

## ICC: A BASIC program for the calculation of intraclass correlations based on fixed effects and random effects models

MICHAEL J STRUBE

*Washington University, St. Louis, Missouri*

The intraclass correlation coefficient (ICC) represents a commonly used measure of reliability among a set of  $k$  judges when data are at least interval in nature (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Hays, 1981). The program described here calculates four ICC estimates based on a subject  $\times$  judge matrix of observations, where each judge observes all subjects. The four estimates correspond to single-judge and  $k$ -judge ICCs based on a fixed effects and a random effects model (see Lahey, Downey, & Saal, 1983; Shrout & Fleiss, 1979). The fixed effects model assumes that the  $k$  judges are the only ones of interest, whereas the random effects model assumes that the  $k$  judges have been randomly sampled from a population to which inferences will be made. The single-judge estimate represents the expected reliability of a single judge and is appropriate when the researcher plans to use only a single judge for measurement purposes (e.g., in order to save resources). The  $k$ -judge estimate is the expected reliability of the group of  $k$  judges and is used when multiple judges are to contribute to measurement (e.g., ratings are to be averaged). This latter estimate is analogous to an internal consistency estimate calculated on a multiple item test.

**Input.** The data can be entered directly into the program or via disk using a previously created file. A missing data indicator can also be specified. This indicator is

used to produce a reduced matrix of complete data upon which the analyses are based (i.e., case deletion of missing data). If data are entered directly into the program, the user has the option of saving the matrix on disk. The data can be listed, and users have the option of correcting errors after the data have been entered.

**Output.** In addition to ICC estimates, the output includes means and standard deviations for all judges, and the ANOVA summary table. When there are more than two judges, the intercorrelations among judges are also printed. These correlations and the ANOVA summary table serve as useful interpretive aids. As Lahey et al. (1983) point out, a low ICC value can result from three different sources: (1) low variation in the data matrix (i.e., nonsignificant subjects effect), (2) low correlations among the  $k$  judges, and (3) a heterogeneous pattern of correlations among the judges (i.e., judge  $\times$  subject interactions). In the case of the random effects model, large variability among the judges can also attenuate the ICC.

**Availability.** The program is written in BASIC and is IBM PC compatible. A listing of the program is available from the author.

### REFERENCES

- CRONBACH, L. J., GLESER, G., NANDA, H., & RAJARATNAM, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- HAYS, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.
- LAHEY, M. A., DOWNEY, R. G., & SAAL, F. E. (1983). Intraclass correlations: There's more than meets the eye. *Psychological Bulletin*, **93**, 586-595.
- SHROUT, P. E., & FLEISS, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420-428.

Address correspondence to Michael J Strube, Department of Psychology, Washington University, St. Louis, MO 63130.

(Revision accepted for publication October 22, 1985.)