

# MULTIPLE CORRESPONDENCE

---

Multiple Correspondence Analysis, also known as homogeneity analysis, quantifies nominal (categorical) data by assigning numerical values to the cases (objects) and categories, such that in the low-dimensional representation of the data, objects within the same category are close together and objects in different categories are far apart. Each object is as close as possible to the category points of categories that apply to the object. In this way, the categories divide the objects into homogeneous subgroups. Variables are considered homogeneous when they classify objects that are in the same categories into the same subgroups.

## Notation

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

$n$	Number of analysis cases (objects)
$n_w$	Weighted number of analysis cases: $\sum_{i=1}^n w_i$
$n_{tot}$	Total number of cases (analysis + supplementary)
$w_i$	Weight of object $i$ ; $w_i = 1$ if cases are unweighted; $w_i = 0$ if object $i$ is supplementary.
$\mathbf{W}$	Diagonal $n_{tot} \times n_{tot}$ matrix, with $w_i$ on the diagonal.
$m$	Number of analysis variables
$m_w$	Weighted number of analysis variables ( $\sum_{j=1}^m v_j$ )
$m_{tot}$	Total number of variables (analysis + supplementary)
$\mathbf{H}$	The data matrix (category indicators), of order $n_{tot} \times m_{tot}$ , after discretization, imputation of missings, and listwise deletion, if applicable.
$p$	Number of dimensions

For variable  $j$ ,  $j = 1, \dots, m_{tot}$

$v_j$  variable weight;  $v_j = 1$  if weight for variable  $j$  is not specified or if variable  $j$  is supplementary

$k_j$  Number of categories of variable  $j$  (number of distinct values in  $\mathbf{h}_j$ , thus, including supplementary objects)

$\mathbf{G}_j$  Indicator matrix for variable  $j$ , of order  $n_{tot} \times k_j$

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

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

$\mathbf{D}_j$  Diagonal  $k_j \times k_j$  matrix, containing the weighted univariate marginals; i.e., the weighted column sums of  $\mathbf{G}_j$  ( $\mathbf{D}_j = \mathbf{G}'_j \mathbf{W} \mathbf{G}_j$ )

$\mathbf{M}_j$  Diagonal  $n_{tot} \times n_{tot}$  matrix, with diagonal elements defined as

$$m(j)_{ii} = \begin{cases} 0 & \text{when the } i\text{th observation is missing and missing strategy variable } j \text{ is passive} \\ 0 & \text{when the } i\text{th object is in } r\text{th category of variable } j \text{ and } r\text{th category is only} \\ & \text{used by supplementary objects (i.e. when } d_{(j)rr} = 0) \\ v_j & \text{otherwise} \end{cases}$$

$$\mathbf{M}_* = \sum_j \mathbf{M}_j$$

The quantification matrices and parameter vectors are:

$\mathbf{X}$  Object scores, of order  $n_{tot} \times p$

$\mathbf{X}_w$  Weighted object scores ( $\mathbf{X}_w = \mathbf{W} \mathbf{X}$ )

$\mathbf{X}^n$              $\mathbf{X}$  normalized according to requested normalization option

$\mathbf{Y}_j$             Category quantifications (centroid coordinates), of order  $k_j \times p$ .

*Note:* The matrices  $\mathbf{W}$ ,  $\mathbf{G}_j$ ,  $\mathbf{M}_j$ ,  $\mathbf{M}^*$ , 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.

## Discretization

Discretization is done on the unweighted data.

### Multiplying

First, the original variable is standardized. Then the standardized values are multiplied by 10 and rounded, and a value is added such that the lowest value is 1.

### Ranking

The original variable is ranked in ascending order, according to the alphanumerical value.

### Grouping into a specified number of categories with a normal distribution

First, the original variable is standardized. Then cases are assigned to categories using intervals as defined in Max (1960).

### Grouping into a specified number of categories with a uniform distribution

First the target frequency is computed as  $n$  divided by the number of specified categories, rounded. Then the original categories are assigned to grouped categories such that the frequencies of the grouped categories are as close to the target frequency as possible.

## Grouping equal intervals of specified size

First the intervals are defined as lowest value + interval size, lowest value + 2\*interval size, etc. Then cases with values in the  $k^{\text{th}}$  interval are assigned to category  $k$ .

## Imputation of Missing Values

When there are variables with missing values specified to be treated as active (impute mode or extra category), then first the  $k_j$ 's for these variables are computed before listwise deletion. Next the category indicator with the highest weighted frequency (mode; the smallest if multiple modes exist), or  $k_j + 1$  (extra category) is imputed. Then listwise deletion is applied if applicable. Finally the  $k_j$ 's are adjusted.

## Configuration

MULTIPLE CORRESPONDENCE can read a configuration from a file to be used as the initial configuration or as a fixed configuration in which to fit variables.

For an initial configuration see step 1 in the Objective Function Optimization section.

A fixed configuration  $\mathbf{X}$  is centered and orthonormalized as described in the optimization section in step 3 (with  $\mathbf{X}$  instead of  $\mathbf{Z}$ ) and step 4 (except for the factor  $n_w^{1/2}$ ), and the result is postmultiplied with  $\mathbf{\Lambda}^{1/2}$  (this leaves the configuration unchanged if it is already centered and orthogonal). The analysis variables are set to supplementary and variable weights are set to one. Then MULTIPLE CORRESPONDENCE proceeds as described in the Supplementary Variables section.

## Objective Function Optimization

### Objective Function

The MULTIPLE CORRESPONDENCE objective is to find object scores  $\mathbf{X}$  and a set of  $\mathbf{Y}_j$  (for  $j = 1, \dots, m$ ) so that the function

$$\sigma(\mathbf{X}; \mathbf{Y}) = (n_w p)^{-1} \sum_j \text{tr} \left( (\mathbf{X} - \mathbf{G}_j \mathbf{Y}_j)' \mathbf{M}_j \mathbf{W} (\mathbf{X} - \mathbf{G}_j \mathbf{Y}_j) \right)$$

is minimal, under the normalization restriction  $\mathbf{X}' \mathbf{M}_* \mathbf{W} \mathbf{X} = n_w m_w \mathbf{I}$  ( $\mathbf{I}$  is the  $p \times p$  identity matrix). The inclusion of  $\mathbf{M}_j$  in  $\sigma(\mathbf{X}; \mathbf{Y})$  ensures that there is no influence of passive missing values (missing values in variables that have missing option passive, or missing option not specified).  $\mathbf{M}_*$  contains the number of active data values for each object. The object scores are also centered; i.e. they satisfy  $\mathbf{u}' \mathbf{M}_* \mathbf{W} \mathbf{X} = \mathbf{0}$  with  $\mathbf{u}$  denoting an  $n$ -vector with ones.

## Optimization

Optimization is achieved by executing the following iteration scheme:

1. Initialization
2. Update category quantifications
3. Update object scores
4. Orthonormalization
5. Convergence test: repeat (2)—(4) or continue
6. Rotation

Steps (1) through (6) are explained below.

### (1) Initialization

If an initial configuration is not specified, the object scores  $\mathbf{X}$  are initialized with random numbers. Then  $\mathbf{X}$  is orthonormalized (see step 4) so that  $\mathbf{u}' \mathbf{M}_* \mathbf{W} \mathbf{X} = \mathbf{0}$  and  $\mathbf{X}' \mathbf{M}_* \mathbf{W} \mathbf{X} = n_w m_w \mathbf{I}$ , yielding  $\mathbf{X}_w^+$ .

### (2) Update category quantifications; loop across variables $j = 1, \dots, m$ (variables $1, \dots, m$ are analysis variables):

With fixed current values  $\mathbf{X}_w^+$  the update of  $\mathbf{Y}_j$  is

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

### (3) Update object scores

First the auxiliary score matrix  $\mathbf{Z}$  is computed as

$$\mathbf{Z} \leftarrow \sum_j \mathbf{M}_j \mathbf{G}_j \mathbf{Y}_j^+$$

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

$$\mathbf{X}^* = (\mathbf{I} - \mathbf{M}_* \mathbf{u} \mathbf{u}' \mathbf{W} / (\mathbf{u}' \mathbf{M}_* \mathbf{W} \mathbf{u})) \mathbf{Z}$$

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

#### (4) Orthonormalization

To find an  $\mathbf{M}_*$ -orthonormal  $\mathbf{X}^+$  that is closest to  $\mathbf{X}^*$  in the least squares sense, we use for the Procrustes rotation (Cliff, 1966) the singular value decomposition  $m_w^{1/2} \mathbf{M}_*^{-1/2} \mathbf{W}^{1/2} \mathbf{X}^* = \mathbf{K} \mathbf{\Lambda}^{1/2} \mathbf{L}'$ ,

then  $n_w^{1/2} m_w^{1/2} \mathbf{M}_*^{-1/2} \mathbf{W}^{1/2} \mathbf{K} \mathbf{L}'$  yields  $\mathbf{M}_*$ -orthonormal weighted object scores:

$$\mathbf{X}_w^+ \leftarrow n_w^{1/2} m_w \mathbf{M}_*^{-1} \mathbf{W} \mathbf{X}^* \mathbf{\Lambda}^{-1/2} \mathbf{L}', \text{ and } \mathbf{X}^+ = \mathbf{W}^{-1} \mathbf{X}_w^+.$$

The calculation of  $\mathbf{L}$  and  $\mathbf{\Lambda}$  is based on tridiagonalization with Householder transformations followed by the implicit QL algorithm (Wilkinson, 1965).

#### (5) Convergence test

The difference between consecutive values of the quantity

$$\text{TFIT} = (pn_w)^{-1} \sum_j v_j \text{tr}(\mathbf{Y}_j' \mathbf{D}_j \mathbf{Y}_j),$$

is compared with the user-specified convergence criterion  $\varepsilon$  - a small positive number. It can be shown that  $\text{TFIT} = m_w - \sigma(\mathbf{X}; \mathbf{Y})$ . Steps (2) through (4) are repeated as long as the loss difference exceeds  $\varepsilon$ .

After convergence TFIT is also equal to  $\text{tr}(\mathbf{\Lambda}^{1/2})$ , with  $\mathbf{\Lambda}$  as computed in step (4) during the last iteration. (See also Model Summary, and Correlations Transformed Variables for interpretation of  $\mathbf{\Lambda}^{1/2}$ ).

#### (6) Rotation

To achieve principal axes orientation,  $\mathbf{X}^+$  is rotated with the matrix  $\mathbf{L}$ . Then step (2) is executed, yielding the rotated quantifications.

## Supplementary Objects

To compute the object scores for supplementary objects, after convergence steps (2) and (3) are repeated, with the zero's in  $\mathbf{W}$  temporarily set to ones in computing  $\mathbf{Z}$  and  $\mathbf{X}^+$ . If a supplementary object has missing values, passive treatment is applied.

## Supplementary Variables

The quantifications for supplementary variables are computed after convergence, by executing step (2) once.

## 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

$$p_{\max} = \min \left( n-1, \left( \sum_j k_j \right) - m \right)$$

if there are no variables with missing values to be treated as passive. If there are variables with missing values to be treated as passive, the maximum rank is

$$p_{\max} = \min \left( n-1, \left( \sum k_j \right) - \max(m_1, 1) \right)$$

with  $m_1$  the number of variables without missing values to be treated as passive.

Here  $k_j$  is exclusive supplementary objects (that is, a category only used by supplementary objects is not counted in computing the maximum rank). Although the number of nontrivial dimensions may be less than  $p_{\max}$  when  $m = 2$ , MULTIPLE CORRESPONDENCE 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.

## Descriptives

The descriptives tables gives the weighted univariate marginals and the weighted number of missing values (system missing, user defined missing, and values  $\leq 0$ ) for each variable.

## Fit and Loss Measures

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

- (a) Fit (VAF). This is the quantity TFIT as defined in step (5).
- (b) Loss. This is  $\sigma(\mathbf{X}; \mathbf{Y})$ .

## Model Summary

### Cronbach's Alpha

Cronbach's Alpha per dimension ( $s = 1, \dots, p$ ):

$$\alpha_s = m_w(\lambda_s^{1/2} - 1) / (\lambda_s^{1/2}(m_w - 1)),$$

Total Cronbach's Alpha is

$$\alpha = m_w \left( \sum_s \lambda_s^{1/2} - 1 \right) / \sum_s \lambda_s^{1/2} (m_w - 1)$$

with  $\lambda_s$  the  $s^{\text{th}}$  diagonal element of  $\mathbf{\Lambda}$  as computed in step (4) during the last iteration.

### Variance Accounted For

Variance Accounted For per dimension ( $s = 1, \dots, p$ ):

$$\text{VAF}_s = n_w^{-1} \sum_j v_j \text{tr}(\mathbf{y}'_{(j)s} \mathbf{D}_j \mathbf{y}_{(j)s}), \text{ (% of variance is } \text{VAF}_s \times 100 / m_w \text{)},$$

### Eigenvalue

Eigenvalue per dimension:

$$\lambda_s^{1/2} = \text{VAF}_s,$$



with  $\lambda_s$  the  $s^{th}$  diagonal element of  $\mathbf{\Lambda}$  as computed in step (4) during the last iteration. (See also Optimization step (5), and Correlations Transformed Variables for interpretation of  $\mathbf{\Lambda}^{1/2}$ ).

The Total Variance Accounted For is the mean over dimensions. So, the total eigenvalue is

$$\text{tr}(\mathbf{\Lambda}^{1/2}) = p^{-1} \sum_s \text{VAF}_s .$$

If there are no passive missing values, the eigenvalues  $\mathbf{\Lambda}^{1/2}$  are those of the correlation matrix (see the Correlations and Eigenvalues section) weighted with variable weights:

$$r_{jj}^w = v_j r_{jj}, \text{ and } r_{jl}^w = r_{jl}^w = v_j^{1/2} r_{jl}$$

If there are passive missing values, then the eigenvalues are those of the matrix  $m_w \mathbf{Q}_c' \mathbf{M}_*^{-1} \mathbf{Q}_c$ , with

$$\mathbf{Q}_c = n_w^{-1/2} (\mathbf{I} - \mathbf{M}_* \mathbf{u} \mathbf{u}' \mathbf{W} / (\mathbf{u}' \mathbf{M}_* \mathbf{W} \mathbf{u})) \mathbf{Q},$$

(for  $\mathbf{Q}$  see the Correlations and Eigenvalues section) which is not necessarily a correlation matrix, although it is positive semi-definite. This matrix is weighted with variable weights in the same way as  $\mathbf{R}$ .

## Inertia

The inertia per dimension is the eigenvalue per dimension divided by  $m_w$ . The total inertia is the total eigenvalue divided by  $m_w$ .

## Correlations and Eigenvalues

### Before transformation

$$\mathbf{R} = n_w^{-1} \mathbf{H}_c' \mathbf{W} \mathbf{H}_c, \text{ with } \mathbf{H}_c \text{ weighted centered and normalized } \mathbf{H} .$$

For the eigenvalue decomposition of  $\mathbf{R}$  (to compute the eigenvalues), first row  $j$  and column  $j$  are removed from  $\mathbf{R}$  if  $j$  is a supplementary variable, and then  $r_{ij}$  is multiplied by  $(v_i v_j)^{1/2}$ .

If passive missing treatment is applicable for a variable, missing values are imputed with the variable mode, regardless of the passive imputation specification.

### After transformation

After transformation  $p$  correlation matrices are computed ( $s = 1, \dots, p$ ):

$$\mathbf{R}_{(s)} = n_w^{-1} \mathbf{Q}'_{(s)} \mathbf{W} \mathbf{Q}_{(s)},$$

$$\text{with } \mathbf{q}_{(s)j} = n_w^{1/2} \mathbf{G}_j \mathbf{Y}_{(j)s} \left( \mathbf{Y}'_{(j)s} \mathbf{D}_j \mathbf{Y}_{(j)s} \right)^{-1/2}.$$

Usually, for the higher eigenvalues, the first eigenvalue of  $\mathbf{R}_{(s)}$  is equal to  $\lambda_s^{1/2}$  (see Model Summary section). The lower values of  $\Lambda^{1/2}$  are in most cases the second or subsequent eigenvalues of  $\mathbf{R}_{(s)}$ .

If there are missing values, specified to be treated as passive, the mode of the quantified variable or the quantification of an extra category (as specified in syntax; if not specified, default (mode) is used) is imputed before computing correlations. Then the eigenvalues of the correlation matrix do not equal  $\Lambda^{1/2}$  (see Model Summary section). The quantification of an extra category is computed as

$$\mathbf{Y}_{(j)(k_{j+1})s} = \left( \sum_{i \in I} w_i \right)^{-1} \sum_{i \in I} w_i x_{is},$$

with  $I$  an index set recording which objects have missing values.

For the eigenvalue decomposition of  $\mathbf{R}$  (to compute the eigenvalues), first row  $j$  and column  $j$  are removed from  $\mathbf{R}$  if  $j$  is a supplementary variable, and then  $r_{jj}$  is multiplied by  $(v_i v_j)^{1/2}$ .

### Discrimination measures

The discrimination measures are the dimensionwise variances of the quantified variables, which are equal to the dimensionwise squared correlations of the quantified variables with the object scores. For variable  $j$  and dimension  $s$  the discrimination measure is

$$\text{Discr}_{js} = n_w^{-1} \mathbf{y}'_{(j)s} \mathbf{D}_j \mathbf{y}_{(j)s},$$

which is equal to the squared correlation between  $\mathbf{G}_j \mathbf{y}_{(j)s}$  and  $\mathbf{x}_s$ .

## Object Scores and quantifications

### Normalization

If  $\Lambda^{1/2}$  gives the eigenvalues, then  $\Lambda^{1/4}$  gives the singular values, that can be used to spread the inertia over the object scores  $\mathbf{X}$  and the category quantifications  $\mathbf{Y}$ .

During the optimization phase, variable principal normalization is used, then  $\mathbf{X}^n = \mathbf{X}$  and  $\mathbf{Y}^n = \mathbf{Y}$ , else

$$\mathbf{X}^n = \mathbf{X}(m_w^{-1}\Lambda)^{a/4}$$

$$\mathbf{Y}^n = \mathbf{Y}(m_w^{-1}\Lambda)^{1/4(b-1)},$$

with  $a = (1+q)/2$ ,  $b = (1-q)/2$ , and  $q$  any real value in the closed interval  $[-1,1]$ , except for independent normalization: then there is no  $q$  value and  $a=b=1$ .  $q=-1$  is equal to variable principal normalization,  $q=1$  is equal to object principal normalization,  $q=0$  is equal to symmetrical normalization.

### Contributions object scores

#### Mass

The mass of object  $i$  is

$$\text{Mass}_i = \frac{m_{*ii}}{\text{tr}(\mathbf{M}_*\mathbf{W})}.$$

#### Inertia

The inertia of object  $i$  is

$$\text{Inertia}_i = \frac{1}{m_{*ii}} \sum_{j, h_{ij} \neq 0} \frac{v_j}{d_{(j)h_{ij}}} - \text{Mass}_i,$$

where  $d_{(j)h_{ij}}$  is the frequency of the category of object  $i$  on variable  $j$ , and  $h_{ij} \neq 0$  indicates to exclude a variable if object  $i$  has a missing value on the variable and the missing option for the variable is passive.

### Contribution of point to inertia of dimension

The contribution of object  $i$  to the inertia of dimension  $s$  is

$$\text{Contribution}_{is} = \frac{m_{*ii} x_{is}^2}{n_w m_w}$$

#### Contribution of dimension to inertia of point

The contribution of dimension  $s$  to the inertia of object  $i$  is

$$\text{Contribution}_{si} = \frac{\frac{m_{*ii} x_{is}^2}{n_w m_w} \text{Inertia}_s}{\text{Inertia}_i}$$

#### Quantifications

The quantifications are the centroid coordinates. If a category is only used by supplementary objects (i.e. treated as a passive missing), the centroid coordinates for this category are computed as

$$\mathbf{y}_{(j)r} = n_w^{1/2} n_{jr}^{-1} \sum_{i \in I} \mathbf{x}_i \Lambda^{1/4(b-1)}$$

where  $\mathbf{y}_{(j)r}$  is the  $r^{\text{th}}$  row of  $\mathbf{Y}_j$ ,  $n_{jr}$  is the number of objects that have category  $r$ , and  $I$  is an index set recording which objects are in category  $r$ .

#### Contributions quantifications

##### Mass

The mass of category  $r$  of variable  $j$  is

$$\text{Mass}_{(j)r} = \frac{d_{(j)rr}}{\text{tr}(\mathbf{M}_* \mathbf{W})}$$

##### Inertia

The inertia of category  $r$  of variable  $j$  is

$$\text{Inertia}_{(j)r} = \frac{\sum_{i=1}^n w_i m_{*ii} g_{(j)ir}}{d_{(j)rr}} - \text{Mass}_{(j)r},$$

if there are no missing values with missing option passive, this is equal to

$$\frac{1}{m_w} - \text{Mass}_{(j)r}, \text{ and then the total inertia for variable } j \text{ is } \frac{v_j(k_j-1)}{m_w}.$$

#### Contribution of point to inertia of dimension

The contribution of category  $r$  of variable  $j$  to the inertia of dimension  $s$  is

$$\text{Contribution}_{(j)rs} = \frac{d_{(j)rr} \frac{y_{(j)rs}^2}{n_w m_w}}{\text{Inertia}_s},$$

the total contribution of variable  $j$  to the inertia of dimension  $s$  is  $\frac{v_j \text{Discr}_{js}}{\lambda_s^{1/2}}$ .

#### Contribution of dimension to inertia of point

The contribution of dimension  $s$  to the inertia of category  $r$  of variable  $j$  is

$$\text{Contribution}_{s(j)r} = \frac{d_{(j)rr} \frac{y_{(j)rs}^2}{n_w m_w}}{\text{Inertia}_{(j)r}}.$$

### Residuals

Residuals gives plots per dimension of  $\mathbf{G}_j \mathbf{y}_{(j)s}^n$  against the approximation  $\mathbf{x}_s^n$ .

### References

- Cliff, N. 1966. Orthogonal rotation to congruence. *Psychometrika*, 31: 33–42.
- Gifi, A. 1990. *Nonlinear multivariate analysis*. Leiden: Department of Data Theory.
- Max, J. (1960), Quantizing for minimum distortion. *IRE Transactions on Information Theory*, 6, 7-12.
- Wilkinson, J. H. 1965. *The algebraic eigenvalue problem*. Oxford: Clarendon Press.