PROGRAM ABSTRACTS/ALGORITHMS

GSCALE: Scalogram analysis for dichotomously scored items

MARILYN D. WANG

*University of Maryland Baltimore County
Baltimore, Maryland 21228*

The technique of scalogram analysis introduced by Guttman (1947, 1950) permits one to test the hypothesis that a set of items forms a unidimensional cumulative scale. When a set of items forms a perfect Guttman scale, it is possible to reproduce a subject's response to each item from knowledge of his total score. Although scalogram analysis has been most extensively used as a method of attitude scaling, the model is also appropriate for the measurement of abilities, particularly when it is assumed that the items form a developmental scale (Anastasi, 1976). In the latter case, items are usually scored on a pass-fail basis and the scalogram model implies that an ideal response pattern is one in which a subject who passes K_i items on a K -item test passes the K_i least difficult items and fails the $K - K_i$ most difficult items.

GSCALE is a FORTRAN IV program which performs a preliminary scalogram analysis on a set of dichotomously scored items. Items are ranked in terms of difficulty (or popularity) and each subject receives a score equal to the number of items passed. Although some have advocated methods of scalogram analysis which maximize the fit between the sample data and the scalogram model (cf. Chevan, 1972; White & Saltz, 1957), Green (1956) has pointed out that, when Guttman's model holds, the items must be ranked by difficulty (or popularity) and that little is usually gained by seeking optimal solutions. GSCALE is therefore most appropriately used when a set of items is initially being screened for scalability.

Input. Input consists of (1) a 75-character title; (2) N , the number of subjects, K , the number of items, and a 0 or 1 to request or suppress optional output; (3) a variable format statement for reading the data matrix; (4) a subjects by items data matrix.

Output. Output consists of (1) a rank ordering of the items with the number of subjects passing each item; (2) a set of summary statistics described below; (3) an optional subject summary consisting of a reordered data matrix, with subjects identified and ranked by total score and items ordered from most to least difficult.

Development of this program was supported by the Computer Science Center of the University of Maryland. Requests for reprints should be sent to the author at the Department of Psychology, University of Maryland Baltimore County, 5401 Wilkens Avenue, Baltimore, Maryland 21228.

Summary Statistics. The most important descriptive statistic in scalogram analysis is the coefficient of reproducibility, R , which is a measure of the extent to which the items form a perfect scale. R is a function of the number of responses that would have to be changed in order for each subject's response pattern to conform to an ideal response pattern. Each such response is considered a scale error and $R = 1 - E/NK$, where N is the number of subjects, K is the number of items, and E is the total number of scale errors for all subjects. A more detailed description of the error counting procedure is given by Green (1956).

The standard error of R is calculated using a formula developed by Goodman (1959, p. 37), and the 95% confidence interval for R is given. This permits a test of the two-tailed hypothesis that the sample R differs significantly from some a priori value. Guttman (1950) recommended that R be at least .9 for a set of items to be considered scalable.

Chilton (1969) has noted that the most widely used computer programs for scalogram analysis do not incorporate statistical tests of the hypothesis that the obtained value of R is significantly higher than the value which would be expected if the items were statistically independent. Exact calculation of chance reproducibility requires generation of 2^K possible response patterns, calculation of the number of scale errors associated with each pattern, and estimation of the number of subjects showing each response pattern when the items are mutually independent. However, a useful approximation has been developed by Green (1956), and Chilton (1969) has shown that Green's coefficient, R_I , gives highly accurate estimates for $K \leq 10$. Although the accuracy of Green's estimate for longer tests has not been determined, R_I , in conjunction with the confidence interval for R , will permit an approximate test of the null hypothesis of item independence. Since the upper limit of R is 1, Green (1956) suggested calculation of an index of consistency $[I = (R - R_I)/(1 - R_I)]$, which equals 1 when the items are perfectly scalable and which has an expected value of 0 when the items are independent. Green suggested that items be considered scalable if $I \geq .5$.

A similar approach to evaluating the scalability of items involves comparing R with its minimum possible value. Minimum marginal reproducibility, R_M , is a function of the number of persons passing each item (Guttman, 1950). Scalability is then measured by $S = (R - R_M)/(1 - R_M)$. Since R_I will invariably be larger than R_M , the index of consistency provides a more conservative measure of scalability.

Program Limitations. The program accepts responses of 200 subjects to 50 items, but these maxima may easily be modified by changing the appropriate dimension statements. Generation of the optional reordered

data matrix requires calculation of a floating point number equal to 2^K , which may exceed the capacity of some small computers. The program was written for a UNIVAC 1108 and requires 30,000 locations. Execution time is primarily a function of the number of subjects and, for 50 items, it ranges from 7 sec for 100 subjects to 23 sec for 200 subjects.

Availability. A documented source listing and sample run are available from the author at no cost.

REFERENCES

- ANASTASI, A. *Psychological testing* (4th ed.). New York: Macmillan, 1976.
- CHEVAN, A. Minimum-error scalogram analysis. *Public Opinion Quarterly*, 1972, **30**, 379-387.
- CHILTON, R. J. A review and comparison of simple statistical tests for scalogram analysis. *American Sociological Review*, 1969, **34**, 238-245.
- GOODMAN, L. A. Simple statistical methods for scalogram analysis. *Psychometrika*, 1959, **24**, 29-43.
- GREEN, B. F. A method of scalogram analysis using summary statistics. *Psychometrika*, 1956, **21**, 79-88.
- GUTTMAN, L. The Cornell technique for scale and intensity analysis. *Educational and Psychological Measurement*, 1947, **7**, 247-279.
- GUTTMAN, L. The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Gardner (Eds.), *Measurement and prediction*. Princeton: Princeton University Press, 1950.
- WHITE, B. W., & SALTZ, E. Measurement of reproducibility. *Psychological Bulletin*, 1957, **54**, 81-99.