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.

developed by J. K. M. Moody and J. Hollis, based on a suggestion by R. M. Needham. But, they observed, "It is obvious that the algorithm is very ill-suited to being programmed in a language such as FORTRAN."

This abstract describes a FORTRAN version of the Needham-Moody-Hollis algorithm which uses list-processing methods not usually associated with the FORTRAN language. Subroutine CLIQUE and its auxiliary subroutines are mainly concerned with building, searching, and modifying simple list structures called "chains." Each chain consists of a set of linked nodes, and each node consists of three integers: the address of the data item stored at that node, the address of the next higher node, and the address of the next lower node.

The algorithm goes through one cycle for each row of the dissimilarity matrix. Each cycle involves dividing an initial Set C of logical vectors (which are used to represent subsets of the points) into three mutually exclusive subsets, D, E, and F. At the end of a cycle, D, E, and F must be pooled to form a new set, C. At the end of the last cycle, C contains the logical vectors representing clusters at the current level.

By treating the four sets as chains, it is easy to add or delete arbitrary elements, to deal efficiently with continually changing set sizes, and to pool D, E, and F by simply linking the appropriate chains.

INPUT

Arguments passed to Subroutine CLIQUE are DIST, a REAL*4 array dimensioned (60,60) in the calling program; NP, the number of points to be clustered; and CRIT, the REAL*4 cutoff value for the current level. DIST (I,J) is the dissimilarity between Points I and J, with $1 \leq I < J \leq NP$. For compatibility with the Jardine and Sibson programs, $DIST(I,J) = \pm 1.1E10$ is treated as $DIST(I,J) = 0.0$. $DIST(I,J) \leq CRIT$ means Points I and J are connected in the graph for Level $h = CRIT$.

OUTPUT

When CLIQUE returns to the calling program, the clusters for the current level are stored as a chain local to CLIQUE. A call to UNPACK returns the next cluster from the chain by means of an INTEGER array INDEX. $INDEX(I) = 1$ if Point I is in the set ($= 0$ if not). UNPACK also returns an INTEGER flag which is positive until the last cluster has been returned. Then it is set to the negative of the maximum number of list storage locations which were used at the current level.

COMPUTER AND LANGUAGE

All subroutines were written in FORTRAN IV. No special hardware or software is required. The subroutines were written and tested on the IBM 360/91 at Princeton University, using the FORTRAN IV (H) compiler. Core requirements for CLIQUE and all associated subroutines and local storage totaled 8340 hexadecimal bytes.

SIZE LIMITATIONS

CLIQUE was written to be compatible with the clustering programs listed in Jardine and Sibson (1971). A maximum of 60 points can be clustered.

AVAILABILITY

A listing of the subroutines, along with the calling program, is available free of charge from the author.

REFERENCES

Jardine, N., & Sibson, R. *Mathematical taxonomy*. New York: Wiley, 1971.

Johnson, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, 32, 241-254.

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

Statistical programs utilizing dynamic allocation of memory on the CDC 6400

PAUL L. DERBY

Colorado State University, Ft. Collins, Colorado 80521

Statistical routines written for general use often require large data storage arrays. A program using three $n \times n$ arrays, for instance, would require 30,000 words of memory for 100 variables. Usually the number of variables in any analysis is less than the maximum number of variables dimensioned in the program. This generally results in a number of unused data storage locations and a waste of computer resources. It would be advantageous to allocate storage for data arrays at execution time rather than at compile time, thereby minimizing storage requirements and turn around time for each program run.

A series of multivariate analysis programs described by Cooley and Lohnes (1971) have been rewritten incorporating a subroutine, CORE, which dynamically allocates memory in blank common during execution according to parameters specified on the first cards preceding the data deck. In multiple problem runs, CORE is capable of either reducing or increasing field length, depending upon the number of variables included in the analysis. As an example, the CORREL program in the Cooley and Lohnes package requires a field length of 43000 words to execute on the CDC 6400 under SCOPE 3.3 without the CORE routine. As few as 16000 words are needed with the CORE routine.

INPUT/OUTPUT

All input and output remains the same as that described by Cooley and Lohnes.

COMPUTER AND PROGRAMMING LANGUAGE

The Cooley and Lohnes programs are written in FORTRAN IV (modified for the CDC FORTRAN extended compiler). The CORE subroutine is written in COMPASS assembly language. The programs are for use on the CDC 6000 series computers operating under the SCOPE monitor.

RESTRICTIONS

CORE can reserve space up to and including the field length requested by the user's job card at execution time. If more space is requested by CORE than initially reserved, the job aborts with an indication of the necessary field length in the system dump.

AVAILABILITY

The programs may be obtained (free) by sending a magnetic tape to the author.

REFERENCE

Cooley, W. W., & Lohnes, P. R. *Multivariate data analysis*. New York: Wiley, 1971.

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