Correspondence Analysis

Jan Graffelman[1]

[1]Department of Statistics and Operations Research
Universitat Politècnica de Catalunya
Barcelona, Spain

jan.graffelman@upc.edu

March 7, 2020

Contents

Some History

- Benzécri, J.P. (1973) *Analyse des Données*, Dunod, Paris.
- Greenacre, M.J. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press.

Objective

- Study the relationships between categorical variables.
  - simple correspondence analysis (CA): two categorical variables.
  - multiple correspondence analysis (MCA): many categorical variables.
- Provide a picture of the association between categorical variables.

Example data set: Dutch calves

	Type of calf		
Ease of delivery	ET	IVP	AI
1	97	150	1686
2	152	183	1339
3	377	249	1209
4	335	227	656
5	42	136	277
6	9	71	62

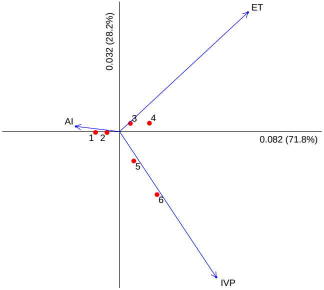
Download Calves.dat

- n = 7257 calves.
- method of production (ET = Embryo Transfer, IVP = In Vitro Production, AI = Artificial Insemination)
- Ease of delivery, scored on a scale from 1 (normal) to 6 (very heavy).
- $\chi^2_{10} = 833.16$ with $p < 0.001$

There is association, but what is the nature of this association?

Result of Correspondence Analysis

	Type of calf		
Ease of delivery	ET	IVP	AI
1	97	150	1686
2	152	183	1339
3	377	249	1209
4	335	227	656
5	42	136	277
6	9	71	62



Some notation

- $\mathbf{N}$ the $I \times J$ contingency table.
- $\mathbf{P} = \mathbf{N}/n$ with $n = \mathbf{1}'\mathbf{N}\mathbf{1}$, and thus $\mathbf{1}'\mathbf{P}\mathbf{1} = 1$.
- $\mathbf{P}$ a matrix of probabilities (the correspondence matrix).

	ET	IVP	AI	**r**
1	0.013	0.021	0.232	0.266
2	0.021	0.025	0.185	0.231
3	0.052	0.034	0.167	0.253
4	0.046	0.031	0.090	0.168
5	0.006	0.019	0.038	0.063
6	0.001	0.010	0.009	0.020
c	0.139	0.140	0.721	1.000

- Row masses

$$r_i = \sum_{j=1}^{J} p_{ij} \qquad \mathbf{r} = \mathbf{P}\mathbf{1} \qquad \mathbf{D}_r = diag(\mathbf{r})$$

- Column masses

$$c_j = \sum_{i=1}^{I} p_{ij} \qquad \mathbf{c} = \mathbf{P}'\mathbf{1} \qquad \mathbf{D}_c = diag(\mathbf{c})$$

Profiles

- A profile is a vector of non-negative elements that sum 1.
- The contingency table can be converted into a matrix of profiles.

Row profiles			
	ET	IVP	AI
1	0.05	0.08	0.87
2	0.09	0.11	0.80
3	0.21	0.14	0.66
4	0.28	0.19	0.54
5	0.09	0.30	0.61
6	0.06	0.50	0.44

Column profiles			
	ET	IVP	AI
1	0.10	0.15	0.32
2	0.15	0.18	0.26
3	0.37	0.25	0.23
4	0.33	0.22	0.13
5	0.04	0.13	0.05
6	0.01	0.07	0.01

Profiles

- Row (column) profiles are obtained by summing the elements of a row (column) in $\mathbf{P}$ and dividing by the total.
- $\mathbf{R} = \mathbf{D}_r{}^{-1}\mathbf{P}$ row profiles $\qquad \mathbf{C} = \mathbf{D}_c{}^{-1}\mathbf{P}'$ column profiles
- Row and column masses turn out be weighted averages of the profiles

$$\mathbf{r}'\mathbf{D}_r{}^{-1}\mathbf{P} = \mathbf{1}'\mathbf{P} = \mathbf{c}' \qquad \mathbf{c}'\mathbf{D}_c{}^{-1}\mathbf{P}' = \mathbf{1}'\mathbf{P}' = \mathbf{r}'$$

Profiles and average profile

Row profiles			
	ET	IVP	AI
1	0.05	0.08	0.87
2	0.09	0.11	0.80
3	0.21	0.14	0.66
4	0.28	0.19	0.54
5	0.09	0.30	0.61
6	0.06	0.50	0.44
c	0.14	0.14	0.72

Centred row profiles:

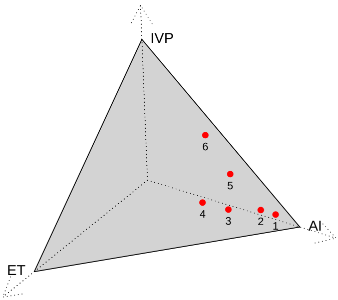
$$\mathbf{D}_r{}^{-1}\mathbf{P} - \mathbf{1}\mathbf{c}'$$

Column profiles				
	ET	IVP	AI	**r**
1	0.10	0.15	0.32	0.27
2	0.15	0.18	0.26	0.23
3	0.37	0.25	0.23	0.25
4	0.33	0.22	0.13	0.17
5	0.04	0.13	0.05	0.06
6	0.01	0.07	0.01	0.02

Centred column profiles:

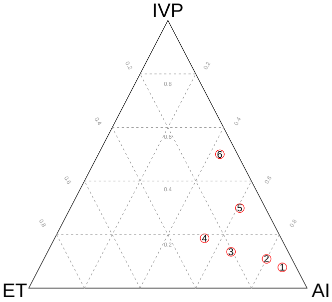
$$\mathbf{D}_c{}^{-1}\mathbf{P}' - \mathbf{1}\mathbf{r}'$$

Profiles of calves data in 3D



Row profiles of the calves data in two dimensions

- The profile matrix for the data under consideration has rank 2.
- In this case, the profiles can be represented in a ternary plot.



Dimensionality

- The column rank of the matrix of row profiles is at most $J - 1$
- The row rank of the matrix of column profiles is at most $I - 1$
- "The" rank of the CA solution is $min(I - 1, J - 1)$

χ^2 statistic

	ET	IVP	AI
1	97	150	1686
	269.56	270.63	1392.81
2	152	183	1339
	233.44	234.36	1206.19
3	377	249	1209
	255.89	256.91	1322.20
4	335	227	656
	169.85	170.52	877.62
5	42	136	277
	63.45	63.70	327.85
6	9	71	62
	19.80	19.88	102.32

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \frac{(97 - 269.56)^2}{269.56} + \cdots + \frac{(62 - 102.32)^2}{102.32} = 833.16$$

Profiles and χ^2 statistic

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i,j} \frac{(np_{ij} - nr_i c_j)^2}{nr_i c_j} = n \sum_{i,j} \frac{(p_{ij} - r_i c_j)^2}{r_i c_j}$$

$$\frac{\chi^2}{n} = \sum_{i,j} \frac{(p_{ij} - r_i c_j)^2}{r_i c_j} = \sum_{i,j} r_i^2 \frac{(\frac{p_{ij}}{r_i} - c_j)^2}{r_i c_j} = \sum_{i,j} r_i \frac{(\frac{p_{ij}}{r_i} - c_j)^2}{c_j} = \sum_i r_i \sum_j \frac{(\frac{p_{ij}}{r_i} - c_j)^2}{c_j}$$

Likewise, for column profiles

$$\frac{\chi^2}{n} = \sum_j c_j \sum_i \frac{(\frac{p_{ij}}{c_j} - r_i)^2}{r_i}$$

- The quantity $\frac{\chi^2}{n}$ is known as the total inertia of the contingency table.
- Note that $\sum_j (\frac{p_{ij}}{r_i} - c_j)^2$ is squared Euclidean distance between profile i and average row profile
- Note that $\sum_j \frac{1}{c_j} (\frac{p_{ij}}{r_i} - c_j)^2$ is weighted squared Euclidean distance between profile i and average row profile (called χ^2 distance)
- Inertia is a weighted average of weigthed squared Euclidean distances.
- Inertia is a measure of spread of the profiles w.r.t. their average.

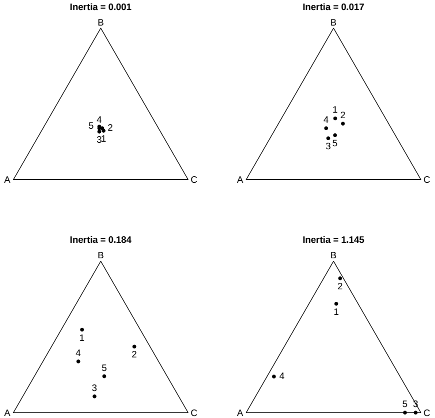
The geometrical interpretation of Inertia

	A	B	C
1	20	20	22
2	19	20	20
3	21	19	20
4	20	20	19
5	20	21	19
	100	100	100

	A	B	C
1	15	21	16
2	17	24	24
3	26	18	22
4	22	20	17
5	20	17	21
	100	100	100

	A	B	C
1	14	23	5
2	7	34	37
3	27	6	23
4	34	25	15
5	18	12	20
	100	100	100

	A	B	C
1	4	23	5
2	1	47	5
3	2	0	65
4	91	30	5
5	2	0	20
	100	100	100



Limiting situations

- Perfect independence: minimal inertia $= 0, \chi^2 = 0$.
- Perfect association: maximal inertia $= min(I-1, J-1)$.

Larger tables

- The profiles of the calves data can be represented exactly in two-dimensional space
- Profiles of $I \times J$ contingency table can be represented exactly in $min(I-1, J-1)$ dimensional space.
- We search for an approximation of the profiles in one, two or at most three dimensions.
- The criterion is to miminize errors in the approximation of the profiles, which is equivalent to maximizing the inertia of the profiles in a k dimensional subspace.
- Equivalently, we do a least-squares approximation to the matrix of deviations from independence
- The optimal solution is obtained by solving an eigenvalue-eigenvector equation, or by doing a singular value decomposition.

Solution as a singular value decomposition

- In CA we do the SVD of the matrix of standardized residuals:

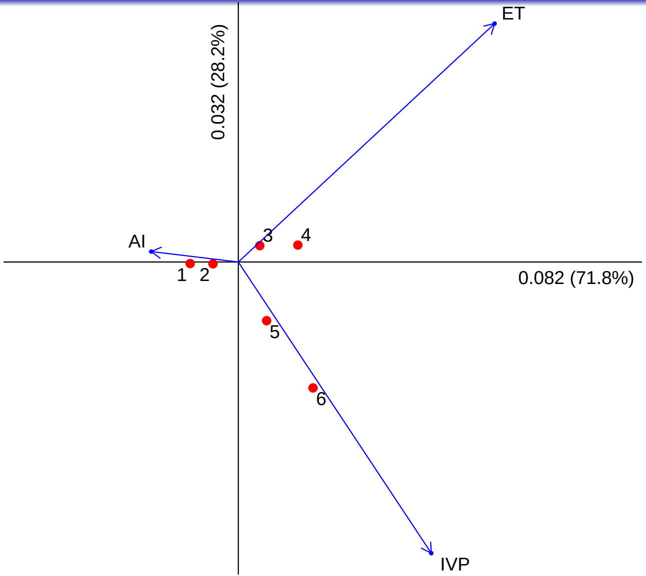
$$\mathbf{D}_r{}^{-1/2}(\mathbf{P} - \mathbf{rc}')\mathbf{D}_c{}^{-1/2} = \mathbf{UDV}'$$

- We approximate residuals in low-dimensional space by using the first two singular values and singular vectors only.
- Biplot coordinates
    - Principal coordinates $\mathbf{F}_p = \mathbf{D}_r{}^{-1/2}\mathbf{UD}$
    - Standard column coordinates $\mathbf{G}_s = \mathbf{D}_c{}^{-1/2}\mathbf{V}$
- Standard and principal coordinates are related
    - $\mathbf{G}_p = \mathbf{G}_s\mathbf{D}_\lambda^{\frac{1}{2}}$
    - $\mathbf{F}_p = \mathbf{F}_s\mathbf{D}_\lambda^{\frac{1}{2}}$

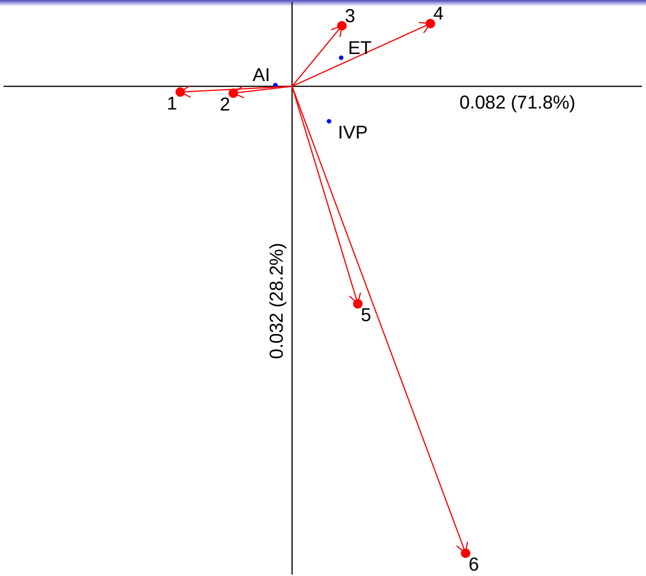
Graphical output of Correspondence analysis

- Joint plot of the rows of $\mathbf{F}_s$ and $\mathbf{G}_p$ (biplot of the row profiles)
- Joint plot of the rows of $\mathbf{F}_p$ and $\mathbf{G}_s$ (biplot of the column profiles)
- We have $\mathbf{F}_s\mathbf{G}_p' = (\mathbf{D}_r{}^{-1}\mathbf{P} - \mathbf{1c}')\mathbf{D}_c{}^{-1}$
- and also $\mathbf{G}_s\mathbf{F}_p' = (\mathbf{D}_c{}^{-1}\mathbf{P}' - \mathbf{1r}')\mathbf{D}_r{}^{-1}$
- Several alternatives to scale CA output have been described in the literature.

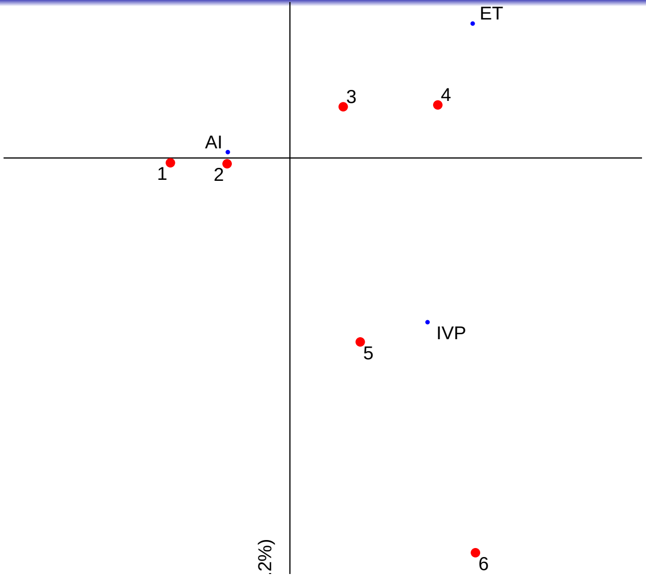
Biplot of the Calves data (row profiles)



Biplot of the Calves data (column profiles)



Biplot of the Calves data (symmetric scaling)



Inertia decomposition Calves data

	1	2
Eigenvalue	0.082	0.032
Proportion	0.718	0.282
Cumulative	0.718	1.000

Note that

$$\frac{\chi^2}{n} = \frac{833.1562}{7257} = 0.114 = 0.082 + 0.032$$

Transition relationships (barycentric relationships)

From previous results

- $\mathbf{F}_p = \mathbf{D}_r{}^{-1}\mathbf{P}\mathbf{G}_s$
- $\mathbf{G}_p = \mathbf{D}_c{}^{-1}\mathbf{P}'\mathbf{F}_s$
- Principal coordinates of the rows are weighted averages of standard coordinates of the columns
- Useful for calculating coordinates of supplementary points

Supplementary points

- Supplementary points or inactive points are rows (columns) of the data matrix, usually collected under different conditions, that do not intervene in the computation of the solution.
- However, their representation in a biplot, posterior to the analysis, can be helpful for interpretation.
- Supplementary points can be situated in CA biplots by expressing them as profiles and using the transition relationships.

Contributions to Inertia

- In PCA we have seen that the total variance of data matrix can be decomposed into contributions made by dimensions (principal components), by variables, and finally by individual observations.
- In CA, a similar decomposition is possible, where the total inertia of a contingency table can be decomposed into contributions made by dimensions (principal axis), by the rows of the table, the columns of the table, and finally, the individual cells of a table.
- Such a decomposition is useful for spotting influential points in the analysis.

Contributions to Inertia

- we had

$$\frac{\chi^2}{n} = \sum_i r_i \sum_j \frac{(\frac{p_{ij}}{r_i} - c_j)^2}{c_j} = \sum_j c_j \sum_i \frac{(\frac{p_{ij}}{c_j} - r_i)^2}{r_i}$$

- each row (and column) make a contribution to the total inertia, these are called row and column inertias.
- note that

$$\frac{\chi^2}{n} = \sum_{i,j} \frac{(p_{ij} - r_i c_j)^2}{r_i c_j} = \mathrm{tr}(\mathbf{D}_r{}^{-1}(\mathbf{P} - \mathbf{rc}')\mathbf{D}_c{}^{-1}(\mathbf{P} - \mathbf{rc})') = \mathrm{tr}(\mathbf{D}_\lambda)$$

- The eigenvalues are called principal inertias and constitute the contribution of each dimension in the solution to the total
- the inertias of each row (column) can be decomposed into contributions made by the principal axis. This allows one to judge how much of the inertia of each row (column) is accounted for by each axis, and to compute goodness-of-fit statistics for each point.

Some R code

```
X <- read.table("http://www-eio.upc.es/~jan/data/calves.dat",header=TRUE)
X <- X[,-1]
library(ca)
out <- ca(X)
out
 Principal inertias (eigenvalues):
           1        2
Value      0.082447 0.03236
Percentage 71.81%   28.19%

 Rows:
                1         2        3        4         5         6
Mass     0.266364  0.230674 0.252859 0.167838  0.062698  0.019567
ChiDist  0.341890  0.180092 0.191439 0.439478  0.462185  1.038777
Inertia  0.031135  0.007481 0.009267 0.032416  0.013393  0.021114
Dim. 1  -1.190087 -0.625554 0.530527 1.472321  0.700072  1.847043
Dim. 2  -0.060380 -0.072450 0.644543 0.667478 -2.313553 -4.965224

 Columns:
              ET       IVP        AI
Mass    0.139452  0.140003  0.720546
ChiDist 0.604587  0.541101  0.178050
Inertia 0.050973  0.040991  0.022843
Dim. 1  1.819589  1.370025 -0.618353
Dim. 2  1.691167 -2.065368  0.074001
```

Some R code

```
> summary(out)

Principal inertias (eigenvalues):

 dim     value       %   cum%   scree plot
 1       0.082447  71.8  71.8   *************************
 2       0.032360  28.2 100.0
         -------- -----
 Total: 0.114807 100.0


Rows:
    name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
1 |    1 |  266 1000  271 | -342 999 377 |  -11   1   1 |
2 |    2 |  231 1000   65 | -180 995  90 |  -13   5   1 |
3 |    3 |  253 1000   81 |  152 633  71 |  116 367 105 |
4 |    4 |  168 1000  282 |  423 925 364 |  120  75  75 |
5 |    5 |   63 1000  117 |  201 189  31 | -416 811 336 |
6 |    6 |   20 1000  184 |  530 261  67 | -893 739 482 |

Columns:
    name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
1 |   ET |  139 1000  444 |  522 747 462 |  304 253 399 |
2 |  IVP |  140 1000  357 |  393 529 263 | -372 471 597 |
3 |   AI |  721 1000  199 | -178 994 276 |   13   6   4 |
```

Another example: Eye and hair colour (Fisher, 1940)

		Eye colour				
		Light	Blue	Medium	Dark	Total
	Fair	688	326	343	98	1455
Hair	Red	116	38	84	48	286
colour	Medium	584	241	909	403	2137
	Dark	188	110	412	681	1391
	Black	4	3	26	85	118
	Total	1580	718	1774	1315	5387

$\chi^2 = 1240.039$, with 12 df p-value $< 2.2e - 16$

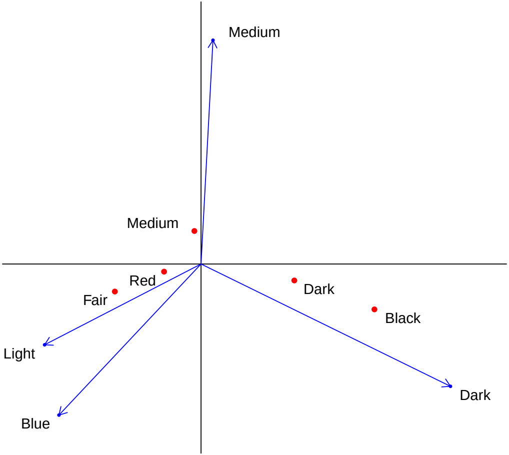
Row profiles and average row profile

	Eye colour			
	Light	Blue	Medium	Dark
Fair	0.473	0.224	0.236	0.067
Red	0.406	0.133	0.294	0.168
Medium	0.273	0.113	0.425	0.189
Dark	0.135	0.079	0.296	0.490
Black	0.034	0.025	0.220	0.720
Average	0.293	0.133	0.329	0.244

Eye and hair colour

```
> out.ca <- ca(X)
> summary(out.ca)

Principal inertias (eigenvalues):

 dim     value       %   cum%   scree plot
 1       0.199245  86.6  86.6   *************************
 2       0.030087  13.1  99.6   ****
 3       0.000859   0.4 100.0
         -------- -----
 Total: 0.230191 100.0


Rows:
    name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
1 | Fair |  270 1000  383 | -544 907 401 | -174  93 271 |
2 |  Red |   53  803   16 | -233 770  14 |  -48  33   4 |
3 | Medm |  397 1000   78 |  -42  39   4 |  208 961 572 |
4 | Dark |  258 1000  401 |  589 969 449 | -104  30  93 |
5 | Blck |   22  998  122 | 1094 934 132 | -286  64  60 |

Columns:
    name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
1 | Lght |  293  995  259 | -441 956 286 |  -88  39  76 |
2 | Blue |  133  979  111 | -400 836 107 | -165 143 121 |
3 | Medm |  329  999   88 |   34  18   2 |  245 981 657 |
4 | Dark |  244 1000  543 |  703 965 605 | -134  35 145 |
```

Biplot of the row profiles



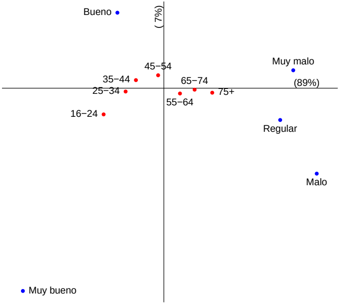
Multiple categorical variables

Approaches for treating multiple categorical variables:

- Interactive coding of categorical variables
- Concatenating tables rowwise or columnwise and analyzing the "broad" or the "long" matrix.
- Multiple correspondence analysis
- ....

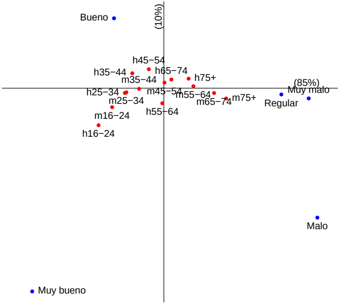
Interactive coding

	Muy malo	Malo	Regular	Bueno	Muy bueno
16-24	0	4	12	95	53
25-34	0	9	47	206	71
35-44	7	13	88	341	87
45-54	6	30	94	269	45
55-64	13	33	111	168	42
65-74	10	36	157	177	30
75+	9	63	184	171	19



Interactive coding

	Muy malo	Malo	Regular	Bueno	Muy bueno
h16-24	0	2	5	50	32
h25-34	0	6	19	98	33
h35-44	4	2	42	182	40
h45-54	2	10	36	131	19
h55-64	4	15	34	74	26
h65-74	3	12	54	100	17
h75+	1	15	59	82	9
m16-24	0	2	7	45	21
m25-34	0	3	28	108	38
m35-44	3	11	46	159	47
m45-54	4	20	58	138	26
m55-64	9	18	77	94	16
m65-74	7	24	103	77	13
m75+	8	48	125	89	10



Example data set: social survey data (2002; Spain)

The categorical variables are questions regarding working women and family. We consider 8 categorical questionnaire variables:

A	a working mother can establish a warm relationship with her child
B	a pre-school child suffers if his or her mother works
C	when a woman works the family life suffers
D	what women really want is a home and kids
E	running a household is just as satisfying as a paid job
F	work is best for a woman's independence
G	a man's job is to work; a woman's job is the household
H	working women should get paid maternity leave

Repondents give an answer in a Likert scale (1 = strongly agree, 2 = agree, 3 = neither agree nor disagree, 4 = disagree, 5 = strongly disagree).

Demographic variables

gender	1 = male, 2 = female
marital status	1 = married, 2 = widowed, 3 = divorced, 4 = separated, 5 = single
education	0 = none, 1 = lowest, 2 = above lowest, 3 = higher secondary, 4 = above h.s., 5 = university degree
age	1 = 16-25, 2 = 26-35, 3 = 36-45, 4 = 46-55, 5 = 56-65, 6 = 66+

Download Women.dat

A look at the data

	A	B	C	D	E	F	G	H	g	m	e	a
1	2	4	3	3	4	1	4	1	1	1	3	4
2	2	4	3	9	4	1	4	1	1	1	3	3
3	3	2	2	3	4	1	3	2	2	2	1	6
4	3	9	2	2	2	1	3	1	1	1	1	6
5	9	1	2	2	3	2	3	1	2	1	1	4
6	2	4	4	4	2	2	5	1	1	5	4	2
7	1	3	2	3	2	4	4	2	2	1	5	5
8	2	9	4	9	9	1	3	1	1	5	3	5
9	2	4	2	3	4	4	5	1	2	1	5	3
10	2	4	2	2	4	2	4	1	2	1	2	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Total sample size: 2471
Missing: 364
Total without missings: 2107

Two types of MCA

MCA is the application of CA to:

- The indicator matrix
- The Burt matrix

Indicator matrix

Categorical variables coded into binary variables.

	Original data								$\mathbf{Z}_1$					$\mathbf{Z}_2$						
	A	B	C	D	E	F	G	H	1	2	3	4	5	1	2	3	4	5	1	2
1	2	4	3	3	4	1	4	1	0	1	0	0	0	0	0	0	1	0	0	0
2	3	2	2	3	4	1	3	2	0	0	1	0	0	0	1	0	0	0	0	1
3	2	4	4	4	2	2	5	1	0	1	0	0	0	0	0	0	1	0	0	0
4	1	3	2	3	2	4	4	2	1	0	0	0	0	0	0	1	0	0	0	1
5	2	4	2	3	4	4	5	1	0	1	0	0	0	0	0	0	1	0	0	1
6	2	4	2	2	4	2	4	1	0	1	0	0	0	0	0	0	1	0	0	1
7	2	2	2	4	4	2	4	1	0	1	0	0	0	0	1	0	0	0	0	1
8	4	2	2	2	4	1	5	1	0	0	0	1	0	0	1	0	0	0	0	1
9	4	2	2	4	4	2	4	1	0	0	0	1	0	0	1	0	0	0	0	1
10	3	3	3	3	2	2	4	2	0	0	1	0	0	0	0	1	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Inertia of the indicator matrix

$$\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_q] \quad \text{with } \mathbf{Z}_{n \times J}$$

Q = number of categoral variables.
J_q = number of categories for variable q.
J = total number of categories.
$In(\cdot)$ = Inertia.

$$J = \sum_{q=1}^{Q} J_q$$

$$In(\mathbf{Z}_q) = J_q - 1$$

$$In(\mathbf{Z}) = \frac{\sum_q \mathbf{Z}_q}{Q} = \frac{J - Q}{Q}$$

Note: the inertia of a concatenated table is the mean of the inertias of all subtables.

Inertia per dimension: $1/Q$

Burt matrix

The Burt matrix is a symmetric $J \times J$ matrix containing all possible two-way tables of the Q categorial variables.

$$\mathbf{B} = \mathbf{Z}'\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1'\mathbf{Z}_1 & \mathbf{Z}_1'\mathbf{Z}_2 & \mathbf{Z}_1'\mathbf{Z}_3 & \cdots & \mathbf{Z}_1'\mathbf{Z}_q \\ \mathbf{Z}_2'\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{Z}_2 & \mathbf{Z}_2'\mathbf{Z}_3 & \cdots & \mathbf{Z}_2'\mathbf{Z}_q \\ \mathbf{Z}_3'\mathbf{Z}_1 & \mathbf{Z}_3'\mathbf{Z}_2 & \mathbf{Z}_3'\mathbf{Z}_3 & \cdots & \mathbf{Z}_3'\mathbf{Z}_q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_q'\mathbf{Z}_1 & \mathbf{Z}_q'\mathbf{Z}_2 & \mathbf{Z}_q'\mathbf{Z}_3 & \cdots & \mathbf{Z}_q'\mathbf{Z}_q \end{bmatrix}$$

Burt matrix for Women data

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5
A1	397	0	0	0	0	19	113	42	132	91	37	91	45	154	70
A2	0	932	0	0	0	18	362	126	405	21	21	405	128	359	19
A3	0	0	91	0	0	2	44	22	21	2	8	44	25	13	1
A4	0	0	0	598	0	40	411	42	101	4	48	422	36	88	4
A5	0	0	0	0	89	45	26	5	6	7	51	26	3	6	3
B1	19	18	2	40	45	124	0	0	0	0	80	34	4	3	3
B2	113	362	44	411	26	0	956	0	0	0	52	673	65	154	12
B3	42	126	22	42	5	0	0	237	0	0	12	82	82	59	2
B4	132	405	21	101	6	0	0	0	665	0	8	182	79	378	18
B5	91	21	2	4	7	0	0	0	0	125	13	17	7	26	62
C1	37	21	8	48	51	80	52	12	8	13	165	0	0	0	0
C2	91	405	44	422	26	34	673	82	182	17	0	988	0	0	0
C3	45	128	25	36	3	4	65	82	79	7	0	0	237	0	0
C4	154	359	13	88	6	3	154	59	378	26	0	0	0	620	0
C5	70	19	1	4	3	3	12	2	18	62	0	0	0	0	97
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Inertia of the Burt matrix

- If all two-way tables have the same margins (no missing data on any of the variables) then the inertia of the Burt matrix is the average of the inertias of all two-way tables.
- If there is missing data, this will be approximately true.

	A	B	C	D	E	F	G	H
A	4.000	0.388	0.369	0.164	0.113	0.154	0.227	0.060
B	0.388	4.000	0.837	0.271	0.182	0.113	0.261	0.038
C	0.369	0.837	4.000	0.427	0.188	0.137	0.229	0.068
D	0.164	0.271	0.427	4.000	0.373	0.160	0.451	0.052
E	0.113	0.182	0.188	0.373	4.000	0.203	0.281	0.087
F	0.154	0.113	0.137	0.160	0.203	4.000	0.154	0.110
G	0.227	0.261	0.229	0.451	0.281	0.154	4.000	0.084
H	0.060	0.038	0.068	0.052	0.087	0.110	0.084	3.000

$$Inertia(\mathbf{B}) = 0.677625$$

Adjusting the inertia of the Burt matrix

- The matrices on the diagonal of the Burt matrix have maximal inertia, $J_q - 1$ each.
- We wish to ignore their contribution to the total inertia, and to take only the inertia in the off-diagonal tables into account.
- Total inertia in the Burt matrix: $Q^2 In(\mathbf{B})$
- Total inertia on the diagonal: $\sum_{q=1}^{Q} (J_q - 1) = J - Q$
- Total off-diagonal inertia: $Q^2 In(\mathbf{B}) - (J - Q)$
- Inertia in off-diagonal part of the Burt matrix:

$$In_{adj}(\mathbf{B}) = \frac{Q}{Q-1}\left(In(\mathbf{B}) - \frac{J-Q}{Q^2}\right)$$

- For the data under study:

$$In_{adj}(\mathbf{B}) = \frac{8}{7}\left(0.677625 - \frac{39-8}{8^2}\right) = 0.22086$$

Adjusting the principal inertias of the Burt matrix

- In the analysis based on the Burt matrix, principal inertias can also be adjusted.
- The adjusted principal inertias sum to the off-diagonal inertia of the Burt matrix.

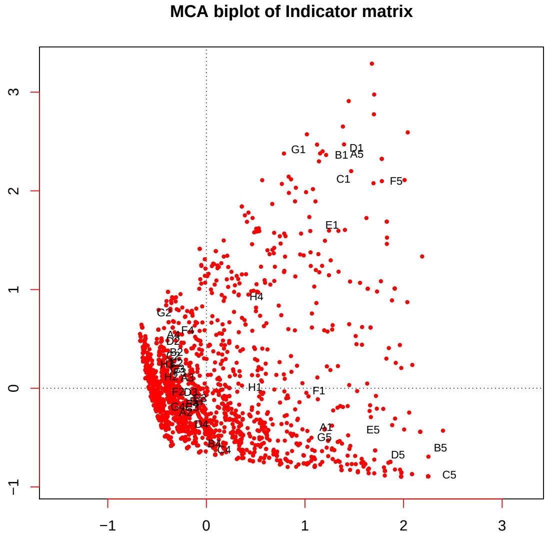
Adjusted principal inertias $\lambda_{k,adj}$ are obtained as:

$$\lambda_{k,adj} = \left(\frac{Q}{Q-1} \left(\sqrt{\lambda_k} - \frac{1}{Q} \right) \right)^2 \text{ for } \sqrt{\lambda_k} > \frac{1}{Q}$$

MCA with **Z** or **B**

- Standard coordinates of MCA with **Z** or **B** are the same.
- Eigenvalues of the Burt matrix are the squares of the eigenvalues of **Z**.
- Percentages of explained inertia are therefore higher when using **B**.
- Principal coordinates of MCA with **B** are shrunk w.r.t. MCA with **Z**.

MCA of indicator matrix (Women)



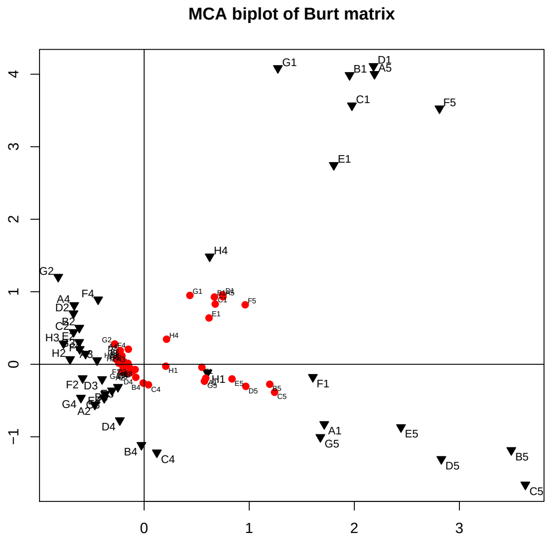
Inertia decomposition using Indicator matrix

	1	2	3	4	5	...	31
λ_i	0.424	0.329	0.234	0.226	0.155	...	0.049
fraction	0.110	0.085	0.060	0.058	0.040	...	0.013
cumul.	0.110	0.194	0.255	0.313	0.353	...	1.000

$$\sum \lambda_i = 3.875 = \frac{J - Q}{Q} = \frac{39 - 8}{8}$$

$$\frac{1}{Q} = 0.125$$

MCA of the Burt matrix



Inertia decomposition using Burt matrix

Unadjusted inertias:

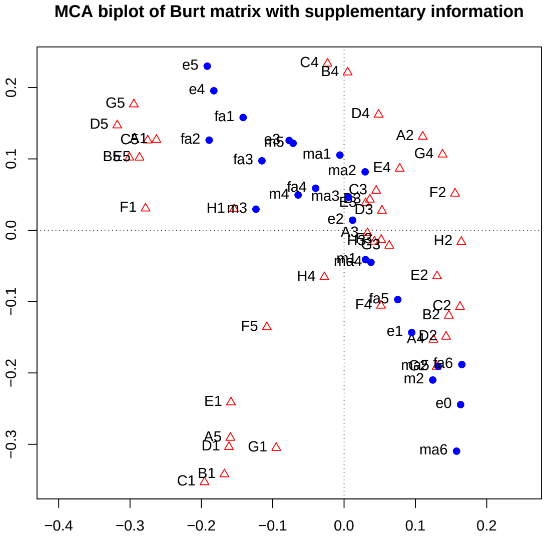
	1	2	3	4	5	...	31
Eigenvalue	0.180	0.108	0.055	0.051	0.024	...	0.002
Proportion	0.266	0.159	0.081	0.075	0.036	...	0.004
Cumulative	0.266	0.425	0.506	0.581	0.616	...	1.000

Adjusted inertias:

	1	2	3	4	5	6	7	8	9
Eigenvalue	0.117	0.054	0.015	0.013	0.001	0.001	0.000	0.000	0.000
Proportion	0.530	0.245	0.070	0.060	0.005	0.003	0.001	0.000	0.000
Cumulative	0.530	0.775	0.845	0.905	0.910	0.913	0.914	0.915	0.915

$$\lambda_1^{adj} = \left(\frac{Q}{Q-1} \left(\sqrt{\lambda_k} - \frac{1}{Q} \right) \right)^2 = \left(\frac{8}{7} (\sqrt{0.18} - 0.125) \right)^2 = 0.117$$

MCA using supplementary demographic information



MCA with R

```
library(ca)
X <- read.csv("http://www-eio.upc.es/~jan/Data/women_Spain2002_original.csv",header=TRUE,sep=";")
X[X==9] <- NA

indmis <- NULL
for(i in 1:nrow(X)) {
  indmis <- c(indmis,any(is.na(X[i,])))
}

X <- X[!indmis,]

X <- X[,1:8]

#
# Combine categories (disagree, strongly disagree for question H)
#

X$H[X$H==5] <- 4

out <- mjca(X[,1:8],lambda="indicator")
plot(out,labels=c(0,1),col=c("red","red"),pch=c(19,19,24,24),main="MCA biplot of Indicator matrix")
```

References

- Benzécri, J. P. (1973) *Analyse des Données*, Dunod, Paris.
- Cuadras, C. (2008) *Nuevos métodos de Análisis Multivariante*. Chapter 9. Download
- Greenacre, M. J. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press.
- Greenacre, M. J. (1993), *Correspondence Analysis in Practice*, Academic Press.
- Greenacre, M. J. (2008), *La práctica del análisis de correspondencias*, Fundación BBVA. Download