

A computer program to determine interobserver reliability statistics

EDWARD BURNS and CLAIRE CAVALLARO
State University of New York, Binghamton, New York 13901

Determining reliability is always an important task when one is using observational data. Because there are several methods for evaluating the reliability of observers and because numerous estimates of reliability across time periods are frequently needed, a BASIC computer program was written that calculates interobserver reliability statistics for a series of time periods and for the sum of all observations across time periods.

Interobserver reliability statistics are generated using the following 2 by 2 data format, in which 0 signifies the nonoccurrence of a target behavior by an observer and 1 indicates the occurrence of a target behavior by an observer:

Observer 2	Observer 1	
	0	1
1	a	b
0	c	d

The five interobserver reliability statistics produced are as follows: (1) percent agreement, or $(b + c)/T * 100$, where T is the sum of all frequencies in a 2 by 2 table; (2) effective percent agreement (occurrence), or $b/(a + b + d) * 100$ (Hartmann, 1977); (3) effective percent agreement (nonoccurrence), or $c/(a + c + d) * 100$; (4) kappa (Cohen, 1960), which is defined as $(p_o - p_c)/(1 - p_c)$, where p_o is determined by $(b + c)/T$ and p_c is determined by $(a + b) * (b + d) / T^2 + (a + c) * (c + d) / T^2$; and (5) phi, which is calculated by $(b * c - a * d) / [(a + b) * (c + d) * (a + c) * (b + d)]^{.5}$.

The program provides interobserver reliability statistics for individual problems composed of n number of time periods and, when desired, total statistics across a series of problems. In addition, there are several program checks used to verify data.

Input. The user specifies whether data are for a new problem or for a new sequence of problems or whether total reliability statistics for all problems within a sequence are desired. Data are entered on one line, according to the following format: a, b, c, d.

Output. The output consists of a 2 by 2 table, with row, column, and total sums and interobserver reliability indexes of percent agreement, effective percent agreement (occurrence), effective percent agreement (non-occurrence), kappa, and phi. Following the output for each problem, a new problem or new sequence of problems can be initiated, total reliability statistics can be requested, or the program can be terminated.

Availability. The computer program is written in IBM VS BASIC (International Business Machines, 1976), using the IF-GOTO option. The program and sample input and output are available upon request at no cost from Claire Cavallaro, School of General Studies and Professional Education, State University of New York, Binghamton, New York 13901.

REFERENCES

- COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, 20, 37-46.
- HARTMANN, D. P. Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, 1977, 10, 103-116.
- INTERNATIONAL BUSINESS MACHINES. *IBM VS BASIC manual* (3rd ed.). San Jose, Calif: Author, 1976.

(Accepted January 20, 1982.)