Explorative factor analysis

Advanced course in questionnaire technique and clinimetrics

Eva Ørnbøl, Statistician



1

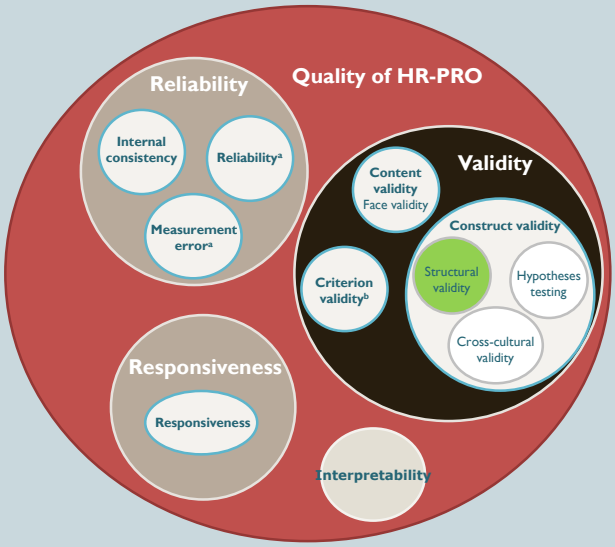
Outline

- The setting: COSMIN and structural validity
- EFA versus CFA
- The idea behind EFA
- The factor analysis model
- Considerations before EFA
- Key steps in EFA
  - Selecting the number of factors
  - Rotation
  - Interpreting factor loadings

2

AARHUS UNIVERSITY

The COSMIN taxonomy



3

AARHUS UNIVERSITY

Structural validity and factor analysis

Structural validity: *"The degree to which the scores of a measurement instrument are an adequate reflection of the dimensionality of the construct to be measured"*

Exploratory Factor Analysis (EFA)

Used when:

- You have no idea of the number of factors involved
- You have no idea of which items belong to which factor

Confirmatory Factor Analysis (CFA)

Used when:

- You know the number factors, based on theory or prior research
- You know which items belong to which factor, again based on theory or prior research

4

AARHUS UNIVERSITY

The idea behind EFA

- EFA starts from the correlation (or covariance) matrix of the items
- Items that measure the same construct should correlate highly with each other
- Items that measure different constructs should correlate less
- So we look for patterns in the correlation matrix: groups of items with
  - high correlations between items within the same group

Try it: can you find groups of items in the correlation matrix on the next slide?

5

AARHUS UNIVERSITY

Correlations – example 1

item	1	2	3	4	5	6
1	1					
2	0.84	1				
3	0.23	0.39	1			
4	0.78	0.69	0.29	1		
5	0.65	0.72	0.32	0.81	1	
6	0.37	0.31	0.86	0.36	0.27	1

6

AARHUS UNIVERSITY

Correlations – example 1

item	1	2	3	4	5	6
1	1					
2	**0.84**	1				
3	0.23	0.39	1			
4	**0.78**	**0.69**	0.29	1		
5	**0.65**	**0.72**	0.32	**0.81**	1	
6	0.37	0.31	**0.86**	0.36	0.27	1

Group 1: items 1, 2, 4, 5

Group 2: items 3, 6

7

AARHUS UNIVERSITY

EFA – purpose

"..achieve parsimony by explaining the maximum amount of common variance in a correlation matrix using the smallest number of *explanatory constructs*"

"These *explanatory constructs* are known as latent variables (or factors) and they represent clusters of variables that correlate highly with each other"

(Field, page 780)

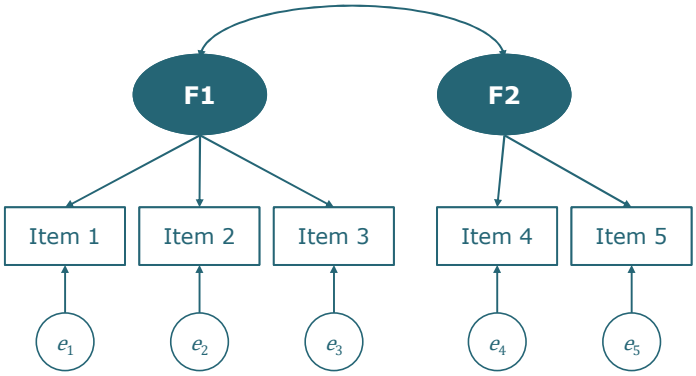
What does that mean: the common variance between measured items is caused by them all being associated to a common latent variable.

The groups we found in the correlation matrix are the intuition – the factors are the model. An item can be related to more than one factor.

8

AARHUS UNIVERSITY

Path diagram – a simple example



9

AARHUS UNIVERSITY

Factor Analysis model – simplified!

$$x_i = \mu_i + \lambda_{i1}F_1 + \lambda_{i2}F_2 + \cdots + \lambda_{iM}F_M + e_i$$

In words: item = mean + a contribution from each common factor + a unique part

- x_i = item i (i = 1, …, p items)
- μ_i = mean of item i
- F_m = factor m (m = 1, …, M – fewer factors than items)
- λ_{im} = factor loading: how strongly item i depends on factor m
- e_i = unique part of item i: measurement error and item-specific variance

The path diagram on the previous slide is this model: arrows from F = loadings, circles e = unique parts.

10

AARHUS UNIVERSITY

Some considerations before EFA

Factorability
Each item should correlate > 0.30 with several other items – but not > 0.8 (overlapping items)

KMO (Kaiser-Meyer-Olkin)
KMO ≥ 0.6: the correlations are suited for EFA
Stata: `estat kmo` R: `psych::KMO()`

Sample size (COSMIN)
≥ 7 × number of items and ≥ 100 (≥ 5 × number of items is adequate)
The exercise: 482 respondents, 19 items

Response scale
< 5 response categories: consider polychoric correlations instead of Pearson
Stata: `polychoric` (user-written)
R: `psych::polychoric()`

11

AARHUS UNIVERSITY

Select number of factors M

- We could extract one factor per item – but that would not reduce complexity
- We want the smallest number of factors that explains most of the common variance between the items
- **Eigenvalue** = how much variance a factor explains, measured in ”items” – an eigenvalue of 1 equals the variance of one item
- **Kaiser’s criterion:** keep factors with eigenvalue > 1, i.e. factors that explain more than one item
- Problems with Kaiser’s criterion:
  - often keeps too many factors when there are many items
  - the cut-off is arbitrary (1.01 vs. 0.99)
  - only makes sense for PCA eigenvalues – not after principal factor extraction

12

AARHUS UNIVERSITY

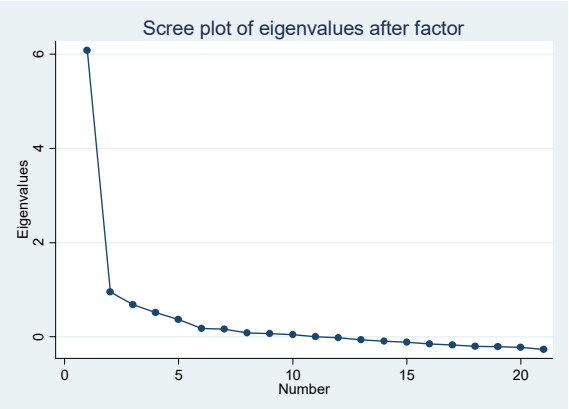
Select number of factors M – scree plot

- Plot the eigenvalue of each factor against the factor number
- The graph declines steeply at first, then levels off and declines slowly (almost linearly)
- Point of inflection (the elbow): the factor where the graph levels off
- Cattell's idea: extract the factors BEFORE the point of inflection
- Where is the elbow? Often a matter of judgement – look at the two examples

13

AARHUS UNIVERSITY

Scree plot – example 1



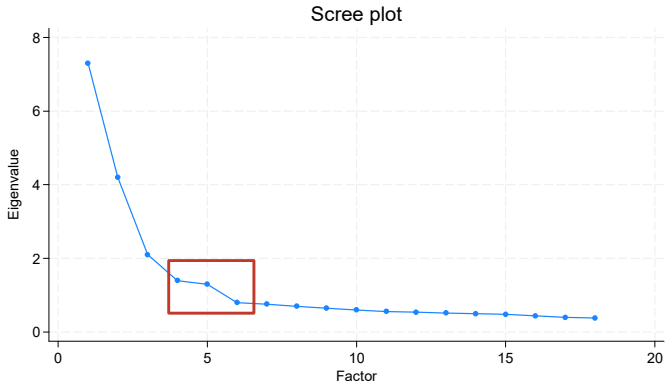
Clear elbow at factor 2 → keep 1 factor

Negative eigenvalues: from principal factor extraction – eigenvalue > 1 does not apply here

14

AARHUS UNIVERSITY

Scree plot – example 2



Where is the elbow – 3 or 5 factors?

15

AARHUS UNIVERSITY

Select number of factors M – other methods

- **Parallel analysis** (recommended): compare the eigenvalues with eigenvalues from random data of the same size – keep the factors whose eigenvalue is larger than from random data

  Stata: `paran` (ssc install paran) R: `psych::fa.parallel()`

- **Maximum likelihood estimation:**
  - likelihood-ratio (LR) test: the smallest number of factors that gives a non-significant test*
  - information criteria AIC or BIC: the number of factors with the lowest value

    Stata: `factor, ml` and `estat factors` R: `factanal()` (LR test), `psych::fa(fm = "ml")` (BIC)

- **The final decision** also rests on theory and interpretability: do the factors make sense?

* With large samples the LR test is almost always significant and points to too many factors – use it together with the other methods

16

AARHUS UNIVERSITY

Rotation

- The factor solution is not unique: the factors can be rotated without changing how well the model fits the data – the first (unrotated) solution is often hard to interpret
- Sticking to the graphical understanding of factors, consider factors as axes in a coordinate system.
- These axes can be rotated to better fit items coordinates – the idea is to make each individual item more related to one axis than to all others (i.e., where the coordinate of the item is close to 1 on one axis and close to 0 on all other axes).

AARHUS UNIVERSITY

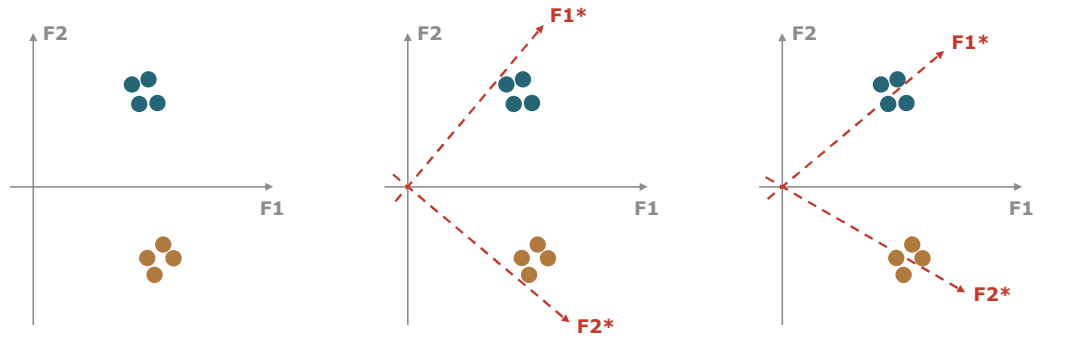
Rotation

- Orthogonal rotation (e.g. varimax): axes are rotated in such a way that they stay orthogonal/uncorrelated, and at the same time overlap clusters of items as much as possible.
- Oblique rotation (e.g. promax): axes are allowed to be correlated (i.e., not being orthogonal), and can for example be squeezed together to better overlap clusters of items.

AARHUS UNIVERSITY

Rotation – a simple example

● items in group A ● items in group B – – – rotated axes



Unrotated

Orthogonal rotation (e.g. varimax): axes stay at 90°

Oblique rotation (e.g. promax): the angle between axes may be < 90°

19

AARHUS UNIVERSITY

Deciding on rotation

- Factors are rarely uncorrelated in real research → start with an oblique rotation (e.g. promax)
- Look at the factor correlations: if small (< about 0.3), an orthogonal rotation (varimax) gives almost the same result
- Interpret the loadings from the rotated solution (for oblique rotation: the pattern matrix)

Stata: `rotate, promax` / `rotate, varimax` R: `psych::fa(rotate = "promax")` / `"varimax"`

20

AARHUS UNIVERSITY

Factor loadings

- **Loading** = how strongly an item is related to a factor (orthogonal rotation: the correlation between item and factor, between −1 and 1)
- **Squared loading** = the proportion of the item's variance explained by the factor: 0.4 → 16%, 0.5 → 25%
- **Uniqueness** (shown in the output) = the proportion of the item's variance not explained by the common factors; communality = 1 − uniqueness
- **Judge the size numerically** – the sign only gives the direction: −0.74 is as strong as 0.74

In the exercise

- loading > 0.5 (numerically): strong
- loading 0.3–0.5: moderate
- loading > 0.3 on more than one factor: cross-loading
- no loading > 0.3: the item does not fit any factor

With oblique rotation, loadings can occasionally exceed 1 – a sign of a problem in the model

AARHUS UNIVERSITY

Factor loadings – ideal world

	F1	F2	F3
item 1	1	0	0
item 2	1	0	0
item 3	0	1	0
item 4	1	0	0
item 5	0	1	0

Which items belong to which factor?

AARHUS UNIVERSITY

Factor loadings – ideal world

	F1	F2	F3
item 1	1	0	0
item 2	1	0	0
item 3	0	1	0
item 4	1	0	0
item 5	0	1	0

F1: items 1, 2, 4

F2: items 3, 5

F3: no items – redundant

Simple structure: each item loads on one factor only – real loadings are never exactly 1 and 0

A factor without any items is not needed – fewer factors will do

AARHUS UNIVERSITY

Factor loadings – example 1

Return to the factor model:

Clear: each item belongs to one factor

$x_i = 3.75 + 0.97 \cdot F_1 + 0.05 \cdot F_2 + e_i$

$x_j = 3.98 + 0.02 \cdot F_1 + 0.91 \cdot F_2 + e_j$

Unclear: both items load 0.36–0.53 on both factors – cross-loadings

$x_i = 3.7 + 0.48 \cdot F_1 + 0.53 \cdot F_2 + e_i$

$x_j = 2.6 + 0.36 \cdot F_1 + 0.49 \cdot F_2 + e_j$

AARHUS UNIVERSITY

Exercise EFA

- Materials:
  - The wording of the exercise
  - Appendix A: eigenvalues and varimax rotated loadings (4, 3 and 2 factors)
  - Appendix B: promax rotated loadings and correlations between factors (4, 3 and 2 factors)
  - The original questionnaire with the items in Danish
  - The English wording of the items

25

AARHUS UNIVERSITY

Take home message from exercise

- Given the same data different researchers will likely interpret the results of an EFA differently. This often originates in:
  - different view on research purpose
  - different background knowledge
- When you read an article about an EFA, it is inevitable that more models, than the one presented, have been considered.

26

AARHUS UNIVERSITY

Extra: more theory of EFA

Assumptions behind the model

- $F_m \sim N(0, 1)$ – standardised factors
- factors and error terms are uncorrelated
- error terms are mutually independent, $e_i \sim N(0, \Psi_i)$

This gives the covariance matrix

$\Sigma = \Lambda\Lambda' + \Psi$

Σ = covariance matrix

Λ = matrix of factor loadings

Ψ = diagonal matrix of unique variances

Principal factor / PCA solution

$\Lambda\Lambda' = (L\sqrt{V})(L\sqrt{V})'$

L = matrix with the eigenvectors in the columns

V = diagonal matrix with the eigenvalues in the diagonal

Rotation

$\Lambda\Lambda' + \Psi = \Lambda^*\Lambda^{*\prime} + \Psi$

$\Lambda = \Lambda^* G$, G = orthogonal matrix

Rotated loadings reproduce the same covariance matrix – the fit is unchanged