PRINCALS

The PRINCALS algorithm was first described in Van Rijckevorsel and De Leeuw (1979) and De Leeuw and Van Rijckevorsel (1980); also see Gifi (1981, 1985). Characteristic features of PRINCALS are the ability to specify any of a number of measurement levels for each variable separately and the treatment of missing values by setting weights in the loss function equal to 0.

Notation

The following notation is used throughout this chapter unless otherwise noted:

n	Number of cases (objects)
m	Number of variables
p	Number of dimensions

For variable j, $j = 1, \ldots, m$

h_j	n-vector with categorical observations
k_j	Number of categories (distinct values) of variable j
$\mathbf{G}_j$	Indicator matrix for variable j, of order $n \times k_j$

The elements of $\mathbf{G}_j$ are defined as $i = 1, \ldots, n;\ r = 1, \ldots, k_j$

$$g_{(j)ir} = \begin{cases} 1 & \text{when the } i\text{th object is in the } r\text{th category of variable } j \\ 0 & \text{when the } i\text{th object is not in the } r\text{th category of variable } j \end{cases}$$

$\mathbf{M}_j$	Binary, diagonal $n \times n$ matrix

The diagonal elements of $\mathbf{M}_j$ are

$$m_{(j)ii} = \begin{cases} 1 & \text{when the } i\text{th observation is within the range } [1, k_j] \\ 0 & \text{when the } i\text{th observation is outside the range } [1, k_j] \end{cases}$$

$\mathbf{D}_j$	Diagonal matrix containing the univariate marginals; that is, the column sums of $\mathbf{G}_j$

The quantification matrices and parameter vectors are:

$\mathbf{X}$	Object scores, of order $n \times p$
$\mathbf{Y}_j$	Multiple category quantifications, of order $k_j \times p$
$\mathbf{y}_j$	Single category quantifications, of order k_j
$\mathbf{a}_j$	Variable weights (equal to component loadings) of order p
$\mathbf{Q}$	Transformed data matrix of order $n \times m$ with columns $\mathbf{q}_j = \mathbf{G}_j \mathbf{y}_j$
$\underline{\mathbf{Y}}$	Collection of multiple and single category quantifications

Note: The matrices $\mathbf{G}_j$, $\mathbf{M}_j$, and $\mathbf{D}_j$ are exclusively notational devices; they are stored in reduced form, and the program fully profits from their sparseness by replacing matrix multiplications with selective accumulation.

Objective Function Optimization

The PRINCALS objective is to find object scores $\mathbf{X}$ and a set of $\underline{\mathbf{Y}}_j$ (for $j = 1, \ldots, m$) — the underlining indicates that they are possibly restricted—so that the function

$$\sigma(\mathbf{X}; \underline{\mathbf{Y}}) = 1/m \sum_j \mathrm{tr}\left(\left(\mathbf{X} - \mathbf{G}_j \underline{\mathbf{Y}}_j \right)' M_j \left(\mathbf{X} - \mathbf{G}_j \underline{\mathbf{Y}}_j \right) \right)$$

is minimal, under the normalization restriction $\mathbf{X}'\mathbf{M}_*\mathbf{X} = mn\mathbf{I}$, where the matrix $\mathbf{M}_* = \sum_j \mathbf{M}_j$, and $\mathbf{I}$ is the $p \times p$ identity matrix. The inclusion of $\mathbf{M}_j$ in $\sigma(\mathbf{X}; \underline{\mathbf{Y}})$ ensures that there is no influence of data values outside the range $[1, k_j]$, which may be really missing or merely regarded as such; $\mathbf{M}_*$ contains the number of "active" data values for each object. The object scores are also centered; that is, they satisfy $\mathbf{u}'\mathbf{M}_*\mathbf{X} = 0$, with $\mathbf{u}$ denoting an n-vector with ones.

The following measurement levels are distinguished in PRINCALS:

Multiple Nominal

$\underline{\mathbf{Y}}_j = \mathbf{Y}_j$ (unrestricted)

Single Nominal

$\underline{\mathbf{Y}}_j = \mathbf{y}_j\mathbf{a}_j'$ (rank-one restrictions only)

(Single) Ordinal

$\underline{\mathbf{Y}}_j = \mathbf{y}_j\mathbf{a}_j'$ and $\mathbf{y}_j \in \mathbf{C}_j$ (rank-one and montonicity restrictions)

(Single) Numerical

$\underline{\mathbf{Y}}_j = \mathbf{y}_j\mathbf{a}_j'$ and $\mathbf{y}_j \in \mathbf{L}_j$ (rank-one and linearity restrictions)

For each variable, these levels can be chosen independently. The general requirement in the "single" options is $\underline{\mathbf{Y}}_j = \mathbf{y}_j\mathbf{a}_j'$; that is, $\underline{\mathbf{Y}}_j$ is of rank one; for identification purposes, $\mathbf{y}_j$ is always normalized so that $\mathbf{y}_j'\mathbf{D}_j\mathbf{y}_j = n$, which implies that the variance of the quantified variable $\mathbf{q}_j = \mathbf{G}_j\mathbf{y}_j$ is 1. In the ordinal case, the additional restriction $\mathbf{y}_j \in \mathbf{C}_j$ means that $\mathbf{y}_j$ must be located in the convex cone of all $\mathbf{k}_j$-vectors with nondecreasing elements. In the numerical case, the additional restriction $\mathbf{y}_j \in \mathbf{L}_j$ means that $\mathbf{y}_j$ must be located in the subspace of all $\mathbf{k}_j$-vectors that are a linear transformation of the vector consisting of $\mathbf{k}_j$ successive integers.

Optimization is achieved through the following iteration scheme twice:

(1) Initialization (a) or (b)

(2) Update object scores

(3) Orthonormalization

(4) Update category quantifications

(5) Convergence test: repeat (2)–(4) or continue

(6) Rotation

The first time (for the initial configuration) initialization (a) described below is used and all single variables are temporarily treated as single numerical. Multiple nominal variables are treated as multiple nominal. The second time (for the final configuration) initialization (b) is used. Steps (1) through (6) are explained below.

(1) Initialization

(a) The object scores X are initialized with random numbers, which are normalized so that $\mathbf{u}'\mathbf{M}_*\mathbf{X} = 0$ and $\mathbf{X}'\mathbf{M}_*\mathbf{X} = mn\mathbf{I}$, yielding $\tilde{\mathbf{X}}$. For multiple variables, the initial category quantifications are obtained as $\tilde{\tilde{\mathbf{Y}}}_j = \mathbf{D}_j^{-1}\mathbf{G}_j'\tilde{\mathbf{X}}$. For single variables, the initial category quantifications are defined as the first k_j successive integers, normalized so that $\mathbf{u}'\mathbf{D}_j\tilde{\mathbf{y}}_j = 0$ and $\tilde{\mathbf{y}}_j'\mathbf{D}_j\tilde{\mathbf{y}}_j = n$, and the initial variable weights are calculated as the vector $\tilde{\mathbf{a}}_j = \tilde{\tilde{\mathbf{X}}}'\mathbf{G}_j\tilde{\mathbf{y}}_j$, rescaled to unit length.

(b) All relevant quantities are copied from the results of the first cycle.

(2) Update object scores

First, the auxiliary score matrix Z is computed as

$$\mathbf{Z} \leftarrow \sum\nolimits_j \mathbf{M}_j\mathbf{G}_j\tilde{\tilde{\mathbf{Y}}}_j$$

and centered with respect to $\mathbf{M}_*$:

$$\tilde{\mathbf{Z}} \leftarrow \left\{\mathbf{M}_* - \left(\mathbf{M}_*\mathbf{u}\mathbf{u}'\mathbf{M}_*/\mathbf{u}'\mathbf{M}_*\mathbf{u}\right)\right\}\mathbf{Z}\ .$$

These two steps yield locally the best updates when there are no orthogonality constraints.

(3) Orthonormalization

The orthonormalization problem is to find an $\mathbf{M}_*$-orthonormal $\mathbf{X}^+$ that is closest to $\tilde{\mathbf{Z}}$ in the least squares sense. In PRINCALS, this is done by setting

$$\mathbf{X}^+ \leftarrow m^{1/2}\mathbf{M}_*^{-1/2}\mathrm{GRAM}\left(\mathbf{M}_*^{-1/2}\tilde{\mathbf{Z}}\right)$$

which is equal to the genuine least squares estimate up to a rotation—see (6). The notation GRAM( ) is used to denote the Gram-Schmidt transformation (Björk and Golub, 1973).

(4) Update category quantifications; loop across variable $j = 1, \ldots, m$

(a) For multiple nominal variables, the new category quantifications are computed as:

$$\mathbf{Y}_j^+ = \mathbf{D}_j^{-1}\mathbf{G}_j'\tilde{\mathbf{X}}$$

(b) For single variables, first an unconstrained update is computed in the same way:

$$\tilde{\mathbf{Y}}_j = \mathbf{D}_j^{-1}\mathbf{G}_j'\mathbf{X}^+$$

Next, one cycle of an ALS algorithm (De Leeuw et al., 1976) is executed for computing a rank-one decomposition of $\tilde{\mathbf{Y}}_j$, with restrictions on the left-hand vector. This cycle starts from the previous single quantifications $\tilde{\mathbf{y}}_j$ with

$$\mathbf{a}_j^+ = \tilde{\mathbf{Y}}_j'\mathbf{D}_j\tilde{\mathbf{y}}_j$$

When the current variable is numerical, we are ready; otherwise we compute

$$\mathbf{y}_j^* = \tilde{\mathbf{Y}}_j\mathbf{a}_j^+$$

Now, when the current variable is single nominal you can simply obtain $\mathbf{y}_j^+$ by normalizing $\mathbf{y}_j^*$ in the way indicated below; otherwise the variable must be ordinal, and you have to insert the weighted monotonic regression process

$$\mathbf{y}_j^* \leftarrow \mathrm{WMON}\left(\mathbf{y}_j^*\right)$$

which make $\mathbf{y}_j^*$ monotonically increasing. The weights used are the diagonal elements of $\mathbf{D}_j$, and the subalgorithm used is the up-and-down-blocks minimum violators algorithm (Kruskal, 1964; Barlow et al., 1972). The result is normalized:

$$\mathbf{y}_j^+ = n^{1/2}\mathbf{y}_j^*\left(\mathbf{y}_j'^*\mathbf{D}_j\mathbf{y}_j^*\right)^{-1/2}$$

Finally, we set

$$\underline{\mathbf{Y}}_j^+ = \mathbf{y}_j^+\mathbf{a}_j'^+$$

(5) Convergence test

The difference between consecutive values of the quantity

$$\text{TFIT} = 1/m \sum_s \left[\sum_{j \in J} \mathbf{y}'_{(j)s} \mathbf{D}_j \mathbf{y}_{(j)s} + \sum_{j \notin J} \mathbf{a}'_j \mathbf{a}_j \right]$$

where $\mathbf{y}_{(j)s}$ denotes the sth column of $\mathbf{Y}_j$ and where J is an index set recording which variables are multiple, is compared with the user-specified convergence criterion ε—a small positive number. It can be shown that $\text{TFIT} = \mathbf{p} - \sigma(\mathbf{X}; \underline{\mathbf{Y}})$. Steps (2) through (4) are repeated as long as the loss difference exceeds ε.

(6) Rotation

As remarked in (3), during iteration the orientation of **X** and **Y** with respect to the coordinate system is not necessarily correct; this also reflects that $\sigma(\mathbf{X}; \underline{\mathbf{Y}})$ is invariant under simultaneous rotations of **X** and **Y**. From the theory of principal components, it is known that if all variables would be single, the matrix **A** — which can be formed by stacking the row vectors $\mathbf{a}'_j$—has the property that $\mathbf{A}'\mathbf{A}$ is diagonal. Therefore you can rotate so that the matrix

$$1/m \mathbf{A}'\mathbf{A} = 1/m \sum_j \mathbf{a}_j \mathbf{a}'_j = 1/m \sum_j \mathbf{Y}'_j \mathbf{D}_j \mathbf{Y}_j$$

becomes diagonal. The corresponding eigenvalues are printed after the convergence message of the program. The calculation involves tridiagonalization with Householder transformations followed by the implicit QL algorithm (Wilkinson, 1965).

Diagnostics

Maximum Rank (may be issued as a warning when exceeded)

The maximum rank $p_{\max}$ indicates the maximum number of dimensions that can be computed for any data set. In general

$$p_{\max} = \min\left\{(n-1), \left(\left(\sum_{j\in J} k_j + m_2\right) - \max\left(m_1, \max\left(0, 1-m_2\right)\right)\right)\right\},$$

where m_1 is the number of multiple variables with no missing values, m_2 is the number of single variables, and J is an index set recording which variables are multiple. Although the number of nontrivial dimensions may be less than $p_{\max}$ when $m = 2$, PRINCALS does allow dimensionalities all the way up to $p_{\max}$. When, due to empty categories in the actual data, the rank deteriorates below the specified dimensionality, the program stops.

Marginal Frequencies

The frequencies table gives the univariate marginals and the number of missing values (that is, values that are regarded as out of range for the current analysis) for each variable. These are computed as the column sums of $\mathbf{D}_j$ and the total sum of $\mathbf{M}_j$.

Fit and Loss Measures

When the HISTORY option is in effect, the following fit and loss measures are reported:

(a) Total fit. This is the quantity TFIT defined in (5).

(b) Total loss. This is $\sigma(\mathbf{X}; \underline{\mathbf{Y}})$, computed as the sum of multiple loss and single loss defined below.

(c) Multiple loss. This measure is computed as

$$\text{TMLOSS} = p - 1/m \sum_j \text{tr}\left[\mathbf{Y}_j'\mathbf{D}_j\mathbf{Y}_j\right]$$

(d) Single loss. This measure is computed only when some of the variables are single:

$$\text{SLOSS} = 1/m \sum_{j\notin J} \text{tr}\left[\mathbf{Y}_j'\mathbf{D}_j\mathbf{Y}_j\right] + \sum_{j\notin J} \mathbf{a}_j'\mathbf{a}_j$$

Eigenvalues and Correlations between Optimally Scaled Variables

If there are no missing data, the eigenvalues printed by PRINCALS are those of $1/m\mathbf{R}(\mathbf{Q})$, where $\mathbf{R}(\mathbf{Q})$ denotes the matrix of correlations between the optimally scaled variables in the columns of $\mathbf{Q}$. For multiple variables, $\mathbf{q}_j$ is defined here as $\mathbf{G}_j\mathbf{y}_{(j)1}$. When all variables are single or when $p = 1$, $\mathbf{R}(\mathbf{Q})$ itself is also printed. If there are missing data, then the eigenvalues are those of the matrix with elements $\mathbf{q}_j'\mathbf{M}_*^{-1}\mathbf{q}_1$, which is not necessarily a correlation matrix, although it is positive semidefinite.

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. 1972. *Statistical inference under order restrictions.* New York: John Wiley & Sons, Inc.

Björk, A., and Golub, G. H. 1973. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27: 579–594.

De Leeuw, J., and Van Rijckevorsel, J. 1980. HOMALS and PRINCALS—Some generalizations of principal components analysis. In: *Data Analysis and Informatics*, E. Diday et al, eds. Amsterdam: North-Holland.

De Leeuw, J., Young, F. W., and Takane, Y. 1976. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika*, 31: 33–42.

Gifi, A. 1981. *Nonlinear multivariate analysis*. Leiden: Department of Data Theory.

Gifi, A. 1985. *PRINCALS*. Leiden: Department of Data Theory, Internal Report UG-85–02.

Kruskal, J. B. 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29: 115–129.

Van Rijckevorsel, J., and De Leeuw, J. 1979. *An outline of PRINCALS*. Leiden: Internal Report RB 002–'79, Department of Data Theory.

Wilkinson, J. H. 1965. *The algebraic eigenvalue problem*. Oxford: Clarendon Press.