
Alexandra Grand
Reliability

- degree to which measures yield consistent results
- reliability = proportion of truth
- measurement error reduces reliability

Spearman approach

\[ X_{\text{observed}} = X_{\text{true}} + X_{\text{error}} \]

\[ V_{\text{observed}} = V_{\text{true}} + V_{\text{error}} \]

**Reliability coefficient** (between 0 and 1)

\[ r = \frac{V_{\text{true}}}{V_{\text{observed}}} \]

\[ r = \frac{V_{\text{observed}} - V_{\text{error}}}{V_{\text{observed}}} \]

**perfect reliability:**

- \( r = 1 \)
- no random error \( X_{\text{observed}} = X_{\text{true}} \)
Measurement error

Where do measurement error come from?
# Reliability methods

<table>
<thead>
<tr>
<th>facet</th>
<th>Test-Retest</th>
<th>Internal Consistency</th>
<th>Alternative Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>subjects, sample</td>
<td>same</td>
<td>same</td>
<td>same</td>
</tr>
<tr>
<td>scale, test</td>
<td>same</td>
<td>items are split in half</td>
<td>two scales</td>
</tr>
<tr>
<td>timpoint</td>
<td>two different time points</td>
<td>same</td>
<td>two different time points</td>
</tr>
</tbody>
</table>

## Information
- stability of the measure
- homogeneity of a set of items
- equivalency of content

## Problem(s)
- length of the time gap
- change in the phenomenon
- correlation of each item with itself
- split of the items; different results
  - Cronbach’s alpha
- development
  - prove
Internal consistency reliability

Cronbach’s alpha

mean reliability coefficient for all possible ways of splitting a set of items in half

$$\alpha = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum_{i=1}^{k} \sigma_i^2}{\sigma_t^2} \right)$$

- $\sigma_i^2 = \text{variance of item } i \text{ (multi-point item)}$
- $\sigma_t^2 = \text{total variance of the scale}$
- $k = \text{number of items in the scale}$

$$\sum_{i=1}^{k} \sigma_i^2 + 2 \sum_{i} \sum_{j} \sigma_{ij}$$

- sum of item variance + 2 * sum of item covariance
Internal consistency reliability

LOWEST COVARIANCE MATRIX FOR SIX PROBABILITY OF LOSS ITEMS

\[ \alpha = \left( \frac{6}{6-1} \right) \left( 1 - \frac{20.08}{20.08 + 45.32} \right) = 0.83 \]
Generalizability theory

Traditional reliability coefficient for one type of facet → but there can be lots of sources for measurement error

Generalizability theory simultaneously analysing multiple sources of variance

condition of measurement = aspect (or „facet“) of measurement procedure
e.g. time, items used in a scale, observers

Are the obtained scores in the sampled conditions of measurement representative of the universe scores for those conditions?
Generalizability study (G)

--> estimation of variance components

e.g. „Brand loyalty measurement“ \( (n_p = 100) \)

two facets of interest:

• **time** (3 time points/occasions; \( n_j = 3 \))

• **instrument** (scale with 10 items; \( n_i = 10 \))

7 different variance components:

**main effects:**

• subjects \( \sigma_p^2 \)

• items \( \sigma_i^2 \)

• occasions \( \sigma_j^2 \)

**two-way interactions:**

• subject-item \( \sigma_{pi}^2 \)

• subject-occasion \( \sigma_{pj}^2 \)

• item-occasion \( \sigma_{ij}^2 \)

**residual** (subject-item-occasion interaction, random error) \( \sigma_{pij,e}^2 \)
## Generalizability study (G)

<table>
<thead>
<tr>
<th>Subject</th>
<th>t1</th>
<th></th>
<th>t2</th>
<th></th>
<th>t3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>l1</td>
<td>...</td>
<td>l9</td>
<td>...</td>
<td>l1</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>l2</td>
<td></td>
<td>l10</td>
<td></td>
<td>l2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

→ three-way analysis of variance

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Sum of squares</th>
<th>d.f.</th>
<th>Mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (p)</td>
<td>5999.400</td>
<td>99</td>
<td>60.600</td>
</tr>
<tr>
<td>Items (i)</td>
<td>1087.200</td>
<td>9</td>
<td>120.800</td>
</tr>
<tr>
<td>Occasions (j)</td>
<td>6540.000</td>
<td>2</td>
<td>3270.000</td>
</tr>
<tr>
<td>pi</td>
<td>1335.609</td>
<td>891</td>
<td>1.499</td>
</tr>
<tr>
<td>pj</td>
<td>712.800</td>
<td>198</td>
<td>3.600</td>
</tr>
<tr>
<td>ij</td>
<td>67.500</td>
<td>18</td>
<td>3.750</td>
</tr>
<tr>
<td>Residual</td>
<td>2227.500</td>
<td>1782</td>
<td>1.250</td>
</tr>
</tbody>
</table>
Generalizability study (G)

obtaining the seven components of variance ...

I) 1. start with the mean square of the residual component: \( \sigma_{pij,e}^2 \)

\[
EMS_{res} = \sigma_{pij,e}^2 = 1.25
\]

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Sum of squares</th>
<th>d.f.</th>
<th>Mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (p)</td>
<td>5999.400</td>
<td>99</td>
<td>60.600</td>
</tr>
<tr>
<td>Items (i)</td>
<td>1087.200</td>
<td>9</td>
<td>120.800</td>
</tr>
<tr>
<td>Occasions (j)</td>
<td>6540.000</td>
<td>2</td>
<td>3270.000</td>
</tr>
<tr>
<td>( p_i )</td>
<td>1335.609</td>
<td>891</td>
<td>1.499</td>
</tr>
<tr>
<td>( pj )</td>
<td>712.800</td>
<td>198</td>
<td>3.600</td>
</tr>
<tr>
<td>( ij )</td>
<td>67.500</td>
<td>18</td>
<td>3.750</td>
</tr>
<tr>
<td>Residual</td>
<td>2227.500</td>
<td>1782</td>
<td><strong>1.250</strong></td>
</tr>
</tbody>
</table>

2. working backward through each equation to compute each of the components of variance

\[
EMS_p = \sigma_{pij,e}^2 + n_i \sigma_{pj}^2 + n_j \sigma_{pi}^2 + n_i n_j \sigma_p^2
\]

\[
: \quad EMS_{res} = \sigma_{pij,e}^2 = 1.25
\]
Generalizability study (G)

obtaining the seven components of variance ...

II) 1. compute the variance component of interest
     e.g. \( \sigma_p^2 \)

\[
EMSP = \frac{MS_p - MS_{pi} - MS_{pj} + MS_{pij}}{n_i n_j} = \frac{60.600 - 1.499 - 3.600 + 1.250}{10 \cdot 3} = 1.892
\]
Generalizability study (G)

ESTIMATED COMPONENTS OF VARIANCE AND PERCENTAGES OF TOTAL VARIANCE FOR HYPOTHETICAL BRAND LOYALTY STUDY

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Mean square</th>
<th>d.f.</th>
<th>Estimate of variance component</th>
<th>Percentage of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (p)</td>
<td>60.60</td>
<td>99</td>
<td>1.892</td>
<td>26.51</td>
</tr>
<tr>
<td>Items (i)</td>
<td>120.80</td>
<td>9</td>
<td>.389</td>
<td>5.45</td>
</tr>
<tr>
<td>Occasions (j)</td>
<td>3270.00</td>
<td>2</td>
<td>3.264</td>
<td>45.73</td>
</tr>
<tr>
<td>pij</td>
<td>4.20</td>
<td>891</td>
<td>.083</td>
<td>1.16</td>
</tr>
<tr>
<td>pji</td>
<td>3.60</td>
<td>198</td>
<td>.235</td>
<td>3.29</td>
</tr>
<tr>
<td>iji</td>
<td>3.75</td>
<td>18</td>
<td>.025</td>
<td>3.5</td>
</tr>
<tr>
<td>Residual</td>
<td>1.25</td>
<td>1782</td>
<td>1.250</td>
<td>17.51</td>
</tr>
</tbody>
</table>

\[ \sigma_t^2 = 7.138 \]  

\[ G = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{pi}^2 + \sigma_{p}^2 + \sigma_{p,ij}^2} \]

\[ G = 1.892 \times (1.892 + 0.083 + 0.235 + 1.250) = 0.547 \]
Generalizability theory / practice

Generalizability (G) study $\rightarrow$ decision (D) study

• very complex
• profitable?

In Practice

400 consumer behavior studies
• no alternative form or multifaced-generalizability approach
• only 19 studies with reliability estimates
• majority of studies with semantic differential or Likert-type items

Problem:
– determination of the appropriate number of items for a scale
  • too many items
  • complex constructs with multiple dimensions

Recommendations:
• reliability coefficient of 0.5 to 0.6; 0.95 “should be the desired standard”
• 10 times as many subjects as items (or at least 5 subjects per item)