Internal structure evidence of validity

Dr Wan Nor Arifin
Lecturer, Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.
E-mail: wnarifin@usm.my

Outlines

Measurement validity and reliability
The classical view of measurement validity
The validity
Factor analysis
Reliability

Measurement validity and reliability

- ***Measurement*** is “the process observing and recording the observations that are collected as part of a research effort.” (Trochim, 2006)
- ***Measurement validity*** is "the degree to which the data measure what they were intended to measure", or in other words, how close the data reflect the true state of what being measured (Fletcher, Fletcher and Wagner, 1996). It is synonymous to **accuracy**.
- ***Measurement reliability*** means **repeatability**, **reproducibility**, **consistency** or **precision** (Fletcher, Fletcher and Wagner, 1996; Gordis, 2009; Trochim, 2006). It is “the extent to which repeated measurements of a stable phenomenon – by different people and instruments, at different times and places – get similar result” (Fletcher, Fletcher and Wagner, 1996).
- Classical way viewing validation process.

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.
- Nowadays, validity is described differently under the unitary concept of validity (Cook, & Beckman, 2006; American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA & NCME], 1999).

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

<u>List of items</u>:

Orange, motorcycle, bus, durian, banana, car

- Do these six items have something in common?

<u>Group the items:</u>

[Orange, durian, banana]

[Motorcycle, bus, car]

Name the groups:

Fruit	**Orange, durian, banana**
Motor-vehicle	**Motorcycle, bus, car**

- 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:

Items	**1**	**2**	**3**	**4**	**5**	**6**
1. Orange	1.00					
2. Durian	***.67***	1.00				
3. Banana	***.70***	***.81***	1.00			
4. Motorcycle	.11	.08	.05	1.00		
5. Bus	.08	.12	.09	***.75***	1.00	
6. Car	.18	.12	.22	***.89***	***.83***	1.00

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

	Factors	
Items	***Fruit***	***Motor vehicle***
1. Orange	X	-
2. Durian	X	-
3. Banana	X	-
4. Motorcycle	-	X
5. Bus	-	X
6. Car	-	X

- 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 <u>a number</u> of outcomes (observed variables), factors are extracted and determined. These factors are <u>unobserved/latent</u> independent factors.
- In contrast to multiple linear regression, the <u>one</u> outcome and many independent factors are <u>measurable</u>.
- 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.

<u>Full Component Model</u>

- 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).

<u>Common factor model</u>

- 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:
  - **<u>Measurement model (CFA)</u>** - 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:

EFA	**CFA**
Explorative procedure.	Confirmatory procedure.
No pre-requisite to specify theoretical factors for a collections of items.	Pre-specified theoretical factors.
Aims to explore the items and extract common ideas. Theory generating based on empirical findings.	Strong theory. Just want to confirm.
Items free loading.	No cross loading of items. Fixed item loadings to pre-specified factors.
Rotation of factors is used to allow simpler solution.	Rotation not used.
Explicit hypothesis is not tested.	Explicit hypothesis testing. Allows assessment of model fit (X^2 GOF, Fit indices).

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 α=1, the items are all identical and perfectly correlated to each other, i.e measure the same thing.
  - → When α=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 ≥ 0.7 (Hair et al., 2010) is acceptable.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: The Guilford Press.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*, *119*, 166.e7-166.e16.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

DeVellis, R. F. (1991). *Scale development: Theory and applications*. California: Sage Publications.

Fletcher, R. H., Fletcher, S. W., & Wagner, E. H. (1996). *Clinical epidemiology: the essentials* (3rd ed.). Maryland: Williams & Wilkins.

Gordis, L. (2009). *Epidemiology* (4th ed.). Philadelphia: Saunders.

Hair Jr., J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.) Upper Saddle River, NJ: Pearson Prentice-Hall.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Publications.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology, 54*, 315-323.

Streiner, D. L. & Norman, G. R. (2008). *Health measurement scales: a practical guide to their development and use.* New York: Oxford University Press.

Trochim, W. M. K. (2006). *Research methods knowledge base*. [Online] Retrieved from http://www.socialresearchmethods.net