

ALSCAL

Initial Configuration

The first step is to estimate an additive constant c_k , which is added to the observed proximity measures (for example, o_{ijk}). Thus,

$$o_{ijk}^* = o_{ijk} + c_k$$

such that for all triples the triangular inequality holds:

$$o_{ijk}^* + o_{jlk}^* \geq o_{ilk}^*$$

and positivity holds $o_{ijk}^* \geq 0$,

where

o_{ijk}^* is the adjusted proximity between stimulus i and stimulus j for subject k

o_{jlk}^* is the adjusted proximity between stimulus j and stimulus l for subject k

o_{ilk}^* is the adjusted proximity between stimulus i and stimulus l for subject k

The constant c_k , which is added, is as small as possible to estimate a zero point for the dissimilarity data, thus bringing the data more nearly in line with the ratio level of measurement. This step is necessary to make the \mathbf{B}_k^* matrix, described below, positive semidefinite (that is, with no imaginary roots).

The next step is to compute a scalar product matrix \mathbf{B}_k^{**} for each subject k by double centering \mathbf{O}_k^* , the adjusted proximity matrix for each subject. An element of the \mathbf{B}_k^{**} matrix b_{ijk}^{**} is computed as follows:

$$b_{ijk}^{**} = -\frac{1}{2} \left(o_{ijk}^{*2} - o_{i.k}^{*2} - o_{.jk}^{*2} + o_{.k}^{*2} \right)$$

2 ALSICAL

where

- $o_{i.k}^*$ are the row means for the adjusted proximities for subject k
- $o_{.jk}^*$ are the column means for the adjusted proximities for subject k
- $o_{..k}^*$ is the grand mean for subject k

Double centering to convert distances to scalar products is necessary because a scalar products matrix is required to compute an initial configuration using the Young-Householder-Torgerson procedure.

Next the individual subject matrices are normalized so that they have the same variance. The normalized matrix \mathbf{B}_k^* is found for each subject. The elements of the matrix are

$$b_{ijk}^* = \frac{b_{ijk}^{**}}{\left[\sum_i \sum_j (b_{ijk}^{**})^2 / (n(n-1)) \right]^{1/2}}$$

where n is the number of stimuli, and $n(n-1)$ is the number of off-diagonal elements in the \mathbf{B}_k^{**} matrix. The denominator is both the root mean square and the standard deviation of the unnormalized scalar products matrix \mathbf{B}^{**} . (It is both because $b_{..k}^{**} = 0$, due to double centering.) \mathbf{B}_k^* is thus a matrix with elements b_{ijk}^* , which are scalar products for individual subject k . Normalization of individual subjects' matrices equates the contribution of each individual to the formation of a mean scalar products matrix and thus the resulting initial configuration.

Next an average scalar products matrix \mathbf{B}^* over the subjects is computed. The elements of this matrix are

$$b_{ij.}^* = \frac{\sum_k b_{ijk}^*}{m}$$

where m is the number of subjects.

The average \mathbf{B}^* matrix used in the previous step is used to compute an initial stimulus configuration using the classical Young-Householder multidimensional scaling procedure

$$\mathbf{B}^* = \mathbf{X}\mathbf{X}'$$

where \mathbf{X} is an $n \times r$ matrix of n stimulus points on r dimensions, and \mathbf{X}' is the transpose of the \mathbf{X} matrix; that is, the rows and columns are interchanged. The \mathbf{X} matrix is the initial configuration.

For the weighted ALSICAL matrix model, initial weight configuration matrices \mathbf{W}_k for each of the m subjects are computed. The initial weight matrices \mathbf{W}_k are $r \times r$ matrices, where r is the number of dimensions. Later the diagonals of \mathbf{W}_k will form rows of the \mathbf{W} matrix, which is an $n \times r$ matrix. The matrices \mathbf{W}_k are determined such that $\mathbf{B}_k^* = \mathbf{Y}\mathbf{W}_k\mathbf{Y}'$, where $\mathbf{Y} = \mathbf{X}\mathbf{T}$ and $\mathbf{T}\mathbf{T}' = \mathbf{I}$, and where \mathbf{T} is an orthogonal rotation of the configuration \mathbf{X} to a new orientation \mathbf{Y} . \mathbf{T} is computed by the Schönemann-de Leeuw procedure discussed by Young, Takane, and Lewyckyj (1978). \mathbf{T} rotates \mathbf{X} so that \mathbf{W}_k is as diagonal as possible (that is, off-diagonal elements are as close to zero as possible on the average over subjects). Off-diagonal elements represent a departure from the model (which assumes that subjects weight only the dimensions of the stimulus space).

Optimization Algorithm

The optimization algorithm is a series of steps which are repeated (iterated) until the final solution is achieved. The steps for the optimization algorithm are performed successively because disparities, weights, and coordinates cannot be solved for simultaneously.

Distance

Distances are computed according to the weighted Euclidean model

$$d_{ijk}^2 = \sum_{a=1}^r w_{ka} (x_{ia} - x_{ja})^2$$

4 ALSICAL

where

- w_{ka} is the weight for subject k on a dimension a ,
- x_{ia} is the coordinate of stimulus i on dimension a ,
- x_{ja} is the coordinate of stimulus j on dimension a .

The first set of distances is computed from the coordinates and weights found in the previous steps. Subsequently, new distances are computed from new coordinates found in the iteration process (described below).

Optimal Scaling

Optimal scaling for ordinal data use Kruskal's least-squares monotonic transformation. This yields disparities that are a monotonic transformation of the data and that are as much like the distances (in a least squares sense) as possible. Ideally, we want the distances to be in exactly the same rank order as the data, but usually they are not. So we locate a new set of numbers, called disparities, which are in the same rank order as the data and which fit the distances as well as possible. When we see an order violation we replace the numbers that are out of order with a block of values that are the mean of the out-of-order numbers.

When there are ties in the data, the optimal scaling process is changed somewhat. Kruskal's primary and secondary procedures are used in ALSICAL.

Normalization

The disparities computed previously are now normalized for technical reasons related to the alternating least squares algorithm (see Young, de Leeuw, and Takane, 1979). During the course of the optimization process, we want to minimize a measure of error called SSTRESS. But the monotone regression procedure described above only minimizes the numerator of the SSTRESS formula. Thus, the formula below is applied to readjust the length of the disparities vector so that SSTRESS is minimized:

$$\mathbf{D}_k^{*N} = \mathbf{D}_k^* \left(\mathbf{D}_k' \mathbf{D}_k \right) \left(\mathbf{D}_k' \mathbf{D}_k^* \right)^{-1}$$

where

\mathbf{D}_k^* is a column vector with $\frac{n(n-1)}{2}$ elements containing all the disparities for subject k ,

\mathbf{D}_k is a column vector with $\frac{n(n-1)}{2}$ elements containing all the distances for subject k ,

$\mathbf{D}_k' \mathbf{D}_k$ is the sum of the squared distances,

$\mathbf{D}_k' \mathbf{D}_k^*$ is the sum of the cross products.

The normalized disparities vector \mathbf{D}_k^{*N} is a conditional least squares estimate for the distances; that is, it is the least squares estimate for a given iteration. The previous \mathbf{D}^* values are replaced by \mathbf{D}^{*N} values, and subsequent steps utilize the normalized disparities.

SSTRESS

The Takane-Young-de Leeuw formula is used:

$$\text{SSTRESS}(1) = S = \left[\frac{1}{m} \sum_{k=1}^m \left[\frac{\sum_i \sum_j (d_{ijk}^2 - d_{ijk}^{*2})^2}{\sum_i \sum_j d_{ijk}^{*4}} \right] \right]^{1/2}$$

where d_{ijk}^* values are the normalized disparity measures computed previously, and d_{ijk} are computed as shown above. Thus SSTRESS is computed from the normalized disparities and the previous set of coordinates and weights.

Termination

The current value of SSTRESS is compared to the value of SSTRESS from the previous iteration. If the improvement is less than a specified value (default equals 0.001), iteration stops and the output stages has been reached. If not, the program proceeds to the next step. (This step is skipped on the first iteration.)

Model Estimation

In ALSICAL the weights and coordinates cannot be solved for simultaneously, so we do it successively. Thus, the model estimation phase consists of two steps: (i) estimation of subject weights and (ii) estimation of stimulus coordinates.

(i) Estimation of subject weights. (This step is skipped for the simple, that is, unweighted, Euclidean model.)

A conditional least-squares estimate of the weights is computed at each iteration:

$$\mathbf{W} \cong \mathbf{D}^* \mathbf{P} (\mathbf{P} \mathbf{P})^{-1}$$

The derivation of the computational formula is as follows:

We have found disparities such that

$$d_{ijk}^* \cong d_{ijk}^2,$$

where

$$d_{ijk}^2 = \sum_{a=1}^r w_{ka} (x_{ia} - x_{ja})^2$$

Let p_{ija} be the unweighted distance between stimuli i and j as projected onto dimension a , that is,

$$p_{ija} = (x_{ia} - x_{ja})^2.$$

Then

$$d_{ijk}^{*2} \cong d_{ijk}^2 = \sum_{a=1}^r w_{ka} p_{ija}.$$

In matrix notation, this is expressed as $\mathbf{D}^* = \mathbf{W}\mathbf{P}'$, where \mathbf{D}^* is now an $m \times \frac{n(n-1)}{2}$ matrix having one row for every subject and one column for each stimulus pair; \mathbf{W} is an $m \times r$ matrix having one row for every subject and one column for each dimension; and \mathbf{P}' has one row for every dimension and one column for every stimulus pair.

We wish to solve for \mathbf{W} , $\mathbf{W}\mathbf{P}' \cong \mathbf{D}^*$, which we do by noting that

$$\mathbf{W}\mathbf{P}'(\mathbf{P}'\mathbf{P}')^{-1} = \mathbf{D}^*\mathbf{P}'(\mathbf{P}'\mathbf{P}')^{-1}.$$

Therefore,

$$\mathbf{W} = \mathbf{D}^*\mathbf{P}'(\mathbf{P}'\mathbf{P}')^{-1}$$

and we have the conditional least squares estimate for \mathbf{W} . We have in fact minimized SSTRESS at this point relative to the previously computed values for stimulus coordinates and optimal scaling. We replace the old subject weights with the newly estimated values.

(ii) Estimation of Stimulus Coordinates. The next step is to estimate coordinates, one at a time, using the previously computed values for \mathbf{D}^* (disparities) and weights. Coordinates are determined one at a time by minimizing SSTRESS with regard to a given coordinate. Equation (2) allows us to solve for a given coordinate x_{le} :

$$\frac{\partial S}{\partial x_{le}} = \frac{1}{m} \sum c_k \frac{\partial S_k}{\partial x_{le}} \quad (1)$$

$$\frac{\partial S_k}{\partial x_{le}} = 4w_{ke}^2 \sum_j (x_{le}^3 - 3x_{le}^2 x_{je} + 2x_{le} x_{je}^2 - b_{lje}^2 x_{le} + b_{lje}^2 x_{je}) \quad (2)$$

Equation (2) can be substituted back into equation (1). This equation with one unknown, x_{le} , is then set equal to zero and solved by standard techniques. All the other coordinates except x_{le} are assumed to be constant while we solve for x_{le} .

Immediately upon solving for x_{le} , we replace the value for x_{le} used on the previous iteration with the newly obtained value, and then proceed to estimate the value for another coordinate. We successively obtain values for each coordinate of

point l , one at a time, replacing old values with new ones. This continues for point l until the estimates stabilize. We then move to a new point and proceed until new coordinates for all stimuli are estimated. We then return to the beginning of the optimization algorithm (the previous step above) and start another iteration.

References

- Bloxom, B. 1978. Constrained multidimensional scaling in n spaces. *Psychometrika*, 43: 397–408.
- Carroll, J. D., and Chang, J. J. 1970. Analysis of individual differences in multidimensional scaling via an N -way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35: 238–319.
- Carroll, J. D., and Chang, J. J. 1972. IDIOSCAL (Individual differences in orientation scaling). Paper presented at the spring meeting of the Psychometric Society, Princeton, New Jersey.
- Carroll, J. D., Pruzansky, S., and Kruskal, J. B. 1980. CANDELINC: A general approach to multidimensional analysis with linear constraints on parameters. *Psychometrika*, 45: 3–24.
- Harshman, R. A. 1970. Foundations of the PARAFAC procedure: Models and conditions for an explanatory multi-model factor analysis. Working Papers in Phonetics No. 16. Los Angeles: University of California.
- Jones, L. E., and Young, F. W. 1972. Structure of a social environment: longitudinal individual differences scaling of an intact group. *Journal of Personality and Social Psychology*, 24: 108–121.
- MacCallum, R. C. 1977. Effects of conditionality on INDSCAL and ALSCAL weights. *Psychometrika*, 42: 297–305.
- Mardia, K. V. 1972. *Statistics of Directional Data*. New York: Academic Press.
- Mardia, K. V., Kent, J. T., and Bibby, J. M. 1979. *Multivariate Analysis*. New York: Academic Press.
- Null, C. H., and Sarle, W. S. 1982. Multidimensional scaling by least squares. *Proceedings of the 7th Annual SAS User's Group International*.

- Schiffman, S. S., Reynolds, M. L., and Young, F. W. 1981. *Introduction to multidimensional scaling: theory, methods and applications*. New York: Academic Press.
- Takane, Y., de Leeuw, J., and Young, F. W. 1977. Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, 42: 7–67.
- Tucker, L. R. 1972. Relations between multidimensional scaling and three mode factor analysis. *Psychometrika*, 37: 3–28.
- Young, F. W., and Lewyckyj, R. 1979. *ALSCAL-4 user's guide*. Carrboro: N.C.: Data Analysis and Theory Associates.
- Young, F. W., Easterling, D. V., and Forsyth, B. N. 1984. The general Euclidean model for scaling three mode dissimilarities: theory and application. In: *Research Methods for Multi-mode Data Analysis in the Behavioral Sciences*, H. G. Law, G. W. Snyder, Jr., J. Hattie, and R. P. McDonald, eds. New York: Praeger.
- Young, F. W., Takane, Y., and Lewyckyj, R. 1978. ALSCAL: A nonmetric multidimensional scaling program with several difference options. *Behavioral Research Methods and Instrumentation*, 10: 451–453.