Correspondence Analysis

In the first part of the course, Principal Component Analysis was seen to be a useful tool to uncover the relationships among continuous variables.

In the next few segments, attention will focus on a similar technique called correspondence analysis (CA) which was designed to explore categorical variables.

Typical data for which CA can be used include

- ✓ contingency tables;
- ✓ answers to multiple choice questionnaires.

Origins

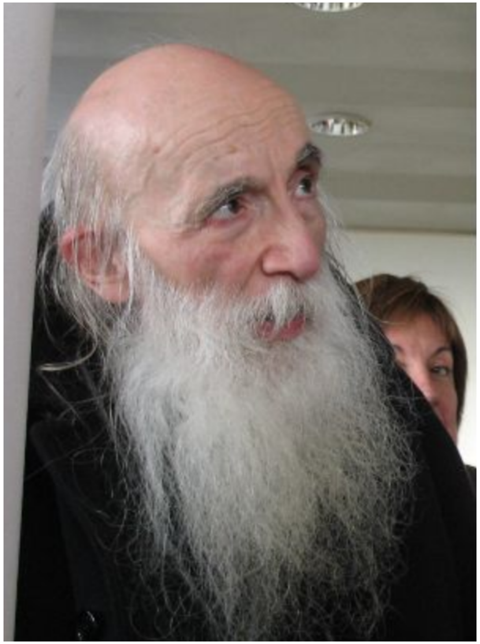
Correspondence analysis was initially proposed by Herman Otto Hirschfeld (*Proc. Camb. Phil. Soc.*, 1935) and later developed by Jean-Paul Benzécri (1973).

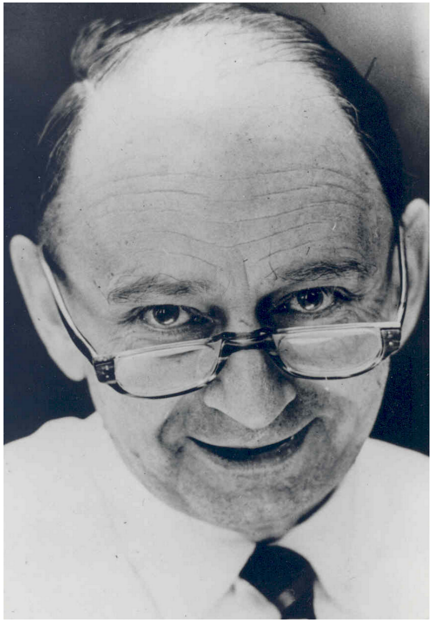
In a similar manner to PCA, it provides a means of displaying or summarizing a set of data in two-dimensional graphical form.

Two special cases of this data analytical technique will be considered:

- ✓ factorial correspondence analysis, which is used for two-way contingency tables;
- ✓ multiple correspondence analysis for multi-way tables.

The Pioneers





Jean-Paul Benzécri (1932–2019) Herman Otto Hirschfeld (1912–1980)

Motivating Example

Consider a hypothetical study exploring the relationship between country of residence and primary language spoken (Sourial et al., 2010).

Data

	Language					
Country	English	French	Spanish	German	Italian	Total
Canada	688	280	10	11	11	1000
USA	730	31	190	8	41	1000
England	798	74	38	31	59	1000
Italy	17	13	11	15	944	1000
Switzerland	15	222	20	648	95	1000
Total	2248	620	269	713	1150	5000

Source

> J Clin Epidemiol. 2010 Jun;63(6):638-46. doi: 10.1016/j.jclinepi.2009.08.008. Epub 2009 Nov 6.

Correspondence analysis is a useful tool to uncover the relationships among categorical variables

Nadia Sourial [1], Christina Wolfson, Bin Zhu, Jacqueline Quail, John Fletcher, Sathya Karunananthan, Karen Bandeen-Roche, François Béland, Howard Bergman

Affiliations + expand

PMID: 19896800 PMCID: PMC3718710 DOI: 10.1016/j.jclinepi.2009.08.008

Free PMC article

Abstract

Objective: Correspondence analysis (CA) is a multivariate graphical technique designed to explore the relationships among categorical variables. Epidemiologists frequently collect data on multiple categorical variables with the goal of examining associations among these variables. Nevertheless, CA appears to be an underused technique in epidemiology. The objective of this article is to present the utility of CA in an epidemiological context.

Study design and setting: The theory and interpretation of CA in the case of two and more than two variables are illustrated through two examples.

Test of Independence (1–5)

Let

$$\mathbf{K} = (k_{ij}),$$

be an $r \times c$ matrix with entries

$$k_{ij} = \text{number of observations in class } i \in \{1, \ldots, r\} \\ \text{and category } j \in \{1, \ldots, c\}.$$

One can deduce from it the table of relative frequencies, viz.

$$\mathbf{F} = (f_{ij}),$$

where

$$f_{ij} = k_{ij}/k_{\bullet\bullet} = k_{ij} \Big/ \sum_{\ell=1}^{r} \sum_{m=1}^{c} k_{\ell m}.$$

Relative frequencies

	Language						
Country	English	French	Spanish	German	Italian	Margin	
Canada	.1376	.0560	.0020	.0022	.0022	.0200	$f_{1\bullet}$
USA	.1460	.0062	.0380	.0016	.0082	.0200	$f_{2\bullet}$
England	.1596	.0148	.0076	.0062	.0118	.0200	
Italy	.0034	.0026	.0022	.0030	.1888	.0200	
Switzerland	.0030	.0444	.0040	.1296	.0190	.0200	
Margin	.4496	.1240	.0538	.1426	.2300	1.0000	
						$k_{\bullet\bullet}$	

\# rows   \# cols

The margins are given, for each $i \in \{1, \ldots, r\}$ and $j \in \{1, \ldots, c\}$, by

$$f_{i\bullet} = \sum_{j=1}^{c} f_{ij} = k_{i\bullet}/k_{\bullet\bullet}\,, \quad f_{\bullet j} = \sum_{i=1}^{r} f_{ij} = k_{\bullet j}/k_{\bullet\bullet}.$$

Test of Independence (3–5)

For each $i \in \{1, \dots, r\}$ and $j \in \{1, \dots, c\}$, f_{ij} is an estimate of

$$\Pr(X = i, Y = j),$$

where X is the country of residence and Y is the language spoken.

The variables X and Y are said to be stochastically independent if and only if the following hypothesis holds.

Hypothesis of independence

For all $i \in \{1, \dots, r\}$ and $j \in \{1, \dots, c\}$, one has

$$\Pr(X = i, Y = j) = \Pr(X = i) \Pr(Y = j). \qquad (\mathcal{H}_0)$$

Test of Independence (4–5)

To test this hypothesis, one can use the chi-squared test, which measures the discrepancy between

- ✓ the observed counts $k_{ij} = k_{\bullet\bullet} \times f_{ij}$,
- ✓ the expected counts under $\mathcal{H}_0$, which are given by

$$\hat{f}_{ij} = k_{\bullet\bullet} \times f_{i\bullet} \times f_{\bullet j} = k_{i\bullet} \times k_{\bullet j} / k_{\bullet\bullet}.$$

The test statistic is

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(k_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(k_{ij} - k_{i\bullet} k_{\bullet j} / k_{\bullet\bullet}\right)^2}{k_{i\bullet} k_{\bullet j} / k_{\bullet\bullet}}.$$

Test of Independence (5–5)

Under the null hypothesis of independence, one has (asymptotically)

$$X^2 \approx \chi^2_\ell, \qquad Pr(X^2 \le t) \underset{\infty}{\to} Pr(\chi^2_\ell \le t)$$

where $\ell = (r-1)(c-1)$ is the number of degrees of freedom.

For the data at hand, one gets

$$X^2 = k_{\bullet\bullet} \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij} - f_{i\bullet} f_{\bullet j})^2}{f_{i\bullet} f_{\bullet j}} \approx 7338.9$$

and the p-value of the test is very small ($p < 2.2 \times 10^{-16}$).

R Code

```
> M <- as.table(rbind(c(688,280,10,11,11), c(730,31,190,8,41),
     c(798,74,38,31,59), c(17,13,11,15,944), c(15,222,20,648,95)))
> M

    A   B   C   D   E
A 688 280  10  11  11
B 730  31 190   8  41
C 798  74  38  31  59
D  17  13  11  15 944
E  15 222  20 648  95

> Xsq <- chisq.test(M)
> Xsq

Pearson's Chi-squared test

data:  M
X-squared = 7338.9, df = 16, p-value < 2.2e-16
```

Additional Information Provided by R

```
> Xsq$observed   # observed counts (same as M)
> Xsq$expected   # expected counts under the null
> Xsq$residuals  # Pearson residuals
> Xsq$stdres     # standardized residuals

> Xsq$expected   # expected counts under the null
      A   B    C     D   E
A 449.6 124 53.8 142.6 230
B 449.6 124 53.8 142.6 230
C 449.6 124 53.8 142.6 230
D 449.6 124 53.8 142.6 230
E 449.6 124 53.8 142.6 230
```

$\hat{f}_{ij} = k_{\bullet\bullet} \hat{f}_i \hat{f}_j$

NOT integers

Profiles

Given the existence of a significant relationship between X and Y, correspondence analysis can help us to visualize it.

The graphical display will be based on the row and column profiles of the data matrix.

Row and column profiles

For each $i \in \{1, \ldots, r\}$, let

$$R_i = \left(\frac{k_{i1}}{k_{i\bullet}}, \ldots, \frac{k_{ic}}{k_{i\bullet}} \right) = \left(\frac{f_{i1}}{f_{i\bullet}}, \ldots, \frac{f_{ic}}{f_{i\bullet}} \right),$$

and for each $j \in \{1, \ldots, c\}$,

$$C_j = \left(\frac{k_{1j}}{k_{\bullet j}}, \ldots, \frac{k_{rj}}{k_{\bullet j}} \right) = \left(\frac{f_{1j}}{f_{\bullet j}}, \ldots, \frac{f_{rj}}{f_{\bullet j}} \right).$$

Row and Column Profiles

Row profiles

Country	Language English	French	Spanish	German	Italian	Margin
Canada	.688	.280	.010	.011	.011	1.000
USA	.730	.031	.190	.008	.041	1.000
England	.798	.074	.038	.031	.059	1.000
Italy	.017	.013	.011	.015	.944	1.000
Switzerland	.015	.222	.020	.648	.095	1.000

Column profiles

Country	Language English	French	Spanish	German	Italian
Canada	.306	.452	.037	.015	.009
USA	.325	.050	.706	.011	.036
England	.355	.119	.141	.044	.051
Italy	.007	.021	.041	.021	.821
Switzerland	.007	.358	.075	.909	.083
Margin	1.000	1.000	1.000	1.000	1.000

Average Profiles

The average row profile is given by

$$\left(\sum_{i=1}^{r} f_{i\bullet} \, \frac{f_{i1}}{f_{i\bullet}}, \ldots, \sum_{i=1}^{r} f_{i\bullet} \, \frac{f_{ic}}{f_{i\bullet}} \right) = (f_{\bullet 1}, \ldots, f_{\bullet c}),$$

i.e., $(0.2, 0.2, 0.2, 0.2, 0.2)$ in the example.

Similarly, the average column profile is given by

$$\left(\sum_{j=1}^{c} f_{\bullet j} \, \frac{f_{1j}}{f_{\bullet j}}, \ldots, \sum_{j=1}^{c} f_{\bullet j} \, \frac{f_{rj}}{f_{\bullet j}} \right) = (f_{1\bullet}, \ldots, f_{r\bullet}),$$

i.e., $(0.4496, 0.1240, 0.0538, 0.1426, 0.2300)$ in the example.

Connection with Independence

In the ideal case of independence where

$$\hat{f}_{ij} = f_{ij} \quad \Leftrightarrow \quad f_{ij} = f_{i\bullet} f_{\bullet j}$$

for all $i \in \{1, \ldots, r\}$ and $j \in \{1, \ldots, c\}$, one would have

$$\left(\frac{f_{i1}}{f_{i\bullet}}, \ldots, \frac{f_{ic}}{f_{i\bullet}} \right) = (f_{\bullet 1}, \ldots, f_{\bullet c})$$

and

$$\left(\frac{f_{1j}}{f_{\bullet j}}, \ldots, \frac{f_{rj}}{f_{\bullet j}} \right) = (f_{1\bullet}, \ldots, f_{r\bullet})$$

for all $i \in \{1, \ldots, r\}$ and $j \in \{1, \ldots, c\}$.

Reformulation of the Concept of Dependence

From the previous discussion, one can see that

the dependence between the two variables depends on the resemblance between the row profiles and the column profiles.

The discrepancy between two row profiles could be computed as

$$d^2(i,i') = \sum_{j=1}^{c} \left(\frac{f_{ij}}{f_{i\bullet}} - \frac{f_{i'j}}{f_{i'\bullet}} \right)^2,$$

but this would fail to take into account the relative importance of the various columns.

Chi-Squared Distance

In contrast, the chi-squared distance between the row profiles, viz.

$$d^2(i,i') = \sum_{j=1}^{c} \frac{1}{f_{\bullet j}} \left(\frac{f_{ij}}{f_{i\bullet}} - \frac{f_{i'j}}{f_{i'\bullet}} \right)^2 ,$$

which does account for the relative importance of the various columns.

In matrix terms, let

$$\underbrace{\mathbf{D}_r}_{r \times r} = \mathrm{diag}(f_{i\bullet}) = \begin{pmatrix} f_{1\cdot} & & 0 \\ & f_{2\cdot} & \\ 0 & & \ddots \\ & & & f_{r\cdot} \end{pmatrix} \quad \text{and} \quad \underbrace{\mathbf{D}_c}_{c \times c} = \mathrm{diag}(f_{\bullet j}).$$

Then

- ✓ the row profiles are the rows of $\mathbf{D}_r^{-1} \underbrace{\mathbf{F}}_{r \times c}$, frequency table
- ✓ the column profiles are the rows of $\mathbf{D}_c^{-1} \mathbf{F}^\top$.

Chi-Squared Formula

In matrix terms, the chi-squared distance is given by

$$\mathbf{x}^\top \mathbf{D}_c^{-1} \mathbf{x} = \sum_{j=1}^{c} x_j^2 / f_{\bullet j}$$

I_c for Euclidean distance

for any column vector $\mathbf{x} \in \mathbb{R}^c$ of the form

$$\mathbf{x}^\top \mathbf{D}_r^{-1} \mathbf{x} = \sum_{i=1}^{r} x_i^2 / f_{i \bullet}$$

for any row vector $\mathbf{x} \in \mathbb{R}^r$.

Distributional Equivalence

The chi-squared distance has two important advantages, viz.

1. The (squared) distance between any two rows is unchanged by merging two columns sharing the same profile.

2. The (squared) distance between any two columns is unchanged by merging two rows sharing the same profile.

To illustrate this point, a synthetic numerical example is given next.

Numerical Example (1–2)

Consider the following hypothetical 2×3 table:

11	22	16	49
16	32	3	51
27	54	19	100

whose first two columns have exactly the same profile.

The (squared) distance between the first two rows is given by

$$d^2(1,2) = \frac{100}{27}\left(\frac{11}{49} - \frac{16}{51}\right)^2 + \frac{100}{54}\left(\frac{22}{49} - \frac{32}{51}\right)^2 + \frac{100}{19}\left(\frac{16}{49} - \frac{3}{51}\right)^2,$$

i.e., $d^2(1,2) = 0.088\,477\,88$.

Numerical Example (1–2)

Now suppose that the first two columns are merged, viz.

$$\begin{array}{ccc|c} 11 & 22 & 16 & 49 \\ 16 & 32 & 3 & 51 \\ \hline 27 & 54 & 19 & 100 \end{array} \quad \begin{array}{c} \Rightarrow \\ \Rightarrow \\ \Rightarrow \end{array} \quad \begin{array}{cc|c} 33 & 16 & 49 \\ 48 & 3 & 51 \\ \hline 81 & 19 & 100 \end{array}$$

The (squared) distance between the first two rows is then

$$d^2(1,2) = \frac{100}{81}\left(\frac{33}{49} - \frac{48}{51}\right)^2 + \frac{100}{19}\left(\frac{16}{49} - \frac{3}{51}\right)^2,$$

i.e., $d^2(1,2) = 0.088\,477\,88$, as before.

Proof in the Case of Columns (1–2)

Assume that there exist two columns $j_1, j_2 \in \{1, \ldots, c\}$ such that, for all $i \in \{1, \ldots, r\}$, one has

$$\frac{f_{ij_1}}{f_{\bullet j_1}} = \frac{f_{ij_2}}{f_{\bullet j_2}} = F_i.$$

Writing $f_{\bullet j_0} = f_{\bullet j_1} + f_{\bullet j_2}$, one then has, for all $i \in \{1, \ldots, r\}$,

$$\frac{f_{ij_0}}{f_{\bullet j_0}} = \frac{f_{ij_1} + f_{ij_2}}{f_{\bullet j_0}} = \frac{F_i(f_{\bullet j_1} + f_{\bullet j_2})}{f_{\bullet j_0}} = F_i.$$

Proof in the Case of Columns (1–2)

It follows that

$$
\begin{aligned}
& d^2(i,i') - d^2(i,i') \\
& \quad = \frac{1}{f_{\bullet j_1}} \left(\frac{f_{ij_1}}{f_{i\bullet}} - \frac{f_{i'j_1}}{f_{i'\bullet}} \right)^2 + \frac{1}{f_{\bullet j_2}} \left(\frac{f_{ij_2}}{f_{i\bullet}} - \frac{f_{i'j_2}}{f_{i'\bullet}} \right)^2 - \frac{1}{f_{\bullet j_0}} \left(\frac{f_{ij_0}}{f_{i\bullet}} - \frac{f_{i'j_0}}{f_{i'\bullet}} \right)^2 .
\end{aligned}
$$

and hence

$$
\Delta = f_{\bullet j_1} \left(\frac{F_i}{f_{i\bullet}} - \frac{F_{i'}}{f_{i'\bullet}} \right)^2 + f_{\bullet j_2} \left(\frac{F_i}{f_{i\bullet}} - \frac{F_{i'}}{f_{i'\bullet}} \right)^2 - f_{\bullet j_0} \left(\frac{F_i}{f_{i\bullet}} - \frac{F_{i'}}{f_{i'\bullet}} \right)^2 ,
$$

i.e., $\Delta = 0$.

Correspondence Analysis

The objective of correspondence analysis is to represent large two-way contingency tables in two dimensions.

Factorial correspondence analysis achieves this goal by representing simultaneously the row profiles in $\mathbb{R}^c$ and the column profiles in $\mathbb{R}^r$.

This representation in the Cartesian plane is obtained by performing a double principal component analysis.

Useful Result

Let $\mathbf{A}$ and $\mathbf{M}$ be two symmetric matrices.

Consider the problem of finding a vector $\mathbf{u}$ such that

$$\mathbf{u}^\top \mathbf{A}\mathbf{u} \text{ is maximized,}$$

given that $\mathbf{u}^\top \mathbf{M}\mathbf{u} = 1$.

The solution is a normed eigenvector $\mathbf{u}$ of $\mathbf{M}^{-1}\mathbf{A}$ corresponding to

$$\lambda = \text{largest eigenvalue of } \mathbf{M}^{-1}\mathbf{A}.$$

It then follows that

$$\mathbf{u}^\top \mathbf{A}\mathbf{u} = \mathbf{u}^\top \lambda \mathbf{M}\mathbf{u} = \lambda(\mathbf{u}^\top \mathbf{M}\mathbf{u}) = \lambda.$$

Proof

Given the constraint and a Lagrange multiplier λ, one must maximize

$$\mathbf{u}^\top \mathbf{A}\mathbf{u} - \lambda(\mathbf{u}^\top \mathbf{M}\mathbf{u} - 1).$$

Differentiate with respect to $\mathbf{u}$ and set equal to $\mathbf{0}$ to find

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{M}\mathbf{u}.$$

Multiply both sides on the left by $\mathbf{M}^{-1}$ to get

$$\mathbf{M}^{-1}\mathbf{A}\mathbf{u} = \lambda \mathbf{u}.$$

Thus λ is an eigenvalue of $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{u}$ is a corresponding eigenvector. Further note that taking λ to be the largest eigenvalue will maximize

$$\mathbf{u}^\top \mathbf{A}\mathbf{u} = \lambda \mathbf{u}^\top \mathbf{M}\mathbf{u} = \lambda.$$

Difference With PCA

Direct Analysis

The rows of $\mathbf{D}_r^{-1}\mathbf{F}$ can be regarded as elements of $\mathbb{R}^c$.

They are represented in this space endowed with the distance function

$$\mathbf{x}^\top \mathbf{D}_c^{-1} \mathbf{x}.$$

Dual Analysis

The rows of $\mathbf{D}_c^{-1}\mathbf{F}^\top$ can be regarded as elements of $\mathbb{R}^r$.

They are represented in this space endowed with the distance function

$$\mathbf{x}^\top \mathbf{D}_r^{-1} \mathbf{x}.$$

Link With Principal Component Analysis

	PCA	FCA
Data	$\mathbf{X}$	$\mathbf{D}_r^{-1}\mathbf{F}$
Weight	$\mathbf{I}$	$\mathbf{D}_r^{-1}$
Distances	$\mathbf{I}$	$\mathbf{D}_c^{-1}$
Projections	$\mathbf{Xu}$	$(\mathbf{D}_r^{-1}\mathbf{F})\ \mathbf{D}_c^{-1}\mathbf{u}$
To maximize	$\mathbf{u}^\top\mathbf{X}^\top\mathbf{Xu}$	$\mathbf{u}^\top\{(\mathbf{D}_r^{-1}\mathbf{F})\mathbf{D}_c^{-1}\}^\top\ \{(\mathbf{D}_r^{-1}\mathbf{F})\mathbf{D}_c^{-1}\}\mathbf{u}$
Constraint	$\mathbf{u}^\top\mathbf{u} = 1$	$\mathbf{u}^\top\mathbf{D}_c^{-1}\mathbf{u} = 1$

First Factorial Axis of the Direct Analysis

We are looking for the vector $\mathbf{u} \in \mathbb{R}^c$ such that

$$(\mathbf{u}^\top \mathbf{D}_c^{-1} \mathbf{F}^\top \mathbf{D}_r^{-1}) \mathbf{D}_r (\mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1} \mathbf{u})$$

be maximized, subject to $\mathbf{u}^\top \mathbf{D}_c^{-1} \mathbf{u} = 1$.

The solution is given by an eigenvector corresponding to the largest eigenvalue of

$$\begin{aligned} \mathbf{D}_c (\mathbf{D}_c^{-1} \mathbf{F}^\top \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1}) &= (\mathbf{D}_c \mathbf{D}_c^{-1}) \mathbf{F}^\top \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1} \\ &= \mathbf{F}^\top \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1} \equiv \mathbf{S} \end{aligned}$$

First Factorial Axis of the Dual Analysis

We are looking for the vector $\mathbf{v} \in \mathbb{R}^r$ such that

$$(\mathbf{v}^\top \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1}) \mathbf{D}_c (\mathbf{D}_c^{-1} \mathbf{F}^\top \mathbf{D}_r^{-1} \mathbf{v})$$

be maximized, subject to $\mathbf{v}^\top \mathbf{D}_r^{-1} \mathbf{v} = 1$.

The solution is given by an eigenvector corresponding to the largest eigenvalue of

$$\mathbf{D}_r (\mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1} \mathbf{F}^\top \mathbf{D}_r^{-1}) = \mathbf{F} \mathbf{D}_c^{-1} \mathbf{F}^\top \mathbf{D}_r^{-1} \equiv \mathbf{T}.$$

Remarks

✓ The matrix $\mathbf{S} = (s_{ij})$ is $c \times c$.

✓ The matrix $\mathbf{T} = (t_{k\ell})$ is $r \times r$.

✓ Neither of them is symmetric, e.g., for all $j, j' \in \{1, \ldots, c\}$, one has

$$s_{jj'} = \sum_{i=1}^{r} \frac{f_{ij} f_{ij'}}{f_{i\bullet} f_{\bullet j'}}.$$

Crucial Observation

$$\boxed{\text{We assume henceforth that } c \le r.}$$

When such is the case, the first c eigenvalues of **S** and **T** coincide.

Indeed, if

$$\mathbf{S}\mathbf{u} = \mathbf{F}^\top \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1} \mathbf{u} = \lambda \mathbf{u},$$

then

$$\mathbf{F}\mathbf{D}_c^{-1}\mathbf{S}\mathbf{u} = \mathbf{F}\mathbf{D}_c^{-1}\underbrace{\mathbf{F}^\top \mathbf{D}_r^{-1}(\mathbf{F}\mathbf{D}_c^{-1}\mathbf{u}}_{\lambda u}) = \lambda(\underbrace{\mathbf{F}\mathbf{D}_c^{-1}\mathbf{u}}_{v}).$$

Therefore, $\mathbf{v} = \mathbf{F}\mathbf{D}_c^{-1}\mathbf{u}$ is an eigenvector of $\mathbf{T} = \mathbf{F}\mathbf{D}_c^{-1}\mathbf{F}^\top \mathbf{D}_r^{-1}$.

Link Between the Eigenvectors

If $\mathbf{u}$ is associated to λ and if

$$\mathbf{u}^\top \mathbf{D}_c^{-1}\mathbf{u} = 1,$$

then $\mathbf{v} = \mathbf{F}\mathbf{D}_c^{-1}\mathbf{u}$ is an eigenvector of $\mathbf{T}$ and

$$\begin{aligned}
\mathbf{v}^\top \mathbf{D}_r^{-1}\mathbf{v} &= \mathbf{u}^\top \mathbf{D}_c^{-1}\mathbf{F}^\top \mathbf{D}_r^{-1}\mathbf{F}\mathbf{D}_c^{-1}\mathbf{u} \\
&= \mathbf{u}^\top \mathbf{D}_c^{-1}(\mathbf{F}^\top \mathbf{D}_r^{-1}\mathbf{F}\mathbf{D}_c^{-1}\mathbf{u}) \\
&= \lambda \mathbf{u}^\top \mathbf{D}_c^{-1}\mathbf{u} \\
&= \lambda.
\end{aligned}$$

Consequence

One must therefore take, for all $j \in \{1, \ldots, c\}$,

$$\mathbf{v}_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{F}\mathbf{D}_c^{-1}\mathbf{u}_j$$

to have a vector of norm 1.

Similarly, one has, for all $j \in \{1, \ldots, c\}$,

$$\mathbf{u}_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{F}^\top \mathbf{D}_r^{-1}\mathbf{v}_j.$$

Definition of the jth Factors

The coordinates of the jth factor of the direct analysis are given, for all $j \in \{1, \ldots, c\}$, by

$$\varphi_j = \mathbf{D}_c^{-1} u_j.$$

The projections of the row profiles on the jth eigenvector u_j are the components of the vector $\mathbf{D}_r^{-1} \mathbf{F} \varphi_j$.

The coordinates of the jth factor of the dual analysis are given, for all $j \in \{1, \ldots, c\}$, by

$$\Psi_j = \mathbf{D}_r^{-1} v_j.$$

The projections of the column profiles on the jth eigenvector v_j are the components of the vector

$$\mathbf{D}_c^{-1} \mathbf{F}^\top \mathbf{D}_r^{-1} v_j = \mathbf{D}_c^{-1} \mathbf{F}^\top \Psi_j.$$

Transition Relations

The following equations establish a link between the two types of analysis:

$$\mathbf{D}_r^{-1}\mathbf{F}\varphi_j = \sqrt{\lambda_j}\Psi_j \equiv \hat{\Psi}_j, \quad \mathbf{D}_c^{-1}\mathbf{F}^\top\Psi_j = \sqrt{\lambda_j}\varphi_j \equiv \hat{\varphi}_j.$$

These relations imply that the coordinates of the points in one space are proportional to the components of the factor in the other space corresponding to the same eigenvalue. Indeed, one has

$$\mathbf{D}_r^{-1}\mathbf{F}\varphi_j = \mathbf{D}_r^{-1}(\mathbf{F}\mathbf{D}_c^{-1}u_j) = \mathbf{D}_r^{-1}(\sqrt{\lambda_j}v_j) = \sqrt{\lambda_j}\Psi_j.$$

The second identity can be shown in a similar way.

Final Remarks

It follows from the previous developments that

$$1 = \lambda_1 \geq \cdots \geq \lambda_c \geq 0.$$

One can also check that

$$(f_{\bullet 1}, \ldots, f_{\bullet c})^\top$$

is an eigenvector corresponding to the eigenvalue $\lambda_1 = 1$ of

$$\mathbf{S} = \mathbf{F}^\top \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1}.$$

Given that the largest eigenvalue is always equal to 1, many statistical software give only the $c - 1$ others.

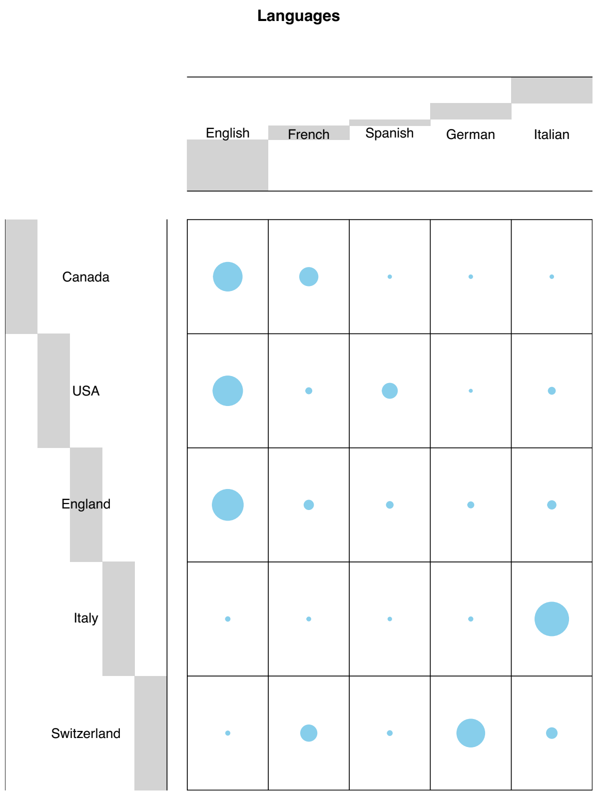
Data

Country	English	French	Language Spanish	German	Italian	Total
Canada	688	280	10	11	11	1000
USA	730	31	190	8	41	1000
England	798	74	38	31	59	1000
Italy	17	13	11	15	944	1000
Switzerland	15	222	20	648	95	1000
Total	2248	620	269	713	1150	5000

Balloon Plot (1–2)

```
head(M)
            English French Spanish German Italian
Canada          688    280      10     11      11
USA             730     31     190      8      41
England         798     74      38     31      59
Italy            17     13      11     15     944
Switzerland      15    222      20    648      95

dt <- as.table(as.matrix(M))

ballonplot(t(dt), main ="Languages", xlab ="", ylab="",
+             label = FALSE, show.margins = FALSE)
```

Balloon Plot (2–2)



Correspondence Analysis (1–4)

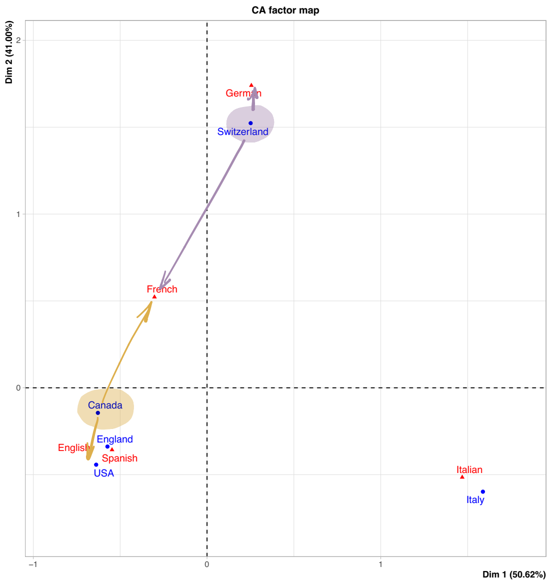
```
Languages<-CA(M, ncp = 2, graph = TRUE)

**Results of the Correspondence Analysis (CA)**
The row variable has  5  categories; the column variable has 5 categories
The chi square of independence between the two variables is equal
                                                  to 7338.936 (p-value =  0).
*The results are available in the following objects:

   name                 description
1  "$eig"               "eigenvalues"
2  "$col"               "results for the columns"
3  "$col$coord"         "coord. for the columns"
4  "$col$cos2"          "cos2 for the columns"
5  "$col$contrib"       "contributions of the columns"
6  "$row"               "results for the rows"
7  "$row$coord"         "coord. for the rows"
8  "$row$cos2"          "cos2 for the rows"
9  "$row$contrib"       "contributions of the rows"
10 "$call"              "summary called parameters"
11 "$call$marge.col" "weights of the columns"
12 "$call$marge.row" "weights of the rows"
```

Correspondence Analysis (2–4)

```
Languages$eig
      eigenvalue percentage of variance cumulative percentage of variance
dim 1 0.74303875              50.623059                          50.62306
dim 2 0.60177457              40.998764                          91.62182
dim 3 0.10392680               7.080509                          98.70233
dim 4 0.01904699               1.297667                         100.00000
```

- ✓ Note that the first eigenvalue (which is equal to 1) is not listed.
- ✓ The first two dimensions explain 91.6% of the variability in the table.

Correspondence Analysis (4–4)



- ✓ Points that are close to one another on the graph are associated.
- ✓ Italian is associated to Italy, Switzerland is between German and French.
- ✓ Canada is a mix of French and English, etc.

Complementary Information

In this segment, we will continue to investigate the nature and meaning of various elements in the standard R output for correspondence analysis.

In particular, technical details will be provided to address the following questions:

- ✓ Why do points that are associated appear close to one another?
- ✓ What does inertia stand for?
- ✓ How are the percentages associated to the axes computed?

Simple Elements Available in the Output (1–2)

Average row profile

```
Languages$call$marge.row

     Canada          USA      England        Italy Switzerland
        0.2          0.2          0.2          0.2         0.2
```

Average column profile

```
Languages$call$marge.col

English  French Spanish  German Italian
 0.4496  0.1240  0.0538  0.1426  0.2300
```

Simple Elements Available in the Output (2–2)

Coordinates of the column points (i.e., the languages):

```
Languages$col

$coord
             Dim 1      Dim 2
English -0.6835356 -0.3897308
French  -0.3024582  0.5223006
Spanish -0.5468139 -0.3570406
German   0.2552338  1.7400844
Italian  1.4688898 -0.5150849
```

The analog information is available for the row points (i.e., the countries).

Gravity Center (1–7)

R and other software produce graphs that are centered at $(0, 0)$.

This is because the analysis is carried out relative to the centers of gravity of the rows and columns.

avg row / col

This is not only common but eminently practical.

Gravity Center (2–7)

The mass associated with the ith row is $f_{i\bullet}$, i.e., the proportion of observations which belong to this row.

Similarly, the mass associated with the jth column is $f_{\bullet j}$.

The center of gravity of the rows is the mean of the row profiles, but weighted by the mass of each row.

Likewise, the center of gravity of the columns is the mean of the column profiles, but weighted by the mass of each column.

The center of gravity of the rows is defined by

$$G_R = (g_1, \ldots, g_c)^\top,$$

where for each $j \in \{1, \ldots, c\}$,

$$g_j = \sum_{i=1}^{r} f_{i\bullet} \frac{f_{ij}}{f_{i\bullet}} = \sum_{i=1}^{r} f_{ij} = f_{\bullet j} \,.$$

Similarly, the center of gravity of the columns is defined by

$$G_C = (f_{1\bullet}, \ldots, f_{r\bullet})^\top.$$

Centering the rows

Rows can be centered by computing

$$\frac{f_{ij}}{f_{i\bullet}} - g_j = \frac{f_{ij}}{f_{i\bullet}} - f_{\bullet j} = \frac{f_{ij} - f_{i\bullet} f_{\bullet j}}{f_{i\bullet}}$$

$(f_{ij} - f_{i\cdot} f_{\cdot j})^2$

observed expected

so that

$$\sum_{j=1}^{c} \frac{f_{ij} - f_{i\bullet} f_{\bullet j}}{f_{i\bullet}} = 0$$

for all $i \in \{1, \ldots, r\}$.

Gravity Center (5–7)

Consequence of the centering of the rows

The analysis is no longer carried out on

$$\mathbf{S} = \mathbf{F}^\top \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1},$$

but rather on $\mathbf{S}^* = (s^*_{jj'})$, where for all $j, j' \in \{1, \ldots, c\}$,

centered

$$s^*_{jj'} = \sum_{i=1}^{r} \frac{(f_{ij} - f_{i\bullet} f_{\bullet j})(f_{ij'} - f_{i\bullet} f_{\bullet j'})}{f_{i\bullet} f_{\bullet j'}}.$$

Connection With the Chi-Squared Statistic (X^2)

By definition,

$$\mathrm{trace}(\mathbf{S}^*) = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij} - f_{i\bullet} f_{\bullet j})^2}{f_{i\bullet} f_{\bullet j}} .$$

This is the formula for the statistic X^2 which is used to test for independence between the two variables.

Connection between $\mathbf{S}$ and $\mathbf{S}^*$

It can be shown that $s^*_{jj'} = s_{jj'} - f_{\bullet j}$, where for all $j, j' \in \{1, \ldots, c\}$,

$$s_{jj'} = \sum_{i=1}^{r} \frac{f_{ij} f_{ij'}}{f_{i\bullet} f_{\bullet j'}}.$$

It follows (still assuming that $c \leq r$) that the matrices $\mathbf{S}$ and $\mathbf{S}^*$ have the same first c normalized eigenvectors.

The projection of the ith row-point on the jth axis is given by

$$\left(\mathbf{D}_r^{-1}\mathbf{F}\varphi_j\right)_i = \frac{1}{f_{i\bullet}} \sum_{j'=1}^{c} f_{ij'}\varphi_{jj'} = \sqrt{\lambda_j}\,\Psi_{ji} \equiv \hat{\Psi}_{ji}.$$

Moreover,

$$\sum_{i=1}^{r} f_{i\bullet}\hat{\Psi}_{ji}^2 = \sum_{i=1}^{r} f_{i\bullet}\left(\sqrt{\lambda_j}\Psi_{ji}\right)^2 = \lambda_j.$$

Coordinates of the Column-Points

The projection of the kth column point on the jth axis is given by

$$\left(\mathbf{D}_c^{-1}\mathbf{F}^\top\psi_j\right)_k = \frac{1}{f_{\bullet k}}\sum_{i=1}^{r} f_{ik}\Psi_{ji} = \sqrt{\lambda_j}\,\varphi_{jk} \equiv \hat{\varphi}_{jk}.$$

Moreover,

$$\sum_{k=1}^{c} f_{\bullet k}\hat{\varphi}_{jk}^2 = \sum_{k=1}^{c} f_{\bullet k}\left(\sqrt{\lambda_j}\varphi_{jk}\right)^2 = \lambda_j.$$

Absolute and Relative Inertia

The absolute inertia of the ith row-point on the jth axis is given by

$$f_{i\bullet}\hat{\Psi}_{ji}^2$$

while the relative inertia of the ith row-point on the jth axis is given by

$$f_{i\bullet}\hat{\Psi}_{ji}^2/\lambda_j.$$

Similarly, the absolute inertia of the kth column-point on the jth axis is

$$f_{\bullet k}\hat{\varphi}_{jk}^2$$

while the relative inertia of the kth column-point on the jth axis is

$$f_{\bullet k}\hat{\varphi}_{jk}^2/\lambda_j.$$

Finally, the total inertia is given by $I = X^2/k_{\bullet\bullet}$, and hence it is related directly to the X^2 statistic for testing independence.

In $\mathbb{R}^r$ endowed with the metric $\mathbf{D}_r^{-1}$,

$$d^2(k, G_C) = \sum_{i=1}^{r} \frac{1}{f_{i\bullet}} \left(\frac{f_{ik}}{f_{\bullet j}} - f_{i\bullet} \right)^2$$

is the squared distance between G_C and the kth column-point.

An Application of Pythagoras' Theorem

The square of the projection of the kth column-point onto axis j is

$$d_j^2(k, G_C) = \left(\sqrt{\lambda_j}\varphi_{jk}\right)^2$$

and

$$\sum_{j=1}^{c} d_j^2(k, G_C) = \sum_{j=1}^{c} \left(\sqrt{\lambda_j}\varphi_{jk}\right)^2 = d^2(k, G_C).$$

Quality of the Representation (1–3)

The quality of the representation of the kth column-point in the jth axis is given by

$$\frac{d_j^2(k, G_C)}{d^2(k, G_C)} = \cos^2(\theta_{kj}),$$

where

$$\theta_{kj} = \text{angle between the point } k \text{ and its projection on axis } j.$$

Interpretation

- ✓ The larger $\cos^2(\theta_{kj})$, the better k is represented on the jth axis.
- ✓ Points that are far from the gravity center are distinctive.
- ✓ A similar interpretation is possible for the row-points.

Quality

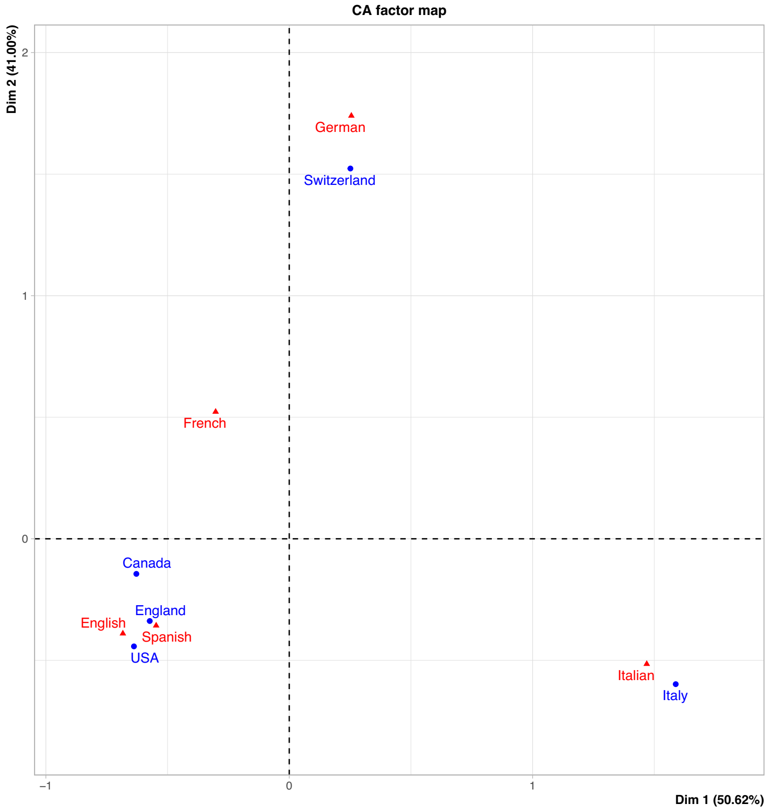
For a fixed number N of dimensions, one has

$$\text{Quality} = \sum_{i=1}^{N} \cos^2(\theta_{ij})$$

for the ith row-point.

A similar measure of quality is available for the kth column-point.

Factor Map:


CA factor map
Dim 2 (41.00%)
Dim 1 (50.62%)
2
1
0
−1
0
1
German
Switzerland
French
Canada
England
English
Spanish
USA
Italian
Italy

Example (2–2)

```
# Cos2 for the columns

Languages$col$cos2
             Dim 1      Dim 2
English 0.74616754 0.24257368   99%
French  0.12250578 0.36531441
Spanish 0.18264056 0.07786699
German  0.02072417 0.96325503
Italian 0.89039764 0.10948719   99%

# Contributions of the columns

Languages$col$contrib
            Dim 1     Dim 2
English 28.270732 11.348070
French   1.526655  5.621197
Spanish  2.164960  1.139682
German   1.250214 71.750730
Italian 66.787438 10.140321
```

Definition of Multiple Correspondence Analysis

Multiple correspondence analysis is an extension of binary correspondence analysis.

It makes it possible to analyze at a glance a multi-way contingency table.

A classical example of multi-way contingency table is an array containing the answers provided by respondents to a multiple choice exam comprising Q questions.

Multiple correspondence analysis is particularly useful to visualize the results of a survey and to attribute scores in order to segment the respondents in homogeneous groups.

Example

To illustrate multiple correspondence analysis, consider the following fictitious example:

ID	Type of Employee					Smoking			Total
	1	2	3	4	5	A	B	C	Q
1	0	0	1	0	0	0	1	0	2
2	0	1	0	0	0	1	0	0	2
3	1	0	0	0	0	0	1	0	2
4	0	0	0	0	1	0	0	1	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
193	0	1	0	0	0	1	0	0	2

In this example, there are $Q = 2$ questions. The table is thus of the form

$$\mathbf{Z} = [\mathbf{Z}_1 \mid \mathbf{Z}_2].$$

For a questionnaire with Q questions, the table would be of the form

$$\mathbf{Z} = [\mathbf{Z}_1 \mid \cdots \mid \mathbf{Z}_Q].$$

Notation

$$\begin{aligned} Q &= \text{number of questions}, \\ n &= \text{number of respondents}, \\ p_q &= \text{number of modalities (choices of answers) for question } q, \\ p &= p_1 + \cdots + p_Q. \end{aligned}$$

Empty Cells

The larger Q, the larger the number of empty cells.

Indeed, the proportion of non-empty cells is

$$\frac{nQ}{np} = \frac{Q}{p}\,.$$

If all the questions have the same number of possible answers, then

$$p_1 = \cdots = p_Q = \frac{p}{Q}\,,$$

and hence

$$\frac{Q}{p} = \frac{1}{p_1} \to 0 \quad \text{as } p_1 \to \infty.$$

Condensed Table

It is an $n \times Q$ table which identifies the answer provided by a respondent to each of the Q questions.

For example, in the table below the first respondent is an employee of Category 3 who is a smoker of Category 2.

ID	Type of Employee	Smoking
1	3	2
2	2	1
3	1	2
4	5	3
$\vdots$	$\vdots$	$\vdots$
193	2	1

Burt Table

A Burt table is another method for coding a contingency table involving more than two variables.

Given a table of responses

$$\mathbf{Z} = [\mathbf{Z}_1 \mid \cdots \mid \mathbf{Z}_Q],$$

the corresponding Burt table is the $p \times p$ matrix given by

$$\mathbf{B} = \mathbf{Z}\mathbf{Z}^\top,$$

viz.

$$\mathbf{B} = \begin{bmatrix} \mathbf{Z}_1^\top \mathbf{Z}_1 & \mathbf{Z}_1^\top \mathbf{Z}_2 & \cdots & \mathbf{Z}_1^\top \mathbf{Z}_Q \\ \mathbf{Z}_2^\top \mathbf{Z}_1 & \mathbf{Z}_2^\top \mathbf{Z}_2 & \cdots & \mathbf{Z}_2^\top \mathbf{Z}_Q \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{Z}_Q^\top \mathbf{Z}_1 & \mathbf{Z}_Q^\top \mathbf{Z}_2 & \cdots & \mathbf{Z}_Q^\top \mathbf{Z}_Q \end{bmatrix}.$$

Cyril Burt (1883–1971)

Sir Cyril L. Burt was an English educational psychologist and geneticist who also made contributions to statistics. He is known for his studies on the heritability of IQ.

Shortly after he died, his studies of inheritance of intelligence were discredited after evidence emerged indicating he had falsified research data, inventing correlations in separated twins which did not exist.

You can read about "The Burt Affair" on Wikipedia.

Characteristics of $Z_q^\top Z_{q'}$

- ✓ $\mathbf{Z}_q^\top \mathbf{Z}_q$ is a $p_q \times p_q$ diagonal matrix containing the answers to question q.
- ✓ The (j, j) element of $\mathbf{Z}_q^\top \mathbf{Z}_q$ is equal to the number d_{jj} of individuals who chose category j for question q.
- ✓ $\mathbf{Z}_q^\top \mathbf{Z}_{q'}$ is a contingency table providing the number of answers to questions q and q'.
- ✓ The (j, j') element of matrix $\mathbf{Z}_q^\top \mathbf{Z}_{q'}$ is equal to the number $d_{jj'}$ of individuals who chose category j for question q and category j' for question q'.

Example (cont'd)

	1	2	3	4	5	A	B	C
1	11	0	0	0	0	4	5	2
2	0	18	0	0	0	4	10	4
3	0	0	51	0	0	25	22	4
4	0	0	0	88	0	18	57	13
5	0	0	0	0	25	10	13	2
A	4	4	25	18	10	61	0	0
B	5	10	22	57	13	0	107	0
C	2	4	4	13	2	0	0	25

This matrix is $(5+3) \times (5+3)$.

Additional Notation

For each $i \in \{1, \ldots, Q\}$, set $\mathbf{D}_i = \mathbf{Z}_i^\top \mathbf{Z}_i$ and

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{D}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{D}_Q \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1^\top \mathbf{Z}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_2^\top \mathbf{Z}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}_Q^\top \mathbf{Z}_Q \end{bmatrix}.$$

Again, these matrices are $p \times p$.

Multiple Correspondence Analysis

A multiple correspondence analysis is a binary correspondence analysis performed either on the matrix $\mathbf{Z}$ or on the Burt table $\mathbf{B}$.

It will be shown that the result of the analysis is the same, whether it is performed on $\mathbf{Z}$ or on $\mathbf{B}$.

Analysis based on $\mathbf{Z}$ (1–3)

For standard binary correspondence analysis, one starts with a matrix

$$\mathbf{F} = (f_{ij}).$$

To carry out a multiple correspondence analysis on the matrix $\mathbf{Z}$, one has

$$\mathbf{F} = \frac{\mathbf{Z}}{nQ}$$

with

$$\sum_{i=1}^{n} \sum_{j=1}^{p} f_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{Z_{ij}}{nQ} = 1.$$

Analysis based on $\mathbf{Z}$ (2–3)

In standard binary correspondence analysis, one has

$$\mathbf{D}_n = \operatorname{diag}\left(f_{i\bullet}\right) \quad \text{and} \quad \mathbf{D}_p = \operatorname{diag}\left(f_{\bullet j}\right).$$

To carry out a multiple correspondence analysis on the matrix $\mathbf{Z}$, the sum of each row equals Q, and hence

$$\mathbf{D}_n = \frac{Q}{nQ}\,\mathbf{I}_n = \frac{\mathbf{I}_n}{n}\,.$$

Moreover,

$$\mathbf{D}_p = \frac{\mathbf{D}}{nQ} = \frac{1}{nQ}\operatorname{diag}(\mathbf{Z}_i^\top \mathbf{Z}_i).$$

Analysis based on $\mathbf{Z}$ (3–3)

As a result, the factors $\varphi_j = \mathbf{D}_p^{-1} u_j$ are such that

$$\mathbf{F}^\top \mathbf{D}_n^{-1} \mathbf{F} \mathbf{D}_p^{-1} u_j = \lambda_j u_j.$$

Therefore,

$$\mathbf{D}_p^{-1} \mathbf{F}^\top \mathbf{D}_n^{-1} \mathbf{F} \varphi_j = \lambda_j \varphi_j$$

ou equivalently,

$$\frac{1}{Q} \mathbf{D}^{-1} \mathbf{Z}^\top \mathbf{Z} \varphi_j = \lambda_j \varphi_j.$$

Analysis based on $\mathbf{B}$ (1–3)

For correspondence analysis with Burt's table, one has

$$\mathbf{F} = \frac{\mathbf{B}}{nQ^2}$$

because each of the Q blocks in $\mathbf{B}$ consists of integers whose sum is equal to n.

Furthermore, $\mathbf{B}$ is a symmetric matrix. Multiple correspondence analysis on Burt's table is thus performed in the case $n = p$.

In this special case, one has

$$\mathbf{D}_n = \mathbf{D}_p = \frac{\mathbf{D}}{nQ}.$$

Analysis based on **B** (2–3)

The factors in the multiple correspondence analysis of the Burt table are given by

$$\varphi_j^* = \mathbf{D}_n^{-1} v_j = nQ\mathbf{D}^{-1} v_j,$$

where

$$\mathbf{F}\mathbf{D}_p^{-1}\mathbf{F}^\top \mathbf{D}_n^{-1} v_j = \lambda_j^* v_j.$$

Equivalently, one has

$$\frac{1}{Q^2}\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top \mathbf{D}^{-1} v_j = \lambda_j^* v_j.$$

The factor φ_j^* is the solution to the equation

$$\frac{1}{Q^2}\,\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top\varphi_j^* = \lambda_j^*\mathbf{D}\varphi_j^*.$$

Upon multiplication on both sides by $\mathbf{D}^{-1}$, the same factor φ_j^* is seen to be a solution to

$$\frac{1}{Q^2}\,\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top\varphi_j^* = \lambda_j^*\varphi_j^*.$$

Comparison between MCA on $\mathbf{Z}$ and $\mathbf{B}$

For the analysis based on $\mathbf{Z}$, one has

$$\frac{1}{Q}\,\mathbf{D}^{-1}\mathbf{B}^\top \varphi_j = \lambda_j \varphi_j,$$

so that upon multiplication on both sides by $\mathbf{D}^{-1}\mathbf{B}/Q$, one finds

$$\frac{1}{Q^2}\,\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top \varphi_j = \lambda_j \frac{\mathbf{D}^{-1}\mathbf{B}\varphi_j}{Q} = \lambda_j^2 \varphi_j.$$

It follows that, for all $j \in \{1, \ldots, p\}$, one has

$$\lambda_j^* = \lambda_j^2 \quad \text{and} \quad \varphi_j^* = \varphi_j.$$

Remark

In the case $Q = 2$, multiple correspondence analysis on $\mathbf{Z}$ is equivalent to binary analysis on the matrix $\mathbf{Z}_2^\top \mathbf{Z}_1$.

In fact the jth vector Φ_j of the multiple correspondence analysis on $\mathbf{Z} = [\mathbf{Z}_1 \mid \mathbf{Z}_2]$ can be expressed in the form

$$\Phi_j = (\varphi_j, \psi_j)^\top ,$$

where φ_j and ψ_j are respectively the jth direct or dual factor of the analysis performed on $\mathbf{Z}_2^\top \mathbf{Z}_1$. Furthermore, given that

$$\lambda_j^* = j\text{th eigenvalue of } \mathbf{Z}_2^\top \mathbf{Z}_1,$$

then, for all $j \in \{1, \ldots, p\}$, one has

$$\lambda_j = \left(1 + \sqrt{\lambda_j^*}\right)/2.$$

Interpretation of the Resulting Graph

A factor map is created in the same way as in binary correspondence analysis.

However, the distance between points and the global geometry of the map can no longer be interpreted as in binary correspondence analysis.

Of particular interest are

- ✓ points which are close to one another or in the same quadrant;
- ✓ points which are in the same direction with respect to the origin.

Example 1: Land Use (1–3)

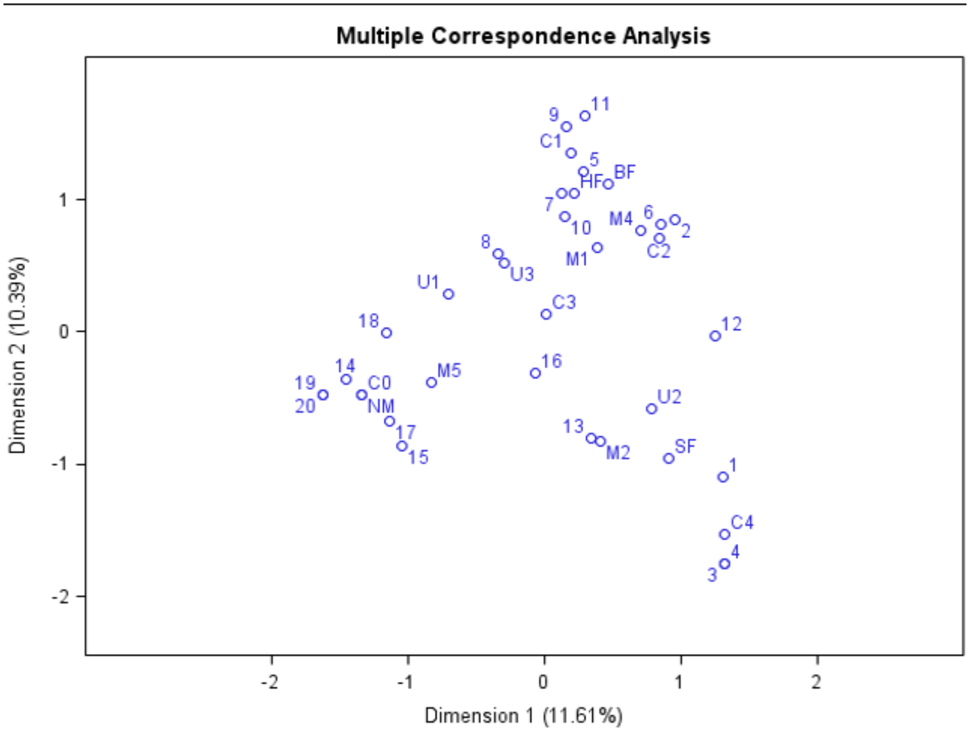
Information was collected about 20 farms in the Netherlands.

```
Humidity
     Ground humidity level (1, 2, 4, 5)

Management
     Land management type (SF = Standard Farm,
     BF = Biological Farm, HF = Holiday Farm,
     NM = Nature Conservation Management)

Production
     Type of production (U1 = Hay,
     U2 = Intermediate Production, U3 = Pasture)

Manure
     Manure use intensity level (C0, C1, C2, C3, C4)
```

Example 1: Land Use (2–3)

Farm	Humidity	Management	Production	Manure
1	M1	SF	U2	C4
2	M1	BF	U2	C2
3	M2	SF	U2	C4
4	M2	SF	U2	C4
5	M1	HF	U1	C2
6	M1	HF	U2	C2
7	M1	HF	U3	C3
8	M5	HF	U3	C3
9	M4	HF	U1	C1
10	M2	BF	U1	C1
11	M1	BF	U3	C1
12	M4	SF	U2	C2
13	M5	SF	U2	C3
14	M5	NM	U3	C0
15	M5	NM	U2	C0
16	M5	SF	U3	C3
17	M2	NM	U1	C0
18	M1	NM	U1	C0
19	M5	NM	U1	C0
20	M5	NM	U1	C0

Example 1: Land Use (3–3)



Example 2: Client Retention (1–2)

Businesses routinely try to understand which factors are associated with clientele mobility. The IBM Watson website provides an example of data analysis on client retention in telecoms.

```
No gender SeniorCitizen Partner Dependents      tenure PhoneService
1 Female              No      Yes         No ShortTenure          No
MultipleLines InternetService OnlineSecurity OnlineBackup
No              DSL                 No           Yes
  DeviceProtection TechSupport StreamingTV StreamingMovies Contract
1                No           No          No               No      MtM
PaperlessBilling  PaymentMethod MonthlyCharges Churn
Yes Electronicheck      LowCharge     No
```

Example 2: Client Retention (2–2)

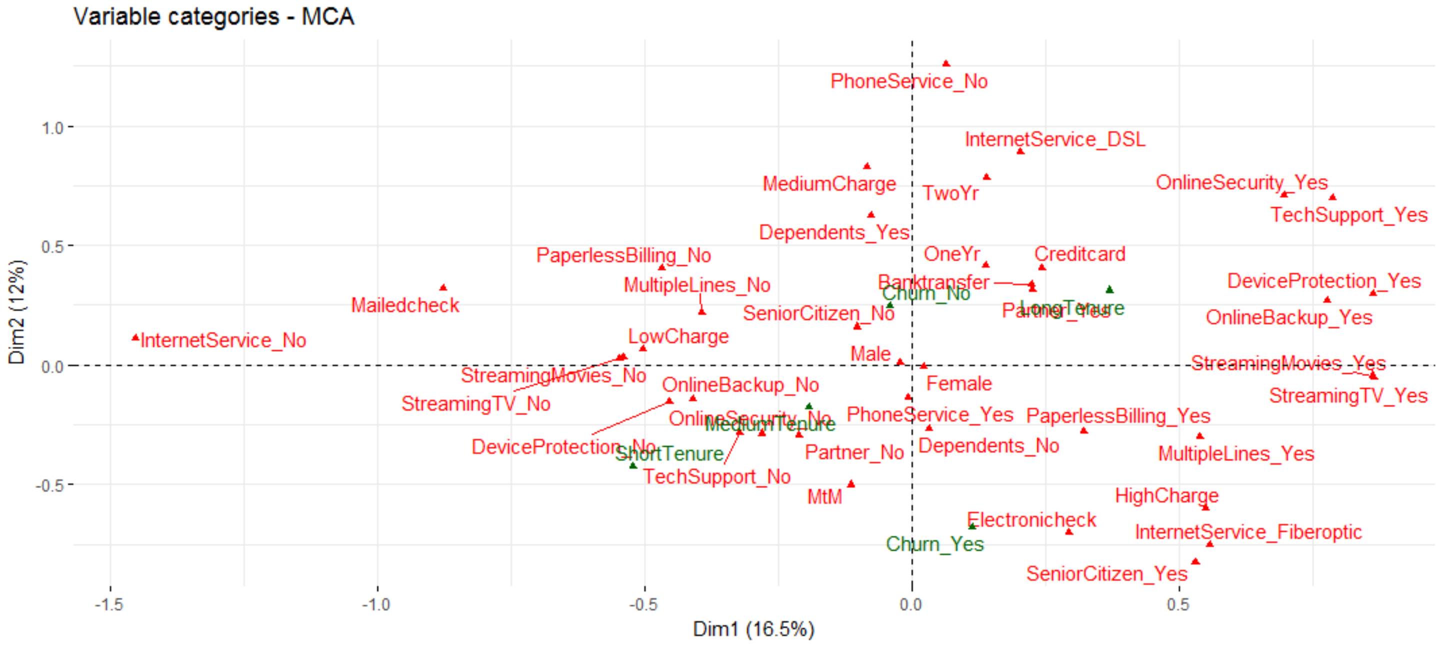


Illustration: Vehicles Sold (1–4)

Data are available on various models of cars sold in the USA in 1993.

```
Manufacturer
    Car manufacturer

Type
     Type of vehicle

Airbags
     Position of the airbags

Traction
     Front-wheel drive, Rear-wheel drive
```

Illustration: Vehicles Sold (2–4)

```
Manufacturer Category Airbags Traction
Acura Small None Front
Acura Midsize DriPas Front
Audi Compact Driver Front
Audi Midsize DriPas Front
BMW Midsize Driver Rear
Buick Midsize Driver Front
Buick Large Driver Front
Buick Large Driver Rear
Buick Midsize Driver Front
Cadillac Large Driver Front
Cadillac Midsize DriPas Front
Chevrolet Compact None Front
Chevrolet Compact Driver Front
Chevrolet Van None 4WD
Chevrolet Large Driver Rear
Chevrolet Sporty Driver Rear
Chrysler Large DriPas Front
Chrysler Compact DriPas Front
Chrysler Large Driver Front
Dodge Small None Front
Dodge Small Driver Front
Dodge Compact Driver Front
.          .        .      .
.          .        .      .
.          .        .      .
```

Illustration: Vehicles Sold (3–4)

Without the manufacturers

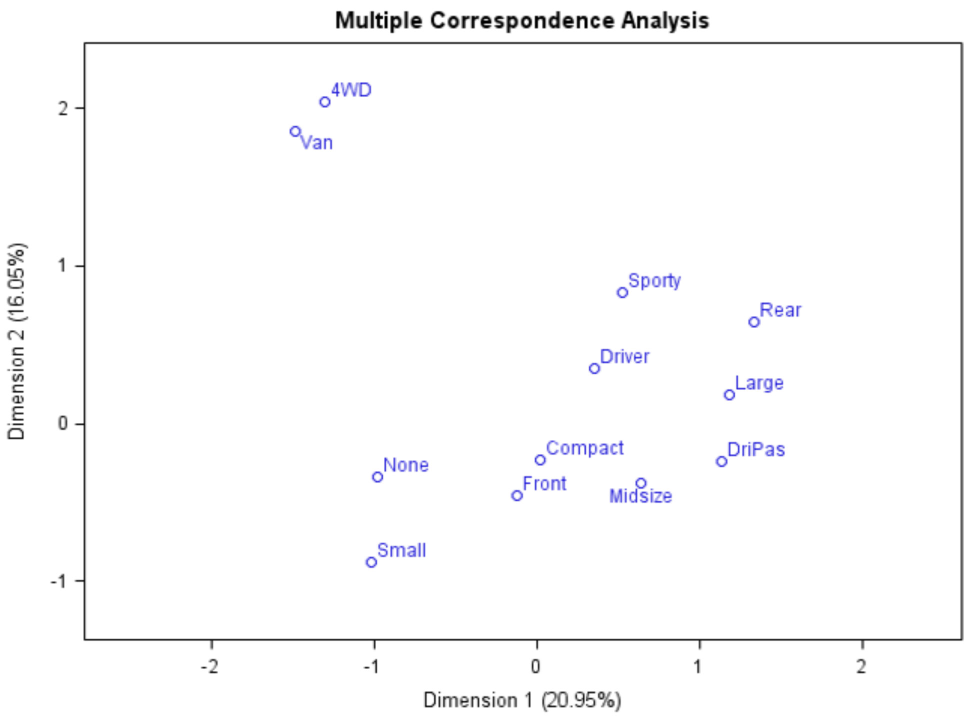


Illustration: Vehicles Sold (4–4)

With the manufacturers