A computer program to compute measures of response agreement for nominal scale data obtained from two judges

RICHARD F. ANTONAK

*Department of Education, University of New Hampshire
Durham, New Hampshire 03824*

Experiments in education and psychology often entail the assignment of a set of objects to one of several unordered categories by a pair of judges. It is assumed that the objects and categories are independent and that the judges function independently. The researcher then proceeds to study the similarities (or discrepancies) of the paired judgments.

Cohen (1960) presented the statistic kappa, a coefficient to measure the level of agreement between two judges, and a series of subsidiary statistics derived from kappa (i.e., confidence limits, maximum value, etc.). Light (1971) proposed various extensions of Cohen's kappa statistic as well as the statistic Δp , a measure to evaluate the pattern of agreement between two judges, and a means to test the significance of the statistic.

The program described in this paper calculates both the kappa (and associated statistics) and the Δp statistics, measures of response agreement for nominal scale data obtained from two judges.

Input. The job deck consists of the following cards: a problem parameters card, title cards, a data format card, data deck, and a last card. The problem parameters card indicates the number of unordered categories, the number of objects categorized, and user preference for optional data output. The next five cards describe the problem in a text which is reproduced on the output. The data format card is an I-type variable-format card which indicates the location, on the subsequent data cards, of the judges' categorization of the objects. The first data card (or set of cards) must contain the data from Judge 1, and the following card (or set of cards) must contain the data from Judge 2. If the user wishes to terminate the program, the last card must contain the word DONE punched in columns 1-4. On the other hand, if the user wishes to analyze another set of data, the last card is blank and the job deck is arranged

sequentially (as described above) beginning with the problem parameters card.

Output. The printed output includes (1) the titling information, (2) a contingency table of the two judges' categorizations of the objects, and (3) the user-requested measures of response agreement. The contingency table contains the number and proportion of objects which fall in each cell, the proportion of objects in each cell expected by chance association (the joint probability of the marginal proportions), and the marginal totals and proportions. If the user requests the kappa statistic, the printed output includes the values of (1) kappa, (2) the standard error of kappa, (3) a normal curve deviate for testing the significance of kappa, (4) 95% and 99% confidence limits, and (5) the maximum value of kappa determined by the marginal distributions. If the user requests the Δp statistic, the printed output includes the value of Δp and a test of the significance of this value.

Computer and Language. The program is written in FORTRAN 10 for processing by computers in the DEC-10 series and adaptable to CDC and IBM systems. Well documented, the program has variable names that are mnemonic and correspond to those used by Cohen (1960) and Light (1971) to facilitate modification by the user.

Restrictions. There cannot be more than 200 objects, and the number of categories, while variable, cannot exceed 11.

Availability. Copies of this paper and a source listing which includes data and printed output for two sample problems described by Cohen (1960) and Light (1971) may be obtained without charge by writing to Richard F. Antonak, Department of Education, University of New Hampshire, Durham, New Hampshire 03824.

REFERENCES

- COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, **20**, 37-46.
LIGHT, R. J. Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin*, 1971, **76**, 365-377.

(Accepted for publication August 26, 1977.)