EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT
1992, 52

THE EQUIVALENCE OF COHEN'S KAPPA AND PEARSON'S CHI-SQUARE STATISTICS IN THE 2 × 2 TABLE

MARCIA FEINGOLD
The University of Michigan, Ann Arbor

With two judges and a two-point rating scale, the test statistic for Kappa is the same as Pearson's chi-square statistic applied to the 2 × 2 table of paired observations. This equivalence allows a quick test of the null hypothesis of no agreement, as Pearson's chi-square statistic is much less cumbersome to compute than the Kappa statistic and its variance. A simple formula for the null hypothesis variance is also derived.

COHEN'S Kappa measures agreement in the situation where each subject in an experiment is rated on a nominal scale by two or more judges (Cohen, 1960). The test statistic is tedious to compute. The purpose of this study was to derive a much simpler, exact, formula that can easily be done with a hand calculator, for the case of two judges and a binary rating scale.

For n subjects, the observed frequencies can be displayed in the format of Table 1.

A gross, or unadjusted, measure of agreement is simply the proportion of times that the two judges agree. Cohen proposed Kappa (κ) in order to adjust the gross agreement by considering the extent of agreement that would occur by chance, because of each judge's overall, or marginal, assignments to each category of the rating scale. Chance agreement is defined as the proportion of times that the two judges would be expected to agree if their ratings were

Marcia Feingold is Visiting Scholar, The University of Michigan, Ann Arbor, MI. Please address correspondence to 352 Hilldale Drive, Ann Arbor, MI 48105.

TABLE 1
Arrangement of Paired Data for Assessing Agreement Between Judges

		Judge 2		
		+	−	Total
Judge 1	+	a	b	t_1
	−	c	d	t_2
	Total	s_1	s_2	n

independent of each other. The amount of chance agreement is also used to scale Kappa to lie between −1 and 1.

An estimate of Kappa, using the sample frequencies, is

$$\hat{\kappa} = (p_o - p_c)/(1 - p_c), \quad (1)$$

where the observed sample agreement is $p_o = (a + d)/n$ and the estimated agreement due to chance is $p_c = (s_1t_1 + s_2t_2)/n^2$. The approximate large sample variance of $\hat{\kappa}$, under the null hypothesis of no agreement, is (Fleiss, Cohen, and Everitt, 1969):

$$\hat{\mathrm{Var}}_o(\hat{\kappa}) = [p_c + p_c^2 - \Sigma\, s_it_i(s_i + t_i)/n^3]/[n(1 - p_c)^2]\,. \quad (2)$$

After some algebra (Appendix A), one finds that the test statistic for H_o is

$$\hat{\kappa}^2/\hat{\mathrm{Var}}_o(\hat{\kappa}) = n(ad - bc)^2/s_1s_2t_1t_2\,, \quad (3)$$

which is, of course, Pearson's chi-square statistic, X^2. Obviously, this formula is not the same as the Pearson's chi-square statistic that would be used to test the association between judges and ratings with non-paired data. For that situation, the data would be arranged as in Table 2, with rows representing the judges; columns, the ratings.

Some more algebra (presented in Appendix B) leads to the further useful relationship:

TABLE 2
Arrangement of Non-Paired Data for Assessing Association of Judges and Ratings

	+	−	Total
Judge 1	t_1	t_2	n
Judge 2	s_1	s_2	n

TABLE 3
Political Party Preferences in Britain

		Earliest Remembered Preferences		
		Conservative	Labour	Total
1964 Preferences	Conservative	15	5	20
	Labour	3	86	89
	Total	18	91	109

$$\hat{\text{Var}}_o^{-1}(\hat{\kappa}) = (n/4)(\hat{\psi} + 1/\hat{\psi} + 2)\,, \tag{4}$$

where $\hat{\psi} = s_2t_1/s_1t_2$, the estimate of the odds ratio if the matching is ignored, as when the data are arranged as in Table 2. Therefore, $\hat{\text{Var}}_o(\hat{\kappa}) \leq 1/n$, with near-equality for estimated odds ratios near one. It should be noted that $1/n$ can be used for a quick, conservative test of the null hypothesis. If $\hat{\kappa}^2 \times n$ is statistically significant, then the exact statistic will be.

Example

A study of political party identification in Great Britain (Butler and Stokes, 1969) contrasted earliest remembered party preferences with 1964 party allegiance (Table 3). The long way of computing Cohen's Kappa test statistic proceeds as follows, by using Equations 1 and 2:

$$p_o = (15 + 86)/109 = .926606,$$
$$p_c = (20 \times 18 + 89 \times 91)/109^2 = .711977, \text{ and}$$
$$\hat{\kappa} = (.926606 - .711977)/(1 - .711977) = .745180.$$
$$\hat{\text{Var}}_o(\hat{\kappa}) = \{.711977 + .711977^2 - [20 \times 18 \times 38 + 89 \times 91 \times 180]/109^3\}/[109(1 - .711977)^2] = .0091370.$$

Therefore, $\hat{\kappa}^2/\hat{\text{Var}}_o(\hat{\kappa}) = 60.7741$. Equation 3 affords a short approach to this result, yielding

$$X^2 = 109(15 \times 86 - 5 \times 3)^2/(20 \times 89 \times 18 \times 91)$$

$$= 60.7733.$$

For a quick computation of $\hat{\text{Var}}_o(\hat{\kappa})$, if needed, one computes

$$\hat{\psi} = (20 \times 91)/(18 \times 89) = 1.136080\,,$$

and, using Equation 4, one finds that

$$\hat{\text{Var}}_o^{-1}(\hat{\kappa}) = (109/4)(1.136080 + 1/1.136080 + 2) = 1/.0091370.$$

Even more quickly, one could have obtained a rough estimate of $\hat{\text{Var}}_o(\hat{\kappa})$ as $1/n = .0091743$.

REFERENCES

Butler, D. and Stokes, D. (1969). *Political change in Britain: Forces shaping electoral choice.* New York: St. Martin's Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 20, 37–46.

Fleiss, J. L., Cohen, J., and Everitt, B. S. (1969). Large sample standard errors of Kappa and weighted Kappa. *Psychological Bulletin*, 72, 323–327.

APPENDIX A

Equivalence of the Test Statistic for κ and Pearson's Chi-Square Statistic

From Equation 1 we have

$$\begin{aligned} p_o - p_c &= (a + d)/n - (s_1t_1 + s_2t_2)/n^2 \\ &= [an + dn - (a + b)(a + c) - (c + d)(b + d)]/n^2 \\ &= 2(ad - bc)/n^2 , \end{aligned}$$

and therefore

$$\hat{\kappa}^2 = 4(ad - bc)^2/[n^4(1 - p_c)^2] . \tag{5}$$

From Equation 2,

$$\begin{aligned} n^5(1 - p_c)^2\hat{\text{Var}}_o(\hat{\kappa}) &= n^4p_c + n^4p_c^2 - \Sigma ns_it_i(s_i + t_i) \\ &= n^2(s_1t_1 + s_2t_2) + s_1^2t_1^2 + 2s_1t_1s_2t_2 + s_2^2t_2^2 - \Sigma\, ns_it_i(s_i + t_i) . \end{aligned}$$

Replace the n^2 term with $(s_1 + s_2)(t_1 + t_2)$ and collect terms:

$$n^5(1-p_c)^2\hat{\mathrm{Var}}_o(\hat{\kappa}) = (s_1+s_2)(t_1+t_2)(s_1t_1+s_2t_2) + 2s_1t_1s_2t_2$$
$$+ \Sigma\, s_it_i(s_it_i - ns_i - nt_i)$$
$$= 4s_1t_1s_2t_2 + \Sigma\, s_it_i(2s_it_i - ns_i - nt_i + s_jt_i + s_it_j)\,.$$

Use the fact that $s_it_i - ns_i = -s_it_j$ and $s_it_i - nt_i = -s_jt_i$ for $i \neq j$, to obtain

$$n^5(1-p_c)^2\hat{\mathrm{Var}}_o(\hat{\kappa}) = 4s_1t_1s_2t_2\,. \tag{6}$$

Combine Equations 5 and 6 to show that

$$\hat{\kappa}^2/\hat{\mathrm{Var}}_o(\hat{\kappa}) = n(ad-bc)^2/s_1t_1s_2t_2\,.$$

APPENDIX B

The Variance of $\hat{\kappa}$ Under H_o

As the expected proportions in all cells of the 2 × 2 table sum to one,

$$1 - p_c = (s_2t_1 + s_1t_2)/n^2\,. \tag{7}$$

Combining Equations 6 and 7, we have

$$\hat{\mathrm{Var}}_o^{-1}(\hat{\kappa}) = n(s_2^2t_1^2 + s_1^2t_2^2 + 2s_1t_1s_2t_2)/4s_1t_1s_2t_2 \tag{8}$$
$$= (n/4)\,(\hat{\psi} + 1/\hat{\psi} + 2)\,,$$

where $\hat{\psi} = s_2t_1/s_1t_2$.