Correct Formulation of the Kappa Coefficient of Agreement

William D. Hudson
Center for Remote Sensing and Department of Forestry, Michigan State University, East Lansing, MI 48824-1111
Carl W. Ramm
Department of Forestry, Michigan State University, East Lansing, MI 48824-1222

INTRODUCTION

SINCE ITS INTRODUCTION to the remote sensing community by Congalton *et al.* (1983), an increasing number of studies have utilized the Kappa coefficient of agreement as a measure of classification accuracy. It was recommended as a standard by Rosenfield and Fitzpatrick-Lins (1986). Their article is an excellent review of the Kappa coefficient, its variance, and its use for testing for significant differences. Unfortunately, a large number of erroneous formulas and incorrect numerical results have been published. This paper briefly reviews the correct formulation of the Kappa statistic.

Although the Kappa statistic was originally developed by Cohen (1960), most articles cite Bishop *et al.* (1975) as a source of formulation:

$$\hat{\kappa} = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+}\, x_{+i}}{N^2 - \sum_{i=1}^{r} x_{i+}\, x_{+i}}$$

where + represents summation over the index.

For computational purposes, the following form is often presented:

$$\hat{\kappa} = \frac{\theta_1 - \theta_2}{1 - \theta_2}$$

$$\text{where } \theta_1 = \sum_{i=1}^{r} x_{ii}/N \text{ and}$$

$$\theta_2 = \sum_{i=1}^{r} x_{i+}\, x_{+i}/N^2.$$

As indicated by Rosenfield and Fitzpatrick-Lins (1986), several earlier versions of the variance of Kappa are incorrect (Cohen, 1960; Spitzer *et al.*, 1967; Cohen, 1968; Everitt, 1968). The correct formulation is given by Fleiss *et al.* (1969). As presented by Bishop *et al.* (1975), the approximate large sample variance of Kappa is

$$\hat{\sigma}^2[\hat{\kappa}] = \frac{1}{N}\left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4}\right]$$

$$\text{where } \theta_1 = \sum_{i=1}^{r} x_{ii}/N,$$

$$\theta_2 = \sum_{i=1}^{r} x_{i+}x_{+i}/N^2,$$

$$\theta_3 = \sum_{i=1}^{r} x_{ii}(x_{i+} + x_{+i})/N^2, \text{ and}$$

$$\theta_4 = \sum_{\substack{i=1 \\ j=1}}^{r} x_{ij}\,(x_{j+} + x_{+i})^2/N^3.$$

Although the formulas as presented by Fleiss *et al.* (1969) and Bishop *et al.* (1975) appear substantially different, they are algebraically equivalent. Also note that the formulas for θ_1, θ_2, θ_3, and θ_4 used by Bishop *et al.* (1975) assume that proportions have been calculated for individual cells of the classification error matrix. Although the formulation of the Kappa statistic and its variance is correct, the numerical example provided by Bishop *et al.* (1975, p. 397) contains a numerical error. The correct value of θ_4 is 0.49536501 and the correct estimated variance is 0.00823495 (R.G. Oderwald, personal communication).

The potential user of the Kappa coefficient of agreement is cautioned that a number of remote sensing articles contain errors in the formula for the Kappa statistic or its variance. Although the erratum (*Photogrammetric Engineering and Remote Sensing*, Vol. 50, No. 10, p. 1477) for an article by Congalton *et al.* (1983) does contain the proper formulas, its appearance ten months after the original article has not been generally referenced when authors cite this work.

A number of published research results (Congalton and Mead, 1983; Congalton *et al.*, 1983; Benson and DeGloria, 1985) contain numerical errors in the reporting of the variance of the Kappa statistic. The errors appear to be caused by the imporper computation of the θ_4 term in a published computer program (Congalton *et al.*, 1981, 1982). Line 53 of this FORTRAN program uses the *i*th row total plus the *j*th column total instead of the *j*th row total plus the *i*th column total as specified by the formula for θ_4: i.e.,

$$\theta_4 = \sum_{\substack{i=1 \\ j=1}}^{r} x_{ij}\,(x_{j+} + x_{+i})^2/N^3$$

(line 53 should be corrected to read TH4 = TH4 + X (I,J) * (SXR (J) + SXC(I))**2; this error is not present in the current version of the program (R. G. Oderwald, personal communication)). Although the error in computing θ_4 is numerically small, and thus changes the variance term by only a very small amount, the continued use of the improper formula should be discouraged.

Although potentially very useful in remote sensing accuracy assessment, the user of the Kappa coefficient of agreement should be conscious of its correct formulation and numerical computation.

REFERENCES

Benson, A. S., and S. D. DeGloria, 1985. Interpretation of Landsat-4 Thematic Mapper and Multispectral Scanner Data for Forest Surveys. *Photogrammetric Engineering and Remote Sensing*, Vol. 51, No. 9, pp. 1281–1289.

Bishop, Y. M. M., S. E. Feinberg, and P. W. Holland, 1975. *Discrete Multivariate Analysis - Theory and Practice*. MIT Press, Cambridge, Mass. 575 p.

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46.

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 53, No. 4, April 1987, pp. 421–422.

0099-1112/87/5304–421$02.25/0

———, 1968. Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin*, Vol. 70, No. 4, pp. 213–220.

Congalton, R. G., and R. A. Mead, 1983. A Quantitative Method to Test for Consistency and Correctness in Photointerpretation. *Photogrammetric Engineering and Remote Sensing*, Vol. 49, No. 1, pp. 69–74.

Congalton, R. G., R. A. Mead, R. G. Oderwald, and J. Heinen, 1981. *Analysis of Forest Classification Accuracy*. Remote Sensing Research Report 81-1, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Congalton, R. G., R. G. Oderwald, and R. A. Mead, 1982. *Accuracy of Remotely Sensed Data: Sampling and Analysis Procedures*. Remote Sensing Research Report 82-1, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

———, 1983. Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques. *Photogrammetric Engineering and Remote Sensing*, Vol. 49, No. 12, pp. 1671–1678.

Everitt, B. S., 1968. Moments of the Statistics Kappa and Weighted Kappa. *The British Journal of Mathematical and Statistical Psychology*, Vol. 21, Part 1, pp. 97–103.

Fleiss, J. L., J. Cohen, and B. S. Everitt, 1969. Large Sample Standard Errors of Kappa and Weighted Kappa. *Psychological Bulletin*, Vol. 72, No. 5, pp. 323–327.

Rosenfield, G. H., and K. Fitzpatrick-Lins, 1986. A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. *Photogrammetric Engineering and Remote Sensing*, Vol. 52, No. 2, pp. 223–227.

Spitzer, R. L., J. Cohen, J. L. Fleiss, and J. Endicott, 1967. Quantification of Agreement In Psychiatric Diagnosis. *Archives of General Psychiatry*, Vol. 17, pp. 83–87.

(Received 25 September 1986; accepted 11 November 1986; revised 5 December 1986)

Comments on the Remote Sensing Brief entitled "Correct Formulation of the Kappa Coefficient of Agreement"

I WAS VERY HAPPY to see this brief written to clear up some of the confusion surrounding the use of the Kappa statistic. Since we introduced these techniques back in 1980 (Congalton *et al.*, 1980), they have enjoyed increasing popularity. Unfortunately, due to a mistake on my part, incorrect versions of the equations were published in the literature. I applaud Hudson and Ramm for writing this paper and settling once and for all this confusion. The corrections to the KAPPA computer program as described by Hudson and Ramm have indeed been made. In addition, all three programs (KAPPA, MARGFIT, and CONTABLE) have been converted to run on an IBM PC. If you would like copies of these programs or just an updated KAPPA program send a request and a blank diskette to me and I will be happy to provide you with this software.

REFERENCES

Congalton, R., R. Mead, and R. Oderwald. 1980. Statistical techniques for the analysis of Landsat classification accuracy. Presented at the ASPRS/ACSM Annual Meeting, St. Louis, Missouri.

—*Russell G. Congalton*
Assistant Professor
Dept. of Forestry and Resource Management,
University of California,
Berkeley, CA 94720

Sixth Thematic Conference on Remote Sensing for Exploration Geology

Houston, Texas
28–31 March 1988

This Sixth Thematic Conference in the continuing series of ERIM Remote Sensing Symposia will focus upon operational development, emphasizing *Applications*, *Technology*, and *Economics*. An industry-oriented technical program, including both conventional plenary sessions and poster presentations, is being formulated by an international program committee to address

- Introduction to Remote Sensing: Technology and Data Availability
- Remote Sensing for Independents
- Frontier Exploration
- Exploration in Mature Basins
- Direct Detection of Hydrocarbons
- Offshore Exploration
- Coal and Minerals
- Advanced Image Processing
- Engineering Logistics and Environmental Applications
- Remote Sensing for the 1990s

The program will also include hands-on workshops, field trips, and an exhibition of commercially available products and services, as well as non-commercial displays by governmental agencies, non-profit research organizations, and academic institutions. In addition, the guest program will feature an opening-night reception, daily refreshments, tours, entertainment, and sightseeing opportunities.

For further information please contact

Thematic Conference
Environmental Research Institute of Michigan
P. O. Box 8618
Ann Arbor, MI 48107
Tele. (313) 994-1200, ext. 3382