis (see [13]), in which the use of the sample biserial correlation coefficient goes unquestioned, no results of a mathematical nature were contributed until a recent paper by Maritz [5]. Much literature exists, however, on the subject of how best to compute Pearson's coefficient. In this connection the reader should see DuBois [2], Dunlap [3], and Royer [7].

The problem of biserial correlation occurs quite often in psychological work, especially in that branch of the subject known as Test Construction and Validation. In connection with an objective test one may be interested in obtaining a measure of the strength of the relation between the ability to answer correctly a particular item in the test and ability to perform at some task, or in testing hypotheses concerning the strength of this relation. Such a measure, together with a random sample of individuals, selected of course from the population for which inferences are to be drawn, helps to determine whether or not the given item should be included in the test. In order to put the problem in the proper form certain assumptions must be made. Suppose that the ability to answer the test item correctly can be represented numerically by a random variable with a normal probability distribution, which however cannot be observed due to the restrictions of the test; in particular, suppose that observations on this ability take the value 0 if the question is answered incorrectly, and the value 1 if the question is answered correctly. If in addition to the underlying normal distribution just postulated, we assume that the ability to perform at the task is also measurable with a normal distribution, and that the two normal variables have correlation coefficient ρ, then we have the problem of biserial correlation.[1] A simple example would occur when a true-false question is included among the questions in a preliminary college aptitude test and then a follow-up study is conducted on the sampled students in order to observe their final grade point average upon graduation four years later. In this case ρ would represent the degree of association between ability to answer the question correctly and ability in school, and would be estimated by some function of the paired observations on the sample students. It should be noted that in such a case acceptable test items are those for which $|\rho|$ is judged to be near 1. We shall refer to this example several times in the sequel in order to illustrate certain points.

[1]Pearson's original formulation was less restrictive than that given here. Anticipating the fact that problems in estimation and testing hypotheses will require assumptions of normality, we make these assumptions at the outset.

In the above example it is assumed that there exists an underlying distribution of ability to answer the true-false question correctly, and that in addition this distribution is normal. Since it is not possible to observe the underlying distribution, it also is not possible to test the assumption of underlying normality, but only that the underlying distribution is normal given that there is an underlying continuous distribution. Therefore, in many situations in which biserial correlation is appealed to, the assumptions involved are open to serious attack, an attack for which there is no adequate statistical defense.

If one wishes, he may give up the assumption of underlying normality and assume instead that the observed bivariate distribution is that of a discrete random variable which takes the values 0 and 1, and a continuous random variable with the property that its conditional distribution be normal for each given value of the discrete variable. With this formulation good results may be obtained if one is willing to make the assumption that the two normal conditional distributions have the same variance. The problem of estimating or testing hypotheses concerning ρ under this set of conditions is known as the problem of point-biserial correlation and is treated in reference [10]. Some of the results obtained there are similar to those of the present paper. Note that this set of conditions may be tested statistically by testing the normality of the conditional distributions. Moreover, no confusion between the two models would occur, since in the case of biserial correlation, it is easily shown to be impossible for the conditional distributions to be normal. In the problem of estimating the correlation between ability to answer a test item correctly and ability in school, the use of point-biserial correlation requires that (i) for each student who answers the test item correctly, the conditional distribution of his grade-point average be normal with, say, mean μ_1, and variance σ^2, and that (ii) for each student who answers the test item incorrectly, the conditional distribution of his grade-point average be normal with mean μ_0 and variance σ^2. Note that it is necessary for the variability of grade-point average to be the same for the two groups of students.

Professor Harold Hotelling realised some years ago that the existing methods for dealing with the problem of biserial correlation were far from satisfactory, and suggested to the author that the whole situation be reconsidered. The results of this examination are contained in the present paper.

Section 2 contains a list of most of the notation which has been adopted, and Section 3 deals with the mathematical model.

In Section 4 the question of maximum likelihood is treated. The maximum likelihood estimator is shown to be asymptotically normal and asymptotically efficient. The asymptotic variances for $\hat{\rho}$ and $\hat{\omega}$, the maximum likelihood estimators of the correlation ρ and point of dichotomy ω, are found by the usual method which employs the information matrix and side-steps the solution of the likelihood equations. A valuable contribution to the theory of biserial correlation was made by Maritz [5]. Comments are made on his work, and on a paper of Tocher [11], in the early part of Section 4.

An evaluation of r^*, the sample biserial correlation coefficient, is given in detail in Section 5. It is shown that r^* has asymptotic efficiency for estimating ρ which is 1 when $\rho = 0$, but which approaches 0 when $|\rho|$ approaches 1. Consistency of r^* was shown by Karl Pearson [6]. The well-known fact that r^* may be greater than 1 is pointed out and some notion of the magnitude of r^* is obtained by a consideration of the product moment correlation coefficient r. Asymptotic normality of r^* is verified by the use of a theorem of Cramér. The asymptotic standard deviation of r^* is tabulated in Table I at the end of the paper. One interesting point in

Section 5 is an intuitively appealing fact which the author discovered is universally assumed, but apparently was never proved: namely, that the asymptotic variance of r^* is a minimum at $\omega = 0$ for each fixed ρ. A proof is given for this result. For the case $\omega = 0$ an approximate variance stabilising transformation is derived. Calculations pertaining to this transformation may be carried out by using Table VB of Fisher [4] for the function $\tanh^{-1}(r)$. This result should prove useful in many situations. Section 5 concludes with a discussion of the preceding results and of the feasibility of using r^* to test the hypothesis H: $\rho = \rho_0$ when $|\rho_0|$ is small.

Section 6 is devoted to a discussion of an iterative method of solution for the likelihood equations. The method is essentially Newton's method for two variables, the calculated values ω^*, r^* being used to start the iteration. The computations are not really prohibitive, considering the importance of the problem, and are to a certain extent organizable for punched-cards methods. An example is given with all of the calculations illustrated. The data consist of a sample of 20 observations taken from an artificially constructed bivariate normal population with $\rho = .707$.

Values of Mills' ratio,

$$\phi(x) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{\int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy}$$

are required for the calculations. For purposes here we shall need a table which gives $\phi(x)$ for x ranging from -3 to $+3$ in steps of .01, so that no transformation or extensive interpolation is required, since it will be necessary to obtain n values of ϕ for a given problem in which a sample of n has been taken. Accordingly, we include Table II at the end of the paper. Interpolation by inspection in Table II should be quite satisfactory.

The subject of biserial correlation is generally given a light treatment in texts on psychological statistics, centering on the unboundedness of r^* and the questionable character of the assumption of underlying normality for the dichotomized variable. A notable exception is the recent book by Walker and Lev [12], which has a more complete treatment of the subject. A few of the results contained in the present paper are referred to, and illustrated in [12].

2. Notation

To eliminate the distraction of searching through the text, we shall list here most of the symbols and notational devices used.

$$\psi(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy - y^2)}$$, the bivariate normal density

$$\lambda(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$, the normal density

$$p(x) = \int_x^\infty \lambda(t)\,dt, \qquad q(x) = 1 - p(x).$$

$$\phi(x) = \frac{\lambda(x)}{p(x)}$$, Mills' ratio.

$$\xi(x,\omega) = \int_\omega^\infty \psi(x,y)\,dy$$

$$\eta(x,\omega) = \int_{-\infty}^\omega \psi(x,y)\,dy$$

X — the undichotomized normal random variable.

Y — the dichotomized normal random variable.

ω — the point of dichotomy of Y, measured in standard units.

Z — the discrete random variable induced by the dichotomization of Y.

$f(x,z)$	the joint density of the random variables X and Z.
$\rho(X,Y)$	the correlation coefficient of the random variables X and Y.
$\hat{\rho}$	the maximum likelihood estimator of ρ.
r^*	the sample biserial correlation coefficient.
r	the ordinary sample correlation coefficient based on the sample (X_i, Z_i), $i = 1,2,\ldots,n$.
$V(r^* \mid \omega, \rho)$	the asymptotic variance of r^*.
$\mathcal{E}(r^* \mid \omega, \rho)$	the asymptotic efficiency of r^* for estimating ρ.
$\mathcal{N}(\mu, \sigma^2)$	a normal random variable with mean μ and variance σ^2.
$\mathcal{BN}(\mu, \nu; \sigma^2, \tau^2; \rho)$	a bivariate normal random vector with means μ, ν, variances σ^2, τ^2, and correlation ρ.
$U_n \sim \mathcal{N}(\mu, \sigma^2)$	denotes the fact that U_n is asymptotically normal with mean μ and variance σ^2.

3. Mathematical Model

Let $(X,Y) = \mathcal{BN}(\mu, \nu; \sigma^2, \tau^2; \rho)$, and ω be any fixed constant. Now let Z be a Bernoulli random variable defined as follows:

(3.1) $\quad Z = 1$ if $\frac{Y-\nu}{\tau} \geq \omega$ and $Z = 0$ if $\frac{Y-\nu}{\tau} < \omega$.

Obviously,

$$P(Z = 1) = \int_{\frac{\omega-\nu}{\tau}}^{\infty} \lambda\left(\frac{y-\nu}{\tau}\right) dy = p(\omega), \quad P(Z = 0) = q(\omega).$$

Consider a sample of n independent random vectors $(X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n)$. The problem of biserial correlation consists in finding a suitable function of (X_i, Z_i) $i = 1,2,\ldots,n$, with which to estimate ρ.

Karl Pearson [6] introduced the estimator r^* ("biserial r"), which we express in the following form:

$$(3.2)\quad r^* = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Z_i-\bar{Z})}{\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right\}^{\frac{1}{2}}\lambda(T)} = r\,\frac{\left\{\frac{1}{n}\sum_{i=1}^{n}(Z_i-\bar{Z})^2\right\}^{\frac{1}{2}}}{\lambda(T)},$$

where r is the product moment correlation coefficient of (X_i, Z_i), and T is the solution of the equation

$$(3.3)\quad \int_T^{\infty}\lambda(y)\,dy = \bar{Z}$$

r^* will be discussed completely in Section 5. For the present we shall merely state the asymptotic variance obtained by Soper [8]:

$$(3.4)\quad V(r^*|\omega,\rho) = \frac{1}{n}\left\{\rho^4 + \rho^2\left[\frac{pq\omega^2}{\lambda^2} + \frac{(2p-1)\omega}{\lambda} - \frac{5}{2}\right] + \frac{pq}{\lambda^2}\right\},$$

where the functions p, q, and λ all have argument ω.

$$\sqrt{n}\cdot\sigma(r^*) = \left\{n\,V(r^*|\omega,\rho)\right\}^{\frac{1}{2}}$$

is given in Table I at the end of the paper. In view of synetry about the values $\rho = 0$ and $p(\omega) = \frac{1}{2}$, the tabulation is given for $\rho = 0$ to 1 in the steps of .10, and for $p = .05$ to .50 in steps of .05.

Since the random variable Z takes the value 0 or 1, the joint density of (X,Z) can be written

$$(3.5)\quad f(x,z) = zf(x,1) + (1-z)f(x,0),$$

where

$$(3.6)\quad f(x,0) = \int_{-\infty}^{\omega\tau+\nu}\Psi(x,y)\,dy,\quad f(x,1) = \int_{\omega\tau+\nu}^{\infty}\Psi(x,y)\,dy,$$

with $\Psi(x,y)$ denoting the density of $BN(\mu,\nu;\sigma^2,\tau^2;\rho)$. Sections 4 and 6 are devoted to a discussion of the likelihood function,

$$(3.7)\quad L = \prod_{i=1}^{n}\left\{(1-z_i)f(x_i,0) + z_i f(x_i,1)\right\}.$$

4. Properties of the Maximum Likelihood Estimators

It may be seen that L is actually independent of ν and τ, since a change of variable $y' = (y-\nu)/\tau$ in the integrals of $f(x,0)$ and $f(x,1)$ removes ν and τ. Hence, in all further work we shall set $\nu = 0$ and $\tau = 1$.

The main stumbling block is the existence of the nuisance parameters μ and σ^2. The 4 likelihood equations in the variables $\omega, \rho, \mu, \sigma^2$ are not hard to write down, but the algebraic difficulties involved in the deviation of asymptotic variances and covariances, and the numerical difficulties involved in solving the equations by an iterative method, prove far too prohibitive. It seems intuitively clear that were we able to solve the 4 likelihood equations for $\hat{\omega}, \hat{\rho}, \hat{\mu}, \hat{\sigma}^2$, we would find that $\hat{\rho} = \hat{\rho}(x_1,x_2,\ldots,x_n; z_1,z_2,\ldots,z_n)$ is invariant under any transformation of the form $x_i' = ax_i+b$. This, of course, does not give us the right to set $\mu = 0$ and $\sigma^2 = 1$, and then expect to find the correct maximum likelihood estimators for μ and σ^2. The following course of action has been adopted as a way out: set $\mu = 0$ and $\sigma^2 = 1$. Then solve the 2 likelihood equations, and in the resulting solution replace x_i by

$$\bar{x} - x_i \left\{ \frac{\sum (x_i - \bar{x})^2}{n-1} \right\}^{\frac{1}{2}} \tag{4.1}$$

If we know the values of μ and σ^2, the problem is naturally avoided by an immediate transformation of the original data. In all future work we shall assume $\mu = 0$ and $\sigma^2 = 1$. $\Psi(x,y)$ now becomes $\Psi(x,y)$, and the likelihood function takes the form

$$L(\omega,\rho) = \prod_{i=1}^{n} \left\{ z_i \, \xi(x_i,\omega) + (1-z_i)\, \eta(x_i,\omega) \right\}. \tag{4.2}$$

We shall pause at this point to discuss the work of Maritz [5]. Using Probit Analysis, together with a result of Tocher [11], he has given a very nice approximation scheme for the solution of the likelihood equations. In what follows we shall give a short outline of his method in terms of the notation of the present paper.

Let $(X,Y) = \mathcal{B}\mathcal{N}(0,0;1,(1-\rho^2)^{-1};\rho)$. In view of the fact that the likelihood equations are invariant under a change of ν or τ^2, this formulation is equivalent to ours. Now introduce a grouping of the observations $\{X_i; i = 1,2,\ldots,n\}$ with a set of k cells of equal width. Denote this grouping by

$$\{I_{j,k}; j = 1,2,\ldots,k\}.$$

Denote the collection of midpoints of these cells by

$$\{\xi_{j,k}; j = 1,2,\ldots,k\}.$$

Let

$$(4.3) \qquad P_{j,k} = P(X \in I_{j,k}), \quad \pi_{j,k} = P(Z = 1 | X = \xi_{j,k})$$

Now let $N_{j,k}$ be the number of observations in the sample $\{X_i; i = 1,2,\ldots,n\}$ which fall in $I_{j,k}$, and $M_{j,k}$ be the number among the $N_{j,k}$ for which the corresponding Z observation is 1. Thus, $\{N_{j,k}; j = 1,2,\ldots,k\}$ have a multinomial distribution with parameters $\{P_{j,k}; j = 1,2,\ldots,k\}$. The conditional distribution of $\{M_{j,k}; j = 1,2,\ldots,k\}$ given $N_{j,k} = n_{j,k}$ is the product of k binomial distributions.

$$(4.4) \qquad \sum_{j=1}^{k} N_{j,k} = n, \quad \sum_{j=1}^{k} P_{j,k} = 1, \quad (k = 1,2,\ldots)$$

Maritz now assumes that the observations $\{X_i; i = 1,2,\ldots,n\}$ are concentrated at their respective cell mid points. Since the marginal distribution of the X_i is independent of ω and ρ, the part of the likelihood function which depends on ρ and ω will be the conditional distribution of the $M_{j,k}$ given $N_{j,k} = n_{j,k}$, which has parameters $\pi_{j,k}$. We are thus ultimately led to two simultaneous equations for $\hat{\omega}$ and $\hat{\rho}$ which contain terms

$$\frac{\partial \pi_{j,k}}{\partial \rho} , \qquad \frac{\partial \pi_{j,k}}{\partial \rho} .$$

These equations are then transformed slightly and probit analysis is used for the solution.

Presumably, as the grouping becomes finer, the estimates $\hat{\omega}(m_{1,1}, m_{2,k}, \ldots, m_{k,k})$ and $\hat{\rho}(m_{1,k}, m_{2,k}, \ldots, m_{k,k})$, together with the asymptotic variances $\sigma^2_{\hat{\omega}}$ and $\sigma^2_{\hat{\rho}}$, will approach the correct values for the original problem. A proof of this result must depend on a close examination of the limiting processes involved. The situation which arises may be described as follows. We assume that the grouping becomes finer and we wish to assert two things:

$$(4.5)\qquad \begin{aligned} &\{\pi_{j,k}; j = 1,2,\ldots,k\} \rightarrow \{\pi_i = P(Z = 1 \mid X = x_i); i = 1,2,\ldots,n\} , \\ &\left\{\frac{M_{j,k}}{N_{j,k}}; j = 1,2,\ldots,k\right\} \rightarrow \{\pi_{j,k}; j = 1,2,\ldots,k\} \end{aligned}$$

in the sense of probability.

The meaning of (4.5) is, then, that as k and n both become large the cell width must become small, but in such a way that each cell still contains sufficiently many observations for $\pi_{j,k}$ to be a valid approximation of $M_{j,k}/N_{j,k}$. This result does indeed appear quite plausible, but a detailed proof would be lengthy.

Instead of attempting a discussion of the above point, we offer an alternative derivation of the asymptotic variances of $\hat{\rho}$ and $\hat{\omega}$ in this section, and in Section 6 and iterative scheme for obtaining $(\hat{\rho}, \hat{\omega})$ which, while more time consuming than that of Maritz, does not require any grouping. It should be noted here that Tocher's exact method (pp. 9-11, [11]), also known as the "scoring" method, doesn't help in this case, owing to the difficulty in obtaining the expectations of the second partial derivatives of L. Again, the plausibility of Maritz' scheme should be emphasized.

Many results have been obtained concerning the asymptotic normality and asymptotic efficiency of maximum likelihood estimators. In each case the parametric family of probability distributions is subject to certain regularity conditions. The density of (X,Z) is f(x,z). We shall not dwell here at any length on the regularity conditions, but shall merely remark that the regularity conditions given by Cramér (Chap. 33.3, [1]) may be easily verified, since f(x,0) and f(x,1) are both integrals of bivariate normal densities. Consequently, $\hat{\omega},\hat{\rho}$ will be asymptotically normal, and asymptotically efficient estimators for ω and ρ respectively. Asymptotic variances of the maximum likelihood estimators may be found by an inversion of the inverse matrix without actually solving the likelihood equations. We now use this technique.

Theorem I.

The asymptotic variance of $\hat{\rho}$ is given by

$$(4.6) \quad V(\hat{\rho}\,|\,\omega,\rho) = \frac{1}{n}(1-\rho^2)^3 \cdot \left\{ \int_{-\infty}^{+\infty} x^2 g(x,\omega,\rho)\,dx - \frac{\left(\int_{-\infty}^{+\infty} x g(x,\omega,\rho)\,dx\right)^2}{\int_{-\infty}^{+\infty} g(x,\omega,\rho)\,dx} \right\}^{-1},$$

where

$$g(x,\omega,\rho) = \lambda(x)\, \phi\left(\frac{\omega-\rho x}{\sqrt{1-\rho^2}}\right) \phi\left(-\frac{\omega-\rho x}{\sqrt{1-\rho^2}}\right)$$

Proof:

$$\log L = \sum_{i=1}^{n} \log\left\{ z_i\, \xi(x_i,\omega) + (1-z_i)\, \eta(x_i,\omega) \right\}.$$

We will need the quantities

$$E\left(\frac{\partial^2 \log L}{\partial \omega^2}\right),\quad E\left(\frac{\partial^2 \log L}{\partial \omega\, \partial \rho}\right),\quad E\left(\frac{\partial^2 \log L}{\partial \rho^2}\right).$$

Letting δ^2 refer to any of the three second order partial operators, we have the fundamental relation

$$(4.7)\quad E\{\delta^2 \log L\} = nqE_0\{\delta^2 \log \eta(X,\omega)\} + npE_1\{\delta^2 \log \xi(X,\omega)\},$$

where E_0 means expectation with respect to the conditional density of X given $Y < \omega$, and E_1 means expectation with respect to the conditional density of X given $Y \geq \omega$. The conditional densities are

$$(4.8)\qquad \psi(x|Y < \omega) = \frac{1}{q}\int_{-\infty}^{\omega} \psi(x,y)\,dy = \frac{1}{q}\eta(x,\omega)$$

$$(4.9)\qquad \psi(x|Y \geq \omega) = \frac{1}{p}\int_{\omega}^{\infty} \psi(x,y)\,dy = \frac{1}{p}\xi(x,\omega).$$

For each of the possible operators δ^2 the calculation of (4.7) procedes in about the same way. As an illustration, we shall compute

$$E\left\{\frac{\partial^2 \log L}{\partial \omega^2}\right\}.$$

For a random variable U with density $h(u;\theta)$, it is well known that

$$(4.10)\qquad E\left\{\frac{\partial^2 \log h(U;\theta)}{\partial \theta^2}\right\} = -E\left\{\left(\frac{1}{h(U;\theta)}\right)^2 \left(\frac{\partial h(U;\theta)}{\partial \theta}\right)^2\right\},$$

provided the expectations exist and differentiation twice under the expectation sign is permissible. It is easily seen that

$$(4.11)\qquad \frac{\partial \eta(x,\omega)}{\partial \omega} = \psi(x,\omega).$$

Consider the first term in the right member of (4.7). From the definition of E_0 and (4.8), together with (4.10), we have the result

$$(4.12)\quad nqE_0\left\{\frac{\partial^2 \log \eta(X,\omega)}{\partial \omega^2}\right\} = -n\int_{-\infty}^{+\infty} \frac{\{\psi(x,\omega)\}^2}{\eta(x,\omega)}\,dx.$$

In a similar manner we show that

$$(4.13)\qquad np\,E_1\left\{\frac{\partial^2 \log \xi(X,\omega)}{\partial \omega^2}\right\} = -n\int_{-\infty}^{+\infty} \frac{\{\psi(x,\omega)\}^2}{\xi(x,\omega)}\,dx.$$

Hence, combining (4.12) and (4.13), we get

$$(4.14)\qquad E\left\{\frac{\partial^2 \log L}{\partial \omega^2}\right\} = -n\int_{-\infty}^{+\infty} \frac{\lambda(x)\,\{\psi(x,\omega)\}^2}{\xi(x,\omega)\,\eta(x,\omega)}\,dx.$$

Using the relations,

$$\begin{cases} \{\psi(x,\omega)\}^2 = \{\lambda(x)\}^2\left\{\lambda\left(\frac{\omega-\rho x}{1-\rho^2}\right)\right\}^2 \dfrac{1}{1-\rho^2}, \\ \xi(x,\omega) = \lambda(x)\displaystyle\int_{\frac{\omega-\rho x}{\sqrt{1-\rho^2}}}^{\infty} \lambda(y)\,dy, \\ \eta(x,\omega) = \lambda(x)\displaystyle\int_{-\infty}^{\frac{\omega-\rho x}{\sqrt{1-\rho^2}}} \lambda(y)\,dy, \end{cases}$$

we have from (4.14) and the definition of $g(x,\omega,\rho)$

$$(4.15)\qquad E\left\{\frac{\partial^2 \log L}{\partial \omega^2}\right\} = -\frac{n}{(1-\rho^2)}\int_{-\infty}^{+\infty} g(x,\omega,\rho)\,dx$$

similarly,

$$(4.16)\qquad E\left\{\frac{\partial^2 \log L}{\partial \rho^2}\right\} = -\frac{n}{(1-\rho)^3}\int_{-\infty}^{+\infty} (x-\rho\omega)^2\, g(x,\omega,\rho)\,dx,$$

$$(4.17)\qquad E\left\{\frac{\partial^2 \log L}{\partial \omega\,\partial \rho}\right\} = \frac{n}{(1-\rho^2)^2}\int_{-\infty}^{+\infty} (x-\rho\omega)\, g(x,\omega,\rho)\,dx.$$

Forming the 2 x 2 information matrix

$$\Lambda = \begin{bmatrix} E\,\dfrac{\partial^2 \log L}{\partial \omega^2} & E\,\dfrac{\partial^2 \log L}{\partial \omega\,\partial \rho} \\ E\,\dfrac{\partial^2 \log L}{\partial \rho\,\partial \omega} & E\,\dfrac{\partial^2 \log L}{\partial \rho^2} \end{bmatrix},$$

and observing that $|\Lambda|$ is non-vanishing because of the Schwarz inequality, we finally obtain

$$(4.18)\begin{cases} V(\hat{\rho}\,|\omega,\rho) = -\lambda_{22}^{-1} = \dfrac{(1-\rho^2)^3}{n}\left\{\displaystyle\int_{-\infty}^{+\infty} x^2 g\,dx - \dfrac{\left(\int_{-\infty}^{+\infty} xg\,dx\right)^2}{\int_{-\infty}^{+\infty} g\,dx}\right\}^{-1}, \\ V(\hat{\omega}\,|\omega,\rho) = -\lambda_{11}^{-1} = \dfrac{(1-\rho^2)}{n}\left\{\displaystyle\int_{-\infty}^{+\infty} g\,dx - \dfrac{\left(\int_{-\infty}^{+\infty} xg\,dx\right)^2}{\int_{-\infty}^{+\infty} x^2 g\,dx}\right\}^{-1}, \end{cases}$$

where λ_{ij}^{-1} is the ij^{th} element of the inverse matrix Λ^{-1}, which proves Theorem I . Expressions (4.18) coincide with the previously mentioned results of Maritz.

5. The Utility of r^*

We shall now present a series of results concerning r^*, which will be followed by a general discussion of its value.

A little later we will need $E(XZ)$. Since, it is not difficult to obtain, we will give the expression for the general moment $\alpha_K = E(X^K Z)$.

Theorem II:

$$\alpha_K = \sum_{j=0}^{K} \binom{K}{j} (1-\rho^2)^{\frac{j}{2}} \rho^{K-j} a_j \int_{\omega}^{\infty} y^{K-j} e^{-\frac{y^2}{2}} dy,$$

where a_j is the j^{th} moment of the random variable $\mathcal{N}(0,1)$.

Proof: Using the definition of E_1 and (4.9), we obtain

$$E(X^K Z) := pE_1(X^K) = \int_{-\infty}^{+\infty}\int_{\omega}^{\infty} x^K \psi(x,y)\,dy\,dx.$$

Make the transformation $t = (x-\rho y)\sqrt{1-\rho^2}$. The above then reduces to

$$\int_{-\infty}^{+\infty}\int_{\omega}^{\infty} (t\sqrt{1-\rho^2} + y\rho)^K \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy\,dt.$$

Using a binomial expansion and integrating with respect to t, we have Theorem II. The integrals contained in Theorem II may be evaluated by a recursion relation. Let

$$b_\nu(\omega) = \int_\omega^\infty \frac{y^\nu}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Then,

$$b_\nu(\omega) = (\nu - 1)\, b_{\nu-2}(\omega) + \omega^{\nu-1} b_1(\omega).$$

We now easily arrive at

$$(5.1) \qquad b_0(\omega) = p(\omega), \; b_1(\omega) = \lambda(\omega),$$

whence

$$(5.2) \qquad \alpha_0 = p(\omega), \; \alpha_1 = \rho\lambda(\omega), \; \alpha_2 = p(\omega) + \omega\lambda(\omega)\rho^2.$$

The relation between $\rho(X,Y)$ and $\rho(X,Z)$ is given by

<u>Theorem III</u>: $\rho(X,Z) = \rho(X,Y) \dfrac{\lambda(\omega)}{\sqrt{pq}}$

Proof: Let $EX = \mu = 0$ and $V(X) = \sigma^2 = 1$. Then, $\rho(X,Z) = (EXZ)\,[V(Z)]^{-\frac{1}{2}} = (EXZ)\,(pq)^{-\frac{1}{2}}$. From (5.2), $\alpha_1 = \rho\,\lambda(\omega)$, which proves the theorem.

It follows from the original definition of biserial correlation, as given by Pearson [6], that r^* is consistent. This fact is also an immediate consequence of relation (3.2) between r^* and r: $r \longrightarrow \rho(X,Z)$ in probability as $n \longrightarrow \infty$. Thus,

$$r^* = \frac{r}{\lambda(T)} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Z_i - \bar{Z})^2 \right\}^{\frac{1}{2}} \longrightarrow \frac{\rho(X,Z)\sqrt{pq}}{\lambda(\omega)} \quad \text{in probability}$$

and hence by Theorem III, $r^* \longrightarrow \rho(X,Y)$ in probability.

With respect to the magnitude of r^*, it is well known that $|r^*|$ can be greater than 1. Something of the nature of this phenomenon can be understood by looking at r. In order to prove a result concerning the magnitude of of r^*, we shall need a result from reference [9].

<u>Theorem IV</u> (Lemma 2 of [9]):

$$p(x)\ q(x) \geq \frac{\pi}{2}\ \lambda^2(x), \qquad (-\infty < x < +\infty)$$

Now we have

<u>Theorem V</u>:

$$\frac{r^*}{r} \geq \sqrt{\frac{\pi}{2}}$$

<u>Proof</u>: Rewriting (3.2) as

$$r^* = r\ \frac{\sqrt{\bar{Z} - \bar{Z}^2}}{\lambda(T)} \quad ,$$

we have, in view of the definition of T,

$$r^* = r\ \frac{\sqrt{p(T)q(T)}}{\lambda(T)} \quad .$$

Theorem IV applies for any T, so Theorem V is proved. As a consequence of Theorem V, we see that

$$r^* \begin{matrix} > 1 \\ < -1 \end{matrix} \text{ according as } r \begin{matrix} > \sqrt{\frac{2}{\pi}} \\ < -\sqrt{\frac{2}{\pi}} \end{matrix}$$

Asymptotic normality of r^*, which will be needed later in this section is a trivial consequence of a theorem of Cramér.

__Theorem VI:__ $r^* \sim \mathcal{N}(\rho, V(r^*|\omega,\rho))$.

__Proof:__ In expression (3.2) the term $\lambda(T)$ is seen to be a continuous function of $\bar{Z}$. Thus, r^* is a continuous function of the sample means $\bar{X}$, $\bar{Z}$, $\overline{X^2}$, $\overline{XZ}$. Applying Cramér's theorem (p. 366 of [1]), we have asymptotic normality with the asymptotic variance (3.4) calculated by Soper [8].

We shall now present two results which are more important than those just preceding. They concern the asymptotic, or large-sample, efficiency of r^*.

__Theorem VII:__ r^* is an asymptotically most efficient estimator of ρ when $\rho = 0$.

__Proof:__ In view of Theorem VI on asymptotic normality, we have a right to inquire about the asymptotic efficiency of r^*, which will be denoted by

$$\mathcal{E}(r^*|\omega,\rho) = \frac{V(\hat{\rho}|\omega,\rho)}{V(r^*|\omega,\rho)} .$$

It may be seen from Theorem I, (4.6), that

$$V(\hat{\rho}|\omega,0) = \frac{p(\omega)q(\omega)}{n[\lambda(\omega)]^2} . \tag{5.3}$$

Now, from (3.4) we observe that (5.3) coincides with $V(r^*|\omega,0)$. The conclusion follows from the definition of an asymptotically most efficient estimator.

__Theorem__ r^* is an asymptotically __least__ efficient estimator of ρ when $|\rho| \to 1$.

__Proof:__ An application of Theorem IV shows that

$$\phi\left(\frac{\omega-\rho x}{\sqrt{1-\rho^2}}\right) \quad \phi\left(-\frac{\omega-\rho x}{\sqrt{1-\rho^2}}\right) \leq \frac{2}{\pi} .$$

Hence, recalling the definition of $g(x,\omega,\rho)$ in Theorem I, we see that all integrals of the form $\int_{-\infty}^{+\infty} x^K g(x,\omega,\rho)dx$ exist. Schwarz' inequality shows that $V(\hat{\rho}|\omega,\rho)$ is such that the terms in braces is non-vanishing. Thus, $V(\hat{\rho}|\omega,\rho) \rightarrow 0$ as $|\rho| \rightarrow 1$. From the fact $V(r^*|\omega,\rho) \rightarrow \frac{2}{\pi}$ as $|\rho| \rightarrow 1$, we conclude that $\mathcal{E}(r^*|\omega,\rho) \rightarrow 0$.

The special case $\omega = 0$ has interesting features which will appear in Theorems X and XI. First we shall need another result from reference [9].

<u>Theorem IX</u> (Lemma 1 of [9])

$$\{1-2p(x)\}\lambda(x) - xp(x)q(x) \geq 0, \qquad x \geq 0.$$

<u>Theorem X</u>:

The asymptotic variance of r^* has its minimum for each ρ at $\omega = 0$.

<u>Proof</u>: We must show

$$V(r^*|\omega,\rho) \geq V(r^*|0,\rho)$$

for each ρ. In view of symmetry, it will be sufficient to show this for $\omega \geq 0$. Let,

$$A(\omega) = \frac{\omega^2 p(\omega)q(\omega)}{\{\lambda(\omega)\}^2} - \frac{\omega\{1-2p(\omega)\}}{\lambda(\omega)}, \qquad B(\omega) = \frac{p(\omega)q(\omega)}{\{\lambda(\omega)\}^2},$$

$$g(\omega) = \{1-2p(\omega)\}\lambda(\omega) - \omega p(\omega)q(\omega), \quad h(\omega) = p(\omega)q(\omega) - \pi\{\lambda(\omega)\}^2/2.$$

From this point until the end of the proof, we shall omit ω whenever it appears as an argument of any function. From reference [9] we have

$g' = 2\lambda^2 - pq$, $g'' = -4\omega\lambda^2 - (2q-1)\lambda$, $g(0) = g(\infty) = 0$, $g'(0) > 0$,

$h' = \lambda(1-2q) + \pi\omega\lambda^2$, $h'' = \lambda^2(\pi-2-2\pi\omega^2) - \omega(1-2q)\lambda$,

$$h(0) = h(\infty) = h'(0) = 0,\ h''(0) > 0.$$

Accordingly, we have $A = -\omega g \lambda^{-2}$, $B = h\lambda^{-2} + \pi/2$, with $A \leq 0$, $B \geq \pi/2$, both equalities holding at $\omega = 0$. The relation $V(r^*|\omega,\rho) \geq V(r^*|0,\rho)$ for all ρ may be written $\rho^2 A + B \geq \pi/2$ for all ρ. Since $A \leq 0$, this last expression is implied by $A + B \geq \pi/2$, which in turn is equivalent to $h \geq \omega g$. Thus, we must show $k = h - \omega g \geq 0$.

$$k' = 2\omega q(1-q) - 2(2q-1)\lambda + \omega(\pi-2)\lambda^2,$$

(5.4) $$k'' = 2q(1-q) - \lambda^2\{6 - \pi + \omega^2(2\pi - 4)\},$$

$$k(0) = k(\infty) = k'(0) = 0,\ k''(0) = 1 - 3/\pi > 0$$

We shall show that there exists no y such such $k'(y) = 0$, $k(y) < 0$. Suppose such a y does exist. Then,

(5.5) $$2(2q-1)\lambda = 2yq(1-q) + (\pi-2)y\lambda^2,$$

$$q(1-q)(1+y^2) - \pi\lambda^2/2 - y(2q-1)\lambda < 0.$$

Substituting the right member of the first expression into the second, we have $2q(1-q) < \lambda^2\{\pi + (\pi-2)y^2\}$. Thus, $k''(y) < \lambda^2\{2\pi-6 - (\pi-2)y^2\}$. A negative maximum must, from (5.4), be followed by a negative minimum. Hence, from the above relation in $k''(y)$, there exist no extrema which exceed $\{(2\pi-6)/(\pi-2)\}^{\frac{1}{2}}$. Assuming there is a negative extremum of k, then there must be a negative minimum in (0,1). Let y be this minimum point. Then $k''(y) > 0$, or from (5.4), $2q(1-q) - \lambda^2\{6-\pi- 12\pi-4)y^2\} > 0$. Substituting the volve of $2q(1-q)$ obtained from the first equation in (5.5), we reach $(2q-1) - y\lambda[2 + (\pi-2)y^2] > 0$.

The left member vanishes at $y = 0$ and has a negative derivative for $0 \leq y^2 \leq 1$. Therefore, there is no negative minimum in (0,1), and from the previous argument $k \geq 0$, which completes the proof.

Since for any fixed ρ, r^* is a better estimate when $\omega = 0$, it will be useful to have something simpler in the way of an asymptotic distribution of r^* than that contained in Theorem VI. We are therefore led to

Theorem XI:

When $\omega = 0$, we have to a close approximation

$$\tanh^{-1} \frac{2r^*}{\sqrt{5}} \sim \mathcal{N}\left(\tanh^{-1} \frac{2\rho}{\sqrt{5}}, \frac{5}{4n}\right).$$

Proof:

$$V(r^*|0,\rho) = \frac{1}{n}\left(\rho^4 - \frac{5}{2}\rho^2 + \frac{\pi}{2}\right) = \frac{1}{n}\left(\frac{5}{4} - \rho^2\right)^2 - \left(\frac{25-8\pi}{16n}\right).$$

Dropping the last term and solving the equation

$$g'(x) = \frac{1}{\left(\frac{5}{4} - x^2\right)},$$

we have $g(x) = (2/\sqrt{5}) \tanh^{-1} (2x/\sqrt{5})$. It is known that

$$\sqrt{n}\left\{g(r^*) - g(\rho)\right\} \sim \mathcal{N}(0,1),$$

so the theorem is proved.

Discussion of Results Concerning r^*

In looking over Theorems V, VI, VII, VIII, X, and XI, several facts stand out. First, even though r^* is consistent and asymptotically normal, it is still inadequate for estimating ρ because of its possible magnitude and its lack of large sample efficiency for large values of $|\rho|$. In the case of testing they hypothesis H: $\rho = \rho_0$ the first defect is not of so much consequence. Even in a problem of estimation, one can always operate under the rule: When $|r^*| < 1$ estimate ρ by r^*, when $r^* \geq 1$ estimate $\rho = 1$, and when $r^* \leq -1$ estimate $\rho = -1$. The gross defect is lack of efficiency.

In practically all applications it is of more interest to detect large values of $|\rho|$ than small values. In just such cases r^* is a "worst" estimator. On the other hand, again speaking in large sample terms, when $\rho = 0$, r^* is a "best" estimator. Hence, if we base a test of H: $\rho = \rho_0$ on r^*, good results should be achieved when $|\rho_0|$ is small. It is then recommended that r^* be used for one and only one purpose, to test H: $\rho = \rho_0$ when $|\rho_0|$ is small. If in addition the assumption $\omega = 0$ is tenable, then the variance stabilizing transformation of Theorem XI may be used, calculations being performed with Table VB of Fisher (p.210, [4]). In such a case advantages of the type discussed by Fisher (pp. 197-204, [4]) will accrue. $\sqrt{n}\,\sigma_{r^*}$ is given in Table I.

In the case of the problem of estimating the value of a particular test item for predicting student performance, $\omega = 0$ would occur when the question is of such difficulty that the average student would have probability .50 of answering it correctly. We could then use r^* and the variance stabilizing transformation of Theorem XI to test the null hypothesis H: $\rho = 0$, which is the hypothesis that the question doesn't add anything to the predictive value of the test. The acceptance of hypothesis H doesn't mean, of course, that the question should be omitted. It is well known that such questions have at times a useful purpose. Note that in view of the above discussion it would be wrong to use r^* to obtain confidence limits for ρ. Also note that according to Theorem X a question for which $\omega = 0$ is a desirable one to have.

6. Solution of the Likelihood Equations

In what follows all summations will be over the domain $i = 1,2,\ldots,n$. From (4.2) it may be seen that the likelihood equations are

$$(6.1) \qquad \sum \left\{ \frac{(1-s_i)\,\delta\eta(x_i,\omega) + s_i\,\delta\xi(x_i,\omega)}{(1-s_i)\,\eta(x_i,\omega) + s_i\,\xi(x_i,\omega)} \right\} = 0,$$

where δ refers to differentiation with respect to ω or ρ. Recall from (4.11) that,

$$\frac{\partial\eta(x_i,\omega)}{\partial\omega} = \psi(x_i,\omega), \quad \frac{\partial\xi(x_i,\omega)}{\partial\omega} = -\psi(x_i,\omega). \tag{6.2}$$

Also,

$$\frac{\partial\eta(x_i,\omega)}{\partial\rho} = -\frac{\psi(x_i,\omega)(x_i-\rho\omega)}{(1-\rho^2)}, \tag{6.3}$$

$$\frac{\partial\xi(x_i,\omega)}{\partial\rho} = \frac{\psi(x_i,\omega)(x_i-\rho\omega)}{(1-\rho^2)},$$

By the use of (6.2) and (6.3) equations (6.1) can be written as

$$\begin{aligned} &\sum (x_i-\rho\omega)(2z_i-1)\ \phi\left\{(2z_i-1)\left(\frac{\omega-\rho x_i}{\sqrt{1-\rho^2}}\right)\right\} = 0, \\ &\sum (2z_i-1))\ \phi\left\{(2z_i-1)\left(\frac{\omega-\rho x_i}{\sqrt{1-\rho^2}}\right)\right\} = 0. \end{aligned} \tag{6.4}$$

Now introduce the notation

$$\delta_i = 2z_i-1, \quad \gamma_i = (\omega-\rho x_i)(1-\rho^2)^{-\frac{1}{2}}, \quad \phi_i = \phi(\delta_i\gamma_i), \tag{6.5}$$

$$A_i = \phi_i(\phi_i - \delta_i\gamma_i).$$

Rewriting (6.4) again, in the new notation, we have

$$\sum \delta_i\phi_i = 0, \quad \sum \delta_i x_i\phi_i = 0. \tag{6.6}$$

Easy differentiation gives $\phi'(x) = \dot{\phi}(x)\{\phi(x) - x\}$. Newton's method in two variables gives the following equations in $\Delta\omega$ and $\Delta\rho$, where $\Delta\omega = \omega - \omega_1$, $\Delta\rho = \rho - \rho_1$, ω_1 and ρ_1 being initial guesses:

$$\left(\sum \frac{A_i}{\sqrt{1-\rho_1^2}}\right)\Delta\omega + \left(\frac{\rho_1\omega_1 \sum A_i - \sum A_i x_i}{(1-\rho_1^2)^{\frac{3}{2}}}\right)\Delta\rho = -\sum \delta_i \phi_i ,$$

(6.7)

$$\left(\sum \frac{A_i\, x_i}{\sqrt{1-\rho_1^2}}\right)\Delta\omega + \left(\frac{\rho_1\omega_1 \sum A_i x_i - \sum A_i x_i^2}{(1-\rho_1^2)^{\frac{3}{2}}}\right)\Delta\rho = -\sum \delta_i x_i \phi_i .$$

Let Δ be the determinant of the coefficients. The method of solution will then be the following:

<u>Method of Solution</u>

i) Compute (ω^*, r^*) from the sample (x_i, z_i), $i = 1,2,\dots,n$, where r^* is the sample biserial correlation coefficient and ω^* is the solution of the equation $p(\omega) = \bar{z}$. Now, let $\omega_1 = \omega^*$ and

$$\rho_1 = \begin{cases} r^* & \text{when } |r^*| < 1 \\ +.90 & \text{when } r^* \geq 1 \\ -.90 & \text{when } r^* \leq -1. \end{cases}$$

ii) Compute δ_i, γ_i, ϕ_i, $\delta_i\phi_i$, $\delta_i x_i \phi_i$, A_i, $A_i x_i$, $A_i x_i^2$ for $i = 1,2,\dots,n$, where δ_i, γ_i, ϕ_i, A_i are defined in (6.5), and Table II is used to obtain numerical values of the ϕ_i.

iii) Evaluate the three determinants

$$\Delta = \begin{vmatrix} \sum A_i & \rho_1\omega_1 \sum A_i - \sum A_i x_i \\ \sum A_i x_i & \rho_1\omega_1 \sum A_i x_i - \sum A_i x_i^2 \end{vmatrix} \cdot \frac{1}{(1-\rho_1^2)^2},$$

$$\Delta\omega = \begin{vmatrix} -\sum \delta_i \phi_i & \rho_1\omega_1 \sum A_i - \sum A_i x_i \\ -\sum \delta_i x_i \phi_i & \rho_1\omega_1 \sum A_i x_i - \sum A_i x_i^2 \end{vmatrix} \cdot \frac{1}{\Delta(1-\rho_1^2)^{3/2}},$$

$$\Delta\rho = \begin{vmatrix} \sum A_i & -\sum \delta_i \phi_i \\ \sum A_i x_i & -\sum \delta_i x_i \phi_i \end{vmatrix} \cdot \frac{1}{\Delta(1-\rho_1^2)^{1/2}}.$$

iv) Obtain (ω, ρ) from $(\Delta\omega, \Delta\rho)$ and (ω_1, ρ_1), and repeat the process using $\omega = \omega_2$, $\rho = \rho_2$ in place of ω_1 and ρ_1.

The rule given in i) is somewhat arbitrary, but is believed to be a good rule of thumb. The longest stage in the scheme outlined above is the determination of ϕ_i, $i = 1,2,\ldots,n$, from Table II.

We shall now present an illustration of the method. In order to have a good vantage point for observing the way the calculations run, we select a random sample from $\mathcal{BN}(0,0;1,1;1/\sqrt{2})$. A table of random numbers from such a population is not available directly, but can be constructed from a table of random numbers from $\mathcal{N}(0,1)$ as follows: Let

$$U = \mathcal{N}(0,1)\ , \quad V = \mathcal{N}(0,1)\ , \quad \omega = .50.$$

now, let

$$X = U \quad , \quad Y = \frac{U + V}{\sqrt{2}}\ .$$

Now dichotomize Y by introducing the Z variable:

$$Z = 1 \text{ if } Y \geqslant .50 \text{ and } Z = 0 \text{ if } Y < .50.$$

The computing scheme for 20 pairs of observations follows:

i	x_i	z_i	γ_i	δ_i	ϕ_i	$\delta_i x_i \phi_i$	$\phi_i - \delta_i \gamma_i$	A_i	$A_i x_i$	$A_i x_i^2$
1	.24	0	.030	-	.779	.187	.809	.630	.151	.036
2	.63	1	-.146		.707	.445	.853	.603	.380	.239
3	-.59	0	.404	-	.560	-.330	.964	.540	-.319	.188
4	1.08	0	-.348	-	1.032	1.115	.684	.706	.762	.822
5	.06	0	.111	-	.729	.044	.840	.451	.027	.002
6	-.01	0	.143	-	.709	-.007	.852	.604	-.006	.000
7	1.59	1	-.578		.470	.747	1.048	.493	.784	1.247
8	-.41	0	.323	-	.604	-.248	.927	.560	-.230	.094
9	-.33	1	.287		.989	-.326	.702	.694	-.229	.076
10	.99	1	-.308		.613	.607	.921	.565	.559	.553
11	.30	0	.003	-	.796	.239	.799	.636	.191	.057
12	-2.07	0	1.070	-	.262	-.542	1.332	.349	-.722	1.495
13	-.21	0	.233	-	.655	-.138	.888	.582	-.122	.026
14	-.47	1	.350		1.033	.486	.683	.706	-.332	.156
15	1.28	1	-.438		.541	.692	.979	.530	.678	.868
16	.82	1	-.231		.656	.538	.867	.582	.477	.391
17	1.06	1	-.339		.595	.631	.934	.556	.589	.624
18	1.09	0	-.353	-	1.035	1.128	.682	.706	.770	.839
19	.57	0	-.119	-	.875	.499	.756	.662	.377	.215
20	-.53	0	.377	-	1.052	-.558	.675	.710	-.376	.199

$\sum x_i = 5.09$ $\quad \sum x_i z_i = 5.04$ $\quad r^* = +.410\ (\rho_1)$ $\quad \gamma_i = 0.138 - 0.450\, x_i$

$\sum x_i^2 = 15.626$ $\quad \sum z_i = 9$ $\quad \omega^* = +.126\ (\omega_1)$

$\sum \delta_i \phi_i = -1.380$ $\quad \sum A_i = 11.865$ $\quad \sum A_i x_i^2 = 8.127$ $\quad \Delta = -122.54$

$\sum \delta_i x_i \phi_i = 0.343$ $\quad \sum A_i x_i = 3.409$ $\quad \Delta\omega = 0.108,$ $\quad \Delta\rho = 0.079$

$\omega_2 = 0.234, \quad \rho_2 = 0.489$

A second iteration resulted in $\omega_3 = 0.251$, $\rho_3 = 0.489$. Since $\hat{\rho}$ remained unchanged in the third place, the results were not included. Recall that the true value of ρ is .707. On the basis of our sample of 20, $\hat{\rho} = .489$ is the best we can do. However, by using the iterative scheme instead of r^* we removed 27% of the error.

7. Summary

The problem of biserial correlation is examined. An attempt is made to touch upon all aspects of the problem, without sacrificing mathematical rigor, and to describe the pertinent literature in its proper setting. Particular attention is paid to the use of maximum likelihood, and to the asymptotic efficiency of the sample biserial correlation coefficient. Results may be summarised as follows.

(1) The likelihood equations for ω, the point of dichotomy, and ρ, the population correlation coefficient, are obtained.

A method for their solution is described and illustrated by an example. Detailed calculations are given.

(2) Asymptotic variances are derived for the maximum likelihood estimators, $\hat{\omega}$ and $\hat{\rho}$, and are found to coincide with expressions given by Maritz [5].

(3) The sample biserial correlation coefficient (biserial r) is shown to be appropriate and very useful for certain problems in testing hypotheses, but essentially worthless in other situations. Several results are given in reference to the limiting distribution and asymptotic efficiency of this coefficient.

(4) Tables are given for the asymptotic standard deviation of the sample biserial correlation coefficient and for Mills' ratio, the latter being useful in solving the likelihood equations.

(5) Practical suggestions are offered, for application of the results of the paper, wherever possible.

TABLE I

The Asymptotic Standard Deviation of r^* (biserial r)

as a Function of p and ρ.

All values must be divided by $\sqrt{N}$

p or 1-p

± ρ	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
0	4.466	2.922	2.345	2.041	1.857	1.737	1.658	1.608	1.580	1.571
.10	2.104	1.699	1.521	1.419	1.353	1.308	1.278	1.258	1.247	1.243
.20	2.077	1.668	1.491	1.389	1.323	1.279	1.248	1.228	1.217	1.213
.30	2.033	1.616	1.440	1.339	1.273	1.229	1.198	1.179	1.167	1.163
.40	1.971	1.543	1.370	1.269	1.203	1.159	1.128	1.109	1.097	1.093
.50	1.893	1.449	1.279	1.179	1.114	1.069	1.038	1.019	1.008	1.004
.60	1.799	1.333	1.167	1.069	1.004	0.960	0.930	0.910	0.898	0.894
.70	1.691	1.194	1.034	0.939	0.875	0.831	0.801	0.781	0.769	0.766
.80	1.569	1.031	0.881	0.789	0.727	0.683	0.653	0.632	0.620	0.616
.90	1.438	0.842	0.705	0.619	0.559	0.517	0.486	0.465	0.453	0.449
1.00	1.302	0.616	0.503	0.429	0.374	0.335	0.304	0.283	0.270	0.266

TABLE II

Mills' Ratio[1,2]

-x	$\phi(x)$	-x	$\phi(x)$
.00	.79788	.40	.56188
.01	.79152	.41	.55649
.02	.78519	.42	.55112
.03	.77887	.43	.54578
.04	.77259	.44	.54047
.05	.76632	.45	.53520
.06	.76008	.46	.52993
.07	.75387	.47	.52471
.08	.74767	.48	.51948
.09	.74148	.49	.51431
.10	.73532	.50	.50917
.11	.72920	.51	.50404
.12	.72309	.52	.49893
.13	.71701	.53	.49387
.14	.71094	.54	.48883
.15	.70491	.55	.48380
.16	.69890	.56	.47883
.17	.69291	.57	.47386
.18	.68694	.58	.46893
.19	.68099	.59	.46402
.20	.67507	.60	.45914
.21	.66917	.61	.45429
.22	.66331	.62	.44947
.23	.65747	.63	.44468
.24	.65165	.64	.43992
.25	.64584	.65	.43518
.26	.64006	.66	.43047
.27	.63431	.67	.42580
.28	.62860	.68	.42114
.29	.62289	.69	.41652
.30	.61723	.70	.41192
.31	.61158	.71	.40736
.32	.60594	.72	.40282
.33	.60035	.73	.39832
.34	.59478	.74	.39383
.35	.58923	.75	.38939
.36	.58371	.76	.38496
.37	.57822	.77	.38056
.38	.57274	.78	.37621
.39	.56731	.79	.37186

1 Mills' ratio: $\phi(x) = \dfrac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{\int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy}$.

2 This table is reproduced with the kind permission of Professor Z. W. Birnbaum of the Laboratory of Statistical Research, University of Washington.

-x	$\phi(x)$	-x	$\phi(x)$
.80	.36756	1.30	.18974
.81	.36329	1.31	.18693
.82	.35904	1.32	.18414
.83	.35481	1.33	.18138
.84	.35062	1.34	.17866
.85	.34646	1.35	.17595
.86	.34234	1.36	.17328
.87	.33823	1.37	.17064
.88	.33416	1.38	.16803
.89	.33012	1.39	.16544
.90	.32611	1.40	.16288
.91	.32213	1.41	.16035
.92	.31818	1.42	.15784
.93	.31425	1.43	.15536
.94	.31035	1.44	.15292
.95	.30649	1.45	.15050
.96	.30264	1.46	.14810
.97	.29884	1.47	.14574
.98	.29506	1.48	.14340
.99	.29132	1.49	.14108
1.00	.28760	1.50	.13879
1.01	.28391	1.51	.13653
1.02	.28025	1.52	.13429
1.03	.27662	1.53	.13208
1.04	.27303	1.54	.12991
1.05	.26945	1.55	.12775
1.06	.26591	1.56	.12562
1.07	.26240	1.57	.12351
1.08	.25892	1.58	.12143
1.09	.25547	1.59	.11938
1.10	.25204	1.60	.11735
1.11	.24866	1.61	.11534
1.12	.24529	1.62	.11338
1.13	.24196	1.63	.11141
1.14	.23865	1.64	.10949
1.15	.23538	1.65	.10758
1.16	.23213	1.66	.10571
1.17	.22891	1.67	.10386
1.18	.22572	1.68	.10202
1.19	.22256	1.69	.10022
1.20	.21944	1.70	.09844
1.21	.21634	1.71	.09668
1.22	.21326	1.72	.09495
1.23	.21023	1.73	.09323
1.24	.20721	1.74	.09155
1.25	.20423	1.75	.08988
1.26	.20127	1.76	.08824
1.27	.19834	1.77	.08661
1.28	.19545	1.78	.08502
1.29	.19257	1.79	.08344

-x	$\phi(x)$
1.80	.08189
1.81	.08036
1.82	.07885
1.83	.07737
1.84	.07591
1.85	.07445
1.86	.07304
1.87	.07163
1.88	.07025
1.89	.06889
1.90	.06756
1.91	.06625
1.92	.06494
1.93	.06366
1.94	.06240
1.95	.06115
1.96	.05994
1.97	.05873
1.98	.05755
1.99	.05639
2.00	.05525
2.01	.05412
2.02	.05301
2.03	.05192
2.04	.05085
2.05	.04979
2.06	.04876
2.07	.04774
2.08	.04674
2.09	.04575
2.10	.04478
2.11	.04383
2.12	.04290
2.13	.04198
2.14	.04107
2.15	.04018
2.16	.03932
2.17	.03846
2.18	.03761
2.19	.03678
2.20	.03597
2.21	.03518
2.22	.03439
2.23	.03362
2.24	.03287
2.25	.03213
2.26	.03140
2.27	.03070
2.28	.02999
2.29	.02930

-x	$\phi(x)$
2.30	.02864
2.31	.02797
2.32	.02733
2.33	.02669
2.34	.02607
2.35	.02546
2.36	.02486
2.37	.02428
2.38	.02370
2.39	.02313
2.40	.02258
2.41	.02204
2.42	.02151
2.43	.02099
2.44	.02048
2.45	.01998
2.46	.01950
2.47	.01902
2.48	.01854
2.49	.01809
2.50	.01764
2.51	.01719
2.52	.01677
2.53	.01634
2.54	.01594
2.55	.01553
2.56	.01514
2.57	.01475
2.58	.01438
2.59	.01401
2.60	.01364
2.61	.01329
2.62	.01295
2.63	.01261
2.64	.01228
2.65	.01196
2.66	.01165
2.67	.01134
2.68	.01104
2.69	.01075
2.70	.01046
2.71	.01017
2.72	.00990
2.73	.00964
2.74	.00938
2.75	.00912
2.76	.00888
2.77	.00863
2.78	.00839
2.79	.00816

-x	$\phi(x)$
2.80	.00794
2.81	.00772
2.82	.00750
2.83	.00729
2.84	.00709
2.85	.00689
2.86	.00669
2.87	.00650
2.88	.00632
2.89	.00614
2.90	.00596
2.91	.00579
2.92	.00563
2.93	.00546
2.94	.00531
2.95	.00515
2.96	.00500
2.97	.00486
2.98	.00472
2.99	.00458
3.00	.00444

x	$\phi(x)$	x	$\phi(x)$
.00	.79788	.50	1.1410
.01	.80426	.51	1.1484
.02	.81066	.52	1.1557
.03	.81708	.53	1.1631
.04	.82351	.54	1.1704
.05	.82998	.55	1.1779
.06	.83646	.56	1.1854
.07	.84298	.57	1.1926
.08	.84950	.58	1.2000
.09	.85605	.59	1.2076
.10	.86262	.60	1.2151
.11	.86923	.61	1.2225
.12	.87582	.62	1.2300
.13	.88246	.63	1.2375
.14	.88909	.64	1.2450
.15	.89577	.65	1.2525
.16	.90246	.66	1.2601
.17	.90916	.67	1.2677
.18	.91589	.68	1.2753
.19	.92266	.69	1.2829
.20	.92941	.70	1.2905
.21	.93621	.71	1.2982
.22	.94300	.72	1.3058
.23	.94984	.73	1.3134
.24	.95668	.74	1.3212
.25	.96357	.75	1.3287
.26	.97043	.76	1.3364
.27	.97734	.77	1.3441
.28	.98427	.78	1.3519
.29	.99119	.79	1.3596
.30	.99816	.80	1.3674
.31	1.00516	.81	1.3751
.32	1.01215	.82	1.3829
.33	1.0192	.83	1.3906
.34	1.0262	.84	1.3986
.35	1.0333	.85	1.4063
.36	1.0404	.86	1.4142
.37	1.0474	.87	1.4221
.38	1.0545	.88	1.4298
.39	1.0616	.89	1.4378
.40	1.0687	.90	1.4457
.41	1.0760	.91	1.4535
.42	1.0831	.92	1.4613
.43	1.0903	.93	1.4693
.44	1.0975	.94	1.4773
.45	1.1047	.95	1.4852
.46	1.1120	.96	1.4932
.47	1.1193	.97	1.5013
.48	1.1265	.98	1.5092
.49	1.1338	.99	1.5170

x	$\phi(x)$	x	$\phi(x)$
1.00	1.5251	1.50	1.9387
1.01	1.5330	1.51	1.9470
1.02	1.5413	1.52	1.9554
1.03	1.5492	1.53	1.9643
1.04	1.5574	1.54	1.9728
1.05	1.5652	1.55	1.9814
1.06	1.5733	1.56	1.9904
1.07	1.5815	1.57	1.9984
1.08	1.5896	1.58	2.0068
1.09	1.5977	1.59	2.0153
1.10	1.6057	1.60	2.0243
1.11	1.6139	1.61	2.0325
1.12	1.6221	1.62	2.0412
1.13	1.6303	1.63	2.0500
1.14	1.6385	1.64	2.0585
1.15	1.6466	1.65	2.0670
1.16	1.6548	1.66	2.0756
1.17	1.6628	1.67	2.0846
1.18	1.6711	1.68	2.0929
1.19	1.6793	1.69	2.1022
1.20	1.6875	1.70	2.1102
1.21	1.6958	1.71	2.1191
1.22	1.7042	1.72	2.1277
1.23	1.7123	1.73	2.1358
1.24	1.7206	1.74	2.1450
1.25	1.7289	1.75	2.1538
1.26	1.7370	1.76	2.1626
1.27	1.7455	1.77	2.1711
1.28	1.7538	1.78	2.1796
1.29	1.7618	1.79	2.1882
1.30	1.7702	1.80	2.1979
1.31	1.7787	1.81	2.2060
1.32	1.7870	1.82	2.2148
1.33	1.7953	1.83	2.2242
1.34	1.8038	1.84	2.2326
1.35	1.8119	1.85	2.2406
1.36	1.8205	1.86	2.2502
1.37	1.8288	1.87	2.2589
1.38	1.8372	1.88	2.2676
1.39	1.8457	1.89	2.2758
1.40	1.8539	1.90	2.2847
1.41	1.8625	1.91	2.2941
1.42	1.8709	1.92	2.3026
1.43	1.8793	1.93	2.3116
1.44	1.8879	1.94	2.3202
1.45	1.8961	1.95	2.3288
1.46	1.9051	1.96	2.3375
1.47	1.9131	1.97	2.3463
1.48	1.9216	1.98	2.3557
1.49	1.9301	1.99	2.3641

x	$\phi(x)$	x	$\phi(x)$
2.00	2.3730	2.50	2.8233
2.01	2.3815	2.51	2.8297
2.02	2.3912	2.52	2.8401
2.03	2.3992	2.53	2.8506
2.04	2.4079	2.54	2.8612
2.05	2.4178	2.55	2.8662
2.06	2.4266	2.56	2.8794
2.07	2.4349	2.57	2.8902
2.08	2.4444	2.58	2.8969
2.09	2.4528	2.59	2.9044
2.10	2.4624	2.60	2.9138
2.11	2.4710	2.61	2.9206
2.12	2.4808	2.62	2.9300
2.13	2.4882	2.63	2.9412
2.14	2.4975	2.64	2.9472
2.15	2.5063	2.65	2.9630
2.16	2.5151	2.66	2.9665
2.17	2.5253	2.67	2.9815
2.18	2.5329	2.68	2.9895
2.19	2.5426	2.69	3.0003
2.20	2.5517	2.70	3.0030
2.21	2.5608	2.71	3.0175
2.22	2.5694	2.72	3.0276
2.23	2.5786	2.73	3.0312
2.24	2.5867	2.74	3.0460
2.25	2.5975	2.75	3.0506
2.26	2. 6055	2.76	3.0618
2.27	2.6157	2.77	3.0750
2.28	2.6240	2.78	3.0769
2.29	2.6323	2.79	3.0836
2.30	2.6427	2.80	3.0941
2.31	2.6511	2.81	3.1046
2.32	2.6596	2.82	3.1162
2.33	2.6695	2.83	3.1201
2.34	2.6781	2.84	3.1279
2.35	2.6860	2.85	3.1368
2.36	2.6947	2.86	3.1506
2.37	2.7064	2.87	3.1656
2.38	2.7122	2.88	3.1706
2.39	2.7248	2.89	3.1766
2.40	2.7307	2.90	3.1817
2.41	2.7390	2.91	3.1939
2.42	2.7503	2.92	3.2113
2.43	2.7586	2.93	3.2248
2.44	2.7701	2.94	3.2321
2.45	2.7786	2.95	3.2331
2.46	2.7855	2.96	3.2404
2.47	2.7941	2.97	3.2552
2.48	2.8035	2.98	3.2712
2.49	2.8121	2.99	3.2873
		3.00	3.2819

References

[1] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

[2] Du Bois, P. H. (1942). A note on the computation of biserial r in item validation. Psychometrika, Vol. 7, pp. 143-146.

[3] Dunlap, J. W. (1936). A nomograph for computing biserial r. Psychometrika, Vol. 1, pp. 59-60.

[4] Fisher, R. A.(1946). Statistical Methods for Research Workers. Oliver and Boyd, (10th edition).

[5] Maritz, J. S. (1953). Estimation of the correlation coefficient of a bivariate normal population when one of the variables is dichotomized. Psychometrika, Vol. 18, pp. 97-110.

[6] Pearson, K. (1909). On a new method for determining the correlation between a measured character A and a character B. Biometrika, Vol. 7, pp. 96-105.

[7] Royer, E. B. (1941). Punched cards methods for determining biserial coefficients. Psychometrika, Vol. 6, pp. 55-59.

[8] Soper, H. E. (1913). On the probable error for the biserial expression for the correlation coefficient. Biometrika, Vol. 10, pp. 384-390.

[9] Tate, R. F. (1953). On a double inequality of the normal distribution. Annals of Mathematical Statistics, Vol. 24, No. 1, pp. 133-134.

[10] Tate, R. F. (1954). Correlation between a discrete and a continuous variable. Annals of Mathematical Statistics, Vol. 25, No. , pp.

[11] Tocher, K. D. (1949). A note on the analysis of grouped probit data. Biometrika, Vol. 36, pp. 9-17.

[12] Walker, H. M. and Lev, J. (1953). Statistical Inference. Henry Holt and Co.

[13] Wherry, R. J. (1947). Multiple biserial and multiple point-biserial correlation. Psychometrika, Vol. 12.

Chief of Naval Research
Office of Naval Research
Washington 25, D. C.
Attn: Code 432
(Mathematics Branch) 5

Dr. E. Lukacs
Head, Statistics Branch
Office of Naval Research
Department of the Navy
Washington 25, D. C. 5

Scientific Section
Office of Naval Research
Department of the Navy
1000 Geary Street
San Francisco 9, Calif. 2
Attn: Dr. J. Wilkes 1

Director, Naval Research
Laboratory
Washington 25, D. C.
Attn: Technical Information
Officer 6

Office of the Assistant Naval
Attache for Research
Naval Attache
American Embassy
Navy No. 100
Fleet Post Office
New York, New York 2

Director
Office of Naval Research
Branch Office
844 North Rush Street
Chicago 11, Illinois 1

Director
Office of Naval Research
Branch Office
346 Broadway
New York 13, New York

Director
Office of Naval Research
Branch Office
1030 East Green Street
Pasadena 1, California

Commander
U.S. Naval Ordnance Test Station
Inyokern, China Lake, California 1

Chief of Naval Operations
Operation Evaluation Group-OP 374
The Pentagon
Washington 25, D. C. 1

Office of Naval Research
Department of the Navy
Washington 25, D. C.
Attn: Code 438
(Mechanics Branch) 2

Office of Naval Research
Logistics Branch - Code 436
T-3 Building
Washington 25, D. C. 1

ASTIA Documents Service Center
Knott Building
Dayton 2, Ohio 5

Headquarter, USAF
Director of Research and
Development
Washington 25, D. C. 1

Assistant Chief of Staff, G-4
for Research and Development
U. S. Army
Washington 25, D. C. 1

Chairman
Research and Development Board
The Pentagon
Washington 25, D. C. 1

Office of Technical Services
Department of Commerce
Washington 25, D. C. 1

Commander, Air Force Personnel
and Training Research Center
Attn: Director, Personnel Research Laboratory
Lackland Air Force Base, Texas 1

National Bureau of Standards
Institute for Numerical Analysis
405 Hilgard Avenue
Los Angeles 24, California 2

Chief, Statistical Engineering
Laboratory
National Bureau of Standards
Washington 25, D. C. 1

RAND Corporation
1500 Fourth Street
Santa Monica, California 1

Applied Mathematics and
Statistics Laboratory
Stanford University
Stanford, California 3

Professor Carl B. Allendoerfer
Department of Mathema tics
University of Washington
Seattle 5, Washington 1

Professor W. G. Cochran
Department of Biostatistics
The Johns Hopkins University
Baltimore S., Maryland 1

Professor Benjamin Epstein
Department of Mathematics
Wayne University
Detroit 1, Michigan 1

Professor Herbert Solomon
Teachers College
Columbia University
New York, New York 1

Professor W. Allen Wallis
Committee on Statistics
University of Chicago
Chicago 37, Illinois 1

Professor J. Wolfowitz
Department of Mathematics
Cornell University
Ithaca, New York 1