Correspondence analysis

David J. Hessen

Utrecht University

Academic year 2022-2023

Frequency data

Two nominal features

The first feature has I levels and the second feature has J levels

An $I \times J$ contingency table (cross tabulation or cross tab)

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1J} \\ x_{21} & x_{22} & \dots & x_{2J} \\ \vdots & \vdots & & \vdots \\ x_{I1} & x_{I2} & \dots & x_{IJ} \end{bmatrix}$$

where x_{ij} is the joint frequency of levels i and j, for all i and j

The total frequency is $N = x_{11} + \ldots + x_{IJ}$

Frequency data

Example: Nobel prize data

	Chemistry	Economics	Literature	Medicine	Peace	Physics
Canada	4	3	2	4	1	4
France	8	3	11	12	10	9
Germany	24	1	8	18	5	24
Italy	1	1	6	5	1	5
Japan	6	0	2	3	1	11
Russia	4	3	5	2	3	10
UK	23	6	7	26	11	20
USA	51	43	8	70	19	66

$I = 8$, $J = 6$, and $N = 570$

Frequency data

Example: Nobel prize data

8×6 matrix

$$\mathbf{X} = \begin{bmatrix} 4 & 3 & 2 & 4 & 1 & 4 \\ 8 & 3 & 11 & 12 & 10 & 9 \\ 24 & 1 & 8 & 18 & 5 & 24 \\ 1 & 1 & 6 & 5 & 1 & 5 \\ 6 & 0 & 2 & 3 & 1 & 11 \\ 4 & 3 & 5 & 2 & 3 & 10 \\ 23 & 6 & 7 & 26 & 11 & 20 \\ 51 & 43 & 8 & 70 & 19 & 66 \end{bmatrix}$$

Joint proportions

The $I \times J$ matrix of joint relative frequencies

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1J} \\ z_{21} & z_{22} & \cdots & z_{2J} \\ \vdots & \vdots & & \vdots \\ z_{I1} & z_{I2} & \cdots & z_{IJ} \end{bmatrix}$$

where $z_{ij} = x_{ij}/N$, for all i and j, so

$$\mathbf{Z} = \mathbf{X}/N$$

$\mathbf{Z}$ is called the *correspondence matrix*

Joint proportions

Example: Nobel prize data

The correspondence matrix

$$\mathbf{Z} = \begin{bmatrix} .007 & .005 & .004 & .007 & .002 & .007 \\ .014 & .005 & .019 & .021 & .018 & .016 \\ .042 & .002 & .014 & .032 & .009 & .042 \\ .002 & .002 & .011 & .009 & .002 & .009 \\ .011 & .000 & .004 & .005 & .002 & .019 \\ .007 & .005 & .009 & .004 & .005 & .018 \\ .040 & .011 & .012 & .046 & .019 & .035 \\ .089 & .075 & .014 & .123 & .033 & .116 \end{bmatrix}$$

Marginal proportions

The vector of row totals of $\mathbf{Z}$ (row marginal proportions)

$$\mathbf{r} = [\, r_1 \; \ldots \; r_I \,]^T$$

where $r_i = z_{i1} + z_{i2} + \ldots + z_{iJ} = (x_{i1} + x_{i2} + \ldots + x_{iJ})/N$, for all i

The vector of column totals of $\mathbf{Z}$ (column marginal proportions)

$$\mathbf{c} = [\, c_1 \; \ldots \; c_J \,]^T$$

where $c_j = z_{1j} + z_{2j} + \ldots + z_{Ij} = (x_{1j} + x_{2j} + \ldots + x_{Ij})/N$, for all j

The elements of $\mathbf{r}$ are called *row masses*

The elements of $\mathbf{c}$ are called *column masses*

Marginal proportions

Example: Nobel prize data

$$\mathbf{Z} = \begin{bmatrix} .007 & .005 & .004 & .007 & .002 & .007 \\ .014 & .005 & .019 & .021 & .018 & .016 \\ .042 & .002 & .014 & .032 & .009 & .042 \\ .002 & .002 & .011 & .009 & .002 & .009 \\ .011 & .000 & .004 & .005 & .002 & .019 \\ .007 & .005 & .009 & .004 & .005 & .018 \\ .040 & .011 & .012 & .046 & .019 & .035 \\ .089 & .075 & .014 & .123 & .033 & .116 \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} .032 \\ .093 \\ .140 \\ .033 \\ .040 \\ .047 \\ .163 \\ .451 \end{bmatrix}$$

$$\mathbf{c}^T = \begin{bmatrix} .212 & .105 & .086 & .246 & .089 & .261 \end{bmatrix}$$

Independence

Expected joint proportions

$$\mathbf{rc}^T = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_I \end{bmatrix} \begin{bmatrix} c_1 & c_2 & \dots & c_J \end{bmatrix} = \begin{bmatrix} r_1c_1 & r_1c_2 & \dots & r_1c_J \\ r_2c_1 & r_2c_2 & \dots & r_2c_J \\ \vdots & \vdots & & \vdots \\ r_Ic_1 & r_Ic_2 & \dots & r_Ic_J \end{bmatrix}$$

So under perfect independence, $\mathbf{Z} = \mathbf{rc}^T$ or $\mathbf{Z} - \mathbf{rc}^T = \mathbf{0}$

$\mathbf{Z} - \mathbf{rc}^T$ is the matrix of residuals

Independence

Example: Nobel prize data

Expected joint proportions

$$\mathbf{rc}^T = \begin{bmatrix} .032 \\ .093 \\ .140 \\ .033 \\ .040 \\ .047 \\ .163 \\ .451 \end{bmatrix} \begin{bmatrix} .212 & .105 & .086 & .246 & .089 & .261 \end{bmatrix}$$

Independence

Example: Nobel prize data

Expected joint proportions

$$\mathbf{rc}^T = \begin{bmatrix} .007 & .003 & .003 & .008 & .003 & .008 \\ .020 & .010 & .008 & .023 & .008 & .024 \\ .030 & .015 & .012 & .034 & .013 & .037 \\ .007 & .004 & .003 & .008 & .003 & .009 \\ .009 & .004 & .003 & .010 & .004 & .011 \\ .010 & .005 & .004 & .012 & .004 & .012 \\ .035 & .017 & .014 & .040 & .015 & .043 \\ .096 & .047 & .039 & .111 & .040 & .118 \end{bmatrix}$$

Independence

Example: Nobel prize data

Matrix of residuals

$$
\mathbf{Z} - \mathbf{r}\mathbf{c}^T = \begin{bmatrix} .000 & .002 & .001 & -.001 & -.001 & -.001 \\ -.006 & -.005 & .011 & -.002 & .009 & -.009 \\ .012 & -.013 & .002 & -.003 & -.004 & .005 \\ -.005 & -.002 & .008 & .001 & -.001 & .000 \\ .002 & -.004 & .000 & -.005 & -.002 & .009 \\ -.003 & .000 & .005 & -.008 & .001 & .005 \\ .006 & -.007 & -.002 & .006 & .005 & -.008 \\ -.006 & .028 & -.025 & .012 & -.007 & -.002 \end{bmatrix}
$$

Conditional proportions

Let $\mathbf{D}_r = \text{diag}\{r_1, r_2, \ldots, r_I\}$ so that $\mathbf{D}_r^{-1} = \text{diag}\left\{\frac{1}{r_1}, \frac{1}{r_2}, \ldots, \frac{1}{r_I}\right\}$

Matrix of row profiles

$$\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{Z} = \begin{bmatrix} \frac{z_{11}}{r_1} & \frac{z_{12}}{r_1} & \ldots & \frac{z_{1J}}{r_1} \\ \frac{z_{21}}{r_2} & \frac{z_{22}}{r_2} & \ldots & \frac{z_{2J}}{r_2} \\ \vdots & \vdots & & \vdots \\ \frac{z_{I1}}{r_I} & \frac{z_{I2}}{r_I} & \ldots & \frac{z_{IJ}}{r_I} \end{bmatrix}$$

Under perfect independence, $\mathbf{R} = \mathbf{1}\mathbf{c}^T$ or $\mathbf{R} - \mathbf{1}\mathbf{c}^T = \mathbf{0}$

Conditional proportions

Example: Nobel prize data

Matrix of row profiles (for countries)

$$\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{Z} = \begin{bmatrix} .222 & .167 & .111 & .222 & .056 & .222 \\ .151 & .057 & .208 & .226 & .189 & .170 \\ .300 & .012 & .100 & .225 & .062 & .300 \\ .053 & .053 & .316 & .263 & .053 & .263 \\ .261 & .000 & .087 & .130 & .043 & .478 \\ .148 & .111 & .185 & .074 & .111 & .370 \\ .247 & .065 & .075 & .280 & .118 & .215 \\ .198 & .167 & .031 & .272 & .074 & .257 \end{bmatrix}$$

'Average' row profile

$$\mathbf{c}^T = \begin{bmatrix} .212 & .105 & .086 & .246 & .089 & .261 \end{bmatrix}$$

Conditional proportions

Let $\mathbf{D}_c = \text{diag}\{c_1, c_2, \ldots, c_J\}$ so that $\mathbf{D}_c^{-1} = \text{diag}\left\{\frac{1}{c_1}, \frac{1}{c_2}, \ldots, \frac{1}{c_J}\right\}$

Matrix of column profiles

$$\mathbf{C} = \mathbf{D}_c^{-1}\mathbf{Z}^T = \begin{bmatrix} \frac{z_{11}}{c_1} & \frac{z_{21}}{c_1} & \cdots & \frac{z_{I1}}{c_1} \\ \frac{z_{12}}{c_2} & \frac{z_{22}}{c_2} & \cdots & \frac{z_{I2}}{c_2} \\ \vdots & \vdots & & \vdots \\ \frac{z_{1J}}{c_J} & \frac{z_{2J}}{c_J} & \cdots & \frac{z_{IJ}}{c_J} \end{bmatrix}$$

Under perfect independence, $\mathbf{C} = \mathbf{1}\mathbf{r}^T$ or $\mathbf{C} - \mathbf{1}\mathbf{r}^T = \mathbf{0}$

Conditional proportions

Example: Nobel prize data

Matrix of column profiles (for Nobel prize categories)

$$\mathbf{C} = \mathbf{D}_c^{-1}\mathbf{Z}^T = \begin{bmatrix} .033 & .066 & .198 & .008 & .050 & .033 & .190 & .421 \\ .050 & .050 & .017 & .017 & .000 & .050 & .100 & .717 \\ .041 & .224 & .163 & .122 & .041 & .102 & .143 & .163 \\ .029 & .086 & .129 & .036 & .021 & .014 & .186 & .500 \\ .020 & .196 & .098 & .020 & .020 & .059 & .216 & .373 \\ .027 & .060 & .161 & .034 & .074 & .067 & .134 & .443 \end{bmatrix}$$

'Average' column profile

$$\mathbf{r}^T = \begin{bmatrix} .032 & .093 & .140 & .033 & .040 & .047 & .163 & .451 \end{bmatrix}$$

Generalized singular value decomposition

GSVD of the $I \times J$ matrix of *residuals*

$$\mathbf{Z} - \mathbf{r}\mathbf{c}^T = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^T$$

where

- $\mathbf{P}^T\mathbf{D}_r^{-1}\mathbf{P} = \mathbf{I}$
- $\mathbf{Q}^T\mathbf{D}_c^{-1}\mathbf{Q} = \mathbf{I}$
- $\mathbf{P}$ is $I \times \min(I, J)$
- $\boldsymbol{\Delta} = \text{diag}\{\delta_1, \ldots, \delta_{\min(I,J)}\}$ and has rank $\min(I-1, J-1)$
- $\mathbf{Q}$ is $J \times \min(I, J)$

Singular value decomposition

SVD of the $I \times J$ matrix of *standardized residuals*

$$\mathbf{D}_r^{-1/2}\big(\mathbf{Z} - \mathbf{r}\mathbf{c}^T\big)\mathbf{D}_c^{-1/2} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T$$

where

- $\mathbf{D}_r^{-1/2} = \text{diag}\Big\{\frac{1}{\sqrt{r_1}}, \frac{1}{\sqrt{r_2}}, \ldots, \frac{1}{\sqrt{r_I}}\Big\}$
- $\mathbf{D}_c^{-1/2} = \text{diag}\Big\{\frac{1}{\sqrt{c_1}}, \frac{1}{\sqrt{c_2}}, \ldots, \frac{1}{\sqrt{c_J}}\Big\}$
- $\mathbf{U} = \mathbf{D}_r^{-1/2}\mathbf{P}$ is $I \times \min(I, J)$ and $\mathbf{U}^T\mathbf{U} = \mathbf{I}$
- $\mathbf{V} = \mathbf{D}_c^{-1/2}\mathbf{Q}$ is $J \times \min(I, J)$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$

Singular value decomposition

Example: Nobel prize data

Matrix of standardized residuals

$$\mathbf{D}_r^{-1/2}\left(\mathbf{Z}-\mathbf{r}\mathbf{c}^T\right)\mathbf{D}_c^{-1/2}=\begin{bmatrix} .004 & .034 & .015 & -.008 & -.020 & -.014 \\ -.041 & -.046 & .126 & -.012 & .101 & -.055 \\ .071 & -.107 & .018 & -.016 & -.034 & .028 \\ -.063 & -.030 & .143 & .006 & -.022 & .001 \\ .021 & -.065 & .001 & -.047 & -.031 & .085 \\ -.030 & .004 & .074 & -.075 & .016 & .046 \\ .031 & -.051 & -.015 & .028 & .039 & -.037 \\ -.020 & .128 & -.126 & .036 & -.035 & -.006 \end{bmatrix}$$

Singular value decomposition

Example: Nobel prize data

Matrix of left singular vectors

$$\mathbf{U} = \begin{bmatrix} -.029 & .104 & -.135 & .139 & .731 & -.220 \\ .526 & .431 & .322 & -.318 & -.052 & .438 \\ .206 & -.619 & .156 & .240 & .344 & .398 \\ .451 & .250 & -.350 & .700 & -.250 & -.135 \\ .122 & -.535 & -.333 & -.212 & -.437 & -.019 \\ .269 & .055 & -.525 & -.534 & .220 & -.256 \\ .050 & -.085 & .562 & -.028 & -.144 & -.676 \\ -.622 & .248 & -.170 & .037 & -.148 & .246 \end{bmatrix}$$

Singular value decomposition

Example: Nobel prize data

Matrix of singular values

$$\boldsymbol{\Delta} = \begin{bmatrix} .289 & .000 & .000 & .000 & .000 & .000 \\ .000 & .194 & .000 & .000 & .000 & .000 \\ .000 & .000 & .147 & .000 & .000 & .000 \\ .000 & .000 & .000 & .089 & .000 & .000 \\ .000 & .000 & .000 & .000 & .044 & .000 \\ .000 & .000 & .000 & .000 & .000 & .000 \end{bmatrix}$$

Singular value decomposition

Example: Nobel prize data

Matrix of right singular vectors

$$\mathbf{V} = \begin{bmatrix} -.093 & -.505 & .334 & -.041 & .641 & .461 \\ -.519 & .589 & -.383 & -.105 & .348 & .324 \\ .802 & .282 & -.234 & .256 & .266 & .293 \\ -.185 & .170 & .388 & .608 & -.413 & .496 \\ .211 & .321 & .460 & -.710 & -.217 & .299 \\ .009 & -.433 & -.570 & -.217 & -.424 & .511 \end{bmatrix}$$

Factor scores

The matrix of *principal coordinates* for the rows is given by

$$\mathbf{F} = \mathbf{D}_r^{-1}\mathbf{P}\boldsymbol{\Delta} = \mathbf{D}_r^{-1}(\mathbf{D}_r^{1/2}\mathbf{U})\boldsymbol{\Delta} = \mathbf{D}_r^{-1/2}\mathbf{U}\boldsymbol{\Delta}$$

The matrix of *principal coordinates* for the columns is given by

$$\mathbf{G} = \mathbf{D}_c^{-1}\mathbf{Q}\boldsymbol{\Delta} = \mathbf{D}_c^{-1}(\mathbf{D}_c^{1/2}\mathbf{V})\boldsymbol{\Delta} = \mathbf{D}_c^{-1/2}\mathbf{V}\boldsymbol{\Delta}$$

The joint plot in principal coordinates, $\mathbf{F}$ and $\mathbf{G}$, is called the *symmetric map* because both row and column profiles are overlaid in the same coordinate system

Factor scores

Example: Nobel prize data

Matrix of principal coordinates for the rows

$$\mathbf{F} = \begin{bmatrix} -.047 & .113 & -.112 & .070 & .179 & .000 \\ .498 & .273 & .155 & -.093 & -.007 & .000 \\ .158 & -.320 & .061 & .057 & .040 & .000 \\ .713 & .265 & -.282 & .340 & -.060 & .000 \\ .176 & -.516 & -.244 & -.094 & -.095 & .000 \\ .356 & .049 & -.355 & -.218 & .044 & .000 \\ .036 & -.041 & .205 & -.006 & -.015 & .000 \\ -.267 & .071 & -.037 & .005 & -.010 & .000 \end{bmatrix}$$

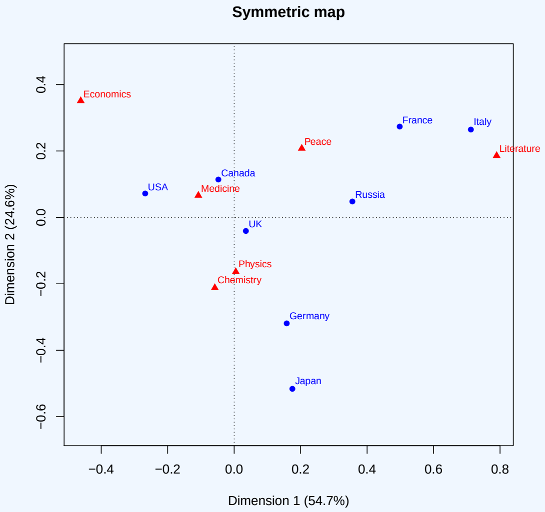
Factor scores

Example: Nobel prize data

Matrix of principal coordinates for the columns

$$\mathbf{G} = \begin{bmatrix} -.058 & -.212 & .107 & -.008 & .061 & .000 \\ -.462 & .351 & -.174 & -.029 & .047 & .000 \\ .790 & .186 & -.117 & .078 & .039 & .000 \\ -.108 & .066 & .115 & .109 & -.036 & .000 \\ .203 & .208 & .226 & -.211 & -.032 & .000 \\ .005 & -.164 & -.164 & -.038 & -.036 & .000 \end{bmatrix}$$

Example: Nobel prize data



Factor scores

The matrix of *standard coordinates* for the rows is given by

$$\mathbf{\Phi} = \mathbf{D}_r^{-1}\mathbf{P} = \mathbf{D}_r^{-1}(\mathbf{D}_r^{1/2}\mathbf{U}) = \mathbf{D}_r^{-1/2}\mathbf{U}$$

The matrix of *standard coordinates* for the columns is given by

$$\mathbf{\Gamma} = \mathbf{D}_c^{-1}\mathbf{Q} = \mathbf{D}_c^{-1}(\mathbf{D}_c^{1/2}\mathbf{V}) = \mathbf{D}_c^{-1/2}\mathbf{V}$$

The joint plot in principal coordinates $\mathbf{F}$ and standard coordinates $\mathbf{\Gamma}$ (or in standard coordinates $\mathbf{\Phi}$ and principal coordinates $\mathbf{G}$) is called an *asymmetric* map (a special biplot)

In an asymmetric map the distances between rows and columns can be interpreted meaningfully, that is, the distance from a row point to a column point reflects their association

Factor scores

Example: Nobel prize data

Matrix of standard coordinates for the rows

$$\mathbf{\Phi} = \begin{bmatrix} -.164 & .586 & -.762 & .784 & 4.114 & -1.239 \\ 1.726 & 1.412 & 1.056 & -1.043 & -.170 & 1.436 \\ .549 & -1.652 & .416 & .641 & .918 & 1.064 \\ 2.469 & 1.371 & -1.917 & 3.834 & -1.369 & -.741 \\ .608 & -2.665 & -1.658 & -1.055 & -2.177 & -.096 \\ 1.234 & .251 & -2.413 & -2.452 & 1.010 & -1.176 \\ .124 & -.211 & 1.390 & -.070 & -.356 & -1.673 \\ -.927 & .369 & -.253 & .055 & -.220 & .366 \end{bmatrix}$$

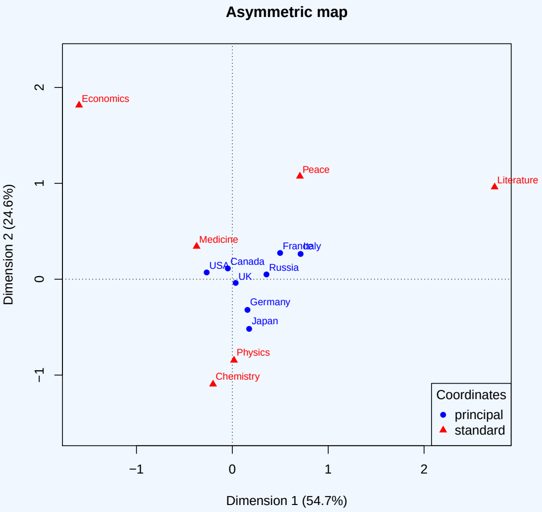
Factor scores

Example: Nobel prize data

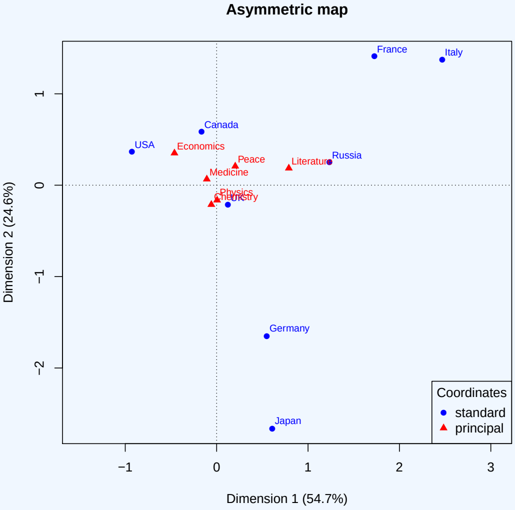
Matrix of standard coordinates for the columns

$$\boldsymbol{\Gamma} = \begin{bmatrix} -.201 & -1.095 & .726 & -.089 & 1.391 & 1.000 \\ -1.599 & 1.815 & -1.180 & -.325 & 1.073 & 1.000 \\ 2.736 & .961 & -.798 & .874 & .907 & 1.000 \\ -.373 & .342 & .783 & 1.228 & -.833 & 1.000 \\ .705 & 1.074 & 1.538 & -2.374 & -.726 & 1.000 \\ .017 & -.846 & -1.114 & -.425 & -.829 & 1.000 \end{bmatrix}$$

Example: Nobel prize data



Example: Nobel prize data



Total inertia explained

Total inertia

$$\mathcal{I} = \mathrm{tr}\left\{\mathbf{D}_r^{-1/2}(\mathbf{Z} - \mathbf{r}\mathbf{c}^T)\mathbf{D}_c^{-1}(\mathbf{Z} - \mathbf{r}\mathbf{c}^T)^T\mathbf{D}_r^{-1/2}\right\} = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(z_{ij} - r_i c_j)^2}{r_i c_j}$$

Note that $\mathcal{I} = X^2/N$, where X^2 is the cross tabs chi-square statistic

If follows that

$$\mathcal{I} = \mathrm{tr}\big(\mathbf{U}\boldsymbol{\Delta}^2\mathbf{U}^T\big) = \mathrm{tr}\big(\mathbf{V}\boldsymbol{\Delta}^2\mathbf{V}^T\big) = \mathrm{tr}\big(\boldsymbol{\Delta}^2\big) = \delta_1^2 + \ldots + \delta_{\min(I-1,J-1)}^2$$

where $\delta_1^2, \ldots, \delta_{\min(I-1,J-1)}^2$ are eigenvalues

The proportion of explained inertia by the first k dimensions is

$$\big(\delta_1^2 + \ldots + \delta_k^2\big)/\mathrm{tr}\big(\boldsymbol{\Delta}^2\big)$$

Total inertia explained

Example: Nobel prize data

Dimension	PIE	CPIE
1	.547	.547
2	.246	.793
3	.142	.936
4	.052	.988
5	.012	1.000

The indicator matrix

A dummy variable for each category

Example: working women

W = full-time, w = part-time, H = at home, ? = non-response

Questions				Question 1				Question 2				Question 3				Question 4			
1	2	3	4	W	w	H	?	W	w	H	?	W	w	H	?	W	w	H	?
1	3	2	2	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
2	3	3	2	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0
4	3	3	2	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0
4	4	4	4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
4	4	4	4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
1	3	2	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Let K be the number of variables and let m_k be the number of categories of variable k, then the number of dummies is $\sum_{k=1}^{K} m_k = M$

The total inertia is $\mathcal{I} = \sum_{k=1}^{K}(m_k - 1)/K = (M - K)/K$

The Burt matrix

The Burt matrix is given by $\mathbf{B} = \mathbf{E}^T\mathbf{E}$, where $\mathbf{E}$ is the indicator matrix

Example: working women

		Question 1				Question 2				Question 3				Question 4			
		W	w	H	?	W	w	H	?	W	w	H	?	W	w	H	?
Question 1	W	2501	0	0	0	172	1107	1131	91	355	1710	345	91	1766	538	40	157
	w	0	476	0	0	7	129	335	5	16	261	181	18	128	293	17	38
	H	0	0	79	0	1	6	72	0	1	17	61	0	14	21	38	6
	?	0	0	0	362	1	57	108	196	7	96	55	204	51	45	2	264
Question 2	W	172	7	1	1	181	0	0	0	127	48	4	2	165	15	0	1
	w	1107	129	6	57	0	1299	0	0	219	997	61	22	972	239	13	75
	H	1131	335	72	108	0	0	1646	0	24	989	573	60	760	616	84	186
	?	91	5	0	196	0	0	0	292	9	50	4	229	62	27	0	203
Question 3	W	355	16	1	7	127	219	24	9	379	0	0	0	360	14	1	4
	w	1710	261	17	96	48	997	989	50	0	2084	0	0	1348	567	23	146
	H	345	181	61	55	4	61	573	4	0	0	642	0	202	286	73	81
	?	91	18	0	204	2	22	60	229	0	0	0	313	49	30	0	234
Question 4	W	1766	128	14	51	165	972	760	62	360	1348	202	49	1959	0	0	0
	w	538	293	21	45	15	239	616	27	14	567	286	30	0	897	0	0
	H	40	17	38	2	0	13	84	0	1	23	73	0	0	0	97	0
	?	157	38	6	264	1	75	186	203	4	146	81	234	0	0	0	465