

Correspondence Analysis and Factor Analysis

Correspondence Analysis (CA)

Computation

Worked Example

Software Example (MVSP)

Arch Effects

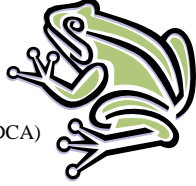
Detrended Correspondence Analysis (DCA)

Detrending by segments

Detrending by polynomials

To detrend or not...

Factor Analysis (FA)



Correspondence Analysis (CA)

Correspondence was developed independently by several authors over a period of ca. 30 years and given many different names in the literature:

Contingency table analysis

RQ-technique

Reciprocal averaging

Correspondence analysis

Reciprocal ordering

Dual scaling

Homogeneity analysis

Correspondence Analysis (CA)

Correspondence analysis was first proposed for analyzing *two-way contingency tables*. In such tables, the states of the first descriptor (rows) are compared to the states of the second descriptor (columns). Data in each cell of the table are *frequencies*. These frequencies are positive integers or zeroes.

In EEB, the most common application of CA is for the analysis of species data (0/1, or abundance) at different sampling sites. A species-site table essentially contains frequencies.

Correspondence Analysis (CA)

In general, CA may be applied to any data table that is dimensionally homogeneous (i.e., the physical dimensions of all variables are the same) and only contains positive integers or zeroes.

The χ^2 distance (D_{16}), which is a coefficient that excludes double-zeroes, is used to quantify the relationship among rows and columns. NB: Some authors have questioned the efficacy of χ^2 distance for certain types of data.

Correspondence Analysis (CA)

CA is primarily a method of ordination. As such it is similar to PCA; it preserves in the space of the principal axes (i.e., after rotation), the Euclidean distance between profiles of weighted conditional probabilities.

In other words, CA preserves the χ^2 distance between the rows and the columns of the contingency table.

CA Computation

Correspondence analysis proceeds along three steps:

- (1) the contingency table is transformed into a table of contributions to the Pearson chi-square statistic after fitting a null model to the table.
- (2) Singular value decomposition (SVD) to that table and the eigenvalues and eigenvectors are computed (as in PCA).
- (3) Further matrix manipulations lead to the tables required for plotting in ordination space.

CA Computation

Let's consider a contingency table with r rows and c columns. Assume that the table is constructed in such a way that $r > c$ (the table can be transposed to meet this condition).

LEXICON:

- Absolute frequencies will be represented as f_{ij} and relative frequencies (proportions) as p_{ij} .
- p_{ij} is the frequency f_{ij} in cell ij divided by its sum f_{++} of the f_{ij} 's over the whole table.
- The table containing the p_{ij} 's is called \mathbf{Q} ; its size is $r \times c$.
- Row weight p_{i+} is equal to f_{i+}/f_{++} , where f_{i+} is the sum of the values in row i . Vector $[p_{i+}]$ is of size r .
- Likewise, column weight p_{+j} is equal to f_{+j}/f_{++} , where f_{+j} is the sum of the values in column j . Vector $[p_{+j}]$ is of size c .

CA Computation - Step 1

The Pearson chi-square statistic, χ^2_p , is a sum of squared χ_{ij} values, computed for every cell ij of the contingency table. Each χ_{ij} value is the standardized residual of a frequency f_{ij} after fitting a null model to the contingency table. The null model states that there is no relationship between the rows and the columns of the table. For each cell:

$$\chi_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}} = \sqrt{f_{++}} \left[\frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} \right]$$

CA Computation - Step 1

CA is based upon the matrix \mathbf{Q} :

$$\bar{\mathbf{Q}} = [q_{ij}] = \left[\frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} \right]$$

where the q_{ij} values are the basis of CA and only differ from the χ_{ij} 's by a constant ($\sqrt{f_{++}}$). This causes all of the eigenvalues to be ≤ 1 .

The sum of squares of all the values in the $\bar{\mathbf{Q}}$, $\sum q_{ij}^2$, measures the total inertia in \mathbf{Q} . It is also equal to the sum of all eigenvalues to be extracted.

CA Computation - Step 2

Singular value decomposition (SVD) is then applied to the $\bar{\mathbf{Q}}$ matrix. While this process is computationally a bit involved, the same answer is more directly derived by applying eigenvalue analysis to the covariance matrix $\bar{\mathbf{Q}}'\bar{\mathbf{Q}}$, which would produce the the matrices of eigenvalues Λ and eigenvectors \mathbf{U} . Once \mathbf{U} is obtained, one can easily solve for $\hat{\mathbf{U}}$ because of the relationship:

$$\hat{\mathbf{U}}_{(r \times c)} = \mathbf{Q} \bar{\mathbf{U}} \Lambda^{-1/2}$$

This process always yields one null eigenvalue due to the centering in the determination of $\bar{\mathbf{Q}}$.

CA Computation - Step 3

Matrices \mathbf{U} and $\hat{\mathbf{U}}$ may be used to plot the positions of the row and column vectors in two separate scatter diagrams. For joint plots, various scaling types of the row and column scores have been proposed.

First, matrices \mathbf{U} and $\hat{\mathbf{U}}$ can be weighted by the inverse of the square roots of the column and row scores, written out in diagonal matrices $\mathbf{D}(\mathbf{p}_{+j})^{-1/2}$ (size $c \times c$) and $\mathbf{D}(\mathbf{p}_{i+})^{-1/2}$ (size $r \times r$), respectively:

$$\begin{aligned} \mathbf{V}_{(c \times c)} &= \mathbf{D}(\mathbf{p}_{+j})^{-1/2} \mathbf{U} \\ \hat{\mathbf{V}}_{(r \times c)} &= \mathbf{D}(\mathbf{p}_{i+})^{-1/2} \hat{\mathbf{U}} \end{aligned}$$

CA Computation - Step 3

Matrix \mathbf{F} , which gives the positions of the rows of the contingency table in the CA space, is obtained from the transformed matrix of eigenvectors \mathbf{V} , which gives the columns in that space. This is done by applying the usual equation for component scores to data matrix \mathbf{Q} , with division by the row weights:

$$\mathbf{F}(r \times c) = \hat{\mathbf{V}} \Lambda^{1/2} \quad \text{or} \quad \mathbf{F}(r \times c) = \mathbf{D}(\mathbf{p}_{i+})^{-1} \mathbf{Q} \mathbf{V}$$

In the same way, matrix $\hat{\mathbf{F}}$, which gives the columns of the contingency table in CA space is derived from the transformed matrix of eigenvectors $\hat{\mathbf{V}}$:

$$\hat{\mathbf{F}}(c \times c) = \mathbf{V} \Lambda^{1/2} \quad \text{or} \quad \hat{\mathbf{F}}(c \times c) = \mathbf{D}(\mathbf{p}_{+j})^{-1} \mathbf{Q}' \hat{\mathbf{V}}$$

CA Computation - Step 3

With this scaling, matrices \mathbf{F} and \mathbf{V} form a pair such that rows (given by matrix \mathbf{F}) are at the centroid of the columns in matrix \mathbf{V} .

In the same way, matrices $\hat{\mathbf{F}}$ and $\hat{\mathbf{V}}$ form a pair such that the columns (given by the matrix $\hat{\mathbf{F}}$) are at the centroids of the rows in matrix $\hat{\mathbf{V}}$.

Matrices \mathbf{F} and \mathbf{V} or $\hat{\mathbf{F}}$ and $\hat{\mathbf{V}}$ can be used to construct scatter diagrams and χ^2 distance (D_{16}) is preserved.

CA Example

Let's use an example other than the stand \times species situation we have been looking at (although we could do this here too) and consider the relative abundance (0, +, ++) of a particular species observed at 100 sites. The temperature at each site was recorded and coded (1,2,3):

Temp. (Descr.-1)	(Descr.-2) Sp. is:	Rare (0)	Abund. (+)	Very Abund. (++)	Row Sums
Cold (1)		10	10	20	40
Med. (2)		10	15	10	35
Warm (3)		15	5	5	25
Col. Sums		35	30	35	100

CA Example

Matrix \mathbf{Q} contains the proportions p_{ij} and the marginal totals p_{i+} and p_{+j} of the rows and columns, respectively. Identifiers of the rows and columns are given outside the matrix brackets in parentheses:

$$\mathbf{Q} = [p_{ij}] = \begin{matrix} & \begin{matrix} (0) & (+) & (++) \end{matrix} & \begin{matrix} [p_{i+}] \\ (1) \begin{bmatrix} 0.10 & 0.10 & 0.20 \\ (2) \begin{bmatrix} 0.10 & 0.15 & 0.10 \\ (3) \begin{bmatrix} 0.15 & 0.05 & 0.05 \end{bmatrix} \end{bmatrix} \end{matrix} \\ \begin{matrix} [p_{+j}] \\ [0.35 & 0.30 & 0.35] \end{matrix} \end{matrix}$$

CA Example

$$\bar{\mathbf{Q}} = [\bar{q}_{ij}] = \left[\frac{p_{ij} - p_{i+} p_{+j}}{\sqrt{p_{i+} p_{+j}}} \right] = \begin{bmatrix} -0.10690 & -0.05774 & 0.16036 \\ -0.06429 & 0.13887 & -0.06429 \\ 0.21129 & -0.09129 & -0.12677 \end{bmatrix}$$

The eigenvalues of $\bar{\mathbf{Q}}\bar{\mathbf{Q}}$ are: $\lambda_1 = 0.09613$ (70.1%) and
 $\lambda_2 = 0.04094$ (29.9%) and
 $\lambda_3 = 0$ (because of centering)

CA Example

The normalized eigenvectors of $\bar{\mathbf{Q}}\bar{\mathbf{Q}}$ are then:

$$\mathbf{U} = \begin{matrix} & (\lambda_1) & (\lambda_2) \\ \begin{matrix} (0) \\ (+) \\ (++) \end{matrix} & \begin{bmatrix} 0.78016 & -0.20336 \\ -0.20383 & 0.81145 \\ -0.59144 & -0.54790 \end{bmatrix} \end{matrix}$$

And the normalized eigenvectors of $\bar{\mathbf{Q}}\bar{\mathbf{Q}}$ are then:

$$\bar{\mathbf{U}} = \begin{matrix} & (\lambda_1) & (\lambda_2) \\ \begin{matrix} (0) \\ (+) \\ (++) \end{matrix} & \begin{bmatrix} -0.53693 & -0.55831 \\ -0.13043 & 0.79561 \\ 0.83349 & -0.23516 \end{bmatrix} \end{matrix}$$

CA Example

In Scaling Type-1, \mathbf{F} and \mathbf{V} are determined to produce a CA joint plot:

$$\mathbf{V} = \mathbf{D}(p_{+j})^{-1/2} \mathbf{U} = \begin{matrix} & (\lambda_1) & (\lambda_2) \\ \begin{matrix} (0) \\ (+) \\ (++) \end{matrix} & \begin{bmatrix} 1.31871 & -0.34374 \\ -0.37215 & 1.48150 \\ -0.99972 & -0.92612 \end{bmatrix} \end{matrix}$$

Now, to put the rows (matrix \mathbf{F}) at the centroids of the columns, the position of each row along an ordination axis is computed as the mean of the column positions, weighted by the relative frequencies of the observations in the various columns of that row...

CA Example

Consider the first row of the original data. The relative frequencies of that row are 0.25, 0.25, 0.50. Multiplying matrix **V** by that vector provides the coordinates of the first row of the ordination diagram:

$$\begin{bmatrix} 0.25 & 0.25 & 0.50 \end{bmatrix} \begin{bmatrix} 1.31871 & -0.34374 \\ -0.37215 & 1.48150 \\ -0.99972 & -0.92612 \end{bmatrix} = \begin{bmatrix} -0.26322 & -0.17862 \end{bmatrix}$$

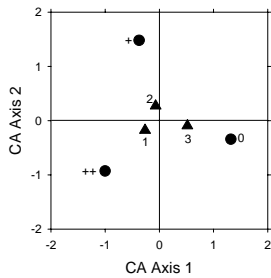
continuing:

$$\mathbf{F} = \mathbf{D}(p_{i+})^{-1} \mathbf{QV} = \begin{matrix} & (\lambda_1) & (\lambda_2) \\ \begin{matrix} (1) \\ (2) \\ (3) \end{matrix} & \begin{bmatrix} -0.26322 & -0.17862 \\ -0.06835 & 0.27211 \\ 0.51685 & -0.09517 \end{bmatrix} \end{matrix}$$



CA Example

Now, using **F** and **V**, we can construct the ordination plot:



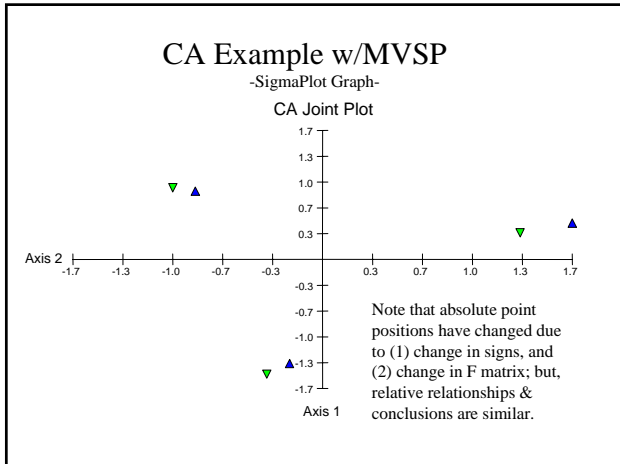
CA Example w/MVSP

	A	B	C	D
1				
2		10,000	10,000	20,000
3		10,000	15,000	10,000
4		15,000	5,000	5,000

CA Example w/MVSP

	Axis 1	Axis 2	Axis 3
Eigenvalues	0.096	0.041	0.000
Percentage	70.132	29.868	0.000
Cum. Percentage	70.132	100.000	100.000
CA variable scores	Axis 1	Axis 2	Axis 3
	1.319	0.344	1.000
	-0.372	-1.491	1.000
	-1.000	0.926	1.000
CA case scores	Axis 1	Axis 2	Axis 3
	-0.049	0.803	1.000
	0.220	-1.345	1.000
	1.667	0.470	1.000

Same Eigenvalues
 Same scores for V
 Different scores for F (uses diff. scaling algorithm)



Data Tables

Correspondence analysis has been applied to many types of data tables other than contingency tables.

However, as a caveat, recognize again that in order for CA to work correctly, the data table must be *dimensionally homogeneous* (i.e., in the same physical units) and *non-negative* (≥ 0).

If the data do not meet these assumptions, they may be transformed or recoded. This is a critical step in CA.

Arch Effect

Let's return to notion of coenocline distortion that we first considered in PCA. Recall that most of these procedures require *linear* (or at least monotonic) responses. Species data, in particular, is usually *unimodally* distributed across a gradient.

Recall that this problem usually manifests itself in the form of an *arch* or *horseshoe* in the data projection.

Some ecologists are willing to tolerate this distortion while others feel that an attempt should be made to recover the original gradient via *detrending*.

Arch Effect

The most extreme form of the arch effect usually occurs while attempting to apply a Euclidean distance measure to species abundance data. A *horseshoe* is formed because the ends actually contract and fold inwards at the ends of Axis-1 and bend along Axis-3. This is because ED considers the extreme sites to be very near each other.



In most instances, CA does not exhibit such a dramatic folding towards the terminal portions, but rather just bends along Axis-1 to form an *arch*.

Detrended Correspondence Analysis (DCA)

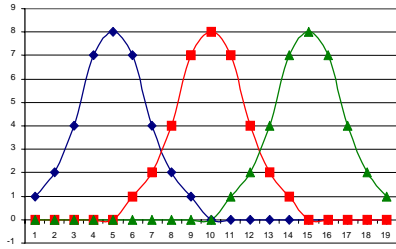
When a single axis is is enough to order the sites and species correctly, a second axis, which is independent of the first, can be obtained by folding the first axis in the middle and bringing the ends together.

Subsequent independent axes can be obtained by folding the first axis in three parts, four parts, etc.

This process is referred to as Detrended Correspondence Analysis (DCA).

Detrended Correspondence Analysis (DCA)

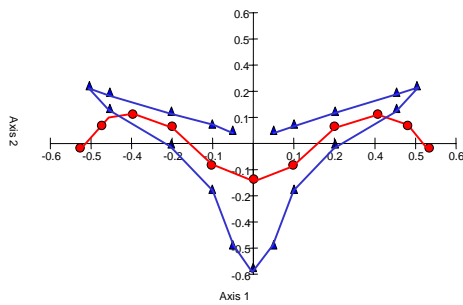
Recall this data set used to evaluate PCA & PCO where 3 species were unimodally distributed across a coenocline:



PCA w/Euclidean Distance

vs.

DCA via quadratic polynomial



Detrended Correspondence Analysis (DCA)

Two main approaches have been proposed to remove the arch effect:

detrending by polynomials (previous example), and
detrending by segments.

Both methods lead to detrended correspondence analysis.

Detrending by Segments

When detrending by segments (Hill and Gauch 1980), axis-I is divided in to a number of "segments" and, within each one, the mean of the scores along axis II is made equal to zero; in other words, data points in each segment are moved along axis-II to make their mean coincide with the abscissa.

Proximities among points should *in no case* be interpreted as meaningful! Segments can generate large differences in scores for points that are near each other in the original ordination but happen to be on either side of a segment division.

The number of segments is *arbitrary*. Different numbers of segments lead to different ordinations.

Detrending by Segments

Various software packages use 10 as a minimum number of segments and ± 46 as a maximum; 26 being a recommended starting or default value. This of course necessitates data sets with considerably more observations than 26. There are no empirical rules for the "correct" number of detrending segments.

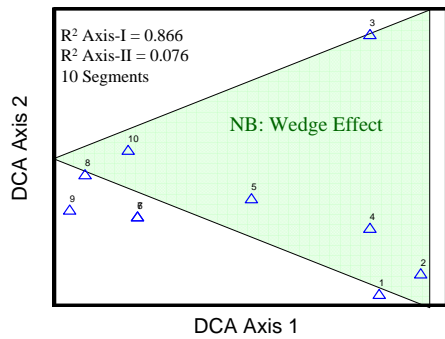
After detrending by segments, the DCA ordination has the interesting property that the axes are scaled in units of the average standard deviation (SD) of species turnover (Gauch 1982). Along a regular gradient, a species typically appears, rises to a modal value, and disappears in 4 SD; similarly, a complete turnover in species composition often occurs over 4 SD. Thus, the length of axis-I is often used as a measure of the length of the ecological gradient.

Detrending by Polynomials

Detrending by *polynomials* (Hill and Gauch 1980) directly follows from the fact that an arch is produced when a gradient of sufficient length is present in the data. When a sufficient number of species are present and replace each other along the gradient, the second axis of the CA approaches a *quadratic function* of the first one (i.e., a second degree polynomial).

When detrending is sought, detrending by polynomials is an attractive method because it results in a *continuous function* of the previous axes, without the discontinuities generated by detrending-by-segments. However, the downside of this method is that it imposes a very specific polynomial *model* that the data must correspond to. It also does not solve terminal gradient compressions at the ends of the ordination axes.

DCA of WI Forest Data



To Detrend or Not To Detrend That is the Question...



The controversy over detrending has raged in the literature for well over a decade now.

Wartenburg et al. (1987) argue that the arch is an important and inherent attribute of the distances among sites, not simply a mathematical artifact. The only effect of DCA is to flatten the distribution of points onto axis-I. They also argue that detrending by segments is completely arbitrary and has no theoretical justification.

Peet et al. (1988) still support DCA on the grounds that detrending and rescaling may facilitate ecological interpretation & called for improved algorithms.

To Detrend or Not To Detrend That is the Question...

Minchin (1987) produced a nice comparison of several ordination techniques and found DCA to perform poorly on most accounts.

He found that DCA actually removed real pattern from the data and produced significant distortion which he referred to as a "tongue" or subsequently a "wedge" in the data and this was a simple artifact of the algorithm.

While DCA may have some applications, my inclination is to avoid its use at this point. NMDS seems to produce superior results with community data.

Factor Analysis

In the social sciences, analysis of the relationships among the descriptors of a multidimensional data matrix is frequently carried out via Factor Analysis (FA).

Recall that the goal of PCA is to account for a maximum amount of the *variance* in the data, whereas the goal of factor analysis is to account for the *covariance* among descriptors.

Put another way, PCA is directed towards reducing the diagonal elements of \mathbf{R} . Factor analysis is directed more towards reducing the off-diagonal elements of \mathbf{R} . Since reducing the diagonal elements reduces the off-diagonal elements and vice versa, both methods achieve much the same thing.

Factor Analysis

To do this, FA assumes that the observed descriptors are linear combinations of hypothetical underlying variables (i.e., the *factors*).

Originally FA was used to evaluate such things as intelligence. Many variables could be measured such as age, parental education, family income, etc. Multiple variables might play out to show that Factor-1 was determined by all variables related to education and Factor-2 to socio-economic conditions (for example).

There are few applications of FA in EEB, so I will not cover it in depth. An excellent treatment can be found in Tabachnick and Fidell (1996).

The text "The End!" is rendered in a 3D, blocky font with a green, marbled texture. The letters are slightly offset and have a perspective view, giving them a three-dimensional appearance.
