British Journal of Mathematical and Statistical Psychology (2004), 57, 235–251

www.bps.org.uk

On the correlation of a naturally and an artificially dichotomized variable

Rolf Ulrich[1]* and Markus Wirtz[2]
[1]University of Tübingen, Germany
[2]University of Freiburg, Germany

A method is suggested for estimating the correlation of a naturally (X) and an artificially (Y) dichotomized variable. It is assumed that a normal random variable (L) underlies the artificially dichotomized variable. The proposed correlation coefficient recovers the product moment correlation coefficient between X and L from a fourfold table of X and Y. The suggested correlation coefficient ν is contrasted with the phi correlation and the biserial η. The biserial η was proposed by Karl Pearson and is conceptually related to the new correlation coefficient. However, in addition, Pearson's biserial η invokes the assumption that the marginal distribution of L is normal, which contradicts its basic assumptions and thus does not recover the true correlation of L and X. Finally, an approximation is provided to simplify the calculation of ν and its standard error.

1. Introduction

Fourfold (2×2 contingency) tables are commonly used to represent the bivariate distribution of two dichotomous variables. When measuring the correlation in these tables, the nature of the dichotomy often plays an important role in the choice of the appropriate correlation coefficient. The term *dichotomy* generally applies to a division of the members of a sample or population into two groups. The division can be based on a qualitative or on a quantitative characteristic. In the former case the dichotomy is sometimes labelled as *natural*,[1] and in the latter case as *artificial*.

A natural dichotomy is simply based on a dichotomous attribute such as gender. Each member in a sample or population is allotted to one of two groups according to whether he or she possesses a specific attribute. In contrast, an artificial dichotomy is created whenever the values of a quantitative variable are recorded only as being greater or less than a specific cutoff value, say γ. For example, the age of an individual is a quantitative variable, say L, and in a study two age groups may be created by assigning individuals with $L \leq 60$ years to a 'younger age group' and individuals with $L > 60$ to an 'older age

* Correspondence should be addressed to Rolf Ulrich, Psychologisches Institut, Universität Tübingen, Friedrichstr, 21, 72072 Tübingen, Germany (e-mail: ulrich@uni-tuebingen.de).
[1] Some authors prefer the adjective 'true' instead of 'natural' (e.g. Kaplan & Saccuzzo, 2001).

group'. Likewise, a sample of students may be divided on the basis of their course grades into 'high' and 'low' achievers. The exact values of a quantitative variable are neglected in an artificial dichotomy, but this loss of information is sometimes accepted because it enhances the analysis and presentation of data.[2]

An artificial dichotomy may also be created if there is a limit to how a quantitative variable can be measured. It is often difficult or even impossible to measure a quantitative variable directly, although it is possible to obtain dichotomous information about it. For example, it is usually difficult to predict the potential of recidivism of a previously violent person, but a variety of indicators might be available to classify an individual as dangerous or harmless (Rice & Harris, 1995). In this case full predictive information about future violence is not accessible although a rough but reliable dichotomous classification (dangerous or harmless) might be possible. The actual underlying quantitative variable on which this classification is based thus has the status of a latent variable, and the created dichotomy may be considered as artificial (Lord & Novick, 1968, pp. 335–349). Hence, an artificial dichotomy may result in a diagnostic situation when one is forced to choose between two alternatives. Such an artificial dichotomy, however, may also arise from measurement limitations. For example, a variety of indicators could be used to classify an individual as an alcoholic or a non-alcoholic, though detailed information about alcohol consumption is not accessible.

There is some agreement in the psychometric literature that the *phi coefficient* ϕ should be used to assess the correlation between two naturally dichotomous variables (e.g. Kaplan & Saccuzzo, 2001; Lord & Novick, 1968; Wherry, 1984), although there are alternative association measures.[3] The formulae for the ordinary product moment correlation can be reduced to the phi coefficient for this type of binary variable, conventionally represented by 0 or 1. Hence the usual product moment correlation is used to measure the relationship between two naturally dichotomized variables.

Although ϕ could also be used to measure the relationship between two artificially dichotomized variables, it would not be possible to infer the actual correlation between the underlying quantitative variables from this measure.[4] This latent relationship, however, may be inferred from a 2×2 contingency table if ϕ is replaced by the *tetrachoric correlation coefficient* developed by Pearson (1900). He assumed that the underlying continuous variables follow a bivariate normal distribution. The tetrachoric correlation estimates the product moment correlation in this bivariate normal distribution. The calculation of the tetrachoric correlation coefficient is cumbersome, but fortunately a computer program (Brown, 1977) and sufficiently accurate approximations (Digby, 1983) are available for its calculation. Detailed discussions of

---

[2] *The problems involved in artificially dichotomizing a continuous variable have been addressed in several studies. For example, dichotomizing leads to an attenuation of statistical power and to an inflated Type I error (Aguinis, 1995; Cohen, 1983; Stone-Romero & Anderson, 1994; Vargha, Rudas, Delaney, & Maxwell, 1996). Therefore, researchers should avoid dichotomizing variables whenever continuous information about them is available.*

[3] *One such alternative measure is Yule's Q (see Bishop, Fienberg, & Holland, 1975, p. 378) and another is the so-called odds ratio (see Tabachnick & Fidell, 2001, p. 548). These alternative measures, however, are not special cases of the product moment correlation.*

[4] *In general, the product moment correlation coefficient of polychotomized variables is attenuated relative to the true correlation coefficient of the underlying continuous variables (Cohen, 1983; Ulrich & Giray, 1989). This attenuation effect increases with a decrease in the number of categories into which the observations are sorted, and since an artificial dichotomy represents the extreme case of polychotomization, a particularly large attenuation effect should be expected for the phi coefficient. In very special cases, however, there are exceptions to this rule (Vargha* et al.*, 1996).*

the tetrachoric correlation coefficient can be found in the works of Lord and Novick (1968, pp. 345–349), Kendall and Stuart (1973, pp. 316-319), and Harris (1988). For an interesting historical survey of the development of the tetrachoric as well as related measures, see Cowles (1989).

The literature available provides no reasonable estimator from which the latent correlation between a naturally and an artificially dichotomized variable can be inferred on the basis of a 2×2 table. As in the case of two artificially dichotomized variables, ϕ is of no use if one wishes to uncover the latent product moment correlation between the naturally dichotomized variable and the continuous variable underlying the artificial dichotomy; since ϕ invokes no assumptions about an underlying continuous variable, there exists no logical basis for such an inference. This gap is almost never mentioned in the psychological and statistical literature available (Bortz & Döring, 2001, p. 509; Frankfort-Nachmias, 1997; Howell, 1997; Sheskin, 2000; see, however, Bortz, 1979, p. 277). Nonetheless, it immediately becomes clear if one tries to teach correlation as a subject in any systematic way.

Although ϕ could also be used to measure the manifest relationship between an artificially dichotomized variable and a naturally dichotomous one, it would not be possible to infer the latent correlation of the natural dichotomous variable with the underlying quantitative variable. It is, however, common practice to employ ϕ to quantify the relationship between both variables. Although more recent works criticize this practice (Fuller & Cowan, 1999; Hasselblad & Hedges, 1995; Mossman, 1994; Rice & Harris, 1995), these studies provide no appropriate estimator of this latent correlation.[5]

A correlation coefficient that appears to close this gap is the special case of the so-called *biserial* η (see Kendall & Stuart, 1973, pp. 319–321 for a detailed discussion). Unfortunately, this coefficient rests on the implicit assumption that there is an underlying univariate normal distribution, which, of course, is an unnecessary restriction. Interestingly, however, the basic assumptions of the biserial η are suitable for this correlation problem; this will become clearer below.

The present paper develops a new correlation coefficient to infer the true correlation between a naturally and an artificially dichotomized variable. This new coefficient involves the basic assumptions of the biserial η while avoiding its implicit assumption of an underlying univariate normal distribution. This new coefficient is compared with the biserial η and ϕ.

2. Assumptions

Some notation is needed to simplify the presentation. Samples (L, X) are taken from a bivariate distribution with correlation $\mathrm{Corr}[L, X]$. X represents the naturally dichotomized variable and L the (latent) continuous variable. Furthermore, let Y be an artificially dichotomized variable that takes the value 0 if $L \leq \gamma$ and the value 1 if $L > \gamma$. Without loss of generality it is assumed that X is a Bernoulli variable taking the values 0 and 1 with the probabilities $\Pr\{X = 0\}$ and $\Pr\{X = 1\} = 1 - \Pr\{X = 0\}$, respectively. Hence, each observation (l, x) can be allocated to one of the cells of a 2×2 contingency table. (For an example of a contingency table, see Table 1). The statistical

[5] *Nevertheless, these studies suggest non-correlational measures. For example, Hasselblad and Hedges (1995) employ d′ from signal detection theory to measure the effect size associated with this latent correlation. The advantage of this measure is that d′ does not depend on the cutoff value on which the artificial dichotomization is based.*

Table 1. A 2×2 contingency table showing the bivariate distribution and marginal distributions of X and Y. Variable X is naturally dichotomized, while Y is artificially dichotomized on the basis of a (latent) continuous variable L.

	X		
Y	0	1	$p_{\cdot y}$
0	p_{00}	p_{10}	$p_{\cdot 0}$
1	p_{01}	p_{11}	$p_{\cdot 1}$
$p_{x\cdot}$	$p_{0\cdot}$	$p_{1\cdot}$	

problem is to infer the (latent) correlation Corr$[L, X]$ between L and X from this type of table.[6] Throughout the text we proceed from the following two assumptions:

Assumption 1.
The conditional density function $f(l/X = x)$ of $L = l$ given $X = x$ is normal.[7]

Assumption 2.
The conditional variance of L does not depend on X, that is,

$$\mathrm{Var}[L|X = 0] = \mathrm{Var}[L|X = 1] = \sigma^2.$$

Assumptions 1 and 2 are not specific to the correlation developed in this paper, because they are also explicitly or implicitly assumed in inferred correlations such as the biserial correlation ρ_b, the tetrachoric, and the biserial η (see Kendall & Stuart, 1973, pp. 316–322).

From these assumptions it is clear that a positive (negative) correlation Corr$[L, X]$ must result if the conditional mean $E[L|X = 1] = \mu_1$ is larger (smaller) than the conditional mean $E[L|X = 0] = \mu_0$. The magnitude of Corr$[L, X]$ increases with the difference between the means, although the conditional variances $\mathrm{Var}[L|X = 0] = \sigma_0^2$ and $\mathrm{Var}[L|X = 1] = \sigma_1^2$ remain constant according to Assumption 2. Of course, if X and L are uncorrelated, then the two conditional means must be equal.

According to the above assumptions, the marginal or unconditional distribution $f(l)$ of L corresponds to the probabilistic mixture distribution

$$f(l) = f(l|X = 0) \cdot \Pr\{X = 0\} + f(l|X = 1) \cdot \Pr\{X = 1\},$$

which implies that the marginal distribution of L cannot correspond to a normal distribution unless $\mu_0 = \mu_1$ and $\sigma_0 = \sigma_1$, because a mixture of normals is not necessarily a normal (cf. Everitt, 1985). This is in contrast to the biserial η, which relies implicitly on the assumption that $f(l)$ is normal, although the basic assumptions of the biserial η are the same as those above. Thus, the biserial η is inherently flawed, since it rests on mutually contradictory assumptions. Because of this flaw, the biserial η provides biased estimates of Corr$[L, X]$, as will be demonstrated below.

[6] *It should be noted that ϕ is equal to Corr[Y, X], but usually not equal to Corr[L, X].*
[7] *This assumption is actually more specific than necessary. The general form of the new correlation coefficient could also be applied to other latent distributions (e.g. the exponential or the binomial distribution).*

The following proposition (proved in Appendix A) can be deduced from the above assumptions. It shows how Corr[L, X] can be calculated from the information provided by a 2×2 contingency table such as Table 1. The deduction rests on the assumption that $f(l)$ is a mixture of normal distributions and thus abolishes the restriction of the biserial η.

Proposition I. *Under Assumptions* 1 *and* 2, *the correlation* $\nu \equiv \mathrm{Corr}[\mathrm{L}, \mathrm{X}]$ *between a naturally dichotomized variable* X *and a continuous variable* L *underlying the artificially dichotomized variable* Y *is*

$$\nu = \frac{\Delta}{\sqrt{\Delta^2 + \frac{1}{p_{1.}(1-p_{1.})}}}, \tag{1}$$

with $\Delta = \Phi^{-1}[\mathrm{p}_{00}/\mathrm{p}_{0.}] - \Phi^{-1}[\mathrm{p}_{10}/\mathrm{p}_{1.}]$, *where* Φ^{-1} *denotes the quantile function of a standard normal variable and* $\mathrm{p}_{1.} = \mathrm{p}_{10} + \mathrm{p}_{11}$.

Note that Δ represents the standardized difference between means

$$\Delta = \frac{\mu_1 - \mu_0}{\sigma}$$

of the conditioned normal distributions. As already mentioned, (1) becomes zero if the two conditional means are equal.

3. Numerical example

To illustrate Proposition 1, we proceed from the probabilities shown in Table 2. To use a concrete example, we may think of $Y = 1$ as a high grade and $Y = 0$ as a low grade, of $X = 0$ as male and $X = 1$ as female. The actual grades, L, are divided according to an unknown cutoff γ into low and high grades. Thus Table 2 illustrates a possible association between an artificial dichotomous variable Y (grades) and a natural dichotomous variable X (gender).

Table 2. An example of a bivariate distribution for X and Y. The natural dichotomy is gender, X. The artificial dichotomy is high ($Y = 1$) versus low ($Y = 0$) grades.

	Gender	
Grade	Male	Female
Low	.15	.10
High	.25	.50
$p_{x.}$	.40	.60

According to Proposition 1, we have

$$\begin{aligned}\Delta &= \Phi^{-1}\left[\frac{.15}{.40}\right] - \Phi^{-1}\left[\frac{.10}{.60}\right] \\ &= (-0.3186) - (-0.9674) \\ &= 0.6488,\end{aligned}$$

and, from (1),

$$\nu = \frac{0.6488}{\sqrt{0.6488^2 + 1/(.6(1-.6))}} \approx .30.$$

The positive sign indicates that, on average, females received better grades.

4. Logistic approximation of the normal distribution

For practical purposes a more convenient formula for computing ν is obtained if the conditional density function $f(l|X = x)$ is assumed to be logistic instead of normal. The cumulative distribution function (cdf) of a logistic random variable X is

$$\Psi(x) = \frac{1}{1 + e^{-1.7x}}. \tag{2}$$

The constant 1.7 in (2) makes $\Psi(x)$ very similar to the standard normal cdf $\Phi(x)$, because $|\Psi(x) - \Phi(x)| < 0.01$ must hold for all values of x (Johnson & Kotz, 1970, p. 6). Thus, for practical purposes, $\Psi(x)$ provides an approximation for $\Phi(x)$.

Under the assumption that $f(l|X = x)$ is logistic, Proposition 1 can be simplified (see Appendix B) to

$$\nu = \frac{\ln\left(\frac{p_{00}p_{11}}{p_{01}p_{10}}\right)}{\sqrt{\ln\left(\frac{p_{00}p_{11}}{p_{01}p_{10}}\right)^2 + \frac{2.89}{p_{1.}(1-p_{1.})}}}, \tag{3}$$

which does not require the calculation of quantile values. Applying this formula to the probabilities in Table 2 yields a correlation of $\nu = .30$, which is virtually identical to that computed on the assumption of a normal distribution.[8]

The assumption of a logistic approximation also leads to a manageable expression for computing the standard error of the estimate $\hat{\nu}$. Its derivation is shown in Appendix C. For the probabilities of Table 2 and a sample size of $N = 100$, the standard error equals 0.121.

Under the assumption of a logistic distribution, the standardized difference between means is equal to $\ln(\frac{p_{00}p_{11}}{p_{01}p_{10}})$, which is also known as the log-odds ratio in the statistical literature (Somes & O'Brien, 1988). Thus, the coefficient ν can also be used to standardize the log-odds ratio to between -1 and $+1$, which enhances the comparison of the ratio with correlation coefficients.

Since (3) is a function of the odds ratio, common coefficients of association measures behave in a similar way to ν in that they are also invariant with respect to the cutoff value γ. For example, Yule's Y and Yule's Q (Chambers, 1982) are monotonic transformations of the odds ratio. Nevertheless, they are developed for different applications and also

[8] *The goodness of the approximation was evaluated for 160 000 contingency tables. For each table, a new set of numerical values was generated for the cell probabilities p_{00}, p_{01}, p_{10}, and p_{11}. Specifically, each probability was systematically varied over the interval [.02, .87] and under the constraint $p_{00} + p_{01} + p_{10} + p_{11} = 1$. The resulting values of ν ranged from $-.86$ to .86. To assess the goodness of the approximation, we calculated for each table the absolute difference between (1) and (3) and then averaged these 160 000 absolute differences. The average absolute difference was 0.009 and the standard deviation of all absolute differences was 0.007. In addition, only 3.4% of all absolute differences were larger than 0.02. This pattern of results clearly suggests that (3) provides a satisfactory approximation in most practical situations. Additional computations showed that for cell probabilities close to zero or close to one, the difference between (1) and (3) can be substantial. In such extreme situations the approximation must be used with caution.*

yield different absolute values. In contrast to ν, these measures cannot be employed to recover the true latent correlation Corr[L, X]. For the probabilities in Table 2, the true latent correlation is $\nu = .30$, whereas $Y = .25$ and $Q = .47$ differ substantially from this value.

5. On the sample estimation of ν

It is common practice to substitute sample estimates for population parameters. In our case the compound probabilities p_{00}, p_{01}, p_{10}, and p_{11} should be replaced by the corresponding sample estimates. Such a practice is not always optimal, but in this case it can be shown (see Appendix D) that this strategy provides maximum likelihood estimates for the parameters Δ and $p_{1.}$ in (1). Since ν is a function of Δ and $p_{1.}$, it follows according to the invariance property of maximum likelihood estimation that this strategy provides a maximum likelihood estimator for ν (see Mood, Graybill, & Boes, 1974, pp. 283–286). It is well known that maximum likelihood estimators possess excellent properties, at least asymptotically.

6. Comparison of ν with biserial η and ϕ

In this section the new correlation coefficient is contrasted with both the ϕ coefficient and the biserial η suggested by Pearson (1909–1910). Pearson proposed the biserial η as a 'method of determining the correlation, when one variable is given by alternative and the other by multiple categories' (p. 248). According to his basic assumptions, the alternative category (Y) emerges from an underlying normal variable (L), whereas the multiple categories (X) might be purely qualitative (e.g. type of crime, type of nationality). Furthermore, the conditional density functions $f(l|X = x)$ of L for $x = 0, 1, 2, \ldots$ were assumed to be normal, differing only in location. The binary variable Y is formed by dividing the conditional distributions at γ into two artificial categories. Thus, the basic assumptions of the biserial η are identical to those from which Proposition 1 emerged.

In the special case that the qualitative variable has only two categories, the biserial η reduces to

$$\eta^2 = 1 - \frac{1 + z_y^2}{1 + z_0^2 \cdot p_{0.} + z_1^2 \cdot p_{1.}}, \tag{4}$$

with $z_y = \Phi^{-1}[p_{.1}]$, $z_0 = \Phi^{-1}[p_{01}/p_{0.}]$, and $z_1 = \Phi^{-1}[p_{11}/p_{1.}]$. The derivation of the biserial η proceeds from the correlation ratio

$$\eta^2 = \mathrm{Corr}[L, X]^2 = \frac{\mathrm{Var}[E[L|X = x]]}{\mathrm{Var}[L]},$$

which can never be negative (for a detailed mathematical treatment of the biserial η, see Kendall & Stuart, 1973, pp. 319–321). This derivation is only valid, however, if the marginal distribution of L is normal, because z_y is computed using the normal distribution function on the marginal distribution of Y.

Applying the biserial η to the numerical example given in Table 2, one computes

$$z_0 = \Phi^{-1}\left[\frac{.25}{.40}\right] = 0.3186,$$

$$z_1 = \Phi^{-1}\left[\frac{.50}{.60}\right] = 0.9674,$$

$$z_y = \Phi^{-1}[.75] = 0.6745.$$

Substituting these values into (4) results in

$$\eta^2 = 1 - \frac{1 + (0.6745)^2}{1 + (0.3186)^2 \cdot (.40) + (0.9674)^2 \cdot (.60)}$$
$$\approx .092,$$

and hence $|\mathrm{Corr}[L, X]| = \eta = .30$. Note that this value is the same as the one given by ν, neglecting possible rounding errors.

In general, however, there is a substantial difference between the coefficients η and ν. This is easily demonstrated by computing η and ν from fourfold tables consistent with Assumptions 1 and 2. Table 3 shows the results of such computations. To generate each fourfold table, the values of $E[L|X = 0]$, $E[L|X = 1]$, and σ were fixed at 0.0, 4.0, and 1.5, respectively and the values of γ and $p_{1.}$ were varied throughout the set (-2.0, 0.0, 2.0, 4.0, 6.0) and (.05, .25, .50, .75, .95), respectively. Separate fourfold tables were computed for each (γ, $p_{1.}$) combination, and the corresponding values of ϕ, η and ν shown in Table 3 were obtained from these tables. Table 3 reveals several interesting features:

(1) By definition, ν is invariant against a variation of cutoff γ. However, both η and ϕ vary considerably with γ; ϕ more so than η.
(2) ν and ϕ can differ considerably.
(3) Interestingly, in some cases ϕ may even be larger than η or ν.

Table 3. The correlation coefficients ϕ, biserial η, and ν as a function of γ and $p_{1.}$. In all cases the values of $E[L|X = 0] = 0$, $E[L|X = 1] = 4$ and $\sigma = 1.5$ are constant. All values are rounded to the nearest hundredth.

		γ				
$p_{1.}$		-2.0	0.0	2.0	4.0	6.0
.05	ϕ	.07	.22	.53	.65	.29
	η	.43	.51	.44	.64	.72
	ν	.50	.50	.50	.50	.50
.25	ϕ	.16	.44	.78	.64	.26
	η	.70	.78	.73	.80	.79
	ν	.76	.76	.76	.76	.76
.50	ϕ	.22	.57	.82	.57	.22
	η	.78	.83	.80	.83	.78
	ν	.80	.80	.80	.80	.80
.75	ϕ	.26	.64	.78	.44	.16
	η	.79	.80	.73	.78	.70
	ν	.76	.76	.76	.76	.76
.95	ϕ	.29	.65	.53	.22	.07
	η	.72	.64	.44	.51	.43
	ν	.50	.50	.50	.50	.50

This demonstration clearly reveals that neither η nor ϕ uncovers the true correlation $\mathrm{Corr}[X, L] = \nu$ between X and L. The strong dependence of ϕ on γ is expected from the theoretical results provided by Carroll (1961), who showed that ϕ always depends on the cutoff points in any bivariate distribution. However, the fact that ϕ may even be larger than ν is remarkable, since one is usually tempted to assume that dichotomizing leads to a loss of information that attenuates the true correlation coefficient (but see

Vargha *et al.*, 1996). Therefore, ϕ cannot generally be viewed as a conservative correlation coefficient that will always provide a lower bound on the true relationship between any two dichotomized variables.

7. Conclusion

In this paper a new method is proposed to estimate the correlation between a naturally (X) and an artificially (Y) dichotomized variable. It is assumed that Y emerges from an underlying variable L, which is dichotomized at an unknown cutoff γ. This new coefficient, ν, recovers the correlation between X and L if certain assumptions about L are met. It is assumed, firstly, that the distribution of L conditioned on X is normal; and secondly, that the conditional variance σ^2 of L does not depend on X (homoscedasticity). Both assumptions are usually invoked in inferred correlation coefficients such as the biserial and the tetrachoric correlation. Violation of either assumption has been studied in detail by Hasselblad and Hedges (1995) for the indices d' and the log-odds ratio on which ν is based. When the violations are modest (e.g. both variances do not differ by a factor more than 4), the bias of ν is neglectable.

The estimation of ν and its standard error is quite cumbersome, if one proceeds from the assumption that L is normally distributed. Hence, we have provided an approximation based on the logistic distribution. This approximation not only yields a simpler formula for computing ν but also allows the derivation of a convenient expression for computing the standard error of the estimate $\hat{\nu}$. Furthermore, we have shown that the suggested equation for estimating ν is a maximum likelihood estimator and thus possesses excellent estimating properties (consistency and high efficiency). As discussed in Appendix C, this information also allows the construction of confidence intervals for the estimate $\hat{\nu}$ when N is relatively large.

The new correlation method is based on assumptions conceptually identical to those of the biserial η. However, the biserial η assumes implicitly that the marginal distribution of L is normal, which actually contradicts the basic assumptions of the biserial η, according to which the marginal distribution of L should be a mixture of two normal distributions. According to the basic assumptions, η should remain constant when the cutoff γ is varied. However, it was demonstrated that the expected invariance of η in this case does not hold. In contrast to the biserial η, the new correlation coefficient ν avoids the assumption of an unconditional univariate normal distribution.

Since ϕ depends on γ, this may lead to false conclusions under certain circumstances. Assume that a researcher in a meta-analysis study compares the ϕ correlation of gender and alcoholism of two countries, say A and B. The reported ϕs are .82 and .22 in A and B, respectively. Does this justify the conclusion that alcoholic consumption depends more on gender in A than in B? Not necessarily, as can be seen in Table 3 under the assumption of $p_{1.} = .5$ in both countries. The actual correlation between L and X could be $\nu = .80$ for both countries, but the reported ϕ values of .82 and .22 could be obtained with the cutoff values 2.0 and 6.0, respectively. Clearly, ν is a more meaningful measure of correlation because of its insensitivity to cutoff values. The same argument applies to two diagnosticians who rely on different criteria for characterizing previously violent people as potential recidivists. As long as their classification is based on the same normally distributed underlying information, ν would be constant.

The new correlation coefficient ν can also be regarded as an analogue of the biserial correlation ρ_b (Kendall & Stuart, 1973, pp. 321–323), which is widely used in psychology (Ferguson, 1976, pp. 418–419). The coefficient ρ_b differs conceptually from

biserial η, because ρ_b recovers the product moment correlation between a normally distributed X and a latent normally distributed variable L that is artificially dichotomized. Thus, there is now a complete set of correlation techniques available to reveal the correlation between non-dichotomous and dichotomous variables. Specifically, the biserial correlation, the tetrachoric and the new correlation coefficient ν should be employed whenever the true correlation between two variables is concealed by an artificial dichotomy. A common attractive feature of these coefficients is that performing a factor analysis based on these correlation coefficients may reveal a more realistic factor structure than a factor analysis based on a conventional Pearson correlation matrix.

Acknowledgements

We thank Willi Nagl for some good advice, and appreciate the helpful comments of two anonymous reviewers on a previous version of this paper.

References

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, *21*, 1141–1158.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge MA: MIT Press.

Bortz, J. (1979). *Lehrbuch der Statistik für Sozialwissenschaftler*. [A textbook of statistics for social scientists] Berlin: Springer-Verlag.

Bortz, J., & Döring, N. (2001). *Forschungsmethoden und Evaluation* [Research methods and evaluation] (3rd ed.). Berlin: Springer-Verlag.

Brown, M. B. (1977). Algorithm AS 116. The tetrachoric correlation and its asymptotic standard error. *Applied Statistics*, *26*, 343–351.

Carroll, J. B. (1961). The nature of data, or how to choose a correlation coefficient. *Psychometrika*, *26*, 347–372.

Chambers, R. G. (1982). Correlation coefficients from 2 × 2 tables and from biserial data. *British Journal of Mathematical and Statistical Psychology*, *35*, 216–227.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychology Measurement*, 7, 249–253.

Cowles, M. (1989). *Statistics in psychology: An historical perspective.* Hillsdale, NJ: Lawrence Erlbaum.

Digby, P. G. N. (1983). Approximating the tetrachoric correlation coefficient. *Biometrics*, *39*, 753–757.

Everitt, B. S. (1985). Mixture distributions. In S. Kotz & N. L. Johnson (Eds), *Encyclopedia of statistical sciences*, Vol 5 (pp. 559–569). New York: Wiley.

Ferguson, G. A. (1976). *Statistical analysis in psychology and education* (4th ed.). New York: McGraw-Hill.

Frankfort-Nachmias, C. (1997). *Social statistics for a diverse society*. Thousand Oaks, CA: Pine Forge Press.

Fuller, J., & Cowan, J. (1999). Risk assessment in a multi-disciplinary forensic setting: Clinical judgement revisited. *Journal of Forensic Psychiatry*, *10*, 276–289.

Harris, B. (1988). Tetrachoric correlation coefficient. In S. Kotz & N. L. Johnson (Eds), *Encyclopedia of statistical sciences*, Vol. 9 (pp. 223–225). New York: Wiley.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, *117*, 167–178.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury Press.

Johnson, N. L., & Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions – 2*. Boston: Houghton Mifflin.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1997). *Discrete multivariate distributions*. New York: Wiley.

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications, and issues* (5th ed.). Monterey, CA: Brooks/Cole.

Kendall, M. G., & Stuart, A. (1973). *The advanced theory of statistics. Volume 2: Interference and relationship* (3rd ed.). New York: Hafner.

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

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Beverly Hills, CA: Sage.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783–792.

Pearson, K. (1900). On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society, Series A*, *195*, 1–47.

Pearson, K. (1909–1910). On a new method of determining correlation, when one variable is given by alternative and the other by multiple categories. *Biometrika*, *7*, 248–257.

Rice, M. E., & Harris, G. D. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737–748.

Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Boca Raton, FL: Chapman & Hall.

Somes, G. W., & O'Brien, K. F. (1988). Odds ratio estimators. In S. Kotz, & N. L. Johnson (Eds), *Encyclopedia of statistical sciences*, Vol. 6 (pp. 407–410). New York: Wiley.

Stone-Romero, E. F., & Anderson, L. E. (1994). Techniques for detecting moderating effects: Relative statistical power of multiple regression and a comparison of subgroup-based correlation coefficients. *Journal of Applied Psychology, 79*, 354–359.

Stuart, A., & Ord, K. (1987). *Kendall's advanced theory of statistics. Volume 1: Distribution theory* (2nd ed.). London: Griffin.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.

Ulrich, R., & Giray, M. (1989). Time resolution of clocks: Effects on reaction time measurement—Good news for bad clocks. *British Journal of Mathematical and Statistical Psychology*, *42*, 1–12.

Vargha, A., Rudas, T., Delaney, H. D., & Maxwell, S. E. (1996). Dichotomization, partial correlation, and conditional independence. *Journal of Educational and Behavioral Statistics*, *21*, 264–282.

Wherry, R. J. Sr. (1984). *Contributions to correlational analysis*. Orlando, FL: Academic Press.

Received 19 May 2002; revised version received 11 February 2003

Appendix A

Proof of proposition 1

First, note that

$$\mathrm{Corr}[L, X] = \frac{\mathrm{Cov}[X, L]}{\sqrt{\mathrm{Var}[X] \cdot \mathrm{Var}[L]}}, \tag{A.1}$$

where $\mathrm{Cov}[X, L]$ denotes the covariance of X and L. The terms $\mathrm{Cov}[X, L]$, $\mathrm{Var}[L]$, and $\mathrm{Var}[X]$ will be analysed separately.

First, the covariance of X and L is

$$\mathrm{Cov}[X, L] = E[X \cdot L] - E[X] \cdot E[L], \tag{A.2}$$

with

$$\begin{aligned} E[X \cdot L] &= E[E[X \cdot L|X]] \\ &= E[0 \cdot L|X = 0]\Pr\{X = 0\} + E[1 \cdot L|X = 1]\Pr\{X = 1\} \\ &= E[L|X = 1]\Pr\{X = 1\} \\ &= \mu_1 \cdot p_{1.}, \end{aligned} \tag{A.3}$$

where μ_1 is the conditional mean of L under $X = 1$. Furthermore, we have

$$E[X] = p_{1.} \tag{A.4}$$

and

$$\begin{aligned} E[L] &= E[E[L|X]] \\ &= \mu_1 \cdot p_{1.} + \mu_0 \cdot (1 - p_{1.}), \end{aligned} \tag{A.5}$$

where μ_0 denotes the conditional mean of L under $X = 0$. Substitution of (A.5), (A.4), and (A.3) into (A.2) yields

$$\mathrm{Cov}[X, L] = p_{1.} \cdot (1 - p_{1.}) \cdot (\mu_1 - \mu_0). \tag{A.6}$$

Second, the variance of L can be written as

$$\mathrm{Var}[L] = \mathrm{Var}[E[L|X]] + E[\mathrm{Var}[L|X]]. \tag{A.7}$$

According to Assumption 2, we may write

$$E[\mathrm{Var}[L|X]] = \sigma^2. \tag{A.8}$$

For the variance of $E[L|X]$ we write

$$\begin{aligned} \mathrm{Var}[E[L|X]] &= E[E[L|X]^2] - E[E[L|X]]^2 \\ &= E[L|X = 0]^2 \cdot p_{0.} + E[L|X = 1]^2 \cdot p_{1.} - E[L]^2 \\ &= \mu_0^2 \cdot p_{0.} + \mu_1^2 \cdot p_{1.} - (\mu_1 \cdot p_{1.} + \mu_0 \cdot p_{0.})^2. \end{aligned} \tag{A.9}$$

Inserting (A.9) and (A.8) into (A.7) yields, after some simplifications,

$$\mathrm{Var}[L] = (1 - p_{1.}) \cdot p_{1.} \cdot (\mu_1 - \mu_0)^2 + \sigma^2. \tag{A.10}$$

Finally, since X is a Bernoulli random variable, we have

$$\mathrm{Var}[X] = (1 - p_{1.}) \cdot p_{1.}. \tag{A.11}$$

Inserting (A.6), (A.10) and (A.11) into (A.1) gives

$$\begin{aligned}\mathrm{Corr}[L,X] &= \frac{p_{1\cdot}\cdot(1-p_{1\cdot})\cdot(\mu_1-\mu_0)}{\sqrt{[p_{1\cdot}\cdot(1-p_{1\cdot})\cdot(\mu_1-\mu_0)^2+\sigma^2]\cdot[p_{1\cdot}\cdot(1-p_{1\cdot})]}}\\ &= \frac{\mu_1-\mu_0}{\sqrt{(\mu_1-\mu_0)^2+\sigma^2/(p_{1\cdot}(1-p_{1\cdot}))}}\\ &= \frac{\Delta}{\sqrt{\Delta^2+1/(p_{1\cdot}(1-p_{1\cdot}))}},\end{aligned}$$

which proves (1).

To compute Δ, note that

$$\begin{aligned}p_{00} &= \Pr\{X=0\cap Y=0\}\\ &= \Pr\{Y=0|X=0\}\cdot\Pr\{X=0\}\\ &= \Pr\{L\le\gamma|X=0\}\cdot\Pr\{X=0\}\\ &= \Pr\left\{\frac{L-\mu_0}{\sigma}\le\frac{\gamma-\mu_0}{\sigma}\right\}\cdot\Pr\{X=0\}.\end{aligned}\tag{A.12}$$

If L is normally distributed under $X=0$, then the fraction $(L-\mu_0)/\sigma$ is a standard normal variable with distribution function Φ. Therefore we can rewrite (A.12) as

$$p_{00}=\Phi\left[\frac{\gamma-\mu_0}{\sigma}\right]\cdot p_{0\cdot}$$

from which one obtains

$$\frac{\gamma-\mu_0}{\sigma}=\Phi^{-1}\left[\frac{p_{00}}{p_{0\cdot}}\right].\tag{A.13}$$

Analogously, one derives

$$\frac{\gamma-\mu_1}{\sigma}=\Phi^{-1}\left[\frac{p_{10}}{p_{1\cdot}}\right].\tag{A.14}$$

From (A.13) and (A.14) one finds that

$$\begin{aligned}\Phi^{-1}\left[\frac{p_{00}}{p_{0\cdot}}\right]-\Phi^{-1}\left[\frac{p_{10}}{p_{1\cdot}}\right] &= \frac{\gamma-\mu_0}{\sigma}-\frac{\gamma-\mu_1}{\sigma}\\ &= \frac{\mu_1-\mu_0}{\sigma}\\ &= \Delta.\end{aligned}$$

Appendix B

Logistic approximation

Note that the inverse of the logistic cdf,

$$\Psi(x)=\frac{1}{1+\mathrm{e}^{-1.7x}}$$

is given by

$$\Psi^{-1}(x) = \frac{1}{1.7}\ln\left(\frac{x}{1-x}\right).$$

Thus the quantity Δ in Proposition 1 is

$$\begin{aligned}\Delta &= \Psi^{-1}(p_{00}/p_{0.}) - \Psi^{-1}(p_{10}/p_{1.}) \\ &= \frac{1}{1.7}\left[\ln\left(\frac{p_{00}/p_{0.}}{1-p_{00}/p_{0.}}\right) - \ln\left(\frac{p_{10}/p_{1.}}{1-p_{10}/p_{1.}}\right)\right] \\ &= \frac{1}{1.7}\ln\left(\frac{p_{00}p_{11}}{p_{10}p_{01}}\right).\end{aligned}$$

Inserting the last expression into (1) of Proposition 1 yields the desired result (3).

Appendix C

Standard error of $\hat{\nu}$

In general the standard error of any random variable Y, which is a composite of random variables $X_1, \ldots, X_n$, can be deduced if the partial derivative $\partial y/\partial x_i$ and the variance–covariance matrix of the X_i is known (Stuart & Ord, 1987, pp. 323–324). This general approach allows the computation of the standard error of $\hat{\nu}$. It can be shown that this standard error (SE) is given by

$$SE^2[\hat{\nu}] = \frac{1}{N}\cdot \boldsymbol{g}'\cdot \boldsymbol{V}\cdot \boldsymbol{g}, \tag{C.1}$$

where the quantities N, $\boldsymbol{V}$, $\boldsymbol{g}'$, and $\boldsymbol{g}$ will be explained below.

First, N is the total number of observations, i.e. $N = n_{00} + n_{01} + n_{10} + n_{11}$ (see Table 1).

Second, $\boldsymbol{V}$ denotes the variance–covariance matrix

$$\boldsymbol{V} = \begin{bmatrix} V_{00} & C_{00,01} & C_{00,10} & C_{00,11} \\ C_{00,01} & V_{01} & C_{01,10} & C_{01,11} \\ C_{00,10} & C_{01,10} & V_{10} & C_{10,11} \\ C_{00,11} & C_{01,11} & C_{10,11} & V_{11} \end{bmatrix}$$

of the estimates $\hat{p}_{00}$, $\hat{p}_{01}$, $\hat{p}_{10}$, and $\hat{p}_{11}$. The variance V_{ij} and covariance $C_{ij,kl}$ of these estimates are (Johnson, Kotz, & Balakrishnan, 1997)

$$V_{ij} = p_{ij} - p_{ij}^2 \tag{C.2}$$

and

$$C_{ij,kl} = -p_{ij}\cdot p_{kl}. \tag{C.3}$$

Finally, $\boldsymbol{g}'$ is the transpose of vector $\boldsymbol{g}$ containing the partial derivatives of ν with respect to the probabilities p_{00}, p_{01}, p_{10}, and p_{11}:

$$\boldsymbol{g}' = \left[\frac{\partial \nu}{\partial p_{00}}, \frac{\partial \nu}{\partial p_{01}}, \frac{\partial \nu}{\partial p_{10}}, \frac{\partial \nu}{\partial p_{11}}\right].$$

These partial derivatives are derived from (3) and it can be shown that they are given by

$$\frac{\partial \nu}{\partial p_{ij}} = \begin{cases} \dfrac{t \cdot (u \cdot p_{ij} + 2 \cdot p_{i.})}{p_{ij} \cdot p_{i.}}, & \text{for } i = j, \\ \dfrac{t \cdot (u \cdot p_{ij} - 2 \cdot p_{i.})}{p_{ij} \cdot p_{i.}}, & \text{for } i \neq j, \end{cases} \tag{C.4}$$

with

$$t = \frac{s \cdot 1.445}{\sqrt{[s^2 \cdot u^2 + 2.89]^3}},$$

$$u = \ln\left(\frac{p_{00} \cdot p_{11}}{p_{01} \cdot p_{10}}\right),$$

$$s = \sqrt{p_{ij} - p_{ij}^2}.$$

The computation of the standard error of $\hat{\nu}$ is demonstrated for the bivariate distribution shown in Table 2. We will assume that a random sample with $N = 100$ subjects is drawn from the population associated with this bivariate distribution. Applying (C.2) and (C.3) to the probabilities in Table 2 yields the variance–covariance matrix

$$\boldsymbol{V} = \begin{bmatrix} 0.13 & -0.04 & -0.02 & -0.08 \\ -0.04 & 0.19 & -0.03 & -0.13 \\ -0.02 & -0.03 & 0.09 & -0.05 \\ -0.08 & -0.13 & -0.05 & 0.25 \end{bmatrix}.$$

The vector $\boldsymbol{g}$ and its transpose $\boldsymbol{g}'$ are obtained from (C.4). Specifically, the transpose is

$$\boldsymbol{g}' = [\,2.01, \quad -0.66, \quad -2.27, \quad 0.73\,].$$

Finally the standard error is given by (C.1):

$$SE[\hat{\nu}] = \sqrt{\tfrac{1}{100} \cdot \boldsymbol{g}' \cdot \boldsymbol{V} \cdot \boldsymbol{g}}$$

$$= \sqrt{\tfrac{1.46}{100}} = 0.121.$$

Since $\hat{\nu}$ is the maximum likelihood estimate, it should be approximately normally distributed. Therefore, the 95% confidence interval for this example ranges from .06 to .54 and thus the observed $\hat{\nu} = .30$ deviates significantly from zero. (To compute the standard error of estimate $\hat{\nu}$ from the data of a random sample, one simply replaces the theoretical probabilities p_{00}, p_{01}, p_{10}, p_{11} by their estimates $\hat{p}_{00}$, $\hat{p}_{01}$, $\hat{p}_{10}$, and $\hat{p}_{11}$).

The accuracy of the standard error was verified by Monte Carlo simulations in which 30 000 independent random samples of size $N = 100$ were generated for each of several sets of bivariate distributions, that is, p_{00}, p_{01}, p_{10}, and p_{11}. For the above numerical example, the estimated standard error in the simulation study was 0.12 and thus was virtually identical to the one computed before. In almost all cases, the simulated and

computed standard errors yielded virtually identical results over a wide range of realistic cell probabilities.

For strongly correlated variables ($\nu > .75$) or very skewed marginal distributions ($p < .1$), however, (C.1) tends to slightly underestimate the true standard error. For example, the computed standard error for $\nu = .8$ (given symmetrical marginal distributions, $N = 100$) is equal to 0.045, whereas simulation yields a value of 0.051. It should be borne in mind, however, that for strong correlations and relatively small sample sizes, the sample distributions are not symmetric and thus the relevance of the standard error for computing confidence intervals is markedly reduced. Although in most applications such strong correlations are seldom encountered, the bootstrapping method (see Mooney & Duval, 1993) may be employed to construct confidence intervals in such extreme cases.

Appendix D

Maximum likelihood estimators

Without loss of generality, we proceed from $\sigma = 1$ and $\mu_0 = 0$, implying $\Delta = \mu_1$. Furthermore, we abbreviate $p \equiv p_{1.}$ and $\mu \equiv \mu_1$.

Proposition 2. *Let* n_{xy} *be the number of observations in cell* (x, y) *of a* 2×2 *contingency table with* $\mathrm{x} \in \{0,1\}$ *and* $\mathrm{y} \in \{0,1\}$. *Then*

$$\hat{\mu} = \Phi^{-1}\left[\frac{\mathrm{n}_{00}}{\mathrm{n}_{0.}}\right] - \Phi^{-1}\left[\frac{\mathrm{n}_{10}}{\mathrm{n}_{1.}}\right], \tag{D.1}$$

$$\hat{\gamma} = \Phi^{-1}\left[\frac{\mathrm{n}_{00}}{\mathrm{n}_{0.}}\right], \tag{D.2}$$

$$\hat{p} = \frac{\mathrm{n}_{1.}}{\mathrm{n}_{0.} + \mathrm{n}_{1.}}, \tag{D.3}$$

provide maximum likelihood estimators for μ, γ, *and* p, *respectively, with* $\mathrm{n}_{0.} = \mathrm{n}_{00} + \mathrm{n}_{01}$ *and* $\mathrm{n}_{1.} = \mathrm{n}_{10} + \mathrm{n}_{11}$.

Proof. Under Assumptions 1 and 2 the likelihood function ℓ is

$$\begin{aligned}\ell(\mu, \gamma, p; n_{00}, n_{01}, n_{10}, n_{11}) = &\left\{\Phi(\gamma) \cdot (1-p)\right\}^{n_{00}} \times \left\{\Phi(\gamma - \mu) \cdot p\right\}^{n_{10}} \\ &\times \left\{[1 - \Phi(\gamma)] \cdot (1-p)\right\}^{n_{01}} \times \left\{[1 - \Phi(\gamma - \mu)] \cdot p\right\}^{n_{11}}.\end{aligned}$$

As is now shown, however, it is easier to work with the logarithm $\ell^* \equiv \ln[\ell(\mu, \gamma, p; n_{00}, n_{01}, n_{10}, n_{11})]$ of ℓ, which can be written as

$$\begin{aligned}\ell^* = &\, n_{00} \cdot \left\{\ln[\Phi(\gamma)] + \ln[1-p]\right\} \\ &+ n_{10} \cdot \left\{\ln[\Phi(\gamma - \mu)] + \ln[p]\right\} \\ &+ n_{01} \cdot \left\{\ln[1 - \Phi(\gamma)] + \ln[1-p]\right\} \\ &+ n_{11} \cdot \left\{\ln[1 - \Phi(\gamma - \mu)] + \ln[p]\right\}.\end{aligned}$$

The point $(\hat{p}, \hat{\gamma}, \hat{\mu})$ where the likelihood function ℓ attains its maximum is a solution of the equations

$$\begin{aligned} \frac{\partial \ell^*}{\partial p} &= 0, \\ \frac{\partial \ell^*}{\partial \mu} &= 0, \\ \frac{\partial \ell^*}{\partial \gamma} &= 0. \end{aligned} \tag{D.4}$$

For p one computes

$$\frac{\partial \ell^*}{\partial p} = \frac{n_{1.}}{p} - \frac{n_{0.}}{1-p}$$

and obtains (D.3) under the constraint of (D.4). For μ one computes

$$\frac{\partial \ell^*}{\partial \mu} = \phi(\gamma - \mu) \cdot \left[\frac{n_{11}}{1 - \Phi(\gamma - \mu)} - \frac{n_{10}}{\Phi(\gamma - \mu)} \right], \tag{D.5}$$

where ϕ denotes the probability density function of a standard normal random variable. Note that under regular conditions the inequality $\phi(\gamma - \mu) > 0$ holds and hence the term in brackets in (D.5) must be equal to zero at the maximum of ℓ. Therefore, we may write

$$\frac{n_{11}}{1 - \Phi(\gamma - \mu)} - \frac{n_{10}}{\Phi(\gamma - \mu)} = 0, \tag{D.6}$$

from which follows the provisional result

$$\gamma - \mu = \Phi^{-1}\left(\frac{n_{10}}{n_{10} + n_{11}} \right). \tag{D.7}$$

Next, differentiate ℓ^* with respect to γ to obtain

$$\frac{\partial \ell^*}{\partial \gamma} = \phi(\gamma) \cdot \left[\frac{n_{00}}{\Phi(\gamma)} - \frac{n_{01}}{1 - \Phi(\gamma)} \right] + \phi(\gamma - \mu) \cdot \left[\frac{n_{10}}{\Phi(\gamma - \mu)} - \frac{n_{11}}{1 - \Phi(\gamma - \mu)} \right].$$

Because of (D.6) and $\phi(\gamma) > 0$, the last equation can only be equal to zero if the term in the first bracket is equal to zero. Hence, we may write

$$\frac{n_{00}}{\Phi(\gamma)} - \frac{n_{01}}{1 - \Phi(\gamma)} = 0,$$

and arrive at

$$\hat{\gamma} = \Phi^{-1}\left[\frac{n_{00}}{n_{0.}} \right], \tag{D.8}$$

which proves (D.2). Subtracting (D.7) from (D.8) yields (D.1).