SEQUIN: A computerized item selection procedure

UDO W. POOCH
Texas A & M University, College Station, Texas 78743

and

WILLIAM MOONAN
U.S. Naval Personnel Research Laboratory
San Diego, California 92152

SEQUIN (Sequential Item Nominator) can be used to determine if the number of items of a test can be reduced without significantly reducing certain desirable psychometric properties of the test, such as validity and internal consistency reliability. This objective can only be realized if desirable properties of the test results are not lost to any significant degree in the process of using fewer items. There even exists, in certain cases, the possibility of increasing the value of the test by reducing the number of items.

The feasibility of making improvements in a test or testing program is established by the existence and availability of digital computers. By combining this device with standard psychometric theory, item reduction possibilities can be examined quickly. Essentially, we use more information than can be or typically is used by psychometricians to develop an item selection procedure. This information has always been available (Gulliksen, 1950), but has been difficult and time consuming to use.

The information referred to is that associated with measuring interitem relationships. Psychometricians usually use item difficulty and some measure of item-criterion relationship (usually point-biserial or biserial correlations to assemble items for a test (Guilford, 1936). The proposed procedure uses this information plus the interitem relationships. Psychometricians usually do not use interitem information. If "I" is the number of items available and one criterion is used, there are $(I + 2)(I + 3)/2$ unique sums, sums of squares and cross products of responses (or transformations of these numbers) of the subjects available. Of these psychometricians usually use $2I + 3$. For about 50 items, this results in about 7.5% use of available information. It might be argued (perhaps with some truth) that the amount used is the most crucial, but this conjecture is difficult to substantiate without a complete analysis.

ITEM SELECTION STRATEGY

The objective of SEQUIN is to investigate the feasibility of reducing assessment times of tests by reducing the number of items in the test. This objective could be stated more directly in the form of an optimization problem. For this problem, we would seek the number of items which, in general, optimizes some function (called the objective function) of validity, internal consistency reliability, and perhaps other variables (Guilford, 1954).

The objective function used in the SEQUIN program is merely a function of the differences of validity of tests constructed from k and $k - 1$ items (Taylor, 1950). This function may be expressed as the difference of product-moment correlations of the tests and the criterion variable (DuBois, 1942).

The problem essentially is a combinatorial one, for there are $2^I - 1$ ways of combining I items into all possible subtests. For $I = 30$, $2^I - 1 = 1{,}073{,}741{,}823$, which is a number commanding some respect even for a computer solution. Obviously, not all of these combinations can be searched. One practical way of proceeding and the method SEQUIN utilizes is to use a strategy similar to, but arithmetically less complicated than, the accretion procedure of statistical multiple regression analysis (Draper & Smith, 1966). This procedure is also suboptional in the sense that it does not necessarily produce optimal combinations, but SEQUIN does determine solutions which are significant improvements over those derived by human examination of basic data or other procedures believed to be commonly in use.

INPUT/OUTPUT

The information contained in the input form consists of user identification, number of external and internal criteria, number of cards per S, S cards, and criteria item difficulty values for the inclusion of items in the analysis. Input data (item responses and criterion variables) for each are in IBM card format or may be placed on magnetic tape.

SEQUIN uses two types of criterion variables. These are described as either "internal" or "external" criteria. Internal criteria are those scores created by SEQUIN from the item responses. These may be so-called "total" scores. External criteria are those scores which are not (necessarily) functions of the item responses and are available as input data. Examples of these are ratings of on-the-job performance of the men, final class standings, or their scores from other measuring instruments. One restriction is that the total number of the criteria must be less than or equal to 10.

SEQUIN also makes a sequence check for the input cards if there are multiple cards per man. Any man whose cards are out of order is eliminated from the analysis. The identification of these cards is printed for the user's information.

SEQUIN allows up to three correct responses for each item that is listed in the so-called KEY CARDS of the input form, and up to 180 items may be analyzed with SEQUIN. The KEY CARDS require the user to specify with which criteria each item is to be processed.

SEQUIN first outputs the identification of the men whose data cards are out of order and the identification of those items, if any, whose difficulties exceed or are less than the difficulty range specified on the input form. Next, the user's name is printed as well as the title, comments, and data of the run request. SEQUIN also specifies the "sample size," that is, the number of men used in the analysis.

The main statistical output of SEQUIN, which is in tabular form, consists of number of items in each "test," designation of the external or internal criteria, item difficulty (Carter, 1942), and point-biserial and biserial correlation coefficients (DuBois, 1942). Additionally, the cumulative correlation, which is equivalent to the validity of the "test" created by the item selection process, internal consistency reliability (Adkins, 1938), mean, standard deviation, and standard error of measurement (Fischer, 1948) complete the table.

COMPUTER

There have not been a sufficient number of runs on IBM 360/65 to develop an accurate timing equation. Execution time depends on the computer used, the sample size, number of items, number of criteria used, and the average item difficulty of the test. At present SEQUIN is operating on the IBM 360/65 computer and is written in FORTRAN IV compiler language.

AVAILABILITY

A complete input form and FORTRAN listing is available free upon request.

REFERENCES

Adkins, D. C. A rational comparison of item-selection techniques. Psychological Bulletin, 1938, 35, 655.

Carter, H. D. How reliable are the common means of difficulty and validity of objective test items? Journal of Psychology, 1942, 13, 31-39.

Draper, N., & Smith, H. *Applied regression analysis*. New York: Wiley, 1966. Pp. 234-260.

DuBois, P. H. A note on the computation of biserial r in item validation. Psychometrika, 1942, 7, 143-146.

Fischer, R. A. *Statistical methods for research workers*. 10th ed. London. Olives & Boyd, 1948. P. 356.

Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936. P. 566.

Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954. Pp. 373-464.

Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950. Pp. 363-394.

Taylor, C. W. Maximizing predictive efficiency for a fixed total testing time. Psychometrika, 1950, 15, 391-406.

(Received for publication November 16, 1973; accepted November 23, 1973.)

SUMRAT9: Generalized item analyses for educational and psychological scales

RICHARD G. LANDRY and KENNETH V. PETERSON

University of North Dakota
Grand Forks, North Dakota 58201

SUMRAT9 is a FORTRAN program designed to provide multiple item analyses for summated rating or Likert-type scales of any format. It can also be used with dichotomously scored items (yes-no; right-wrong), semantic differential-type scales, and various types of Q or R sorts. The uniqueness of this program stems from the fact that it provides (1) as many item analyses as requested by the investigator, (2) two forms of discrimination indices, and (3) printouts of all steps in the calculation process.

INPUT

SUMRAT can be used with any type of scale data, provided individual item responses are on a numerical continuum (e.g., 1-0, 1-5, or 1-11). Selector cards at the end of the data deck indicate how many item analyses are to be performed and which items belong on which scale (an individual item may be used on as many scales as required).

OUTPUT

The output includes for each scale: (1) total scores for each individual and his rank; (2) mean, standard deviation, and Cronbach's alpha (reliability); (3) item means, item standard deviations, and item discriminations (correlation of each item with total score); (4) item means for top third of group and item means for bottom third of group; (5) ranking of items in order of significance.

COMPUTER AND LANGUAGE

SUMRAT9 was written in FORTRAN IV, Level F, for an IBM 360/70 with no auxiliary components.

RESTRICTIONS

There basically are no data restrictions on the program, as any type of data may be used. However, the program as written may not exceed 500 items or 500 observations, with a maximum of 65 items per selection on the scales. The parameters can easily be modified to adapt to local restrictions.

AVAILABILITY

A copy of the program, the documentation for the program, and sample output from the program are available free on request and can be obtained from the author, Richard G. Landry, Bureau of Educational Research, University of North Dakota, Grand Forks, N. Dak. 58201.

(Received for publication November 16, 1973; accepted November 23, 1973.)

CLIQUE: A FORTRAN IV program for the Needham-Moody-Hollis cluster-listing algorithm*

MICHAEL SHAFTO

Princeton University, Princeton, New Jersey 08540

Jardine & Sibson (1971) presented a general axiomatic treatment of the class of data-simplification techniques known as *cluster analysis*. They defined two classes of methods, one for rank-order data, the other for interval-scaled data. Both classes of methods yield stratified (multilevel) clusterings based on a square symmetric matrix of dissimilarities. The rows and columns of the input matrix correspond to the points or items to be clustered. Entries in the matrix represent the dissimilarity or distance between each pair of points.

The clustering methods isolate sets of points such that points within the same set are relatively similar, while points not in the same set are relatively dissimilar. Both methods of Jardine and Sibson allow the user to control the amount of overlap between clusters at the same level, with zero overlap producing hierarchical solutions. The rank-order method is a generalization of the well known "single-link" method [cf. Johnson's (1967) "connectedness method"].

In computing the clustering solution for an input dissimilarities matrix, that matrix is transformed, and the clusters at each level are defined with respect to the transformed matrix. The clusters at Level h are defined as follows: (a) Construct an undirected graph from the transformed dissimilarities matrix by letting all pairs of points as close as h units to each other be connected and those further than h units apart be disconnected. (b) List the vertex sets of the maximal complete subgraphs of this graph. These sets are called the *ML sets* at Level h.

(In an undirected graph, a complete subgraph is a subgraph in which any two vertices are connected. A complete subgraph is maximal if it is not contained in another complete subgraph.)

The usefulness of the cluster-analysis methods mentioned above is "crucially dependent on the availability of an efficient cluster listing method [Jardine & Sibson, 1971, p. 238]," i.e., a method for finding the ML sets for a given dissimilarity matrix and Cutoff Value h. Jardine and Sibson described such an algorithm,

*This work was supported by Research Grant 19223 from the National Institute of Mental Health, United States Public Health Service, to T. Trabasso, Principal Investigator. Computer time was made available through the Department of Psychology, Princeton University.