The method of Comparative Judgement for assessment and research
And data analysis in R
San Verhavert
Sven De Maeyer
Vincent Donche
san.verhavert@uantwerpen.be
www.d-pac.be

What?
• What is D-PAC?
• What is Comparative Judgement (CJ)?

How?
• Analysis in R
• Applications
• Research in D-PAC: in short

What is D-PAC NOT!
What is D-PAC?

• Digital Platform for the Assessment of Competences

1. Creating awareness
2. Developing a tool
3. Providing feedback

What is D-PAC?

• What is Comparative Judgement (CJ)?
  – Terminology
  – Theoretical background
• Analysis in R
• Applications
• Research in D-PAC: in short

Terminology

• Comparative Judgement or CJ
• Assessment
• Competence
• Assessor, judge: individual who assesses
• Assessee: individual who is assessed
• Representation [of competence]: product that is assessed
Comparative Judgement (CJ)

• How do we assess?
• Assessment is difficult!
  – Consistent: over time, assessors, ...
• Making an absolute judgement is difficult or even impossible (Laming, 1990)
• Every judgement is a comparison (Laming, 2003)

Comparative Judgement

• Every judgement is a comparison (Laming, 2003)
• Thurstone (1927): The Law of Comparative Judgement

\[ X_{AB} = v_A - v_B \]

\[ p(A > B | v_A, v_B, \sigma_{AB}^2) = p(X_{AB} = 1 | v_A, v_B, \sigma_{AB}^2) \]

\[ = \frac{1}{\sqrt{2\pi} \sigma_{AB}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} dx \]

Comparative Judgement

• Every judgement is a comparison (Laming, 2003)
• Thurstone (1927): The Law of Comparative Judgement
• Bradley-Terry-Luce model (BTL model) (Bradley & Terry, 1952; Luce, 1959)
• Rasch model (Rasch, 1960)

\[ p(A > B | v_A, v_B) = p(X_{AB} = 1 | v_A, v_B) = \frac{\exp(v_A - v_B)}{1 + \exp(v_A - v_B)} \]

\[ p(x_{ij} = 1 | \theta_i, \theta_j) = \frac{\exp(\theta_i - \theta_j)}{1 + \exp(\theta_i - \theta_j)} \]

(Andrich, 1978)
Comparative Judgement

- Writing and IQ testing: Thurstone, 1925; 1927
- Language assessment: Pollitt & Murray 1995

- Alternative assessment method (Pollitt, 2004; 2008)
- Facilitating evaluation of open-ended tasks (Jones, Swan & Pollitt, 2015)

- Inter-rater reliability $r = 0.86$ (Jones & Inglis, 2015)
- Efficiency
• What is D-PAC?
• What is Comparative Judgement (CJ)?
• Analysis in R
  – The data structure
  – Estimating the BTL model in R
  – A simulation and estimation results
• Applications
• Research in D-PAC: in short
Estimating Bradley-Terry-Luce Model in R

- **Our own function**
  - Conditional Maximal Likelihood (CML) procedure (Rasch, 1960; Andersen, 1970)
    - Sum scores as initial estimates
    - Iteratively updated: Newton-Raphson procedure
Estimating Bradley-Terry-Luce Model in R

- Our own function
  - CML
- btm function from sirt (Robitzsch, 2016)
  - Minorization-Maximization procedure (Hunter, 2004)

- BTm function from BradleyTerry2 (Turner & Firth, 2012)
  - glm function
  - Least Squares Maximum Likelihood
- vglm and Brat functions from VGAM (Yee, 2010)
  - glm like function
  - Least Squares Maximum Likelihood

vglm from VGAM

vglm( Brat( DataMat ) ~ 1,
    brat( refgp = refCatNo, refvalue = 1 ),
    trace = T )

<table>
<thead>
<tr>
<th></th>
<th>Paper 1</th>
<th>Paper 2</th>
<th>Paper 3</th>
<th>Paper 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper 1</td>
<td>NA</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Paper 2</td>
<td>2</td>
<td>NA</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Paper 3</td>
<td>0</td>
<td>2</td>
<td>NA</td>
<td>10</td>
</tr>
<tr>
<td>Paper 4</td>
<td>10</td>
<td>8</td>
<td>0</td>
<td>NA</td>
</tr>
</tbody>
</table>
Estimating Bradley-Terry-Luce Model in R

- Our own function
  - CML
- btm function from sirt (Robitzsch, 2016)
- BTm function from BradleyTerry2 (Turner & Firth, 2012)
  - glm function
    - Least Squares Maximum Likelihood
- vglm and brat functions from VGAM (Yee, 2010)
  - glm like function
    - Least Squares Maximum Likelihood
  
llbtPC.fit function in prefmod, btmodel function in psychotools, ...

Bias

Estimating Bradley-Terry-Luce Model in R

- Our own function
  - CML
- btm function from sirt (Robitzsch, 2016)
- BTm function from BradleyTerry2 (Turner & Firth, 2012)
  - glm like function
    - Least Squares Maximum Likelihood
  - vglm and brat functions from VGAM (Yee, 2010)
    - Least Squares Maximum Likelihood

Bias

Unbound likelihood

Scale '0'

Estimating Bradley-Terry-Luce Model in R

- Our own function
  - ε bias correction
- btm function from sirt (Robitzsch, 2016)
  - ε bias correction
  - bias reduction by function modification
- BTm function from BradleyTerry2 (Turner & Firth, 2012)
- vglm and brat functions from VGAM (Yee, 2010)
A simulation and estimation results

- N infinity: 35
- Mean iterations:
  - Own no BC/BC: 4
  - Sirt no BC/BC: 100
  - BT2 no BR: 10
  - BT2 BR: 1
  - VGAM: 9
- Reliability

Mean reliability* (1 SD error bars)

(* only rank orders without infinity included; n=65)
A simulation and estimation results

- N infinity: 35
- Mean iterations:
  - Own no BC/BC: 4
  - Sirt no BC/BC: 100
  - BT2 no BR: 10
  - BT2 BR: 1
  - VGAM: 9
- Reliability
- Bias

Mean bias (1 SD error bars)

<table>
<thead>
<tr>
<th>SE=</th>
<th>0.52</th>
<th>0.54</th>
<th>0.69</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>own no BC</td>
<td>own BC</td>
<td>sirt no BC</td>
<td>sirt BC</td>
<td>BT2 no BR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SE=</th>
<th>0.52</th>
<th>0.54</th>
<th>0.69</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>own no BC</td>
<td>own BC</td>
<td>sirt no BC</td>
<td>sirt BC</td>
<td>BT2 no BR</td>
</tr>
</tbody>
</table>

A simulation and estimation results

- Conclusion
  - Our own function
  - btm from sirt with c bias correction
• What is D-PAC?
• What is Comparative Judgement (CJ)?
• Analysis in R
• Applications
• Research in D-PAC: in short

Applications
Within the D-PAC project

Assessments
- Writing skills
- Reporting and reasoning
- ER models
- Mindboards

Jury’s
- Selection based on rank order
- Consultation/meeting based on rank order
- Accountability (FB)

Research
26 try-outs

Assessment tool
10 try-outs

Learning tool
11 try-outs

Professionalization
1 try-out

Data collection and Scale development:
- Input PhD’s
- Writing skills
- Reading skills
- Visual skills
- Mathematical problem solving
- Speech (audiology)
- Number estimation
- Delphi study

Peer assessment
- Learning effect of comparing
- Learning effect of FB

(Maarten Goossens)
Research in D-PAC: in short

1. CJ method: reliability, efficiency and validity
2. Feedback to assessees
3. Feedback to organizations
4. Feedback to assessors
5. Design research: userfriendly?

Thank you!
San Verhavert
san.verhavert@uantwerpen.be

D-PAC

"Because human judgements are comparisons"
Thurstone's Law of Comparative Judgment

Thurstone's Law of Comparative Judgment

Thurstone's Law of Comparative Judgment

Thurstone's Law of Comparative Judgment
Thurstone's Law of Comparative Judgment

\[ p(A > B | \nu_a, \nu_b, \sigma_{ab}^2) = p(\nu_{ab} = 1 | \nu_a, \nu_b, \sigma_{ab}^2) \]

\[ = \frac{1}{\sqrt{2\pi}\sigma_{ab}} \int_{-\infty}^{\infty} \exp\left( -\frac{(X - \mu)^2}{2\sigma^2} \right) \, dx \]

\[ X_{ab} = \frac{\nu_a - \nu_b}{\sigma} \]

\[ X_{ab} = \nu_a - \nu_b \]

(Thurstone, 1927)

Bradley-Terry-Luce Model (BTL)

\[ p(A > B | \nu_a, \nu_b) = p(\nu_{ab} = 1 | \nu_a, \nu_b) = \frac{\exp(\nu_a - \nu_b)}{1 + \exp(\nu_a - \nu_b)} \]

\[ p(\nu_{ab} = 1 | \nu_a, \nu_b) = \frac{\exp(\nu_a - \nu_b)}{1 + \exp(\nu_a - \nu_b)} \]

\[ X_{ab} = \nu_a - \nu_b \]

(Paech, 1960) (Andrich, 1978)

Difference between methods

(own v
sirt no BC
own v
sirt BC
own v
BT2 no BR
own v
BT2 BR
vgam
sirt v
BT2 no BR
vgam
BT2 v)

Rasch, 1960)

(Andrich, 1978)