COM 425--Analysis of Face-to-Face Communication

Cohen's Kappa

Cohen's Kappa is a statistic that assesses interjudge agreement for nominally coded data. It can be applied at both the global level (i.e. for the coding system as a whole) and the local level (i.e. for individual categories). In either case, the formula is

$$\text{kappa} = \frac{p_o - p_c}{1 - p_c}$$

where p_o is the proportion of units that the two judges coded the same, and p_c is the proportion expected by chance. An equivalent formula, using frequencies, is

$$\text{kappa} = \frac{f_o - f_c}{N - f_c}$$

where f_o denotes the number (not proportion) of units coded similarly, f_c represents number of units that would be expected to be coded the same way by chance alone, and N is the number of units coded by either coder (i.e., if they code 50 units each, N = 50, not 100).

Attached, you will find Table 2 from Jacob Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, 1960, *20*, 37-46. This table gives a numerical example for both proportions and frequencies. In the example, the global reliability of a three-category coding system is calculated. If you look at the upper part of the table, you will see that 88 of the 200 units (44%) were coded Category 1 by both coders. Similarly, 14 of the units (7%) were judged to be Category 1 by Judge B, but Category 2 by Judge A. As can be seen, the units on which the judges agreed are displayed along the main diagonals, so p_o = .44 + .20 + .06 = .70, and f_o = 88 + 40 + 12 = 140. The numbers in parentheses are the expected proportions for the left-hand matrix, and the expected frequencies for the right-hand matrix. These numbers are obtained by multiplying the marginals and dividing by the grand total (1 for proportions, N for frequencies). Thus, for the proportions, .30 = (.60)(.50)/1; .09 = (.30)(.30)/1; .02 = (.10)(.20)/1. For frequencies, 60 = (120)(100)/200; 18 = (60)(60)/200; 4 = (20)(40)/200. Then, the total proportion expected by chance is p_c = .30 + .09 + .02 = .41; the total frequency expected by chance is f_c = 60 + 18 + 4 = 82. To get the final kappa, simply plug these numbers into the appropriate formula.

Observe that given expected frequencies, you can obtain expected proportions by dividing again by N. Thus .30 = 60/200; .09 = 18/200; .02 = 4/200. This is what I was doing in class, because my example was stated in terms of frequencies, but the formula I had handy was stated in terms of proportions. Now that you've got both formulas, there's no need to do the conversion: If your problem is stated in terms of frequencies, use the frequency formula; if it is stated in proportions, use the proportion formula.

To get local (category-by-category) reliabilities, you follow the same procedure, except that your matrices will only be 2 X 2 (presence or absence of the category). So if Judge A coded ten units as follows: 1-1-2-1-3-3-1-1-3-3, and Judge B obtained 1-1-1-2-3-1-1-2-1-1, the frequency matrix for Category 1 would be

		Judge A	
		1	not-1
Judge B	1	3	4
	not-1	2	1

A Cohen's kappa calculated on this matrix would tell you how reliable Category 1 was. Categories 2 and 3 would be handled similarly.

TABLE 2

Illustrative Agreement Matrix

a) Proportions

	Category	Judge A 1	Judge A 2	Judge A 3	p_{iB}
Judge B	1	.44(.30)*	.07	.09	.60
	2	.05	.20(.09)	.05	.30
	3	.01	.03	.06(.02)	.10
	p_{iA}	.50	.30	.20	$\sum p_i = 1.00$

b) Frequencies

	Category	Judge A 1	Judge A 2	Judge A 3	f_{iB}
Judge B	1	88(60)*	14	18	120
	2	10	40(18)	10	60
	3	2	6	12(4)	20
	f_{iA}	100	60	40	$N = 200$

$$p_o = .44 + .20 + .06 = .70$$

$$p_c = .30 + .09 + .02 = .41$$

$$\kappa = \frac{.70 - .41}{1 - .41} = .492 \quad \text{(Eq. 1)}$$

$$\kappa_M = \frac{(.50 + .30 + .10) - .41}{1 - .41} = .831 \quad \text{(Eq. 6)}$$

$$\sigma_\kappa = \sqrt{\frac{.70\,(1 - .70)}{200\,(1 - .41)^2}} = .055 \quad \text{(Eq. 7)}$$

95% confidence limits = .492 ± 1.96 (.055) = .384↔ .600

$$\sigma_{\kappa_0} = \sqrt{\frac{.41}{200\,(1 - .41)}} = .059 \quad \text{(Eq. 10)}$$

$$z = \frac{.492}{.059} = 8.34; \quad \kappa \text{ significant at } P < .001$$

$$f_o = 88 + 40 + 12 = 140$$

$$f_c = 60 + 18 + 4 = 82$$

$$\kappa = \frac{140 - 82}{200 - 82} = .492 \quad \text{(Eq. 2)}$$

$$\sigma_\kappa = \sqrt{\frac{140\,(200 - 140)}{200\,(200 - 82)^2}} = .055 \quad \text{(Eq. 8)}$$

$$\sigma_{\kappa_0} = \sqrt{\frac{82}{200\,(200 - 82)}} = .059 \quad \text{(Eq. 11)}$$

* Chance expectancy