INTRODUCTION TO LOG-LINEAR MODELING

Raymond Sin-Kwok Wong
University of California-Santa Barbara
September 8-12
Academia Sinica
Taipei, Taiwan
Hypothetical Data for Admission to Graduate School

<table>
<thead>
<tr>
<th>Department 1</th>
<th>Accept</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Department 2</th>
<th>Accept</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>72</td>
</tr>
</tbody>
</table>

Department 1: Science & Engineering
Department 2: Social Sciences & Liberal Arts
## Data Collapsed Over Department

<table>
<thead>
<tr>
<th></th>
<th>Accept</th>
<th>Reject</th>
<th>Acceptance Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>26</td>
<td>41</td>
<td>67</td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>51</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>92</td>
<td>132</td>
</tr>
</tbody>
</table>

Odds-Ratio, $\theta = \frac{(26)(51)}{(14)(41)} = 2.31$

$X^2 = 4.6577$, $G^2 = 4.7148$, 1 df, $p = 0.0299$

Acceptance Rate for Male > Female

Files: 2way01.inp, 2way01.out
Data Collapsed Over Acceptance Outcome

<table>
<thead>
<tr>
<th></th>
<th>Accept</th>
<th>Reject</th>
<th>% Male Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept 1</td>
<td>39</td>
<td>11</td>
<td>78.0%</td>
</tr>
<tr>
<td>Dept 2</td>
<td>28</td>
<td>54</td>
<td>34.1%</td>
</tr>
</tbody>
</table>

Odds-Ratio, $\theta = (39)(54)/(28)(11) = 6.84$

$X^2 = 23.8991$, $G^2 = 24.9815$, 1 df, $p = 0.0000$

Male Applicants in Dept 1 > Male Applicants in Dept 2

Files: 2way02.inp, 2way02.out
Data Collapsed Over Sex of Applicant

<table>
<thead>
<tr>
<th></th>
<th>Accept</th>
<th>Reject</th>
<th>Acceptance Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept 1</td>
<td>30</td>
<td>20</td>
<td>60.0%</td>
</tr>
<tr>
<td>Dept 2</td>
<td>10</td>
<td>72</td>
<td>12.2%</td>
</tr>
</tbody>
</table>

Odds-Ratio, $\theta = (30)(72)/(10)(20) = 10.80$

$X^2 = 33.6089$, $G^2 = 32.8288$, 1 df, $p = 0.0000$

Acceptance Rate in Dept 1 > Acceptance Rate in Dept 2

Files: 2way03.inp, 2way03.out
Hypothetical Data for Admission to Graduate School

<table>
<thead>
<tr>
<th></th>
<th>Department 1</th>
<th></th>
<th>Department 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accept</td>
<td>Reject</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23</td>
<td>16</td>
<td>39</td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>20</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Accept</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>72</td>
</tr>
</tbody>
</table>

$X^2 = 0.0777, G^2 = 0.0783$  
$\sum G^2 = 0.1670, \text{ 2 df, } p = 0.9199$
Fail to reject $H_0$: Conditional Independence
Hierarchical Log-Linear Models for School Admission Example

**SEX x ADMIT x DEPT (2 x 2 x 2)**

<table>
<thead>
<tr>
<th>Model Description</th>
<th>df</th>
<th>$L^2$</th>
<th>$p$</th>
<th>FILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SEX+ADMINT+DEPT</td>
<td>4</td>
<td>58.9773</td>
<td>0.000</td>
<td>3way01</td>
</tr>
<tr>
<td>2. SEX*ADMINT+DEPT</td>
<td>3</td>
<td>54.2625</td>
<td>0.000</td>
<td>3way02</td>
</tr>
<tr>
<td>3. SEX*DEPT+ADMINT</td>
<td>3</td>
<td>33.9958</td>
<td>0.000</td>
<td>3way03</td>
</tr>
<tr>
<td>4. SEX+ADMINT*DEPT</td>
<td>3</td>
<td>25.1486</td>
<td>0.000</td>
<td>3way04</td>
</tr>
<tr>
<td>5. SEX<em>DEPT+ADMINT</em>DEPT</td>
<td>2</td>
<td>0.1670</td>
<td>0.920</td>
<td>3way05</td>
</tr>
<tr>
<td>6. SEX<em>DEPT+ADMINTY</em>DEPT+SEX*ADMINT</td>
<td>1</td>
<td>0.0004</td>
<td>0.984</td>
<td>3way06</td>
</tr>
</tbody>
</table>

**Contrasts**

| (2) vs (1): SEX.ADMINT                  | 1  | 4.7148   | 0.040|        |
| (3) vs (1): SEX.DEPT                   | 1  | 24.9815  | 0.000|        |
| (4) vs (1): ADMIT.DEPT                 | 1  | 33.8287  | 0.000|        |
| (5) vs (4): SEX.DEPT                   | 1  | 24.9816  | 0.000|        |
| (6) vs (5): SEX.ADMINT                 | 1  | 0.1666   | 0.683|        |
General Log-Linear Models

- The general log-linear model does not distinguish between independent and dependent variables. We simply study the mutual association between variables.

- Two-way table: suppose that there is a multinomial sample of size \( n \) over the \( N=ij \) cells of an \( I \times J \) contingency table for variables A and B. Suppose further that the expected frequency is denoted as \( m_{ij} \). We can specify the independence model in the following manner:

\[
m_{ij} = \eta \tau_i^A \tau_j^B
\]

Where \( \eta \) is the grand mean, \( \tau_i^A \) and \( \tau_j^B \) are marginal parameters.
General Log-Linear Model

Take natural logarithm on both sides

\[ \log m_{ij} = u + u_i^A + u_j^B \]

In order to make the above log-linear model identifiable, some restrictions are needed. Two possible restrictions are:

(a) \( \Sigma u_i^A = \Sigma u_j^B = 0 \): \( u \) is the grand mean and \( u_i^A \) and \( u_j^B \) represent deviation from the grand mean for row and column categories, respectively.

(b) \( u_1^A = u_1^B = 0 \): \( u \) is the \((1,1)\) cell and \( u_i^A \) and \( u_j^B \) are deviations from the \(1^{st}\) row and column marginals, respectively.

The independence model has \((I-1)(J-1)\) df.

ECTA and CDAS adopts the first normalization, GLIM adopts the second one, and \( \mathcal{L}_{EM} \) can use either normalization.
General Log-Linear Model

What would the odds ratios be when the independence model holds?

\[
\theta_{ij} = \frac{m_{ij} m_{i+1,j+1}}{m_{i+1,j} m_{i,j+1}} = 1
\]

\[
\log \theta_{ij} = 0
\]

Interaction Model

\[
\log m_{ij} = u + u_i^A + u_j^B + u_{ij}^{AB}
\]

\[df = 0 \text{ (saturated model).}\]

Note that when we include interaction effects, the main effects, \(u_i^A\) and \(u_j^B\), are no longer a simple transformation of row and column marginals. We cannot use the main effects to study marginal or differences in marginals unless the model is constrained in special ways (QS, see Sobel, Hout, and Duncan 1985; Hout, Duncan, and Sobel 1987 for details).
Relationship Between Log-Linear and Logit/Multinomial Logit Models

- When one variable is the response (dependent) variable, then there is a clear relationship between log-linear and logit models.
- Consider a three-way table, A x B x C where variable A is the response variable (A=1 or 2), B and C are explanatory variables.
Relationship Between Log-Linear and Logit/Multinomial Logit Models

General log-linear model:

\[ \log m_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{jk}^{AC} + u_{ik}^{BC} + u_{ijk}^{ABC} \]

If we believe in the structure, that is, the pattern of response and explanatory variables, then certain effects have to be included in the log-linear model a priori. In other words, \( u, u_i^A, u_j^B, u_k^C \), and \( u_{jk}^{BC} \) have to be there whereas \( u_{ij}^{AB}, u_{ik}^{AC} \), and \( u_{ijk}^{ABC} \) need to be tested.

Even though \( u_{jk}^{BC} \) is insignificant, it still needs to be included.
Relationship Between Log-Linear and Logit/Multinomial Logit Models

If more than two explanatory variables, then one must include the main and interaction effects between all explanatory variables.

\[ B, C, D \rightarrow B, C, D, BC, BD, CD, BCD \]

\[ B, C, D, E \rightarrow B, C, D, E, BC, BD, BE, CD, CE, DE, BCD, BDE, BCE, CDE, BCDE \]

so on and so forth. As explanatory variables, we are not at all interested in how they are interrelated with each other. [Analogy: In MR framework, we permit all X’s to be correlated, and do not care how they are related to each other, unless they are highly collinear.]
Relationship Between Log-Linear and Logit/Multinomial Logit Models

Rewrite the fully interactive model as two equations:

\[
\begin{align*}
\log m_{1jk} &= u + u_1^A + u_j^B + u_k^C + u_{1j}^{AC} + u_{1k}^{AC} + u_{jk}^{BC} + u_{jk}^{ABC} \\
\log m_{2jk} &= u + u_2^A + u_j^B + u_k^C + u_{2j}^{AC} + u_{2k}^{AC} + u_{jk}^{BC} + u_{2jk}^{ABC}
\end{align*}
\]

The difference between the two equations:

\[
\log m_{2jk} - \log m_{1jk} = (u_2^A - u_1^A) + (u_{2j}^{AB} - u_{1j}^{AB}) + (u_{2k}^{AC} - u_{1k}^{AC}) + (u_{2jk}^{ABC} - u_{1jk}^{ABC})
\]

\[= w + w_j^B + w_k^C + w_{jk}^{BC}\]

Define \(P_{jk} = P(A=2|B=j, C=k)\), then

\[
\log \left[\frac{P_{jk}}{1-P_{jk}}\right] = w + w_j + w_k + w_{jk}
\]

Log-linear model with no 3-way interaction=logit additive model

Log-linear model with 3-way interaction=logit saturated model
Relationship Between Log-Linear and Logit/Multinomial Logit Models

If we use the normalization that $\sum u_i^A = \sum u_j^B = \sum u_k^C = \sum u_{ij}^{AB} = \sum u_{ik}^{AC} = \sum u_{jk}^{BC} = \sum u_{ijk}^{ABC} = 0$, then

logit parameters = log-linear parameters $\times 2$

However, if we use the normalization that uses the first category all 1-way, 2-way, and 3-way marginal parameters as 0, then

logit parameters = log-linear parameters

By the same token, the relationship can be generalized when the response variable has more than two categories (multinomial logit model).

The choice of logit and log-linear models is a matter of convenience. Logit and multinomial logit models are generally faster to estimate as there are fewer number of parameters and can incorporate categorical as well continuous covariates.
Quasi-Independence (QI) Model

Sometimes, the lack of fit in contingency tables is because certain part of the table has association whereas independence holds for the remaining portion → Quasi-Independence Model

In the case of square tables (R x R or R x R x L), clustering on the main diagonal is common.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Create a design matrix and weigh it in the analysis and include either one

Create factors, DIAG and INH, and include either one

Note that the topological models (to be discussed later) are special cases of QI (Hauser).
Square Tables

When there is a one-to-one correspondence between row and column categories, e.g. father’s and respondent’s occupation, husband’s and wife’s characteristics, international trade flows data, etc., then the following models may be of interests.

(a) Symmetry (S)

\[
m_{ij} = m_{ji}
\]

\[
df = l(l - 1)/2
\]

(b) Quasi-Symmetry (QS)

\[
\log m_{ij} = u + u_i + u_j + \delta_{ij}\text{ where } \delta_{ij} = \delta_{ji}
\]

\[
df = (l - 1)^2 - l(l - 1)/2
\]

\[
\begin{align*}
1 & 2 & 3 & 4 & 5 \\
2 & 1 & 6 & 7 & 8 \\
3 & 6 & 1 & 9 & 10 \\
4 & 7 & 9 & 1 & 11 \\
5 & 8 & 10 & 11 & 1 \\
\end{align*}
\]

Model = SYM  
Model = A + B + SYM
Square Tables

(C) Marginal Homogeneity (MH) Model
Let $\tau_{i+}^{AB}$ represents the row marginal total for $i$-th row and $\tau_{+j}^{AB}$ represents the column marginal total for $j$-th column, then

$$\tau_{i+}^{AB} = \tau_{+j}^{AB} \text{ when } i = j, \quad df = l - 1$$

Note that MH is not a log-linear model and therefore cannot be estimated directly.
However, since S implies MH, and S implies QS,

$$S = MH + QS$$

Note that QI also implies QS.
Topological (Levels) Model

Let $H_k$ be a mutually exclusive and exhaustive partition of the pairs $(i,j)$ in which

$$\log m_{ij} = u + u_i + u_j + \delta_{ij}$$

where $\delta_{ij} = \delta_k$ for $(i,j) \in H_k$. $\delta_{ij}$ is the interaction effect and the cells $(i,j)$ are assigned to $K$ mutually exclusive and exhaustive subsets, and each of these sets shares a common interaction parameter, $\delta_k$. The number of levels ($K$) should be substantially less than the number of cells in the table. One can interpret the model as within each level of $H_k$, the model of independence holds.

$$df = (I - 1)(J - 1) - (K - 1)$$

Topological matrix:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

TOP1 = 4 4 4 5 5  TOP2 = 4 4 4 5 5

5 5 5 6 5  5 5 5 6 5

6 6 5 5 4  6 6 5 5 4
More Refined Topological Models (Overlapping)

CASMIN mobility analysis: 7 classes (I/II, III, IVa/b, IVc, V/VI, VIIa, and VIIb) with 8 topological effects to capture social mobility

Mobility Effects:
(a) Hierarchy: vertical effects (HI1 and HI2). The class structure is believed to be partition into a threefold division.
   I/II vs III, IVa/b, IVc, V/VI vs VIIa, VIIb
(b) Inheritance: IN1, IN2, and IN3
(c) Sector: intersectoral movement between three sectors (SE)
   I/II, III, IVa/b vs IVc vs V/VI, VIIa, VIIb
(d) Affinity: residual effects (AF1, AF2)

\[
\log m_{ij} = u + u_i + u_j + HI1 + HI2 + IN1 + IN2 + IN3 + SE + AF1 + AF2
\]
<table>
<thead>
<tr>
<th>I/II</th>
<th>I/II</th>
<th>III</th>
<th>IVa/b</th>
<th>IVc</th>
<th>V/VI</th>
<th>VIIa</th>
<th>VIIb</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN1+IN2</td>
<td>HI1+AF2</td>
<td>HI1+AF2</td>
<td>HI1+SE</td>
<td>HI1</td>
<td>HI1+</td>
<td>HI1+HI2+</td>
<td></td>
</tr>
<tr>
<td>HI1+AF2</td>
<td>IN1</td>
<td>-----</td>
<td>SE</td>
<td>-----</td>
<td>HI1</td>
<td>HI1+SE</td>
<td></td>
</tr>
<tr>
<td>HI1+AF2</td>
<td>----</td>
<td>IN1+IN2</td>
<td>SE+AF2</td>
<td>-----</td>
<td>HI1</td>
<td>HI1+SE</td>
<td></td>
</tr>
<tr>
<td>HI1+HI2+</td>
<td>SE</td>
<td>SE</td>
<td>+AF2</td>
<td>IN2+IN3</td>
<td>SE</td>
<td>AF2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V/VI</th>
<th>V/VI</th>
<th>VIIa</th>
<th>VIIb</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI1</td>
<td>-----</td>
<td>-----</td>
<td>SE</td>
</tr>
<tr>
<td>HI1+HI2</td>
<td>HI1</td>
<td>HI1</td>
<td>HI1+SE</td>
</tr>
<tr>
<td>HI1+HI2+</td>
<td>HI1+SE</td>
<td>HI1+SE</td>
<td>HI1</td>
</tr>
</tbody>
</table>

where __________ = hierarchical divisions (shown vertically)
-------------- = sectoral divisions (shown horizontally)

7 asymmetries: (1,4) (4,1), (2,4) (4,2), (3,4) (4,3), (4,5) (5,4), (4,6) (6,4), (4,7) (7,4), (6,7) (7,6)
Ordinal Variables? How to solve that?

(1) Impose scale restrictions: U, R, C, R+C, linear by linear association (Duncan, Goodman, Hout)

(2) Log-multiplicative scaled association models (instead of imposing scales, estimate the scores posteriori) (Goodman, Wong)

(3) Conditional odds (continuation ratio) model (Mare)

(4) Cumulative odds (ordered logit/probit) models (Winship and Mare)

(1) & (2) are for cross-classified tables only and (3) & (4) are for both cross-classified and individual data.
Uniform Association (U) Model

Let $u_i$ and $v_j$ be the scores of the row and column categories, respectively, then the uniform association model can be defined as:

$$\log m_{ij} = u + u_i^A + u_j^B + \beta u_i v_j$$

Let’s defined adjacent odds-ratios as:

$$\log \theta_{ij} = \left[ F_{ij} F_{i+1,j+1} \right] / \left[ F_{i+1,j} F_{i,j+1} \right]$$

It can be easily that

$$\log \theta_{ij} = \beta (u_{i+1} - u_i) (v_{j+1} - v_j)$$

Suppose that $u_i = i$ and $v_j = j$, then the uniform association model becomes the following:

$$\log m_{ij} = u + u_i^A + u_j^B + \beta ij$$

and the adjacent odds-ratios,

$$\log \theta_{ij} = \beta \left[ (i+1) - 1 \right] \left[ (j+1) - 1 \right] = \beta$$

$$\text{df} = (I - 1)(J - 1) - 1 = IJ - I - J$$
Uniform Association (\(U\)) Model

When \(u_i = i\) and \(v_j = j\), Goodman labels the model as \(U\). When no restrictions are imposed on \(u_i\) and \(v_j\), Goodman labels the model as \(U^0\). Therefore, \(U\) is a special case of \(U^0\).

What happens to the odds-ratios when they are two or more steps away?

2 steps: \(\log \theta_{ij}^* = \beta(u_{i+2} - u_i)(v_{j+2} - v_j) = 4 \beta\)
3 steps: \(\log \theta_{ij}^{**} = \beta(u_{i+3} - u_i)(v_{j+3} - v_j) = 9 \beta\)
4 steps: \(\log \theta_{ij}^{***} = \beta(u_{i+4} - u_i)(v_{j+4} - v_j) = 16 \beta\)
so on and so forth.
Uniform Association (U) Model

Multiple Groups:

\[
\log m_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{jk}^{AC} + \beta_k^{ij}
\]

\[
\log \theta_{ij(k)} = \beta_k
\]

\[
df = K(I - 1) (J - 1) - K = K(IJ - I - J)
\]

Additional tests can be imposed to test for whether \( \beta_k = \beta \), \( \beta_k = \beta (1 + at) \) for linear trend, and \( \beta_k = \beta (1 + at + bt^2) \) for quadratic non-linear trend.

If the grouping variable consists of more than one type of groups (country by gender, or race by cohort), more complex constraints such as \( \beta_k = \beta_x + \beta_y \) or \( \beta_k = \beta_x (1 + at) \)
Row Effects (R) Model

Let’s assume that the column categories are ordered and have scores, $v_j$, then the row effects (R) model can be defined as:

$$\log m_{ij} = u + u_i^A + u_j^B + \tau_i^A v_j$$

where $\tau_i^A$ are the row effects parameters.

$df = (I - 1)(J - 1) - (I - 1) = (I - 1)(J - 2)$

The adjacent odds-ratios can be written as:

$$\log \theta_{ij} = \tau_{i+1}^A v_{j+1} + \tau_i^A v_j - \tau_i^A v_j - \tau_{i+1}^A v_{j+1}$$

When $v_j = j$, then $v_{j+1} - v_j = 1$ and

$$\log \theta_{ij} = \tau_{i+1}^A - \tau_i^A$$
Column Effects (C) Model

Let’s assume that the row categories are ordered and have scores, $u_i$, then the column effects (C) model can be defined as:

$$
\log m_{ij} = u + u_i^A + u_j^B + \tau_j^B u_i
$$

where $\tau_j^B$ are the column effects parameters.

$$
df = (I - 1)(J - 1) - (J - 1) = (I - 2)(J - 1)
$$

The adjacent odds-ratios can be written as:

$$
\log \theta_{ij} = \tau_{j+1}^B u_{i+1} + \tau_j^B u_i - \tau_{j+1}^B u_i - \tau_j^B u_{i+1}
$$

When $u_j = i$, then $u_{i+1} - u_i = 1$ and

$$
\log \theta_{ij} = \tau_{j+1}^B - \tau_j^B
$$
Row and Column Effects (R+C) Model

Let’s assume that both row and column categories are ordered and have scores $u_i$ and $v_j$, respectively, then the log-linear row and column effects (R+C) model can be written as:

$$\log m_{ij} = u + u_i^A + u_j^B + \tau_i^A v_j + \tau_j^B u_i$$

Alternatively, one can rewrite the R+C model as:

$$\log m_{ij} = u + u_i^A + u_j^B + \tau_i^A v_j + \tau_j^B u_i + \beta u_i v_j$$

Where $\tau_1^A = \tau_i^A = \tau_1^B = \tau_j^B = 0$.

$$df = (I - 1)(J - 1) - (I - 2) - (J - 2) - 1 = (I - 2)(J - 2)$$

Similarly, the adjacent odds-ratios can be written as:

$$\log \theta_{ij} = (u_{i+1} - u_i) (\tau_{j+1}^B - \tau_j^B) + (\tau_{i+1}^A - \tau_i^A) (v_{j+1} - v_j)$$

If $u_i = I$ and $v_j = J$, then $u_{i+1} - u_i = v_{j+1} - v_j = 1$, and

$$\log \theta_{ij} = (\tau_{j+1}^B - \tau_j^B) + (\tau_{i+1}^A - \tau_i^A)$$
Equal Row and Column Effects (Equal R+C) Model

Suppose we have a square table with one-to-one correspondence between row and column categories (e.g. husband’s and wife’s characteristics), then we can further impose restrictions to row and column effects parameters to form an even more parsimonious model.

\[ \log m_{ij} = u + u_i^A + u_j^B + \tau_i^A \nu_j + \tau_j^B u_i \]

where \( \tau_i^A = \tau_j^B \) when \( i = j \) and \( u_i = i \), and \( \nu_j = j \).

\[ df = (I - 1)(J - 1) - (I - 1) = (I - 1)(I - 2) \] because \( I = J \).

If \( u_i = i \) and \( \nu_j = j \), it can be shown easily that the adjacent odds-ratios can be written as:

\[ \log \theta_{ij} = 2(\tau_{i+1}^A - \tau_i^A) = 2(\tau_{j+1}^B - \tau_j^B) \]

Note that the equal R+C model imposes symmetrical association pattern, and if it fits well, so will be the model of quasi-symmetry. That is, equal R+C implies QS.
SAT Model (Michael Hout)

\[ \log m_{ij} = u + u_i^A + u_j^B + \beta_1 S_i S_j + \beta_2 A_i A_j + \beta_3 T_i T_j + \beta_4 \delta_1 S_i^2 + \beta_5 \delta_1 A_i^2 + \beta_6 \delta_1 T_i^2 \]

Where \( S_i \) = status scores, \( A_i \) = autonomy scores, \( T_i \) = training scores, and \( \delta_1 = 1 \) for all diagonal cells, 0 otherwise.

Note that the model has \((I - 1)^2 - 6\) df and the model is to capture the multidimensional aspect of intergenerational mobility.

Later, I’ll show how to model multidimensional association by estimating the dimensions \emph{posteriori}. 

Log-Multiplicative Row and Column Effects (RC) Model

Instead of imposing row and column scores a priori, it is possible to estimate the row and column scores posteriori from the data. Consider the following model:

\[ \log m_{ij} = u + u_i^A + u_j^B + \phi \mu_i \nu_j \]

where \( \phi \) = intrinsic association parameter;
\( \mu_i \) = row score parameters; and
\( \nu_j \) = column score parameters.

The model has \((I - 2)(J - 2)\) df.

Under the RC model, the adjacent odds-ratios can be written as:

\[ \log \theta_{ij} = \phi (\mu_{i+1} - \mu_i) (\nu_{j+1} - \nu_j) \]
Log-Multiplicative Row and Column Effects \((RC)\) Model

Also \(\mu_i\) and \(\nu_j\) are subject to the following normalizations:

\[
\sum \mu_i = \sum \nu_j = 0; \quad \sum \mu_i^2 = \sum \nu_j^2 = 1.
\]

In other words, \(\mu_i\) and \(\nu_j\) are unit standardized scores. Note that other normalizations such as marginal weights are possible (see the works of Goodman and Clogg for details).

The adoption of different weights alters the point estimates but the fit statistics and the relative distance between categories are unaffected. For single table analysis, the choice between weighted and unweighted solutions is inconsequential. However, for multiple tables with different marginal distributions, the choice of weights lead to different conclusions. Standard weights (such as the unit standardized scores above) are more desirable in such circumstances.
Log-Multiplicative Row and Column Effects (RC) Model

What is the difference between $R+C$ and $RC$ models? In addition to the ability to estimate row and column distances posteriori rather than a priori, the $RC$ model has one desirable characteristic. Any interchange(s) between row and/or column categories will not affect the likelihood ratio test statistics and the estimated row and column scores. However, different results will occur under the $R+C$ model. Thus, the $RC$ model is extremely useful when the rank ordering of row and/or column categories are unknown or unclear.
Log-Multiplicative Row and Column Effects (RC) Model

Because the RC model is not log-linear but log-multiplicative, the parameters cannot be estimated directly. Goodman (1979) provides details about steps to estimate the RC model through an iterative procedure (using an EM algorithm).

The solutions can be obtained rather quickly from the EM algorithm. Unfortunately, the algorithm does not yield standard errors of the parameters as a by-product. Standard errors can be obtained by using the jackknifing procedure (see Clogg and Shihadeh for details).
Multidimensional $RC$ Association

$RC(M)$ Model

The $RC(M)$ model increases the dimensionality of the association model:

$$\log m_{ij} = u + u_i^A + u_j^B + \sum \phi_{mim} v_{jm}$$

Where $0 \leq M \leq \min(I - 1, J - 1)$. If $M^* < M$, then the model is an unsaturated one. If $M^* = M$, the model becomes saturated.

The model has $(I - M - 1)(J - M - 1)\ df$. Similar to the constraints applied to normalize row and column scores, that is, $\Sigma \mu_{im} = \Sigma v_{jm} = 0$; $\Sigma \mu_{im}^2 = \Sigma v_{jm}^2 = 1$, the following constraints also hold:

$$\Sigma \mu_{im} \mu_{im'} = \Sigma v_{jm} v_{jm'} = 0 \text{ where } m \neq m'.$$

Note that the cross-dimensional constraints will ensure that the row and column scores are orthogonal to each other. Without any loss of generality, one can rearrange $\phi_m$ such that

$$\phi_1 \geq \phi_2 \geq \ldots \geq \phi_m \geq \ldots \geq \phi_M$$
Multidimensional $RC$ Association

$RC(M)$ Model

Special cases:

1. When $M = 0$, $RC(M) \Rightarrow$ independence model
2. When $M = 1$, $RC(M) \Rightarrow RC(1)$ association model
3. When $M = 2$, $RC(M) \Rightarrow RC(2)$ association model

$RC(M)$ models can be estimated in GLIM, CDAS, and $\ell_{EM}$ (with no cross-dimensional constraints imposed)

Some extensions:

1. $U + RC$
2. $R + RC$
3. $C + RC$
4. $(R+C)+RC$
Modeling Group Differences in Association

(a) Log-linear layer effect model (LL1) (Yamaguchi, Wong)

\[ \log m_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC} + u_{ij}^{AB} + \beta_k^i j \]

with \((I - 1)(J - 1)(K - 1) - (K - 1) = (IJ - I - J)(K - 1)\) df and \(\beta_1 = 0\)
(normalization)

\[ \log \theta_{ij(k)} = \log \theta_{ij}^* + \beta_k \] . Thus, \(\beta_k\) is the deviation from the common set of odds-ratios for layer \(k\) from layer 1.

It can be shown easily that:

\[ \frac{\log \theta_{ij(k)}}{\log \theta_{ij(k)}} = \frac{\log \theta_{ij}^* + \beta_k}{\log \theta_{ij} + \beta_k^*} \]

but \(\log \theta_{ij(k)} - \log \theta_{ij} = \beta_k - \beta_k^*\).

This explains why the model is known as the log-linear uniform difference or log-linear layer effect model and it is related to the uniform association model.
(b) Log-multiplicative layer effect model (LL2)

\[ \log m_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC} + \phi_k \psi_{ij} \]

where \( \psi_{ij} \) represents the full-interaction between A and B. \( \psi_{ij} \)'s are identified by \( \sum_j \psi_{ij} = \sum_j \psi_{ij} = 0 \)

and \( \phi_k \) are identified by \( \sum \phi_k^2 = 1 \). However, it is possible to simplify \( \psi_{ij} \) as more restrictive structure to the AB association.

It can be shown easily that

\[
\frac{\log \theta_{ij(k)}}{\log \theta_{ij(k')}} = \frac{\phi_k}{\phi_{k'}} \quad \text{but}
\]

\[
\log \theta_{ij(k)} - \log \theta_{ij(k')} = \log \theta_{ij} (\phi_k - \phi_{k'})
\]

In other words, the ratio of the log-odds-ratio for some pair \((i,j)\) in one layer, \(k\), and the corresponding log-odds-ratio for that same pair \((i,j)\) in some other layer, \(k'\), is constant for all \((i,j)\) pairs \(\Rightarrow\) log-multiplicative uniform difference or layer effect model.
Modeling Group Differences in Association

(c) Modified regression-type approach (MR)

\[ \log m_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC} + u_{ij}^{AB} + \phi_k \psi_{ij} \]

Note that the only difference between the modified approach and the log-multiplicative model is the inclusion of \( u_{ij}^{AB} \) into the model.

\[ \text{df} = (IJ - 1 - J)(K - 2) \]

What’s good about this approach?

\[ \log \theta_{ij(k)} = \mu_{ij} + \mu_{ij} \phi_k \]

where \( \mu_{ij} = u_{ij}^{AB} + u_{i+1,j+1}^{AB} - u_{i+1,j}^{AB} - u_{i,j+1}^{AB} \) and

\[ \mu_{ij} = \psi_{ij} + \psi_{i+1,j+1} - \psi_{i+1,j} - \psi_{i,j+1} \]

Recall that in OLS, \( E(y|x) = \beta_0 + \beta_1 x \). The above formulation has similar form and Goodman and Hout label their model as the modified regression-type approach.
Modeling Group Differences in Association

- Note that
  - (a) \( \log \theta_{ij(k)} - \log \theta_{ij(k')} = \mu_{ij} (\phi_k - \phi_{k'}) \)
  - (b) \( \log \theta_{ij(k)} / \log \theta_{ij(k')} = [\mu_{ij} + \mu_{ij} \phi_k] / [\mu_{ij} + \mu_{ij} \phi_{k'}] \)

Neither the differences between the log-odds-ratios nor the ratios of the log-odds-ratios are uniform across different (i,j). However, the cross-layer differences in log-odds-ratios are all proportional. That is,

\[
\frac{\log \theta_{ij(k)} - \log \theta_{ij(k')}}{\log \theta_{ij(k)} - \log \theta_{ij(k')}} = \frac{\phi_k - \phi_{k'}}{\phi_k - \phi_{k'}}
\]

For \( k \neq k' \neq k^* \)
Modeling Group Differences in Association

- Merits of the regression-type approach
  - (a) $u_{ij}^{AB}$ establish the baseline pattern of association and $\psi_{ij}$ and $\phi_k$ adjust that baseline pattern for layer k.
  - (b) The baseline pattern of association, which is established by $u_{ij}^{AB}$, may or may not apply to a particular layer, depending on the parameterization of $\psi_{ij}$ and $\phi_k$. For instance, one can use full-interaction, uniform association, RC association, topological models or alike constraints either to $u_{ij}^{AB}$ or $\phi_k \psi_{ij}$ or both. This provides the best test for the invariance thesis ever existed!
Multidimensional Association Models

Conditional independence between variables $A$ and $B$, given variable $C$:

$$\log F_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC} \quad (3)$$

$$df = (I - 1)(J - 1)K$$

The full three-way interaction model:

$$\log F_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} +$$

$$u_{jk}^{BC} + u_{ijk}^{ABC} \quad (4)$$

$$df = 0.$$
Partial Association Models

Case I: Decomposition without Three-Way Interaction

Partial association model

\[
\log F_{ijk} = u + u^A_i + u^B_j + u^C_k + \phi_1^{AB} \mu_{i1} \nu_{j1} + \phi_1^{AC} \mu^*_{i1} \eta_{k1} + \phi_1^{BC} \nu^*_{j1} \eta^*_{k1}
\]  

\[df = IJK - 3I - 3J - 3K + 11. \quad (5)\]

Both centering and scaling constraints on row, column, and layer scores are needed to identify the model. I label Equation (5) as the unrestricted RC(1) + RL(1) + CL(1) model.
Partial Association Models

If we restrict $\mu_{i1} = \mu_{i1}^*$, $v_{j1} = v_{j1}^*$, and $\eta_{k1} = \eta_{k1}^*$, then the more restricted model is labeled as the restricted $RC(1)+RL(1)+CL(1)$ model with consistent score restrictions. It has $IJK - 2I - 2J - 2K + 5$ df.

The contrast of the likelihood test statistics between the two models yields a $\chi^2$ statistic with $I + J + K - 6$ df that can be used to test whether consistent score restrictions are indeed congruent with the data.
Partial Association Models

The most general form can be labeled as the $RC(M_1) + RL(M_2) + CL(M_3)$ where $M_1$, $M_2$, and $M_3$ are the dimensionality of the $AB$, $AC$, and $BC$ partial association, respectively; and $0 \leq M_1 \leq \min(I - 1, J - 1)$, $0 \leq M_2 \leq \min(I - 1, K - 1)$, $0 \leq M_3 \leq \min(J - 1, K - 1)$ and can be written as:

$$
\log F_{ijk} = u + u_i^A + u_j^B + u_k^C + \sum_{m=1}^{M_1} \phi_{im}^{AB} \mu_{im} \nu_{jm} + \sum_{m=1}^{M_2} \phi_{im}^{AC} \mu_{im} \eta_{km} + \sum_{m=1}^{M_3} \phi_{im}^{BC} \nu_{jm}^* \eta_{km}^* 
$$

(7)

$$
df = IJK - I - J - K + 2 - M_1(I + J - M_1 - 2) - M_2(I + K - M_2 - 2) - M_3(J + K - M_3 - 2).
$$
Partial Association Models

To uniquely identify the parameters, centering and scaling restrictions as well as cross-dimensional constraints on row, column, and layer score parameters within each two-way interaction are needed. Centering restrictions:

$$\sum_{i=1}^{I} \mu_{im} = \sum_{i=1}^{I} \mu_{im}^* = \sum_{j=1}^{J} \nu_{jm} = \sum_{j=1}^{J} \nu_{jm}^* = \sum_{k=1}^{K} \eta_{km} = \sum_{k=1}^{K} \eta_{km}^* = 0.$$  

Scaling and cross-dimensional constraints:

$$\sum_{i=1}^{I} \mu_{im} \mu_{im'} = \sum_{i=1}^{I} \mu_{im}^* \mu_{im'} = \sum_{j=1}^{J} \nu_{jm} \nu_{jm'} = \sum_{j=1}^{J} \nu_{jm}^* \nu_{jm'} = \sum_{k=1}^{K} \eta_{km} \eta_{km'} = \sum_{k=1}^{K} \eta_{km}^* \eta_{km'} = \delta_{mm'},$$

where $$\delta_{mm'}$$ is the Kronecker delta with $$\delta_{mm'} = 1$$ if $$m = m'$$, 0 otherwise.
Partial Association Models

Under the $RC(M_1) + RL(M_2) + CL(M_3)$ model, the conditional local odds-ratios for $A$ and $B$ given $C$, $\theta_{i(j)k}$ for $A$ and $C$ given $B$, and $\theta_{(i)jk}$ for $B$ and $C$ given $A$; and the ratio of the conditional local odds-ratios or the local odds-ratios for $A$, $B$, and $C$, $\theta_{ijk}$ have the following terms:

$$\log \theta_{ij(k)} = \sum_{m=1}^{M_1} \phi_{m}^{AB} (\mu_{i+1,m} - \mu_{im}) (\nu_{j+1,m} - \nu_{jm})$$

$$\log \theta_{i(j)k} = \sum_{m=1}^{M_2} \phi_{m}^{AC} (\mu_{i+1,m}^* - \mu_{im}^*) (\eta_{k+1,m} - \eta_{km})$$

$$\log \theta_{(i)jk} = \sum_{m=1}^{M_3} \phi_{m}^{BC} (\nu_{j+1,m}^* - \nu_{jm}^*) (\eta_{k+1,m}^* - \eta_{km}^*)$$

and

$$\log \theta_{ijk} = 0.$$
Partial Association Models

If \( M_1 = M_2 = M_3 = M \), then the \( df \) for the unrestricted \( RC(M) + RL(M) + CL(M) \) model can be simplified to
\[
\]

Problem: For some restricted \( RC(M) + RL(M) + CL(M) \) models with consistent score restrictions, not all cross-dimensional constraints are needed in order to uniquely identify these models.
Partial Association Models

Generally speaking, only one set of cross-dimensional constraints on either row, column, or layer scores will be sufficient to uniquely identify the model for restricted $RC(M) + RL(M) + CL(M)$ models with consistent score restrictions. For example, only one, not three, cross-dimensional constraint is needed on $\{\mu_{i1}, \mu_{i2}\}$, $\{v_{j1}, v_{j2}\}$, or $\{\eta_{k1}, \eta_{k2}\}$ for the restricted $RC(2) + RL(2) + CL(2)$ model with consistent scores. That is, either (i) $\sum_{i=1}^{I} \mu_{i1} \mu_{i2} = 0$, (ii) $\sum_{j=1}^{J} v_{j1} v_{j2} = 0$, or (iii) $\sum_{k=1}^{K} \eta_{k1} \eta_{k2} = 0$ can be imposed.
Partial Association Models

Similarly, for the restricted $RC(3)+RL(3)+CL(3)$ model, only three cross-dimensional constraints are needed on $\{\mu_{i1}, \mu_{i2}, \mu_{i3}\}$, $\{v_{j1}, v_{j2}, v_{j3}\}$, or $\{\eta_{k1}, \eta_{k2}, \eta_{k3}\}$, that is, either (i) $\sum_{i=1}^{I} \mu_{i1} \mu_{i2} = \sum_{i=1}^{I} \mu_{i1} \mu_{i3} = \sum_{i=1}^{I} \mu_{i2} \mu_{i3} = 0$, (ii) $\sum_{j=1}^{J} v_{j1} v_{j2} = \sum_{j=1}^{J} v_{j1} v_{j3} = \sum_{j=1}^{J} v_{j2} v_{j3} = 0$, or (iii) $\sum_{k=1}^{K} \eta_{k1} \eta_{k2} = \sum_{k=1}^{K} \eta_{k1} \eta_{k3} = \sum_{k=1}^{K} \eta_{k2} \eta_{k3} = 0$.

In general, the restricted $RC(M) + RL(M) + CL(M)$ model with consistent score restrictions in all dimensions has $IJK - I - J - K + 2 - M(I + J + K - 3) + M(M - 1)/2$ df.
Partial Association Models

For some restricted $RC(M) + RL(M) + CL(M)$ models with consistent score restrictions on a subset but not all dimensions, it is even possible that *no* cross-dimensional constraint is required.

Consider the $RC(2) + RL(2) + CL(2)$ model again, there is no need to impose any orthogonal restriction if consistent score restrictions are applied to the first dimension only.
Multidimensional Association Models

Case II. Decomposition with Three-Way Interaction
This can be further differentiated into situations where the following parameters are decomposed:
(i) the three-way interaction term \( u_{ijk}^{ABC} \);
(ii) some but not all two-way marginal parameters.
For example, when the third variable, \( C \), represents either groups, birth cohorts, or years of observation and we are only interested to decompose \( AB \) and \( ABC \) interaction \( (u_{ij}^{AB} \text{ and } u_{ijk}^{ABC}) \); and
(iii) all two-way \( (AB, AC, BC) \) and three-way \( (ABC) \) interactions \( (u_{ij}^{AB}, u_{ik}^{AC}, u_{jk}^{BC}, \text{ and } u_{ijk}^{ABC}) \).
Multidimensional Association Models

(a) Decompose $u_{ij}^{AB}$ and $u_{ijk}^{ABC}$ (Conditional Association Models)

The $RC(M)$-L model, a generalization of the $RC(M)$ model in two-way table, has been proposed to study group differences in association across layers (Becker 1989; Becker and Clogg 1989; Clogg and Shihadeh 1994). The model has the following form:

$$
\log F_{ijk} = u + u_A^i + u_B^j + u_C^k + u_{ik}^{AC} + u_{jk}^{BC} + \sum_{m=1}^{M} \phi_{mk}^{ABC} \mu_{imk} \nu_{jmk}
$$

(9)
Multidimensional Association Models

with the following constraints: $\Sigma_{i=1}^{I} \mu_{imk} = \Sigma_{j=1}^{J} v_{jmk} = 0$, and $\Sigma_{i=1}^{I} \mu_{imk} \mu_{im'k} = \Sigma_{j=1}^{J} v_{jmk} v_{jm'k} = \delta_{mm'}$ where $\delta_{mm'}$ is the Kronecker delta with $\delta_{mm'} = 1$ if $m = m'$, 0 otherwise. The model has $(I - M - 1)(J - M - 1)K$ df. I label equation (9) as the heterogeneous RC(M)-L model.

The homogeneous RC(M)-L model that imposes equality restrictions in all $\phi_{mk}$, $\mu_{imk}$, and $v_{jmk}$ parameters across layers reduces Equation (9) as:

$$\log F_{ijk} = u + u_{i}^{A} + u_{j}^{B} + u_{k}^{C} + u_{ik}^{AC} + u_{jk}^{BC} + \Sigma_{m=1}^{M} \phi_{m}^{ABC} \mu_{im} v_{jm}$$

(11)
Multidimensional Association Models

Again, some partial homogeneous or partial heterogeneous models may not require cross-dimensional restrictions whereas others do.

Consider the model with homogeneity restrictions on $\mu_{imk}$ and $\nu_{jmk}$. The partial homogeneous RC(M)-L with equal score restrictions can be written as:

$$
\log F_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC} + \sum_{m=1}^{M} \Phi_{mk}^{ABC} \mu_{im} \nu_{jm} \quad (12)
$$

Note that there is no need to impose any cross-dimensional constraints on this model because the
Multidimensional Association Models

converged parameters estimates are rotationally unique. Therefore, the model has \((I - 1)(J - 1)K - (I + J + K - 4)M\) df.

Why equation (12) does not require cross-dimensional constraints? Rewrite the model as:

\[
\log F_{ijk} = u + u^A_i + u^B_j + u^C_k + u^{AC}_{i,k} + u^{BC}_{j,k} + \sum_{m=1}^{M} \Phi_{m}^{ABC} \mu_{im} \nu_{jm} \eta_{km}
\]  

(13)

where \(\sum_{i=1}^{I} \mu_{im} = \sum_{j=1}^{J} \nu_{jm} = 0\) and \(\sum_{i=1}^{I} \mu_{im}^2 = \sum_{j=1}^{J} \nu_{jm}^2 = \sum_{k=1}^{K} \eta_{km}^2 = 1\).
Multidimensional Association Models

This model bears a strong resemblance of the log-trilinear model using the parallel factor or canonical decomposition method for metric data in the psychometric literature. It is well known that solutions obtained from the PARAFAC/CANDECOMP decomposition are unique and require no rotational restrictions. I label equation (13) as the $(AB + ABC)$ PARAFAC/CANDECOMP $RCL(M)$ model where $(AB + ABC)$ highlights the terms under decomposition and $RCL$ highlights the log-trilinear relationship between row, column, and layer variables.
Multidimensional Association Models

The \((AB + ABC)\) Tucker 3-mode \(RCL(R, S, T)\) model can be written as:

\[
\log F_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC} + \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} \phi_{rst}^{ABC} \mu_{ir} \nu_{js} \eta_{kt} \tag{14}
\]

where \(\sum_{i=1}^{I} \mu_{ir} = \sum_{j=1}^{J} \nu_{js} = 0\); \(\sum_{i=1}^{I} \mu_{ir}^2 = \sum_{j=1}^{J} \nu_{js}^2 = \sum_{k=1}^{K} \eta_{kt}^2 = 1\); \(\sum_{i=1}^{I} \mu_{ir} \mu_{ir'} = \sum_{j=1}^{J} \nu_{js} \nu_{js'} = \sum_{k=1}^{K} \eta_{kt} \eta_{kt'} = 0\); and \(\sum_{s=1}^{S} \sum_{t=1}^{T} \phi_{rst} \phi_{rst'} = \sum_{r=1}^{R} \sum_{t=1}^{T} \phi_{rst} \phi_{rst'} = \sum_{r=1}^{R} \sum_{s=1}^{S} \phi_{rst} \phi_{rst'} = 0\) where \(r \neq r'\), \(s \neq s'\), and \(t \neq t'\).

Note that centering restriction on \(\eta_{kt}\) is not required because of the omission of \(u_{ij}^{AB}\) term in the model.
Multidimensional Association Models

The set of restrictions on the $\phi_{rst}^{ABC}$ terms is known as the principal component rotation restrictions. Since there is no confusion in the $\phi_{rst}^{ABC}$ terms, I will write them as $\phi_{rst}$ interchangeably. The model has $(I - 1)(J - 1)K - R(I - R - 1) - S(J - S - 1) - T(K - T) - RST \ df$. 
Multidimensional Association Models

Relationship between PARAFAC/CANDECOMP $RCL(M)$ and Tucker 3-mode $RCL(M)$:
(a) Equivalence between PARAFAC/ CANDECOMP $RCL(1)$ and Tucker 3-mode $RCL(1)$ and between PARAFAC/CANDECOMP $RCL(M)$ and the restricted Tucker 3-mode $RCL(M, M, M)$ with non-zero values in the superdiagonal cells.
(b) The equivalence of PARAFAC/CANDECOMP $RCL(2)$ model with Tucker 3-mode $RCL(2, 2, 2)$. The two models yield identical degrees of freedom, goodness-of-fit statistics, and expected frequencies,
Multidimensional Association Models

but their parameter estimates differ because of different normalization.
(c) The restricted Tucker $RCL(3, 3, 3)$ model with all six unique $\phi_{rst}$ terms ($\phi_{123}, \phi_{132}, \phi_{213}, \phi_{231}, \phi_{312},$ and $\phi_{321}$) is equivalent to the PARAFAC/CANDECOMP $RCL(3)$ model.
Multidimensional Association Models

Steps to determine whether cross-dimensional restrictions are needed:
(a) estimate the model without cross-dimensional constraints in the iterative stage, record the log-likelihood test statistic ($L^2_0$) and parameter estimates ($\hat{\beta}_0$).
(b) estimate the model again with different (random) start values and record the log-likelihood test statistics ($L^2_i$) and parameter estimates ($\hat{\beta}_i$). If the test statistics and parameter estimates remain the same, no cross-dimensional constraint is needed and the estimates are rotationally unique. However, if only the test statistics remain the same whereas the parameter estimates differ, they imply that some cross-dimensional constraints are needed.
(c) To locate specific cross-dimensional restrictions, compile a list of potential candidates. Add only one restriction in the iterative stage and compare the converged test statistics, $L^2_i$, with $L^2_0$. If $L^2_i = L^2_0$, this particular cross-dimensional restriction is needed. Otherwise, proceed with another restriction until all potential restrictions have been tested.
(d) Based on the results from step (c), include two "required" restrictions simultaneously and then compare the test statistics $L^2_i$ with $L^2_0$. If the results are identical, both restrictions are required. Proceed to incorporate additional restrictions incrementally until the list is completely exhausted. This incremental procedure is necessary because it is possible that for some models, cross-dimensional restrictions can be imposed individually on either row, column, or layer score parameters but not simultaneously.
(e) Finally, adjust the degrees of freedom accordingly by the number of valid cross-dimensional restrictions obtained from step (d).