A very short introduction to Mokken Scaling
Edps/Soc 584 & Psych 594

Carolyn J. Anderson

Department of Educational Psychology

Illinois
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Spring 2013
Overview

- First proposed by Mokken (1971)
- Non-parametric scaling procedure for dichotomous & polytomous items.
- Many parametric IRT models are special cases of Mokken.
- Mainly used for scaling test and questionnaire data.
- For our purposes, we’ll use it for item analysis.

References:

Components

- An automated item selection procedure (AISP)
  - Partitions of a set of ordinal items into scales or “Mokken Scale”.
  - Some items might not be selected. These are “unscalable”.

- Methods to check goodness-of-fit of non-parametric IRT model for each Mokken Scale or the strength of scales.
Notation:

- $\theta_a$ is a value of the latent variable.

- $X_j$ is item $j$ and $x$ is the response to variable $j$.

- $P(X_j = x_j | \theta)$ equals to probability that response $x_j$ is made to $i$ given the value on the latent variable.
Assumptions

The “Monotone Homogeneity Model”:

- **Unidimensionality**: There is only one latent being measured (i.e., one underlying variable explaining association between responses to items).

- **Local independence**.

\[
P(X_1 = x_1, X_2 = x_2, \ldots, X_J = x_J | \theta) = \prod_{j=1}^{J} P(X_j = x_j | \theta)
\]

- **Monotonicity**: 

\[
P(X_1 \geq x | \theta) \text{ is a nondecreasing function of } \theta
\]
Scalability Coefficients

These are computed for pairs of items (i.e., $H_{ij}$), each item (i.e., $H_i$) and the total (i.e., $H$)

The item pair scalability coefficients are

- Based on the frequencies of responses in a cross-classification of responses to two items.
- Coefficients are weighted by responses inconsistent with Guttman’s model; that is, “Guttman errors”.
- Are defined as
  \[
  H_{ij} = \frac{\text{cov}(X_i, X_j)}{\text{cov}(X_i, X_j)_{\text{max}}}
  \]

- $-\infty \geq H_{ij} \geq 1$
- If there are no Guttman errors, then $H_{ij} = 1$.
- The Monotone model implies that $0 \geq H_{ij} \geq 1$.
- If $H_{ij} < 0$, then an item does not fit the model.
Scalability Coefficient for Item and Scale

- The item scalability coefficient $H_i$ equals

$$H_i = \frac{\text{cov}(X_i, R(i))}{\text{cov}(X_i, R(i))^{\text{max}}}$$

where $R(i)$ is a rest-score.

- The test-scalability coefficient is $H$

$$H = \frac{\sum_{i=1}^{l} \text{cov}(X_i, R(i))}{\sum_{i=1}^{l} \text{cov}(X_i, R(i))^{\text{max}}}$$

C.J. Anderson (Illinois)
Rules of Thumb

- If $H < .3$ the items are unscalable.
- If $.3 \geq H \geq .4$, a scale is weakly scalable.
- If $.4 \geq H \geq .5$, a scale is moderately scalable.
- If $.5 \geq H$, a scale is strongly scaleable.
R for Victimization

R code:

1. Install `mokken` package from your favorite site

2. `library(mokken)`

3. Read in data.

4. `X = cbind(v1, v2, v3, v4)`

5. `coefh(X)`
## Output for Victimization

### Hij

<table>
<thead>
<tr>
<th></th>
<th>v1</th>
<th>se</th>
<th>v2</th>
<th>se</th>
<th>v3</th>
<th>se</th>
<th>v4</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1</td>
<td>0.891</td>
<td>(0.026)</td>
<td>0.771</td>
<td>(0.040)</td>
<td>0.541</td>
<td>(0.066)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v2</td>
<td>0.891</td>
<td>(0.026)</td>
<td>0.843</td>
<td>(0.030)</td>
<td>0.530</td>
<td>(0.070)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v3</td>
<td>0.771</td>
<td>(0.040)</td>
<td>0.843</td>
<td>(0.030)</td>
<td>0.507</td>
<td>(0.067)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v4</td>
<td>0.541</td>
<td>(0.065)</td>
<td>0.530</td>
<td>(0.070)</td>
<td>0.507</td>
<td>(0.067)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Hi

<table>
<thead>
<tr>
<th>Item</th>
<th>H</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1</td>
<td>0.759</td>
<td>(0.030)</td>
</tr>
<tr>
<td>v2</td>
<td>0.784</td>
<td>(0.027)</td>
</tr>
<tr>
<td>v3</td>
<td>0.732</td>
<td>(0.033)</td>
</tr>
<tr>
<td>v4</td>
<td>0.526</td>
<td>(0.061)</td>
</tr>
</tbody>
</table>
Partitioning into Scales

Algorithm for the automated Item Selection Procedure or AISP.

Two ones that have been implemented in mokken are

- Hierarchical clustering algorithm:
  - Stars by taking 2 items having largest value of $H_{ij}$ (significantly different from 0).
  - Items adding that meet criteria until no more can be added.
  - Takes unselected items are starts another Mokken scale
  - Continues until no more scales can be created.

- Genetic algorithm
R for Partitioning into Scales

Using victimization and fight items:

1. Take a look at these: coefH(vfX)
2. Using the default (HCA): hca ¡- aisp(vfX)
3. Using the genetic algorithm: ga ¡- asip(vfX, search=”ga”)
4. Combine to can compare: cbind(hc,ga)

Demonstrate in R.... Do we have 3 scales?