A FORTRAN IV program to compute reliability of ratings for two or more judges

GERARD J. SMITS

Boston University, Boston, Massachusetts 02215

The program described here is based on procedures outlined in Winer (1971) in which the reliability of the ratings (or scaling) of stimuli is obtained. An analysis of variance approach is used to partition variance into that attributable to the rated objects and to the raters. The larger the variance attributable to the objects, the larger is the resulting reliability coefficient.

Input. Two parameters, the number of raters and the number of objects to be rated, must be specified. Data is input one card per subject, with each card containing a rating for each of the stimulus objects. The format statement is presently set to handle up to a 10-point rating scale, although this may be easily modified.

Output. An analysis of variance table is printed, which includes SS, df, and MS for the several sources

of variance. In addition, the average rater reliability of the individual rater, as well as the reliability of the mean ratings of all raters, is printed.

Computer. The program was tested on an IBM 370/158. The core memory requirement is approximately 4K bytes, and takes approximately 1.2 sec to run.

Limitations. The present program was set up to handle the ratings of 10 judges on 50 objects. This, however, may be easily modified, and is limited only by the size of the available core memory.

Availability. A program listing can be obtained without charge from Gerard J. Smits, Department of Psychology, Boston University, 64 Cummings Street, Boston, Massachusetts 02215.

REFERENCE

- WINER, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971. Pp. 283-289.

(Accepted for publication July 6, 1978.)