A new statistic to detect segmentation or unequal variance in 2-Alternative Choice (2-AC) testing

Rune H B Christensen\textsuperscript{1,*} John M Ennis\textsuperscript{2} Daniel M Ennis\textsuperscript{2} Per B Brockhoff\textsuperscript{1}

\textsuperscript{1}DTU Informatics, IMM, Section for Statistics, Technical University of Denmark
\textsuperscript{2}The Institute for Perception, Richmond, VA, USA

*Contact author: rhbc@imm.dtu.dk

July 13th 2012
Paired preference testing

2 products:

A  Chocolate bar (standard)

B  Chocolate bar with darker chocolate
Paired preference testing

2 products:
A  Chocolate bar (standard)
B  Chocolate bar with darker chocolate

2-Alternative Forced Choice (2-AFC):
Do you prefer A or B?

Prefer A  □  Prefer B  □
Paired preference testing

2 products:
A  Chocolate bar (standard)
B  Chocolate bar with darker chocolate

2-Alternative Forced Choice (2-AFC):

- Do you prefer A or B?

Prefer A  Prefer B
□□

2-Alternative Forced Choice (2-AC):

- Do you prefer A or B, or do you have no preference?

Prefer A  No Preference  Prefer B
□□□
Paired preference with a *no preference* option

Terminology:

No preference \(\sim\) No difference \(\sim\) Ties

Why allow for a no preference option?
- More information and greater resolution in data
- Products may actually be equally liked
  - No preference counts may support non-inferiority claims

Why avoid a no preference option?
- Statistical methods less well-known
Paired preference with a *no preference* option

**Terminology:**
- *No preference*  
- *No difference*  
- *Ties*

*Why allow for a no preference option?*
- More information and greater resolution in data
- Products may actually be equally liked
- *No preference* counts may support non-inferiority claims

*Why avoid a no preference option?*
- Statistical methods less well-known
Paired preference with a *no preference* option

Terminology:  *No preference*  ~  *No difference*  ~  *Ties*

Why allow for a *no preference* option?

More information and greater resolution in data
Products may actually be equally liked
*No preference* counts may support non-inferiority claims

Why avoid a *no preference* option?

Statistical methods less well-known
Paired preference with a *no preference* option

Terminology: *No preference*  \(\sim\)  *No difference*  \(\sim\)  Ties

Why allow for a *no preference* option?

- More information and greater resolution in data
Paired preference with a *no preference* option

**Terminology:**  
*No preference*  \sim  *No difference*  \sim  *Ties*

**Why allow for a *no preference* option?**

- More information and greater resolution in data
- Products may actually be equally liked
Paired preference with a *no preference* option

Terminology:  *No preference*  \(\sim\)  *No difference*  \(\sim\)  Ties

Why allow for a *no preference* option?
- More information and greater resolution in data
- Products may actually be equally liked
- *No preference* counts may support non-inferiority claims
Paired preference with a *no preference* option

Terminology: \( \text{No preference} \sim \text{No difference} \sim \text{Ties} \)

Why allow for a *no preference* option?
- More information and greater resolution in data
- Products may actually be equally liked
- *No preference* counts may support non-inferiority claims

Why avoid a *no preference* option?
Paired preference with a *no preference* option

Terminology:  
*No preference* $\sim$ *No difference* $\sim$ *Ties*

Why allow for a *no preference* option?
- More information and greater resolution in data
- Products may actually be equally liked
- *No preference* counts may support non-inferiority claims

Why avoid a *no preference* option?
- Statistical methods less well-known
Placebo experiments and identicality norms

Consider the data:

<table>
<thead>
<tr>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All counts</td>
<td>90</td>
<td>20</td>
<td>90</td>
</tr>
</tbody>
</table>
Placebo experiments and identicality norms

Consider the data:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All counts</td>
<td>90</td>
<td>20</td>
<td>90</td>
<td>200</td>
</tr>
</tbody>
</table>

Are there no differences wrt. preference in the population?
Consider the data:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All counts</td>
<td>90</td>
<td>20</td>
<td>90</td>
<td>200</td>
</tr>
</tbody>
</table>

- Are there no differences wrt. preference in the population?
- What if there are two opposing segments?
Consider the data:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All counts</td>
<td>90</td>
<td>20</td>
<td>90</td>
<td>200</td>
</tr>
<tr>
<td>Segment 1</td>
<td>8</td>
<td>10</td>
<td>82</td>
<td>100</td>
</tr>
<tr>
<td>Segment 2</td>
<td>82</td>
<td>10</td>
<td>8</td>
<td>100</td>
</tr>
</tbody>
</table>

- Are there no differences wrt. preference in the population?
- What if there are two opposing segments?
Placebo experiments and identicality norms

Consider the data:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All counts</td>
<td>90</td>
<td>20</td>
<td>90</td>
<td>200</td>
</tr>
<tr>
<td>Segment 1</td>
<td>8</td>
<td>10</td>
<td>82</td>
<td>100</td>
</tr>
<tr>
<td>Segment 2</td>
<td>82</td>
<td>10</td>
<td>8</td>
<td>100</td>
</tr>
</tbody>
</table>

- Are there no differences wrt. preference in the population?
- What if there are two opposing segments?

Ennis and Ennis (2012) suggest:

1. Perform placebo experiment
2. Estimate the *identicality norm*:

   *The expected proportion of counts for identical products*

Example: Comparing data with an identicality norm

Ennis’ Approach:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>25</td>
<td>15</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Identicality norm</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>—</td>
</tr>
</tbody>
</table>

\[ \chi^2 = \frac{(25-40)^2}{40} + \frac{(15-20)^2}{20} + \frac{(60-40)^2}{40} = 5.625 + 1.25 + 10 = 16.875 \]

\[ p\text{-value} = 0.00022 \]

Assumes identicality norm known without error

Uncertainty in the placebo experiment not taken into account!
Example: Comparing data with an identicality norm

Ennis’ Approach:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>25</td>
<td>15</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Identicality norm</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>—</td>
</tr>
</tbody>
</table>

\[
X^2_2 = \frac{(25 - 40)^2}{40} + \frac{(15 - 20)^2}{20} + \frac{(60 - 40)^2}{40} \\
= 5.625 + 1.250 + 10.00 = 16.875 \\
p-value = 0.00022
\]
Example: Comparing data with an identicality norm

Ennis’ Approach:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>25</td>
<td>15</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Identicality norm</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>—</td>
</tr>
</tbody>
</table>

\[
X^2_2 = \frac{(25 - 40)^2}{40} + \frac{(15 - 20)^2}{20} + \frac{(60 - 40)^2}{40} \\
= 5.625 + 1.250 + 10.00 = 16.875
\]

\[
p\text{-value} = 0.00022
\]

- Assumes identicality norm known without error
Example: Comparing data with an identicality norm

Ennis’ Approach:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>25</td>
<td>15</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Identicality norm</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>—</td>
</tr>
</tbody>
</table>

\[
X^2 = \frac{(25 - 40)^2}{40} + \frac{(15 - 20)^2}{20} + \frac{(60 - 40)^2}{40}
\]

\[
= 5.625 + 1.250 + 10.00 = 16.875
\]

\[p\text{-value} = 0.00022\]

- Assumes identicality norm known without error
- Uncertainty in the placebo experiment not taken into account!
Example: Comparing data with an identicality norm

How do we take the uncertainty in the placebo experiment into account?
Example: Comparing data with an identicality norm

How do we take the uncertainty in the placebo experiment into account?

Assume $n = 100$ in placebo experiment:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>25</td>
<td>15</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Placebo data</td>
<td>40</td>
<td>20</td>
<td>40</td>
<td>100</td>
</tr>
</tbody>
</table>

Expected counts:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>32.5</td>
<td>17.5</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Placebo data</td>
<td>32.5</td>
<td>17.5</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

The standard (genuine) Pearson $\chi^2$ test:

$$X^2 = \frac{(25 - 32.5)^2}{32.5} + \frac{(40 - 32.5)^2}{32.5} = 8.18$$

$p$-value $= 0.0168$ (previous $p$-value $= 0.00022$)
Example: Comparing data with an identicality norm

How do we take the uncertainty in the placebo experiment into account?

Assume \( n = 100 \) in placebo experiment:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>25</td>
<td>15</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Placebo data</td>
<td>40</td>
<td>20</td>
<td>40</td>
<td>100</td>
</tr>
</tbody>
</table>

Expected counts:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>32.5</td>
<td>17.5</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Placebo data</td>
<td>32.5</td>
<td>17.5</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

The standard (genuine) Pearson \( \chi^2 \) test:

\[
\chi^2 = \frac{(25-32.5)^2}{32.5} + \frac{(40-32.5)^2}{32.5} + \ldots + \frac{(40-50)^2}{50} = 8.18
\]

\( p \)-value = 0.0168 (previous \( p \)-value = 0.00022)
Example: Comparing data with an identicality norm

How do we take the uncertainty in the placebo experiment into account?

Assume \( n = 100 \) in placebo experiment:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>25</td>
<td>15</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Placebo data</td>
<td>40</td>
<td>20</td>
<td>40</td>
<td>100</td>
</tr>
</tbody>
</table>

Expected counts:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>32.5</td>
<td>17.5</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Placebo data</td>
<td>32.5</td>
<td>17.5</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

The standard (genuine) Pearson \( \chi^2 \) test:

\[
X_2^2 = \frac{(25 - 32.5)^2}{32.5} + \frac{(40 - 32.5)^2}{32.5} + \ldots + \frac{(40 - 50)^2}{50} = 8.18
\]

\( p \)-value = 0.0168 \( \) (previous \( p \)-value = 0.00022)
Effect of sample size in placebo experiment

Standard Pearson test on $2 \times 3$ table:

<table>
<thead>
<tr>
<th>$n_{\text{placebo}}$</th>
<th>$\chi^2$ statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2.80</td>
<td>0.24619</td>
</tr>
<tr>
<td>50</td>
<td>5.50</td>
<td>0.06393</td>
</tr>
<tr>
<td>100</td>
<td>8.18</td>
<td>0.01677</td>
</tr>
<tr>
<td>1000</td>
<td>15.15</td>
<td>0.00051</td>
</tr>
<tr>
<td>$10^9$</td>
<td>16.87</td>
<td>0.00022</td>
</tr>
</tbody>
</table>
Effect of sample size in placebo experiment

Standard Pearson test on $2 \times 3$ table:

<table>
<thead>
<tr>
<th>$n_{placebo}$</th>
<th>$\chi^2$ statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2.80</td>
<td>0.24619</td>
</tr>
<tr>
<td>50</td>
<td>5.50</td>
<td>0.06393</td>
</tr>
<tr>
<td>100</td>
<td>8.18</td>
<td>0.01677</td>
</tr>
<tr>
<td>1000</td>
<td>15.15</td>
<td>0.00051</td>
</tr>
<tr>
<td>$10^9$</td>
<td>16.87</td>
<td>0.00022</td>
</tr>
</tbody>
</table>

Ennis & Ennis (2012):

$X^2 = 16.87$ and $p$-value = 0.00022
Effect of sample size in placebo experiment

Standard Pearson test on $2 \times 3$ table:

<table>
<thead>
<tr>
<th>$n_{placebo}$</th>
<th>$\chi^2$ statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2.80</td>
<td>0.24619</td>
</tr>
<tr>
<td>50</td>
<td>5.50</td>
<td>0.06393</td>
</tr>
<tr>
<td>100</td>
<td>8.18</td>
<td>0.01677</td>
</tr>
<tr>
<td>1000</td>
<td>15.15</td>
<td>0.00051</td>
</tr>
<tr>
<td>$10^9$</td>
<td>16.87</td>
<td>0.00022</td>
</tr>
</tbody>
</table>

Ennis & Ennis (2012):
$X^2 = 16.87$ and $p$-value $= 0.00022$
Preliminary results and purpose of this work

Preliminary results:

No preference votes contain information. Don’t ignore the uncertainty in the placebo data. The genuine Pearson test on the $2 \times 3$ table is a better option.

Purpose of this work: Find a good test for 2-AC testing. Desirable properties of a good test:

- Appropriate type I error
- High power
- Insightful interpretation
- Easy to compute
Preliminary results and purpose of this work

Preliminary results:

- *No preference* votes contain information
Preliminary results and purpose of this work

Preliminary results:

- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data

Purpose of this work: Find a good test for 2-AC testing

Desirable properties of a good test:
- Appropriate type I error
- High power
- Insightful interpretation
- Easy to compute
Preliminary results and purpose of this work

Preliminary results:

- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data
- The genuine Pearson test on the $2 \times 3$ table is a better option

Purpose of this work: Find a good test for 2-AC testing

Desirable properties of a good test:

- Appropriate type I error
- High power
- Insightful interpretation
- Easy to compute
Preliminary results and purpose of this work

Preliminary results:
- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data
- The genuine Pearson test on the $2 \times 3$ table is a better option

Are there even better tests?
Preliminary results and purpose of this work

Preliminary results:
- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data
- The genuine Pearson test on the $2 \times 3$ table is a better option

Are there even better tests?

Purpose of this work: Find a good test for 2-AC testing
Preliminary results and purpose of this work

Preliminary results:
- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data
- The genuine Pearson test on the $2 \times 3$ table is a better option

Are there even better tests?

Purpose of this work: Find a good test for 2-AC testing

Desirable properties of a good test:
Preliminary results and purpose of this work

Preliminary results:
- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data
- The genuine Pearson test on the $2 \times 3$ table is a better option

Are there even better tests?

Purpose of this work: Find a good test for 2-AC testing

Desirable properties of a good test:
- Appropriate type I error
Preliminary results and purpose of this work

Preliminary results:
- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data
- The genuine Pearson test on the $2 \times 3$ table is a better option

Are there even better tests?

Purpose of this work: Find a good test for 2-AC testing

Desirable properties of a good test:
- Appropriate type I error
- High power
Preliminary results and purpose of this work

Preliminary results:
- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data
- The genuine Pearson test on the $2 \times 3$ table is a better option

Are there even better tests?

Purpose of this work: Find a good test for 2-AC testing

Desirable properties of a good test:
- Appropriate type I error
- High power
- Insightful interpretation
Preliminary results and purpose of this work

Preliminary results:
- *No preference* votes contain information
- Don’t ignore the uncertainty in the placebo data
- The genuine Pearson test on the $2 \times 3$ table is a better option

Are there even better tests?

Purpose of this work: Find a good test for 2-AC testing

Desirable properties of a good test:
- Appropriate type I error
- High power
- Insightful interpretation
- Easy to compute
Approach

1. Consider 5 test statistics

2. Compare the power of the 5 tests in a simulation study
Approach

1. Consider 5 test statistics

2. Compare the power of the 5 tests in a simulation study
Parameterization and test statistics

Parameterization:

Note: $p_0 = 0.5$ is given by the design.
Parameterization and test statistics

Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>$p_0(1 - s_0)$</td>
<td>$s_0$</td>
<td>$(1 - p_0)(1 - s_0)$</td>
</tr>
</tbody>
</table>

Note: $p_0 = 0.5$ is given by the design!
Parameterization and test statistics

Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>$p_0(1 - s_0)$</td>
<td>$s_0$</td>
<td>$(1 - p_0)(1 - s_0)$</td>
</tr>
<tr>
<td>Preference</td>
<td>$p_1(1 - s_1)$</td>
<td>$s_1$</td>
<td>$(1 - p_1)(1 - s_1)$</td>
</tr>
</tbody>
</table>

Note: $p_0 = 0.5$ is given by the design!
Parameterization and test statistics

Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>$p_0 (1 - s_0)$</td>
<td>$s_0$</td>
<td>$(1 - p_0)(1 - s_0)$</td>
</tr>
<tr>
<td>Preference</td>
<td>$p_1 (1 - s_1)$</td>
<td>$s_1$</td>
<td>$(1 - p_1)(1 - s_1)$</td>
</tr>
</tbody>
</table>

Test statistics:

<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tie effects</td>
<td>$s_0 = s_1$</td>
<td>$s_0 \neq s_1$</td>
<td>1</td>
</tr>
</tbody>
</table>
### Parameterization and test statistics

#### Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>(p_0(1 - s_0))</td>
<td>(s_0)</td>
<td>((1 - p_0)(1 - s_0))</td>
</tr>
<tr>
<td>Preference</td>
<td>(p_1(1 - s_1))</td>
<td>(s_1)</td>
<td>((1 - p_1)(1 - s_1))</td>
</tr>
</tbody>
</table>

#### Test statistics:

<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tie effects</td>
<td>(s_0 = s_1)</td>
<td>(s_0 \neq s_1)</td>
<td>1</td>
</tr>
<tr>
<td>Directional effects</td>
<td>(p_1 = 0.5)</td>
<td>(p_1 \neq 0.5)</td>
<td>1</td>
</tr>
</tbody>
</table>
Parameterization and test statistics

Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>$p_0(1 - s_0)$</td>
<td>$s_0$</td>
<td>$(1 - p_0)(1 - s_0)$</td>
</tr>
<tr>
<td>Preference</td>
<td>$p_1(1 - s_1)$</td>
<td>$s_1$</td>
<td>$(1 - p_1)(1 - s_1)$</td>
</tr>
</tbody>
</table>

Test statistics:

<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tie effects</td>
<td>$s_0 = s_1$</td>
<td>$s_0 \neq s_1$</td>
<td>1</td>
</tr>
<tr>
<td>Directional effects</td>
<td>$p_1 = 0.5$</td>
<td>$p_1 \neq 0.5$</td>
<td>1</td>
</tr>
<tr>
<td>Genuine Pearson</td>
<td>$s_0 = s_1$ and $p_0 = p_1$</td>
<td>$s_0 \neq s_1$ or $p_0 \neq p_1$</td>
<td>2</td>
</tr>
</tbody>
</table>
Parameterization and test statistics

Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>$p_0(1 - s_0)$</td>
<td>$s_0$</td>
<td>$(1 - p_0)(1 - s_0)$</td>
</tr>
<tr>
<td>Preference</td>
<td>$p_1(1 - s_1)$</td>
<td>$s_1$</td>
<td>$(1 - p_1)(1 - s_1)$</td>
</tr>
</tbody>
</table>

Test statistics:

<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tie effects</td>
<td>$s_0 = s_1$</td>
<td>$s_0 \neq s_1$</td>
<td>1</td>
</tr>
<tr>
<td>Directional effects</td>
<td>$p_1 = 0.5$</td>
<td>$p_1 \neq 0.5$</td>
<td>1</td>
</tr>
<tr>
<td>Genuine Pearson</td>
<td>$s_0 = s_1$ and $p_0 = p_1$</td>
<td>$s_0 \neq s_1$ or $p_0 \neq p_1$</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: $p_0 = 0.5$ is given by the design!
Parameterization and test statistics

Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>( p_0(1 - s_0) )</td>
<td>( s_0 )</td>
<td>( (1 - p_0)(1 - s_0) )</td>
</tr>
<tr>
<td>Preference</td>
<td>( p_1(1 - s_1) )</td>
<td>( s_1 )</td>
<td>( (1 - p_1)(1 - s_1) )</td>
</tr>
</tbody>
</table>

Test statistics:

<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tie effects</td>
<td>( s_0 = s_1 )</td>
<td>( s_0 \neq s_1 )</td>
<td>1</td>
</tr>
<tr>
<td>Directional effects</td>
<td>( p_1 = 0.5 )</td>
<td>( p_1 \neq 0.5 )</td>
<td>1</td>
</tr>
<tr>
<td>Genuine Pearson</td>
<td>( s_0 = s_1 ) and ( p_0 = p_1 )</td>
<td>( s_0 \neq s_1 ) or ( p_0 \neq p_1 )</td>
<td>2</td>
</tr>
</tbody>
</table>

- Note: \( p_0 = 0.5 \) is given by the design!
- The Genuine Pearson test is NOT the right test
Parameterization and test statistics

Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>$p_0(1 - s_0)$</td>
<td>$s_0$</td>
<td>$(1 - p_0)(1 - s_0)$</td>
</tr>
<tr>
<td>Preference</td>
<td>$p_1(1 - s_1)$</td>
<td>$s_1$</td>
<td>$(1 - p_1)(1 - s_1)$</td>
</tr>
</tbody>
</table>

Test statistics:

<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tie effects</td>
<td>$s_0 = s_1$</td>
<td>$s_0 \neq s_1$</td>
<td>1</td>
</tr>
<tr>
<td>Directional effects</td>
<td>$p_1 = 0.5$</td>
<td>$p_1 \neq 0.5$</td>
<td>1</td>
</tr>
<tr>
<td>Genuine Pearson</td>
<td>$s_0 = s_1$ and $p_0 = p_1$</td>
<td>$s_0 \neq s_1$ or $p_0 \neq p_1$</td>
<td>2</td>
</tr>
<tr>
<td>Modified Pearson</td>
<td>$s_0 = s_1$ and $p_1 = 0.5$</td>
<td>$s_0 \neq s_1$ or $p_1 \neq 0.5$</td>
<td>2</td>
</tr>
</tbody>
</table>

- **Note:** $p_0 = 0.5$ is given by the design!
- The Genuine Pearson test is NOT the right test
Parameterization and test statistics

Parameterization:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>$p_0(1 - s_0)$</td>
<td>$s_0$</td>
<td>$(1 - p_0)(1 - s_0)$</td>
</tr>
<tr>
<td>Preference</td>
<td>$p_1(1 - s_1)$</td>
<td>$s_1$</td>
<td>$(1 - p_1)(1 - s_1)$</td>
</tr>
</tbody>
</table>

Test statistics:

<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tie effects</td>
<td>$s_0 = s_1$</td>
<td>$s_0 \neq s_1$</td>
<td>1</td>
</tr>
<tr>
<td>Directional effects</td>
<td>$p_1 = 0.5$</td>
<td>$p_1 \neq 0.5$</td>
<td>1</td>
</tr>
<tr>
<td>Genuine Pearson</td>
<td>$s_0 = s_1 \text{ and } p_0 = p_1$</td>
<td>$s_0 \neq s_1 \text{ or } p_0 \neq p_1$</td>
<td>2</td>
</tr>
<tr>
<td>Modified Pearson</td>
<td>$s_0 = s_1 \text{ and } p_1 = 0.5$</td>
<td>$s_0 \neq s_1 \text{ or } p_1 \neq 0.5$</td>
<td>2</td>
</tr>
<tr>
<td>Pooled Test</td>
<td>$s_0 = s_1 \text{ and } p_1 = 0.5$</td>
<td>$s_0 \neq s_1 \text{ or } p_1 \neq 0.5$</td>
<td>2</td>
</tr>
</tbody>
</table>

- Note: $p_0 = 0.5$ is given by the design!
- The Genuine Pearson test is NOT the right test
Approach

1. Consider 5 test statistics

2. Compare the power of the 5 tests in a simulation study
Settings for power simulations

Placebo experiment (true identicality norm):

<table>
<thead>
<tr>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Power simulations in 6 settings:

<table>
<thead>
<tr>
<th>Placebo sample size</th>
<th>Structures in preference data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tie effects</td>
</tr>
<tr>
<td>100</td>
<td>1A</td>
</tr>
<tr>
<td>1.000.000</td>
<td>2A</td>
</tr>
</tbody>
</table>

- $n_{\text{preference}} = 100$
- 10,000 simulations at each point
Tie effects

Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test

Power

Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15
Pr(tie) 0.2 0.3 0.4 0.5 0.6 0.7
Pr(B) 0.4 0.35 0.3 0.25 0.2 0.15
Tie effects

Power

Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15
Pr(tie) 0.2 0.3 0.4 0.5 0.6 0.7
Pr(B) 0.4 0.35 0.3 0.25 0.2 0.15

Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test
Tie effects

Directional effects

Joint effects

Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test

Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15
Pr(tie) 0.2 0.3 0.4 0.5 0.6 0.7
Pr(B) 0.4 0.35 0.3 0.25 0.2 0.15

Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15
Pr(tie) 0.2 0.25 0.3 0.25 0.25 0.25
Pr(B) 0.4 0.4 0.4 0.4 0.4 0.4

Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15
Pr(tie) 0.2 0.25 0.3 0.25 0.25 0.25
Pr(B) 0.4 0.4 0.4 0.4 0.4 0.4
Example — new insights

Example data:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo exp.</td>
<td>81</td>
<td>45</td>
<td>74</td>
<td>200</td>
</tr>
<tr>
<td>Preference exp.</td>
<td>37</td>
<td>12</td>
<td>51</td>
<td>100</td>
</tr>
</tbody>
</table>
Example — new insights

Example data:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo exp.</td>
<td>81</td>
<td>45</td>
<td>74</td>
<td>200</td>
</tr>
<tr>
<td>Preference exp.</td>
<td>37</td>
<td>12</td>
<td>51</td>
<td>100</td>
</tr>
</tbody>
</table>

ANOVA-like analysis:

<table>
<thead>
<tr>
<th>Test</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled test</td>
<td>7.00</td>
<td>2</td>
<td>0.030</td>
</tr>
<tr>
<td>Tie effects</td>
<td>4.78</td>
<td>1</td>
<td>0.029</td>
</tr>
<tr>
<td>Directional effects</td>
<td>2.23</td>
<td>1</td>
<td>0.136</td>
</tr>
</tbody>
</table>

© Rune H B Christensen (DTU)
Example — new insights

Example data:

<table>
<thead>
<tr>
<th></th>
<th>Prefer A</th>
<th>No Preference</th>
<th>Prefer B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo exp.</td>
<td>81</td>
<td>45</td>
<td>74</td>
<td>200</td>
</tr>
<tr>
<td>Preference exp.</td>
<td>37</td>
<td>12</td>
<td>51</td>
<td>100</td>
</tr>
</tbody>
</table>

ANOVA-like analysis:

<table>
<thead>
<tr>
<th>Test</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled test</td>
<td>7.00</td>
<td>2</td>
<td>0.030</td>
</tr>
<tr>
<td>Tie effects</td>
<td>4.78</td>
<td>1</td>
<td>0.029</td>
</tr>
<tr>
<td>Directional effects</td>
<td>2.23</td>
<td>1</td>
<td>0.136</td>
</tr>
<tr>
<td>Modified Pearson</td>
<td>7.20</td>
<td>2</td>
<td>0.027</td>
</tr>
</tbody>
</table>
Final points

Conclusions and recommendations:

Placebo data contain valuable information

Don’t ignore the uncertainty in the placebo data

The modified Pearson and Pooled statistics have the highest power against general alternatives

Use the Pooled statistic to provide insight into the structure of the data

Open questions:

What may cause tie-effects?

Segmentation

Heterogeneity in preference

Unequal variances in the underlying perceptual distributions
Final points

Conclusions and recommendations:
- Placebo data contain valuable information
Conclusions and recommendations:

- Placebo data contain valuable information
- Don’t ignore the uncertainty in the placebo data
Final points

Conclusions and recommendations:

- Placebo data contain valuable information
- Don’t ignore the uncertainty in the placebo data
- The modified Pearson and Pooled statistics have the highest power against general alternatives
Final points

Conclusions and recommendations:

- Placebo data contain valuable information
- Don’t ignore the uncertainty in the placebo data
- The modified Pearson and Pooled statistics have the highest power against general alternatives
- Use the Pooled statistic to provide insight into the structure of the data
Final points

Conclusions and recommendations:

- Placebo data contain valuable information
- Don’t ignore the uncertainty in the placebo data
- The modified Pearson and Pooled statistics have the highest power against general alternatives
- Use the Pooled statistic to provide insight into the structure of the data

Open questions:
Final points

Conclusions and recommendations:

- Placebo data contain valuable information
- Don’t ignore the uncertainty in the placebo data
- The modified Pearson and Pooled statistics have the highest power against general alternatives
- Use the Pooled statistic to provide insight into the structure of the data

Open questions:

- What may cause tie-effects?
Final points

Conclusions and recommendations:

- Placebo data contain valuable information
- Don’t ignore the uncertainty in the placebo data
- The modified Pearson and Pooled statistics have the highest power against general alternatives
- Use the Pooled statistic to provide insight into the structure of the data

Open questions:

- What may cause tie-effects?
  - Segmentation
Conclusions and recommendations:

- Placebo data contain valuable information
- Don’t ignore the uncertainty in the placebo data
- The modified Pearson and Pooled statistics have the highest power against general alternatives
- Use the Pooled statistic to provide insight into the structure of the data

Open questions:

- What may cause tie-effects?
  - Segmentation
  - Heterogeneity in preference
Final points

Conclusions and recommendations:

- Placebo data contain valuable information
- Don’t ignore the uncertainty in the placebo data
- The modified Pearson and Pooled statistics have the highest power against general alternatives
- Use the Pooled statistic to provide insight into the structure of the data

Open questions:

- What may cause tie-effects?
  - Segmentation
  - Heterogeneity in preference
  - Unequal variances in the underlying perceptual distributions
A new statistic to detect segmentation or unequal variance in 2-Alternative Choice (2-AC) testing

Rune H B Christensen¹,*  John M Ennis²  Daniel M Ennis²  Per B Brockhoff¹

¹DTU Informatics, IMM, Section for Statistics, Technical University of Denmark
²The Institute for Perception, Richmond, VA, USA

*Contact author: rhbc@imm.dtu.dk

July 13th 2012