Questionnaire Analysis Using R

Determining the Validity of Internal Structure Using Factor Analysis

Wan Nor Arifin, Universiti Sains Malaysia
wnarifin@usm.my & wnarifin.github.io

About Me

1. A medical doctor (long... time ago).
2. A lecturer at Biostatistics & Research Methodology Unit, School of Medical Sciences, USM.
3. A PhD candidate at School of Computer Sciences, USM.
4. Questionnaire validation research.
5. I “eat” R and Python everyday.

Outlines

1. Overview of Validity
2. Exploratory Factor Analysis
3. Confirmatory Factor Analysis

Download workshop materials @

wnarifin.github.io/workshop.html

2019/10/05 Questionnaire analysis using R @ confeRence 2019 (Sunway University)

Overview of Validity

Measurement validity & reliability

- Measurement ⟶ Process of observing & recording.
- Measurement validity ⟶ Accuracy.
- Measurement reliability ⟶ Precision, consistency, repeatability.

Classical validity

3Cs:

1. Content
2. Criterion
3. Construct

The validity

- Unitary concept.
- Degree of evidence ⟶ Purpose & Intended use of a tool.
- Evidence from 5 sources:
  1. Content.
  2. Internal structure.
  3. Relations to other variables
  4. Response process.
  5. Consequences.

The validity

- Construct – Concept to be measured by a tool.
- Internal structure evidence of validity.
- How relationship between items and factors reflect construct.
- Analysis:
  1. Factor analysis.
  2. Reliability.

Factoring

- Group things that have common concept.
- Simplify.
- Factoring = Grouping.
- Factor = Construct = Domain = Concept.

Factoring

Intuitive factoring:

Orange, motorcycle, bus, durian, banana, car

Anything in common?

Factoring

Group them

[Orange, durian, banana]
[Motorcycle, bus, car]

into two groups

Factoring

Name the group

Fruits	Motor vehicle
Orange	Motorcycle
Durian	Bus
Banana	Car

factor out the common concept

Factoring

- Find out correlated variables from correlation matrix.
- Manageable for small number of variables.
- Impossible for large number of variables.

Factor analysis

- In fields like psychology, we cannot observe directly (latent) psychological states, thus measured indirectly in of form items.

Factor analysis

- e.g. Depression:
    - depression causes symptoms of depression.
    - depression (latent) is measured indirectly by its symptoms (items).
    - prove the symptoms are correlated to each other, representing the concept of depression → factor analysis.

Factor analysis

- Multivariate analysis > 1 outcomes.
- Numerical items, e.g. Likert scale, VAS scores, laboratory results etc.
- Group correlated items (in a measurement scale).
- Factor out latent (unobserved) factors cause the correlation between the items.
- Latent variable model analysis.

Factor analysis

Common factor model:

Common Factors + Measurement Error

Classification:

- Exploratory factor analysis (EFA).
- Confirmatory factor analysis (CFA).

Internal consistency reliability

- Consistent responses in a construct.
- Homogenous ⟶ ↑Reliability.
- Heterogenous ⟶ ↓Reliability.
- Advantage: Measure 1x only.

Internal consistency reliability

- EFA: Cronbach's alpha coefficient.
- CFA: Raykov’s rho coefficient.
- Range: Not reliable 0 ⟶ 1 Perfectly reliable.
- Aim > 0.7.

Exploratory Factor Analysis (EFA)

EFA

- explorative method.
- e.g. at early of questionnaire development.
- theory generating.

EFA

Research questions:

- How many factors are there?
- Strength of relationship between items and the factors?
- Factor correlations?
- % variance explained by the extracted factors?

EFA

Applications:

- Psychological scales/questionnaires, e.g. personality, depression, stress etc.
- Explore the number of common factors in personality items.

EFA

Extraction methods:

- classical: Principal axis factoring.
- other methods: Maximum likelihood, image analysis, alpha analysis.

Factor loading:

- item-factor relationship.
- values > 0.3.

EFA

To simplify EFA results, need factor rotation to obtain clear factors and factor loadings.

Types of rotation:

- Orthogonal – uncorrelated factors.
  - Varimax, Quartimax, Equamax.
- Oblique – correlated factors.
  - Oblimin, Promax.

EFA - preliminaries

Suitability of data for the analysis:

- there must be correlations between the variables.

Judged by:

- Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy (MSA). KMO > 0.7 required.
- Bartlet’s test of sphericity. P-value < 0.05 indicates presence of correlations.

EFA - preliminaries

Number of factors to extract:

- Factors with Eigenvalues > 1 (Kaiser’s rule).
- Cattell’s scree test.
- Parallel analysis.

EFA - Practical [20 minutes]

efa.R

1. Data exploration.
2. EFA.
3. Reliability Cronbach's alpha.

```r
# libraries
library(foreign)  # for importing SPSS data
library(psych)  # for psychometrics

# descriptive
describe(data)
response.frequencies(data)
mardia(data)

# preliminaries
KMO(data)
cortest.bartlett(data)
scree = scree(data); print(scree)
parallel = fa.parallel(data, fa = both); print(parallel)
```

```r
# run efa
fa = fa(data, nfactors = k, fm = "pa", rotate = "oblimin"); print(fa)

# reliability
alpha = alpha(data[FACTOR]); print(alpha)
```

Confirmatory Factor Analysis (CFA)

CFA

Structural equation modeling (SEM):

- measurement model – CFA.
- structural model – path analysis.

CFA

- Confirmatory method – e.g. at final stage of questionnaire development.
- Theory confirmation.
- Based on common factor model – similar to EFA.
- Accounts for measurement errors.
- Analysis done on variance-covariance matrix.
- Allows assessment of model fit.

CFA vs EFA

CFA items:

- I love fast food
- I hate vegetable
- I hate eating fruits
- I hate exercise

→ Obesity

CFA vs EFA

EFA items:

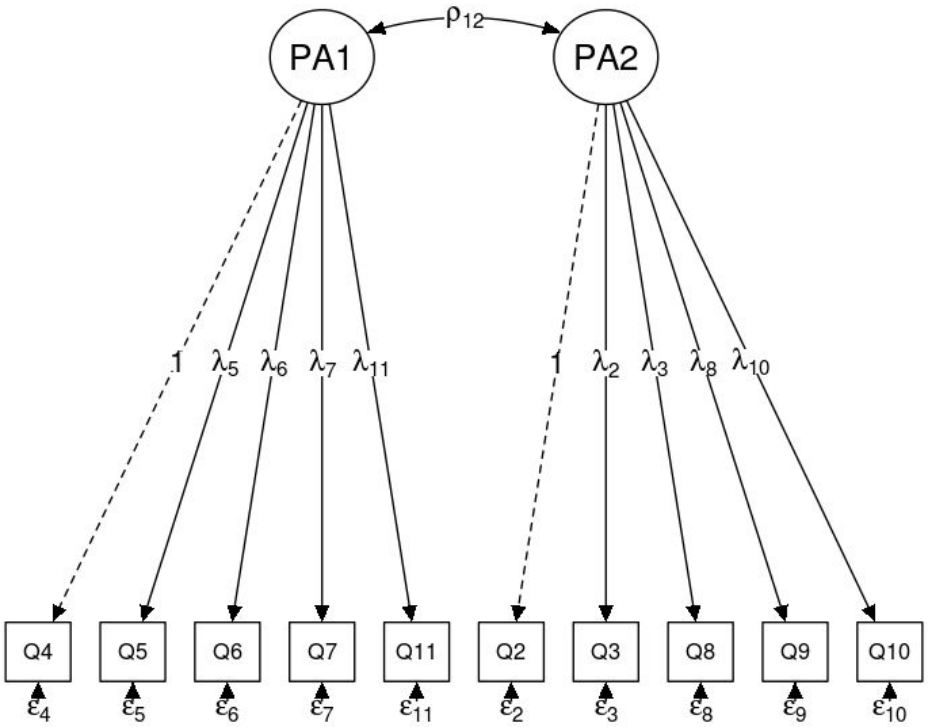
- I love cat
- I love statistics
- I love snorkelling
- I love driving car
- I love computer game
- I like to have everything normally distributed
- I love nasi ayam
- I eat a lot of pisang goreng
- I spend most of my time in front of computer

→ What factors???

CFA vs EFA

EFA	**CFA**
Exploratory.	Confirmatory.
Not necessary to specify factors.	Pre-specified factors.
Theory generating.	Theory confirmation.
Items not fixed to factors.	Items fixed to factors.
Model fit not tested.	Model fit assessment.

CFA path diagram



CFA

Research question:

- Does our theoretical measurement model (factor + item + correlation) fit the data?
- Strength of relationship between items and the factors?
- Factor correlations?

⟶ confirm the theory.

CFA

Applications:

- confirm the measurement model of newly developed questionnaires in target populations.
- confirm the measurement model of existing questionnaires in new populations.
- confirm the measurement model translated questionnaires in new populations.

⟶ confirmatory role of the analysis.

CFA

Estimations methods:

- numerical scale + multivariate normal data ⟶ maximum likelihood.
- numerical scale + non-multivariate normal data ⟶ robust maximum likelihood.
- categorical scale ⟶ weighted least squares.

Many more . . . http://lavaan.ugent.be/tutorial/est.html

CFA

Results to focus on:

1. Overall model fit – by fit indices.
2. Parameter estimates
   a. Factor loadings.
   b. Factor correlations.
3. Localized areas of misfit - Modification indices.

1. Fit indices

Category	Fit index	Cut-off
Absolute fit	Chi-square	P>0.05
	Standardized root mean square (SRMR)	≤ 0.08
Parsimony correction	Root mean square error of approximation (RMSEA)	RMSEA (90% CI) ≤ 0.08, CFit P>0.05
Comparative fit	Comparative fit index (CFI)	≥ 0.95
	Tucker-Lewis index (TLI)	≥ 0.95

2.a Factor loadings (FLs)

- The guideline for EFA is applicable also to CFA (FL > 0.3).
- In addition, the P-values of the FLs must be significant (at $\alpha = 0.05$).
- Also look for out-of-range values – FLs should be in range of 0 to 1 (absolute values).

2.b Factor correlations

- Similar to EFA, a factor correlation must be < 0.85, which indicates that the factors are distinct.
- A correlation > 0.85 indicates multicollinearity problem.
- Also look for out-of-range values – factor correlations should be in range of 0 to 1 (absolute values).

Modification indices (MIs)

- MI indicates the expected parameter change if we include a particular specification in the model.
- e.g. by correlating between errors of Q1 and Q2.
- MIs > 3.84 should be investigated.

Model revision

To improve model fit.

Causes of poor fit:

- Item – low FL, wrong factor.
- Factor – multicollinearity.
- Correlated error – items with similar wording/meaning.
- Model not supported by data.

Model-to-model comparison

To choose the best model.

Compare models:

- AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion).
- Better model = Smaller AIC/BIC.

CFA - Practical [20 minutes]

cfa.R

1. Data exploration.
2. CFA.
3. Path diagram.
4. Reliability - Raykov's rho.

```r
# libraries
library(foreign)  # for importing SPSS data
library(psych)  # for psychometrics
library(lavaan)  # for CFA
library(semTools)  # for additional functions in SEM
library(semPlot)  # for path diagram
```

```r
# lavaan model specification
model = "
FACTOR1 =~ Q1 + Q2 + Q3
FACTOR2 =~ Q4 + Q5 + Q6
Q1 ~~ Q2
"

# fit cfa model
cfa.model = cfa(model, data, estimator = "MLR")
summary(cfa.model, fit.measures = T, standardized = T)
# modification indices
mi = modificationIndices(cfa.model); subset(mi, mi > 3.84)
# model comparison
anova(cfa.model, cfa.model1)
```

```r
# path diagram
semPaths(cfa.model, 'path', 'std', style = 'lisrel',
         edge.color = 'black', intercepts = F)

# reliability
reliability(cfa.model)
```

Thank You

Q&A Session