APPLECOR: An Apple II implementation of hierarchical clustering analysis using the CONCOR algorithm

ANDREW R. GILPIN

University of Northern Iowa, Cedar Falls, Iowa

Despite the existence of numerous programs to perform hierarchical clustering analysis on mainframe computers (Blashfield, Aldenderfer, & Morey, 1982; Shepard & Arabie, 1979), there are relatively few implementations of such programs on microcomputers (Whaley, 1982). Program APPLECORN performs hierarchical clustering analyses of up to 30 objects or variables on the Apple II family of computers. Input to the program is similarity data in the form of a correlation matrix; the program produces a dendrogram and a version of the matrix as permuted therein.

Requirements. The program is written in Applesoft BASIC. It was developed on a 48K Apple II+ microcomputer, but should run without changes on other members of the Apple II family (II, IIe, and IIfx). A printer is desirable but not required because APPLECORN does not access external storage devices. It can be implemented either on tape- or disk-based systems. Conversion to other dialects of Microsoft BASIC should be straightforward (the manual contains a section intended to facilitate conversion to other machines). On the Apple II+, APPLECORN runs in 16K of RAM, although it would require more in systems whose BASICs do not permit integer arrays.

Input/Output Options. The correlation matrix that forms the raw data is entered into DATA statements at the end of the program. Since the matrix will be symmetrical, only the upper triangular half must be entered. There are two modes for output: (1) the 40-column screen and (2) a printer connected to any slot. If the former option is selected, the user may elect to have the computer await a keypress after displaying each line of output, to prevent information's scrolling off the screen.

Methodology. The CONCOR algorithm as first reported by McQuitty (1968) and McQuitty and Clark

(1968), and subsequently developed by Breiger, Boorman, and Arabie (1975), involves computing correlation coefficients columnwise on the original matrix; a new matrix composed of these coefficients is then formed, and the process is repeated. Breiger et al. reported that the process will converge on a matrix in which variables may be permuted to a blocked form in which two clusters of objects will be identified. These clusters may in turn be partitioned using the same procedure; that is, the algorithm is divisive (in contrast to most approaches, which are agglomerative). The program APPLECORN applies the CONCOR algorithm repeatedly, producing a dendrogram output that indicates the partition of each cluster.

Several parameters may be set, including the convergence cutoff criterion (Kruskal, 1978), the maximum number of iterations permitted, and the approximate number of clusters to be partitioned; default values should be adequate for most situations.

Availability. A user's manual, including program listing and annotation, is available without charge from the author. Copies of the program on diskette (DOS 3.3) may be obtained by sending a blank diskette plus \$2 (to cover postage and handling costs).

REFERENCES

- BLASHFIELD, R. K., ALDENDERFER, M. S., & MOREY, L. C. (1982). Cluster analysis software. In P. R. Krishnaiah & L. N. Kanal (Eds.), *Handbook of statistics: Vol. 2. Classification, pattern recognition, and reduction of dimensionality* (pp. 245-266). Amsterdam: North-Holland.
- BREIGER, R. L., BOORMAN, S. A., & ARABIE, P. (1975). An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12, 328-383.
- KRUSKAL, J. B. (1978, May). *Some results about CONCOR, a method of clustering*. Paper presented at the meeting of the Classification Society, Clemson, SC.
- MCQUITT, L. L. (1968). Multiple clusters, types, and dimensions from interactive intercolumnar correlation analysis. *Multivariate Behavioral Research*, 3, 465-477.
- MCQUITT, L. L., & CLARK, J. A. (1968). Clusters from iterative, intercolumnar correlational analysis. *Educational and Psychological Measurement*, 28, 211-238.
- SHEPARD, R. N., & ARABIE, P. (1979). Additive clustering: Representations of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86, 87-123.
- WHALEY, C. P. (1982). Interactive clustering. *Behavior Research Methods & Instrumentation*, 14, 170-175.

(Revision accepted for publication February 14, 1985.)

The author's mailing address is: Department of Psychology, University of Northern Iowa, Cedar Falls, IA 50614.