PSYCHOMETRIKA—VOL. 75, NO. 2, 328–330
JUNE 2010
DOI: 10.1007/S11336-010-9155-7

A KRAEMER-TYPE RESCALING THAT TRANSFORMS THE ODDS RATIO INTO THE WEIGHTED KAPPA COEFFICIENT

MATTHIJS J. WARRENS

LEIDEN UNIVERSITY, THE NETHERLANDS

This paper presents a simple rescaling of the odds ratio that transforms the association measure into the weighted kappa statistic for a 2×2 table.

Key words: Cohen's kappa, 2×2 association measure.

1. Measures of 2×2 Association

In a validity study a dichotomous variable Y is often compared to a 'gold standard' variable X. For example, in a medical test evaluation one has a 'gold standard' evaluation of the presence/absence or type of a disease against which a test is assessed. A 2×2 study can be summarized in a table like Table 1 (Warrens, 2008a, 2008b, 2009). In Table 1, the four proportions a, b, c, and d characterize the joint distribution of the variables X and Y. The row and column totals are the marginal distributions that result from summing the joint proportions. We denote these by p_1 and q_1 for variable X and by p_2 and q_2 for variable Y (Warrens, 2008c, 2008d).

The odds ratio is a widely used measure of 2×2 association, and probably the most widely used measure in epidemiology (Edwards, 1963; Fleiss, 2003; Kraemer, 2004). The formula of the odds ratio in terms of proportions a, b, c, and d is $\mathrm{OR} = ad/bc$. The odds ratio is the ratio of the odds of an event occurring in one group to the odds of it occurring in another group. These groups might be any other dichotomous classification. An odds ratio of 1 indicates that the condition or event under study is equally likely in both groups. An odds ratio greater than 1 indicates that the event is more likely in the first group.

Another statistic of 2×2 association is the weighted kappa index (Spitzer, Cohen, Fleis, & Endicott, 1967; Vanbelle & Albert, 2009). It is the unique measure that is based on an acknowledgment that the clinical consequences of a false negative may be quite different from the clinical consequences of a false positive (Bloch & Kraemer, 1989; Kraemer, Periyakoil, & Noda, 2004). A real number $r \in [0, 1]$ must be specified a priori indicating the relative importance of false negatives to false positives. The sample estimator of the weighted kappa (Bloch & Kraemer, 1989) is

$$\kappa(r) = \frac{ad - bc}{rp_1q_2 + (1-r)p_2q_1}. \tag{1}$$

The measure $\kappa(1/2)$ is also known as Cohen's (1960) kappa (see also, Kraemer, 1979; Warrens, 2008e). Since the denominator of (1) can be written as

$$\begin{aligned} &rp_1q_2 + (1-r)p_2q_1 \\ &\quad = r(a+b)(b+d) + (1-r)(a+c)(c+d) \end{aligned}$$

Requests for reprints should be sent to Matthijs J. Warrens, Institute of Psychology, Unit Methodology and Statistics, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: warrens@fsw.leidenuniv.nl

TABLE 1.
Break-down of proportions for binary variables X and Y.

Proportions	$Y=1$	$Y=0$	Totals
$X=1$	a	b	p_1
$X=0$	c	d	q_1
Totals	p_2	q_2	1

$$
\begin{aligned}
&= rad + r\big(ab + b^2 + bd\big) + (1-r)ad + (1-r)\big(ac + c^2 + cd\big) \\
&= rad + rb(1-c) + (1-r)ad + (1-r)c(1-b) \\
&= ad - bc + rb + (1-r)c,
\end{aligned}
$$

Equation (1) is equivalent to

$$\kappa(r) = \frac{ad - bc}{ad - bc + rb + (1-r)c}. \tag{2}$$

2. Rescaling the Odds Ratio

The odds ratio is a measure without fixed endpoints. Under statistical independence the value of the odds ratio is 1, but all other values of the odds ratio lie between 0 and ∞. Several authors have therefore proposed rescalings of the odds ratio that transform the measure to a correlation-like codomain (Yule, 1900, 1912; Digby, 1983). The association measures by Yule (1900, 1912) and Digby (1983) have value 0 when two variables are statistically independent and maximum value 1 (perfect association).

Kraemer (1988) showed how 2×2 association measures like the sensitivity (a/p_1), specificity (d/q_1), predictive value of a positive Y (a/p_2), predictive value of a negative Y (d/q_2) and the four risk ratios can be transformed such that the new index has value 0 under statistical independence and a maximum value of 1. In each case, the new 2×2 measure coincides with a special case of the weighted kappa $\kappa(r)$.

Let $p > 0$ and $e > -1$ be real numbers. Kraemer (1988, p. 46) noted that any rescaling of the odds ratio of the form

$$\frac{\mathrm{OR}^p - 1}{\mathrm{OR}^p + e} \tag{3}$$

has fixed endpoints at 0 (random test) and 1 (perfect association). Kraemer (1988, p. 47) conjectured that "it is not possible to derive some simple rescaling of the odds ratio that simultaneously fixes the endpoints and corresponds to some quality index". However, it turns out that

$$\kappa(r) = \frac{\mathrm{OR} - 1}{\mathrm{OR} - 1 + (r/c) + ((1-r)/b)}. \tag{4}$$

The proof is as follows. Using $p = 1$ and $e = -1 + (r/c) + ((1-r)/b)$ in (3) we obtain (4). Multiplying both the numerator and denominator of (4) by bc we obtain (2), which is equivalent to (1). Hence, Equation (4) is a simple Kraemer-type rescaling of the odds ratio that transforms the association measure into the weighted kappa statistic for a 2×2 table, effectively proving Kraemer's conjecture to be false.

References

Bloch, D.A., & & Kraemer, H.C. (1989). 2×2 Kappa coefficients: Measures of agreement or association. *Biometrics*, *45*, 269–287.

Cohen, J.A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 213–220.

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

Edwards, A.W.F. (1963). The measure of association in a 2×2 table. *Journal of the Royal Statistical Society, Series A*, *126*, 109–114.

Fleiss, J.L. (2003). *Statistical methods for rates and proportions*. New York: Wiley.

Kraemer, H.C. (1979). Ramifications of a population model for κ as a coefficient of reliability. *Psychometrika*, *44*, 461–472.

Kraemer, H.C. (1988). Assessment of 2×2 associations: Generalization of signal-detection methodology. *The American Statistician*, *42*, 37–49.

Kraemer, H.C. (2004). Reconsidering the odds ratio as a measure of 2×2 association in a population. *Statistics in Medicine*, *23*, 257–270.

Kraemer, H.C., Periyakoil, V.S., & Noda, A. (2004). Tutorial in biostatistics: Kappa coefficients in medical research. *Statistics in Medicine*, *21*, 2109–2129.

Spitzer, R.L., Cohen, J., Fleiss, J.L., & & Endicott, J. (1967). Quantification of agreement in psychiatric diagnosis. *Archives of General Psychiatry*, *17*, 83–87.

Vanbelle, S., & Albert, A. (2009). Agreement between two independent groups of raters. *Psychometrika*, *74*, 477–491.

Warrens, M.J. (2008a). On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions. *Psychometrika*, *73*, 777–789.

Warrens, M.J. (2008b). On similarity coefficients for 2×2 tables and correction for chance. *Psychometrika*, *73*, 487–502.

Warrens, M.J. (2008c). Bounds of resemblance measures for binary (presence/absence) variables. *Journal of Classification*, *25*, 195–208.

Warrens, M.J. (2008d). On the indeterminacy of resemblance measures for (presence/absence) data. *Journal of Classification*, *25*, 125–136.

Warrens, M.J. (2008e). On the equivalence of Cohen's kappa and the Hubert-Arabie adjusted Rand index. *Journal of Classification*, *25*, 177–183.

Warrens, M.J. (2009). k-adic similarity coefficients for binary (presence/absence) data. *Journal of Classification*, *26*, 227–245.

Yule, G.U. (1900). On the association of attributes in statistics. *Philosophical Transactions of the Royal Society A*, *75*, 257–319.

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

Manuscript Received: 1 DEC 2009
Final Version Received: 12 DEC 2009
Published Online Date: 23 MAR 2010