Behavior Research Methods & Instrumentation
1980, Vol. 12(4), 466

Interrater agreement statistics with the microcomputer

MARLEY W. WATKINS
Deer Valley School District, Phoenix, Arizona 85007

and

LEON D. LARIMER
North Central Kansas Guidance Center, Manhattan, Kansas 66502

The simple percentage of agreement measure of interrater reliability has retained its popularity (Kelly, 1977) despite its failure to take into account the proportion of agreement due to chance. Consequently, misleading or insufficient reliability information has entered the professional literature. Demonstration of the role of chance in professional research was provided by Spitzer and Fleiss (1974) with a reanalysis of psychiatric diagnostic studies and by Watkins (1979) in a reanalysis of manuscript-reviewer data.

Chance-corrected solutions to the nominal scale classification problem have been developed by Cohen (1960) and Scott (1955). Both solutions utilize the formula Po – Pc/1 – Pc, where Po is the observed proportion of agreement and Pc is the probability of agreeing by chance alone. Scott's (1955) Pi and Cohen's (1960) Kappa are both widely used (Krippendorff, 1970), but they differ in the way expected frequencies (Pc) are computed.

A number of useful programs have been developed for computer applications of Kappa (Antonak, 1977; Berk & Campbell, 1976) and of Pi (Thornton & Croskey, 1975), but they are restricted to main-frame computers operating with the FORTRAN language. This paper presents a program that utilizes the BASIC language and a microcomputer to calculate both Pi and Kappa.

Description. The program is written in Applesoft BASIC for the Apple II microcomputer. It occupies 3.2K of RAM memory and will accommodate 20 variables for each 16K of user RAM memory. Variables are in mnemonic form, and the program is fully documented to allow adaptation to other popular microcomputers. Input is interactive and consists of the cross-tabulation matrix of ratings. Output includes Pi, Kappa, and critical values of Z for both statistics.

Availability. A source listing, complete with sample input and output, may be obtained from Marley Watkins, 1313 West Latham, Phoenix, Arizona 85007, upon receipt of a self-addressed envelope with 30 cents postage affixed.

REFERENCES

ANTONAK, R. F. A computer program to compute measures of response agreement for nominal scale data obtained from two judges. *Behavior Research Methods & Instrumentation*, 1977, **9**, 553.

BERK, R. A., & CAMPBELL, K. L. A FORTRAN program for Cohen's kappa coefficient of observer agreement. *Behavior Research Methods & Instrumentation*, 1976, **8**, 396.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, **20**, 37-46.

KELLY, M. B. A review of the observational data collection and reliability procedures reported in the *Journal of Applied Behavior Analysis*. *Journal of Applied Behavior Analysis*, 1977, **10**, 97-101.

KRIPPENDORFF, K. Bivariate agreement coefficients for reliability of data. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), *Sociological methodology*. San Francisco: Jossey-Bass, 1970.

SCOTT, W. A. Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, 1955, **19**, 321-325.

SPITZER, R. L., & FLEISS, J. L. A re-analysis of the reliability of psychiatric diagnosis. *British Journal of Psychiatry*, 1974, **125**, 341-347.

THORNTON, B. W., & CROSKEY, F. L. A computer program for calculating an index of interobserver reliability from timeseries data. *Educational and Psychological Measurement*, 1975, **35**, 735-737.

WATKINS, M. W. Chance and interrater agreement on manuscripts. *American Psychologist*, 1979, **34**, 796-798.

(Accepted for publication July 1, 1980.)

0005-7878/80/040466-01$00.35/0