Multiple Correspondence $K$-means

a new approach for simultaneous dimension reduction and clustering

M. Fordellone$^1$ and M. Vichi$^1$

$^1$ Department of Statistical Science
Sapienza University of Rome

Giornata del dottorato 2016
Rome, 9th February 2016
Outline

Introduction

Sequential approach
  Tandem Analysis type 1
  Tandem Analysis type 2

Simultaneous approach
  Reduced K-means (RKM) and Factorial K-means (FKM)
  Multiple Correspondence K-means (MCKM)

Sequential vs simultaneous approaches
  Quantitative data example
  Qualitative data example

Application of MCKM on real data
  South Korean underwear manufacturer (1997)

Conclusions
Often, given a $N \times J$ data matrix, we could be interested to explore the relationships within the set of $N$ objects through a discrete classification model.
Often, given a $N \times J$ data matrix, we could be interested to explore the relationships within the set of $N$ objects through a discrete classification model.

When the number of variables is large, for the dimension reduction, we could apply a factorial model on the $J$ variables and a classification model on the computed object scores: this sequential approach is called *tandem analysis* (Arabie and Hubert, 1994).
Introduction

Often, given a $N \times J$ data matrix, we could be interested to explore the relationships within the set of $N$ objects through a discrete classification model. When the number of variables is large, for the dimension reduction, we could apply a factorial model on the $J$ variables and a classification model on the computed object scores: this sequential approach is called *tandem analysis* (Arabie and Hubert, 1994). However, in some case this approach is not reliable because the factorial methods may identify dimensions that do not necessarily represent the clustering structure in the data (De Sarbo et al., 1990).
Often, given a $N \times J$ data matrix, we could be interested to explore the relationships within the set of $N$ objects through a discrete classification model. When the number of variables is large, for the dimension reduction, we could apply a factorial model on the $J$ variables and a classification model on the computed object scores: this sequential approach is called tandem analysis (Arabie and Hubert, 1994). However, in some case this approach is not reliable because the factorial methods may identify dimensions that do not necessarily represent the clustering structure in the data (De Sarbo et al., 1990). A solution to this problem is the statistical methods that use the simultaneous dimension reduction and cluster analysis on the data matrix $X$ (De Soete and Carroll, 1994; Vichi and Kiers, 2001).
...what are the aims of the work?

The principal aims of the work are:

- Provide a background on the statistical methods usually used for dimension reduction and clustering.
- Present a new approach for simultaneous dimension reduction and clustering for categorical data.
- Talk about the differences between the sequential and simultaneous approaches through the applications on simulated data.
- Show an empirical application of the new approach on the real categorical data.
Outline

Introduction

Sequential approach
  Tandem Analysis type 1
  Tandem Analysis type 2

Simultaneous approach
  Reduced K-means (RKM) and Factorial K-means (FKM)
  Multiple Correspondence K-means (MCKM)

Sequential vs simultaneous approaches
  Quantitative data example
  Qualitative data example

Application of MCKM on real data
  South Korean underwear manufacturer (1997)

Conclusions
Tandem Analysis type 1: Apply a factorial method on the data matrix $\mathbf{X}$ and, sequentially, $K$-means on score matrix $\mathbf{Y}$. 
The loss-function of the Tandem Analysis type 1 can be seen as the mean of the objective functions of Principal Component Analysis (PCA) and \( K \)-means:
Tandem Analysis type 1

The loss-function of the Tandem Analysis type 1 can be seen as the mean of the objective functions of Principal Component Analysis (PCA) and K-means:

$$f(Y, A, U, \bar{Y}) = \frac{1}{2} \left( \left\| X - YA' \right\|^2_{\text{PCA}} + \left\| YA' - U\bar{Y}A' \right\|^2_{\text{K-means}} \right)$$

where:

- $X$ is the $N \times J$ data matrix,
- $Y$ is the $N \times P$ scores matrix of PCA,
- $\bar{Y}$ is the $K \times P$ centroid matrix in reduced space,
- $U$ is the $N \times K$ memberships matrix of K-means,
- $A$ is the $J \times P$ loadings matrix of PCA,

note that $Y = XA$ and $\bar{Y} = \bar{X}A$. 
Outline

Introduction

Sequential approach
  Tandem Analysis type 1
  Tandem Analysis type 2

Simultaneous approach
  Reduced K-means (RKM) and Factorial K-means (FKM)
  Multiple Correspondence K-means (MCKM)

Sequential vs simultaneous approaches
  Quantitative data example
  Qualitative data example

Application of MCKM on real data
  South Korean underwear manufacturer (1997)

Conclusions
Tandem Analysis type 2: Apply $K$-means on the data matrix $\mathbf{X}$ and, sequentially, a factorial method on the centroid matrix $\bar{\mathbf{X}}$. 

$\mathbf{X} = (N \times J)$

$K$-means

$\bar{\mathbf{X}} = (K \times J)$

ACP/FA/MCA

$\bar{\mathbf{Y}} = (K \times P)$
Tandem Analysis type 2

The loss-function of the Tandem Analysis type 2 can be seen as the mean of the objective functions of $K$-means and Principal Component Analysis (PCA):

$$f(U, \bar{X}, \bar{Y}, A) = \frac{1}{2} \left( \| X - U \bar{X} \|_2^2 + \| U \bar{X} - U \bar{Y} A' \|_2^2 \right)$$

where:
- $X$ is the $N \times J$ data matrix,
- $\bar{X}$ is the $K \times J$ centroid matrix in original space,
- $\bar{Y}$ is the $K \times P$ centroid matrix in reduced space,
- $U$ is the $N \times K$ memberships matrix of $K$-means,
- $A$ is the $J \times P$ loadings matrix of PCA,

so that $Y = XA$ and $\bar{Y} = \bar{X}A$. 
Tandem Analysis type 2

The loss-function of the Tandem Analysis type 2 can be seen as the mean of the objective functions of $K$-means and Principal Component Analysis (PCA):

$$f(U, \bar{X}, \bar{Y}, A) = \frac{1}{2} \left( \|X - U\bar{X}\|^2 + \|U\bar{X} - U\bar{Y}A'\|^2 \right)$$

where:
- $X$ is the $N \times J$ data matrix,
- $\bar{X}$ is the $K \times J$ centroid matrix in original space,
- $\bar{Y}$ is the $K \times P$ centroid matrix in reduced space,
- $U$ is the $N \times K$ memberships matrix of $K$-means,
- $A$ is the $J \times P$ loadings matrix of PCA,
- note that $Y = XA$ and $\bar{Y} = \bar{X}A$. 
Outline

Introduction

Sequential approach
  Tandem Analysis type 1
  Tandem Analysis type 2

Simultaneous approach
  Reduced K-means (RKM) and Factorial K-means (FKM)
  Multiple Correspondence K-means (MCKM)

Sequential vs simultaneous approaches
  Quantitative data example
  Qualitative data example

Application of MCKM on real data
  South Korean underwear manufacturer (1997)

Conclusions
RKM and FKM

PCA:
\[ X = XAA' + E_1 \]
subject to:
\[ A' A = I_P \text{ (orthogonal)} \]

K-means:
\[ XA = U\bar{X}A + E_2 \]
subject to:
\[ U[u_{ik} \in \{0, 1\}] \text{ (binary)} \]
\[ U1_K = 1_N \text{ (row stochastic)} \]

\[ E_1 \text{ and } E_2 \text{ are the errors matrix} \]
**PCA:**  
\[ X = XAA' + E_1 \]  
subject to:  
\[ A'A = I_P \text{ (orthogonal)} \]

**K-means:**  
\[ XA = U\tilde{X}A + E_2 \]  
subject to:  
\[ U[u_{ik} \in \{0, 1\}] \text{ (binary)} \]  
\[ U1_K = 1_N \text{ (row stochastic)} \]

**Reduced K-means (RKM)**  
If \( E_{\text{RKM}} = E_2A' + E_1 \)  
\[ X = (U\tilde{X}A + E_2)A' + E_1 \]  
\[ X = U\tilde{X}AA' + E_{\text{RKM}} \]  

\( E_1 \) and \( E_2 \) are the errors matrix.
RKM and FKM

**PCA:**
\[ X = XAA' + E_1 \]
subject to:
\[ A'A = I_p \text{ (orthogonal)} \]

\[ E_1 \text{ and } E_2 \text{ are the errors matrix} \]

**K-means:**
\[ XA = U\tilde{X}A + E_2 \]
subject to:
\[ U[u_{ik} \in \{0, 1\}] \text{ (binary)} \]
\[ U1_K = 1_N \text{ (row stochastic)} \]

**Reduced K-means (RKM)**
\[ X = (U\tilde{X}A + E_2)A' + E_1 \]
if \( E_{RKM} = E_2A' + E_1 \)

\[ X = U\tilde{X}AA' + E_{RKM} \]

by post-multiplying both members of the previous equation by \( A \) and rewriting the error \( E_{FKM} = E_{RKM} \), we have the Factorial K-means (FKM):

\[ XA = U\tilde{X}A + E_{FKM} \]

subject to:
\[ U[u_{ik} \in \{0, 1\}] \text{ (binary)} \]
\[ U1_K = 1_N \text{ (row stochastic)} \]
\[ A'A = I_p \text{ (orthogonal)} \]
RKM and FKM

Objective function Reduced K-means
\[ \min_{\mathbf{U}, \bar{\mathbf{X}}, \mathbf{A}} \| \mathbf{X} - \mathbf{U}\bar{\mathbf{X}}\mathbf{A}'\|_2^2 \text{ or } \max_{\mathbf{U}, \bar{\mathbf{X}}, \mathbf{A}} \| \mathbf{U}\bar{\mathbf{X}}\mathbf{A}'\|_2^2 \]

subject to
\[ \mathbf{U}[u_{ik} \in \{0, 1\}]; \ i = 1, \ldots, N; \ k = 1, \ldots, K \]
\[ \mathbf{U}1_K = \mathbf{1}_N \]
\[ \mathbf{A}'\mathbf{A} = \mathbf{I}_P \]
RKM and FKM

Objective function Reduced K-means

\[
\min_{U, \bar{X}, A} \|X - U\bar{X}AA'\|^2 \quad \text{or} \quad \max_{U, \bar{X}, A} \|U\bar{X}AA'\|^2
\]

subject to
\[
U[u_{ik} \in \{0, 1\}] ; \quad i = 1, \ldots, N ; \quad k = 1, \ldots, K \\
U1_K = 1_N \\
A'A = I_p
\]

Alternative least-squares Reduced K-means algorithm

- **Step 0**: Initial random values are chosen for \(\hat{A}\) and \(\hat{U}\).
- **Step 1**: Update \(\bar{X}\) by \(\bar{X} = (\hat{U}'\hat{U})^{-1}\hat{U}'X\), given \(\hat{U}\).
- **Step 2**: Update \(U\) by \(F([u_{ik}]) = \|X - U\hat{X}\hat{A}\|^2\), given the current values of \(\hat{A}\) and \(\hat{X}\). The problem is solved by taking \(u_{ik} = 1\), if \(F([u_{ik}]) = \min\{F([u_{iv}]) : v = 1, \ldots, P; (v \neq k)\}; u_{ik} = 0\), otherwise.
- **Step 3**: Update \(A\), given \(\hat{U}\) and \(\hat{X}\), by minimizing \(\|X - \hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}'XAA'\|^2\). The problem is solved by taking the first \(p\) eigenvectors of \(X'(\hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}')X\) (e.g., see Ten Berge, 1993).
- **Stopping Rule**: compute the objective function for the current values of \(\hat{U}\), \(\hat{X}\) and \(\hat{A}\). If the function decreases more than an arbitrary small constant, it return to **Step 1**; otherwise, stop the process.
RKM and FKM

**Objective function Factorial K-means**

\[
\min_{U, \bar{X}, A} \| XA - U\bar{X}A \|^2
\]

subject to

- \( U[u_{ik} \in \{0, 1\}]; \ i = 1, \ldots, N; \ k = 1, \ldots, K \)
- \( U1_K = 1_N \)
- \( A' A = I_P \)
**Objective function** Factorial K-means

\[
\min_{U, \bar{X}, A} \|XA - U\bar{X}A\|^2
\]

subject to

- \(U[u_{ik} \in \{0, 1\}]; i = 1, \ldots, N; k = 1, \ldots, K\)
- \(U1_K = 1_N\)
- \(A'\bar{A} = I_p\)

**Alternative least-squares Factorial K-means algorithm**

- **Step 0**: Initial random values are chosen for \(\hat{A}\) and \(\hat{U}\).
- **Step 1**: Update \(\bar{X}\) by \(\bar{X} = (\hat{U}'\hat{U})^{-1}\hat{U}'X\), given \(\hat{U}\).
- **Step 2**: Update \(U\) by \(F([u_{ik}]) = \|XA - U\hat{X}\hat{A}\|^2\), given the current values of \(\hat{A}\) and \(\hat{X}\). The problem is solved by taking \(u_{ik} = 1\), if \(F([u_{ik}]) = min\{F([u_{iv}]): v = 1, \ldots, P; (v \neq k)\}; u_{ik} = 0\), otherwise.
- **Step 3**: Update \(A\), given \(\hat{U}\) and \(\hat{X}\), by minimizing \(\|XA - \hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}'XA\|^2\). The problem is solved by taking the first \(p\) eigenvectors of \(X'(\hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}')X\) (e.g., see Ten Berge, 1993).
- **Stopping Rule**: compute the objective function for the current values of \(\hat{U}, \hat{X}\) and \(\hat{A}\). If the function decreases more than an arbitrary small constant, it return to **Step 1**; otherwise, stop the process.
Outline

Introduction

Sequential approach
- Tandem Analysis type 1
- Tandem Analysis type 2

Simultaneous approach
- Reduced K-means (RKM) and Factorial K-means (FKM)
- Multiple Correspondence K-means (MCKM)

Sequential vs simultaneous approaches
- Quantitative data example
- Qualitative data example

Application of MCKM on real data
- South Korean underwear manufacturer (1997)

Conclusions
MCKM

**MCA:**

\[ J^{-1/2} J B L^{-1/2} = Y A' + E_1 \]

subject to:

\[ A'A = I_P \text{ (orthogonal)} \]

**K-means:**

\[ Y = U \bar{Y} + E_2 \]

subject to:

\[ U[u_{ik} \in \{0, 1\}] \text{ (binary)} \]

\[ U 1_K = 1_N \text{ (row stochastic)} \]

where:

- \( B = [B_1, B_2, \ldots, B_J] \) is the matrix of indicator matrices of the \( J \) categorical variables,
- \( L = \text{diag}(B'1_N) \),
- \( J \) is the idempotent centering matrix with \( 1_N \) the vector of unitary elements,
- \( U \) is the \( N \times K \) memberships matrix of \( K \)-means,
- \( \bar{Y} \) is the \( K \times P \) scores matrix in reduced space.

\( E_1 \) and \( E_2 \) are the errors matrix.
Multiple Correspondence K-means (MCKM)

\[ J^{-1/2} JBL^{-1/2} = (U\bar{Y} + E_2)A' + E_1 \]

if \( E_{MCKM} = E_2A' + E_1 \)

\[ J^{-1/2} JBL^{-1/2} = U\bar{Y}A' + E_{MCKM} \]
**Multiple Correspondence K-means (MCKM)**

\[ J^{-1/2} JBL^{-1/2} = (U\bar{Y} + E_2)A' + E_1 \]

if \( E_{MCKM} = E_2A' + E_1 \)

\[ J^{-1/2} JBL^{-1/2} = U\bar{Y}A' + E_{MCKM} \]

**Objective function**

\[ \min_{U,\bar{Y},A} \left\| J^{-1/2} JBL^{-1/2} - U\bar{Y}A' \right\|^2 \quad \text{or} \quad \max_{U,\bar{Y},A} \left\| U\bar{Y}A' \right\|^2 \]

**subject to**

\[ U[u_{ik} \in \{0, 1\}; \ i = 1, \ldots, N; \ k = 1, \ldots, K \]

\[ U1_K = 1_N \]

\[ A'A = I_P \]
Outline

Introduction

Sequential approach
   Tandem Analysis type 1
   Tandem Analysis type 2

Simultaneous approach
   Reduced K-means (RKM) and Factorial K-means (FKM)
   Multiple Correspondence K-means (MCKM)

Sequential vs simultaneous approaches
   Quantitative data example
   Qualitative data example

Application of MCKM on real data
   South Korean underwear manufacturer (1997)

Conclusions
Quantitative data example

Criterion

900 units sampled by three bivariate Gaussian distribution with a structure of three groups.

\[
\begin{align*}
\mu_1 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\Sigma_1 &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \\
\mu_2 &= \begin{bmatrix} 6 \\ 0 \end{bmatrix} \\
\Sigma_2 &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \\
\mu_3 &= \begin{bmatrix} 3 \\ 5 \end{bmatrix} \\
\Sigma_3 &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}
\end{align*}
\]
Quantitative data example

Criterion

900 units sampled by three bivariate Gaussian distribution with a structure of three groups. Four univariate Gaussian distribution white noise and with the same variance have been added to data.
Quantitative data example

**Criterion**

900 units sampled by three bivariate Gaussian distribution with a structure of three groups. Four univariate Gaussian distribution white noise and with the same variance have been added to data. Then we have a $900 \times 6$ data matrix.
Quantitative data example

**Tandem Analysis 1**

<table>
<thead>
<tr>
<th>Group</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>171</td>
<td>12</td>
<td>117</td>
<td>300</td>
</tr>
<tr>
<td>Group 2</td>
<td>18</td>
<td>147</td>
<td>135</td>
<td>300</td>
</tr>
<tr>
<td>Group 3</td>
<td>25</td>
<td>128</td>
<td>147</td>
<td>300</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>214</strong></td>
<td><strong>287</strong></td>
<td><strong>399</strong></td>
<td><strong>900</strong></td>
</tr>
</tbody>
</table>

- Misclassification rate = 48.33%
- Rand-Adj = 0.1407

**Factorial K-means**

<table>
<thead>
<tr>
<th>Group</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>300</td>
<td>0</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>Group 2</td>
<td>0</td>
<td>300</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>Group 3</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>300</strong></td>
<td><strong>300</strong></td>
<td><strong>300</strong></td>
<td><strong>900</strong></td>
</tr>
</tbody>
</table>

- Misclassification rate = 0.00%
- Rand-Adj = 1.0000
### Quantitative data example

#### LOADINGS MATRIX PCA

<table>
<thead>
<tr>
<th></th>
<th>Var 1</th>
<th>Var 2</th>
<th>Var 3</th>
<th>Var 4</th>
<th>Var 5</th>
<th>Var 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp 1</td>
<td>-0.300</td>
<td>0.635</td>
<td>-0.564</td>
<td>0.090</td>
<td>0.396</td>
<td>-0.152</td>
</tr>
<tr>
<td>Comp 2</td>
<td>-0.518</td>
<td>-0.330</td>
<td>-0.092</td>
<td>0.573</td>
<td>-0.296</td>
<td>-0.445</td>
</tr>
<tr>
<td>Comp 3</td>
<td>0.736</td>
<td>-0.094</td>
<td>-0.245</td>
<td>0.158</td>
<td>0.095</td>
<td>-0.597</td>
</tr>
<tr>
<td>Comp 4</td>
<td>0.014</td>
<td>0.599</td>
<td>0.698</td>
<td>0.138</td>
<td>-0.121</td>
<td>-0.346</td>
</tr>
<tr>
<td>Comp 5</td>
<td>0.295</td>
<td>0.156</td>
<td>-0.024</td>
<td>0.766</td>
<td>-0.025</td>
<td>0.548</td>
</tr>
<tr>
<td>Comp 6</td>
<td>-0.112</td>
<td>-0.308</td>
<td>0.355</td>
<td>0.181</td>
<td>0.855</td>
<td>-0.050</td>
</tr>
</tbody>
</table>

#### LOADINGS MATRIX FACTORIAL K-MEANS

<table>
<thead>
<tr>
<th></th>
<th>Var 1</th>
<th>Var 2</th>
<th>Var 3</th>
<th>Var 4</th>
<th>Var 5</th>
<th>Var 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp 1</td>
<td>0.064</td>
<td>0.991</td>
<td>-0.046</td>
<td>-0.003</td>
<td>0.107</td>
<td>0.005</td>
</tr>
<tr>
<td>Comp 2</td>
<td>0.997</td>
<td>-0.066</td>
<td>0.009</td>
<td>-0.037</td>
<td>0.014</td>
<td>-0.011</td>
</tr>
</tbody>
</table>
Outline

Introduction

Sequential approach
  Tandem Analysis type 1
  Tandem Analysis type 2

Simultaneous approach
  Reduced K-means (RKM) and Factorial K-means (FKM)
  Multiple Correspondence K-means (MCKM)

Sequential vs simultaneous approaches
  Quantitative data example
  Qualitative data example

Application of MCKM on real data
  South Korean underwear manufacturer (1997)

Conclusions
Qualitative data example

Criterion

90 units sampled by two multinomial distributions with nine categories and a structure of three groups.
Qualitative data example

Criterion

90 units sampled by two multinomial distributions with nine categories and a structure of three groups. Others four multinomial distributions with the same number of categories have been added to data as noise.
Qualitative data example

Criterion

90 units sampled by two multinomial distributions with nine categories and a structure of three groups.
Others four multinomial distributions with the same number of categories have been added to data as noise.
Then we have a $90 \times 54$ indicators data matrix.
Qualitative data example

**Tandem Analysis 1**

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>29</td>
<td>0</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Group 2</td>
<td>12</td>
<td>12</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>Group 3</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>12</td>
<td>37</td>
<td>90</td>
</tr>
</tbody>
</table>

**Multiple Correspondence K-means**

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Group 2</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Group 3</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>90</td>
</tr>
</tbody>
</table>

Missclassification rate = 21.11%
Rand-Adj = 0.5226

Missclassification rate = 0.00%
Rand-Adj = 1.0000
## Qualitative data example

### LOADINGS MATRIX MCA

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var 1</td>
<td>-0.009</td>
<td>0.072</td>
<td>0.093</td>
<td>-0.047</td>
<td>0.159</td>
<td>0.313</td>
<td>-0.176</td>
<td>-0.214</td>
<td>-0.190</td>
</tr>
<tr>
<td>Var 2</td>
<td>0.129</td>
<td>0.040</td>
<td>0.255</td>
<td>0.062</td>
<td>-0.019</td>
<td>0.113</td>
<td>-0.163</td>
<td>-0.183</td>
<td>-0.235</td>
</tr>
<tr>
<td>Var 3</td>
<td>-0.059</td>
<td>-0.077</td>
<td>-0.151</td>
<td>-0.142</td>
<td>0.125</td>
<td>0.001</td>
<td>0.182</td>
<td>-0.069</td>
<td>0.171</td>
</tr>
<tr>
<td>Var 4</td>
<td>-0.068</td>
<td>-0.040</td>
<td>-0.091</td>
<td>0.225</td>
<td>0.103</td>
<td>-0.113</td>
<td>-0.073</td>
<td>0.088</td>
<td>-0.006</td>
</tr>
<tr>
<td>Var 5</td>
<td>0.010</td>
<td>0.006</td>
<td>0.037</td>
<td>0.108</td>
<td>-0.002</td>
<td>-0.012</td>
<td>0.187</td>
<td>-0.110</td>
<td>-0.262</td>
</tr>
<tr>
<td>Var 6</td>
<td>-0.181</td>
<td>-0.062</td>
<td>0.041</td>
<td>0.094</td>
<td>-0.067</td>
<td>0.329</td>
<td>-0.016</td>
<td>-0.048</td>
<td>0.061</td>
</tr>
</tbody>
</table>

### LOADINGS MATRIX MULTIPLE CORRESPONDENCE K-MEANS

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var 1</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
<td>-0.302</td>
<td>-0.302</td>
<td>-0.302</td>
<td>0.206</td>
<td>0.206</td>
<td>0.206</td>
</tr>
<tr>
<td>Var 2</td>
<td>-0.302</td>
<td>-0.302</td>
<td>-0.302</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
<td>0.206</td>
<td>0.206</td>
<td>0.206</td>
</tr>
<tr>
<td>Var 3</td>
<td>-0.042</td>
<td>-0.071</td>
<td>-0.019</td>
<td>0.043</td>
<td>-0.046</td>
<td>0.014</td>
<td>0.016</td>
<td>0.115</td>
<td>-0.009</td>
</tr>
<tr>
<td>Var 4</td>
<td>0.087</td>
<td>-0.017</td>
<td>0.027</td>
<td>-0.068</td>
<td>-0.073</td>
<td>0.002</td>
<td>0.058</td>
<td>0.000</td>
<td>-0.059</td>
</tr>
<tr>
<td>Var 5</td>
<td>-0.024</td>
<td>-0.036</td>
<td>-0.066</td>
<td>0.000</td>
<td>0.012</td>
<td>0.034</td>
<td>-0.060</td>
<td>0.071</td>
<td>0.088</td>
</tr>
<tr>
<td>Var 6</td>
<td>0.062</td>
<td>0.032</td>
<td>-0.066</td>
<td>0.000</td>
<td>0.021</td>
<td>-0.214</td>
<td>0.011</td>
<td>0.074</td>
<td>-0.060</td>
</tr>
</tbody>
</table>

Comp 1  Comp 2  Comp 1  Comp 2

LOADINGS MATRIX MCA  LOADINGS MATRIX MULTIPLE CORRESPONDENCE K-MEANS
Outline

Introduction

Sequential approach
  Tandem Analysis type 1
  Tandem Analysis type 2

Simultaneous approach
  Reduced K-means (RKM) and Factorial K-means (FKM)
  Multiple Correspondence K-means (MCKM)

Sequential vs simultaneous approaches
  Quantitative data example
  Qualitative data example

Application of MCKM on real data
  South Korean underwear manufacturer (1997)

Conclusions
About data

The present example is part of a large survey conducted by a South Korean underwear manufacturer in 1997 (Yang, 1997). In particular, 664 South Korean consumers were asked to provide responses for three multiple-choice items: preferred brand of underwear (8 brands), attributes when considering a brand of underwear to purchase (15 attributes) and consumer age (3 levels).
### Brands of underwear

- A01. BYC
- A02. TRY
- A03. VICMAN
- A04. James Dean
- A05. Michiko-London
- A06. Benetton
- A07. Bodyguard
- A08. Calvin Klein
About data

Attributes of underwear

B01. Comfortable
B02. Smooth
B03. Superior fabrics
B04. Reasonable price
B05. Fashionable design
B06. Favorable advertisements
B07. Trendy color
B08. Good design
B09. Various colors
B10. Elastic
B11. Store is near
B12. Excellent fit
B13. Design quality
B14. Youth appeal
B15. Various sizes
About data

Age category of the consumers

C01. 10 - 29
C02. 30 - 49
C03. 50 and over
Results

MCKM with P=2 and K=3
Results

MCKM with P=2 and K=3

Inertia Component 1 \( \approx 15\% \)
Inertia Component 2 \( \approx 12\% \)
Total Inertia \( \approx 27\% \)
Results

MCKM with $P=2$ and $K=3$

Inertia Component 1 $\approx 15\%$
Inertia Component 2 $\approx 12\%$
Total Inertia $\approx 27\%$

Group 1: $n = 240$
Group 2: $n = 241$
Group 3: $n = 183$
## Results

<table>
<thead>
<tr>
<th></th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BRA</td>
<td>ATT</td>
</tr>
<tr>
<td>1</td>
<td>0.148</td>
<td>0.060</td>
</tr>
<tr>
<td>2</td>
<td>0.180</td>
<td>0.020</td>
</tr>
<tr>
<td>3</td>
<td>0.068</td>
<td>0.021</td>
</tr>
<tr>
<td>4</td>
<td>-0.079</td>
<td>0.037</td>
</tr>
<tr>
<td>5</td>
<td>-0.092</td>
<td>-0.134</td>
</tr>
<tr>
<td>6</td>
<td>-0.082</td>
<td>-0.001</td>
</tr>
<tr>
<td>7</td>
<td>-0.208</td>
<td>-0.054</td>
</tr>
<tr>
<td>8</td>
<td>-0.093</td>
<td>-0.060</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>0.010</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>0.022</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>-0.025</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>-0.016</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-0.085</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>-0.044</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>0.010</td>
</tr>
</tbody>
</table>
...in summary

We have proposed a new simultaneous model for dimension reduction and clustering in the case of categorical data... but what are the principal advantages?

1. Simultaneous approach identifies the best partition of the objects, described by the best orthogonal linear combinations of the variables according to an optimization criterion.

2. We have seen that in some case the simultaneous approach computed more homogeneous and better-separated groups of objects with respect the tandem approach.

3. The new proposed model, with respect to tandem analysis minimizing a single objective function solving the quadratic constrained problem through a simple ALS algorithm.
References

Greenacre M., Pardo R.
Multiple correspondence analysis of a subset of response categories
SSRN 847647, 2005

De Sarbo W., Jedidi K., Cool K., Schendel D.
Simultaneous multidimensional unfolding and cluster analysis: An investigation of strategic groups.

Hwang H., Dillon W.R., Takane Y.
An extension of multiple correspondence analysis for identifying heterogeneous subgroups of respondents

Timmerman M. E., Ceulemans E., Kiers H. A., Vichi, M.
Factorial and reduced K-means reconsidered
Computational Statistics & Data Analysis 54(7), 2010: pp. 1858–1871

Van De Velden M., D’Enza A. I., Palumbo F.
Cluster Correspondence Analysis

Vichi M., Kiers H. A.
Factorial K-means analysis for two-way data
Computational Statistics & Data Analysis 37(1), 2001: pp. 49–64

Yamamoto, M., Hwang, H.
A general formulation of cluster analysis with dimension reduction and subspace separation
Behaviormetrika 41(1), 2014: pp. 115–129