

MS-E2112 Multivariate Statistical Analysis (5cr)

Lecture 7: Multiple Correspondence Analysis

Lecturer: Pauliina Ilmonen
Slides: Ilmonen/Kantala

- Multiple Correspondence Analysis
- Frequency Tables
- Row Profiles
- Column Profiles
- Attraction Repulsion Indices
- Multiple Correspondence Analysis
- Graphical Presentation
- Example
- Some Remarks
- References

Contents

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple Correspondence Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion Indices

Multiple Correspondence Analysis

Graphical Presentation

Example

Some Remarks

References

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Multiple Correspondence Analysis

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Multiple Correspondence Analysis

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple correspondence analysis (MCA) is an extension of bivariate correspondence analysis to more than 2 variables.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Example: Gender, Civil Status and Education

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

In this lecture, we consider an example where we examine dependencies of categorical variables **gender**, **civil status** and **education**.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Frequency Tables

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Contingency Tables

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

We consider a sample of size n described by P qualitative variables Y_1, \dots, Y_P . The variable Y_p has K_p modalities (categories), and $\sum_{p=1}^P K_p$ is the total number of the categories. The number of individuals having the modality l of the variable Y_p is denoted by n_{pl} . We set a variable $x_{ipl} = 1$ if individual i has modality l of Y_p , and we set $x_{ipl} = 0$ otherwise. Now

$$\sum_{l=1}^{K_p} n_{pl} = n,$$

and

$$\sum_{p=1}^P \sum_{l=1}^{K_p} n_{pl} = nP.$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Table of Dummy Variables

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The table of K_p dummy variables associated with variable Y_p .

	1	2	...	K_p	
1	x_{1p1}	x_{1p2}	...	x_{1pK_p}	1
2	x_{2p1}	x_{2p2}	...	x_{2pK_p}	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	x_{np1}	x_{np2}	...	x_{npK_p}	1
	n_{p1}	n_{p2}	...	n_{pK_p}	n

Table: Table of dummy variables

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Complete Disjunctive Table

Now we introduce the $n \times K$ table/matrix $X = [X_1, \dots, X_P]$, called the **complete disjunctive table**.

	X_1			\dots	X_P			$\sum_{p=1}^P \sum_{l=1}^{K_p} X_{ipl}$
1	X_{111}	\dots	X_{11K_1}	\dots	X_{1P1}	\dots	X_{1PK_P}	P
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
i	X_{i11}	\dots	X_{i1K_1}	\dots	X_{iP1}	\dots	X_{iPK_P}	P
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	X_{n11}	\dots	X_{n1K_1}	\dots	X_{nP1}	\dots	X_{nPK_P}	P
$\sum_{i=1}^n X_{ipl}$	n_{11}	\dots	n_{1K_1}	\dots	n_{P1}	\dots	n_{PK_P}	nP

Table: Complete disjunctive table

Example: Gender, Civil Status and Education

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

We consider a sample of 4 individuals and 3 variables —
 $n = 4, P = 3$.

- Variable X_1 gender has two modalities/categories — male (1) and female (2).
- Variable X_2 civil status has three modalities — single (1), married (2), divorced/widowed (3).
- Variable X_3 education has two modalities — low education (1), at least high school diploma (2).

Now $K = K_1 + K_2 + K_3 = 2 + 3 + 2 = 7$.

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

Example: Gender, Civil Status and Education

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

We display the gender, civil status and education data as a complete disjunctive table.

	X_{11}	X_{12}	X_{21}	X_{22}	X_{23}	X_{31}	X_{32}	$\sum_{p=1}^7 \sum_{l=1}^{K_p} X_{ipl}$
1	0	1	1	0	0	1	0	3
2	0	1	1	0	0	0	1	3
3	1	0	0	0	1	1	0	3
4	0	1	0	1	0	0	1	3
$\sum_{i=1}^n X_{ipl}$	1	3	2	1	1	2	2	12

Table: Complete disjunctive table

- The first individual is female, single, and has low education.
- The third individual is male, divorced/widowed, and has low education.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles
Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Multiple Correspondence Analysis

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Bivariate correspondence analysis is now applied to the complete disjunctive table!

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Relative Frequency Tables

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

From the complete disjunctive table, it is straightforward to compute the associated relative frequency table (F), where the elements of the complete disjunctive table are divided by the total sum nP leading to

$$f_{ipl} = \frac{x_{ipl}}{nP} \quad (i = 1, \dots, n; p = 1, \dots, P; l = 1, \dots, K_p).$$

The marginal relative frequencies are computed as

$$f_{i..} = \frac{1}{n} \quad (i = 1, \dots, n) \text{ and } f_{.pl} = \frac{n_{pl}}{nP} \quad (p = 1, \dots, P; l = 1, \dots, K_p).$$

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

Example: Gender, Civil Status and Education

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

We display the gender, civil status and education data as a complete disjunctive table.

	X_{11}	X_{12}	X_{21}	X_{22}	X_{23}	X_{31}	X_{32}	$f_{i..}$
1	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	$\frac{1}{12}$	0	$\frac{1}{4}$
2	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	$\frac{1}{12}$	$\frac{1}{4}$
3	$\frac{1}{12}$	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	$\frac{1}{4}$
4	0	$\frac{1}{12}$	0	$\frac{1}{12}$	0	0	$\frac{1}{12}$	$\frac{1}{4}$
$f_{.pl}$	$\frac{1}{12}$	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	1

Table: Relative frequency table

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Row Profiles

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

The idea behind MCA, like in bivariate correspondence analysis, is to apply PCA on one hand to the row profiles, and on the other hand to the column profiles of the relative frequencies table F . The coordinate pl of the row profile $l_i(1 \times K)$ associated with individual i is given as

$$(l_i)_{pl} = \frac{f_{ipl}}{f_{i..}} = \frac{x_{ipl}}{P}, \quad i = 1, \dots, n.$$

The n row profiles weighted equally ($1/n$) compose a point cloud in \mathbb{R}^K with a center given by the relative marginal profile

$$G_l = \left(\frac{n_{11}}{nP}, \dots, \frac{n_{1K_1}}{nP}, \dots, \frac{n_{P1}}{nP}, \dots, \frac{n_{PK_P}}{nP} \right).$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Example: Gender, Civil Status and Education

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The row profiles of the gender, civil status and education data is given as follows.

	X_{11}	X_{12}	X_{21}	X_{22}	X_{23}	X_{31}	X_{32}	
1	0	$\frac{1}{3}$	$\frac{1}{3}$	0	0	$\frac{1}{3}$	0	1
2	0	$\frac{1}{3}$	$\frac{1}{3}$	0	0	0	$\frac{1}{3}$	1
3	$\frac{1}{3}$	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	0	1
4	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0	0	$\frac{1}{3}$	1

Table: Row profiles

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Row Profiles

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Intuitively, the distance between two individuals is small if they have many modalities in common, and the distance between the individual i and the center increases as the modalities taking by the individual i becomes rare ($x_{ipl} = 1$ for n_{pl} small).

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

More formally, the chi-square distances between two row profiles l_{i_1} and l_{i_2} can be given as

$$\begin{aligned}d^2(l_{i_1}, l_{i_2}) &= \sum_{p=1}^P \sum_{l=1}^{K_p} \frac{1}{f_{.pl}} ((l_{i_1})_{pl} - (l_{i_2})_{pl})^2 \\ &= \frac{n}{P} \sum_{p=1}^P \sum_{l=1}^{K_p} \frac{1}{n_{pl}} (x_{i_1 pl} - x_{i_2 pl})^2.\end{aligned}$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Example: Gender, Civil Status and Education

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The distance between individual 1 and individual 2 is

$$\left(\frac{n}{P} \sum_{p=1}^P \sum_{k=1}^{K_p} \frac{1}{n_{pk}} (x_{i_1, pk} - x_{i_2, pk})^2\right)$$

$$= \left(\frac{4}{3} \sum_{p=1}^3 \sum_{k=1}^{K_p} \frac{1}{n_{pk}} (x_{i_1, pk} - x_{i_2, pk})^2\right)$$

$$= \left(\frac{4}{3} (1(0-0)^2 + \frac{1}{3} (1-1)^2 + \frac{1}{2} (1-1)^2 + 1(0-0)^2 + 1(0-0)^2 + \frac{1}{2} (1-0)^2 + \frac{1}{2} (0-1)^2)\right)$$
$$= \frac{4}{3} \approx 1.33.$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

Column Profiles

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

The coordinate i of the column profile c_{pl} ($n \times 1$) associated with the modality l of Y_p is given as

$$(c_{pl})_i = \frac{f_{ipl}}{f_{.pl}} = \frac{x_{ipl}}{n_{pl}}, \quad p = 1, \dots, P; l = 1, \dots, K_p.$$

The weight of each column profiles is proportional to its frequency and given by $f_{.pl} = \frac{n_{pl}}{nP}$. The K column profiles compose a point cloud in \mathbb{R}^n with the center given by the relative marginal profile $G_c = (\frac{1}{n}, \dots, \frac{1}{n})$.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Example: Gender, Civil Status and Education

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The column profiles of the gender, civil status and education is given as follows.

	X_{11}	X_{12}	X_{21}	X_{22}	X_{23}	X_{31}	X_{32}
1	0	$\frac{1}{3}$	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0
2	0	$\frac{1}{3}$	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$
3	1	0	0	0	1	$\frac{1}{2}$	0
4	0	$\frac{1}{3}$	0	1	0	0	$\frac{1}{2}$
	1	1	1	1	1	1	1

Table: Column profiles

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Column Profiles

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Intuitively, the χ^2 distance between two modalities is small if the same individuals take these two modalities together, and the distance between the modality l of Y_p and the center increases as the modality becomes more rare (n_{pl} small).

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

More formally, the chi-square distances between two column profiles $c_{p_1 l_1}$ and $c_{p_2 l_2}$ can be given as

$$\begin{aligned}d^2(c_{p_1 l_1}, c_{p_2 l_2}) &= \sum_{i=1}^n \frac{1}{f_{i..}} ((c_{p_1 l_1})_i - (c_{p_2 l_2})_i)^2 \\ &= n \sum_{i=1}^n \left(\frac{x_{ip_1 l_1}}{n_{p_1 l_1}} - \frac{x_{ip_2 l_2}}{n_{p_2 l_2}} \right)^2.\end{aligned}$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Example: Gender, Civil Status and Education

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The distance between modality 1 of Y_1 (male) and modality 2 of Y_2 (married) is

$$\sum_{i=1}^n \frac{1}{f_{i..}} ((c_{p_1 l_1})_i - (c_{p_2 l_2})_i)^2$$
$$= 4((0 - 0)^2 + (0 - 0)^2 + (1 - 0)^2 + (0 - 1)^2) = 8$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Attraction Repulsion Indices

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Attraction Repulsion Indices

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

With categorical variables, it is usual to test, whether there is a significant association between the variables, with the chi-square test of independence. It is also interesting to compare the association at the level of the modalities instead of the variables. Let $n_{p_1 l_1, p_2 l_2}$ be the number of individuals having the modality l_1 of the variable Y_{p_1} and the modality l_2 of the variable Y_{p_2} . Now the attraction repulsion index $d_{p_1 l_1, p_2 l_2}$ between the modality l_1 of the variable Y_{p_1} and the modality l_2 of the variable Y_{p_2} is given by

$$d_{p_1 l_1, p_2 l_2} = \frac{n_{p_1 l_1, p_2 l_2} / n}{n_{p_1 l_1} / n \cdot n_{p_2 l_2} / n} = \frac{n_{p_1 l_1, p_2 l_2}}{\frac{n_{p_1 l_1} n_{p_2 l_2}}{n}}.$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Attraction Repulsion Indices

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

It is clear that if the attraction repulsion index is larger than one, the individuals are more inclined to take both modalities simultaneously than under the hypothesis of independence. And vice-versa, if the attraction repulsion index is smaller than one, the individuals are less inclined to take both modalities simultaneously than under the hypothesis of independence. The aim of the MCA is to produce graphical display in lower dimension which reproduce, without losing too much information, the associations between the modalities through the attraction repulsion index.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Attraction Repulsion Indices

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The attraction repulsion index $d_{i,pl}$ between the individual i and the modality l of the variable Y_p is defined as follows.

$$d_{i,pl} = \frac{f_{ipl}}{f_{i..} f_{.pl}} = \frac{x_{ipl}}{n_{pl}/n}.$$

Now, clearly

$$d_{i,pl} = 0,$$

if $x_{ipl} = 0$ and

$$d_{i,pl} = \frac{n}{n_{pl}},$$

if $x_{ipl} = 1$. Thus, if the individual i does not have the modality l of the variable Y_p , then the attraction repulsion index $d_{i,pl}$ is equal to 0, and if the individual i does have the modality l of Y_p , then the attraction repulsion index $d_{i,pl}$ increases as the l of Y_p becomes rare.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Multiple Correspondence Analysis

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Multiple Correspondence Analysis

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

To obtain a representation in lower dimension, PCA is applied on the two data clouds: the rows and column profiles. A transformation of the profiles is necessary to center the variables, and to be able to use euclidian distances instead of χ^2 distances. Finally, PCA is applied to the transformed row and column profiles:

$$(l_i^{\circ})_{pl} = \frac{(l_i)_{pl}}{\sqrt{f_{.pl}}} - \sqrt{f_{.pl}} \text{ and } (c_{pl}^{\circ})_i = \frac{(c_{pl})_i}{\sqrt{f_{i..}}} - \sqrt{f_{i..}}$$

The solution of the problem of maximization associated with the PCA on the transformed row and column profiles is given respectively by the eigenvalues and the eigenvectors of the matrices $V(K \times K)$ and $W(n \times n)$ where

$$V = T^T T \text{ and } W = T T^T \text{ where } T_{i,pl} = \frac{x_{ipl} - n_{pl}/n}{\sqrt{Pn_{pl}}}$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Multiple Correspondence Analysis

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The principal components for the individuals are derived from the eigenvectors of the matrix V , and the principal components for the modalities from the eigenvectors of the matrix W .

Let $H = \text{rank}(V) = \text{rank}(W)$. The new system of coordinates (scores) for the individuals is defined through the H principal components

$$\phi_{h,i} = \sum_{k=1}^K u_{h,k} (l_i^{\circ})_k \quad h = 1, \dots, H,$$

where $u_{h,k}$ is the k th element of the eigenvector associated with the h th largest eigenvalues of V .

The new system of coordinates (scores) for the modalities is defined through the principal components

$$\psi_{h,pl} = \sum_{i=1}^n v_{h,i} (c_{pl}^{\circ})_i \quad h = 1, \dots, H.$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Contribution of the Modalities

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Contribution of the modality l of Y_p on the variance of the new variable ψ_h is given by

$$C(pl, h) = \frac{f_{.pl}\psi_{h,pl}^2}{\lambda_h} = \frac{n_{pl}\psi_{h,pl}^2}{nP\lambda_h}.$$

Global contribution of the variable Y_p is given by

$$C(p, h) = \sum_{l=1}^{K_p} C(pl, h).$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Graphical Presentation

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple

Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple

Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Comparison of the Modalities

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The attraction repulsion index

$$d_{p_1 h, p_2 h} = 1 + \sum_{h=1}^H \psi_{h, p_1 h} \psi_{h, p_2 h}.$$

The graphical output of MCA is the approximation of the previous formula using few dimensions. Suppose that the modalities are well represented in two dimensions, then we can plot the two first principal components and interpret the proximity between the points on the first principal plan with the following approximation

$$d_{p_1 h, p_2 h} \approx 1 + \sum_{h=1}^2 \psi_{h, p_1 h} \psi_{h, p_2 h}.$$

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Comparison of the Individuals

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The proximity between two individuals i_1 and i_2 is defined as

$$d_{i_1, i_2} = 1 + \sum_{h=1}^H \phi_{h, i_1} \phi_{h, i_2}.$$

Two individuals are close if they have in general the same modalities.

Now d_{i_1, i_2} can be approximated by

$$d_{i_1, i_2} \approx 1 + \sum_{h=1}^2 \phi_{h, i_1} \phi_{h, i_2}.$$

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

Simultaneous Comparison

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The attraction repulsion index

$$d_{i,pl} = 1 + \sum_{h=1}^H \frac{1}{\sqrt{\lambda_h}} \phi_{h,i} \psi_{h,pl},$$

and thus again

$$d_{i,pl} \approx 1 + \sum_{h=1}^2 \frac{1}{\sqrt{\lambda_h}} \phi_{h,i} \psi_{h,pl}.$$

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

Simultaneous Comparison

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

The components are often standardized defining

$$\hat{\phi}_{1,j} = \frac{1}{\sqrt{\lambda_1}} \phi_{1,j}$$

and

$$\hat{\phi}_{2,j} = \frac{1}{\sqrt{\lambda_2}} \phi_{2,j}.$$

Then

$$d_{i,pl} \approx 1 + \sum_{h=1}^2 \hat{\phi}_{h,i} \psi_{h,pl},$$

and the final graphical representation can be given simultaneously as a double biplot.

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

Example

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Example of MCA: Extended Gender, Civil Status and Education Data

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Disclaimer: This example data set is randomly generated.
Please do not draw real life conclusions from it.

	X_{11}	X_{12}	X_{21}	X_{22}	X_{23}	X_{31}	X_{32}	$\sum_{p=1}^7 \sum_{l=1}^{K_p} X_{jpl}$
1	0	1	1	0	0	1	0	3
2	0	1	1	0	0	0	1	3
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
25	1	0	0	0	1	0	1	3
$\sum_{i=1}^n X_{ijl}$	16	9	9	6	10	14	11	125

Table: Complete disjunctive table

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

$\sum_{p=1}^7 \sum_{l=1}^{K_p} X_{jpl}$
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Example of MCA

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

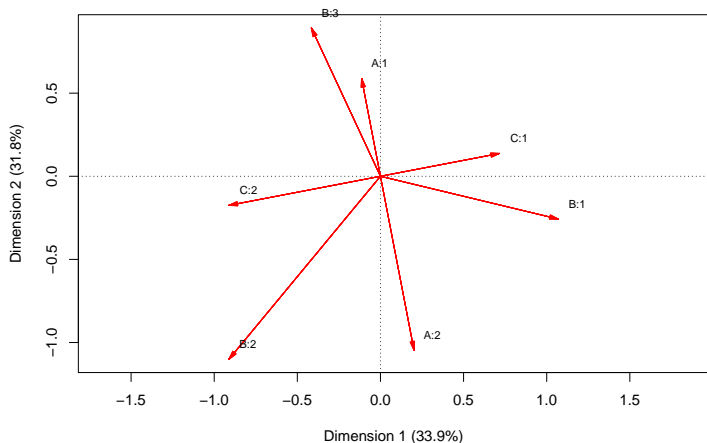


Figure: Result of MCA (A1=male, A2=female, B1=single, B2=married, B3=divorced/widowed, C1=low education, C2=at least high school diploma.)

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

Some Remarks

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Some Remarks

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

When performing MCA, it is better to take into account variables that have more or less the same number of modalities. (The number of modalities have an effect on analysis.) It is also advised to avoid having rare modalities. (Rare modalities have a big impact on analysis, and that makes MCA quite nonrobust method.) One can preprocess the data by grouping modalities if necessary.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

Next Week

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Next week we will talk about canonical correlation analysis.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

References

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices


Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References

 K. V. Mardia, J. T. Kent, J. M. Bibby, *Multivariate Analysis*, Academic Press, London, 2003 (reprint of 1979).

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

References II

Lecturer:
Pauliina Ilmonen
Slides:
Ilmonen/Kantala

-  R. V. Hogg, J. W. McKean, A. T. Craig, Introduction to Mathematical Statistics, Pearson Education, Upper Saddle River, 2005.
-  R. A. Horn, C. R. Johnson, Matrix Analysis, Cambridge University Press, New York, 1985.
-  R. A. Horn, C. R. Johnson, Topics in Matrix Analysis, Cambridge University Press, New York, 1991.

Multiple
Correspondence
Analysis
Frequency Tables
Row Profiles
Column Profiles
Attraction Repulsion
Indices
Multiple
Correspondence
Analysis
Graphical Presentation
Example
Some Remarks
References

 L. Simar, An Introduction to Multivariate Data Analysis,
Université Catholique de Louvain Press, 2008.

Multiple
Correspondence
Analysis

Frequency Tables

Row Profiles

Column Profiles

Attraction Repulsion
Indices

Multiple
Correspondence
Analysis

Graphical Presentation

Example

Some Remarks

References