Discrimination Tests With Sureness: Thurstonian and R-Index Analysis

Graham Cleaver
Consumer Perception & Behaviour
Unilever Food & Health Research Institute Vlaardingen

Sensometrics
July 2008
# Unilever Categories & Brands

## Foods
- Savoury & Dressings
- Spreads
- Weight Management
- Tea
- Ice Cream

## Home and Personal Care
- Skin
- Deodorants
- Laundry #1 in D&E
- Daily Hair Care #1 in D&E
- Household Care
- Oral Care

## Our 12 €1 billion brands
- Knorr
- Sunsilk
- Lipton
- Blue Band
- Becel
- Hellmann’s
- Dove
- Surf
- Rexona
- Lux
- Dirt is Good

World Number 1
World Number 2
Local Strength
Content

A / Not-A Test With Sureness
- R-Index Calculation
- Thurstonian Analysis

Power Analysis
- Alternative R-Index Analyses
- vs Thurstonian

Power Charts
- Difference Test
- Similarity Test

Issues & New Developments
- Difference Between Individuals
- Alternative Analysis
A / Not-A Test Procedure

Familiarisation Of Control (‘A’)

Blind Product 1
Is It An ‘A’ or ‘Not A’? How Sure?

Re-familiarisation with ‘A’

Optional

Blind Product 2
Is It An ‘A’ or ‘Not A’? How Sure?

Re-familiarisation with ‘A’

and so on...

<table>
<thead>
<tr>
<th>Product Identity</th>
<th>‘A’</th>
<th>Response</th>
<th>‘Not A’</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sure</td>
<td>Not Sure</td>
<td>Guess</td>
<td>Guess</td>
</tr>
<tr>
<td>Control (A)</td>
<td>4</td>
<td>11</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Test (Not A)</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

R Index is the percentage of all possible ‘theoretical’ pairwise comparisons between each Control and each Test product that would be classified correctly.

Pure Chance: R Index = 50%

Perfect Discrimination: R Index = 100%

In this example, R Index = 82%
1. R-Index as binomial proportion
   Bi J. & O’Mahony M. (1995) Table for testing the significance of the R-Index
   Journal of Sensory Studies 10 (4) , 341–347

2. R-Index equivalent to Mann-Whitney U Statistic, with correction for ties
   Mann H.B. & Whitney D.R. (1947) On a test of whether one of two random
   variables is stochastically larger than the other. Ann. Math. Stat. 18, 50-56

3. R-Index with revised variance estimator
   Bi J. & O’Mahony M. (2007) Updated and extended table for testing the
   significance of the R-Index   Journal of Sensory Studies 22 (6) , 713-720

Aim:
- Establish which is most appropriate / most powerful significance test for the R-Index
- Compare with Thurstonian analysis in terms of power
- Create power curves for difference testing and similarity testing
### Thurstonian Model

#### Product Identity

<table>
<thead>
<tr>
<th>Response</th>
<th>'A'</th>
<th>'Not A'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sure</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Not Sure</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Guess</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Guess</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Not Sure</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Sure</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Boundary Criteria

![Diagram showing two overlapping normal distributions with a 'd' indicating the distance between them and labeled as 'Boundary Criteria'.]
### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>D F</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-1.1642</td>
<td>0.2832</td>
<td>16.9001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>-0.0437</td>
<td>0.2175</td>
<td>0.0404</td>
<td>0.8407</td>
</tr>
<tr>
<td>Intercept</td>
<td>3</td>
<td>0.6879</td>
<td>0.2297</td>
<td>8.9681</td>
<td>0.0027</td>
</tr>
<tr>
<td>Intercept</td>
<td>4</td>
<td>1.3468</td>
<td>0.2595</td>
<td>26.9415</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>5</td>
<td>2.4484</td>
<td>0.3499</td>
<td>48.9596</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>PRODUCT</td>
<td>1</td>
<td>-1.3401</td>
<td>0.2971</td>
<td>20.3437</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**SAS Code**

```sas
DATA DAT;
    INPUT PRODUCT SURENESS COUNT;
    LINES;
    0 1 4
    0 2 11
    0 3 8
    0 4 4
    0 5 2
    0 6 1
    1 1 0
    1 2 2
    1 3 5
    1 4 8
    1 5 12
    1 6 3
    ;
RUN;

PROC LOGISTIC DATA=DAT;
    WEIGHT COUNT;
    MODEL SURENESS=PRODUCT
        / AGGREGATE SCALE=NONE
        LINK=PROBIT;
RUN;
```

**Boundary Criteria**

[Diagram showing boundary criteria with d']
PROC NLIN DATA=DAT;
   PARMS C1=-2 D1=0.5 D2=0.5 D3=0.5 D4=0.5 B=1.0 A=0.05;
   BOUNDS D1>0, D2>0, D3>0, D4>0, B>0;
   AX=EXP(A*STIM);
   IF SURENESS=1 THEN DO;
      ZJ=(C1-B*STIM)*AX;
      MODEL RP=PROBNORM(ZJ);
   END;
   IF SURENESS>1 AND SURENESS<NC THEN DO;
      IF SURENESS=2 THEN DO;
         KJ=C1+D1;
         K0=C1;
      END;
      IF SURENESS=3 THEN DO;
         KJ=C1+D1+D2;
         K0=C1+D1;
      END;
      IF SURENESS=4 THEN DO;
         KJ=C1+D1+D2+D3;
         K0=C1+D1+D2;
      END;
      IF SURENESS=5 THEN DO;
         KJ=C1+D1+D2+D3+D4;
         K0=C1+D1+D2+D3;
      END;
      ZJ=(KJ-B*STIM)*AX;
      Z0=(K0-B*STIM)*AX;
      PJ=PROBNORM(ZJ);
      P0=PROBNORM(Z0);
      MODEL RP=PJ-P0;
   END;
   IF SURENESS=6 THEN DO;
      KJ=C1+D1+D2+D3+D4;
      ZJ=(KJ-B*STIM)*AX;
      MODEL RP=1-PROBNORM(ZJ);
   END;
   WEIGHT_=NT/MODEL.RP;
   DEV=-2*NR*LOG(MODEL.RP);
   IF RP>0 AND RP<1 THEN
      DEV=DEV+2*NR*LOG(RP);
   _WEIGHT_=DEV/WEIGHT_;
   PR=MODEL.RP;
RUN;

<table>
<thead>
<tr>
<th>Obs</th>
<th>STIM</th>
<th>SURENESS</th>
<th>NR</th>
<th>NT</th>
<th>RP</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>50</td>
<td>0.22</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>19</td>
<td>50</td>
<td>0.38</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
<td>10</td>
<td>50</td>
<td>0.20</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>50</td>
<td>0.14</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>50</td>
<td>0.04</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>50</td>
<td>0.02</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>50</td>
<td>0.08</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>50</td>
<td>0.12</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>50</td>
<td>0.18</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>50</td>
<td>0.24</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>50</td>
<td>0.18</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>6</td>
<td>10</td>
<td>50</td>
<td>0.20</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Approx Std Error</th>
<th>Approximate 95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>-0.7355</td>
<td>0.0884</td>
<td>-0.9627 to -0.5083</td>
</tr>
<tr>
<td>D1</td>
<td>0.9346</td>
<td>0.0783</td>
<td>0.7334 to 1.1357</td>
</tr>
<tr>
<td>D2</td>
<td>0.6266</td>
<td>0.0641</td>
<td>0.4618 to 0.7914</td>
</tr>
<tr>
<td>D3</td>
<td>0.7520</td>
<td>0.0849</td>
<td>0.5337 to 0.9703</td>
</tr>
<tr>
<td>D4</td>
<td>0.6445</td>
<td>0.0988</td>
<td>0.3906 to 0.8983</td>
</tr>
<tr>
<td>B</td>
<td>1.1825</td>
<td>0.1281</td>
<td>0.8532 to 1.5117</td>
</tr>
<tr>
<td>A</td>
<td>-0.2451</td>
<td>0.0899</td>
<td>-0.4763 to -0.0139</td>
</tr>
</tbody>
</table>

\[d' = 1.18\]
Power Depends On Scale Usage

Power Depends On Scale Usage

Boundary criteria
- evenly separated
- close together
- far apart
Power Determination: Simulation

Select ‘true’ underlying product difference

Simulate set of boundary criteria

Simulate data set of n replicates of each product based on Thurstonian model

Thurstonian Analysis d’  R-Index Analysis 1  R-Index Analysis 2  R-Index Analysis 3

Significant?  Significant?  Significant?  Significant?

Power  Power  Power  Power

Distribution of Power  Distribution of Power  Distribution of Power  Distribution of Power

Power Curve  Power Curve  Power Curve  Power Curve

Repeat x1000  Repeat x100

Repeat Varying True Difference and Varying Replication n
To detect difference of R-Index=65% (d'=0.6)
40 replicates per product
Sign level = 0.05
One-sided test

Box plot shows mean, middle 50%, middle 90% and range
No product difference: R-Index=50% (d’=0.0)
40 replicates per product
Sign level = 0.05
One-sided test

Power Comparison

Analysis Method
Thurstonian  R Index (U Stat)  R Index (Bi2007)  R Index (Bi1995)

Significance level = 0.05
Power Curve Comparison

To detect difference of R-Index=65% (d’=0.6)
Sign level = 0.05
One-sided test

Thurstonian Analysis
R-Index (U Test)
R-Index (Bi 2007)
R-Index (Bi 1995)
To detect difference of R-Index=65% (d’=0.6)
Sign level = 0.05
One-sided test

Thurstonian Analysis
R-Index (U Test)
R-Index (Bi 2007)
R-Index (Bi 1995)

Power = 0.8
Study Objective

<table>
<thead>
<tr>
<th>Product Improvement</th>
<th>Sensory Differentiation vs Competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim</td>
<td>Demonstrate that the product difference is significant</td>
</tr>
<tr>
<td>Difference test</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changing the formulation for cost-savings</th>
<th>Changing suppliers of raw materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim</td>
<td>Demonstrate, with confidence, that the product difference is less than pre-specified value</td>
</tr>
<tr>
<td>Similarity test</td>
<td></td>
</tr>
</tbody>
</table>
40 reps per product will give an 80% chance detecting a ‘true’ R Index of 65% as significant.

60 reps per product will give a 65% chance detecting a ‘true’ R Index of 60% as significant.

If the ‘true’ R Index is 60% then 30 reps per product will give a 40% chance of getting a significant difference.
If (a) the ‘action standard’ is: demonstrate with confidence that the R Index is less than 65% and (b) We estimate that really is no difference between the products (‘True’ R Index=50) then 60 reps per product will give 90% chance of being able to conclude that the products are ‘acceptably similar’.
If (a) the ‘action standard’ is: demonstrate with confidence that the R Index is less than 60%

and (b) We estimate that really is a small difference between the products (‘True’ R Index=55)

Then 600 reps per product are required to give 90% chance of being able to conclude that the products are ‘acceptably similar’
### Individual vs Pooled Analysis

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sure</th>
<th>Not Sure</th>
<th>Guess</th>
<th>Same as Reference</th>
<th>Different to Reference</th>
<th>Total</th>
<th>R Index</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>78.5</td>
</tr>
<tr>
<td>Test</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>76.5</td>
</tr>
<tr>
<td>Ref</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>76.5</td>
</tr>
<tr>
<td>Test</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>10</td>
<td>76.5</td>
</tr>
<tr>
<td>Ref</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>84.0</td>
</tr>
<tr>
<td>Test</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>84.0</td>
</tr>
<tr>
<td>Ref</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>77.0</td>
</tr>
<tr>
<td>Test</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>77.0</td>
</tr>
<tr>
<td>Ref</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>84.0</td>
</tr>
<tr>
<td>Test</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>84.0</td>
</tr>
<tr>
<td>Pooled</td>
<td>4</td>
<td>15</td>
<td>11</td>
<td>12</td>
<td>7</td>
<td>1</td>
<td>50</td>
<td>73.9</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>3</td>
<td>11</td>
<td>13</td>
<td>16</td>
<td>7</td>
<td>50</td>
<td>73.9</td>
</tr>
</tbody>
</table>

Mean = 80.0  SEM = 5.67  
\[ z = 5.29 \text{ ***} \]

Mean = 73.9  SEM = 5.7  
\[ z = 4.21 \text{ ***} \]
Alternative Analyses

Pooled Analysis

- *Most suited to analysis of one individual, pooling over replicate assessments*
- *Or where it can be safely assumed that individual assessors are the same and using the scale in the same way*

Individual Analysis For Each Assessor Before Pooling

- *Protects against bias caused by differences in usage of scale*
- *Requires some replication of each product per assessor*
- *Initial results indicate 4+ reps per assessor*

New Directions

- *Aim: Overall analysis that allows for individual differences in sensitivity and boundary criteria, and influences of other factors*
  *Poster P13: A statistical model for A-Not A data with and without sureness. R.H.B. Christensen G. J. Cleaver and P. B. Brockhoff*
Summary

R-Index
- Important to use the correct analysis for significance test
- Recommended: Test based on U-Statistic with ties
- Or test in Bi (2007) if table of critical values required

Thurstonian Analysis
- Based on assumption about underlying nature of perception
- More insights on perceptual difference and scoring
- Similar power to R-Index analysis

Power Charts
- To be used according to objective of study
  * Difference Test
  * Similarity Test

New Directions
- Potential bias with simple pooling of data
- Analysis based on individual R-Index values safer
- New Generalised NL Model – see Poster (Christensen et al)