

## A computer program to determine interrater reliability for dichotomous-ordinal rating scales

ROGER N. BLOOR

*Neuropsychiatric Centre, Princess Alexandra Hospital Wroughton  
Swindon, Wiltshire, England*

Assessing reliability using standard statistical tests involves the adoption of assumptions of normality of distribution, meaningful agreement, negligible chance agreement, and meaningful total scores (Hall, 1974). The assumptions may not be valid in the particular case of rating scales, and Hall suggested the adoption of weighted Kappa (Cohen, 1968), using a linear weighting system rather than a quadratic system when the data are ordinal.

Cicchetti (1976) outlined the need to modify the linear weights in the case of ordinal-dichotomous systems when a point of "absence" is used on the scale. This modification allows a higher weighting for agreement in cases in which absence-presence confusion is not involved. Cicchetti also described a method of calculation of the minimal sample size needed to ensure the validity of Kappa.

The significance of a given weighted Kappa can be found from calculation of the variance of the coefficient. Fleiss, Cohen, and Everitt (1969) described errors in previously published methods for calculation of the variance of Kappa and outlined a valid formula for computing the variance of Kappa and weighted Kappa.

A number of BASIC programs have been developed for computer application of Kappa (Burns & Cavallaro, 1982; Watkins & Larimer, 1980). This paper presents a program in BASIC that utilizes the methods described by Hall (1974), Cicchetti (1976), and Fleiss et al. (1969) to calculate weighted Kappa for data from ordinal-dichotomous rating scales such as ward-rating scales or standardized interviews.

**Input.** The user specifies the number of points on the rating scale and the number of observer ratings to be considered. The user is then offered the option of enter-

The author's mailing address is: Neuropsychiatric Centre, Princess Alexandra Hospital Wroughton, Royal Air Force Wroughton, Swindon, Wilts SN4 0QJ, England.

ing data in the form of cell frequencies in a cross-tabulation matrix or as rated values for each observer pairing.

**Output.** The output consists of an  $n \times n$  matrix of observed frequency cells, with marginals, and a similar matrix of calculated linear agreement weights ( $n$  = number of rating points on scale). The following information is also displayed: weighted Kappa; variance of weighted Kappa; suggested minimal sample size for validity; percentage agreement (exact); and percentage agreement ( $\pm 1\%$ ).

**Limitations.** (1) The total amount of memory required is determined by the number of points on the rating scale: A 3-point scale requires 5.2K of RAM; A 15-point scale requires 12.5K of RAM. (2) The program is only valid for scales in which there is a point of "absence." Continuous ordinal scales require a separate weighting system.

**Language and Computer.** The program was written in MICROSOFT BASIC and was developed on a 16K 6502-based OSI machine. It uses no machine-specific routines and has been run on a 48K Z-80-based system with no modification.

**Availability.** Program listings are available at no charge from: Administration Office, Neuropsychiatric Centre, Princess Alexandra Hospital Wroughton, Royal Air Force Wroughton, Swindon, Wiltshire SN4 0QJ, England.

### REFERENCES

- BURNS, E., & CAVALLARO, C. A computer program to determine interobserver reliability statistics. *Behavior Research Methods & Instrumentation*, 1982, 14, 42.
- CICCHETTI, D. V. Assessing inter-rater reliability for rating scales: Resolving some basic issues. *British Journal of Psychiatry*, 1976, 129, 452-456.
- COHEN, J. Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 1968, 70, 213-220.
- FLEISS, J. L., COHEN, J., & EVERITT, B. S. Large sample standard errors of Kappa and weighted Kappa. *Psychological Bulletin*, 1969, 72, 323-327.
- HALL, J. N. Inter-rater reliability of ward rating scales. *British Journal of Psychiatry*, 1974, 125, 248-255.
- WATKINS, M. W., & LARIMER, L. D. Interrater agreement statistics with the microcomputer. *Behavior Research Methods & Instrumentation*, 1980, 12, 466.

(Revision accepted for publication October 26, 1983.)