Handling Missing Values with Regularized Iterative Multiple Correspondence Analysis

J. Josse\textsuperscript{1}, M. Chavent\textsuperscript{2}, B. Liquet\textsuperscript{3} & F. Husson\textsuperscript{1}

\textsuperscript{1} Agrocampus Rennes
\textsuperscript{2} University of Bordeaux 2
\textsuperscript{3} ISPED Bordeaux

CARME conference, Rennes, 9 February 2011
**Type of missing values**

- “Really missing” and “not really missing”
- MCAR, MAR, MNAR (Rubin, 1976)

⇒ In MCA, van der Heijden & Escofier (1987) discussed which method is well suited for which kind of missing data
Missing single

- **Missing single:** a new category is added for missing values
  ⇒ well-adapted for “not really missing” or MNAR
Missing passive modified margin

- Missing passive (Benzécri, 1973; Meulman, 1982)

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ind 1</td>
<td>a</td>
<td>NA</td>
<td>g</td>
</tr>
<tr>
<td>ind 2</td>
<td>NA</td>
<td>f</td>
<td>g</td>
</tr>
<tr>
<td>ind 3</td>
<td>a</td>
<td>e</td>
<td>h</td>
</tr>
<tr>
<td>ind 4</td>
<td>a</td>
<td>e</td>
<td>h</td>
</tr>
<tr>
<td>ind 5</td>
<td>b</td>
<td>f</td>
<td>h</td>
</tr>
<tr>
<td>ind 6</td>
<td>c</td>
<td>f</td>
<td>h</td>
</tr>
<tr>
<td>ind 7</td>
<td>c</td>
<td>f</td>
<td>NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>V1_a</th>
<th>V1_b</th>
<th>V1_c</th>
<th>V2_e</th>
<th>V2_f</th>
<th>V3_g</th>
<th>V3_h</th>
</tr>
</thead>
<tbody>
<tr>
<td>ind 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ind 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ind 3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ind 4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ind 5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ind 6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ind 7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Missing values are skipped
Row margins are not equal ⇒ many properties of MCA are lost

- Missing passive modified margin (Escofier, 1987)
  ⇒ row margins are fixed to $J$
  ⇒ Good properties: $f_s$ maximises $\sum_{j=1}^J \hat{\eta}_{f_s}^2 |v_j$
  ⇒ Equivalence with subset MCA (Greenacre & Pardo, 2006)
Handling missing values in exploratory multivariate analysis

The method consists to find the components $\mathbf{F}$ and the axes $\mathbf{U}$ that minimize the reconstruction error:

$$C = \| \mathbf{X} - \mathbf{F} \mathbf{U}' \|_{M,D}^2$$

With missing values, a matrix of weights $\mathbf{W}$ is introduced:

$$C = \| \mathbf{W} \ast (\mathbf{X} - \mathbf{F} \mathbf{U}') \|_{M,D}^2$$

with $w_{ik} = 0$ if $x_{ik}$ is missing and $w_{ij} = 1$ otherwise.

$\Rightarrow$ Use of iterative algorithms
Iterative algorithms

- **Initialization:** missing values in $X$ are imputed with initial values (such as the mean of each variable)
- **Estimation step:** the analysis is performed on the completed data set
- **Imputation step:** missing values are imputed with the reconstruction formulae with $S$ dimensions

$$X = W \times X + (1 - W) \times (\hat{F} \hat{U}')$$

- Steps E and M are repeated until convergence

$\Rightarrow$ EM type algorithms
$\Rightarrow$ The number of dimensions $S$ has to be chosen *a priori*
$\Rightarrow$ Nora-Chouteau in CA (1974); Kiers in PCA (1997)
Iterative MCA

MCA can be seen as the SVD of (data, metric, row masses)

\[
\left( /X D^{-1}_\Sigma, \frac{1}{I J} D \Sigma, \frac{1}{I} \mathbb{I} \right)
\]

with \( X \) the indicator matrix and \( D \Sigma \) the diagonal matrix of the column margins of \( X \),

\[
X = \begin{pmatrix}
I_1 & & \\
& I_k & \\
& & I_K
\end{pmatrix}
\]

\[
D \Sigma = \begin{pmatrix}
\cdots & & 0 \\
& I_k & \\
0 & & \cdots
\end{pmatrix}
\]
Iterative MCA

1. initialization \( \ell = 0 \): \( X^0 \) missing values are imputed with the proportion of the category (the sum must equal one) \( \Rightarrow D^0_\Sigma \);

2. step \( \ell \):
   a) MCA on \( X^{\ell-1} \): \( \hat{F} \) and \( \hat{U} \) are obtained from a PCA on
   \[
   \left( \frac{1}{l}X^{\ell-1}(D^{-1}_\Sigma)^{-1}, \frac{1}{lj}D^{-1}_\Sigma, \frac{1}{l}I_l \right)
   \]
   b) Impute the indicator matrix using the reconstruction formulae:
   \[
   \hat{x}_{ik}^{\ell} = \frac{1}{l} \left( 1 + \sum_{s=1}^{S} \hat{f}_{is}^{\ell} \hat{u}_{ks}^{\ell} \right) D^{-1}_\Sigma
   \]
   The new imputed dataset is \( X^{\ell} = W \ast X + (1 - W) \ast \hat{X}^{\ell} \)
   c) \( D^{\ell}_\Sigma \) is updated with the new column margins \( I_k^{\ell} \) of \( X^{\ell} \);

3. steps (2.a), (2.b) and (2.c) are repeated until convergence
Iterative MCA

- Step 0: missing fuzzy average = reconstruction of order 0
- The algorithm can return a completed indicator matrix

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ind 1</td>
<td>a</td>
<td>NA</td>
<td>g</td>
</tr>
<tr>
<td>ind 2</td>
<td>NA</td>
<td>f</td>
<td>g</td>
</tr>
<tr>
<td>ind 3</td>
<td>a</td>
<td>e</td>
<td>h</td>
</tr>
<tr>
<td>ind 4</td>
<td>a</td>
<td>e</td>
<td>h</td>
</tr>
<tr>
<td>ind 5</td>
<td>b</td>
<td>f</td>
<td>h</td>
</tr>
<tr>
<td>ind 6</td>
<td>c</td>
<td>f</td>
<td>h</td>
</tr>
<tr>
<td>ind 7</td>
<td>c</td>
<td>f</td>
<td>NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>V1_a</th>
<th>V1_b</th>
<th>V1_c</th>
<th>V2_e</th>
<th>V2_f</th>
<th>V3_g</th>
<th>V3_h</th>
</tr>
</thead>
<tbody>
<tr>
<td>ind 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.71</td>
<td>0.29</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ind 2</td>
<td>0.13</td>
<td>0.29</td>
<td>0.59</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ind 3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ind 4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ind 5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ind 6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ind 7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.37</td>
<td>0.63</td>
</tr>
</tbody>
</table>

- Imputed values can be seen as degree of membership
Overfitting
Overfitting

True configuration

Dim 1 (19.69%)

Dim 2 (16.73%)

iterative MCA

Dim 1 (29.34%)

Dim 2 (18.38%)

Observed values are well-tted but missing ones are badly predicted

⇒ Regularization methods
Overfitting

\[
\text{mean}(x_{ik} - \hat{x}_{ik})^2 = 0.03 \quad \text{whereas} \quad \text{mean}(x_{ik} - \hat{x}_{ik})^2 = 0.34
\]

Observed values are well-fitted but missing ones are badly predicted 
... and consequently axes and components are badly predicted 
⇒ Regularization methods
Regularized Iterative MCA

\[
\sum_{s=1}^{S} \hat{f}_{is}^\ell \hat{u}_{ks} = \sum_{s=1}^{S} \frac{\hat{f}_{is}^\ell}{\|\hat{f}_{s}^\ell\|_D} (\sqrt{\lambda_s}) \hat{u}_{ks}
\]

The eigenvalues can be shrunk in the reconstruction step:

\[
\sum_{s=1}^{S} \frac{\hat{f}_{is}^\ell}{\|\hat{f}_{s}^\ell\|_D} \left( \sqrt{\lambda_s} - \frac{\hat{\sigma}^2}{\sqrt{\lambda_s}} \right) \hat{u}_{ks}
\]

with \( \hat{\sigma}^2 = \frac{1}{K-J-S} \sum_{s=S+1}^{K-J} \lambda_s \)

\[\Rightarrow\] remove the noise to avoid instability on the predictions
Simulations

Many scenarios are considered:

- percentage of missing values: small, medium
- missing values mechanism: MCAR, MAR
- pattern of missing values: random or not random
- relationship between variables: low or strong
- 1000 simulations

The simulated data:

- 100 individuals
- 10 variables from a normal distribution
- each variable is cut in 3 equal-count categories
  ⇒ By construction, 4 underlying dimensions
The criterion used is the RV coefficient between the configuration without missing values and the one obtained from the algorithm.

<table>
<thead>
<tr>
<th>Missing</th>
<th>Link</th>
<th>Missing Passive Margin R - NR</th>
<th>Missing Fuzzy Average R - NR</th>
<th>Missing single R - NR</th>
<th>RiMCA R - NR</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% MCAR</td>
<td>low</td>
<td>0.94 - 0.91</td>
<td>0.94 - 0.92</td>
<td>0.87 - 0.47</td>
<td>0.94 - 0.93</td>
</tr>
<tr>
<td>10% MCAR</td>
<td>strong</td>
<td>0.97 - 0.94</td>
<td>0.97 - 0.95</td>
<td>0.96 - 0.68</td>
<td>0.98 - 0.97</td>
</tr>
<tr>
<td>30% MCAR</td>
<td>low</td>
<td>0.77 - 0.44</td>
<td>0.77 - 0.77</td>
<td>0.67 - 0.32</td>
<td>0.76 - 0.78</td>
</tr>
<tr>
<td>30% MCAR</td>
<td>strong</td>
<td>0.88 - 0.71</td>
<td>0.88 - 0.91</td>
<td>0.86 - 0.46</td>
<td>0.91 - 0.90</td>
</tr>
<tr>
<td>8% MAR</td>
<td>low</td>
<td>0.94 - 0.91</td>
<td>0.94 - 0.91</td>
<td>0.72 - 0.28</td>
<td>0.95 - 0.92</td>
</tr>
<tr>
<td>8% MAR</td>
<td>strong</td>
<td>0.96 - 0.91</td>
<td>0.96 - 0.90</td>
<td>0.96 - 0.54</td>
<td>0.98 - 0.96</td>
</tr>
<tr>
<td>16% MAR</td>
<td>low</td>
<td>0.86 - 0.80</td>
<td>0.83 - 0.79</td>
<td>0.50 - 0.29</td>
<td>0.88 - 0.83</td>
</tr>
<tr>
<td>16% MAR</td>
<td>strong</td>
<td>0.89 - 0.80</td>
<td>0.84 - 0.78</td>
<td>0.88 - 0.55</td>
<td>0.95 - 0.90</td>
</tr>
</tbody>
</table>
A real example

- 1232 respondents, 14 questions, 35 categories, 9% of missing values concerning 42% of respondents
A real example

- 1232 respondents, 14 questions, 35 categories, 9% of missing values concerning 42% of respondents
A real example
A real example
Conclusion

Regularized iterative MCA

- gives “good” results
- is efficient when strong relationships between variables (you learn from the other variables) ...
- ... but needs tuning parameters
- can be used as an imputation method?
- can be used to perform a clustering on categorical variables with missing values
- is available in the missMDA package that imputes the indicator matrix and the FactoMineR package that performs the MCA from an indicator matrix