SINGD: A PL/I program for Eckart-Young singular decomposition*

RICARDO DOBSON and SALVATORE J. HEPPER
Transportation and Urban Analysis Department
General Motors Research Laboratories
Warren, Michigan 48090

This program implements an Eckart-Young (1936) decomposition of a data matrix, X, by successive factor residual matrix and principal axes factor matrix methods (Horst, 1965). Given an n x m X matrix with $n \geqslant m$, the program finds $\hat{X}$ such that

$$\hat{X} = P \Lambda Q'$$

where $\hat{X}$ has the same dimensions as X, but P is n x r, Λ is r x r, and Q is m x r. P and Q are both orthonormal and Λ is diagonal. The diagonal elements of Λ, λ_i, are ordered so that $\lambda_i \geqslant \lambda_{i+1}$ for i equals 1 through $r-1$. P, Λ, and Q are derived so that

$$\text{trace}\,[(X-\hat{X})'\,(X-\hat{X})] = \text{minimum}.$$

After the completion of the decomposition, two rotations of the solutions are possible. The successive factor varimax rotation (Horst, 1965) finds a transformation matrix, T, so that the columns of L are orthogonal but they have maximum variance subject to the simple structure constraint (Thurstone, 1947); $L = (Q\Lambda)T$. Since T is an r x r orthonormal matrix, $\hat{X} = (PT)(T'\Lambda Q') = P\Lambda Q'$. An oblique transformation matrix, C, may also be applied to $Q\Lambda$ to bring it into congruence with H, a hypothesis matrix, so that

$$(Q\Lambda)C = H \,\text{diag}\,(H'Q\Lambda^{-2}Q'H)^{-1/2}.$$

H is an m x r matrix which represents some notion about simple structure or some other structural hypothesis. C is an r x r nonsingular matrix such that $(C'C)$ contains 1 in its diagonal elements and the correlations among the r latent axes in its nondiagonal elements; $\hat{X} = (PC'^{-1})(C'\Lambda Q') = P\Lambda Q'$.

The program was designed for a wide variety of social science applications. A singular decomposition may be performed on the data matrix, X, or X may be row centered, column centered, or row and column centered before decomposition. In addition, the data matrix may be row or column standardized before decomposition. By selecting the appropriate options, the user may execute a classical principal components factor analysis of the R or Q type (Harman, 1967), Carroll's (1972) vector preference model, or a decomposition of the interaction matrix in Gollob's (1968) FANOVA model. The rotation to a hypothesis matrix provides a method for investigating alternative structural assumptions about the relations among the m variables.

*The mathematical advice of David Nash is gratefully acknowledged.

INPUT

The program requires an n x m X matrix, and it allows the optional input of an m x r H matrix. Centering and standardizing options as well as input/output options are selected by a series of numerical codes which precede X and H, if it is included.

OUTPUT

The output consists of: (a) means and standard deviations for rows and columns of X; (b) the inner product matrix which is decomposed; (c) Q and $Q\Lambda$, the right orthonormal and orthogonal matrices, respectively; (d) the singular values, λ_i, the percentage and cumulative percentage of trace accounted for by the eigenvalues, and plots of the singular values and cumulative percentage of trace for each eigenvalue; (e) the transformation matrix for the varimax rotation or both transformation matrices (C and C'^{-1}) as well as the correlations among latent dimensions for the rotation to hypothesis. The new eigenvalues are listed for the varimax solution along with the percentage and cumulative percentage of trace for the eigenvalues; (f) the rotated versions of $Q\Lambda$ and P, the right orthogonal and left orthonormal vectors, respectively; (g) plots of the columns of $Q\Lambda$, the matrix of right orthogonal vectors, or its rotated version if a rotation is requested, vs each other.

COMPUTER

The program has been implemented on an IBM 370/165 with O.S. Version 20.6 and MVT under HASP.

PROGRAM CAPACITY

Since the program allocates memory dynamically, its capacity is limited only by the size of the memory. Another consequence of dynamic allocation, of course, is that the core requirements for execution reduce automatically as the dimensions of X are reduced.

LANGUAGE

PL/I F compiler or PL/I optimizing compiler.

AVAILABILITY

Copies of a user's manual and source deck may be obtained from the authors without charge at General Motors Research Laboratories, Warren, Michigan 48090.

REFERENCES

Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, and S. B. Nerlove (Eds.), *Multidimensional scaling.* New York: Seminar Press, 1972, Pp. 105-157.

Eckart, C., & Young, G. The approximation of one matrix by another of lower rank. Psychometrika, 1936, 1, 211-218.

Gollob, H. F. A statistical model which combines features of factor analysis and analysis of variance techniques. Psychometrika, 1968, 33, 73-113.

Harman, H. *Modern factor analysis.* Chicago: University of Chicago Press, 1967.

Horst, P. *Factor analysis of data matrices.* New York: Holt, Rinehart, & Winston, 1965.

Thurstone, L. L. *Multiple factor analysis.* Chicago: University of Chicago Press, 1947.

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