Measurement validity and reliability

The classical view of measurement validity

Validity used to be divided into 3Cs (DeVellis, 1991; Fletcher, Fletcher and Wagner, 1996):
1. Content validity.
2. Criterion validity.
3. Construct validity.

The validity

- Validity is “the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose” (AERA, APA & NCME, 1999).
- The validity evidence can be obtained from five sources (AERA, APA & NCME, 1999; Cook, & Beckman, 2006):
  1. Content.
  2. **Internal structure.**
  3. Relations to other variables
  4. Response process.
  5. Consequences.
- Our focus → **Internal structure**.
- **Construct** is “the concept or characteristic that a test is designed to measure” (AERA, APA & NCME, 1999).
- **Construct = Domain = Concept = Idea**
- **Internal structure evidence** → the extent of how the relationships between the test items and components reflect the construct (AERA, APA & NCME, 1999).
- Evidence based on internal structure can be obtained from (Cook, & Beckman, 2006):
  1. **Factor analysis.**
  2. **Reliability.**

Factor analysis

**Factoring**

- We tend to group things that have something in common.
- Simplify long list of items into smaller groups.
- **Factoring = Grouping = Clustering.**
- The factor/group may represent the construct.

**Intuitive factoring**

**List of items:**

**Orange, motorcycle, bus, durian, banana, car**

- Do these six items have something in common?

  **Group the items:**

  [Orange, durian, banana]

  [Motorcycle, bus, car]
Name the groups:

<table>
<thead>
<tr>
<th>Fruit</th>
<th>Orange, durian, banana</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor-vehicle</td>
<td>Motorcycle, bus, car</td>
</tr>
</tbody>
</table>

- Finding something in common among the items, factoring the items and naming the factors are basically factor analysis!
- Factor out the common idea from the items.

Correlation matrix:

- Let say the same items are rated on Likert-scale responses from 1 to 5 on their characteristics of being fruit or motor vehicle. Then the Pearson's correlation coefficients among the items are tabulated:

<table>
<thead>
<tr>
<th>Items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Orange</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Durian</td>
<td>.67</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Banana</td>
<td>.70</td>
<td>.81</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Motorcycle</td>
<td>.11</td>
<td>.08</td>
<td>.05</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Bus</td>
<td>.08</td>
<td>.12</td>
<td>.09</td>
<td>.75</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>6. Car</td>
<td>.18</td>
<td>.12</td>
<td>.22</td>
<td>.89</td>
<td>.83</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- We examine pattern of correlation in the correlation matrix, then group highly correlated items into factors.

Factors

<table>
<thead>
<tr>
<th>Items</th>
<th>Fruit</th>
<th>Motor vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Orange</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>2. Durian</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>3. Banana</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>4. Motorcycle</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>5. Bus</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>6. Car</td>
<td>-</td>
<td>X</td>
</tr>
</tbody>
</table>

- However such approach is tedious for a large number of items, for example for 43 items, we must examine 43(43-1)/2 = 903 correlations.
- Factor analysis enables objective assessment of these correlations and factor/group the items.

Factor analysis

- It is a multivariate statistical analysis i.e. many outcomes.
- Factors can be determined in mathematical way.
- The basis is the determination of number and nature of factors that are responsible for the
correlations among items (Brown, 2006).

- From a number of outcomes (observed variables), factors are extracted and determined. These factors are unobserved/latent independent factors.
- In contrast to multiple linear regression, the one outcome and many independent factors are measurable.
- The analysis can be (Brown, 2006):
  - Exploratory – Exploratory Factor Analysis (EFA).
  - Confirmatory – Confirmatory Factor Analysis (CFA).

**Exploratory factor analysis (EFA)**

- An exploratory procedure.
- Aims to explore the items, factor common concepts and generate theory.
- Rotation of factors is used to allow simpler solution.
  - **Orthogonal method** – uncorrelated factors.
    - Varimax, Quartimax, Equamax
  - **Oblique method** – correlated factors.
    - Promax, Direct Oblimin
- Generally there are two models (Gorsuch, 1983):
  - Full Component Model.
  - Common Factor Model.
- The types of EFA determine extraction methods.

**Full Component Model**

- Extraction method: Principal component analysis (PCA)
- Account for all variances, suitable for data reduction, e.g. items are condensed into a factor then used as a single variable for other statistical analysis.
- Does not account for error in measurement.
- Not the 'real' factor analysis (Gorsuch, 1983; Brown, 2006).

**Common factor model**

- Extraction methods:
  - Classical: Principal axis analysis.
  - Other variants: Image analysis, alpha analysis, maximum likelihood analysis.
- Attempts to account for common variance and error variance.
  - Common variance - variance shared between the related items.
  - Error/Unique variance - variance specific to the item. It can be further partitioned into systematic error and random error variances.
- The 'real' factor analysis.
- The maximum likelihood variant allows assessment of model fit.
- The common factor model will be used throughout the workshop.
Confirmatory factor analysis (CFA)

- A confirmatory procedure.
- Also based on common factor model.
- A type of Structural Equation Modeling (SEM) analysis:
  - **Measurement model (CFA)** - dealing with latent variables (factors) and the relationships between the items and the factors.
  - Structural model (path analysis) - dealing with how latent variables are related to each other.
- Maximum likelihood method is commonly used for estimation.
- Allows assessment of measurement model fit, as well as other aspects of the validity.
- The main difference between EFA and CFA is that by using CFA, the researcher already established the factors and which items belong to the factors No longer exploratory.
- For example, CFA items:
  
  I love fast food
  I hate vegetable
  I hate eating fruits
  I hate exercise

  Obesity

- The items are probably based on his exploratory procedure (EFA), literature reviews, theories, or experience – strong theoretical basis for the items and factors.
- For example, EFA items:

  I love cat
  I hate snake
  I love traveling
  I love snorkeling
  I support ABC football team
  I love driving car
  I love computer game
  I like to have everything normally distributed
  I am a strong believer of central limit theorem
  My favorite food is nasi ayam
  I enjoy eating pisang goreng
  I spend most of my time in front of computer
  I love R

- No idea? Use EFA.
The differences between EFA and CFA can be summarized in the table below:

<table>
<thead>
<tr>
<th>EFA</th>
<th>CFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explorative procedure.</td>
<td>Confirmatory procedure.</td>
</tr>
<tr>
<td>No pre-requisite to specify theoretical factors for a collections of items.</td>
<td>Pre-specified theoretical factors.</td>
</tr>
<tr>
<td>Aims to explore the items and extract common ideas. Theory generating based on empirical findings.</td>
<td>Strong theory. Just want to confirm.</td>
</tr>
<tr>
<td>Items free loading.</td>
<td>No cross loading of items. Fixed item loadings to pre-specified factors.</td>
</tr>
<tr>
<td>Rotation of factors is used to allow simpler solution.</td>
<td>Rotation not used.</td>
</tr>
<tr>
<td>Explicit hypothesis is not tested.</td>
<td>Explicit hypothesis testing. Allows assessment of model fit (X² GOF, Fit indices).</td>
</tr>
</tbody>
</table>

Reliability

- In the current framework, part of validity evidence from internal structure source.
- Reliability are generally divided into types (Trochim, 2006; Kline, 2011):
  1. Test-retest reliability
  2. Parallel-forms reliability
  3. Interrater reliability
  4. Internal consistency reliability

Internal Consistency

- It is the degree to which responses are consistent across the items within a construct i.e. measure the same thing (Kline, 2011) in similar direction for a particular subject. In other words, how homogenous the items in a construct in term of their variance.
- Low internal consistency means that the items are heterogeneous within a construct i.e. do not measure the same factor, thus the total score is not the best way to summarize the construct (Kline, 2011).
- When responses for items within a construct are positively correlated to each other, they may measure the same factor. In this case, high internal consistency is obtained.
- In comparison to the rest of reliability types, it only requires measurement on a single occasion.
Cronbach's Alpha

- **Cronbach's alpha coefficient** is a common way to indicate internal consistency of a construct.
- Ranges from 0 to 1.
  - When $\alpha=1$, the items are all identical and perfectly correlated to each other, i.e. measure the same thing.
  - When $\alpha=0$, the items are all independent and none related to each other, i.e. do not measure the same thing.
- A generally acceptable cutoff value is 0.7 and above, while 0.6 is acceptable in exploratory research (Hair et al., 2010). However, it should not exceed 0.9 (Streiner, 2003).

Raykov's rho

- For a CFA model with good fit, it indicates the construct/composite reliability of a factor.
- Reliability by Raykov's rho (Raykov, 2001) is one of the reliability indices in CFA context.
- It also accounts for correlated errors, if specified in the model.
- Construct reliability $\geq 0.7$ (Hair et al., 2010) is acceptable.

References


