

## 6.5 Correspondence Analysis

### Introduction

Correspondence analysis is an *exploratory data analytic* technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. As opposed to traditional hypothesis testing designed to verify *a priori* hypotheses about relations between variables, exploratory data analysis is used to identify systematic relations between variables when there are no (or rather incomplete) *a priori* expectations as to the nature of those relations.

Correspondence analysis is also a (*multivariate*) *descriptive data analytic* technique. Even the most commonly used statistics for simplification of data may not be adequate for description or understanding of the data. Simplification of data provides useful information about the data, but that should not be at the expense of valuable information. Correspondence analysis remarkably simplifies complex data and provides a detailed description of practically every bit of information in the data, yielding a simple, yet exhaustive analysis.

Correspondence analysis has several features that distinguish it from other techniques of data analysis. An important feature of correspondence analysis is the multivariate treatment of the data through simultaneous consideration of multiple categorical variables. The multivariate nature of correspondence analysis can reveal relationships that would not be detected in a series of pair wise comparisons of variable. Another important feature is the graphical display of row and column points in biplots, which can help in detecting structural relationships among the variable categories and objects (*i.e.*, cases). Finally, correspondence analysis has highly flexible data requirements. The only strict data requirement is a rectangular data matrix with non-negative entries. Correspondence analysis is most effective if the following conditions are satisfied:

- The data matrix is large enough, so that visual inspection or simple statistical analysis cannot reveal its structure
- The variables are homogeneous, so that it makes sense to calculate the statistical distances between the rows or columns.
- The data matrix is *a priori* "amorphous", *i.e.*, its structure is either unknown or poorly understood.

A distinct advantage of correspondence analysis over other methods yielding joint graphical displays is that it produces two dual displays whose row and column geometries have similar interpretations, facilitating analysis and detection of relationships. In other multivariate approaches to graphical data representation, this duality is not present.

*In a nutshell*, correspondence analysis (CA) may be defined as a special case of principal components analysis (PCA) of the rows and columns of a table, especially applicable to a cross-tabulation. However CA and PCA are used under different circumstances. Principal components analysis is used for tables consisting of continuous measurement, whereas correspondence analysis is applied to contingency tables (*i.e.* cross-tabulations). Its primary goal is to transform a table of numerical information into a graphical display, in which each row and each column is depicted as a point.

The usual procedure for analyzing a cross-tabulation is to determine the probability of *global association* between rows and columns. The significance of association is tested by the *Chi-square* test, but this test provides no information as to which are the significant individual associations between row-column pairs of the data matrix. Correspondence analysis shows how the variables are related, not just that a relationship exists.

## Basic Concepts and Definitions

There are certain fundamental concepts in correspondence analysis: which are described below.

### Primitive matrix

The original data matrix,  $N(I, J)$ , or contingency table, is called the primitive matrix or primitive table. The elements of this matrix are  $n_{ij}$ .

### Profiles

While interpreting a cross-tabulation, it makes little sense to compare the actual frequencies in each cell. Each row and each column has a different number of respondents, called the *base* of respondents. For comparison it is essential to reduce either the rows or columns to the same base.

Consider a contingency table  $N(I, J)$  with  $I$  rows ( $i=1, 2, I$ ) and  $J$  columns ( $j=1, 2, \dots, J$ ) having frequencies  $n_{ij}$ . Marginal frequencies are denoted by  $n_{i+}$  and  $n_{+j}$

$$n_{i+} = \sum_j n_{ij}$$

$$n_{+j} = \sum_i n_{ij}$$

Total frequency is given by

$$n = \sum_j \sum_i n_{ij}$$

### Row profiles

The profile of each row  $i$  is a vector of conditional densities:

$$n_{ij} / n_{i+} (j=1, 2, \dots, J)$$

$$i = 1, 2, \dots, I$$

The complete set of the row profile may be denoted by  $I \times J$  matrix **R**.

**Matrix of Row Profiles**

ROWS	COLUMNS		TOTAL
	1	2                      J	
1.	$n_{11}/n_{1+}$ $n_{21}/n_{2+}$	$n_{12}/n_{1+}$ ..... $n_{1j}/n_{1+}$	1
2.	$n_{31}/n_{3+}$	$n_{22}/n_{2+}$ ..... $n_{2j}/n_{2+}$	1
3.	.	$n_{32}/n_{3+}$ ..... $n_{3j}/n_{3+}$	1
.	$n_{i1}/n_{i+}$	.	
I		$n_{i2}/n_{i+}$ ..... $n_{ij}/n_{i+}$	1
<i>Column mass</i>	$n_{+1}/n_{++}$	$n_{+2}/n_{++}$ $n_{+j}/n_{++}$	1

**Column Profiles**

The profile of each column  $j$  is a vector of conditional densities  $n_{ij}/n_{+j} (i=1,2,\dots,I)$ . The complete set of the column profiles may be denoted by  $(i \text{ ' } j)$  matrix **C**.

**Matrix of Column Profiles**

ROWS	COLUMNS		ROW MASS
	1	2                      J	
1.	$n_{11}/n_{+1}$ $n_{21}/n_{+1}$	$n_{12}/n_{+2}$ ..... $n_{1j}/n_{+j}$	$n_{+1}/n_{++}$ $n_{+2}/n_{++}$
2.	$n_{31}/n_{+1}$	$n_{22}/n_{+2}$ ..... $n_{2j}/n_{+j}$	
3.	.	$n_{32}/n_{+2}$ ..... $n_{3j}/n_{+j}$	
.	$n_{i1}/n_{+1}$	.	$n_{+i}/n_{++}$
I		$n_{i2}/n_{+2}$ ..... $n_{ij}/n_{+j}$	
<i>Column mass</i>	1	...1                      1	1

Average row profile                       $\bar{r} = n_{+j}/N \quad (j=1,2,\dots,J)$

Average column profile                       $\bar{c} = n_{i+}/N \quad (i=1,2,\dots,I)$

## Masses

Another fundamental concept in correspondence analysis is the concept of mass. The mass of the  $i$ th row =

Marginal frequency of the  $i$ th row/Grand total

$$= n_{+i}/n$$

Similarly the mass of the  $j$ th column =

Marginal frequency of the  $j$ th column/Grand total

$$= n_{j+}/n$$

## Correspondence matrix

The correspondence matrix  $\mathbf{P}$  is defined as the original table  $\mathbf{N}$  divided by the grand total  $n$ ,  $\mathbf{P} = (1/n) \mathbf{N}$ . Thus, each cell of the correspondence matrix is given by the cell frequency divided by the grand total.

The correspondence matrix shows how one unit of mass is distributed across the cells. The row and column totals of the correspondence matrix are the row mass and column mass, respectively.

### Clouds of Points $N(I)$ and $N(J)$

The cloud of points  $N(I)$  is the set of elements of points  $i \hat{I}$ , whose coordinates are the components of the profile and whose mass is  $n_{+i}/n_{++}$ .

The cloud of points  $N(J)$  is the set of elements of points  $j \hat{J}$ , whose coordinates are the components of the profile and whose mass is  $n_{j+}/n_{++}$ .

## Distances

A variant of Euclidean distance, called the weighted Euclidean distance, is used to measure and thereby depict the distances between profile points. Here, the weighting refers to differential weighting of the dimensions of the space and not to the weighting of the profiles.

Distance between two rows  $i$  and  $i'$  is given by

$$d^2(i, i') = \sum_{j=1}^J \frac{1}{n_{+j}} \left( \frac{n_{ij}}{n_{i+}} - \frac{n_{i'j}}{n_{i'+}} \right)^2$$

In a symmetric fashion, the distance between two columns  $j$  and  $j'$  is given by

$$d^2(j, j') = \sum_{i=1}^I \frac{1}{n_{i+}} \left( \frac{n_{ij}}{n_{+j}} - \frac{n_{ij'}}{n_{+j'}} \right)^2$$

The distance thus obtained is called the *Chi-square distance*. The Chi-square distance differs from the usual Euclidean distance in that each square is weighted by the inverse of the frequency corresponding to each term.

The division of each squared term by the expected frequency is "variance – standardizing" and compensates for the larger variance in high frequencies and the smaller variance in low frequencies. If no such standardization were performed, the differences between larger proportions would tend to be large and thus dominate the distance calculation, while the differences between the smaller proportions would tend to be swamped. The weighting factors are used to equalize these differences.

Essentially, the reason for choosing the Chi-square distance is that it satisfies the principle of *distributional equivalence*, expressed as follows:

- If two rows  $i$  and  $i'$  of  $I$  of  $N(I, J)$  are proportioned and if they are replaced by only one, which is the sum, column-by-column, then the distances between columns are not changed in  $N(J)$ .
- If two columns  $j$  and  $j'$  of  $J$  of  $N(I, J)$  are proportioned and if they are replaced by only one, which is the sum, row-by-row, then the distances between rows are not changed in  $N(I)$ .

### **Inertia**

Inertia is a term borrowed from the "moment of inertia" in mechanics. A physical object has a center of gravity (or centroid). Every particle of the object has a certain mass  $m$  and a certain distance  $d$  from the centroid. The moment of inertia of the object is the quantity  $md^2$  summed over all the particles that constitute the object.

$$\text{Moment of inertia} = \sum m d^2$$

This concept has an analogy in correspondence analysis. There is a cloud of profile points with masses adding up to 1. These points have a centroid (i.e., the average profile) and a distance (Chi-square distance) between profile points. Each profile point contributes to the inertia of the whole cloud. The inertia of a profile point can be computed by the following formula.

For the  $i$ th row profile,

$$\text{Inertia} = m_i \sum_j \frac{(r_{ij} - \bar{r}_j)^2}{\bar{r}_j}$$

where  $r_{ij}$  is the ratio  $n_{ij}/n_{i+}$  and  $\bar{r}_j$  is  $n_{.j}/n$

The inertia of the  $j$ th column profile is computed similarly.

The total inertia of the contingency table is given by:

$$\text{Total inertia} = \sum_i \sum_j \frac{(p_{ij} - r_i c_j)^2}{r_i c_j}$$

which is the Chi-square statistic divided by  $n$ ?

### Reduction of Dimensionality

Another way of looking at correspondence analysis is to consider it as a method for decomposing the overall inertia by identifying a small number of dimensions in which the deviations from the expected values can be represented. This is similar to the goal of factor analysis, where the total variance is decomposed, so as to arrive at a lower-dimensional representation of variables that allows one to reconstruct most of the variance/covariance matrix of variables.

### Criterion for Dimensionality Reduction

In correspondence analysis, we are essentially looking for a low-dimensional subspace, which is as close as possible to the set of profile points in the high-dimensional true space. Let  $S$  denote any candidate subspace. For the  $i$ :th profile point, we can compute the Chi-square distance between the profile point and  $S$ , denoted by  $d_i(S)$ . The weighted measure of the distance of the profile point and the subspace is given by:

$$r_i [d_i(S)]^2$$

The distance of all the profiles to the subspace  $S$  is given by:

$$\sum r_i [d_i(S)]^2$$

The objective of correspondence analysis is to discover which subspace  $S$  minimizes the above criterion.

The criterion used for dimensionality reduction implies that the inertia of a cloud in the optimal subspace is maximum, but that would still be less than that in the true space. What is lost in this process is the knowledge of how far and in which direction the profiles lie off this subspace. What is gained is a view of the profiles, which otherwise would not be possible. The ratio of inertia inside the subspace to the total inertia gives a measure of the accuracy of representation of a cloud in the subspace.

Correspondence analysis determines the principal axes of inertia and for each axis the corresponding eigenvalue, which is the same as the inertia of the cloud in the direction of the axis. The first factorial axis is the line in the direction of which the inertia of the cloud is a maximum. The second factorial axis is, among all the lines that are perpendicular to the first factorial axis, the one in whose direction the inertia of the cloud is a maximum. The third factorial axis is, among all the lines that are perpendicular to both the first and second factorial axes, the line in whose direction the inertia of the cloud is a maximum, and so on. The optimal subspace is a subspace spanned by the principal axes. The inertia of a profile along a principal axis is called the *Principal Inertia*.

Geometrically, the principal inertia is the weighted average of the Chi-squared distances from the centroid to the

projections of the row profiles on the respective principal axis. It is an absolute measure of the dispersion of the row profiles in the direction of that axis. Each principal inertia can be decomposed into components due to each row profile (or column profile). Rows, which contribute highly to a principal axis, largely determine the orientation and the identity of the corresponding principal axis.

The cosines of the row profiles' deviation vectors from the centroid and the principal axis describe how closely each profile vector lies or correlates with a principal axis. Thus, they measure how well the display approximates the profile's true position.

The eigenvalues ( $l_i$ ), corresponding to the sequence of the principal axes are in the decreasing order of magnitude:

$$l_1 > l_2 > l_3 > \dots > l_k$$

### Row and Column Analyses

The row analysis of a matrix consists in situating the row profiles in a multidimensional space and finding the low-dimensional subspace, which comes closest to the profile points. The row profiles are projected onto such a subspace for interpretation of the inter-profile positions. Similarly, the analysis of column profiles involves situating the column profiles in a multidimensional space and finding the low-dimensional subspace, which comes closest to the profile points.

The row and column analyses are intimately connected. If a row analysis is performed, the column analysis is also *ipso facto* performed, and *vice versa*. The two analyses are equivalent in the sense that each has the same total inertia, the same dimensionality and the same decomposition of inertia into principal inertias along principal axes.

#### *Row and Column Contributions to Inertia*

- The total inertia of a table quantifies how much variation is present in the row profiles or in the column profiles.
- Each row and each column makes a contribution to the total inertia, respectively called row inertia and column inertia. The principal inertia of the row (or column) points is the inertia of the row (or column) points projected onto the axis. Thus, each row or column makes a contribution to the principal inertia. The component of row inertia or column inertia along a principal axis is called the principal inertia.

These contributions can be expressed in relative terms:

- The contribution of a row (or column) to a – axis, relative to the corresponding principal inertia. This is the relative contribution of a row (column) to the composition of the a – axis, usually denoted by *CTR* (a), which allows diagnosing as to which points play a major role in the orientation of a principal axis.
- The contribution of a row (column) to a – axis, relative to the corresponding point's inertia. This is called the contribution of a point to the eccentricity of the axis. denoted as *COR* (a ). This allows diagnosing the position of each point whether it is well represented or poorly represented on a given axis

#### *Maximum number of dimensions*

Since the sums of the frequencies across the columns must be equal to the row totals, and the sums across the rows equal to the column totals, there are in a sense only (number,  $J$ , of columns-1) independent entries in each row, and (number,  $I$ , of rows-1) independent entries in each column of the contingency table. Thus, the maximum number of eigenvalues that can be extracted from a two-way table is equal to the minimum of [ the number of columns minus 1, and the number of rows minus 1 ] . If we choose to extract (*i.e.*, interpret) the maximum number of dimensions that can be extracted, then we can reproduce exactly all the information contained in the table.

### Interpretation of correspondence analysis

The interpretation of the results of correspondence analysis comprises the interpretation of numerical results and factor graphics, yielded by CA. The former implies selection of significant axes and significant points.

### Selection of Significant Axes

How many axes are significant and should be retained for further analysis or interpretation? Here significant means 'necessary to study in detail' – not in terms of statistical significance tests. Two types of factor axes are considered: *First order* factor axes and *Second order* factor axes. First order factor axes are considered on the basis of contributions to the total variance (or inertia), whereas the second order factor axes are considered on the basis of contributions to the eccentricity, that is.  $\text{COS}^2 j$  .

Correspondence analysis issues eigenvalues for the  $\min[(I, J)-1]$  factor axes; the eigenvalues are ranked in the decreasing order of magnitude.

*First order factor axes:*

The number of (significant) axes,  $M$ , can be determined by any of the following rules:

1. Sum of the inertia explained by the first  $M$  axes exceeds a certain threshold, typically 80% of the total inertia.
2. Choose all the axes whose eigenvalues exceed  $1/[\min(I, J)-1]$

*Second order factor axes:*

After having selected the first order factor axes, the second order factor axes are selected as follows:

Let  $M_i$  be the rank of a factor axis for which a point  $i$  of  $N(I)$  and or  $j$  of  $N(J)$  exists, such that

$$\text{COS}^2 j(i) \geq k$$

or

$$\text{COS}^2 j(j) \geq k$$

where  $k$  is typically = 0.25.

Thus, the number of axes chosen for interpretation =  $M + M_i$ .

## Rules for interpreting factorial axes by individual points

### Explicative points

An explicative is a point whose absolute contribution  $CTR_a(i)$  (for  $i \in I$ ) or  $CTR_a(j)$  (for  $j \in J$ ) are distinctly higher than the contributions of other points. The points  $i \in I$  whose contributions are higher than the average of the whole contribution are considered as *explicative*. The explicative points can be selected according to any of the following criteria:

- $CTR_a(i) \geq \text{average } CTR_a$  of all points
- The points  $i \in I$  are ordered by their contribution to  $CTR_a(i)$ , in the decreasing order. Then, the sum  $\{ \sum CTR_a(i) \geq p \}$  is truncated at the lowest value  $i_0 \in I$  such that the truncated sum is  $\geq p$ . The set  $\{i \in I \mid i \leq i_0\}$  is the set of explicative points. The same procedure is followed for  $J$ .

### Explained points

The points explained by an  $a$ -axis are the variable points  $i$  of  $N(I)$  [ or  $j$  of  $N(J)$  ], whose contributions to the eccentricity are greater than a certain threshold. The contributions are similar to a squared coefficient of correlation ( $COS^2_j$ ). Usually a threshold of 0.25 is used.

A point  $j$  can be an explained point (by an  $a$ -axis) without being an explicative point. Suppose that point  $i$  has an absolute contribution 40% and a squared correlation of 0.15 to an axis. This means that it contributes strongly to the creation of the axis, but it probably participates in the building of many other axes.

Thus, two sets of coefficients are calculated for each axis. These coefficients apply equally to the rows and columns of the data matrix.

*Absolute contributions*, which indicate the proportion of variance (i.e., inertia) explained by each variable in relation to each principal axis. This proportion is calculated with respect to the entire set of variables.

*The squared correlations*, which indicate the part of the variance of a variable explained by a principal axis.

The interpretation of absolute contributions is opposite to that of the relative contributions ( $COS^2_j$ ). The latter indicate the extent to which each row category and each column category is described by the axis. The contribution to inertia, on the other hand, indicates the extent to which the geometric orientation of the axis is determined by the single variable categories

## Quality of representation

The quality of representation of a point in the coordinate system, defined by the chosen numbers of dimensions, is defined as the ratio of the squared distance of the point from the origin in the chosen number of dimensions, over the squared distance from the origin in the space defined by the maximum number of dimensions. It is also equal to the sum of  $COS^2_j$ .

$$\text{Quality} = \sum \text{COS}^2 j$$

A low quality means that the current number of dimensions does not represent well the respective column or row point.

### Supplementary elements

The most distinguishing feature of correspondence analysis is the possibility of introducing supplementary elements (variables or objects) into factor graphics. The supplementary elements do not contribute to the orientation of the factorial axis, but their relative contributions to the factorial axes and their coordinates are computed by the correspondence analysis. A simple way to think of such points is that they have a position in the full space, but no mass.

Supplementary points are additional rows and columns of a contingency table, which have meaningful profiles and which exist in the full space of row and column profiles. They can be projected onto the low-dimensional subspace and their positions relative to the active elements can be determined.

The relative contribution of a supplementary point to the eccentricity of an axis (*i.e.*,  $\text{COS}^2 j$ ) can be used to judge whether the supplementary point lies to a larger or lesser extent in the plot rather than out of it. This procedure is used:

- To suppress a particular point in a factor analysis graphic and then to re-introduce it as a supplementary point. This is usually done when a particular point is an outlier
- To classify elements, whose description in terms of profiles, is missing or incomplete. In that case the data elements are estimated and then these points are re-introduced as supplementary elements into the graphics.
- To compare similar data matrices for two different time points or for two different countries or regions, etc.

### Outlier points

Outlier points plague correspondence analysis. Occasionally, a row or column profile is rare in its set of points that it has a minor role in the determination of the higher order axes. This situation can be discerned easily by considering the point's contribution to the axes. When a point has a large contribution (*CTR*), at a large principal coordinate at a major principal axis, it is called an outlier. Outlier points should be treated as supplementary variables.

### Graphics

As in principal components analysis, the results of correspondence analysis are presented on graphs that represent the configurations of points in projection planes, formed by the first principal axes taken two at a time. It is customary to summarize the row and column coordinates in a single plot. However, it is important to remember that in such plots, one can only interpret the distances between row points, and the distances between column points, but not the distances between row points and column points. However, it is legitimate to interpret the relative positions of one point of one set with respect to all the points of the other set

The joint display of row and column points shows the relation between a point from one set and all points of another set, not between individual points between each set. Except in special cases, it is extremely dangerous to interpret

the proximity of two points corresponding to different sets of points.

Some keys for interpreting the factorial maps are:

- Points near the origin have undifferentiated profile distribution as a consequence of the origin being placed at the center of gravity of both clouds N (I) and N (J).
- The points, which do not contribute essentially to the inertia of each axis, are virtually identical to the average profile.
- Points of a cloud (or set) situated away from the origin, but close to each other have similar profiles
- Geometrically, a particular row profile would be attracted to a position in its subspace that corresponds to column variable categories prominent in that row profile.
- When correspondence analysis has more than two dimensions. Proximity with one pair of axes may disappear when other axes are (added) plotted.
- It is customary to summarize the row and column coordinates in a single plot. However, it is important to remember that in such plots, one can only interpret the distances between row points, and the distances between column points, but not the distances between row points and column points. cannot be interpreted. The joint display of coordinates shows the relation between a point from one set and *all* points of the other set and not between individual points between each set.
- A point makes a high contribution to the inertia of a principal axis in two ways –when it has a large distance from the barycenter, even if it has a small mass, or when it has a large mass, but a small distance. Considering all these points, it is necessary that the numerical results of correspondence analysis, *viz.* mass. Absolute contribution (CTR) and relative contribution  $\text{COS}^2_j$  are all taken into account for interpreting the results of correspondence analysis.

## Mathematics of Correspondence Analysis

### Notation

Contingency table N (I' J)

Row mass = row sums/grand total =  $n_{i+}/n$

Column mass = column sums/grand total =  $n_{+j}/n$

Correspondence matrix is defined as the original table (or matrix) N divided by the grand total  $n$ .

$$P = \left( \frac{1}{n} \right) N$$

The matrix of row profiles can also be defined as the rows of the correspondence matrix  $\mathbf{P}$  divided by their respective row sums (*i.e.* row masses), which can be written as:

$$\text{Matrix of row profile} = \mathbf{D}_r^{-1} \mathbf{P}$$

where  $\mathbf{D}_r$  is the diagonal matrix of row masses.

The matrix of column profiles consists of the columns of the correspondence matrix  $\mathbf{P}$  divided by their respective column sums.

$$\text{Matrix of column profiles} = \mathbf{D}_c^{-1} \mathbf{P}$$

where  $\mathbf{D}_c$  is the diagonal matrix of the column masses.

The correspondence analysis problem is to find a low-dimensional approximation to the original data matrix that represents both the row and column profiles

$$\mathbf{R} = \mathbf{D}_r^{-1} \mathbf{P}$$

$$\mathbf{C} = \mathbf{D}_c^{-1} \mathbf{P}$$

In a low  $k$ -dimensional subspace, where  $k$  is less than  $I$  or  $J$ . These two  $k$ -dimensional subspaces (one for the row profiles and one for the column profiles) have a geometric correspondence that enables us to represent both the rows and columns in the same display.

Since we wish to graphically represent the distances between row (or column) profiles, we orient the configuration of points at the center of gravity of both sets. The centroid of the set of row points in its space is the vector of column masses. The centroid of the set of column point in its space is  $\mathbf{r}$ , the vector of row masses. This is the average column profile.

To perform the analysis with respect to the center of gravity,  $\mathbf{P}$  is centered "symmetrically" by rows and columns, *i.e.*,  $\mathbf{P} - \mathbf{rc}^T$  so that it correspondence to the average profiles of both sets of points. The solution to finding a representation of both sets of points is the singular value decomposition of the matrix of standardized residuals *i.e.*,  $I \times J$  matrix with elements:

$$a_{ij} = (p_{ij} - r_i c_j) / \sqrt{r_i c_j}$$

The singular value decomposition (SVD) is defined as the decomposition of an  $I \times J$  matrix  $\mathbf{A}$  as the product of three matrices

$$\mathbf{A} = \mathbf{U} \mathbf{G} \mathbf{V}^T \tag{1}$$

where the matrix  $\mathbf{G}$  is a diagonal matrix of positive numbers in decreasing order:

$$g_1 \geq g_2 \geq \dots \geq g_k > 0 \tag{2}$$

where  $k$  is the rank of  $\mathbf{A}$ , and the columns of the matrices  $\mathbf{U}$  and  $\mathbf{V}$  are orthonormal, *i.e.*,

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \quad \mathbf{V}^T \mathbf{V} = \mathbf{I} \quad (3)$$

where  $\mathbf{U}^T$  is the transpose of  $\mathbf{U}$ , and  $\mathbf{V}^T$  is the transpose of  $\mathbf{V}$ .

$g_1, g_2, \dots, g_k$  are called singular values.

Columns of  $\mathbf{U}$  ( $u_1, u_2, \dots, u_k$ ) are called left singular vectors.

Columns of  $\mathbf{V}$  ( $v_1, v_2, \dots, v_k$ ) are called right singular vectors.

Consider a set of  $I$  points in  $J$ -dimensional space, where coordinates are in the rows of the matrix  $\mathbf{Y}$  with masses  $m_1, m_2, \dots, m_I$  assigned to the respective points, where the space is structured by the weighted Euclidean (with dimension weights  $q_1, q_2, \dots, q_J$  associated with the respective dimensions). In other words, the distance between any two points, say  $\mathbf{x}$  and  $\mathbf{y}$ , is equal to

$$[ (\mathbf{x} - \mathbf{y})^T \mathbf{D}_q (\mathbf{x} - \mathbf{y}) ]^{1/2} \quad (4)$$

Let  $\mathbf{D}_m$  and  $\mathbf{D}_q$  be the diagonal matrices of point masses and dimension weights respectively

Let  $\mathbf{m}$  be the vector of point masses (we have already assumed that  $\sum m_i = 1$ ):

$$\mathbf{1}^T \mathbf{m} = 1$$

where  $\mathbf{1}$  is the vector of ones.

Any low-dimensional configuration of the points can be derived directly from the singular value decomposition of the matrix:

$$\mathbf{A} = \mathbf{D}_m^{1/2} (\mathbf{Y} - \mathbf{Y} \bar{\mathbf{x}}^T) \mathbf{D}_q^{1/2} \quad (5)$$

where  $\bar{\mathbf{x}}^T$  is the centroid of the rows of  $\mathbf{Y}$ .

Applying singular value decomposition to the above equation, we find that principal coordinates of row points (i.e. projections of row profiles onto principal axes) are contained in the following matrix:

$$\mathbf{F} = \mathbf{D}_m^{-1/2} \mathbf{U} \mathbf{G} \quad (6)$$

The coordinates of the points in an optimal  $a$ -dimensional subspace are contained in the first  $a$  columns. The principal axes of this space are contained in the matrix

$$\mathbf{A} = \mathbf{D}_q^{-1/2} \mathbf{V}$$

Here, we have two special cases of the above general result, *viz.* *Row problem* and *Column problem*. These problems involve the reduction of dimensionality of the row profiles and the column profiles, where each set of points has its associated masses and Chi-square distances. Both these problems reduce to singular value decomposition of the same matrix of standardized residuals.

### **Row problem**

The row problem consists of a set of  $I$  profiles in the rows of  $\mathbf{D}_r^{-1} \mathbf{P}$  with masses  $\mathbf{r}$  in the diagonal matrix  $\mathbf{D}_r$  in a space with distance defined by the diagonal matrix  $\mathbf{D}_r^{-1}$ . The centroid of the row profiles can be derived as follows

$$\mathbf{r}^T \mathbf{D}_r^{-1} \mathbf{P} = \mathbf{I}^T \mathbf{P} = \mathbf{c}^T$$

where  $\mathbf{c}^T$  is the row vector of the column masses

The matrix  $\mathbf{A}$  in (Equation 5) can be written as

$$\mathbf{A} = \mathbf{D}_r^{1/2} (\mathbf{D}_r^{-1} \mathbf{P} - \mathbf{I} \mathbf{c}^T) \mathbf{D}_c^{-1/2} \quad (7)$$

which can be rewritten as

$$\mathbf{A} = \mathbf{D}_r^{-1/2} (\mathbf{P} - \mathbf{y} \mathbf{c}^T) \mathbf{D}_r^{-1/2} \quad (8)$$

### **Column problem**

The column problem consists of a set of  $J$  profiles in the columns of  $\mathbf{P} \mathbf{D}_c^{-1}$  with masses  $\mathbf{c}$  in the diagonal of  $\mathbf{D}_c$  in a space with distance defined by the diagonal matrix  $\mathbf{D}_c^{-1}$ .

By transposing the matrix  $\mathbf{P} \mathbf{D}_c^{-1}$  of column profiles, we obtain  $\mathbf{D}_c^{-1} \mathbf{P}^T$ . The centroid of these profiles is  $\mathbf{c}^T \mathbf{D}_c^{-1} \mathbf{P}^T = \mathbf{I}^T \mathbf{P}^T = \mathbf{r}^T$  (*i.e.* the row vector of row masses).

The matrix in Equation (5)

$$\mathbf{A} = \mathbf{D}_m^{1/2} (\mathbf{Y} - \mathbf{I} \mathbf{r}^T) \mathbf{D}_y^{-1/2} \quad (9)$$

can be written as

$$\mathbf{A} = \mathbf{D}_c^{1/2} (\mathbf{D}_c^{-1} \mathbf{P}^T - \mathbf{I} \mathbf{r}^T) \mathbf{D}_y^{-1/2}$$

$$A = D_r^{-1/2}(P^T - cr^T)D_c^{-1/2}$$

This is the transpose of the matrix derived for **A**, the row problem. It follows that both the row and column problems can be solved by singular value decomposition of the same matrix of standardized residuals:

$$A = D_r^{-1/2}(P - rc^T)D_c^{-1/2} \quad (10)$$

The elements of this  $I \times J$  matrix are:

$$a_{ij} = (p_{ij} - r_i c_j) \sqrt{r_i c_j} \quad (11)$$

It can be easily seen that the centroid of these profiles is:

$$c^T D_c^{-1} P^T = I^T P^T = r^T \quad (\text{the row vector of } r \text{ masses})$$

The matrix in Equation 5 is thus reduced to

$$\begin{aligned} A &= D_c^{1/2}(D_c^{-1} P^T - I r^T) D_r^{-1/2} \\ &= D_r^{-1/2}(P^T - cr^T) D_c^{-1/2} \end{aligned} \quad (12)$$

It can be easily seen that the matrix **A** is the transpose of the matrix derived for the row problem. These results imply that both the row problem and column problems are solved by computing the singular value decomposition of the same matrix (i.e. the matrix of the standard residuals).

$$A = D_r^{-1/2}(P - rc^T)D_c^{-1/2} \quad (13)$$

whose elements are:

$$a_{ij} = (p_{ij} - r_i c_j) \sqrt{r_i c_j} \quad (14)$$

It follows from Equation ( 10 ) that the Chi-square statistic can be decomposed into  $I \times J$  components of the form:

$$\frac{n(p_{ij} - r_i c_j)^2}{r_i c_j}$$

The sum of squares of the elements of **A** is the total inertia of the contingency table.

$$\text{Total inertia} = \sum_i \sum_j \frac{(p_{ij} - r_i c_j)^2}{r_i c_j}$$

which is the chi-square statistic divided by  $n$ .

Thus, there are  $k = \min [I-1, J-1]$  dimensions in the solution. The squares of the singular values of  $\mathbf{A}$  i.e. the eigenvalues of  $\mathbf{A}\mathbf{A}$  or  $\mathbf{A}\mathbf{A}^T$  also decompose the total inertia. These are denoted by  $\lambda_1, \lambda_2, \dots, \lambda_k$  and are called the principal inertias.

The principal coordinates of the row problem are:

$$\mathbf{F} = D_r^{-1/2} \mathbf{u} \mathbf{G} \quad (15)$$

or in the scalar notation:

$$f_{ik} = u_{ik} r_k / \sqrt{r_i} \quad (16)$$

The principal coordinates of the columns are obtained from:

$$\mathbf{G} = D_c^{-1/2} \mathbf{v} \mathbf{G}$$

or in the scalar notation:

$$g_{kj} = v_{jk} r_k / \sqrt{c_j}$$

The standard coordinates of the rows are the principal coordinates divided by their respective singular values, i.e.

$$\mathbf{X} = \mathbf{F} \mathbf{G}^{-1} = D_r^{-1/2} \mathbf{U} \quad (17)$$

or in the scalar notation

$$x_{ik} = f_{ik} / r_k$$

The standard coordinates of the columns are the principal coordinates divided by their respective singular values:

$$\mathbf{Y} = \mathbf{G} \mathbf{G}^{-1} = D_c^{-1/2} \mathbf{V} \quad (18)$$

i.e.

$$y_{jk} = g_{jk} / r_k$$

Each principal inertia  $\lambda_k$  is decomposed into components  $x_i f_{ik}^2$  for each row  $i$ :

$$\lambda_k = \sum r_i f_{ik}^2$$

or in the matrix notation

$$D_k = \mathbf{F}^T D_r \mathbf{F} \quad (19)$$

The contribution of the rows to the principal inertia  $l_k$  is equal to:

$$r_i f_{ik}^2 / \lambda_k$$

For the  $i$ th row, the inertia components for all  $k$  axes sum up to the row inertia of the  $i$ th row:

$$\sum_j (p_{ij} / r_i - c_j)^2 / c_j = \sum_k r_i f_{ik}^2$$

The left hand side of the above equation is identical to the sum of squared elements in the  $i$ th row of **A**

$$\sum_j s_{ij}^2 = \sum_j (p_{ij} - r_i c_j)^2 / (r_i c_j)$$

or

$$D_r = G^T D_c G \quad (20)$$

There are  $k = \min [I-1, J-1]$  dimensions in the solution. The square of the singular values of A, are denoted by are called singular values.

The principal coordinates of the rows are obtained using [Equation (6)], for the row problem.

$$F = D_r^{-1/2} U \Gamma \quad (21)$$

or in scalar notation:

$$f_{ik} = u_{ik} / \sqrt{r_i}$$

Similarly the principal coordinates of the columns are obtained using Equation (6), for the column problem.

$$G = D_c^{-1/2} V \Gamma \quad (22)$$

*i.e.*

$$g_{jk} = v_{jk} / \sqrt{c_j}$$

The standard coordinates of the rows are the principal coordinates divided by their respective singular values:

$$X = F \Gamma^{-1} = D_r^{-1/2} U \quad (23)$$

*i.e.*

$$x_{ik} = f_{ik} / \lambda_k$$

The standard coordinates of the columns are the principal coordinates divided by their respective singular values:

$$Y = G\Gamma^{-1} = D_c^{-1/2}V$$

*i.e.*

$$y_{ik} = g_{ik} / \gamma_k$$