PSYCHOMETRIKA—VOL. 75, NO. 1, 176–185
MARCH 2010
DOI: 10.1007/S11336-009-9138-8

INEQUALITIES BETWEEN KAPPA AND KAPPA-LIKE STATISTICS FOR $k \times k$ TABLES

MATTHIJS J. WARRENS

LEIDEN UNIVERSITY

The paper presents inequalities between four descriptive statistics that can be expressed in the form $[P - E(P)]/[1 - E(P)]$, where P is the observed proportion of agreement of a $k \times k$ table with identical categories, and $E(P)$ is a function of the marginal probabilities. Scott's π is an upper bound of Goodman and Kruskal's λ and a lower bound of both Bennett et al. S and Cohen's κ. We introduce a concept for the marginal probabilities of the $k \times k$ table called weak marginal symmetry. Using the rearrangement inequality, it is shown that Bennett et al. S is an upper bound of Cohen's κ if the $k \times k$ table is weakly marginal symmetric.

Key words: Cohen's kappa, Bennett, Alpert and Goldstein's S, Goodman and Kruskal's lambda, Scott's pi, upper bound, rearrangement inequality, nominal agreement.

1. Introduction

In this paper, we prove some inequalities between four statistics of rater agreement for nominal categories, namely Cohen's (1960) κ, Bennett, Alpert and Goldstein's (1954) S, Scott's (1955) π, and Goodman and Kruskal's (1954) λ. In general, these indices may be used to summarize the cross classification of two nominal variables with identical categories (Brennan & Prediger, 1981; Zwick, 1988; Krippendorff, 2004; De Mast, 2007). These $k \times k$ tables occur in various fields of science, including psychometrics, educational measurement, biometrics (Fleiss, 1975), map comparison (Visser & De Nijs, 2006), and content analysis (Krippendorff, 2004). We introduce the four statistics in the context of rater agreement.

Suppose that two raters each distribute m objects (individuals, things) among a set of k mutually exclusive categories. In addition, suppose that the categories are defined in advance. To measure the agreement among the two raters, a first step is to obtain a contingency table $\mathbf{N}$ with elements n_{ij}, where n_{ij} indicates the number of objects placed in category i by the first rater and in category j by the second rater. For notational convenience, let $\mathbf{P}$ be the table of the same size as $\mathbf{N}$ with elements $p_{ij} = n_{ij}/m$. Row and column totals

$$p_{i+} = \sum_{j=1}^{k} p_{ij} \quad \text{and} \quad p_{+j} = \sum_{i=1}^{k} p_{ij}$$

are the marginal probabilities of $\mathbf{P}$.

Suppose that the categories of the raters are in the same order, so that the diagonal elements p_{ii} of $\mathbf{P}$ reflect the proportion of objects put in the same categories by both raters. A straightforward and crude measure of agreement between the raters is the observed proportion of agreement

$$P = \sum_{i=1}^{k} p_{ii}.$$

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

© 2009 The Psychometric Society. This article is published with open access at Springerlink.com

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

TABLE 1.
Definitions of $E(P)$ for λ, S $(= C = \kappa_n)$, π and κ.

Statistic	Symbol	Definition
Goodman and Kruskal's (1954) λ	$E(P)_{\mathrm{G}}$	$\max_i\left(\frac{p_{i+}+p_{+i}}{2}\right)$
Bennett et al. (1954) S	$E(P)_{\mathrm{B}}$	$\frac{1}{k}$
Janson and Vegelius' (1979) C		
Brennan and Prediger's (1981) κ_n		
Scott's (1955) π	$E(P)_{\mathrm{S}}$	$\sum_{i=1}^{k}\left(\frac{p_{i+}+p_{+i}}{2}\right)^2$
Cohen's (1960) κ	$E(P)_{\mathrm{C}}$	$\sum_{i=1}^{k} p_{i+}p_{+i}$

There is general consensus in the fields of science where $k \times k$ tables are encountered that P is artificially high and should be corrected for agreement due to chance. The statistics studied in this paper incorporate chance agreement, and can be expressed in the form

$$\frac{P - E(P)}{1 - E(P)}, \tag{1}$$

where $E(P)$, called the expected proportion of agreement, is conditional on fixed marginals of $\mathbf{P}$, and 1 is the maximum value of P. Four definitions of $E(P)$ are presented in Table 1. Using $E(P)_{\mathrm{G}}$, $E(P)_{\mathrm{B}}$, $E(P)_{\mathrm{S}}$, and $E(P)_{\mathrm{C}}$ in (1), we obtain, respectively, Goodman and Kruskal's λ, Bennett et al. S, Scott's π, and Cohen's κ.

An inequality is a statement about the relative size of two statistics, e.g., $S \geq \pi$. In this paper, we prove several inequalities between λ, S, π, and κ. An ordering between the values of these statistics for rater agreement is frequently observed in practice. Some authors (Blackman & Koval, 1993; Warrens, 2008a, 2008b) have proved inequalities between λ, π, and κ for 2×2 tables (Warrens, 2008c, 2008d). In this paper, we formally prove the double inequalities $S \geq \pi \geq \lambda$ and $\kappa \geq \pi \geq \lambda$ for $k \times k$ tables.

The paper is organized as follows. In Sect. 2, some background of the statistics is discussed. In Sect. 3, the double inequalities $S \geq \pi \geq \lambda$ and $\kappa \geq \pi \geq \lambda$ are proved for $k \times k$ tables. In Sect. 4, the concept of weak marginal symmetry is introduced. Bennett et al. S is an upper bound of Cohen's κ if the marginals of $\mathbf{P}$ are weakly symmetric. Section 5 contains a discussion and an illustration of the derived inequalities.

2. Background

Although often used as merely descriptive measures, λ, S, π, and κ are based on different assumptions, and may therefore not be appropriate in all contexts. The assumptions are hidden in the different definitions of $E(P)$ (Table 1). An excellent review of the rationales behind S, π, and κ can be found in Zwick (1988). Following Krippendorff (1987) and Warrens (2008a), we distinguish three ways in which chance factors may operate: two, one, or no underlying continua.

Suppose the data are a product of chance concerning two different frequency distributions (Cohen, 1960; Krippendorff, 1987), one for each nominal variable. $E(P)_{\mathrm{C}}$ is the value of P under statistical independence. The expectation of p_{ii} under statistical independence is defined by the product of the marginal probabilities. $E(P)_{\mathrm{C}}$ can be obtained by considering all permutations of the observations of one of the nominal variables, while preserving the order of the observations of the other variable. For each permutation the value of P can be determined. The arithmetic mean of these values is $\sum_{i=1}^{k} p_{i+}p_{+i}$.

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

A second possibility is that there are no relevant underlying continua. $E(P)_{\mathrm{G}}$ simply focuses on the most abundant category. Alternatively, if each rater randomly allocates objects to categories, then for each rater, the expected marginal probability for each category is $1/k$. The probability that two raters assign, by chance, any object to the same category is $(1/k)(1/k) = 1/k^2$. Summing these probabilities over all categories, we obtain $k/k^2 = 1/k = E(P)_{\mathrm{B}}$.

Finally, there may be only one frequency distribution involved. First, suppose it is assumed that the frequency distribution underlying the two nominal variables is the same for both variables (Scott, 1955; Krippendorff, 1987). The expectation of p_{ii} must be either known or it must be estimated from p_{i+} and p_{+i}. Different functions may be used. For example, Scott (1955) proposed the arithmetic mean $(p_{i+} + p_{+i})/2$. If one would use the geometric mean $\sqrt{p_{i+}p_{+i}}$ instead, one obtains $E(P)_{\mathrm{C}}$. Alternatively, Brennan and Prediger (1981, p. 693) show that if only one rater randomly allocates objects to categories, the probability of chance agreement is also given by $E(P)_{\mathrm{B}} = 1/k$.

Although λ, S, π, and κ are based on different assumptions, Cohen's κ is by far the most popular index of rater agreement for nominal categories. Warrens (2008e) proved that the 2×2 kappa is equivalent to the Hubert–Arabie (1985) adjusted Rand index for cluster validation (cf. Steinley, 2004). The popularity of κ has led to the development of many extensions, e.g., multirater kappa (Fleiss, 1971; Conger, 1980), or fuzzy kappa (Dou, Ren, Wu, Ruan, Chen, Bloyet, & Constans, 2007). However, several authors have identified difficulties or paradoxes with κ's interpretation (see, e.g., Brennan & Prediger, 1981, Feinstein & Cicchetti, 1990, or Byrt, Bishop & Carlin, 1993 and the references therein).

Zwick (1988) notes that Bennett et al. S is equivalent to coefficient C proposed in Janson and Vegelius (1979, p. 260) and κ_n proposed in Brennan and Prediger (1981, p. 693). Furthermore, for $k = 2$, Bennett et al. S is equivalent to statistics discussed in Holley and Guilford (1964), Maxwell (1977) and Krippendorff (1987).

Brennan and Prediger (1981) argue that Cohen's κ and Scott's π on the one hand, and Bennett et al. S on the other hand, are appropriate in different contexts. These authors make a distinction between studies where the marginal probabilities are fixed a priori, or free to vary. Marginals are said to be "fixed" whenever the marginal probabilities are known to the rater before classifying the objects into categories. Brennan and Prediger (1981) find Cohen's κ appropriate in reliability studies, when marginal probabilities are fixed. When either or both of the marginals are free to vary, Brennan and Prediger (1981) suggest that κ is replaced by S.

3. Inequalities

In this section, we prove the double inequalities $S \geq \pi \geq \lambda$ and $\kappa \geq \pi \geq \lambda$. Three lemmas will be used; especially Lemma 1 will be used repeatedly. The result is similar to Proposition 4 in Warrens (2008a, p. 496).

Lemma 1. *Equation* (1) *is a decreasing function of* $E(P)$.

Proof: Let P_1^* and P_2^* be two chance-corrected versions of P with expectations $E(P)_1$ and $E(P)_2$, respectively. We have

$$
\begin{aligned}
&P_1^* \geq P_2^*,\\
&\frac{P - E(P)_1}{1 - E(P)_1} \geq \frac{P - E(P)_2}{1 - E(P)_2},\\
&(1 - P)E(P)_1 \leq (1 - P)E(P)_2,\\
&E(P)_1 \leq E(P)_2.
\end{aligned}
$$

This completes the proof. □

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

Lemma 1 is used in the proofs of Theorems 1 to 5. We first prove the inequality $\pi \geq \lambda$ in Theorem 1. Theorem 1 is believed to be new. Lemma 2 is used in the proof of Theorem 1. Inequality (3) is a special case of Abel's inequality (see, e.g., Mitrinović, 1964, p. 18).

Lemma 2. *If $a_1, \ldots, a_k$ are nonnegative real numbers that satisfy*

$$\sum_{i=1}^{k} a_i = 1, \tag{2}$$

then

$$\sum_{i=1}^{k} a_i^2 \leq \max_i(a_i). \tag{3}$$

Proof: Let the numbers $s_1, \ldots, s_k$ be given by

$$s_j = \sum_{i=1}^{j} a_i.$$

Note that $s_j \leq 1$ for all j, due to (2).

The left-hand side of (3) may be written in the form

$$\begin{aligned} \sum_{i=1}^{k} a_i^2 &= s_1 a_1 + (s_2 - s_1)a_2 + \cdots + (s_k - s_{k-1})a_k \\ &= s_1(a_1 - a_2) + s_2(a_2 - a_3) + \cdots + s_{k-1}(a_{k-1} - a_k) + s_k a_k. \end{aligned} \tag{4}$$

Without loss of generality, assume $a_1 \geq \cdots \geq a_k$. Since $s_j \leq 1$ for all j, and by assumption,

$$a_1 - a_2 \geq 0, \quad a_2 - a_3 \geq 0, \quad \ldots, \quad a_{k-1} - a_k \geq 0, \quad \text{and} \quad a_k \geq 0,$$

we have the sequence of inequalities

$$\begin{aligned} &s_1(a_1 - a_2) \leq a_1 - a_2, \\ &s_2(a_2 - a_3) \leq a_2 - a_3, \\ &\vdots \\ &s_{k-1}(a_{k-1} - a_k) \leq a_{k-1} - a_k, \\ &s_k a_k \leq a_k. \end{aligned}$$

Adding these k inequalities, we obtain

$$s_1(a_1 - a_2) + s_2(a_2 - a_3) + \cdots + s_{k-1}(a_{k-1} - a_k) + s_k a_k \leq a_1.$$

Using (4), we arrive at the inequality $\sum_{i=1}^{k} a_i^2 \leq a_1$. □

Theorem 1. $\pi \geq \lambda$.

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

Proof: Due to Lemma 1, it must be shown that

$$E(P)_{\mathrm{S}} \leq E(P)_{\mathrm{G}},$$

$$\sum_{i=1}^{k}\left(\frac{p_{i+}+p_{+i}}{2}\right)^2 \leq \max_i\left(\frac{p_{i+}+p_{+i}}{2}\right). \tag{5}$$

Using $a_i = (p_{i+}+p_{+i})/2$ in (3), we obtain (5). □

Using Lemma 1, it is not difficult to show that $\kappa \geq \pi$. The proof of Theorem 2 for $k = 2$, can be found in Blackman and Koval (1993, p. 216). Inequality (6) also follows from the arithmetic-geometric mean inequality (see, e.g., Mitrinović, 1964, p. 9; Hardy, Littlewood, & Pólya, 1988).

Theorem 2. $\kappa \geq \pi \geq \lambda$.

Proof: Since inequality $\pi \geq \lambda$ is proved in Theorem 1, the proof is limited to $\kappa \geq \pi$.

We have for all i,

$$\left(\frac{p_{i+}-p_{+i}}{2}\right)^2 \geq 0,$$

$$\left(\frac{p_{i+}+p_{+i}}{2}\right)^2 \geq p_{i+}p_{+i}. \tag{6}$$

Using (6), we have

$$\sum_{i=1}^{k} p_{i+}p_{+i} \leq \sum_{i=1}^{k}\left(\frac{p_{i+}+p_{+i}}{2}\right)^2,$$

$$E(P)_{\mathrm{C}} \leq E(P)_{\mathrm{S}}.$$

The desired inequality then follows from application of Lemma 1. □

Inequality (7) is used in the proofs of Theorems 3, 4, and 5. Lemma 3 is also known as the rearrangement inequality (see, e.g., Hardy, Littlewood, & Pólya, 1988, p. 261).

Lemma 3. *For two sets of nonnegative real numbers* $a_1 \leq \cdots \leq a_k$ *and* $b_1 \leq \cdots \leq b_k$ *and every permutation* $a_{\sigma(1)}, \ldots, a_{\sigma(k)}$ *of* $a_1, \ldots, a_k$, *it holds that*

$$a_k b_1 + \cdots + a_1 b_k \leq a_{\sigma(1)}b_1 + \cdots + a_{\sigma(k)}b_k \leq \sum_{i=1}^{k} a_i b_i. \tag{7}$$

We end this section by showing that Bennett et al. S is an upper bound of Scott's π. Theorem 3 and its proof are believed to be new. Inequality (9) is also know as the sum of squares inequality, and is a special case of the Cauchy–Schwarz inequality (see, e.g., Mitrinović, 1964, p. 20).

Theorem 3. $S \geq \pi \geq \lambda$.

Proof: Since inequality $\pi \geq \lambda$ is proved in Theorem 1, the proof is limited to inequality $S \geq \pi$.

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

Using $b_i = a_i$ in (7), we obtain

$$\sum_{i=1}^{k} a_i^2 \geq a_{\sigma(1)}a_1 + \cdots + a_{\sigma(k)}a_k. \tag{8}$$

Consider the $k-1$ variants of (8) such that each product $a_i a_j$ for all $i \neq j$ on the right-hand side occurs exactly twice. Adding these $k-1$ variants, and adding $\sum_{i=1}^{k} a_i^2$ to both sides of the result, we obtain

$$k\sum_{i=1}^{k} a_i^2 \geq \left(\sum_{i=1}^{k} a_i\right)^2. \tag{9}$$

Using (2) in (9), and dividing the result by k, we obtain

$$\sum_{i=1}^{k} a_i^2 \geq \frac{1}{k}. \tag{10}$$

Using $a_i = (p_{i+} + p_{+i})/2$ in (10), we obtain

$$\sum_{i=1}^{k}\left(\frac{p_{i+} + p_{+i}}{2}\right)^2 \geq \frac{1}{k}$$

$$E(P)_{\mathrm{S}} \geq E(P)_{\mathrm{B}}.$$

The result then follows from application of Lemma 1. □

4. Marginal Symmetry and Asymmetry

The inequalities presented in the previous section are valid for all $k \times k$ tables. In this section, we consider inequalities that are only valid if certain requirements are met. The conditions that we need for Theorems 4 and 5 are specified in the following definitions on marginal symmetry.

Definition 1. Table **P** is weakly marginal symmetric if the permutation that orders the marginal probabilities from lowest to highest is the same for the p_{i+} and the p_{+i}.

With regard to Definition 1, the term strong marginal symmetry may be used in the case that $p_{i+} = p_{+i}$ for all i. In contrast to symmetry, asymmetry has many (more than two) faces. Only the following definition of marginal asymmetry will be used.

Definition 2. Table **P** is marginal asymmetric if the permutation that orders the marginal probabilities p_{i+} from lowest to highest, orders the p_{+i} from highest to lowest.

First, we show that Bennett et al. S is an upper bound of Cohen's κ if **P** is weakly marginal symmetric (Definition 1). Theorem 4 and its proof are believed to be new.

Theorem 4. *If* **P** *is weakly marginal symmetric, then* $S \geq \kappa \geq \pi \geq \lambda$.

Proof: Since inequality $\kappa \geq \pi \geq \lambda$ is proved in Theorem 2, the proof is limited to $S \geq \kappa$.

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

Without loss of generality, assume that $p_{1+} \leq \cdots \leq p_{k+}$ and $p_{+1} \leq \cdots \leq p_{+k}$. Using $a_i = p_{i+}$ and $b_i = p_{+i}$ in (4), we obtain

$$\sum_{i=1}^{k} p_{i+}p_{+i} \geq p_{\sigma(1)+}p_{+1} + \cdots + p_{\sigma(k)+}p_{+k}. \tag{11}$$

Consider the k variants of (11) such that each product $p_{i+}p_{+j}$ for $i, j = 1, 2, \ldots, k$ on the right-hand side occurs exactly once. Adding these k variants and dividing the result by k, we obtain

$$\begin{aligned} &\sum_{i=1}^{k} p_{i+}p_{+i} \geq \frac{1}{k}\left(\sum_{i=1}^{k} p_{i+}\right)\left(\sum_{i=1}^{k} p_{+i}\right), \qquad (12)\\ &\sum_{i=1}^{k} p_{i+}p_{+i} \geq \frac{1}{k},\\ &E(P)_{\mathrm{C}} \geq E(P)_{\mathrm{B}}. \end{aligned}$$

The result then follows from application of Lemma 1. □

Next, we show that κ is an upper bound of S if **P** is marginal asymmetric (Definition 2). Theorem 5 and its proof are believed to be new.

Theorem 5. *If* **P** *is marginal asymmetric, then* $\kappa \geq S \geq \pi \geq \lambda$.

Proof: Since inequality $S \geq \pi \geq \lambda$ is proved in Theorem 3, the proof is limited to $\kappa \geq S$.

Without loss of generality, assume that $p_{1+} \leq \cdots \leq p_{k+}$ and $p_{+1} \geq \cdots \geq p_{+k}$. Using similar arguments as in the proof of Theorem 4, we obtain

$$\sum_{i=1}^{k} p_{i+}p_{+i} \leq \frac{1}{k}\left(\sum_{i=1}^{k} p_{i+}\right)\left(\sum_{i=1}^{k} p_{+i}\right),$$

instead of (12). The desired inequality then follows from application of Lemma 1. □

5. Discussion

Inequalities were derived between four descriptive statistics that can be expressed in the form $[P - E(P)]/[1 - E(P)]$, where P is the observed proportion of agreement of a $k \times k$ table with identical categories, and $E(P)$ is a function of the marginal probabilities. Scott's π is an upper bound of Goodman and Kruskal's λ (Theorem 1) and a lower bound of both Cohen's κ (Theorem 2) and Bennett et al. S (Theorem 3). Although the double inequalities $S \geq \pi \geq \lambda$ and $\kappa \geq \pi \geq \lambda$ have been observed frequently in applications, they have never been formally proved for $k \times k$ tables. References were provided if an inequality was already known for 2×2 tables.

In addition to inequalities $S \geq \pi \geq \lambda$ and $\kappa \geq \pi \geq \lambda$, two conditional inequalities between Bennett et al. S and Cohen's κ were derived. First, two concepts for the marginal probabilities of the $k \times k$ table were introduced. The $k \times k$ table is said to be weakly marginal symmetric if the permutation that orders the marginal probabilities from lowest to highest is the same for the row and column marginals. If the agreement table is weakly marginal symmetric, then $S \geq \kappa$ (Theorem 4). The $k \times k$ table is said to be marginal asymmetric if the permutation that orders the marginal probabilities of the rows from lowest to highest, orders the marginal probabilities of

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

TABLE 2.
Personality descriptions of oldest child by 200 sets of fathers and mothers (Cohen, 1960).

Father	Mother			Row
	Type 1	Type 2	Type 3	marginals
Type 1	0.44	0.05	0.01	0.50
Type 2	0.07	0.20	0.03	0.30
Type 3	0.09	0.05	0.06	0.20
Column marginals	0.60	0.30	0.10	1.00

TABLE 3.
Values of P, λ, S, π and κ and the corresponding $E(P)$'s, for the data presented in Table 2.

Statistic	P	$E(P)$	Value
Goodman and Kruskal's λ	0.70	0.55	0.333
Bennett et al. S	0.70	0.33	0.552
Scott's π	0.70	0.42	0.487
Cohen's κ	0.70	0.41	0.492

the columns from highest to lowest. If the agreement table is marginal asymmetric, then $\kappa \geq S$ (Theorem 5).

The paper is summarized in Theorems 4 and 5. If the agreement table is weakly marginal symmetric, then $S \geq \kappa \geq \pi \geq \lambda$. If the agreement table is marginal asymmetric, then $\kappa \geq S \geq \pi \geq \lambda$. To see statistics λ, S, π, and κ in action, we use the data in Table 2 from Cohen (1960). Two hundred sets of fathers and mothers were asked to identify which of three personality descriptions best describes their oldest child. Table 2 is the probability table of the cross classification of the fathers description and mothers description of the oldest child. For the data in Table 2, the values of the statistics are presented in Table 3. Note that the requirement of Theorem 4, Table 2 is weakly marginal symmetric, is satisfied. We have $S \geq \kappa \geq \pi \geq \lambda$, which illustrates Theorem 4.

The four chance-corrected statistics were originally derived using different assumptions and are thus appropriate in different situations. Cohen's κ is based on the assumption that the data are a product of chance concerning two different frequency distributions, one for each nominal variable, whereas for Scott's π it is assumed that the frequency distribution is the same for both nominal variables. The assumption of one underlying continuum is more restrictive than two underlying continua, and this is reflected in the inequality $\kappa \geq \pi$ (Theorem 2). The assumption of no relevant underlying continua is not necessarily a stronger condition than the assumption of one or two distributions. The expected proportion of agreement proposed in Goodman and Kruskal (1954) is the largest of the expectations in Table 1, and this is reflected in the inequality $\pi \geq \lambda$ (Theorem 1). Since Goodman and Kruskal's λ is the most conservative agreement statistic, it can be used as a lower bound to agreement if it is unclear what assumption is appropriate for the data at hand.

Section 4 introduced the concepts of weak marginal symmetry and marginal asymmetry for the marginal probabilities of the $k \times k$ table. Recall that Table 2 is weakly marginal symmetric. Furthermore, of the 9 square tables in Chapter 10 of Agresti (1990), 6 are weakly marginal symmetric and none are marginal asymmetric. An anonymous reviewer pointed out the paper by Agresti and Winner (1997). These authors evaluate agreement among 8 widely renowned movie reviewers and report kappa for all 28 pairs of reviewers. Of the 28 pairwise tables, 10 are weakly marginal symmetric and only 1 is marginal asymmetric. Thus, it appears that weak marginal symmetry is commonly observed in practice, but marginal asymmetry is not.

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

The study presented here was limited to four statistics that can be expressed in the form $[P - E(P)]/[1 - E(P)]$. Due to Lemma 1, comparing these four statistics is relatively easy. For future work, the four statistics can be compared to other statistics for $k \times k$ tables that cannot be expressed in the form $[P - E(P)]/[1 - E(P)]$. For example, Janson and Vegelius (1979) compare Cohen's κ to their coefficient S (Janson & Vegelius, 1979, p. 263), which is a generalization of the Phi coefficient (see, e.g., Warrens, 2008b) to $k \times k$ tables. They claim (p. 265) that the absolute value of Cohen's κ never exceeds the absolute value of their coefficient S.

Acknowledgements

The author would like to thank three anonymous reviewers for their helpful comments and valuable suggestions on an earlier version of this article.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
Agresti, A., & Winner, L. (1997). Evaluating agreement and disagreement among movie reviewers. *Chance*, *10*, 10–14.
Bennett, E.M., Alpert, R., & Goldstein, A.C. (1954). Communications through limited response questioning. *Public Opinion Quarterly*, *18*, 303–308.
Blackman, N.J.-M., & Koval, J.J. (1993). Estimating rater agreement in 2×2 tables: Correction for chance and intraclass correlation. *Applied Psychological Measurement*, *17*, 211–223.
Brennan, R.L., & Prediger, D.J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, *41*, 687–699.
Byrt, T., Bishop, J., & Carlin, J.B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, *46*, 423–429.
Cohen, J.A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 213–220.
Conger, A.J. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin*, *88*, 322–328.
De Mast, J. (2007). Agreement and kappa-type indices. *The American Statistician*, *61*, 148–153.
Dou, W., Ren, Y., Wu, Q., Ruan, S., Chen, Y., Bloyet, D., & Constans, J.-M. (2007). Fuzzy kappa for the agreement measure of fuzzy classifications. *Neurocomputing*, *70*, 726–734.
Feinstein, A.R., & Cicchetti, D.V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*, 543–548.
Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*, 378–382.
Fleiss, J.L. (1975). Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*, *31*, 651–659.
Goodman, G.D., & Kruskal, W.H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732–764.
Hardy, G.H., Littlewood, J.E., & Polya, G. (1988). *Inequalities* (2nd ed.). Cambridge: Cambridge University Press.
Holley, J.W., & Guilford, J.P. (1964). A note on the *G* index of agreement. *Educational and Psychological Measurement*, *24*, 749–753.
Hubert, L.J., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.
Janson, S., & Vegelius, J. (1979). On generalizations of the *G* index and the Phi coefficient to nominal scales. *Multivariate Behavioral Research*, *14*, 255–269.
Krippendorff, K. (1987). Association, agreement, and equity. *Quality and Quantity*, *21*, 109–123.
Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30*, 411–433.
Maxwell, A.E. (1977). Coefficients between observers and their interpretation. *British Journal of Psychiatry*, *116*, 651–655.
Mitrinović, D.S. (1964). *Elementary inequalities*. Noordhoff: Groningen.
Scott, W.A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, *19*, 321–325.
Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods*, *9*, 386–396.
Visser, H., & De Nijs, T. (2006). The map comparison kit. *Environmental Modelling & Software*, *21*, 346–358.
Warrens, M.J. (2008a). On similarity coefficients for 2×2 tables and correction for chance. *Psychometrika*, *73*, 487–502.
Warrens, M.J. (2008b). Bounds of resemblance measures for binary (presence/absence) variables. *Journal of Classification*, *25*, 195–208.

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press

Warrens, M.J. (2008c). On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions. *Psychometrika*, *73*, 777–789.
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.
Zwick, R. (1988). Another look at interrater agreement. *Psychological Bulletin*, *103*, 374–378.

Manuscript Received: 17 NOV 2008
Final Version Received: 27 MAY 2009
Published Online Date: 23 SEP 2009

https://doi.org/10.1007/s11336-009-9138-8 Published online by Cambridge University Press