Multiple categorical variables

Multiple Correspondence Analysis

Data set "Science&Environment" (ISSP,1993)

Q.4 SCIENCE AND ENVIRONMENT

How much do you agree or disagree with each of these statements?

Q.4a We believe too often in science, and not enough in feelings and faith.
Q.4b Over all, modern science does more harm than good.
Q.4c Any change humans cause in nature - no matter how scientific - is likely to make things worse.
Q.4d Modern science will solve our environmental problems with little change to our way of life.

Response categories

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree
8. Can't choose, don't know
9. NA, refused

We are interested now in the relationship between the four variables, not so much as the differences between countries. Since the relationship between the four variables might change across the countries, we shall restrict our attention for the moment to one country, say Germany (dataset **wg93** in the **ca** package). Missing values have been removed in this initial example of MCA. The sample size is n = 871.

Indicator matrix

Original responses (Q = 4 questions)

4a	4b	4c	4d
2	3	4	3
3	4	2	3
2	3	2	4
2	2	2	2
3	3	3	3
3	4	4	5
3	4	2	4
3	4	4	2
3	2	2	1
3	3	2	2
.	.	.	.
.	.	.	.
etc.	.	.	

Indicator matrix (J = 20 categories)

A					B					C					D				
1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0
0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0
0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
etc. . (871 rows)																			

Burt matrix

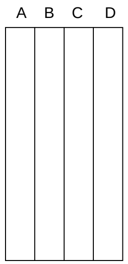
	V4a(A)					V4b(B)					V4c(C)					V4d(D)				
V4a	119	0	0	0	0	27	28	30	22	12	49	40	18	7	5	15	25	17	34	28
	0	322	0	0	0	38	74	84	96	30	67	142	60	41	12	22	102	76	68	54
	0	0	204	0	0	3	48	63	73	17	18	75	70	34	7	10	44	68	58	24
	0	0	0	178	0	3	21	23	79	52	16	50	40	56	16	9	52	28	54	35
	0	0	0	0	48	0	3	5	11	29	2	9	9	16	12	4	9	13	12	10
V4b	27	38	3	3	0	71	0	0	0	0	43	19	4	3	2	9	17	10	10	25
	28	74	48	21	3	0	174	0	0	0	36	88	34	15	1	16	51	42	45	20
	30	84	63	23	5	0	0	205	0	0	37	90	57	19	2	10	53	63	51	28
	22	96	73	79	11	0	0	0	281	0	27	88	75	74	17	6	66	70	92	47
	12	30	17	52	29	0	0	0	0	140	9	31	27	43	30	19	45	17	28	31
V4c	49	67	18	16	2	43	36	37	27	9	152	0	0	0	0	25	24	15	38	50
	40	142	75	50	9	19	88	90	88	31	0	316	0	0	0	15	97	67	89	48
	18	60	70	40	9	4	34	57	75	27	0	0	197	0	0	5	51	83	41	17
	7	41	34	56	16	3	15	19	74	43	0	0	0	154	0	6	44	30	51	23
	5	12	7	16	12	2	1	2	17	30	0	0	0	0	52	9	16	7	7	13
V4d	15	22	10	9	4	9	16	10	6	19	25	15	5	6	9	60	0	0	0	0
	25	102	44	52	9	17	51	53	66	45	24	97	51	44	16	0	232	0	0	0
	17	76	68	28	13	10	42	63	70	17	15	67	83	30	7	0	0	202	0	0
	34	68	58	54	12	10	45	51	92	28	38	89	41	51	7	0	0	0	226	0
	28	54	24	35	10	25	20	28	47	31	50	48	17	23	13	0	0	0	0	151

B can be thought of as the "covariance matrix" between the 4 variables

Two equivalent forms of MCA

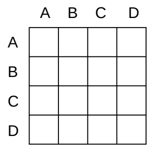
Indicator

$\mathbf{Z}$ = $(n \times J)$



Burt

$\mathbf{B} = \mathbf{Z}^\mathsf{T}\mathbf{Z}$ = $(J \times J)$

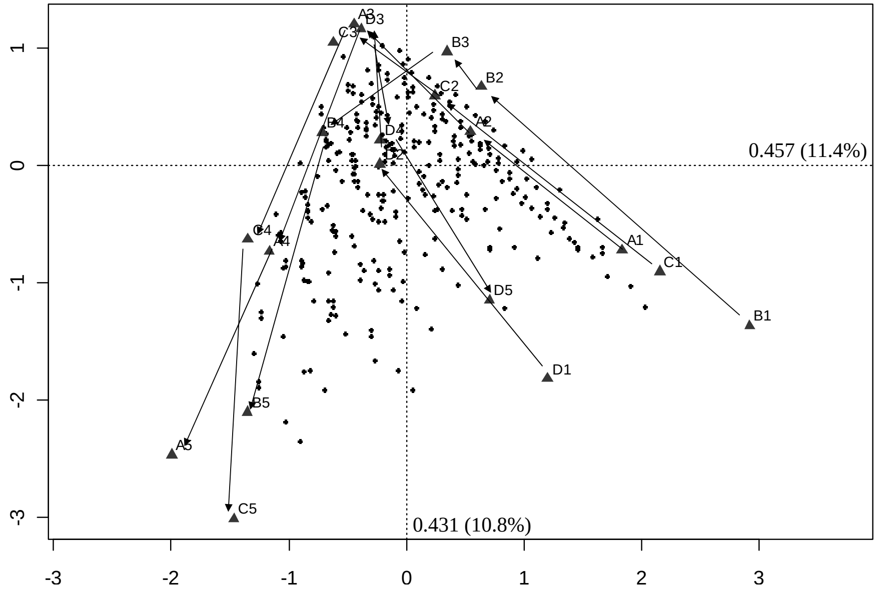


- CA of $\mathbf{Z}$
- Column standard coordinates Γ
- Row profiles are 0s with $1/Q$ = ¼ in the response positions; mass of each row is $1/n$.
- In asymmetric map the rows (cases) are at ordinary averages of their reponses.
- Principal inertias μ_1, μ_2, …

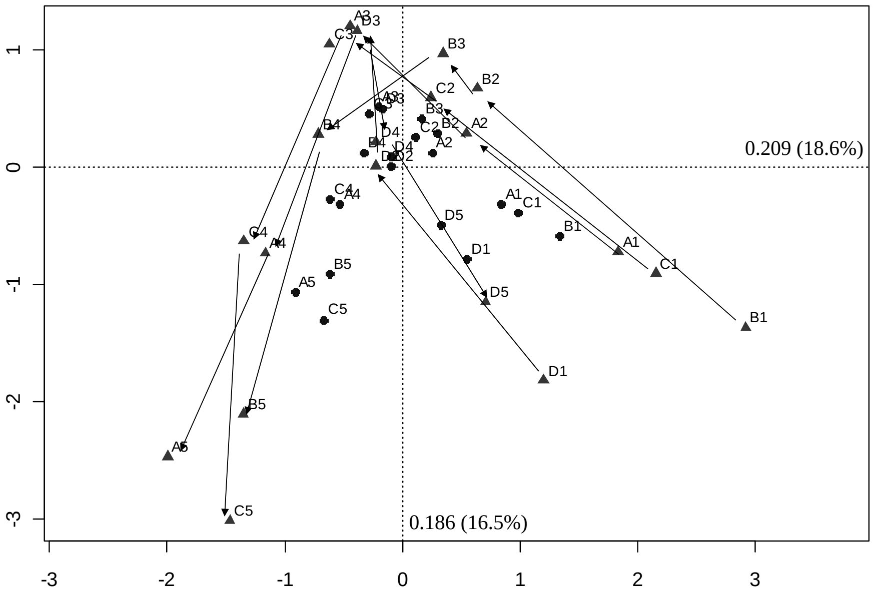
- CA of $\mathbf{B}$
- Column standard coordinates Γ
- Row standard coordinates Φ also Γ
- Row profiles are $1/Q$ = ¼ times the "2-way" profiles, but including a "diagonal profile" (which is a vertex point); masses proportional to marginal frequencies…
- Principal inertias $\lambda_1 = \mu_1^2$, $\lambda_2 = \mu_2^2$, …

In both versions, the percentages of inertia are very low (less so for $\mathbf{B}$).

CA of indicator matrix



CA of Burt matrix



R code for MCA

```
# use wg93 data set on science & environment
data(wg93)

# CA of idicator matrix using mjca function
ca.wg93.Z <- mjca(wg93[,1:4], lambda="indicator")
plot(ca.wg93.Z, labels=c(0,2), map="rowprincipal")

# inertias and percentages of inertia
ca.wg93.Z$sv^2
100*ca.wg93.Z$sv^2/sum(ca.wg93.Z$sv^2)

# Burt matrix
wg93.B <- mjca(wg93)$Burt

# CA of Burt matrix using ca function
ca.wg93.B <- ca(wg93.B[1:20,1:20])
ca.wg93.B$rowcoord <- - ca.wg93.B$rowcoord
ca.wg93.B$colcoord <- - ca.wg93.B$colcoord
plot(ca.wg93.B, map="rowprincipal")

# inertias and percentages of inertia
ca.wg93.B$sv^2
100*ca.wg93.B$sv^2/sum(ca.wg93.B$sv^2)
```

Burt matrix – subtable inertias

	V4a(A)	V4b(B)	V4c(C)	V4d(D)
V4a	119 0 0 0 0 0 322 0 0 0 0 0 204 0 0 0 0 0 178 0 0 0 0 0 48 **4**	**0.238**	**0.194**	**0.053**
V4b	27 38 3 3 0 28 74 48 21 3 30 84 63 23 5 22 96 73 79 11 12 30 17 52 29	71 0 0 0 0 0 174 0 0 0 0 0 205 0 0 0 0 0 281 0 0 0 0 0 140 **4**	**0.297**	**0.085**
V4c	49 67 18 16 2 40 142 75 50 9 18 60 70 40 9 7 41 34 56 16 5 12 7 16 12	43 36 37 27 9 19 88 90 88 31 4 34 57 75 27 3 15 19 74 43 2 1 2 17 30	152 0 0 0 0 0 316 0 0 0 0 0 197 0 0 0 0 0 154 0 0 0 0 0 52 **4**	**0.155**
V4d	15 22 10 9 4 25 102 44 52 9 17 76 68 28 13 34 68 58 54 12 28 54 24 35 10	9 16 10 6 19 17 51 53 66 45 10 42 63 70 17 10 45 51 92 28 25 20 28 47 31	25 15 5 6 9 24 97 51 44 16 15 67 83 30 7 38 89 41 51 7 50 48 17 23 13	60 0 0 0 0 0 232 0 0 0 0 0 202 0 0 0 0 0 226 0 0 0 0 0 151 **4**

Inertia of **B** is the average of the 16 inertias: 1.128

Inertia of off-diagonal blocks of Burt matrix is the average of the 12 inertias: 0.170

Adjustment of principal inertias (eigenvalues)

We can rescale an existing MCA solution quite simply in order to best fit the off-diagonal tables. All we need is the total inertia of the Burt matrix, *inertia*(**B**), and the principal inertias λ_k of the Burt matrix in the solution space.

If the solution has been computed on the indicator matrix **Z**, the eigenvalues calculated are $\lambda_k^{½}$ so all the squares of the principal inertias of **Z** need to be summed in order to get *inertia*(**B**). If the Burt matrix **B** has been analysed, the λ_k's are available and *inertia*(**B**) is the total inertia, the sum of the λ_k's.

Here are the steps to rescale the solution:

1. Calculate the average off-diagonal inertia :

   $$\text{average off-diagonal inertia} = \frac{Q}{Q-1}\left(inertia\,(\mathbf{B}) - \frac{J-Q}{Q^2}\right)$$

2. Calculate the adjusted principal inertias :

   $$\text{adjusted principal inertias} = \left(\frac{Q}{Q-1}\right)^2\left(\lambda_k^{1/2} - \frac{1}{Q}\right)^2 \quad \text{only for } \lambda_k^{1/2} > \frac{1}{Q}$$

3. Calculate adjusted percentages of inertia :

   $$\text{adjusted percentages of inertia} = \frac{\text{adjusted principal inertias}}{\text{average off-diagonal inertia}}$$

Adjustment of inertias: wg93 data

average off-diagonal inertia $= (4/3)\,(1.1277 - 16/16) = 0.1703$

adjusted principal inertias in first two dimensions

$= (4/3)^2(0.4574 - ¼)^2 = 0.07647$ (dimension 1)

$= (4/3)^2(0.4310 - ¼)^2 = 0.05824$ (dimension 2)

adjusted percentages of inertia

$= 0.07647 / 0.1703 = 0.4490$ i.e. 44.9 %

$= 0.05824 / 0.1703 = 0.3420$ i.e. 34.2 %

Here are the steps to rescale the solution:

1. Calculate the average off-diagonal inertia :

   $$\text{average off-diagonal inertia} = \frac{Q}{Q-1}\left(inertia\,(\mathbf{B}) - \frac{J-Q}{Q^2}\right)$$

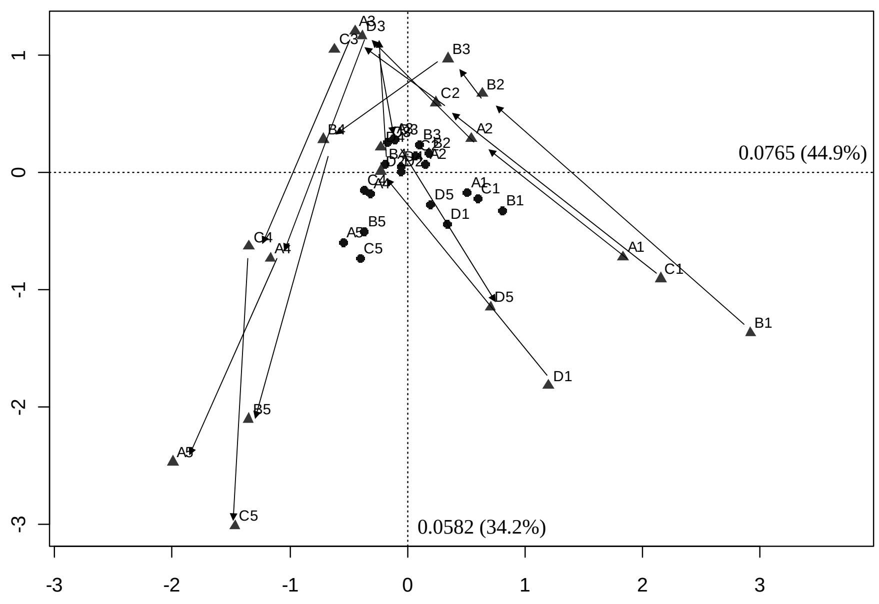
2. Calculate the adjusted principal inertias :

   $$\text{adjusted principal inertias} = \left(\frac{Q}{Q-1}\right)^2\left(\lambda_k^{1/2} - \frac{1}{Q}\right)^2 \quad \text{only for } \lambda_k^{1/2} > \frac{1}{Q}$$

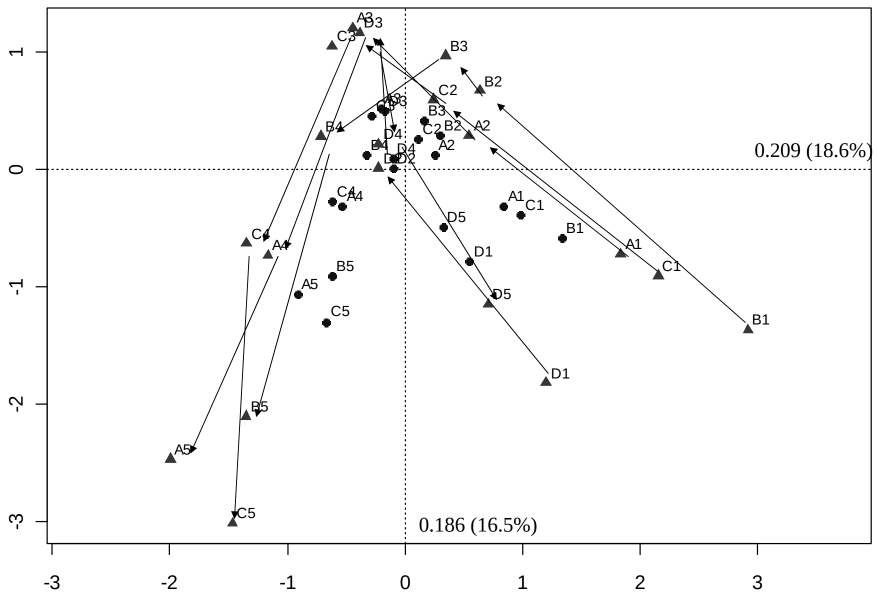
3. Calculate adjusted percentages of inertia :

   $$\text{adjusted percentages of inertia} = \frac{\text{adjusted principal inertias}}{\text{average off-diagonal inertia}}$$

Analysis with adjusted eigenvalues



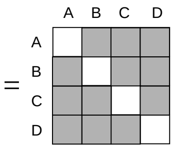
CA of Burt matrix



Joint Correspondence Analysis

Burt

$$\mathbf{B} = \mathbf{Z}^{\mathsf{T}}\mathbf{Z} = \quad (J \times J)$$



Adjusted MCA:

- Do CA of $\mathbf{B}$ and then adjust the coordinates so that the off-diagonal ▢ are optimally fitted.

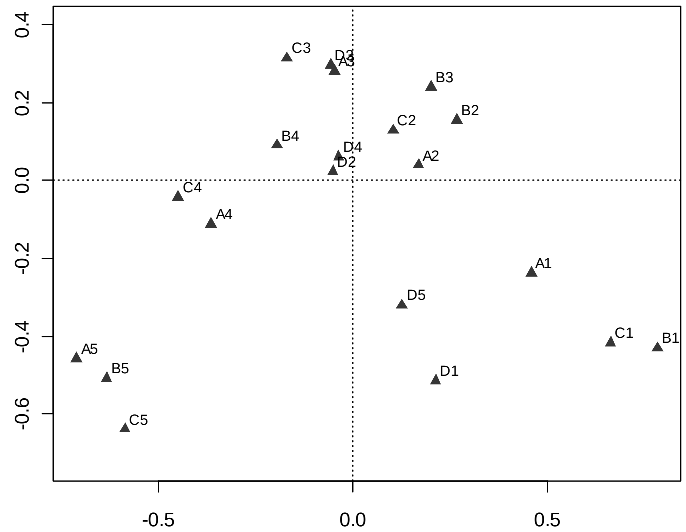
Joint CA (JCA)

- Fit the off-diagonal ▢ optimally.
- This is a different algorithm, performed iteratively.
- Advantage: maximum inertia in ▢ is explained.
- Disadvantages: axes are not nested, scale optimality degraded.

R output for JCA

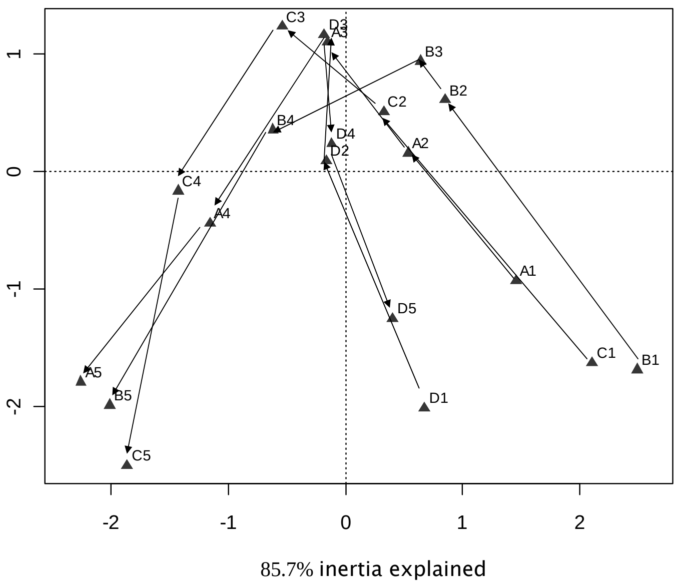
```
> summary(mjca(wg93[,1:4], lambda="JCA")

Principal inertias (eigenvalues):

     dim    value
[1,] 1      0.099091
[2,] 2      0.065033
[3,] 3      0.005459
[4,] 4      0.003345
[5,] 5      0.001176
[6,] 6      0.000207
[7,]        --------
[8,] Total: 0.182425

 Diagonal inertia discounted from eigenvalues: 0.0547405
 Percentage explained by JCA in 2 dimensions: 85.7%
 (Eigenvalues are not nested)
 [Iterations in JCA: 44 , epsilon = 9.91e-05]
```

Joint Correspondence Analysis

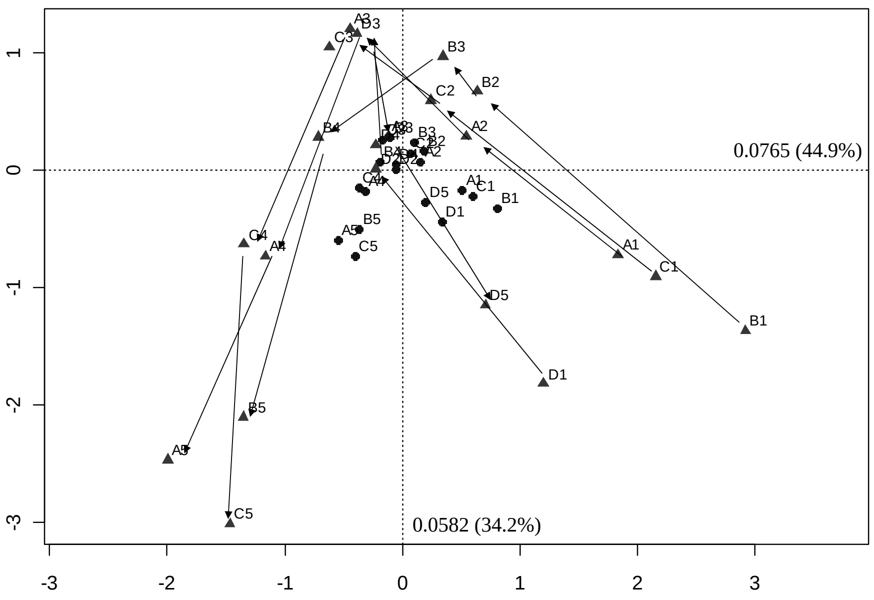


- % of inertia explained by axes 1 & 2: 85.7% (79.1% before, with adjustments)
- Axes are not “nested”, as in CA and MCA.

Joint Correspondence Analysis



Analysis with adjusted eigenvalues



Quantification as a goal – Homogeneity Analysis (HOMALS)

Original responses

```
2 2 1 2
2 2 2 5
4 3 2 5
2 5 4 2
4 2 1 5
1 4 1 5
1 2 2 3
1 3 2 4
3 2 2 4
3 5 5 2
. . . .
. . . .
```

→

Quantified responses

```
a2 b2 c1 d2
a2 b2 c2 d5
a4 b3 c2 d5
a2 b5 c4 d2
a4 b2 c1 d5
a1 b4 c1 d5
a1 b2 c2 d3
a1 b3 c2 d4
a3 b2 c2 d4
a3 b5 c5 d2
.  .  .  .
.  .  .  .
```

→

Scores (averaged or summated

```
(a2+b2+c1+d2)/4
(a2+b2+c2+d5)/4
(a4+b3+c2+d5)/4
(a2+b5+c4+d2)/4
(a4+b2+c1+d5)/4
(a1+b4+c1+d5)/4
(a1+b2+c2+d3)/4
(a1+b3+c2+d4)/4
(a3+b2+c2+d4)/4
(a3+b5+c5+d2)/4
       .
       .
```

Objective of HOMALS: to determine the set of J scale values **a1,a2,...,** **b1,b2,...,** etc... so that the implied scores for each individual are as "close" as possible to that individual's particular set of Q scale values.

"Closeness" is defined in terms of squared sum of differences, and the solution is obtained by least-squares – this is mathematically equivalent to maximising the sum of squared correlations between the scores and the quantified responses.

Scaling properties of MCA

The category quantifications in MCA maximize the average squared correlation between items scores and summated scores:

- Principal inertia of indicator matrix (square root of principal inertia of Burt matrix) is this average squared correlation:

$$\lambda_1^{½} = ¼\,[\mathrm{cor}^2(A,A+B+C+D) + \mathrm{cor}^2(B,A+B+C+D) + \mathrm{cor}^2(C,A+B+C+D) + \mathrm{cor}^2(D,A+B+C+D)\,]$$

The category quantifications in MCA maximize the Cronbach α reliability coefficient

$$\alpha = \frac{Q-1}{Q}\left(1-\frac{1}{Q\lambda_1^{1/2}}\right)$$

= (4/3)[1−1/(4×0.4574)]

= 0.605

(dropping variable D) = (4/3)[1−1/(4×0.6018)]

= 0.779